ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

"ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ ΓΙΑ ΕΠΙΚΟΙΝΩΝΙΕΣ ΚΑΙ ΠΟΛΥΜΕΣΑ"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μελέτη και Αξιολόγηση Τεχνικών Συμπίεσης Ολοκληρωτικών Εικόνων

Αντώνης Μ. Χαλάς

Επιβλέποντες: **Εμμανουήλ Σαγκριώτης,** Αναπληρωτής Καθηγητής ΕΚΠΑ
Νικόλαος Σγούρος, Διδάκτωρ ΕΚΠΑ

ΑΘΗΝΑ

ΙΑΝΟΥΑΡΙΟΣ 2012

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μελέτη και Αξιολόγηση Τεχνικών Συμπίεσης Ολοκληρωτικών Εικόνων

Αντώνης Μ. Χαλάς

Α.Μ.: Μ1083

ΕΠΙΒΛΕΠΟΝΤΕΣ: **Εμμανουήλ Σαγκριώτης,** Αναπληρωτής Καθηγητής ΕΚΠΑ
Νικόλαος Σγούρος, Διδάκτωρ ΕΚΠΑ

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Εμμανουήλ Σαγκριώτης,** Αναπληρωτής Καθηγητής ΕΚΠΑ
Νικόλαος Σγούρος, Διδάκτωρ ΕΚΠΑ

Ιανουάριος 2012

ΠΕΡΙΛΗΨΗ

Στην εποχή της πληροφορίας, την οποία διανύουμε, οι πολυμεσικές εφαρμογές εξελίσσονται με την πάροδο του χρόνου σε ολοένα και πιο απαιτητικές. Τη μεγαλύτερη πρόκληση αποτελεί η λήψη, η κωδικοποίηση και η ανακατασκευή τρισδιάστατων στατικών και κινούμενων εικόνων. Στην παρούσα διπλωματική εργασία θα πραγματοποιηθεί αναφορά στη μέθοδο ολοκληρωτικής φωτογράφισης και θα μελετηθούν σε βάθος όλες οι τεχνικές συμπίεσης των ολοκληρωτικών εικόνων. Πρόκειται για μία από τις δημοφιλέστερες μεθόδους τρισδιάστατης καταγραφής και απεικόνισης. Χαρακτηριστικό των ολοκληρωτικών φωτογραφιών αποτελεί η υψηλή συσχέτιση η οποία παρουσιάζεται εντός των εικόνων, εξαιτίας του τρόπου λήψεως τους, γεγονός το οποίο μπορεί να εκμεταλλευτεί ο κωδικοποιητής για να επιτύχει αποδοτικότερη συμπίεση. Το μεγαλύτερο μέρος της εργασίας αφιερώνεται στην εκτενή ανάλυση των βασικών αρχών και μεθόδων συμπίεσης εικόνας προκειμένου ο αναγνώστης να κατανοήσει τον τρόπο με τον οποίο επιτυγχάνεται η συμπίεση των δεδομένων με και χωρίς απώλειες. Στα τελευταία κεφάλαια θα συζητηθούν εξειδικευμένες μέθοδοι συμπίεσης ολοκληρωτικών εικόνων και θα πραγματοποιηθούν σχετικές μετρήσεις προκειμένου να εξαχθούν συμπεράσματα και να αναδειχθούν οι αποδοτικότεροι κωδικοποιητές.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία Εικόνας

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Τεχνικές Συμπίεσης Εικόνας, Διακριτός Μετασχηματισμός Συνημίτονου, H.264, Διακριτός Μετασχηματισμός Κυματιδίου, Ολοκληρωτικές Εικόνες.

ABSTRACT

In the information age, that we are going through, multimedia applications becoming more demanding with the passage of time. The capturing, coding and reconstruction of three-dimensional static and moving pictures constitute some of the greatest challenges. In this thesis we discus one of the most popular three-dimensional techniques, known as integral photography, providing an in depth analysis of integral imaging compression methods. The main characteristic in this type of pictures appears to be the high correlation which observed in neighboring regions of the image due to the nature of the capturing process. The encoder can exploit this correlation in order to achieve more efficient compression. The biggest part of this thesis is dedicated to fundamental principles of data, image and video compression methods to give the reader a deep understanding of lossy and lossless compression theory. In the last chapters are discussed integral imaging-specific compression methods, performing experiments with the most popular image and video encoders, in order to conclude about the most efficient compression method.

SUBJECT AREA: Image Processing

KEYWORDS: Image Compression Methods, DCT, H.264, DWT, Integral Imaging

Η διπλωματική μου εργασία είναι αφιερωμένη στην οικογένεια μου

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τους καθηγητές μου Εμμανουήλ Σαγκριώτη και Ιωάννη Έλληνα για την εμπιστοσύνη που έδειξαν στο πρόσωπο μου καθώς και για τις γνώσεις που μου προσέφεραν κατά τη διάρκεια των μεταπτυχιακών και προπτυχιακών σπουδών μου αντίστοιχα, τόσο στη συμπίεση δεδομένων όσο και στην ανάλυση εικόνας. Επίσης θα ήθελα να ευχαριστήσω τον καθηγητή, συνάδελφο και φίλο μου, Νικόλαο Σγούρο καθώς είναι ο άνθρωπος ο οποίος με ενέπνευσε να ακολουθήσω, αλλά και με βοήθησε να σταθώ, στο δρόμο της επεξεργασίας σήματος και ανάλυσης εικόνας που τελικά τόσο αγάπησα. Τέλος ευχαριστώ την οικογένεια μου, η οποία βρίσκεται πάντα στο πλευρό μου και με στηρίζει με κάθε τρόπο, καθώς επίσης και τους φίλους που είναι πάντα εδώ, στα εύκολα και στα δύσκολα.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ **33**

1. ΕΙΣΑΓΩΓΗ **35**

1.1 Ψηφιακή Φωτογραφία **35**

1.2 Τεχνική Ολοκληρωτικής Φωτογράφισης **36**

1.2.1 Καταγραφή 39

1.2.2 Ανακατασκευή 41

2. ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΠΛΗΡΟΦΟΡΙΑΣ **45**

2.1 Πληροφορία **45**

2.1.1 Εντροπία 45

2.1.2 Ποσότητα Πληροφορίας 45

2.1.3 Μέση Πληροφορία ανά Σύμβολο 45

2.2 Θεώρημα Κωδικοποίησης του Shannon για Πηγή Χωρίς Θόρυβο **46**

2.3 Θεώρημα Κωδικοποίησης του Shannon σε Ενθόρυβο Κανάλι **47**

2.4 Θεώρημα Κωδικοποίησης Πηγής του Shannon **48**

2.5 Θεώρημα Μετάδοσης της Πληροφορίας **49**

3. ΒΑΣΙΚΕΣ ΤΕΧΝΙΚΕΣ ΣΥΜΠΙΕΣΗΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ **51**

3.1 Κωδικοποίηση Τρέχοντος Μήκους (Run Length Encoding) **51**

3.2 Κώδικες Huffman **52**

3.2.1 Κωδικοποίηση 52

3.2.2 Περιορισμοί 53

3.2.3 Αποκωδικοποίηση 54

3.3 Αριθμητικοί Κώδικες **55**

3.3.1 Κωδικοποίηση 55

3.3.2 Αποκωδικοποίηση 57

4. ΚΒΑΝΤΙΣΗ **59**

4.1 Η έννοια της Κβάντισης 59

4.2 Βαθμωτή Κβάντιση 59

4.2.1 Ομοιόμορφη Κβάντιση 59

4.2.2 Παραμόρφωση Κβάντισης 62

4.2.3 Σχεδίαση Καβντιστή 64

4.2.4 Ομοιόμορφος Κβαντιστής με Ομοιόμορφα Κατανεμημένη Είσοδο 65

4.2.5 Συνθήκες του Βέλτιστου κβαντιστή 67

4.2.6 Βέλτιστος Ομοιόμορφης Κβαντιστής με Διαφορετικές Κατανομές Εισόδου 68

4.2.7 Βέλτιστος μη-ομοιόμορφος κβαντιστής 69

4.2.8 Κβάντιση με Διεύρυνση του Δυναμικού Εύρος (Companding) 71

4.2.9 Βασικές Εφαρμογές του Βαθμωτού Κβαντιστή στις Εικόνες 75

4.3 Διανυσματική Κβάντιση 76

4.3.1 Αρχικές Προσεγγίσεις Διανυσματικής Κβάντισης 76

4.3.2 Ο Αλγόριθμος LBG 79

4.3.3 Αλγόριθμος Επιλογής Αρχικών Εγγραφών του Κωδικού-Βιβλίου 83

4.3.4 Δενδροειδής Μορφή VQ 85

4.3.5 Προσαρμοστική Διανυσματική Κβάντιση 86

4.3.6 Παράδειγμα 92

5. ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΕΙΚΟΝΑΣ 95

5.1 Η Βασική Ιδέα των Μετασχηματισμών 95

5.2 Ορθογώνιοι Μετασχηματισμοί 102

5.3 Δισδιάστατοι Ορθογώνιοι Μετασχηματισμοί 107

5.3.1 Walsh-Hadamard Transform 108

5.3.2 Μετασχηματισμός Haar 110

5.3.3 Μετασχηματισμός Karhunen-Loeve 112

5.4 Ο Διακριτός Μετασχηματισμός Συνημίτονου 114

5.4.1 Ο Μετασχηματισμός DCT ως βάση 127

5.4.2 Ο DCT ως Μετασχηματισμός Περιστροφής 134

5.4.3 Οι τέσσερεις τύποι DCT 137

6. ΤΟ ΠΡΟΤΥΠΟ ΣΥΜΠΙΕΣΗΣ JPEG 141

6.1 Βασικά Βήματα JPEG 142

6.1.1 Ο Μετασχηματισμός DCT στο JPEG 143

6.1.2 Κβάντιση 145

6.1.3 Κωδικοποίηση της DC Συνιστώσας και Σάρωση zig-zag 147
6.1.4 Κωδικοποίηση Εντροπίας 148

6.2 Συμπίεση και Ποιότητα Εικόνας 149

6.3 Βήματα Επεξεργασίας στη Συμπίεση με Πρόγνωση για Συμπίεση χωρίς Απώλειες 149

6.4 Στάδια Κωδικοποίησης Εντροπίας 151
6.4.1 Ενδιάμεση Αναπαράσταση Συμβόλων 151
6.4.2 Κωδικοποίηση Εντροπίας Μεταβλητού Μήκους 153

6.5 Παράδειγμα στον Βασικό Ακολουθιακό Κωδικοποιητή 154

7. ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΚΥΜΑΤΙΔΙΟΥ 159

7.1 Ο Συνεχής Μετασχηματισμός Κυματιδίου και ο Αντίστροφος του 159

7.2 Ο Μετασχηματισμός Κυματιδίου Haar 166

7.3 Εφαρμογή του Μετασχηματισμού Κυματιδίου Haar 172
7.3.1 Ο Δισδιάστατος Μετασχηματισμός Haar στη Συμπίεση Εικόνας 172
7.3.2 Προσέγγιση του Μετασχηματισμού Haar με τη Χρήση Πινάκων 176

7.4 Τράπεζες Φίλτρων 180
7.4.1 Η Έννοια και ο Ρόλος των Φίλτρων στο Μετασχηματισμό Υποζώνης 180
7.4.2 Φίλτρα στο Πεδίο της Συχνότητας 184
7.4.3 Εύρεση των συντελεστών των φίλτρων 190

7.5 Ο Διακριτός Μετασχηματισμός Κυματιδίου 192

7.6 Διάφοροι τύποι Αποδόμησης Εικόνας 195

7.7 Ο Ακέραιος Μετασχηματισμός Κυματιδίου 201

8. ΤΟ ΠΡΟΤΥΠΟ ΣΥΜΠΙΕΣΗΣ JPEG2000 205

8.1 Στοιχεία Χρώματος 206

8.2 Πλακίδια 207

8.3 Μετασχηματισμός Κυματιδίου 208

8.4 Κβάντιση 210

8.5 Τομείς και Κωδικά-Μπλοκ 211

8.6 Κωδικοποίηση Εντροπίας 212

8.7 Πακέτα 217

8.8 Επίπεδα 218

8.9 Προοδευτική Μετάδοση 218

8.10 Περιοχές Ενδιαφέροντος (ROIs) 219

9. ΜΕΘΟΔΟΙ ΣΥΜΠΙΕΣΗΣ ΤΡΙΣΔΙΑΣΤΑΤΩΝ ΕΙΚΟΝΩΝ 221

9.1 Συμπίεση Ολοκληρωτικής Εικόνας ως Βίντεο-Ροή με Χρήση του Προτύπου H.264 222

9.1.1 Βασικές Αρχές Συμπίεσης Βίντεο 222

9.1.2 Τεμαχισμός Εικόνας σε Μακρο-Μπλοκ 224

9.1.3 Εκτίμηση Κίνησης 224

9.1.4 Αντιστάθμιση Κίνησης 225

9.1.5 Ανάπτυξη Μεθόδου 226

9.1.6 Κωδικοποίηση Ολοκληρωτικής Εικόνας ως Καρέ Τύπου I (Intra) 228

9.2 Υβριδική Μέθοδος Συμπίεσης με τη Χρήση των Μετασχηματισμών Κυματιδίου και DCT 231

9.2.1 Τεμαχισμός Ολοκληρωτικής Εικόνας 231

9.2.2 Εφαρμογή του Μετασχηματισμού Κυματιδίου 232

9.2.3 Διαχωρισμός Πακέτων 233

9.2.4 Εφαρμογή του Μετασχηματισμού DCT 234

9.2.5 Ολοκλήρωση της Μεθόδου 234

9.3 Μέθοδος Προ-Επεξεργασίας Ολοκληρωτικών Εικόνων 235

9.3.1 Εισαγωγή στη Μέθοδο 235

9.3.2 Ανάπτυξη της Μεθόδου 235

10. ΜΕΤΡΗΣΕΙΣ ΑΠΟΔΟΣΗΣ ΚΩΔΙΚΟΠΟΙΗΤΩΝ 239

10.1 Συμπίεση Τεχνιτών Ολοκληρωτικών Εικόνων Δημιουργημένων με Προγράμματα Προσομοίωσης 239

10.1.1 Συμπίεση της Εικόνας Ferrari 239

10.1.2 Συμπίεση της Εικόνας Cube 245

10.1.3 Συμπίεση της Εικόνας Cat 250

10.2 Συμπίεση Πραγματικών Ολοκληρωτικών Εικόνων 254

10.2.1 Συμπίεση της Εικόνας Birds 255

10.2.2 Συμπίεση της Εικόνας Seagull 256

10.2.3 Συμπίεση της Εικόνας Sports 257

11. ΣΥΜΠΕΡΑΣΜΑΤΑ 259

11.1 Μετρικές και Στατιστικά Μεγέθη 260

11.1.1 Τυπική Απόκλιση 260

11.1.2 Ετεροσυσχέτιση (Cross-Correlation) 260

11.1.3 Αμοιβαία Πληροφορία 261

11.1.4 Μέση Απόλυτη Διαφορά εντός της Στοιχειώδους Εικόνας (IntraMAD) 261

11.1.5 Μέση Απόλυτη Διαφορά μεταξύ των Στοιχειωδών Εικόνων (InterMAD) 262

11.2 Γραφική Αναπαράσταση των Μέτρων και των Στατιστικών Μεγεθών των Τεχνητά Κατασκευασμένων Ολοκληρωτικών Εικόνων 263

11.2.1 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Ferrari 263

11.2.2 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Cube 267

11.2.3 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Cat 271

11.2.4 Συμπεράσματα που Απορρέουν από τις Μετρήσεις 277

11.3 Γραφική Αναπαράσταση των Μέτρων και των Στατιστικών Μεγεθών των Φυσικών Ολοκληρωτικών Φωτογραφιών 278

11.3.1 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Birds 278

11.3.2 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Seagull 280

11.3.3 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Sport 282

11.4 Προτεινόμενη Μέθοδος Βάσει των Συμπερασμάτων 285

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ 291

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ 293

ΑΝΑΦΟΡΕΣ 295

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1.1 : Καταγραφή Εικόνας με τη Μέθοδο της Ολοκληρωτικής Φωτογράφισης 37

Σχήμα 1.2 : Ανακατασκευή της Τρισδιάστατης Εικόνας.....38

Σχήμα 1.3 : Διαδικασία Καταγραφής με Χρήση (a) Μικρών Οπών και (b) Μικρο-Φακών 39

Σχήμα 1.4 : Η Απεικόνιση ενός Σημείου του Αντικειμένου Τοποθετημένο σε Διάφορες Αποστάσεις από τη Συστοιχία Μικρο-Φακών. (a) $\boldsymbol{l = 100mm}$ (b) $\boldsymbol{l = 143mm}$ (c) $\boldsymbol{l = 200mm}$ (d) $\boldsymbol{l = 333mm}$ 41

Σχήμα 1.5 : Στοιχειώδεις Εικόνες Ληφθείσες από Διάφορους Μικρο-Φακούς της Συστοιχίας. Τα Αντικείμενα Βρίσκονται Τοποθετημένα στα $\boldsymbol{100mm, 143mm, 200mm}$ και $\boldsymbol{333mm}$ αντίστοιχα. (a) Μικρο-Φακός $(\boldsymbol{0}, \boldsymbol{-2})$, (b) Κεντρικός Μικρο-Φακός (0,0), (c) Μικρο-Φακός $(\boldsymbol{0}, \boldsymbol{+2})$. 41

Σχήμα 1.6 : Διαδικασία Ανακατασκευής με Χρήση (a) Μικρών Οπών και (b) Μικρο-Φακών 42

Σχήμα 1.7 : Ανακατασκευασμένη Εικόνα η οποία Προέκυψε από Συστοιχία Μικρο-Φακών για Διαφορετικές Αποστάσεις Ανακατασκευής (a) $\boldsymbol{l = 100mm}$ (b) $\boldsymbol{l = 143mm}$ (c) $\boldsymbol{l = 200mm}$ (d) $\boldsymbol{l = 333mm}$.....43

Σχήμα 2.1 : Συνάρτηση Ρυθμού-Παραμόρφωσης 49

Σχήμα 3.1 : Δένδρο Huffman.....53

Σχήμα 3.2 : Γραφική Αναπαράσταση Αριθμητικής Κωδικοποίησης.....56

Σχήμα 4.1 : Χαρακτηριστική Εισόδου – Εξόδου Ομοιόμορφου Midtread Κβαντιστή 60

Σχήμα 4.2 : Χαρακτηριστική Εισόδου – Εξόδου Ομοιόμορφου midrise Κβαντιστή.....61

Σχήμα 4.3 : Θόρυβος κβάντισης του Ομοιόμορφου midtread Κβαντιστή.....63

Σχήμα 4.4 : Χαρακτηριστική Εισόδου – Εξόδου ενός Ομοιόμορφου midtread Κβαντιστή με Ομοιόμορφα Κατανεμημένη Είσοδο στο Διάστημα $\boldsymbol{-4.5, 4.5}$ με N=9.....65

Σχήμα 4.5 : Θόρυβος Κβάντισης του Κβαντιστή που Παρουσιάστηκε στο Σχήμα 4.4 ...65

Σχήμα 4.6 : Ο Λόγος Ανάμεσα στο Σφάλμα του Βέλτιστου Κβαντιστή και το Σφάλμα του Βέλτιστου Ομοιόμορφου Κβαντιστή προς το Πλήθος των Σταθμών Ανακατασκευής N.71

Σχήμα 4.7 : Τεχνική Κβάντισης Διεύρυνσης του Δυναμικού Εύρους 72

Σχήμα 4.8 : Οι Χαρακτηριστικές της Τεχνικής Κβάντισης Διεύρυνσης του Δυναμικού

Εύρους. 73

Σχήμα 4.9 : Χαρακτηριστική Συμπίεσης – Κανόνας μ.....74

Σχήμα 4.10 : Χαρακτηριστική Συμπίεσης – Κανόνας A 74

Σχήμα 4.11 : Δώδεκα Σημεία και Τέσσερεις Εγγραφές $\boldsymbol{Ci(0)}$ στο Κωδικό-Βιβλίο.....81

Σχήμα 4.12 : Δώδεκα Σημεία και Τέσσερεις Εγγραφές $\boldsymbol{Ci(1)}$ στο Κωδικό-Βιβλίο.....82

Σχήμα 4.13 : Αρίθμηση τεταρτημορίων.....85

Σχήμα 4.14 : (a): Κυματοειδής Κάλυψη. (b): Κυκλική Κάλυψη. (c): Διαγώνια κάλυψη ...89

Σχήμα 4.15 : Επάνω Αριστερό Τμήμα Εικόνας.....93

Σχήμα 4.16 : Εγγραφές με τις οποίες Ενημερώνεται το Λεξικό.....93

Σχήμα 5.1 : Ζεύγη των Τιμών Φωτεινοτήτων Γειτονικών Εικονοστοιχείων ως Συντεταγμένες ενός Δισδιάστατου Χώρου 96

Σχήμα 5.2 : Περιστροφή των σημείων κατά 45 μοίρες.....97

Σχήμα 5.3 : Κατανομή των Τιμών στους Άξονες x και y αντίστοιχα 98

Σχήμα 5.4 : Κατανομή των Τιμών στους Άξονες x και y αντίστοιχα μετά την στροφή των αξόνων 98

Σχήμα 5.5 : Ζεύγη των Τιμών Φωτεινοτήτων Γειτονικών Εικονοστοιχείων ως Συντεταγμένες ενός Τρισδιάστατου Χώρου 100

Σχήμα 5.6 : Περιστροφή των σημείων κατά 45 μοίρες.....101

Σχήμα 5.7 : Συχνότητες Εικόνας.....103

Σχήμα 5.8 : Ο Πυρήνας WHT για N=4 110

Σχήμα 5.9 : Η Εικόνα Βάσης του μετασχηματισμού Haar για $\boldsymbol{n = 8}$ 112

Σχήμα 5.10 : (a) Μονοδιάστατα Δεδομένα Εισόδου και (b) ο Αντίστροφος Μετασχηματισμός Fourier μετά την Κβάντιση.....119

Σχήμα 5.11 : Αυξανόμενες Συχνότητες.....130

Σχήμα 5.12 : Οι Γωνίες και οι Τιμές του Συνημίτονου για τον Μετασχηματισμό DCT 8-Σημείων 132

Σχήμα 5.13 : Γραφική Αναπαράσταση του Μονοδιάστατου DCT 133

Σχήμα 5.14 : Οι 64 Εικόνες Βάσης του Δισδιάστατου Μετασχηματισμού DCT.....134

Σχήμα 5.15 : Ο Δισδιάστατος Μετασχηματισμός DCT ως Διπλή Περιστροφή 137

Σχήμα 6.1 : Μπλοκ Διάγραμμα Κωδικοποιητή JPEG Βασισμένο στον Μετασχηματισμό DCT 142

Σχήμα 6.2 : Μπλοκ Διάγραμμα Αποκωδικοποιητή JPEG Βασισμένο στον Μετασχηματισμό DCT 143

Σχήμα 6.3 : Διαφορική Κωδικοποίηση των Συντελεστών DC 147

Σχήμα 6.4 : Ακολουθία Σάρωσης Zig-Zag 147

Σχήμα 6.5 : Μπλοκ Διάγραμμα Κωδικοποιητή JPEG Χωρίς Απώλειες.....150

Σχήμα 6.6 : Πρόγνωση του Δείγματος X από τα Γειτονικά Δείγματα A, B και C 150

Σχήμα 7.1 : (a) Το Κυματίδιο Τύπου Morlet. (b) Το Κυματίδιο Τύπου Mecixan Hat.....160

Σχήμα 7.2 : Ημιτονοειδής Κυματομορφή με Δύο Αντίγραφα του Κυματιδίου Mexican Hat 162

Σχήμα 7.3 : Παραχθέν Κύμα από τη Μετατόπιση του Κυματιδίου 163

Σχήμα 7.4 : Διαστολή Κυματιδίου 163

Σχήμα 7.5 : (a) Ο Συνεχής Μετασχηματισμός Κυματιδίου μίας Ημιτονοειδούς Κυματομορφής και (b) το Περίγραμμα του.....164

Σχήμα 7.6 : Ημιτονοειδής Κυματομορφή Αυξανόμενης Συχνότητας 165

Σχήμα 7.7 : (a) Ο Συνεχής Μετασχηματισμός Κυματιδίου της Ημιτονοειδούς Κυματομορφής Αυξανόμενης Συχνότητας και (b) το Περίγραμμα του 165

Σχήμα 7.8 : Κυματίδιο Haar 166

Σχήμα 7.9 : Μετατοπισμένα Αντίγραφα το Κυματιδίου Harr.....167

Σχήμα 7.10 : Κυματίδια Haar για $\boldsymbol{k = 0, 1, 2, 3}$ 168

Σχήμα 7.11 : Υπολογισμός του Κανονικοποιημένου Μετασχηματισμού Κυματιδίου Haar 171

Σχήμα 7.12 : Ανακατασκευή από τους Κανονικοποιημένους Συντελεστές του Μετασχηματισμού Κυματιδίου Haar.....171

Σχήμα 7.13 : Ο Τυπικός Μετασχηματισμός Κυματιδίου και η Αποδόμηση του.....172

Σχήμα 7.14 : Ο Μετασχηματισμός Κυματιδίου Πυραμίδας 173

Σχήμα 7.15 : Επίπεδα και Υποζώνες Κατά την Αποδόμηση του Κυματιδίου.....175

Σχήμα 7.16 : Παράδειγμα του Μετασχηματισμού Κυματιδίου Πυραμίδας.....176

Σχήμα 7.17 : Παράδειγμα Αρχικής Εικόνας 177

Σχήμα 7.18 : Φίλτρο δύο Καναλιών 181

Σχήμα 7.19 : Ορθογώνια Τράπεζα Φίλτρων με τέσσερεις Συντελεστές.....183

Σχήμα 7.20 : Ημιορθογώνιο Φίλτρο με Τέλεια Ανακατασκευή 183

Σχήμα 7.21 : Συνάρτηση Κλιμάκωσης και Κυματιδίου ως Όρια ενός Λογαριθμικού Δένδρου.....184

Σχήμα 7.22 : Τα πλάτη ενός Βαθυπερατού (a) και ενός Υψηπερατού Φίλτρου 189

Σχήμα 7.23 : Μια Γενική Τράπεζα Φίλτρων 190

Σχήμα 7.24 : Μέθοδος Αποδόμησης Γραμμής 196

Σχήμα 7.25 : Μέθοδος Αποδόμησης Quincunx 197

Σχήμα 7.26 : Μέθοδος Αποδόμησης Πυραμίδας 198

Σχήμα 7.27 : Τυπική Μέθοδος Αποδόμισης 199

Σχήμα 7.28 : Ομοιόμορφη Μέθοδος Αποδόμησης 200

Σχήμα 7.29 : Προσαρμοστική Μέθοδος Αποδόμησης Πακέτου Κυματιδίων.....201

Σχήμα 8.1 : Πλακίδια στο Πρότυπο JPEG 2000 207

Σχήμα 8.2 : Επέκταση μίας Γραμμής Εικονοστοιχείων.....208

Σχήμα 8.3 : Οργάνωση Υποζωνών στο JEPG 2000 210

Σχήμα 8.4 : Υποζώνες, Τομείς και Κωδικά-Μπλοκ.....211

Σχήμα 8.5 : Λωρίδες σε ένα Κωδικό-Μπλοκ.....212

Σχήμα 8.6 : Επίπεδα bit των Τεσσάρων Συντελεστών ενός Κωδικό-Μπλοκ 215

Σχήμα 8.7 : Οι 8 Γείτονες.....216

Σχήμα 9.1 : Σειρά Κωδικοποίησης των Καρέ.....223

Σχήμα 9.2 : Σειρά Προβολής των Καρέ 224

Σχήμα 9.3 : Διαδικασία Εκτίμησης Κίνησης 225

Σχήμα 9.4 : Τεμαχισμός Εικόνας σε Μκρο-Μπλοκ Διαφορετικών Μεγεθών και Αντιστάθμιση τους 226

Σχήμα 9.5 : Δημοφιλή Σχήματα Σάρωσης των Στοιχειωδών Εικόνων Ολοκληρωτικής

Εικόνας (a) Παράλληλο Οριζόντιο, (b) Παράλληλο Κάθετο, (c) Σπιράλ και (d) Hilbert 228

Σχήμα 9.6 : Σάρωση των DC Συνιστωσών των $\mathbf{4} \times \mathbf{4}$ Μπλοκ για Εφαρμογή του Μετασχηματισμού Hadamard 229

Σχήμα 9.7 : Σχήματα πρόγνωσης του κωδικοποιητή H.264 σε Μπλοκ Μεγέθους $\mathbf{4} \times \mathbf{4}$ 230

Σχήμα 9.8 : Σχήματα πρόγνωσης του κωδικοποιητή H.264 σε Μπλοκ Μεγέθους $\mathbf{16} \times \mathbf{16}$ 230

Σχήμα 9.9 : (a) Ολοκληρωτική Εικόνα και (b) Τμηματοποίηση της σε Στοιχειώδεις Εικόνες 231

Σχήμα 9.10 : (a) Ολοκληρωτική Εικόνα και (b) Τμηματοποίηση της σε Στοιχειώδεις Εικόνες στη Γενική Περίπτωση 232

Σχήμα 9.11 : Εφαρμογή Μετασχηματισμού Κυματιδίου σε κάθε Στοιχειώδη Εικόνα ... 233

Σχήμα 9.12 : Διαδικασία Διαχωρισμού Πακέτου 234

Σχήμα 9.13 : Ολοκληρωτική Εικόνα και η Τμηματοποίηση της σε Υπο-Εικόνες 236

Σχήμα 9.14 : Οθόνη Ολοκληρωτικής Απεικόνισης με Συστοιχία Μικρο-Φακών. Οι Ακτίνες οι οποίες Τέμνονται από τις Παράλληλες γραμμές του Ίδιου Τύπου Ανήκουν στην Ίδια Υπο-Εικόνα 237

Σχήμα 10.1 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα της Ferrari και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 240

Σχήμα 10.2 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Ferrari η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 240

Σχήμα 10.3 : (a) Η Αρχική Ολοκληρωτική Εικόνα της Ferrari και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI) 241

Σχήμα 10.4 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα της Ferrari 241

Σχήμα 10.5 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα της Ferrari 242

Σχήμα 10.6 : (a) Η Αρχική Ολοκληρωτική Εικόνα της Ferrari και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI) 242

Σχήμα 10.7 : Οι καλύτερες Επιδόσεις και στα Τρία Σχήματα Προ-Επεξεργασίας

Επιτυγχάνονται από τον H.264 όταν Κωδικοποιηθεί η Ολοκληρωτική Εικόνα ως Βίντεο-Ροή. Το Σχήμα Προ-Επεξεργασίας EI Αποδίδει Καλύτερα και με Σημαντική Διαφορά Έναντι των Υπολοίπων 243

Σχήμα 10.8 : $\mathbf{2} \times \mathbf{2}$ Στοιχειώδεις Εικόνες της (a) Αρχικής Ολοκληρωτικής Εικόνας και (b) της Ανασυσταμένης Εικόνας μετά την Κωδικοποίηση με τον Υβριδικό Κωδικοποιητή με Ρυθμό $\mathbf{0.15}\boldsymbol{bpp}$ στα 30.65db 244

Σχήμα 10.9 : Η Χειρότερη Επίδοση στο σχήμα EI (Jpeg) Αποδίδει Καλύτερα από την Καλύτερη Επίδοση των Σχημάτων Προ-Επεξεργασίας SI και RI 245

Σχήμα 10.10 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα Cube και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 245

Σχήμα 10.11 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Cube η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 246

Σχήμα 10.12 : Αποδοτικότητα Συμπίεσης στην Εικόνα με τη Σκηνή της Ferrari και του Κύβου 246

Σχήμα 10.13 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cube και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI) 247

Σχήμα 10.14 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα 247

Σχήμα 10.15 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cube και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI) 248

Σχήμα 10.16 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα 248

Σχήμα 10.17 : Σύγκριση των Τριών Καλύτερων Μεθόδων Συμπίεσης (H.264) σε Κάθε Σχήμα Προ-Επεξεργασίας 249

Σχήμα 10.18 : Στοιχειώδης Εικόνα Ολοκληρωτικής Φωτογραφίας που Απεικονίζει μία Γάτα 250

Σχήμα 10.19 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα Cat και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 250

Σχήμα 10.20 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Cat η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 251

Σχήμα 10.21 : Αποδοτικότητα Συμπίεσης της Εικόνας της Ferrari του Κύβου και της

Γάτας 251

Σχήμα 10.22 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cat και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI) 252

Σχήμα 10.23 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα Cat 252

Σχήμα 10.24 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cat και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI) 253

Σχήμα 10.25 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα 253

Σχήμα 10.26 : Οι καλύτερες Επιδόσεις για Κάθε Σχήμα Προ-Επεξεργασίας 254

Σχήμα 10.27 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Birds η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 255

Σχήμα 10.28 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Seagull η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 257

Σχήμα 10.29 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Sport η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες 258

Σχήμα 11.1 : Τα Γειτονικά Εικονοστοιχεία τα οποία Χρησιμοποιούνται για να Προσδιοριστεί η Συσχέτιση τους με το Εικονοστοιχείο (1,1) 262

Σχήμα 11.2 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Ferrari 263

Σχήμα 11.3 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG 264

Σχήμα 11.4 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Ferrari 265

Σχήμα 11.5 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Ferrari 265

Σχήμα 11.6 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Ferrari και (b) στις Επιμέρους Στοιχειώδεις Εικόνες.....266

Σχήμα 11.7 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube 267

Σχήμα 11.8 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube 268

Σχήμα 11.9 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG 269

Σχήμα 11.10 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Cube.....270

Σχήμα 11.11 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Cube και (b) στις Επιμέρους Στοιχειώδεις Εικόνες.....271

Σχήμα 11.12 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cat 271

Σχήμα 11.13 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube 272

Σχήμα 11.14 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Cat.....273

Σχήμα 11.15 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG 273

Σχήμα 11.16 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Cat και (b) στις Επιμέρους Στοιχειώδεις Εικόνες 274

Σχήμα 11.17 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων 276

Σχήμα 11.18 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων ανά Στοιχειώδη Εικόνα 276

Σχήμα 11.19 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Birds 278

Σχήμα 11.20 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Birds 278

Σχήμα 11.21 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Birds.....279

Σχήμα 11.22 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Birds και (b) στις Επιμέρους Στοιχειώδεις Εικόνες 279

Σχήμα 11.23 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Seagull.....280

Σχήμα 11.24 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Seagull.....280

Σχήμα 11.25 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Seagull 281

Σχήμα 11.26 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Seagull και (b) στις Επιμέρους Στοιχειώδεις Εικόνες 281

Σχήμα 11.27 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Sport 282

Σχήμα 11.28 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Sport 282

Σχήμα 11.29 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Sport 283

Σχήμα 11.30 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Sport και (b) στις Επιμέρους Στοιχειώδεις Εικόνες.....283

Σχήμα 11.31 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων 285

Σχήμα 11.32 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων ανά Στοιχειώδη Εικόνα 285

Σχήμα 11.33 : Η Ολοκληρωτική Εικόνα της Ferrari και η Γραφική Αναπαράσταση της Τυπικής Απόκλισης της Κάθε Στοιχειώδους Εικόνας με Μέση Τιμή Διασποράς $\mathbf{18.3}$.286

Σχήμα 11.34 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Ferrari σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Τυπικής Απόκλισης) 287

Σχήμα 11.35 : Η Ολοκληρωτική Εικόνα Cube και η Γραφική Αναπαράσταση της Τυπικής Απόκλισης της Κάθε Στοιχειώδους Εικόνας με Μέση Τιμή Διασποράς $\mathbf{10.7}$.....288

Σχήμα 11.36 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Cube σε Σχέση με

τον Προτεινόμενο Κωδικοποιητή (Μέτρο Τυπικής Απόκλισης) 288

Σχήμα 11.37 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Ferrari σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Intra-MAD) 289

Σχήμα 11.38 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Cube σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Intra-MAD) 290

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1.1 : Τα εικονοστοιχεία τα οποία βρίσκονται σε μία μικρή περιοχή της οθόνης υγρών κρυστάλλων ενός ηλεκτρονικού υπολογιστή 35

Εικόνα 1.2 : Φωτογραφία υψηλής ανάλυσης (αριστερά) και χαμηλής ανάλυσης (δεξιά)35

Εικόνα 1.3 : Παράδειγμα Ολοκληρωτικής Φωτογραφίας με Χρήση $\mathbf{4} \times \mathbf{5}$ Μικρο-Φακών 38

Εικόνα 9.1 : Εικόνα Ολοκληρωτικής Φωτογράφισης 221

Εικόνα 9.2 : Παράδειγμα Σάρωσης Ολοκληρωτικής Εικόνας με το σχήμα των Παράλληλων Γραμμών 227

Εικόνα 9.3 : Στοιχειώδεις Εικόνες ως Καρέ Βίντεο-Ροής Μετά το Σάρωμα που Απεικονίζεται στην Εικόνα 9.2. Στην εικόνα Παρουσιάζονται τα 5 Πρώτα Καρέ. 227

Εικόνα 9.4 : (a) Παράδειγμα Ολοκληρωτικής Εικόνας (b) Ο Μετασχηματισμός της σε Υπο-Εικόνες και (c) Ο Μετασχηματισμός της στο Πεδίο των Ακτίνων 237

Εικόνα 10.1 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Birds και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 255

Εικόνα 10.2 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Seagull και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 256

Εικόνα 10.3 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Sports και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι 257

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 3.1 : Πιθανότητες Εμφάνισης Συμβόλων και Διαστήματα 55

Πίνακας 3.2 : Υπολογισμός Διαστημάτων μετά την Επεξεργασία Κάθε Συμβόλου.....56

Πίνακας 4.1 : Ο Βέλτιστος, Συμμετρικός, Ομοιόμορφος Κβαντιστής για την Ομοιόμορφη, τη Gaussian, τη Laplacian και τη Gamma κατανομή 69

Πίνακας 4.2 : Ο Βέλτιστος, Συμμετρικός, μη-Ομοιόμορφος Κβαντιστής για την Ομοιόμορφη, τη Gaussian, τη Laplacian και τη Gamma κατανομή 70

Πίνακας 4.3 : Υπολογισμός των 12 παραμορφώσεων για k=0.....81

Πίνακας 4.4 : Υπολογισμός των 12 παραμορφώσεων για k=1.....83

Πίνακας 5.1 : Τα Αρχικά Δεδομένα του Πρώτου (a) και του Δεύτερου (b) Μπλοκ 121

Πίνακας 5.2 : Οι DCT Συντελεστές του Πρώτου (a) και του Δεύτερου (b) Μπλοκ.....122

Πίνακας 5.3 : Οι Κβαντισμένοι Συντελεστές του Πρώτου (a) και του Δεύτερου (b) Μπλοκ 122

Πίνακας 5.4 : Τα ανακατασκευασμένα Δείγματα του Πρώτου (a) και του Δεύτερου (b) Μπλοκ 123

Πίνακας 5.5 : Εικόνα Συνεχούς Τόνου Μεγέθους $\mathbf{8} \times \mathbf{8}$ 123

Πίνακας 5.6 : Οι DCT Συντελεστές της Εικόνας του Πίνακας 5.5.....124

Πίνακας 5.7 : Οι Κβαντισμένοι DCT Συντελεστές της Εικόνας του Πίνακας 5.5.....124

Πίνακας 5.8 : Ανακατασκευασμένη Εικόνα με τον Αντίστροφο DCT στους Κβαντισμένους Συντελεστές του Πίνακας 5.7 125

Πίνακας 5.9 : Εικόνα Διακριτού Τόνου Μεγέθους $\mathbf{8} \times \mathbf{8}$.....125

Πίνακας 5.10 : Οι DCT Συντελεστές της Εικόνας που Παρουσιάζει ο Πίνακας 5.9 126

Πίνακας 5.11 : Οι Κβαντισμένοι DCT Συντελεστές της Εικόνας που Παρουσιάζει ο Πίνακας 5.9.....126

Πίνακας 5.12 : Ανακατασκευασμένη Εικόνα με τον Αντίστροφο DCT στους Κβαντισμένους Συντελεστές του Πίνακας 5.11 127

Πίνακας 5.13 : Ο Πίνακας Μετασχηματισμού DCT για $\boldsymbol{n} = \mathbf{3}$ 128

Πίνακας 5.14 : Ο μη-Κανομικοποιημένος πίνακας DCT στη μία Διάσταση για $\boldsymbol{n} = \mathbf{8}$ ..131

Πίνακας 5.15 : Οι Τιμές των Επιλεχθέντων Γωνιών του Μετασχηματισμού DCT-1

$(\mathbf{9} \times \mathbf{9})$ 138

Πίνακας 5.16 : Οι Τιμές των Επιλεχθέντων Γωνιών του Μετασχηματισμού DCT-2 $(\mathbf{8} \times \mathbf{8})$ 139

Πίνακας 6.1 : Τρόπος Υπολογισμού της Πρόγνωσης Ανάλογα με την Τιμή Επιλογής 151

Πίνακας 6.2 : (a) Δομή του Συμβόλου-1 κατά την Κωδικοποίηση Huffman στον Βασικό Κωδικοποιητή. (b) Δομή του Συμβόλου-2 κατά την Κωδικοποίηση στον Βασικό Κωδικοποιητή 153

Πίνακας 6.3 : Δείγματα Πηγαίας Εικόνας 154

Πίνακας 6.4 : Οι Συντελεστές DCT που Προέκυψαν από το Μετασχηματισμό της Πηγαίας Εικόνας 155

Πίνακας 6.5 : Προτεινόμενος Πίνακας Κβάντισης 155

Πίνακας 6.6 : Κανονικοποιημένοι Συντελεστές Κβάντισης 156

Πίνακας 6.7 : Αποκανονικοποιημένοι Συντελεστές Κβάντισης 156

Πίνακας 6.8 : Ανακατασκευασμένα Δείγματα Εικόνας 157

Πίνακας 7.1 : Μία Εικόνα Μεγέθους $\mathbf{8} \times \mathbf{8}$ και η Αποδόμησή της σε Υποζώνες 174

Πίνακας 8.1 : Ελάχιστες τιμές $\boldsymbol{l}$ και $\boldsymbol{r}$ ως Συνάρτηση των Τιμών $\boldsymbol{k}$ και $\boldsymbol{m}$ 209

Πίνακας 8.2 : Τα εννέα Πλαίσια του Περάσματος Διάδοσης Σημαντικότητας 217

Πίνακας 11.1 : Οι Αριθμητικές Τιμές των Μεγεθών Που Χρησιμοποιήθηκαν για την Περιγραφή των Ολοκληρωτικών Εικόνων της υποενότητας 10.1 275

Πίνακας 11.2 : Οι Αριθμητικές Τιμές των Μεγεθών Που Χρησιμοποιήθηκαν για την Περιγραφή των Ολοκληρωτικών Φωτογραφιών 284

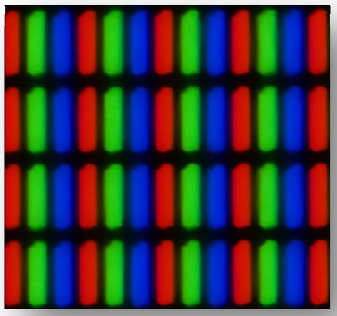
ΠΡΟΛΟΓΟΣ

Η παρούσα εργασία πραγματεύεται τη συμπίεση τρισδιάστατων εικόνων με τη μέθοδο της ολοκληρωτικής φωτογράφισης. Σημαντικό μέρος της εργασίας με βαρύνουσα σημασία αποτελεί η διεξαγωγή μετρήσεων για μελέτη και ανάλυση της απόδοσης των κωδικοποιητών. Οι μετρήσεις αυτές δε θα ήταν εφικτό να πραγματοποιηθούν εάν δεν υπήρχαν στη διάθεση μου ολοκληρωτικές εικόνες με συγκεκριμένα χαρακτηριστικά κατασκευασμένες με πρόγραμμα προσομοίωσης. Τις εικόνες αυτές κατασκεύασε και μου διέθεσε ο υποψήφιος διδάκτορας του Πανεπιστημίου Αθηνών του τμήματος Πληροφορικής και Τηλεπικοινωνιών Θεοφάνης Καλογεράκης τον οποίο και θα ήθελα να ευχαριστήσω.

1. ΕΙΣΑΓΩΓΗ

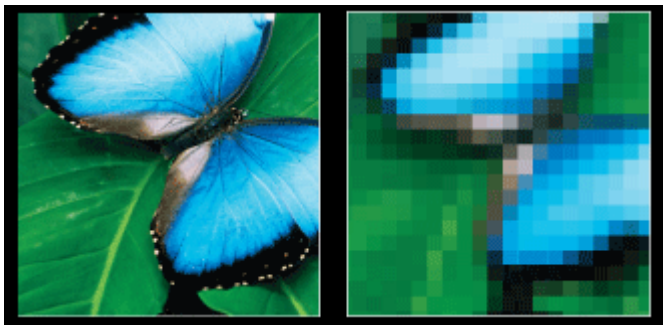
1.1 Ψηφιακή Φωτογραφία

Όλες οι ψηφιακές συσκευές οι οποίες προβάλουν ψηφιακή εικόνα (στατική ή κινούμενη) χρησιμοποιούν ως δομικό στοιχείο για την αναπαράσταση τους το εικονοστοιχείο (pixel). Κατά την προβολή μιας εικόνας από οποιαδήποτε ψηφιακή συσκευή, κάθε εικονοστοιχείο καταλαμβάνει μία πολύ μικρή, σε μέγεθος, ορθογώνια περιοχή της οθόνης προβάλλοντας ένα συγκεκριμένο χρώμα κάθε χρονική στιγμή.



Εικόνα 1.1 : Τα εικονοστοιχεία τα οποία βρίσκονται σε μία μικρή περιοχή της οθόνης υγρών κρυστάλλων ενός ηλεκτρονικού υπολογιστή

Τα εικονοστοιχεία τοποθετούνται με τέτοιο τρόπο ώστε να σχηματίζουν έναν πίνακα 2 διαστάσεων. Οι εικόνες που προβάλλονται στην οθόνη σχηματίζονται από τα διαφορετικά χρώματα που λαμβάνουν τα εικονοστοιχεία που το αποτελούν κάθε χρονική στιγμή. Η αναπαράσταση του εικονοστοιχείου όμως έχει οδοντωτές ακμές οι οποίες γίνονται αντιληπτές από το ανθρώπινο μάτι όταν το πλήθος των εικονοστοιχείων δεν είναι μεγάλο (χαμηλή ανάλυση). Για το λόγο αυτό όσο περισσότερα εικονοστοιχεία αποτελούν μία εικόνα (υψηλή ανάλυση) τόσο λιγότερο παρατηρούνται οι ακμές αυτές.



Εικόνα 1.2 : Φωτογραφία υψηλής ανάλυσης (αριστερά) και χαμηλής ανάλυσης (δεξιά)

Με την πάροδο του χρόνου το πλήθος των εικονοστοιχείων σε όλες τις ψηφιακές

συσκευές καταγραφής εικόνας (φωτογραφικές μηχανές, βίντεο-κάμερες) αυξήθηκε δραματικά, με αποτέλεσμα, ακόμα και φορητές συσκευές καταγραφής εικόνας, χαμηλού κόστους να επιτυγχάνουν αναλύσεις της τάξης των 5 Mega-Pixels. Αυτό σημαίνει πως κάθε φωτογραφία ενδέχεται να αποτελείται από 5.000.000 εικονοστοιχεία. Άμεσα γίνεται αντιληπτό το πρόβλημα σχετικά με των χώρο αποθήκευσης των δεδομένων αυτών. Κάθε εικονοστοιχείο σε μία τυπική, έγχρωμη ψηφιακή συσκευή καταγραφής εικόνας χρησιμοποιεί 8 bit για την αναπαράσταση καθενός εκ των τριών βασικών χρωμάτων. Τα βασικά χρώματα τα οποία χρησιμοποιούνται για να δημιουργηθούν όλες οι αποχρώσεις που αντιλαμβάνεται το ανθρώπινο μάτι είναι το κόκκινο, το πράσινο και το μπλε. Επομένως κάθε εικονοστοιχείο καταλαμβάνει $8bit \times 3\ \beta\alpha\sigma\iota\kappa\acute{\alpha}\ \chi\rho\acute{\omega}\mu\alpha\tau\alpha = 3bytes$. Για τα 5.000.000 εικονοστοιχεία θα χρειαστούν επομένως περίπου $14.3MB$ αποθηκευτικού χώρου για κάθε εικόνα. Συνεπώς η ανάγκη για συμπίεση των δεδομένων αυτών είναι επιτακτική. Το πρόβλημα γίνεται ακόμα μεγαλύτερο όταν τα δεδομένα πρέπει να μεταφερθούν μέσω δια-δικτύου. Συμπίεση μπορεί να επιτευχθεί είτε με κάποιες απώλειες (αποδεκτές αναλόγως των απαιτήσεων της εφαρμογής) είτε χωρίς απώλειες. Διαισθητικά αντιλαμβανόμαστε πως υπάρχει κάποια σχέση ανάμεσα στο ρυθμό (πλήθος bit που απαιτούνται για την αναπαράσταση ενός εικονοστοιχείου) και την παραμόρφωση (απώλειες στην ποιότητα της εικόνας που αντιλαμβάνεται ο χρήστης).

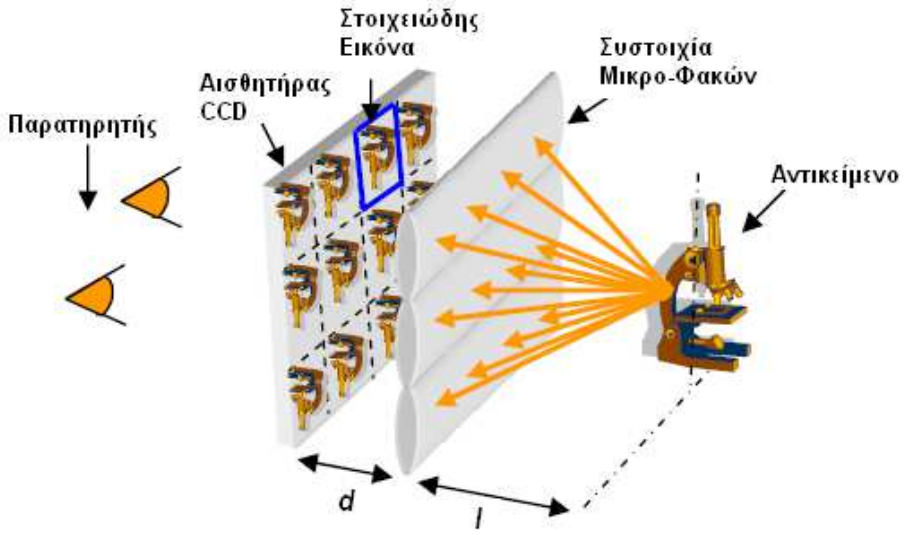
Η θεωρεία που βρίσκεται πίσω από τις τεχνικές συμπίεσης κυρίως των εικόνων καθώς και σχετικοί αλγόριθμοι, θα μελετηθούν εκτενώς στην παρούσα εργασία. Στο κεφάλαιο 10 θα διεξαχθούν μετρήσεις σχετικά με την απόδοση συμπίεσης των εικόνων ολοκληρωτικής φωτογράφισης με χρήση των σημαντικότερων τεχνικών συμπίεσης εικόνας.

1.2 Τεχνική Ολοκληρωτικής Φωτογράφισης

Μία από τις μεγάλες προκλήσεις της σύγχρονης κοινωνίας της πληροφορίας και των πολυμέσων είναι η ανάπτυξη και υλοποίηση τεχνολογιών λήψης και αναπαραγωγής τρισδιάστατων εικόνων και ταινιών αφού μεταξύ άλλων υπάρχουν προοπτικές για εφαρμογές στην ιατρική, την ασφάλεια, το στρατό καθώς και στη βιομηχανία της ψυχαγωγίας. Για το λόγο αυτό έχει υπάρξει σημαντική ερευνητική δραστηριότητα στη συγκεκριμένη περιοχή προκειμένου να βρεθούν οι καλύτερες δυνατές τεχνικές λήψης και προβολής οι οποίες με τη σειρά τους οδηγούν στην αναζήτηση των ιδανικών τεχνικών συμπίεσης[3] και μετάδοσης των δεδομένων αυτών αφού ο όγκος της τρισδιάστατης πληροφορίας θα είναι σημαντικά μεγαλύτερος. Στο φυσικό κόσμο συλλέγουμε την τρισδιάστατη πληροφορία από ένα σύνολο ερεθισμάτων. Τα

σημαντικότερα ερεθίσματα είναι η στερεοσκοπική παράλλαξη, δηλαδή η λήψη διαφορετικής εικόνας από κάθε μάτι και η παράλλαξη κίνησης, δηλαδή η λήψη διαφορετικών εικόνων καθώς ο παρατηρητής μεταβάλλει τη γωνία θέασης.

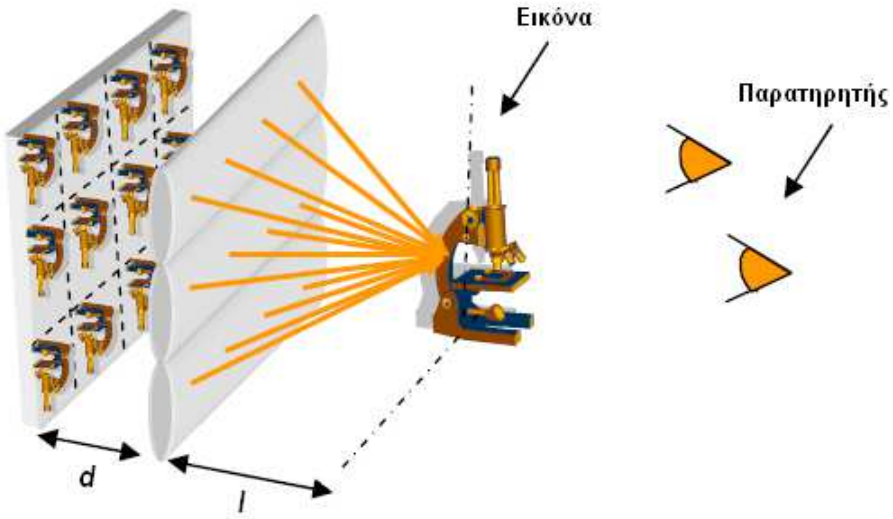
Από τις υπάρχουσες μεθόδους, η μέθοδος ολοκληρωτικής φωτογράφισης φαίνεται να είναι μία από τις πολλά υποσχόμενες τεχνικές για εφαρμογή σε τρισδιάστατες οθόνες[1][2]. Αντίθετα με άλλες τεχνικές δεν απαιτεί τη χρήση κάποιου ειδικού εξωτερικού εξοπλισμού όπως γυαλιά. Ένα επιπλέον πλεονέκτημα είναι η δυνατότητα που παρέχει στον παρατηρητή να κινείται ελεύθερα βλέποντας διαφορετικές εικόνες κατά την αλλαγή της γωνίας θέασης καθώς και της δυνατότητας να παρακολουθούν παράλληλα περισσότεροι του ενός παρατηρητές την οθόνη. Παρόλα αυτά, όπως κάθε παραδοσιακό σύστημα οπτικής, η ολοκληρωτική απεικόνιση παρουσιάζει αδυναμία στην εστίαση του βάθους καθώς και περιορισμένη αναπαράσταση του έχοντας ως συνέπεια, σκηνές με έντονο παρασκήνιο να κινδυνεύουν να βρεθούν εκτός εστίασης[60]. Υπήρξαν πολλές προσεγγίσεις με σκοπό την αύξηση του πεδίου βάθους στα συστήματα ολοκληρωτικής απεικόνισης. Χαρακτηριστικό παράδειγμα ο Jang[4], και άλλοι, οι οποίοι εισήγαγαν την ιδέα των πραγματικών και εικονικών πεδίων εικόνας. Μία άλλη τεχνική χρησιμοποιεί καμπύλες συσκευές κατά την καταγραφή και απεικόνιση των στοιχειωδών εικόνων (EI) [5] ενώ σε πολλές περιπτώσεις χρησιμοποιούνται συστοιχίες μικρο-φακών με διαφορετικές αποστάσεις εστίασης[6].



Σχήμα 1.1 : Καταγραφή Εικόνας με τη Μέθοδο της Ολοκληρωτικής Φωτογράφισης

Η βασική ιδέα της ολοκληρωτικής απεικόνισης[7] είναι η καταγραφή της τρισδιάστατης πληροφορίας μίας σκηνής μέσα από μία δισδιάστατη συστοιχία μικρο-φακών. Κατά τη μέθοδο της καταγραφής κάθε μικρο-φακός σχηματίζει μία στοιχειώδη εικόνα

λαμβανόμενη από ένα συγκεκριμένο σημείο θέασης. Το σύνολο των στοιχειωδών εικόνων καταγράφεται από έναν ψηφιακό αισθητήρα καταγραφής CCD[8] όπως παρουσιάζεται στο Σχήμα 1.1. Κατά την ανακατασκευή της εικόνας οι ληφθείσες εικόνες προβάλλονται στις αρχικές του θέσεις. Όταν οι εικόνες αυτές θεαθούν μέσα από την ίδια συστοιχία μικρο-φακών με αυτή που έγινε η καταγραφή, ο παρατηρητής αποκτά την αίσθηση των τριών διαστάσεων. Στη συνέχεια περιγράφεται η διαδικασία λήψης και ανακατασκευής των εικόνων καθώς και τα προβλήματα τα οποία προκύπτουν.



Σχήμα 1.2 : Ανακατασκευή της Τρισδιάστατης Εικόνας

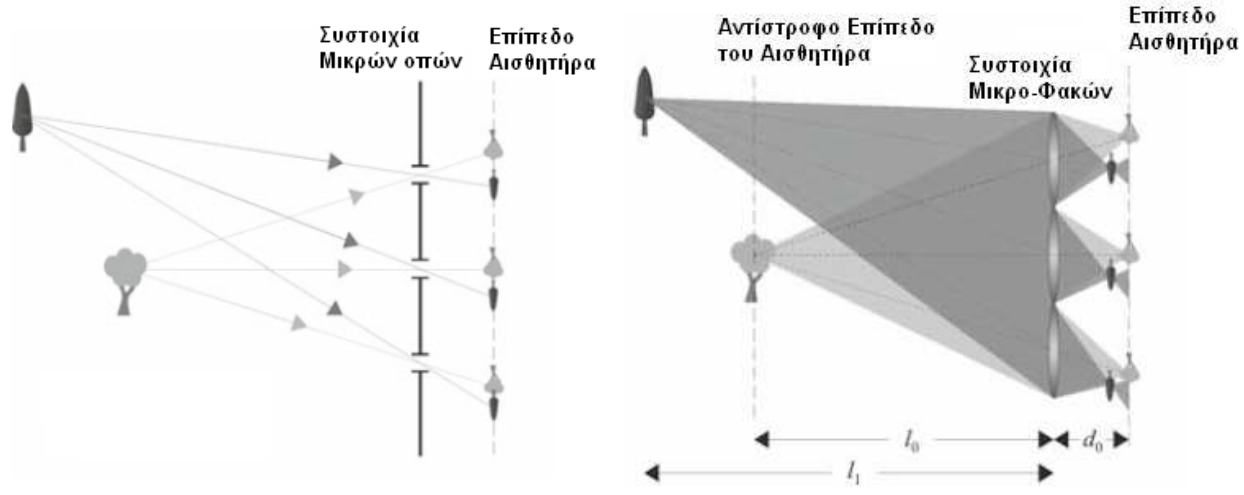
Ένα τυπικό παράδειγμα ολοκληρωτικής φωτογραφίας παρουσιάζεται στην Εικόνα 1.3

Εικόνα 1.3 : Παράδειγμα Ολοκληρωτικής Φωτογραφίας με Χρήση 4×5 Μικρο-Φακών

1.2.1 Καταγραφή

Αρχικά μπορεί να θεωρηθεί η απλούστερη περίπτωση ολοκληρωτικής απεικόνισης με μία συστοιχία μικρών οπών. Κατά το στάδιο της καταγραφής, κάθε οπή προβάλει ένα συγκεκριμένο σημείο του αντικειμένου στο επίπεδο των στοιχειωδών εικόνων. Οι θέσεις των προβολών εξαρτώνται από την πλευρική και διαμήκη θέση του αρχικού αντικειμένου. Συγκεκριμένα, η απόσταση μεταξύ των διαφορετικών προβολών του ίδιου σημείου εξαρτάται από την απόσταση του σημείου αυτού. Η πληροφορία που σχετίζεται με το βάθος καταγράφεται μέσω των συσχετίσεων ανάμεσα στις στοιχειώδεις εικόνες. Επειδή μία μόνο ακτίνα περνά από κάθε μικρή οπή, τα προβαλλόμενα σημεία λαμβάνονται όμοια για κάθε απόσταση του αντικειμένου.

Αν και το μοντέλο με χρήση οπών είναι το απλούστερο, δεν είναι ιδιαιτέρως χρήσιμο στην πράξη καθώς οι οπές μειώνουν σημαντικά το ποσό συγκέντρωσης του φωτός. Στην πράξη οι στοιχειώδεις εικόνες λαμβάνονται από συστοιχίες μικρο-φακών. Για την πρώτη αυτή περιγραφή θα θεωρηθεί πως τα οπτικά δεν προκαλούν παρεκκλίσεις πέραν της μη σωστής εστίασης. Κάθε μικρο-φακός συλλέγει μία ολόκληρη δέσμη ακτινών οι οποίες φθάνουν από ένα σημείο του αντικειμένου και τις συγκεντρώνει ώστε να σχηματιστεί ένα σημείο της εικόνας.



Σχήμα 1.3 : Διαδικασία Καταγραφής με Χρήση (a) Μικρών Οπών και (b) Μικρο-Φακών

Η σχέση ανάμεσα στην απόσταση l που βρίσκεται το αντικείμενο και την απόσταση d του επιπέδου της αντίστροφης εικόνας δίνεται από την εξίσωση των φακών:

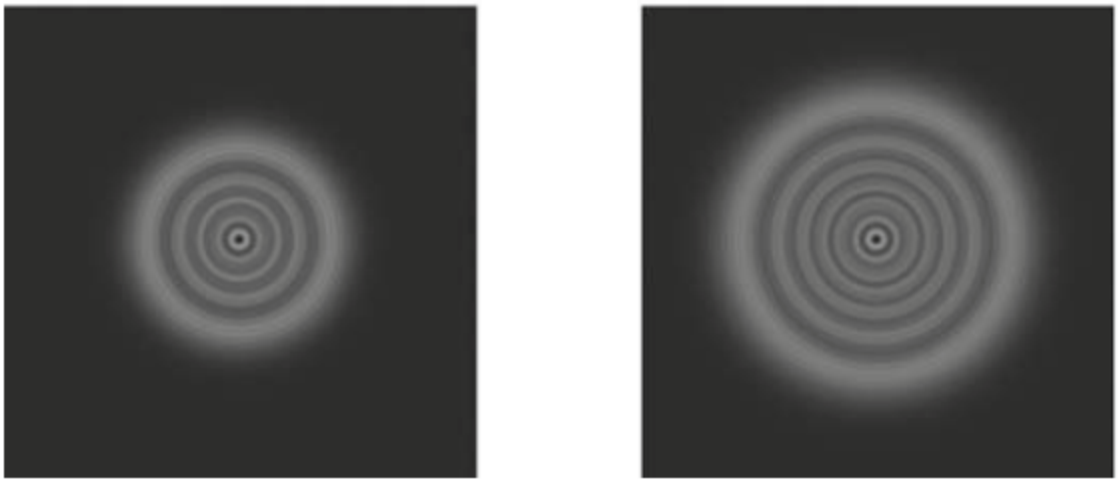
$$\frac{1}{l} + \frac{1}{d} = \frac{1}{f} \tag{1.1}$$

όπου f η εστιακή απόσταση του μικρο-φακού. Δυστυχώς, η απόσταση της αντίστροφης εικόνας d δεν ανταποκρίνεται απαραίτητα στην απόσταση του αισθητήρα d_0. Σε αυτή την περίπτωση, η προβολή του σημείου του αντικειμένου στο επίπεδο του αισθητήρα επεκτείνεται χωρικά κάτι που έχει ως αποτέλεσμα τη θολερότητα της στοιχειώδους εικόνας. Μόνο τα αντικείμενα σε απόσταση l_0 παρουσιάζονται ανεστραμμένα και πλήρως εστιασμένα στο d_0.

Το Σχήμα 1.4 παρουσιάζει την ένταση της φωτεινότητας ενός σημείου κάποιου αντικειμένου για διαφορετικές αποστάσεις και την απώλεια στην εστίαση που υπάρχει για $W = 0,\ W = 3\lambda,\ W = 5\lambda$ και $W = 7\lambda$ αντίστοιχα. Είναι προφανές ότι το φαινόμενο της θολερότητας είναι όλο και εντονότερο όσο το αντικείμενο απομακρύνεται από το σημείο εστίασης. Το πεδίο βάθους μπορεί να οριστεί ως το εύρος των αποστάσεων l γύρω από το l_0 όπου η απόκλιση στην εστίαση να είναι σχεδόν μη αντιληπτή. Στο Σχήμα 1.5 παρουσιάζονται τρείς στοιχειώδεις εικόνες της ληφθείσας εικόνας όπως προέκυψε από την προσομοίωση ενός συστήματος ολοκληρωτικής απεικόνισης. Είναι εμφανές πως δεν απεικονίζονται όλα τα αντικείμενα εξίσου καθαρά. Δεδομένου ότι στόχος της ολοκληρωτικής φωτογράφισης είναι η ακριβής καταγραφή του βάθους των σκηνών, όπως παρατηρούμε, η δυνατότητα καταγραφής εικόνων που θα εστιάζουν το ίδιο αντικείμενα σε όλες τις αποστάσεις, δεν υφίσταται και αποτελεί ένα κρίσιμο ζήτημα. Προφανώς, εάν οι στοιχειώδεις εικόνες καταγραφούν θολές, η τρισδιάστατη απεικόνιση τους στο στάδιο της ανακατασκευής θα είναι εξίσου θολή.

(a)

(b)



(c) (d)

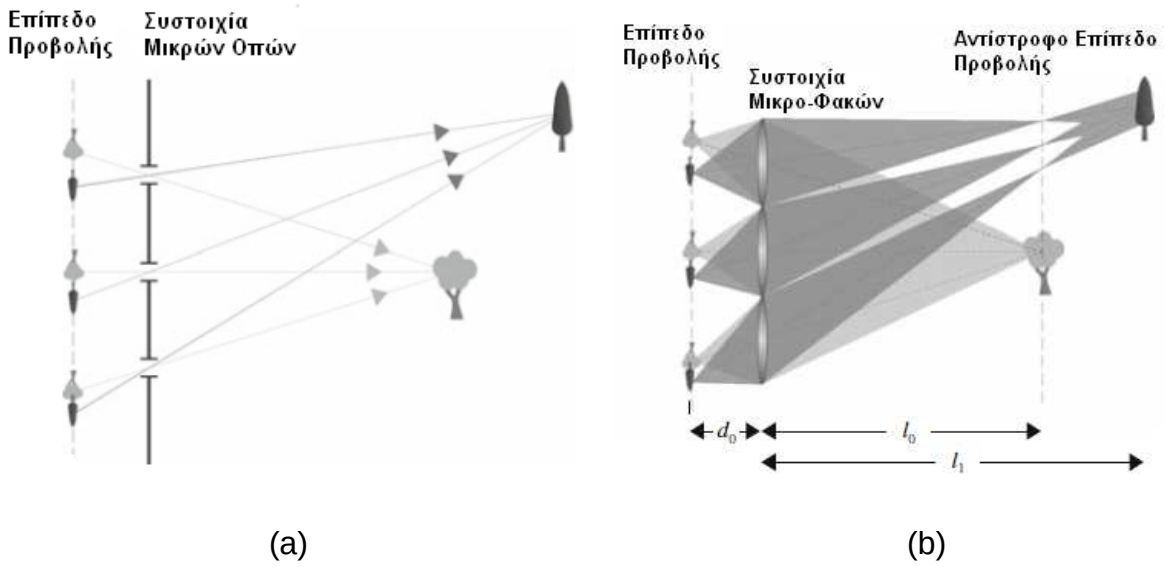
Σχήμα 1.4 : Η Απεικόνιση ενός Σημείου του Αντικειμένου Τοποθετημένο σε Διάφορες Αποστάσεις από τη Συστοιχία Μικρο-Φακών. (a) $l = 100mm$ (b) $l = 143mm$ (c) $l = 200mm$ (d) $l = 333mm$

(a) (b) (c)

Σχήμα 1.5 : Στοιχειώδεις Εικόνες Ληφθείσες από Διάφορους Μικρο-Φακούς της Συστοιχίας. Τα Αντικείμενα Βρίσκονται Τοποθετημένα στα $100mm, 143mm, 200mm$ και $333mm$ αντίστοιχα. (a) Μικρο-Φακός $(0, -2)$, (b) Κεντρικός Μικρο-Φακός (0,0), (c) Μικρο-Φακός $(0, +2)$.

1.2.2 Ανακατασκευή

Για την επεξήγηση της διαδικασίας ανακατασκευής θα θεωρήσουμε πως το πρόβλημα του πεδίου του βάθους κατά την καταγραφή δεν υφίσταται. Συνεπώς υποθέτουμε πως κάθε στοιχειώδης εικόνα είναι σωστά εστιασμένη και ακριβής.



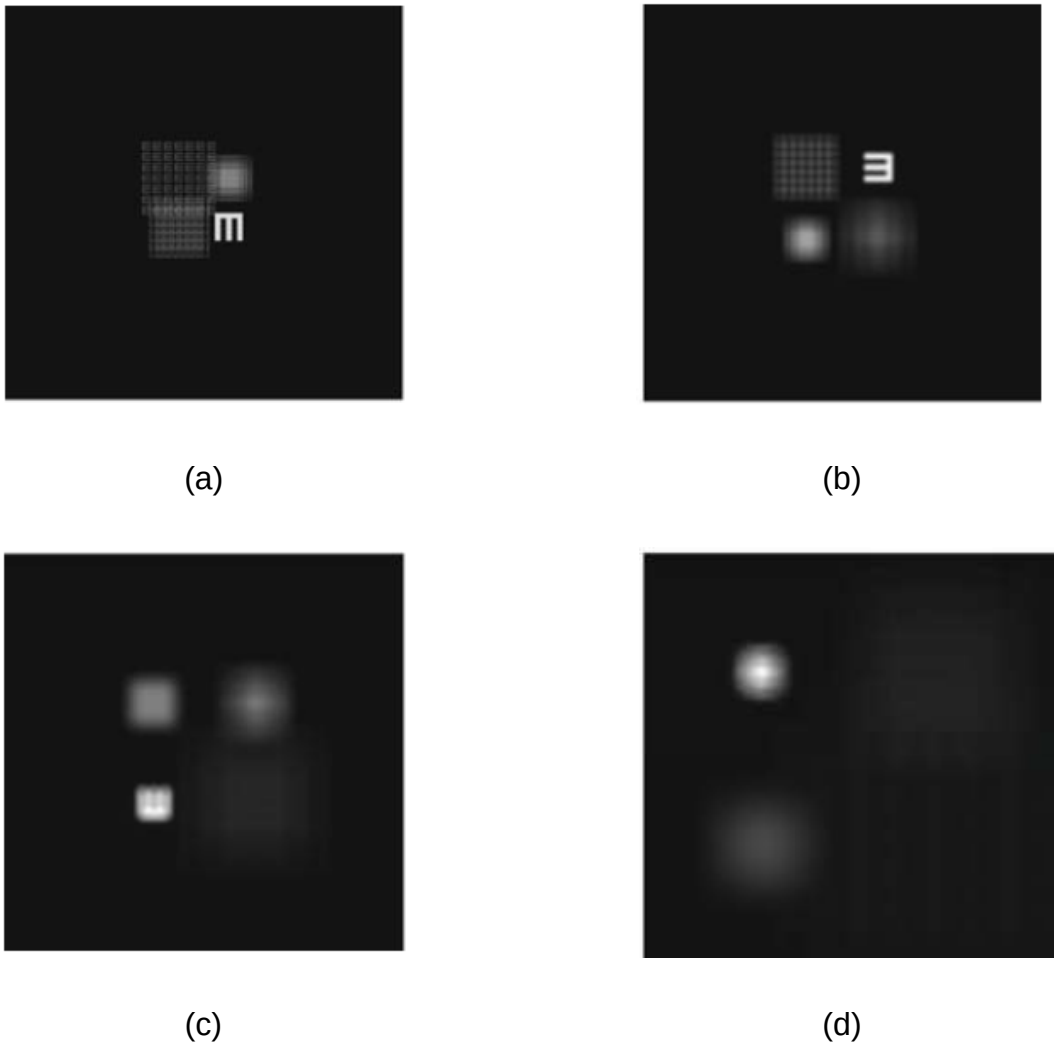
Σχήμα 1.6 : Διαδικασία Ανακατασκευής με Χρήση (a) Μικρών Οπών και (b) Μικρο-Φακών

Αυτό μπορεί να επιτευχθεί είτε με εικόνες που έχουν δημιουργηθεί τεχνητά από προγράμματα προσομοίωσης είτε με τις μικρές οπές. Στη δεύτερη περίπτωση, οι στοιχειώδεις εικόνες προβάλλονται σε μία απόσταση d_0 από τη συστοιχία των οπών. Όταν η ανακατασκευή ολοκληρωθεί οι ακτίνες οι οποίες προέρχονται από διαφορετικές στοιχειώδεις εικόνες, συνθέτουν ένα σημείο του αρχικού αντικειμένου στον τρισδιάστατο χώρο, σε απόσταση l από τη συστοιχία όπως φαίνεται στο Σχήμα 1.6.

Όταν η ανακατασκευή πραγματοποιηθεί με τη χρήση συστοιχίας μικρο-φακών, κάθε μικρο-φακός μεταφέρει μία ολόκληρη δέσμη ακτινών (αντίθετα με τη μία ακτίνα των μικρών οπών) για κάθε σημείο των στοιχειωδών εικόνων. Στο σημείο αυτό λαμβάνουν χώρα δύο ανταγωνιστικές διαδικασίες. Από τη μία πλευρά οι δέσμες οι οποίες μεταδίδονται από διαφορετικούς μικρο-φακούς συνεχίζουν να εστιάζουν στη θέση του αρχικού αντικειμένου σε απόσταση l. Από την άλλη οι δέσμες που μεταφέρονται από έναν μικρο-φακό, συγκλίνουν στο αντίστροφο επίπεδο του επιπέδου προβολής σε απόσταση l_0. Η απόσταση αυτή εξαρτάται από τη θέση του επιπέδου προβολής η οποία είναι σταθερή ενώ αντίθετα η απόσταση l εξαρτάται από τη θέση των αρχικών αντικειμένων. Προφανώς οι δύο αποστάσεις στη γενική τους περίπτωση είναι διαφορετικές. Το αποτέλεσμα αυτής της ασυμφωνίας είναι ότι το αντικείμενο ναι μεν ανακατασκευάζεται στη σωστή θέση, εμφανίζοντας όμως θολερότητα καθώς οι δέσμες δεν είναι εστιασμένες στο σημείο αυτό. Μόνο τα αντικείμενα των οποίων η αρχική θέση ήταν $l = l_0$ ανακατασκευάζονται πλήρως εστιασμένα. Στο Σχήμα 1.7 παρουσιάζονται τέσσερεις εικόνες που απεικονίζουν την ανακατασκευή μίας σκηνής σε διαφορετικά επίπεδα. Επαναλαμβάνουμε ότι η ανακατασκευή θα πραγματοποιηθεί στο σωστό

σημείο όμως η ανακατασκευασμένη εικόνα θα παρουσιάζει θολερότητα όσο το σημείο αυτό βρίσκεται μακρύτερα από το αντίστροφο επίπεδο του επιπέδου προβολής $l_0 = 100mm$. Υπάρχει ένα εύρος αποστάσεων γύρω από το σημείο εστίασης l_0 όπου η απουσία εστίασης είναι σχεδόν μη αντιληπτή. Το εύρος αυτό ονομάζεται βάθος εστίασης και ορίζεται αντίστοιχα με το πεδίο βάθους της προηγούμενης υποενότητας.

Η εντατική έρευνα στο χώρο της ολοκληρωτικής απεικόνισης απέφερε πολλές λύσεις για τη βελτίωση του πεδίου βάθους καθώς και του βάθους εστίασης οι οποίες έχουν βελτιώσει σημαντικά την απόδοση αυτών των συστημάτων. Η μελέτη τέτοιων λύσεων ξεφεύγει από τα πλαίσια της παρούσας εργασίας που σκοπό έχει την ανάδειξη των αποδοτικότερων μεθόδων συμπίεσης εικόνων ολοκληρωτικής φωτογράφισης.



Σχήμα 1.7 : Ανακατασκευασμένη Εικόνα η οποία Προέκυψε από Συστοιχία Μικρο-Φακών για Διαφορετικές Αποστάσεις Ανακατασκευής (a) $l = 100mm$ (b) $l = 143mm$ (c) $l = 200mm$ (d) $l = 333mm$

2. Βασικά Στοιχεία Θεωρίας Πληροφορίας

2.1 Πληροφορία

Ο όρος πληροφορία μπορεί να θεωρηθεί ως μία από τις θεμελιώδεις έννοιες στη συμπίεση εικόνας. Στην παρούσα ενότητα θα εξεταστούν κάποια βασικά στοιχεία της θεωρίας πληροφορίας τα θεμέλια της οποίας τέθηκαν από τον Claude Shannon[9]. Αρχικά θα μελετηθεί η ποσότητα της πληροφορίας καθώς και η εντροπία της πηγής ενώ στη συνέχεια θα εισαχθούν κάποια θεωρήματα κωδικοποίησης τα οποία έχουν σημαντικό ρόλο στη μελέτη της συμπίεσης εικόνας.

2.1.1 Εντροπία

Η εντροπία είναι ένα πολύ σημαντικό μέγεθος στη θεωρία πληροφορίας και κατ επέκταση στη συμπίεση δεδομένων. Αρχικά θα οριστεί το πληροφοριακό περιεχόμενο ενός πηγαίου συμβόλου και στη συνέχεια θα εκφραστεί η εντροπία ως το μέσο πληροφοριακό περιεχόμενο ανά σύμβολο για ένα διακριτό κανάλι χωρίς μνήμη.

2.1.2 Ποσότητα Πληροφορίας

Η πληροφορία μεταφράζεται ως γνώση και μπορεί να μετρηθεί ποσοτικά. Οι φορείς της πληροφορίας είναι τα σύμβολα. Θεωρούμε ένα σύμβολο με πιθανότητα εμφάνισης p. Το πληροφοριακό του περιεχόμενο I ορίζεται ως

$$I = \log_2 \frac{1}{p} bit \quad \text{ή} \quad I = -\log_{2p} bit \tag{2.1}$$

Δεδομένου ότι αναφερόμαστε σε δυαδική κωδικοποίηση, ως βάση του λογαρίθμου χρησιμοποιείται ο αριθμός 2. Σύμφωνα με την παραπάνω σχέση, η πληροφορία η οποία περιλαμβάνεται σε ένα σύμβολο είναι λογαριθμική συνάρτηση της πιθανότητας εμφάνισης του. Όσο μικρότερη η πιθανότητα, τόσο μεγαλύτερο το πληροφοριακό περιεχόμενο του συμβόλου. Η πιθανότητα εμφάνισης σχετίζεται με την αβεβαιότητα εμφάνισης ενός συμβόλου. Μικρή πιθανότητα εμφάνισης ενός συμβόλου σημαίνει μεγάλη αβεβαιότητα. Αξίζει να σημειωθεί πως ο ορισμός αυτός της πληροφορίας ισχύει τόσο για ισοπίθανα όσο και μη-ισοπίθανα σύμβολα.

2.1.3 Μέση Πληροφορία ανά Σύμβολο

Θεωρούμε μία διακριτή πηγή χωρίς μνήμη. Με την έννοια διακριτή εννοείται ένα μετρήσιμο πλήθος συμβόλων ενώ χωρίς μνήμη σημαίνει ότι η παρουσία ενός συμβόλου είναι ανεξάρτητη των προηγηθέντων συμβόλων. Ας υποθέσουμε μία πηγή η οποία

διαθέτει m πιθανά σύμβολα $\{s_i, i = 1,2, \dots, m\}$. Οι αντίστοιχες πιθανότητες εμφάνισης των συμβόλων αυτών συμβολίζονται με $\{p_i, i = 1,2, \dots m\}$. Σύμφωνα με όσα περιγράφηκαν, το πληροφοριακό περιεχόμενο I_i ενός συμβόλου s_i ισούται με $I_i = -\log_2 p_i \ bit$. Η εντροπία ορίζεται ως το μέσο πληροφοριακό περιεχόμενο ανά πηγαίο σύμβολο. Η εντροπία μπορεί να γραφεί ως

$$H = -\sum_{i=1}^{m} p_i \log_2 p_i \ bit \tag{2.2}$$

Από τον ορισμό αυτό προκύπτει ότι η εντροπία μίας πηγής είναι μία συνάρτηση πιθανοτήτων εμφάνισης. Εύκολα αποδεικνύεται ότι η εντροπία λαμβάνει τη μέγιστη τιμή της όταν όλα τα σύμβολα μίας πηγής είναι ισοπίθανα.

2.2 Θεώρημα Κωδικοποίησης του Shannon για Πηγή Χωρίς Θόρυβο

Θεωρούμε διακριτή, στατική πηγή, χωρίς μνήμη. Με τον όρο κωδικοποίηση εννοείται η απόδοση μίας κωδικής λέξης σε κάθε σύμβολο της πηγής. Το πλήθος των bit που αποδίδεται σε μία κωδική λέξη ονομάζονται μήκος της κωδικής λέξης. Το μέσο μήκος των κωδικών λέξεων ονομάζεται ρυθμός μετάδοσης και εκφράζεται ως ο αριθμός των bit ανά σύμβολο.

Το θεώρημα κωδικοποίησης του Shannon για πηγή χωρίς θόρυβο αναφέρει πως σε μία διακριτή στατική πηγή, χωρίς μνήμη, ο ελάχιστος ρυθμός δεδομένων που απαιτείται για την κωδικοποίηση ενός συμβόλου, κατά μέση τιμή, ισούται με την εντροπία της πηγής. Το θεώρημα αυτό παρέχει ένα κάτω όριο στην κωδικοποίηση πηγής. Ο Shannon απέδειξε ότι το κατώτερο φράγμα μπορεί να επιτευχθεί όταν η καθυστέρηση κωδικοποίησης τείνει στο άπειρο. Με τον όρο κωδικοποίηση καθυστέρησης, εννοούμε ότι ο κωδικοποιητής περιμένει και στη συνέχεια αποκωδικοποιεί ένα συγκεκριμένο αριθμό συμβόλων απευθείας. Είναι σημαντικό ότι με πεπερασμένη καθυστέρηση κωδικοποιητή επιτυγχάνεται μέση κωδική λέξη πολύ κοντά στην εντροπία. Συνεπώς δεν χρειάζεται να θυσιαστεί ο ρυθμός δεδομένων σημαντικά προκειμένου να αποφευχθεί μεγάλη καθυστέρηση κωδικοποίησης, η οποία απαιτεί μεγάλη υπολογιστική πολυπλοκότητα και μνήμη.

Σημειώνεται ότι η παραδοχή της διακριτής πηγής δεν είναι απαραίτητη. Θεωρήσαμε διακριτή πηγή μόνο διότι στην παρούσα εργασία αναφερόμαστε σε ψηφιακές εικόνες. Αντίθετα η παραδοχή της στατικότητας είναι απαραίτητη αν και στην πράξη ενδέχεται να μην ικανοποιείται. Παρόλα αυτά, το θεώρημα του Shannon δεν είναι παρά μία

θεωρητική καθοδήγηση η οποία όμως χωρίς αμφιβολία αποτελεί ένα θεμελιώδες θεωρητικό αποτέλεσμα στην θεωρία πληροφορίας.

Ένας τρόπος εκτίμησης της αποδοτικότητας ενός σχήματος κωδικοποίησης είναι να καθοριστεί η αποδοτικότητα του σε σχέση με το κάτω φράγμα, για παράδειγμα, την εντροπία. Η αποδοτικότητα n ορίζεται ως

$$n = \frac{H}{L_{avg}} \tag{2.3}$$

όπου H η εντροπία και L_{avg} το μέσο μήκος των κωδικών λέξεων του κώδικα. Δεδομένου ότι η εντροπία αποτελεί το κατώτερο φράγμα, η αποδοτικότητα δεν ξεπερνά ποτέ τη μονάδα, δηλαδή $n \leq 1$. Ο ορισμός αυτός μπορεί να χρησιμοποιηθεί και για την αποδοτικότητα ανάμεσα σε δύο κώδικες, δηλαδή

$$n = \frac{L_{avg,1}}{L_{avg,2}} \tag{2.4}$$

Όπου $L_{avg,1}$ και $L_{avg,2}$ αναπαριστούν το μέσο μήκος κωδικής λέξης του κώδικα 1 και κώδικα 2 αντίστοιχα. Συνήθως τοποθετείται η μεγαλύτερη τιμή στον παρονομαστή και n ονομάζεται η αποδοτικότητα του κώδικα 2 σε σχέση με τον κώδικα 1. Μία συμπληρωματική ποσότητα της αποδοτικότητα του κώδικα είναι ο πλεονασμός ζ ο οποίος ορίζεται ως

$$\zeta = 1 - n \tag{2.5}$$

2.3 Θεώρημα Κωδικοποίησης του Shannon σε Ενθόρυβο Κανάλι

Εάν ένας κώδικας έχει αποδοτικότητα $n = 1$ επιτυγχάνει το κατώτερο όριο κωδικοποίησης πηγής, κατ επέκταση ο πλεονασμός της πληροφορίας είναι $\zeta = 0$. Εάν θεωρήσουμε ενθόρυβο κανάλι, τότε κατά τη μετάδοση ενός συμβόλου μέσα από το κανάλι αυτό ενδέχεται να παρουσιαστεί σφάλμα με αποτέλεσμα το ληφθέν σύμβολο από το δέκτη να είναι λανθασμένο λόγω της έλλειψης πλεονασμού. Αντίθετα, είναι γνωστό πως με κατάλληλη κωδικοποίηση καναλιού, για παράδειγμα με την προσθήκη κάποιων bit πλεονασμού, υπάρχει η δυνατότητα να διορθωθούν κάποια από τα λάθη τα οποία προκύπτουν κατά τη μετάδοση μέσα από ένα ενθόρυβο κανάλι. Με τον τρόπο αυτό ο θόρυβος μπορεί να καταπολεμηθεί.

Το θεώρημα του Shannon για κωδικοποίηση σε ενθόρυβο κανάλι αναφέρει πως είναι πιθανή η μετάδοση συμβόλων σε ένα ενθόρυβο κανάλι χωρίς σφάλματα εάν ο ρυθμός δεδομένων είναι μικρότερος της χωρητικότητας του καναλιού C. Δηλαδή

$$R < C \tag{2.6}$$

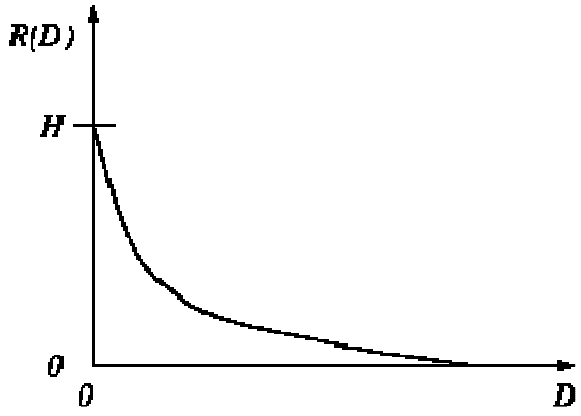
, όπου το R συμβολίζει το ρυθμό μετάδοσης. Η χωρητικότητα του καναλιού καθορίζεται από το θόρυβο και την ισχύ του σήματος. Το θεώρημα αυτό αποτελεί το δεύτερο θεώρημα του Shannon και αφόρα ενθόρυβο κανάλι χωρίς μνήμη. Επομένως, στοχεύοντας σε μία αξιόπιστη επικοινωνία, χωρίς σφάλματα, ο ρυθμός δεδομένων δε θα πρέπει να ξεπερνά τη χωρητικότητα του καναλιού.

2.4 Θεώρημα Κωδικοποίησης Πηγής του Shannon

Μέχρι στιγμής πργματοποιήθηκε αναφορά στο θεώρημα κωδικοποίησης πηγής χωρίς θόρυβο το οποίο και καθορίζει το θεωρητικό κάτω φράγμα του ρυθμού δεδομένων για κανάλι χωρίς θόρυβο καθώς και το θεώρημα κωδικοποίησης καναλιού με θόρυβο το οποίο ορίζει τον μέγιστο ρυθμό δεδομένων που μπορεί να επιτευχθεί σε κανάλι με θόρυβο χωρίς σφάλματα. Συνεπώς και τα δύο θεωρήματα αφορούν την αξιόπιστη μετάδοση της πληροφορίας. Στην παρούσα ενότητα συζητείται η διακριτή πηγή χωρίς μνήμη στην περίπτωση όπου πραγματοποιείται κωδικοποίηση με απώλειες. Αυτό έχει ως αποτέλεσμα την παραμόρφωση της πληροφορίας. Για παράδειγμα η κβάντιση, της οποίας αναφορά θα πραγματοποιηθεί σε επόμενο κεφάλαιο, προκαλεί απώλεια πληροφορίας. Επομένως από τη στιγμή που σε μία διαδικασία υπεισέρχεται η έννοια της κβάντισης τότε κάνουμε λόγο για κωδικοποίηση με απώλειες. Αυτό πρακτικά σημαίνει πως θα υπάρξει απώλεια πληροφορίας κατά την κωδικοποίηση ακόμα και εάν το κανάλι δεν προκαλεί κανένα σφάλμα. Αυτό που μας ενδιαφέρει είναι η εύρεση ενός κάτω φράγματος για το ρυθμό δεδομένων σε αυτή την περίπτωση.

Το θεώρημα κωδικοποίησης πηγής του Shannon αναφέρει ότι για δεδομένη παραμόρφωση D υπάρχει μία συνάρτηση ρυθμού-παραμόρφωσης $R(D)$ (Σχήμα 2.1) η οποία αποτελεί τον ελάχιστο ρυθμό δεδομένων με τον οποίο απαιτείται να αποστέλλει τα δεδομένα μία πηγή προκειμένου να επιτευχθεί παραμόρφωση μικρότερη ή ίση του D. Άρα για την επίτευξη παραμόρφωσης μικρότερης του D[10] ο ρυθμός R θα πρέπει να ικανοποιεί τη συνθήκη

$$R \geq R(D) \tag{2.7}$$



Σχήμα 2.1 : Συνάρτηση Ρυθμού-Παραμόρφωσης

2.5 Θεώρημα Μετάδοσης της Πληροφορίας

Είναι προφανές ότι συνδυάζοντας το θεώρημα κωδικοποίησης ενθόρυβου καναλιού και το θεώρημα κωδικοποίησης της πηγής προκύπτει η ακόλουθη σχέση

$$C \geq R(D) \tag{2.8}$$

Αυτό ονομάζεται θεώρημα μετάδοσης της πληροφορίας και ορίζει πως εάν η χωρητικότητα ενός ενθόρυβου καναλιού C είναι μεγαλύτερη από την τιμή της συνάρτησης ρυθμού-παραμόρφωσης για κάποιο D τότε είναι εφικτή η μετάδοση της πληροφορίας με παραμόρφωση D μέσα από το ενθόρυβο αυτό κανάλι.

3. Βασικές Τεχνικές Συμπίεσης και Στατιστικές Μέθοδοι

Στο κεφάλαιο 0, συζητήθηκε η έννοια της πληροφορίας και ορίστηκε η εντροπία της ως το μέσο πληροφοριακό περιεχόμενο ανά πηγαίο σύμβολο, δηλαδή ουσιαστικά εκφράζει τον ελάχιστο αριθμό bit που απαιτείται για να κωδικοποιηθούν τα σύμβολα μίας πηγής.

Στο κεφάλαιο αυτό θα μελετηθούν οι μέθοδοι που αναπτύχθηκαν προκειμένου να καταστεί δυνατή η δημιουργία κωδικοποιητών οι οποίοι θα επιτυγχάνουν μέσο μήκος κωδικής λέξης ανά σύμβολο ίσο με την εντροπία της πηγής ή πολύ κοντά σε αυτή. Πριν τη μελέτη των μεθόδων κωδικοποίησης εντροπίας θα πραγματοποιηθεί αναφορά στην πλέον χρησιμοποιημένη τεχνική συμπίεσης τόσο σε κωδικοποιητές ψηφιακών δεδομένων όσο και σε κωδικοποιητές εικόνας, βίντεο και ήχου. Πρόκειται για την τεχνική κωδικοποίησης τρέχοντος μήκους RLE[11][12] η οποία περιγράφεται στην ακόλουθη ενότητα.

3.1 Κωδικοποίηση Τρέχοντος Μήκους (Run Length Encoding)

Ο όρος «τρέχων» χρησιμοποιείται για να δηλώσει την επανάληψη ενός συμβόλου, ενώ ο όρος «τρέχων μήκος» υποδηλώνει το πλήθος των επανειλημμένων συμβόλων, δηλαδή το πλήθος των διαδοχικών όμοιων συμβόλων. Η βασική ιδέα βρίσκεται στο γεγονός ότι ενδέχεται να είναι αποδοτικότερο αντί να κωδικοποιηθούν τα διαδοχικά σύμβολα ξεχωριστά, να κωδικοποιηθεί το πλήθος των όμοιων διαδοχικών συμβόλων συνοδευόμενο από τη τιμή τους. Είναι χαρακτηριστικό πως η μέθοδος αυτή συμπίεσης χρησιμοποιήθηκε από την αρχή της ανάπτυξης της θεωρίας πληροφορίας και βρίσκει εφαρμογή ακόμα και σήμερα. Όπως θα συζητηθεί και στη συνέχεια κατά τη μελέτη του προτύπου JPEG στο κεφάλαιο 6, οι συντελεστές που προκύπτουν από το μετασχηματισμό DCT σε ένα μπλοκ 8×8, αφού πρώτα κβαντιστούν, σαρώνονται με το σχήμα zig-zag. Ο λόγος που συμβαίνει αυτό είναι επειδή μετά την κβάντιση, μεγάλο μέρος των συντελεστών θα έχει τιμή 0. Οι μη-μηδενικοί συντελεστές DCT καθώς και η θέση του στο 8×8 μπλοκ, θα πρέπει κωδικοποιηθούν και να μεταδοθούν. Η θέση ενός μη-μηδενικού συντελεστή DCT προσδιορίζεται από το τρέχων μήκος των μηδενικών ανάμεσα στους μη-μηδενικούς συντελεστές DCT. Το σχήμα zig-zag επιτυγχάνει τη ομαδοποίηση των διαδοχικών μηδενικών αποδοτικότερα σε σχέση με ένα σχήμα σάρωσης κατά γραμμές. Στη συνέχεια τόσο οι συντελεστές όσο και τιμές τρέχοντος μήκους κωδικοποιούνται με τη μέθοδο Huffman. Η μέθοδος καθώς και λεπτομέρειες σχετικά με τον τρόπο που εφαρμόζεται στα πρότυπα συμπίεσης όπου εφαρμόζεται θα ακολουθήσουν σε επόμενα κεφάλαια.

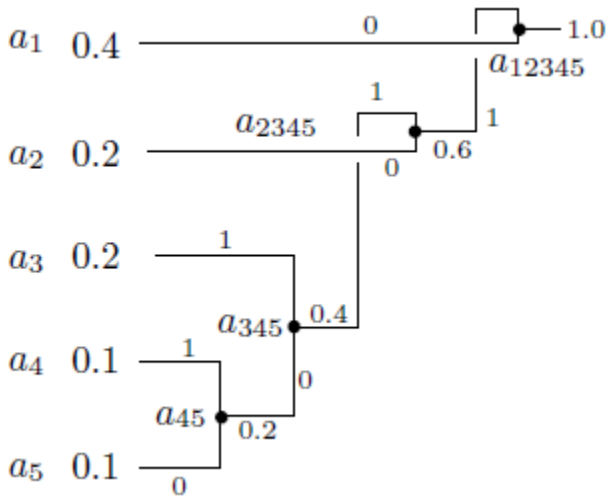
3.2 Κώδικες Huffman

3.2.1 Κωδικοποίηση

Μία μέθοδος η οποία συναντάται συχνά σε εφαρμογές συμπίεσης δεδομένων είναι η κωδικοποίηση Huffman[13]. Ο κώδικας Huffman είναι κώδικας εντροπίας και επιτυγχάνει τα βέλτιστα αποτελέσματα όταν οι πιθανότητες εμφάνισης των συμβόλων είναι αρνητικές δυνάμεις του 2. Η μέθοδος Huffman κατασκευάζει μία λίστα με κάθε σύμβολο του αλφαβήτου και την ταξινομεί κατά φθίνουσα σειρά. Στη συνέχεια κατασκευάζεται ένα δένδρο με ένα σύμβολο σε κάθε κόμβο του, από κάτω προς τα επάνω. Αυτό επιτυγχάνεται με μία σειρά βημάτων όπου σε κάθε βήμα επιλέγονται τα δύο σύμβολα με τις μικρότερες πιθανότητες εμφάνισης και το άθροισμα τους τοποθετείται στην κορυφή του μερικού δένδρου, διαγράφεται από τη λίστα και τα δύο σύμβολα αντικαθίστανται από ένα ενιαίο σύμβολο. Όταν στη λίστα απομείνει ένα μόνο σύμβολο τότε το δένδρο έχει ολοκληρωθεί. Στο τέλος το δένδρο σαρώνεται από πάνω προς τα κάτω ώστε να καθοριστούν οι κωδικές λέξεις κάθε συμβόλου. Η διαδικασία θα γίνει καλύτερα αντιληπτή με ένα παράδειγμα. Δίνονται 5 σύμβολα με πιθανότητες εμφάνισης $a_1 = 0.4,\ a_2 = 0.2,\ a_3 = 0.2,\ a_4 = 0.1\ \kappa\alpha\iota\ a_5 = 0.1$. Τα ζευγάρια τα οποία θα προκύψουν θα έχουν εξής:

1. Επιλέγονται τα σύμβολα a_4 και a_5 και αντικαθίστανται από το σύμβολο a_{45} με πιθανότητα 0.2
2. Πλέον υπάρχουν 4 σύμβολα. Το a_1 με πιθανότητα εμφάνισης 0.4 και τα a_2, a_3 και a_{45} με πιθανότητα 0.2 το καθένα. Αυθαίρετα επιλέγουμε τα σύμβολα a_3 και a_{45}, τα οποία αντικαθίστανται από το σύμβολο a_{345} με πιθανότητα εμφάνισης 0.4.
3. Σε αυτό το σημείο υπάρχουν διαθέσιμα 3 σύμβολα. Τα a_1, a_2 και a_{345} με πιθανότητες 0.4, 0.2 και 0.4 αντίστοιχα. Επιλέγεται το σύμβολο a_2 και αυθαίρετα μεταξύ των a_1 και a_{345} επιλέγουμε το a_{345} τα οποία θα αντικατασταθούν από το σύμβολο a_{2345} με πιθανότητα 0.6.
4. Τέλος, τα σύμβολα a_1 και a_{2345} αντικαθίστανται από το σύμβολο a_{12345} το οποίο έχει πιθανότητα εμφάνισης 1 και αποτελεί και τη ρίζα του τελικού δένδρου.

Το δένδρο πλέον έχει ολοκληρωθεί. Στο Σχήμα 3.1 εμφανίζεται η ρίζα του δένδρου επάνω δεξιά και οι πέντε κόμβοι προς τα αριστερά.



Σχήμα 3.1 : Δένδρο Huffman

Στους κόμβους, οι κωδικοί ανατίθενται αυθαίρετα (κατά σύμβαση), το bit 0 στον ένα κόμβο και το bit 1 στον άλλο, για όλους τους κόμβους. Το αποτέλεσμα του παραδείγματος είναι οι κωδικές λέξεις 0, 10, 111, 1101 και 1100 αλλά προφανώς θα μπορούσαν να ήταν και διαφορετικές εάν γινόταν ανάποδα η ανάθεση των bit στους κόμβους του δένδρου ή εάν επιλέγαμε διαφορετικά ζεύγη ισοπίθανων κόμβων κατά τα βήματα 2 και 3. Το μέσο μέγεθος του κώδικα είναι $0.4 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.1 \times 4 + 0.1 \times 4 = 2.2$ bit/σύμβολο. Αξίζει να σημειωθεί ότι και διαφορετικά να είχαν ανατεθεί τα bit, το μέσο μέγεθος του κώδικα θα ήταν και πάλι το ίδιο.

3.2.2 Περιορισμοί

Η μέθοδος Huffman δεν μπορεί να συμπιέσει δεδομένα τα οποία είναι ισοπίθανα. Αυτό γίνεται άμεσα αντιληπτό αφού ισοπίθανα σύμβολα σημαίνει ότι τα σύμβολα είναι τοποθετημένα τυχαία και σε τυχαία δεδομένα δεν υπάρχει συμπίεση. Υπάρχουν κάποιες ειδικές περιπτώσεις όπου τα σύμβολα με ίσες πιθανότητες δεν είναι τοποθετημένα εντελώς τυχαία και μπορούν να συμπιεστούν. Ένα τυπικό παράδειγμα θα μπορούσε να αποτελεί η ακολουθία $a_1a_1a_1a_1 \dots a_1a_2a_2a_2 \dots a_2a_3a_3a_3 \dots$ όπου όπως παρατηρούμε το ίδιο σύμβολο εμφανίζεται πολλές φορές διαδοχικά. Σε αυτή την περίπτωση μπορεί να υπάρξει συμπίεση με τη χρήση της μεθόδου RLE αλλά σε καμία περίπτωση με κώδικα Huffman. Η μέθοδος Huffman επίσης, δε μπορεί να εφαρμοστεί σε πηγή με αλφάβητο δύο συμβόλων. Σε ένα αλφάβητο αυτής της μορφής, ο κωδικός 0 θα δινόταν στο ένα σύμβολο και ο κωδικός 1 στο άλλο. Η κωδικοποίηση Huffman δε μπορεί να αποδώσει σε κανένα σύμβολο, κωδικό μικρότερο του ενός bit, επομένως δεν έχει τη δυνατότητα να βελτιώσει έναν τέτοιο απλό κώδικα. Εάν θεωρήσουμε μία δυαδική

εικόνα η οποία αποτελείται από μεμονωμένα bit, είναι δυνατό να συνδυαστούν αυτά κατά τετράδες ή οκτάδες προκειμένου να δημιουργήσουν ένα νέο σύμβολο, θεωρώντας ότι το νέο αλφάβητο αποτελείται από τα 16 ή 256 αντίστοιχα σύμβολα. Το πρόβλημα με τη μέθοδο αυτή είναι ότι τα αρχικά δυαδικά δεδομένα είναι πολύ πιθανό να παρουσιάζουν ισχυρή συσχέτιση μεταξύ τους και μέρος της συσχέτισης αυτής θα χαθεί όταν τα bit σχηματίσουν σύμβολα. Σε μία τυπική δυαδική εικόνα, ένα εικονοστοιχείο είναι πιθανότερο να ακολουθείται από ένα όμοιο εικονοστοιχείο παρά από ένα ανόμοιο. Επομένως, ίσως σε μία περίπτωση σαν και αυτή να ήταν αποδοτικότερη μία μέθοδος συμπίεσης RLE παρά οι κώδικες Huffman.

3.2.3 Αποκωδικοποίηση

Προτού ξεκινήσει η κωδικοποίηση των δεδομένων, ο κωδικοποιητής θα πρέπει να καθορίσει τον κώδικα Huffman. Για να καθοριστεί ο κώδικας θα πρέπει πρώτα να γίνουν γνωστές οι πιθανότητες εμφάνισης των συμβόλων. Οι πιθανότητες εμφάνισης θα πρέπει να εμφανίζονται μέσα στη συμπιεσμένη ροή δεδομένων ώστε ο αποκωδικοποιητής να είναι σε θέση να αποκωδικοποιήσει την ακολουθία Huffman. Αυτό επιτυγχάνεται εύκολα καθώς οι συχνότητες είναι ακέραιοι αριθμοί και επομένως οι πιθανότητες εμφάνισης μπορούν να γραφούν ως ακέραιοι υπό κλίμακα. Η διαδικασία αυτή συνήθως προσθέτει μερικές εκατοντάδες byte στη συμπιεσμένη ροή δεδομένων.

Η αποκωδικοποίηση μπορεί να ξεκινήσει μόνο εφόσον ο αποκωδικοποιητής κατασκευάσει από τις πιθανότητες του αλφάβητου τον κώδικα Huffman. Ο αλγόριθμος αποκωδικοποίησης είναι απλός. Διαβάζεται το πρώτο bit της ροής δεδομένων. Εάν είναι μηδέν ακολουθείται ο αριστερός κόμβος του δένδρου αλλιώς εάν είναι άσσος διαβάζεται ο δεξιός (κατά συνθήκη που ορίζεται τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή). Η διαδικασία συνεχίζεται μέχρι κάποιο bit να οδηγήσει σε φύλλο του δένδρου. Όταν ο αποκωδικοποιητής φτάσει σε φύλλο τότε εντοπίζεται το αρχικό ασυμπίεστο σύμβολο που αντιστοιχεί στην ακολουθία των bit τα οποία οδήγησαν από τη ρίζα στο φύλλο του δένδρου. Ο αλγόριθμός επαναλαμβάνεται και ξεκινά από τη ρίζα του δένδρου ακολουθώντας την ίδια διαδικασία προκειμένου να αποκωδικοποιηθεί κάθε σύμβολο που περιλαμβάνεται στην ακολουθία.

3.3 Αριθμητικοί Κώδικες

3.3.1 Κωδικοποίηση

Ο κώδικας Huffman είναι απλός, αποδοτικός και παράγει καλούς κώδικες για μεμονωμένα σύμβολα. Παρόλα αυτά όπως συζητήθηκε και σε προηγούμενη ενότητα, η μοναδική περίπτωση στην οποία παράγει ιδανικούς κώδικες μεταβλητού μήκους, δηλαδή κώδικες όπου το μέσο μήκος του είναι ίδιο με την εντροπία της πηγής, είναι όταν τα σύμβολα έχουν πιθανότητα εμφάνισης που είναι αρνητικές δυνάμεις του 2. Αυτό συμβαίνει καθώς η μέθοδος Huffman αναθέτει μία κωδική λέξη με ακέραιο πλήθος bit σε κάθε σύμβολο του αλφαβήτου. Η θεωρεία πληροφορίας έχει δείξει ότι σε ένα σύμβολο με πιθανότητα εμφάνισης 0.4 θα έπρεπε ιδανικά να ανατεθεί μία κωδική λέξη με μήκος 1.32 bit, καθώς $-\log_2 0.4 = 1.32$. Παρόλα αυτά η μέθοδος Huffman θα ανέθετε σε ένα τέτοιο σύμβολο 1 ή 2 bit. Η αριθμητική κωδικοποίηση[14][15] ξεπερνά το πρόβλημα ανάθεσης ακέραιων κωδικών λέξεων σε κάθε ξεχωριστό σύμβολο, αναθέτοντας μία (συνήθως μεγάλη) κωδική λέξη σε ολόκληρο το αρχείο. Η μέθοδος ξεκινά με ένα συγκεκριμένο διάστημα, διαβάζει το αρχείο εισόδου σύμβολο προς σύμβολο και χρησιμοποιεί την πιθανότητα εμφάνισης κάθε συμβόλου προκειμένου να περιορίσει το διάστημα αυτό. Ο προσδιορισμός ενός στενότερου διαστήματος απαιτεί περισσότερα bit. Για να επιτευχθεί συμπίεση, ο αλγόριθμος σχεδιάστηκε με τέτοιο τρόπο ώστε τα σύμβολα με μεγάλη πιθανότητα εμφάνισης να περιορίζουν το διάστημα λιγότερο σε σχέση με τα σύμβολα που εμφανίζονται με μικρότερες πιθανότητες και επομένως τα σύμβολα με μεγάλη πιθανότητα εμφάνισης να συνεισφέρουν λιγότερα bit στην έξοδο. Το αρχικό διάστημα είναι το $[0, 1)$, δηλαδή $0 \leq x < 1$. Η διαδικασία της κωδικοποίησης θα γίνει περισσότερο αντιληπτή με τη χρήση ενός παραδείγματος.

Θεωρείται το αλφάβητο $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ με πιθανότητες που ορίζει ο Πίνακας 3.1

Πίνακας 3.1 : Πιθανότητες Εμφάνισης Συμβόλων και Διαστήματα

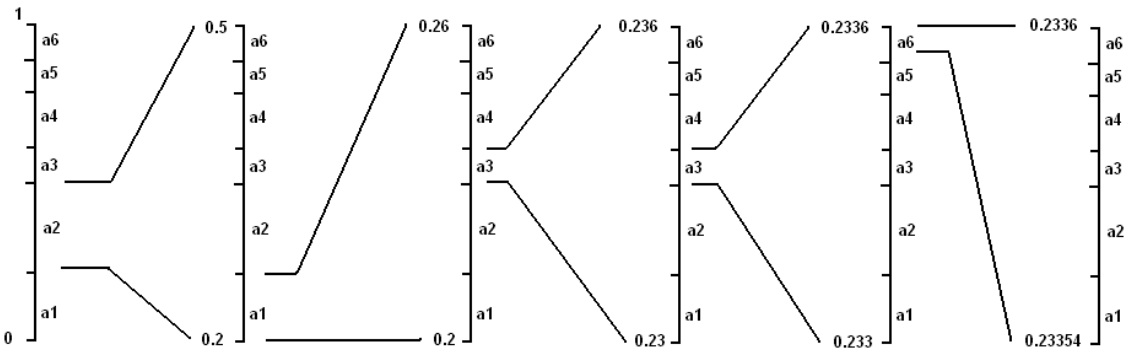
Σύμβολο	Πιθανότητα Εμφάνισης	Εύρος
a_1	0.2	$[0, 0.2)$
a_2	0.3	$[0.2, 0.5)$
a_3	0.1	$[0.5, 0.6)$
a_4	0.2	$[0.6, 0.8)$
a_5	0.1	$[0.8, 0.9)$
a_6	0.1	$[0.9, 1)$

Θεωρούμε επίσης ότι επιθυμούμε να κωδικοποιηθεί η ακολουθία $a_2a_1a_3a_3a_6$. Αρχικά τόσο ο κωδικοποιητής όσο και ο αποκωδικοποιητής γνωρίζουν ότι το εύρος είναι το $[0,1)$. Μετά την εμφάνιση του πρώτου συμβόλου, a_2, ο κωδικοποιητής περιορίζει το διάστημα στο $[0.2,0.5)$ και το δεσμεύει στο σύμβολο αυτό. Το δεύτερο σύμβολο, α_1, περιορίζει το διάστημα αυτό στο 1/5 του αρχικού διαστήματος, καθώς το α_1 βρισκόταν στο διάστημα $[0,0.2)$ με μήκος 0.2=1/5. Έτσι δημιουργείται το διάστημα $[0.2,0.26)$, καθώς το προηγούμενο διάστημα ήταν μήκους 0.3 μονάδων και το 1/5 αυτού είναι το 0.06. Το επόμενο σύμβολο είναι το σύμβολο a_3 και βρίσκεται στο διάστημα $[0.5,0.6)$, το οποίο όταν εφαρμοστεί στο διάστημα $[0.2,0.26)$ παράγει το ακόμα μικρότερο διάστημα $[0.23,0.236)$. Ακολουθώντας την ίδια διαδικασία ολοκληρώνεται η κατασκευή των διαστημάτων όπως παρουσιάζει ο Πίνακας 3.2.

Πίνακας 3.2 : Υπολογισμός Διαστημάτων μετά την Επεξεργασία Κάθε Συμβόλου

	$[0 \quad 1)$
a_2	$[0.2 \quad 0.5)$
a_1	$[0.2 \quad 0.26)$
a_3	$[0.23 \quad 0.236)$
a_3	$[0.233 \quad 0.2336)$
α_6	$[0.23354 \quad 0.2336)$

Το Σχήμα 3.2 παρουσιάζει έναν εναλλακτικό τρόπο αναπαράστασης της διαδικασίας κωδικοποίησης. Οι κάθετες ράβδοι αναπαριστούν τις πιθανότητες των συμβόλων. Μετά την επεξεργασία του πρώτου συμβόλου το διάστημα στο οποίο εργαζόμαστε περιορίζεται στο $[0.2,0.5)$. Μετά το δεύτερο σύμβολο, το διάστημα περιορίζεται στο $[0.2,0.26)$ ενώ όλα τα διαστήματα τα παρουσιάζει τόσο το Σχήμα 3.2 όσο και ο Πίνακας 3.2.



Σχήμα 3.2 : Γραφική Αναπαράσταση Αριθμητικής Κωδικοποίησης

3.3.2 Αποκωδικοποίηση

Για την αποκωδικοποίηση του αποτελέσματος ακολουθείται η ακριβώς αντίστροφη διαδικασία. Υποθέτουμε ότι ο αποκωδικοποιητής γνωρίζει ότι το τελικό μήνυμα βρίσκεται στο διάστημα $[0.23354, 0.2336)$. Αμέσως μπορεί να αποκωδικοποιήσει το πρώτο σύμβολο και να καταλήξει πως είναι το a_2, καθώς το διάστημα αυτό βρίσκεται μέσα στο διάστημα $[0.2, 0.5]$ όπως ορίζει ο Πίνακας 3.1. Για να υπολογιστεί το νέο διάστημα επιλέγεται ένας οποιοσδήποτε αριθμός από το διάστημα $[0.23354, 0.2336)$ από τον οποίο αφαιρείται η τιμή της αρχής του διαστήματος του προηγούμενου συμβόλου και διαιρείται με το εύρος του διαστήματος αυτού. Στο παράδειγμα μας ισχύει ότι:

Έστω ότι επιλέγεται ο τυχαίος αριθμός 0.23356 από το διάστημα $[0.23354, 0.2336)$ οπότε έχουμε ότι:

$$\frac{0.23356 - 0.2}{0.23354 - 0.2336} = \frac{0.0336}{0.3} = 0.1119$$

Ο αριθμός 0.1119 βρίσκεται στο διάστημα $[0, 0.2)$, συνεπώς ο δεύτερος χαρακτήρας όπως ορίζει ο Πίνακας 3.1 είναι το a_1. Η διαδικασία επαναλαμβάνεται μέχρι να αποκωδικοποιηθούν όλα τα σύμβολα. Ένα πρόβλημα το οποίο θα αντιμετωπίσει ο αποκωδικοποιητής είναι άγνοια σχετικά με τη στιγμή ολοκλήρωσης της διαδικασίας αποκωδικοποίησης όλων των συμβόλων. Το τέλος της ροής δεδομένων θα πρέπει να υποδηλώνεται με ένα ειδικό σύμβολο γνωστό τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή.

4. Κβάντιση

4.1 Η έννοια της Κβάντισης

Η κβάντιση[16] αποτελεί μία απαραίτητη διαδικασία στη συμπίεση δεδομένων με απώλειες και έχει άμεση επίδραση στο ρυθμό δεδομένων και την παραμόρφωση της ανακατασκευασμένης εικόνας. Με τον όρο κβάντιση αναφερόμαστε θεμελιωδώς στην ψηφιοποίηση μίας ποσότητας, βήμα πολύ σημαντικό στη συμπίεση ψηφιακής εικόνας με απώλειες. Η είσοδος και η έξοδος ενός κβαντιστή μπορεί να είναι είτε βαθμωτά είτε διανύσματα. Αρχικά θα εξεταστεί η βαθμωτή κβάντιση και στη συνέχεια η διανυσματική. Μετά την κβάντιση, κωδικές λέξεις αποδίδονται στις διάφορες τιμές των συμβόλων της πηγής από την έξοδο του κβαντιστή. Ο φυσικός δυαδικός κώδικας (NBC) οι κώδικες μεταβλητού μήκους (VLC) καθώς και ο κώδικας τρέχοντος μήκους μηδενικών (RLC) αποτελούν παραδείγματα τέτοιων κωδίκων.

Η έννοια της κβάντισης επιφέρει κάποιες επιπτώσεις στην πληροφορία. Καταρχήν η κβάντιση είναι μία μη αντιστρέψιμη διαδικασία και επομένως στη γενική περίπτωση δεν υπάρχει μέθοδος για την εύρεση της αρχικής τιμής από την κβαντισμένη. Δεύτερον η κβάντιση αποτελεί μία πηγή απώλειας της πληροφορίας και στην ουσία πρόκειται για ένα κρίσιμο βήμα στη συμπίεση δεδομένων. Η επίδραση της είναι σημαντική τόσο στην παραμόρφωση της ανακατασκευασμένης εικόνας όσο στο ρυθμό μετάδοσης. Προφανώς ισχυρή κβάντιση έχει ως αποτέλεσμα μεγαλύτερες παραμορφώσεις και χαμηλότερους ρυθμούς από μία λιγότερο επιθετική κβάντιση.

4.2 Βαθμωτή Κβάντιση

Δεδομένου ότι στο σημείο αυτό θα ασχοληθούμε με τη βαθμωτή κβάντιση, θα θεωρήσουμε ότι τα δεδομένα προς συμπίεση είναι σε μορφή αριθμών οι οποίοι δίνονται ως είσοδος από μία πηγή. Στη συνέχεια θα γενικεύσουμε τη θεώρηση της κβάντισης με περιπτώσεις όπου τα δεδομένα αποτελούνται από σύνολα αριθμών, γνωστά ως διανύσματα και όχι από μεμονωμένους αριθμούς, προκειμένου να μελετηθεί η διανυσματική κβάντιση.

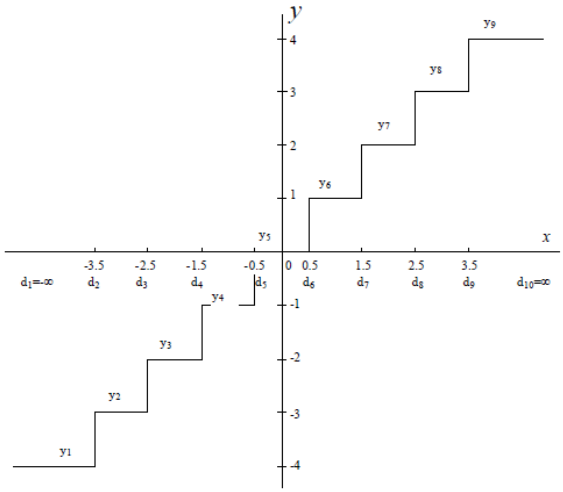
Στην παρούσα ενότητα θα αναλυθεί πλήθος αρχών και τεχνικών κβάντισης συμπεριλαμβανομένου της ομοιόμορφης, μη-ομοιόμορφης και βέλτιστης κβάντισης.

4.2.1 Ομοιόμορφη Κβάντιση

Η ομοιόμορφη κβάντιση είναι η απλούστερη και πιο δημοφιλής τεχνική κβάντισης. Η θεωρητική της σημασία είναι πολύ μεγάλη. Για το λόγο αυτό η μελέτη σχετικά με την

κβάντιση ξεκινά με τον ομοιόμορφο κβαντιστή και στη συνέχεια θα εξεταστούν και οι υπόλοιποι τύποι κβαντιστών.

Στο Σχήμα 4.1 ο οριζόντιος άξονας παριστάνει την είσοδο στον κβαντιστή ενώ ο κατακόρυφος άξονας παριστάνει την έξοδο του. Η σχέση μεταξύ εισόδου – εξόδου είναι εκείνη που χαρακτηρίζει τον κβαντιστή και το γράφημα αυτό αποτελεί την χαρακτηριστική εισόδου – εξόδου του κβαντιστή. Στο Σχήμα 4.1 παρατηρούνται 9 διαστήματα κατά μήκος του άξονα X. Όταν η τιμή της εισόδου βρίσκεται σε κάποιο από τα διαστήματα αυτά τότε η τιμή της εξόδου λαμβάνει την αντίστοιχη τιμή στον άξονα Y. Η χαρακτηριστική εισόδου – εξόδου του κβαντιστή έχει μια κλιμακοειδή μορφή η οποία είναι εμφανώς μη γραμμική.



Σχήμα 4.1 : Χαρακτηριστική Εισόδου – Εξόδου Ομοιόμορφου Midtread Κβαντιστή

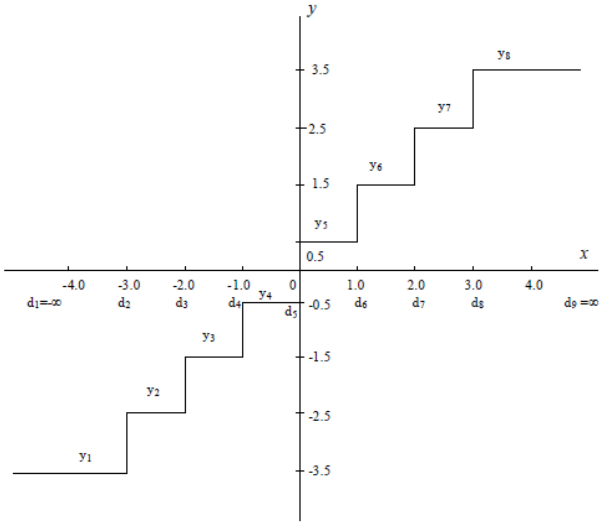
Τα όρια των διαστημάτων ονομάζονται επίπεδα αποφάσεων και σημειώνονται με d_i όπου i ο δείκτης των διαστημάτων. Η έξοδος του κβαντιστή αναφέρεται ως στάθμη ανακατασκευής (γνωστή και ως στάθμη κβάντισης) και σημειώνεται ως y_i όπου i ο δείκτης του. Το μήκος του διαστήματος ονομάζεται βήμα του κβαντιστή και συμβολίζεται με Δ. Έχοντας καθορίσει πλέον τους παραπάνω όρους μπορεί σε αυτό το σημείο να καθοριστεί μαθηματικά η συνάρτηση του κβαντιστή του σχήματος ως:

$$y_i = Q(x), \qquad \varepsilon\acute{\alpha}\nu\ x \in (d_i, d_{i+1}) \tag{4.1}$$

Όπου $i = 1,2,\dots,9$ και $Q(x)$ η έξοδος του κβαντιστή σε σχέση με την είσοδο x. Σημειώνεται ότι στο σχήμα $\Delta = 1$. Τα επίπεδα απόφασης και οι στάθμες ανακατασκευής είναι ισοκατανεμημένα. Πρόκειται για έναν ομοιόμορφο κβαντιστή λόγω του ότι διαθέτει τα εξής δύο χαρακτηριστικά:

1. Εκτός από το, κατά πάσα πιθανότητα, αριστερό και δεξί διάστημα, όλα τα υπόλοιπα διαστήματα κατά μήκος του άξονα x έχουν όμοια απόσταση μεταξύ τους. Αυτό σημαίνει πως κάθε εσωτερικό διάστημα έχει το ίδιο μήκος.
2. Εκτός από τα κατά πάσα πιθανότητα ακραία διαστήματα, οι στάθμες ανακατασκευής του κβαντιστή έχουν επίσης όμοιες αποστάσεις. Παράλληλα, κάθε εσωτερική στάθμη ανακατασκευής είναι ο αριθμητικός μέσος όρος των δύο επιπέδων απόφασης του αντίστοιχου διαστήματος κατά μήκος του άξονα x.

Ο ομοιόμορφος κβαντιστής που εμφανίζεται στο Σχήμα 4.1 ονομάζεται midtread κβαντιστής. Κάθε κβαντιστής αντίστοιχης μορφής σύμφωνα όμως με τον οποίο οι στάθμες ανακατασκευής δεν περιέχουν την τιμή του μηδενός ονομάζεται midrise κβαντιστής. Ένας midrise κβαντιστής έχει βήμα κβάντισης $\Delta = 1$ όπως φαίνεται στο Σχήμα 4.2. Οι midtread κβαντιστές χρησιμοποιούνται συνήθως για περιττό αριθμό σταθμών ανακατασκευής ενώ οι midrise για άρτιο αριθμό σταθμών κβάντισης.



Σχήμα 4.2 : Χαρακτηριστική Εισόδου – Εξόδου Ομοιόμορφου midrise Κβαντιστή

Σημειώνεται ότι η χαρακτηριστική εισόδου-εξόδου τόσο του midtread όσο και του midrise ομοιόμορφου κβαντιστή που παρουσιάζουν το Σχήμα 4.1 και Σχήμα 4.2 αντίστοιχα, έχουν περιττή συμμετρία ως προς τον κατακόρυφο άξονα στο $\chi = 0$. Η μελέτη θα συνεχιστεί κάτω από τη θεώρηση αυτής της συμμετρίας αφού τα συμπεράσματα που προκύπτουν δεν προκαλούν βλάβη της γενικότητας.

Θεωρώντας με N το πλήθος των σταθμών ανακατασκευής ενός κβαντιστή και παρατηρώντας το Σχήμα 4.1 και Σχήμα 4.2, εμφανίζεται πως εάν το N είναι άρτιος αριθμός τότε το επίπεδο απόφασης $d_{\left(\frac{N}{2}\right)+1}$ είναι τοποθετημένο στο μέσον του άξονα x. Εάν το N είναι περιττός αριθμός τότε η στάθμη ανακατασκευής $y_{(N+1)/2}$ ισούται με

μηδέν. Η θεώρηση αυτή είναι σημαντική για την περαιτέρω κατανόηση των πινάκων σχεδιάσης κβαντιστών.

4.2.2 Παραμόρφωση Κβάντισης

Το θεώρημα κωδικοποίησης πηγής, στο οποία πραγματοποιήθηκε αναφορά σε ενότητα 0, ισχυρίζεται ότι για δεδομένη παραμόρφωση D, υπάρχει μία συνάρτηση ρυθμού – παραμόρφωσης $R(D)$ τέτοια ώστε όσο ο ρυθμός δεδομένων που χρησιμοποιείται είναι μεγαλύτερος από $R(D)$ τότε είναι δυνατή η μετάδοση των συμβόλων της πηγής με παραμόρφωση μικρότερη του D. Δεδομένου ότι δεν είναι εφικτός ένας άπειρος ρυθμός δεδομένων προκειμένου να αναπαρασταθούν τα σύμβολα της πηγής, η εισαγωγή παραμόρφωσης είναι αναπόφευκτη. Ουσιαστικά από τη στιγμή που η κβάντιση προκαλεί μη αντιστρέψιμη απώλεια πληροφορίας αποδεχόμαστε και ένα σφάλμα το οποίο υπεισέρχεται. Το γεγονός αυτό δημιουργεί και ένα ερώτημα σχετικά με την αξιολόγηση της ποιότητας. Στο πλαίσιο της αντικειμενικής αξιολόγησης θα καθοριστούν κάποιες μετρικές όπως η μέση τετραγωνική τιμή (MSE) η ριζική μέση τετραγωνική τιμή (RMSE) και ο λόγος σήματος προς θόρυβο (SNR). Σε σχέση με την κβάντιση ορίζεται ως σφάλμα κβάντισης e_q, η διαφορά ανάμεσα στο σήμα εισόδου και την κβαντισμένη έξοδο:

$$e_q = x - Q(x) \tag{4.2}$$

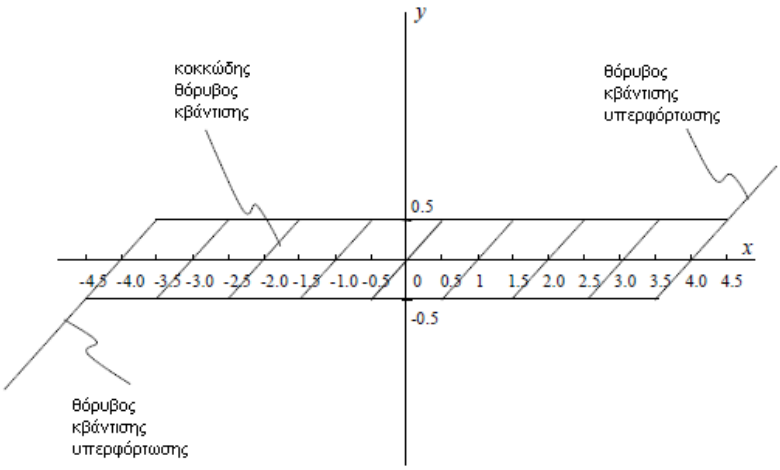
Όπου x και $Q(x)$ είναι η είσοδος και η κβαντισμένη έξοδος αντίστοιχα. Το σφάλμα κβάντισης συχνά αναφέρεται και ως θόρυβος κβάντισης. Συνηθίζεται να θεωρείται η είσοδος x ως μία τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας (pdf) $f_x(x)$. Συνεπώς το μέσο τετραγωνικό σφάλμα κβάντισης, MSE_q, μπορεί να εκφραστεί ως

$$MSE_q = \sum_{i=1}^{N} \int_{d_i}^{d_{i+1}} \left(x - Q(x)\right)^2 f_x(x)dx \tag{4.3}$$

, όπου N το πλήθος των σταθμών ανακατασκευής. Είναι προφανές πως όταν η συνάρτηση πυκνότητας πιθανότητας παραμένει αμετάβλητη τότε λιγότερες στάθμες ανακατασκευής (μικρότερο N) οδηγούν σε μεγαλύτερη παραμόρφωση. Δηλαδή, πιο επιθετική κβάντιση έχει ως αποτέλεσμα μεγαλύτερο θόρυβο κβάντισης. Το γεγονός αυτό επιβεβαιώνει το γεγονός ότι η κβάντιση αποτελεί ένα κρίσιμο βήμα κωδικοποίησης

της πηγής έχοντας σημαντική επιρροή τόσο στο ρυθμό μετάδοσης όσο και στην παραμόρφωση του κωδικοποιητή. Όπως ήδη αναφέρθηκε νωρίτερα, η θεώρηση που πραγματοποιήθηκε σχετικά με την περιττή συμμετρία της χαρακτηριστικής εισόδου – εξόδου ως προς τον άξονα χ, συνεπάγεται ότι η μέση τιμή της τυχαίας μεταβλητής x είναι ίση με μηδέν, δηλαδή, $E(x) = 0$. Επομένως το μέσο τετραγωνικό σφάλμα κβάντισης MSE_q είναι η διασπορά της συνάρτησης του θορύβου κβάντισης, δηλαδή, $MSE_q = \sigma_q^2$.

Ο θόρυβος κβάντισης που σχετίζεται με τον midtread κβαντιστή που εμφανίστηκε στο Σχήμα 4.1 παρουσιάζεται στο Σχήμα 4.3. Είναι εμφανές ότι ο θόρυβος κβάντισης εξαρτάται από το σήμα. Παρατηρούμε ότι σε συνδυασμό με τα εσωτερικά διαστήματα, ο θόρυβος κβάντισης φράσσεται από το $\pm\Delta$. Ο τύπος αυτός θορύβου κβάντισης αναφέρεται ως κοκκώδης θόρυβος (granular noise). Ο θόρυβος που σχετίζεται με το δεξιότερο και αριστερότερο διάστημα δεν φράσσεται καθώς η είσοδος x τείνει στο $\pm\infty$. Αυτός ο τύπος κβάντισης ονομάζεται θόρυβος υπερφόρτωσης (overload noise).



Σχήμα 4.3 : Θόρυβος κβάντισης του Ομοιόμορφου midtread Κβαντιστή

Συμβολίζοντας τον μέσο τετραγωνικό κοκκώδη θόρυβο με $MSE_{q,g}$ και τον μέσο τετραγωνικό θόρυβο υπερφόρτωσης με $MSE_{q,o}$ ισχύουν οι ακόλουθες σχέσεις:

$$MSE_q = MSE_{q,g} + MSE_{q,o} \tag{4.4}$$

Καθώς επίσης και

$$MSE_{q,g} = \sum_{i=2}^{N-1} \int_{d_i}^{d_{i+1}} \left(x - Q(x)\right)^2 f_x(x)dx \tag{4.5}$$

$$MSE_{q,o} = 2 \int_{d_1}^{d_2} \left(x - Q(x)\right)^2 f_x(x)dx \tag{4.6}$$

4.2.3 Σχεδίαση Καβντιστή

Η σχεδίαση ενός κβαντιστή (ομοιόμορφου ή μη-ομοιόμορφου) περιλαμβάνει την επιλογή του πλήθους των σταθμών ανακατασκευής, N (συνεπώς και το πλήθος των επιπέδων απόφασης $N + 1$) καθώς και την επιλογή των τιμών των επιπέδων απόφασης (όσων αφορά το σημείο τοποθέτησης τους. Απλούστερα, ο σχεδιασμός ενός κβαντιστή ισούται με τον καθορισμό της χαρακτηριστικής εισόδου – εξόδου.

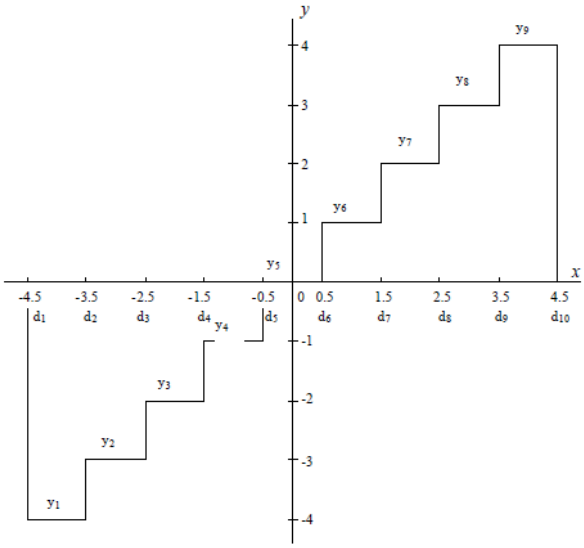
Το πρόβλημα σχεδιασμού ενός βέλτιστου κβαντιστή μπορεί να οριστεί ως εξής. Δοθείσας της συνάρτησης πυκνότητας πιθανότητας μίας τυχαίας μεταβλητής εισόδου $f_x(x)$, καθορίστε το πλήθος των σταθμών ανακατασκευής, N, επιλέξτε ένα σύνολο επιπέδων απόφασης $\{d_i, i = 1, \dots, N + 1\}$ και ενός συνόλου σταθμών ανακατασκευής $\{y_i, i = 1, \dots, N\}$ έτσι ώστε το μέσο τετραγωνικό σφάλμα κβάντισης MSE_q να ελαχιστοποιείται[18].

Στη σχεδίαση του ομοιόμορφου κβαντιστή, ο συνολικός αριθμός επιπέδων κβάντισης, N, συνήθως είναι δεδομένος. Σύμφωνα με τα δύο χαρακτηριστικά του ομοιόμορφου κβαντιστή, είναι γνωστό ότι οι στάθμες ανακατασκευής στον ομοιόμορφο κβαντιστή μπορούν να υπολογιστούν από τα επίπεδα απόφασης. Επομένως, μόνο ένα εκ των δύο συνόλων είναι ανεξάρτητο. Επιπλέον, τόσο τα επίπεδα απόφασης όσο και οι στάθμες ανακατασκευής έχουν τοποθετηθεί σε όμοια διαστήματα εκτός ίσως από τα ακραία διαστήματα. Οι περιορισμοί αυτοί σε συνδυασμό με την θεώρηση συμμετρίας οδηγούν στην εξής παρατήρηση: Υπάρχει μόνο μία παράμετρος που χρειάζεται να καθοριστεί στον ομοιόμορφο κβαντιστή και αυτή είναι το βήμα κβάντισης Δ. Συνεπώς στη σχεδίαση του βέλτιστου ομοιόμορφου κβαντιστή, διαφορετική συνάρτηση πυκνότητας πιθανότητας οδηγεί σε διαφορετικό βήμα κβάντισης.

Στη συνέχεια θα μελετηθεί ο βέλτιστος ομοιόμορφος κβαντιστής τόσο όταν η είσοδος ακολουθεί την ομοιόμορφη κατανομή όσο και διαφορετικές πιθανοτικές κατανομές.

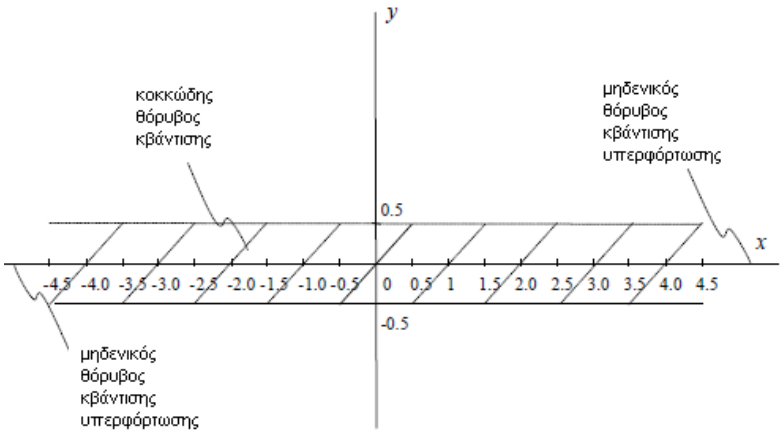
4.2.4 Ομοιόμορφος Κβαντιστής με Ομοιόμορφα Κατανεμημένη Είσοδο

Θεωρούμε τη χαρακτηριστική εισόδου – εξόδου του midtread κβαντιστή με 9 στάθμες ανακατασκευής όπως αυτός παρουσιάστηκε στο Σχήμα 4.1. Θεωρούμε επίσης ότι η είσοδος x είναι μία ομοιόμορφα κατανεμημένη τυχαία μεταβλητή. Η χαρακτηριστική εισόδου – εξόδου παρουσιάζεται στο Σχήμα 4.4.



Σχήμα 4.4 : Χαρακτηριστική Εισόδου – Εξόδου ενός Ομοιόμορφου midtread Κβαντιστή με Ομοιόμορφα Κατανεμημένη Είσοδο στο Διάστημα $[-4.5, 4.5]$ με N=9

Παρατηρείται ότι η νέα χαρακτηριστική περιορίζεται εντός του πεπερασμένου διαστήματος του x το οποίο είναι $-4.5 \leq x \leq 4.5$. Αυτό προκύπτει από τον ορισμό της ομοιόμορφης κατανομής. Συνεπώς ο θόρυβος κβάντισης υπερφόρτωσης δεν υφίσταται στην περίπτωση αυτή όπως φαίνεται και στο Σχήμα 4.5.



Σχήμα 4.5 : Θόρυβος Κβάντισης του Κβαντιστή που Παρουσιάστηκε στο Σχήμα 4.4

Το μέσο τετραγωνικό σφάλμα κβάντισης δίνεται από τις σχέσεις:

$$MSE_q = N \int_{d_1}^{d_2} \left(x - Q(x)\right)^2 \frac{1}{N\Delta} dx \tag{4.7}$$

$$MSE_q = \frac{\Delta^2}{12} \tag{4.8}$$

Το αποτέλεσμα αυτό δείχνει πως εάν η είσοδος ενός ομοιόμορφου κβαντιστή έχει ομοιόμορφη κατανομή και το πλήθος των σταθμών ανακατασκευής είναι δεδομένο, τότε το μέσο τετραγωνικό σφάλμα κβάντισης είναι ανάλογο του τετραγώνου του βήματος κβάντισης. Όσο μεγαλύτερο το βήμα τόσο μεγαλύτερο το σφάλμα κβάντισης γεγονός που συμφωνεί με την αρχική θεώρηση πως ισχυρή κβάντιση οδηγεί σε μεγάλα σφάλματα κβάντισης.

Όπως προαναφέρθηκε, το μέσο τετραγωνικό σφάλμα κβάντισης ισούται με τη διασπορά του θορύβου κβάντισης, δηλαδή $MSE_q = \sigma_q^2$. Προκειμένου να βρεθεί ο λόγος σήματος προς θόρυβο στην περίπτωση του ομοιόμορφου κβαντιστή, απαιτείται να καθοριστεί η διασπορά της εισόδου x. Επισημαίνεται ότι η είσοδος x είναι μία ομοιόμορφη τυχαία μεταβλητή με μηδενική μέση τιμή. Επομένως σύμφωνα με τη θεωρία πιθανοτήτων ισχύει ότι:

$$\sigma_x^2 = \frac{(N\Delta)^2}{12} \tag{4.9}$$

Συνεπώς ο λόγος του μέσου τετραγωνικού σήματος προς θόρυβο SNR_{ms} είναι:

$$SNR_{ms} = 10 \log_{10} \frac{\sigma_x^2}{\sigma_q^2} = 10 \log_{10} N^2 \tag{4.10}$$

Ο δείκτης ms χρησιμοποιείται για να υποδηλώσει ότι αναφερόμαστε στον λόγο σήματος προς θόρυβο υπό την έννοια του μέσου τετραγώνου. Εάν θεωρηθεί ότι $N = 2^n$ τότε:

$$SNR_{ms} = 20 \log_{10} 2^n = 6.02 \cdot n\, db$$

Η ερμηνεία των παραπάνω αποτελεσμάτων είναι η εξής. Εάν γινόταν χρήση του φυσικού δυαδικού κώδικα, οι στάθμες ανακατασκευής του ομοιόμορφου κβαντιστή με ομοιόμορφη κατανομή της πηγής εισόδου τότε κάθε bit που αυξάνεται στην κωδική λέξη προκαλεί μία αύξηση της τάξης των $6.02 \cdot n\, db$ στο SNR_{ms}. Αντίστοιχο συμπέρασμα προκύπτει και την εξίσωση (4.8) σύμφωνα με την οποία όταν το βήμα κβάντισης ενός

ομοιόμορφου κβαντιστή υποδιπλασιάζεται τότε το μέσο τετραγωνικό σφάλμα κβάντισης υπό-τετραπλασιάζεται.

4.2.5 Συνθήκες του Βέλτιστου κβαντιστή

Οι συνθήκες κάτω από τις οποίες το μέσο τετραγωνικό σφάλμα κβάντισης MSE_q ελαχιστοποιείται, βρίσκεται δοθείσης της συνάρτησης πυκνότητας πιθανότητας της εισόδου του κβαντιστή $f_x(x)$.

Αναφέρθηκε νωρίτερα η σχέση σύμφωνα με την οποία υπολογίζεται το μέσο τετραγωνικό σφάλμα κβάντισης. Οι απαραίτητες συνθήκες για τον βέλτιστο κβαντιστή ως προς μέσο τετραγωνικό σφάλμα υπολογίζονται εξισώνοντας την παράγωγο του MSE_q με το μηδέν ως προς d_i και y_i.

$$(d_i - y_{i-1})^2 f_x(d_i) - (d_i - y_i)^2 f_x(d_i) = 0, \qquad i = 2, \dots, N \tag{4.11}$$

$$-\int_{d_i}^{d_i+1} (x - y_i) f_x(x) dx = 0, \qquad i = 1, \dots, N \tag{4.12}$$

Οι ικανές συνθήκες μπορούν επίσης να υπολογιστούν χρησιμοποιώντας τη δεύτερη παράγωγο και η παραδοχή της συμμετρίας υπεισέρχεται και σε αυτό το σημείο. Οι συνθήκες είναι οι ακόλουθες:

$$x_1 = -\infty \; \kappa\alpha\iota \; x_{N+1} = +\infty \tag{4.13}$$

$$\int_{d_i}^{d_{i+1}} (x - y_i) f_x(x) dx = 0, \qquad i = 1,2,..,N \tag{4.14}$$

$$d_i = \frac{1}{2}(y_{i-1} + y_i), \qquad i = 1,2, \dots, N \tag{4.15}$$

Σημειώνεται ότι η πρώτη συνθήκη ισχύει για είσοδο με εύρος $-\infty < x < +\infty$. Οι παραπάνω συνθήκες ερμηνεύονται ως εξής. Κάθε επίπεδο απόφασης (εκτός των ακραίων) είναι ο αριθμητικός μέσος όρος των δύο γειτονικών σταθμών ανακατασκευής και κάθε στάθμη ανακατασκευής είναι το κεντροειδές του εμβαδού κάτω από τη συνάρτηση πυκνότητας πιθανότητας $f_x(x)$ ανάμεσα σε δύο γειτονικά επίπεδα απόφασης. Οι παραπάνω συνθήκες είναι γενικές και επιβάλλονται περιορισμοί σε ότι αφορά τη συνάρτηση πυκνότητας πιθανότητας. Στη συνέχεια θα πραγματοποιηθεί αναφορά στο βέλτιστο ομοιόμορφο κβαντιστή όταν η είσοδος ακολουθεί διαφορετικές

κατανομές.

4.2.6 Βέλτιστος Ομοιόμορφης Κβαντιστής με Διαφορετικές Κατανομές Εισόδου

Στην περίπτωση σχεδίασης του βέλτιστου κβαντιστή με είσοδο που ακολουθεί την ομοιόμορφη κατανομή, τα ακραία διαστήματα είναι πεπερασμένα. Στην ομοιόμορφη κατανομή η εξίσωση (4.14) συνεπάγεται ότι κάθε στάθμη ανακατασκευής είναι ο αριθμητικός μέσος όρος των δύο αντίστοιχων επιπέδων απόφασης. Ανατρέχοντας στα δύο χαρακτηριστικά του ομοιόμορφου κβαντιστή παρατηρούμε ότι αυτός ο κβαντιστής είναι βέλτιστος όταν η είσοδος ακολουθεί την ομοιόμορφη κατανομή.

Όταν η είσοδος x είναι ομοιόμορφα κατανεμημένη στο διάστημα $[-1, 1]$, το βήμα κβάντισης Δ του βέλτιστου ομοιόμορφου κβαντιστή παρουσιάζει ο Πίνακας 4.1 όταν οι στάθμες ανακατασκευής, N, είναι $2, 4, 8, 16\ \kappa\alpha\iota\ 32$. Από τον πίνακα αυτόν παρατηρείται ότι το MSE_q ενός ομοιόμορφου κβαντιστή με ομοιόμορφη κατανομή της εισόδου υποτετραπλασιάζεται καθώς το N διπλασιάζεται. Όπως ήδη αναφέρθηκε αυτό ισοδυναμεί με αύξηση του SNR_{ms} κατά $6.02db$ για κάθε διπλασιασμό του N.

Η παραπάνω είναι μία ειδική περίπτωση, δηλαδή ο ομοιόμορφης κβαντιστής είναι βέλτιστος εφόσον η είσοδος ακολουθεί την ομοιόμορφη κατανομή. Συνήθως, εάν συνάρτηση πυκνότητας πιθανότητας δεν είναι ομοιόμορφη τότε ο βέλτιστος κβαντιστής δεν είναι ένας ομοιόμορφος κβαντιστής. Παρόλα αυτά σε κάποιες περιπτώσεις, λόγω της απλότητας στη σχεδίαση του ομοιόμορφου κβαντιστή, χρησιμοποιείται ο βέλτιστος ομοιόμορφος κβαντιστής για εισόδους που ακολουθούν διαφορετικές κατανομές.

Έχοντας υπόψη τα παραπάνω αντιλαμβανόμαστε ότι οι εξισώσεις (4.13), (4.14) και (4.15) δεν αποτελούν ένα σύστημα εξισώσεων το οποίο μπορεί να λυθεί με ευκολία. Αριθμητικές διαδικασίες έχουν προταθεί για την εύρεση λύσης του σχεδιασμού του βέλτιστου ομοιόμορφου κβαντιστή. Ο Max το 1960 υπολόγισε το βήμα ομοιόμορφης κβάντισης Δ όταν η είσοδος ακολουθεί Gaussian κατανομή. Οι Paez και Glisson[20] το 1972 υπολόγισαν το βήμα Δ για σήματα εισόδου που ακολουθούν τη Laplacian και την κατανομή Gamma. Τα αποτελέσματα αυτά παρουσιάζει ο Πίνακας 4.1. Σημειώνεται ότι και οι τρείς κατανομές έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Εάν η μέση τιμή δεν ήταν μηδενική, απαιτείται μόνο μία ολίσθηση της εισόδου όταν εφαρμόζονται τα αποτελέσματα αυτά. Εάν η τυπική απόκλιση δεν είναι μοναδιαία τότε το βήμα που αναγράφεται στον πίνακα πρέπει να πολλαπλασιαστεί με την τυπική απόκλιση. Το θεωρητικό MSE_q παρατίθεται επίσης στον πίνακα.

Πίνακας 4.1 : Ο Βέλτιστος, Συμμετρικός, Ομοιόμορφος Κβαντιστής για την Ομοιόμορφη, τη Gaussian, τη Laplacian και τη Gamma κατανομή

	Uniform			Gaussian			Laplacian			Gamma		
N	d_i	y_i	MSE	d_i	y_i	MSE	d_i	y_i	MSE	d_i	y_i	MSE
2	-1.000 0.000 1.000	-0.500 0.500	8.33 $\times 10^{-2}$	-1.596 0.000 1.596	-0.798 0.798	0.363	-1.414 0.000 1.414	-0.707 0.707	0.500	-1.154 0.000 1.154	-0.577 0.577	0.668
4	-1.000 -0.500 0.000 0.500 1.000	-0.750 -0.250 0.250 0.750	2.08 $\times 10^{-2}$	-1.991 -0.996 0.000 0.996 1.991	-1.494 -0.498 0.498 1.494	0.119	-2.174 -1.087 0.000 1.087 2.174	-1.631 -0.544 0.544 1.631	1.963 $\times 10^{-1}$	-2.120 -1.060 0.000 1.060 2.120	-1.590 -0.530 0.500 1.590	0.320
8	-1.000 -0.750 -0.500 -0.250 0.000 0.250 0.500 0.750 1.000	-0.875 -0.625 -0.375 -0.125 0.125 0.375 0.625 0.875	5.21 $\times 10^{-3}$	-2.344 -1.758 -1.172 -0.586 0.000 0.586 1.172 1.758 2.344	-2.051 -1.465 -0.879 -0.293 0.293 0.879 1.465 2.051	3.74 $\times 10^{-2}$	-2.924 -2.193 -1.462 -0.731 0.000 0.731 1.462 2.193 2.924	-2.559 -1.828 -1.097 -0.366 0.366 1.097 1.828 2.559	7.17 $\times 10^{-2}$	-3.184 -2.388 -1.592 -0.796 0.000 0.796 1.592 2.388 3.184	-2.786 -1.990 -1.194 -0.398 0.398 1.194 1.990 2.786	0.132
16	-1.000 -0.875 -0.750 -0.625 -0.500 -0.375 -0.250 -0.125 0.000 0.125 0.250 0.375 0.500 0.625 0.750 0.875 1.000	-0.938 -0.813 -0.688 -0.563 -0.438 -0.313 -0.188 -0.063 0.063 0.188 0.313 0.438 0.563 0.688 0.813 0.938	1.30 $\times 10^{-3}$	-2.680 -2.345 -2.010 -1.675 -1.340 -1.005 -0.670 -0.335 0.000 0.335 0.670 1.005 1.340 1.675 2.010 2.345 2.680	-2.513 -2.178 -1.843 -1.508 -1.173 -0.838 -0.503 -0.168 0.168 0.503 0.838 1.173 1.508 1.843 2.178 2.513	1.15 $\times 10^{-2}$	-3.648 -3.192 -2.736 -2.280 -1.824 -1.368 -0.912 -0.456 0.000 0.456 0.912 1.368 1.824 2.280 2.736 3.192 3.648	-3.420 -2.964 -2.508 -2.052 -1.596 -1.140 -0.684 -0.228 0.228 0.684 1.140 1.596 2.052 2.508 2.964 3.420	2.54 $\times 10^{-2}$	-4.320 -3.780 -3.240 -2.700 -2.160 -1.620 -1.080 -0.540 0.000 0.540 1.080 1.620 2.160 2.700 3.240 3.780 4.320	-4.050 -3.510 -2.970 -2.430 -1.890 -1.350 -0.810 -0.270 0.270 0.810 1.350 1.890 2.430 2.970 3.510 4.050	5.01 $\times 10^{-2}$

4.2.7 Βέλτιστος μη-ομοιόμορφος κβαντιστής

Δεν είναι δύσκολο να αντιληφθεί κανείς πως εκτός από την ειδική περίπτωση όπου η μεταβλητή εισόδου x ακολουθεί την ομοιόμορφη κατανομή, οι βέλτιστοι κβαντιστές θα είναι μη-ομοιόμορφοι.

Θεωρούμε την περίπτωση όπου η τυχαία μεταβλητή εισόδου ακολουθεί την Gaussian κατανομή με μηδενική μέση τιμή και μοναδιαία διασπορά ενώ το πλήθος των σταθμών ανακατασκευής είναι πεπερασμένο. Λογικά σκεπτόμενοι, εάν επιλεχθεί να τοποθετηθούν πιο πυκνά τα επίπεδα απόφασης στο κέντρο του άξονα $x = 0$ (περιοχή μεγάλης πυκνότητας πιθανότητας) και πιο αραιά όσο απομακρυνόμαστε από το κέντρο του άξονα x (περιοχή μικρής πυκνότητας πιθανότητας) θα οδηγηθούμε σε μικρότερο MSE_q. Συνεπώς αυτή είναι μία στρατηγική η οποία υιοθετήθηκε.

Οι συνθήκες του βέλτιστου κβαντιστή παρουσιάστηκαν παραπάνω. Στη γενική περίπτωση οι εξισώσεις αυτές δε λύνονται σε κλειστό τύπο. Οι Lloyd[19] και Max[18] πρότειναν μία επαναληπτική διαδικασία προκειμένου να λυθούν αριθμητικά οι εξισώσεις αυτές. Για το λόγο αυτό οι βέλτιστοι αυτοί κβαντιστές ονομάζονται κβαντιστές Lloyd-

Max.

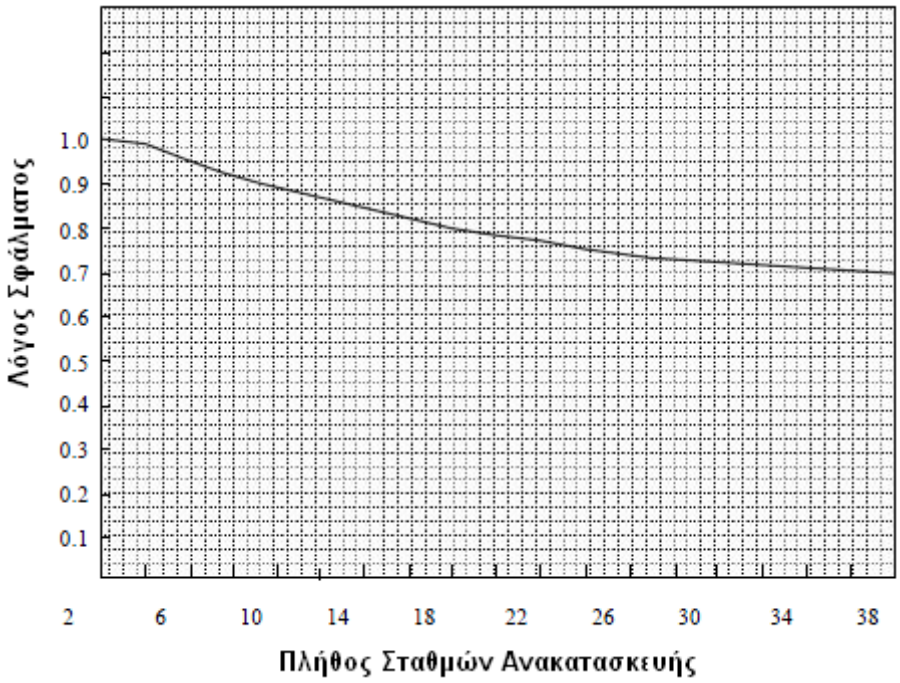
Η λύση στη σχεδίαση του βέλτιστου κβαντιστή βρέθηκε για ένα πλήθος διαφορετικών σταθμών ανακατασκευής N, όταν η είσοδος x ακολουθεί την Gaussian κατανομή. Αυτό πρακτικά σημαίνει πως υπολογίστηκαν τα επίπεδα απόφασης και οι στάθμες ανακατασκευής σε συνδυασμό με το ελάχιστο MSE_q και το βέλτιστο SNR. Χρησιμοποιώντας τη μέθοδο αυτή οι παράμετροι για τη σχεδίαση των κβαντιστών όταν η είσοδος ακολουθεί τις κατανομές Laplacian και Gamma καθορίστηκαν από τους Paez και Glisson το 1972 και τα αποτελέσματα παρουσιάζει ο Πίνακας 4.2. Όπως παρατηρήθηκε και νωρίτερα η ομοιόμορφη κβάντιση είναι βέλτιστη όταν η είσοδος x είναι μία τυχαία μεταβλητή που ακολουθεί την ομοιόμορφη κατανομή.

Πίνακας 4.2 : Ο Βέλτιστος, Συμμετρικός, μη-Ομοιόμορφος Κβαντιστής για την Ομοιόμορφη, τη Gaussian, τη Laplacian και τη Gamma κατανομή

	Uniform			Gaussian			Laplacian			Gamma		
N	d_i	y_i	MSE	d_i	y_i	MSE	d_i	y_i	MSE	d_i	y_i	MSE
2	-1.000 0.000 1.000	-0.500 0.500	8.33×10^{-2}	-∞ 0.000 ∞	-0.799 0.799	0.363	-∞ 0.000 ∞	-0.707 0.707	0.500	-∞ 0.000 ∞	-0.577 0.577	0.668
4	-1.000 -0.500 0.000 0.500 1.000	-0.750 -0.250 0.250 0.750	2.08×10^{-2}	-∞ -0.982 0.000 -0.982 ∞	-1.510 -0.453 0.453 1.510	0.118	-∞ -1.127 0.000 1.127 ∞	-1.834 -0.420 0.420 1.834	1.765×10^{-1}	-∞ -1.205 0.000 1.205 ∞	-2.108 -0.302 0.302 2.108	0.233
8	-1.000 -0.750 -0.500 -0.250 0.000 0.250 0.500 0.750 1.000	-0.875 -0.625 -0.375 -0.125 0.125 0.375 0.625 0.875	5.21×10^{-3}	-∞ -1.748 -1.050 -0.501 0.000 0.501 1.050 1.748 ∞	-2.152 -1.344 -0.756 -0.245 0.245 0.756 1.344 2.152	3.45×10^{-2}	-∞ -2.377 -1.253 -0.533 0.000 0.533 1.253 2.377 ∞	-3.087 -1.673 -0.833 -0.233 0.233 0.833 1.673 3.087	5.48×10^{-2}	-∞ -2.872 -1.401 -0.504 0.000 0.504 1.401 2.872 ∞	-3.799 -1.944 -0.859 -0.149 0.149 0.859 1.944 3.799	7.12×10^{-2}
16	-1.000 -0.875 -0.750 -0.625 -0.500 -0.375 -0.250 -0.125 0.000 0.125 0.250 0.375 0.500 0.625 0.750 0.875 1.000	-0.938 -0.813 -0.688 -0.563 -0.438 -0.313 -0.188 -0.063 0.063 0.188 0.313 0.438 0.563 0.688 0.813 0.938	1.30×10^{-3}	-∞ -2.401 -1.844 -1.437 -1.099 -0.800 -0.522 -0.258 0.000 0.258 0.522 0.800 1.099 1.437 1.844 2.401 ∞	-2.733 -2.069 -1.618 -1.256 -0.942 -0.657 -0.388 -0.128 0.128 0.388 0.657 0.942 1.256 1.618 2.069 2.733	9.50×10^{-3}	-∞ -3.605 -2.499 -1.821 -1.317 -0.910 -0.566 -0.266 0.000 0.266 0.566 0.910 1.317 1.821 2.499 3.605 ∞	-4.316 -2.895 -2.103 -1.540 -1.095 -0.726 -0.407 -0.126 0.126 0.407 0.726 1.095 1.540 2.103 2.895 4.316	1.54×10^{-2}	-∞ -5.050 -3.407 -2.372 -1.623 -1.045 -0.588 -0.229 0.000 0.229 0.588 1.045 1.623 2.372 3.407 5.050 ∞	-6.085 -4.015 -2.798 -1.945 -1.300 -0.791 -0.386 -0.072 0.072 0.386 0.791 1.300 1.945 2.798 4.015 6.085	1.96×10^{-2}

Το Σχήμα 4.6 παρέχει τη συγκριτική απόδοση ανάμεσα στον βέλτιστο ομοιόμορφο κβαντιστή και τον βέλτιστο κβαντιστή στην περίπτωση όπου η είσοδος ακολουθεί τη Gaussian κατανομή με μέση τιμή μηδέν και μοναδιαία διασπορά. Η τετμημένη αναπαριστά το πλήθος των σταθμών ανακατασκευής, N, και η τεταγμένη το λόγο ανάμεσα στο σφάλμα του βέλτιστου κβαντιστή και στο σφάλμα του βέλτιστου ομοιόμορφου κβαντιστή. Μπορεί να δειχθεί πως όταν το N είναι μικρός αριθμός ο λόγος

πλησιάζει στη μονάδα. Αυτό συμβαίνει γιατί σε αυτή την περίπτωση οι επιδόσεις κυμαίνονται σε παραπλήσια επίπεδα. Όταν το N αυξάνει ο λόγος μειώνεται. Πιο συγκεκριμένα, όταν το N είναι μεγάλο ο μη-ομοιόμορφος κβαντιστής είναι κατά 20% με 30% πιο αποδοτικός σε σχέση με τον βέλτιστο ομοιόμορφο κβαντιστή για τη Gaussian κατανομή εισόδου με μηδενική μέση τιμή και μοναδιαία διασπορά.



Σχήμα 4.6 : Ο Λόγος Ανάμεσα στο Σφάλμα του Βέλτιστου Κβαντιστή και το Σφάλμα του Βέλτιστου Ομοιόμορφου Κβαντιστή προς το Πλήθος των Σταθμών Ανακατασκευής N

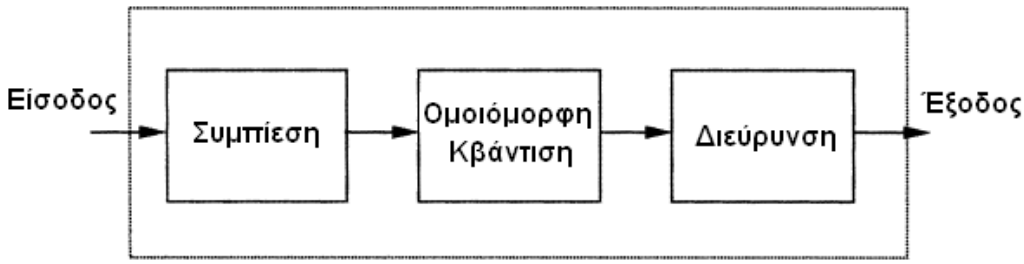
4.2.8 Κβάντιση με Διεύρυνση του Δυναμικού Εύρος (Companding)

Είναι γνωστό ότι το σήμα ομιλίας έχει συνήθως μεγάλο δυναμικό εύρος. Επιπλέον η στατιστική του κατανομή φανερώνει ότι συνήθως σε κάθε επικοινωνία με χρήση φωνής επικρατούν χαμηλές εντάσεις. Πιο συγκεκριμένα υπάρχει 50% πιθανότητα η τάση που χαρακτηρίζει την ενέργεια του ανιχνευθέντος σήματος να είναι μικρότερη από 25% της ρίζας της μέσης τετραγωνικής τιμής του σήματος. Μεγάλες τιμές εντάσεων παρατηρούνται μόλις στο 15% των περιπτώσεων. Τα στατιστικά αυτά οδηγούν συνήθως στην ανάγκη για μη-ομοιόμορφη κβάντιση με σχετικά πυκνά επίπεδα αποφάσεων σε μικρό εύρος τιμών και σχετικά αραιά επίπεδα απόφασης σε μεγάλο εύρος τιμών.

Όταν ο ρυθμός δεδομένων είναι 8 bit ανά δείγμα η ακόλουθη τεχνική διεύρυνσης[22] του δυναμικού εύρους η οποία θεωρεί μη-ομοιόμορφη κβάντιση εμφανίζεται εξαιρετικά χρήσιμη. Καθότι στο πλαίσιο αυτής της εργασίας δε θα ασχοληθούμε με την

κωδικοποίηση ομιλίας, η τεχνική της διεύρυνσης του δυναμικού εύρους θα περιγραφεί επιγραμματικά ως άλλη μια τεχνική επίτευξης μη-ομοιόμορφης κβάντισης.

Η τεχνική διεύρυνσης του δυναμικού εύρους, γνωστή επίσης και ως λογαριθμική κβάντιση, αποτελείται από τα εξής τρία στάδια: Συμπίεση, ομοιόμορφη κβάντιση και διεύρυνση όπως φαίνεται στο Σχήμα 4.7. Αρχικά το σήμα εισόδου συμπιέζεται με μία λογαριθμική χαρακτηριστική και στη συνέχεια κβαντίζεται χρησιμοποιώντας έναν ομοιόμορφο κβαντιστή. Τέλος, τα αποτελέσματα που προέκυψαν κατά την ομοιόμορφη κβάντιση επεκτείνονται αντιστρόφως. Παρουσίαση των χαρακτηριστικών των τριών αυτών σταδίων καθώς και του αποτελέσματος της μη-ομοιόμορφης κβάντισης πραγματοποιείται στο Σχήμα 4.8.



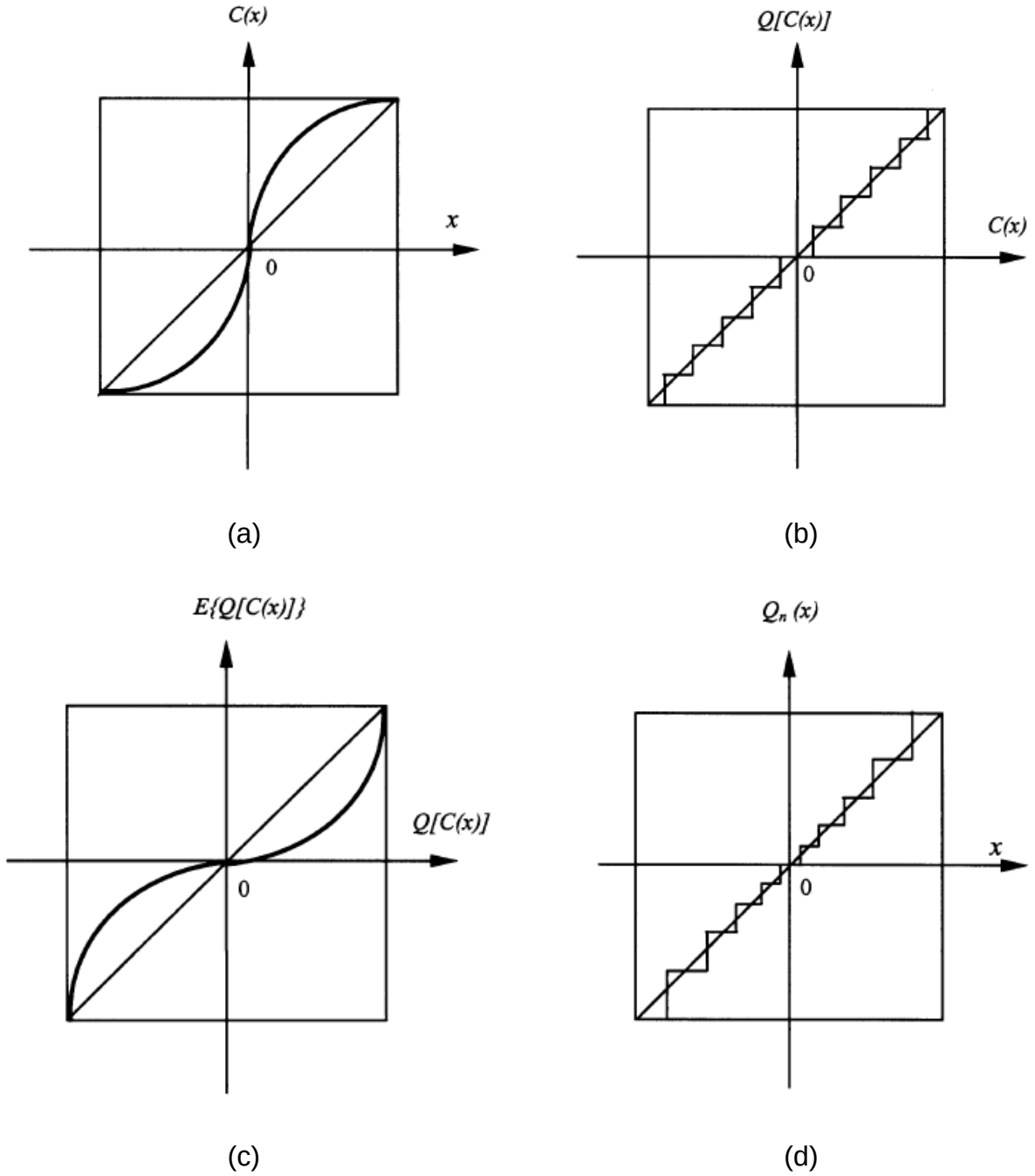
Σχήμα 4.7 : Τεχνική Κβάντισης Διεύρυνσης του Δυναμικού Εύρους

Στην πράξη, χρησιμοποιείται μία τμηματικά γραμμική προσέγγιση της χαρακτηριστικής λογαριθμικής συμπίεσης. Υπάρχουν δύο διαφορετικές μέθοδοι. Στη νότια Αμερική, η χαρακτηριστική που χρησιμοποιείται βασίζεται στον κανόνα-μ ορίζεται ως εξής:

$$c(x) = x_{max} \frac{\ln\left[1 + \mu\left(\frac{|x|}{x_{max}}\right)\right]}{\ln(1+\mu)} sgn\, x, \tag{4.16}$$

Όπου sgn είναι η συνάρτηση που ορίζεται ως

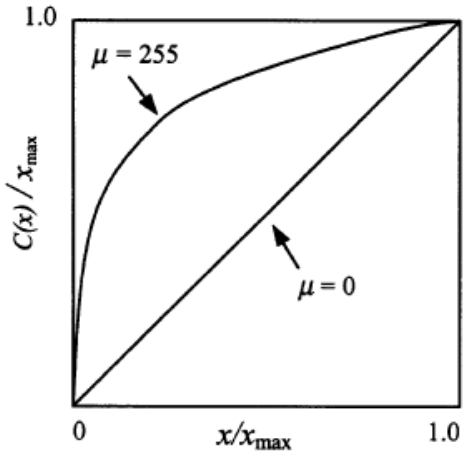
$$sgn\, x = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \tag{4.17}$$



Σχήμα 4.8 : Οι Χαρακτηριστικές της Τεχνικής Κβάντισης Διεύρυνσης του Δυναμικού Εύρους.

(a) Χαρακτηριστική Συμπίεσης. (b) Χαρακτηριστική Ομοιόμορφου Κβαντιστή. (c) Χαρακτηριστική Διεύρυνσης. (d) Χαρακτηριστική μη-ομοιόμορφου κβαντιστή.

Η χαρακτηριστική συμπίεσης που βασίζεται στον κανόνα μ παρουσιάζεται στο Σχήμα 4.9. Η τυπική τιμή για το μ είναι 255. Σημειώνεται όπως προκύπτει και από το σχήμα ότι το $\mu = 0$ αντιστοιχεί στον ομοιόμορφο κβαντιστή.

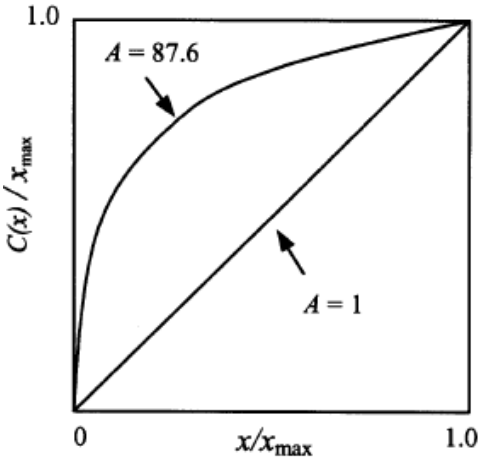


Σχήμα 4.9 : Χαρακτηριστική Συμπίεσης – Κανόνας μ

Στην Ευρώπη χρησιμοποιείται η χαρακτηριστική που βασίζεται στον κανόνα-Α. Η χαρακτηριστική αυτή παρουσιάζεται στο Σχήμα 4.10 και ορίζεται ως εξής:

$$c(x) = \begin{cases} x_{max} \dfrac{\mathrm{A}\left(\dfrac{|x|}{x_{max}}\right)}{1+\ln \mathrm{A}} sgn\ x, & 0 < \dfrac{|x|}{x_{max}} < \dfrac{1}{A} \\ x_{max} \dfrac{1+\ln\left[A\left(\dfrac{|x|}{x_{max}}\right)\right]}{1+\ln \mathrm{A}} sgn\ x, & \dfrac{1}{A} < \dfrac{|x|}{x_{max}} < 1 \end{cases} \tag{4.18}$$

Σημειώνεται ότι η τυπική τιμή του Α είναι 87.6. Η περίπτωση $A = 1$ αντιστοιχεί στην ομοιόμορφη κβάντιση.



Σχήμα 4.10 : Χαρακτηριστική Συμπίεσης – Κανόνας Α

4.2.9 Βασικές Εφαρμογές του Βαθμωτού Κβαντιστή στις Εικόνες

Εφόσον αναφερόμαστε στον βαθμωτό κβαντιστή υπενθυμίζεται η θεώρηση σχετικά με τα προς συμπίεση δεδομένα τα οποία δίνονται υπό τη μορφή αριθμών από μία πηγή.

Ένα πρώτο παράδειγμα αποτελεί η αφελής διακριτή κβάντιση μίας ροής εισόδου αποτελούμενη από αριθμούς των 8 bit. Στην περίπτωση αυτή μπορούν να διαγραφούν τα 4 λιγότερο σημαντικά bit κάθε αριθμού. Αυτή είναι μία σπάνια περίπτωση όπου ο λόγος συμπίεσης θα είναι πάντοτε 2 ανεξαρτήτως από τη στατιστική κατανομή των δεδομένων. Η είσοδος αποτελείται από $2^8 = 256$ διαφορετικά σύμβολα ενώ η έξοδος μόνο από $2^4 = 16$ σύμβολα. Η μέθοδος αυτή είναι απλή αλλά καθόλου πρακτική εξαιτίας της πολύ μεγάλης απώλειας προκειμένου να επιτευχθεί ένας πολύ μέτριος παράγοντας συμπίεσης.

Προκειμένου να αναπτυχθούν καλύτερες προσεγγίσεις[25][26], θεωρούνται και πάλι δεδομένα εισόδου, αριθμοί μη προσημασμένοι μήκους 8 bit. Συνεπώς το εύρος των τιμών είναι $[0, 255]$. Επιλέγεται μία παράμετρος απόστασης s και υπολογίζεται η ακολουθία των ομοιόμορφα κβαντισμένων τιμών $0, s, 2s, \dots, k \cdot s, 255$ με $(k+1)s > 255$ και $k \cdot s \leq 255$. Κάθε σύμβολο εισόδου s κβαντίζεται λαμβάνοντας την τιμή της πλησιέστερης τιμής της ακολουθίας. Επιλέγοντας $s = 3$ παράγεται η ομοιόμορφη ακολουθία $0,3,6,9,12, \dots ,252,255$. Με $s = 4$ έχουμε $0,4,8,12,251,255$ (καθώς $252 + 4 = 256$ οπότε κβαντίζεται στο 255).

Μία παρεμφερής προσέγγιση είναι η επιλογή τιμών κβάντισης οι οποίες για οποιοδήποτε αριθμό εισόδου στο διάστημα $[0, 255]$ δε θα απέχουν περισσότερο από d μονάδες από την τιμή στην οποία θα κβαντιστούν. Αυτό επιτυγχάνεται διαιρώντας το εύρος τιμών σε τμήματα μεγέθους $2d + 1$ κεντράροντας τα στο εύρος $[0, 255]$. Εάν για παράδειγμα το $d = 16$ τότε το διάστημα $[0, 255]$ διαιρείται σε 7 τμήματα μεγέθους 33 μονάδων το καθένα και απομένουν άλλες 25 μονάδες. Επομένως μπορεί να θεωρηθεί ως πρώτο σημείο ο αριθμός 12 και έτσι έχουμε τελικά την ακολουθία 10 αριθμών $12, 33, 45, 78, 111, 144, 177, 210, 243, 255$. Κάθε αριθμός πλέον στο διάστημα $[0, 255]$ απέχει το πολύ 16 μονάδες από κάθε σημείο κβάντισης. Με τη μέθοδο αυτή μπορεί να περιοριστεί ο αριθμός των σταθμών κβάντισης σε 8 (δηλαδή με χρήση τριών bit) λαμβάνοντας την ακολουθία $8, 41, 74, 107, 140, 173, 206, 239$.

Οι κβαντισμένες ακολουθίες που αναφέρθηκαν παραπάνω έχουν νόημα όταν κάθε σύμβολο εισόδου εμφανίζεται με ίση πιθανότητα. Εάν τα δεδομένα εισόδου δεν είναι ομοιόμορφα κατανεμημένα τότε η ακολουθία των κβαντισμένων τιμών πρέπει να

κατανέμεται με τον ίδιο τρόπο όπως τα δεδομένα.

Η βαθμωτή κβάντιση αποτελεί ένα παράδειγμα συμπίεσης με απώλειες όπου είναι εύκολος ο έλεγχος ανάμεσα στο λόγο συμπίεσης και της απώλειας πληροφορίας. Παρόλα αυτά, λόγω της απλότητας της, χρησιμοποιείται μόνο σε πολύ ειδικές περιπτώσεις όπου η μεγάλη απώλεια πληροφορίας είναι ανεκτή. Πολλές μέθοδοι συμπίεσης εικόνας επιτρέπουν την απώλεια πληροφορίας όμως η βαθμωτή κβάντιση δεν είναι κατάλληλη για συμπίεση εικόνας καθώς δημιουργία ενοχλητικές ατέλειες κατά την αποσυμπίεση της εικόνας. Για παράδειγμα, μία εικόνα με σχεδόν ομοιόμορφη κατανομή όπου όλα τα εικονοστοιχεία έχουν τιμές φωτεινότητας 127 ή 128. Εάν η τιμή 127 κβαντιστεί στο 111 και η τιμή 128 στο 144 τότε μετά την αποσυμπίεση της εικόνας ενδέχεται να παρουσιαστεί το φαινόμενο της σκακιέρας αφού γειτονικά εικονοστοιχεία θα εναλλάσσουν την τιμή τους ανάμεσα σε 111 και 144. Αυτός είναι και ο λόγος που στην πράξη χρησιμοποιείται η διανυσματική κβάντιση αντί της βαθμωτής και η οποία θα μελετηθεί στη συνέχεια.

4.3 Διανυσματική Κβάντιση

Η διανυσματική κβάντιση[17] αποτελεί τη γενίκευση της βαθμωτής κβάντισης. Χρησιμοποιείται τόσο στη συμπίεση ήχου όσο και εικόνας. Πρακτικά η διανυσματική κβάντιση χρησιμοποιείται ευρέως στη συμπίεση δεδομένων[24] τα οποία έχουν προκύψει από ψηφιοποίηση μίας αναλογικής πηγής, όπως δειγματοληπτημένο ηχητικό σήμα ή σαρωμένες εικόνες. Τα δεδομένα αυτά ονομάζονται “αναλογικά δεδομένα ψηφιακά δειγματοληπτιμένα” (DSAD). Η διανυσματική κβάντιση βασίζεται σε δύο διαπιστώσεις.

1. Μέθοδοι συμπίεσης που κωδικοποιούν αλφαριθμητικά αποδίδουν καλύτερα σε σχέση με μεθόδους που χρησιμοποιούν μεμονωμένα σύμβολα.
2. Γειτονικά δείγματα δεδομένων σε μία εικόνα (εικονοστοιχεία) ή στον ψηφιακό ήχο είναι συσχετισμένα.

4.3.1 Αρχικές Προσεγγίσεις Διανυσματικής Κβάντισης

Ας ξεκινήσουμε με μία απλή διαισθητική μέθοδο διανυσματικής κβάντισης στη συμπίεση εικόνας. Δοθείσης μίας εικόνας, τη διαιρούμε σε μικρά μπλοκ μεγέθους 2×2 ή 4×4. Κάθε μπλοκ θεωρείται ως ένα διάνυσμα. Ο κωδικοποιητής διατηρεί μία λίστα, γνωστή ως κωδικό-βιβλίο, διανυσμάτων και συμπιέζει κάθε μπλοκ γράφοντας στη συμπιεσμένη ροή δεδομένων έναν δείκτη στο αντίστοιχο μπλοκ του κωδικού-βιβλίου. Ο αποκωδικοποιητής δεν έχει παρά να διαβάζει τους δείκτες και ανατρέχοντας στις

αντίστοιχες θέσεις του κωδικού-βιβλίου να συνθέτει το μπλοκ της εικόνας. Η διανυσματική κβάντιση αποτελεί μία ασύμμετρη μέθοδο συμπίεσης.

Στην περίπτωση όπου χρησιμοποιηθούν μπλοκ μεγέθους 2×2, κάθε μπλοκ (διάνυσμα) αποτελείται από τέσσερα εικονοστοιχεία. Εάν θεωρήσουμε δυαδική εικόνα τότε η τιμή κάθε εικονοστοιχείου είναι ένα bit και επομένως το μπλοκ αποτελείται από τέσσερα bit. Συνολικά υπάρχουν $2^4 = 16$ διαφορετικά μπλοκ. Όταν το κωδικό-βιβλίο είναι τόσο μικρό μπορεί εύκολα να αποθηκευτεί τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Ωστόσο, σε αυτή την περίπτωση για να διευθυνσιοδοτήσουμε κάθε μπλοκ του κωδικού-βιβλιου απαιτούνται 4 bit, δηλαδή όσα ακριβώς απαιτούνται για να κωδικοποιήσουμε το μπλοκ χωρίς δείκτες. Εάν κάθε εικονοστοιχείο αποτελείται από $k\ bit$, τότε κάθε μπλοκ αναπαρίσταται με $4k\ bit$ και συνεπώς υπάρχουν 2^{4k} διαφορετικά μπλοκ. Το κωδικό-βιβλίο αυξάνεται γρήγορα σε σχέση με το k αφού για $k = 8$ υπάρχουν $256^4 = 2^{32} = 4\ Terra$ εγγραφές. Παρόλα αυτά και πάλι εάν αντικατασταθεί ένα μπλοκ των $4k\ bits$ με έναν δείκτη των $4k$ bit προφανώς και πάλι δεν επιτυγχάνεται κάποιο κέρδος συμπίεσης. Αυτό ισχύει για κάθε μέγεθος μπλοκ.

Στο σημείο αυτό είναι πλέον ξεκάθαρο ότι αυτή η απλή μέθοδος δε μπορεί να αποδώσει. Σε μία εικόνα όμως στην πράξη δεν εμφανίζεται κάθε πιθανό μπλοκ. Σε μία εικόνα των 8 bit ανά εικονοστοιχείο το πλήθος των 2×2 μπλοκ είναι $2^{2 \cdot 2 \cdot 8} = 2^{32} \approx 4.3\ \delta\iota\sigma\varepsilon\kappa\tau o\mu\acute{\upsilon}\rho\iota\alpha$ όμως στην πράξη η εικόνα συνήθως δεν ξεπερνά τα μερικά χιλιάδες διαφορετικά μπλοκ. Επομένως η εκδοχή του διανυσματικού κωδικοποιητή που θα δοκιμαστεί θα ξεκινά με ένα κενό κωδικό-βιβλίο και θα σαρώνεται η εικόνα μπλοκ προς μπλοκ. Το κωδικό-βιβλίο θα σαρώνεται προκειμένου να εξετάσει την ύπαρξη κάποιου μπλοκ. Εάν το μπλοκ βρεθεί στο κωδικό-βιβλίο τότε ως έξοδος δίνεται ένας δείκτης στο μπλοκ του κωδικό-βιβλίου. Εάν το μπλοκ δε βρεθεί στο κωδικό-βιβλίο τότε προστίθεται και ο αντίστοιχος δείκτης δίνεται ως έξοδος.

Το πρόβλημα της απλής αυτής μεθόδου είναι ότι κάθε μπλοκ που προστίθεται στο κωδικό-βιβλίο θα πρέπει να εγγράφεται και στη συμπιεσμένη ροή δεδομένων. Αυτό μειώνει σημαντικά την αποδοτικότητα της μεθόδου και ενδέχεται όχι απλώς να επιτευχθεί χαμηλή συμπίεση αλλά ακόμα και να προκύψει ροή δεδομένων μεγαλύτερη από την αρχική. Επίσης υπάρχει μία επιπρόσθετη πολυπλοκότητα καθώς το κωδικό-βιβλίο μεγαλώνει κατά τη διαδικασία της συμπίεσης με αποτέλεσμα να αυξάνεται και το μήκος των δεικτών, διαδικασία όμως την οποία δεν είναι δύσκολο να διαχειριστεί ο αποκωδικοποιητής.

Τα προβλήματα που μόλις περιγράφηκαν αποτελούν ουσιαστικά τον λόγο για τον οποίο η διανυσματική κβάντιση είναι κβάντιση με απώλειες. Εάν δεχτούμε τη συμπίεση με απώλειες τότε το μέγεθος του κωδικού-βιβλίου μπορεί να μειωθεί. Στο σημείο αυτό θα περιγραφεί μία διαισθητική μέθοδος συμπίεσης εικόνας με απώλειες χρησιμοποιώντας διανυσματική κβάντιση. Θεωρούμε ένα πλήθος εικόνων οι οποίες θα χρησιμοποιηθούν ως εικόνες εκπαίδευσης και βρίσκουμε τα n πιο δημοφιλή μπλοκ. Στη συνέχεια δημιουργούμε το κωδικό-βιβλίο με τα n δημοφιλέστερα μπλοκ τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Κάθε εγγραφή στο βιβλίο αναπαριστά και ένα μπλοκ. Προκειμένου να συμπιεστεί μία εικόνα, σαρώνουμε την εικόνα μπλοκ προς μπλοκ και για κάθε ένα μπλοκ ανιχνεύουμε την εγγραφή στο κωδικό-βιβλίο η οποία ταιριάζει περισσότερο στο συγκεκριμένο μπλοκ της εικόνας λαμβάνοντας ως έξοδο έναν δείκτη σε αυτή. Προφανώς το μέγεθος του δείκτη είναι $\lceil \log_2 n \rceil$ συνεπώς ο λόγος συμπίεσης σε αυτή την περίπτωση είναι γνωστός εκ των προτέρων και ισούται με

$$\frac{\lceil \log_2 n \rceil}{\mu\acute{\varepsilon}\gamma\varepsilon\theta o\varsigma\ \mu\pi\lambda o\kappa}$$

Ένα πρόβλημα που προκύπτει από την προσέγγιση αυτή είναι ο τρόπος με τον οποίο θα επιλεγεί η εγγραφή του κωδικό-βιβλίου που ταιριάζει περισσότερο στο μπλοκ της εικόνας. Υπάρχει ένα πλήθος μετρικών με τον οποίο καθορίζεται η έννοια της απόστασης δύο μπλοκ. Έστω $B = (b_1, b_2, \dots, b_n)$ και $C = (c_1, c_2, \dots, c_n)$ τα οποία αναπαριστούν ένα μπλοκ από εικονοστοιχεία μίας εικόνας και μία εγγραφή του κωδικό-βιβλίου αντίστοιχα (καθένα ένα διάνυσμα των n bit). Συμβολίζουμε την απόσταση μεταξύ τους ως $d(B, C)$ και την υπολογίζουμε ως:

$$d_1(B, C) = \sum_{i=0}^{n} |b_i - c_i|, \tag{4.19}$$

$$d_2(B, C) = \sum_{i=0}^{n} (b_i - c_i)^2, \tag{4.20}$$

$$d_3(B, C) = MAX_{i=0}^{n} |b_i - c_i| \tag{4.21}$$

Η τρίτη μετρική $d_3(B, C)$ ερμηνεύεται εύκολα καθώς βρίσκει το στοιχείο εκείνο του οποίου η διαφορά B και C είναι μέγιστη και επιστρέφει τη διαφορά τους. Οι δύο πρώτες μετρικές μπορούν εύκολα να αναπαρασταθούν στην περίπτωση για $n = 3$. Η μετρική $d_1(B, C)$ σε αυτή την περίπτωση γίνεται ουσιαστικά η απόσταση ανάμεσα στα δύο

τρισδιάστατα διανύσματα B και C όταν κινούμαστε κατά μήκος των αξόνων. Η μετρική $d_2(B,C)$ γίνεται σε αυτή την περίπτωση η Ευκλείδεια απόσταση μεταξύ των δύο διανυσμάτων. Οι ποσότητες $d_i(B,C)$ μπορούν επίσης να θεωρηθούν ως μέτρα παραμόρφωσης.

Ένα άλλο πρόβλημα σε αυτή την προσέγγιση είναι η ποιότητα του κωδικού-βιβλίου. Στην περίπτωση όπου τα μπλοκ έχουν μέγεθος 2×2 με 8 bit ανά εικονοστοιχείο τότε το πλήθος των μπλοκ είναι $2^{23} \approx 4.3$ δισεκατομύρια. Εφόσον αποφασιστεί να περιοριστεί το μέγεθος του κωδικού-βιβλίου, για παράδειγμα σε 1 εκατομμύριο εγγραφές, τότε θα περιέχεται μόνο το 0.023% του συνολικού πλήθους των μπλοκ (όπου ακόμα σε αυτή την περίπτωση το μέγεθος του κωδικού-βιβλίου θα είναι 32 εκατομμύρια bit ή αλλιώς 4MB). Χρησιμοποιώντας αυτό το κωδικό-βιβλίο για τη συμπίεση μίας μη-τυπικής εικόνας μπορεί να οδηγηθούμε σε συμπίεση με μεγάλες παραμορφώσεις ανεξαρτήτως της μετρικής που θα επιλέξουμε. Όταν η συμπιεσμένη εικόνα αποσυμπιεστεί είναι πιθανό να διαφέρει σημαντικά σε σχέση με την αρχική καθιστώντας πρακτικά τη μέθοδο ακατάλληλη. Ένας τρόπος για να ξεπεραστεί το πρόβλημα αυτό είναι να τροποποιηθεί το αρχικό κωδικό-βιβλίο προκειμένου να προσαρμοστεί στα μπλοκ της συγκεκριμένης εικόνας. Ασφαλώς το κωδικό-βιβλίο σε αυτή την περίπτωση θα πρέπει να συμπεριληφθεί στη ροή δεδομένων όμως λόγω του περιορισμού του μεγέθους που επιβάλλουμε δε θα είναι ιδιαιτέρως μεγάλο και επομένως η μέθοδος μπορεί να προσφέρει μια ικανοποιητική απόδοση συμπίεσης σε σχετικά ικανοποιητική ποιότητα.

4.3.2 Ο Αλγόριθμος LBG

Έναν αλγόριθμο αυτής της μορφής ανέπτυξαν οι Linde, Buzo και Gray[23]. Είναι γνωστός ως ο αλγόριθμος LBG και αποτελεί τη βάση πολλών μεθόδων διανυσματικής κβάντισης τόσο για συμπίεση εικόνας όσο και ήχου. Τα βασικά του βήματα είναι τα ακόλουθα:

Βήμα 1^{ο}: Επέλεξε την τιμή κατωφλίου ε και θέσε $k = 0$ και $D^{(-1)} = \infty$. Αρχικοποίηση κωδικού-βιβλίου με $C_i^{(k)}$ εγγραφές και $k = 0$ το οποίο θα αυξάνεται σε κάθε επανάληψη. Συμβολίζουμε τα μπλοκ της εικόνας με B_i (τα μπλοκ αυτά ονομάζονται επίσης και διανύσματα εκπαίδευσης αφού ο αλγόριθμος τα χρησιμοποιεί προκειμένου να βρει τις εγγραφές που ταιριάζουν περισσότερο στο κωδικό-βιβλίο).

Βήμα 2^{ο}: Επιλέγουμε μία εγγραφή $C_i^{(k)}$ του κωδικό-βιβλίου. Βρίσκουμε όλα τα μπλοκ B_m της εικόνας τα οποία είναι πλησιέστερα στο C_i σε σχέση με οποιοδήποτε άλλο C_j. Πιο

συγκεκριμένα, βρίσκουμε το σύνολο των B_m που ικανοποιεί τη σχέση

$$d(B_m, C_i) < d\left(B_m, C_j\right), \qquad \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ j \neq i$$

Το σύνολο αυτό συμβολίζεται με $P_i^{(k)}$. Επαναλαμβάνουμε για κάθε i. Υπάρχει το ενδεχόμενο κάποια σύνολα να είναι κενά. Η περίπτωση αυτή αντιμετωπίζεται σε επόμενο βήμα.

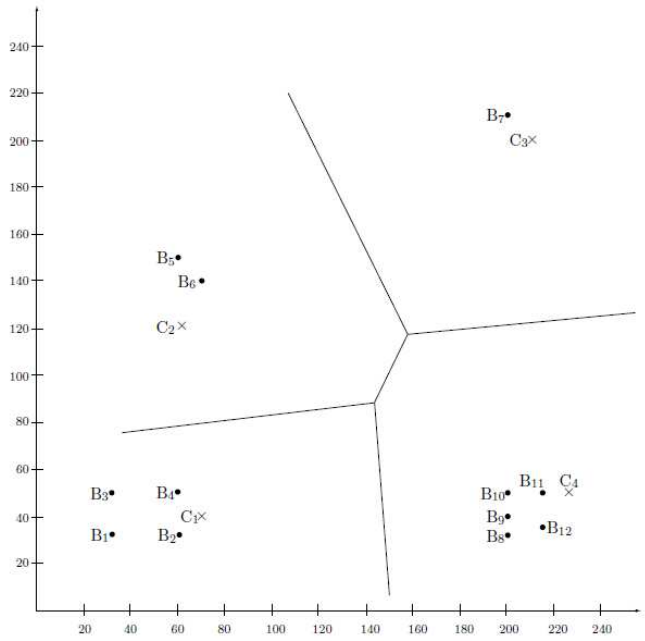
Βήμα 3ο: Επιλέγουμε ένα i και υπολογίζουμε την παραμόρφωση $D_i^{(k)}$ ανάμεσα στην εγγραφή του κωδικού-βιβλίου $C_i^{(k)}$ και στο σύνολο των διανυσμάτων εκπαίδευσης $P_i^{(k)}$ που βρέθηκαν στο βήμα 2. Επαναλαμβάνουμε για κάθε i και στη συνέχεια υπολογίζουμε τη μέση τιμή $D^{(k)}$ για όλα τα $D_i^{(k)}$. Η τιμή της παραμόρφωσης $D_i^{(k)}$ για κάποιο i βρίσκεται υπολογίζοντας τις αποστάσεις $d\left(C_i^{(k)}, B_m\right)$ για κάθε μπλοκ B_m του συνόλου $P_i^{(k)}$ και στη συνέχεια υπολογίζοντας τη μέση τιμή των αποστάσεων. Εναλλακτικά, το $D_i^{(k)}$ μπορεί να τεθεί ως η ελάχιστη απόσταση $d\left(C_i^{(k)}, B_m\right)$.

Βήμα 4ο: Αυξάνουμε το k κατά 1 και υπολογίζουμε τις νέες εγγραφές $C_i^{(k)}$ όπου η κάθε μία ισούται με τη μέση τιμή των μπλοκ της εικόνας (διανύσματα εκπαίδευσης) του συνόλου $P_i^{(k-1)}$ που υπολογίστηκε στο βήμα 2. Με τον τρόπο αυτό το κωδικό-βιβλίο προσαρμόζεται στην εκάστοτε εικόνα. Επαναλαμβάνουμε το βήμα 2.

Για την πλήρη κατανόηση του αλγορίθμου θα ακολουθήσει ένα αναλυτικό παράδειγμα. Προκειμένου να οπτικοποιηθεί ευκολότερα το παράδειγμα, θα θεωρηθεί εικόνα 8-bit ανά εικονοστοιχείο και μπλοκ μεγέθους 2 εικονοστοιχείων. Τυπικά τα μπλοκ πρέπει να είναι τετράγωνα όμως το πλεονέκτημα σε στην περίπτωση των μπλοκ 2 εικονοστοιχείων είναι ότι μπορούμε να τα σχεδιάσουμε ως σημεία του δισδιάστατου χώρου και επομένως να τα αντιληφθούμε καλύτερα. Οι τιμές των δύο εικονοστοιχείων του μπλοκ θα αποτελέσουν τις συντεταγμένες ενός σημείου (x, y).

Στο παράδειγμα θεωρείται εικόνα 24ων εικονοστοιχείων οργανωμένη σε 12 μπλοκ: $B_1 = (32, 32), B_2 = (60, 32), B_3 = (32, 50), B_4 = (60, 50), B_5 = (60, 150), B_6 = (70, 140), B_7 = (200, 210), B_8 = (200, 32), B_9 = (200, 40), B_{10} = (200, 50), B_{11} = (215, 50) \kappa\alpha\iota\ B_{12} = (215, 35)$. Παρατηρώντας και το Σχήμα 4.11 είναι ξεκάθαρο πως τα 12 σημεία συγκεντρώνονται σε τέσσερεις περιοχές. Επιλέγουμε αρχικά ένα κωδικό-βιβλίο με τέσσερεις εγγραφές $C_1^{(0)} = (70, 40), C_2^{(0)} = (60, 120), C_3^{(0)} = (210, 200)$ και $C_4^{(0)} = (225, 50)$ (εμφανίζονται στο σχήμα με το σύμβολο X). Οι εγγραφές αυτές

επιλέχθηκαν τυχαία όμως στη συνέχεια θα παρουσιαστεί ο τρόπος με τον οποίο επιλέγονται οι τιμές αυτές μεθοδικά. Από το σχήμα προκύπτουν πολύ εύκολα τα τέσσερα αρχικά σύνολα τα οποία είναι $P_1^{(0)} = (B_1, B_2, B_3, B_4)$, $P_2^{(0)} = (B_5, B_6)$, $P_3^{(0)} = (B_7)$ και $P_4^{(0)} = (B_8, B_9, B_{10}, B_{11}, B_{12})$.



Σχήμα 4.11 : Δώδεκα Σημεία και Τέσσερεις Εγγραφές $C_i^{(0)}$ στο Κωδικό-Βιβλίο

Ο Πίνακας 4.3 παρουσιάζει τον τρόπο με τον οποίο υπολογίζεται η μέση παραμόρφωση $D^{(0)}$ κατά την πρώτη επανάληψη (με χρήση του μέτρου της ευκλείδειας απόστασης).

Πίνακας 4.3 : Υπολογισμός των 12 παραμορφώσεων για k=0

I:	$(70-32)^2 + (40-32)^2 = 1508$	$(70-60)^2 + (40-32)^2 = 164$
	$(70-32)^2 + (40-50)^2 = 1544$	$(70-60)^2 + (40-50)^2 = 200$
II:	$(60-60)^2 + (120-150)^2 = 900$	$(60-70)^2 + (120-140)^2 = 500$
III:	$(210-200)^2 + (200-210)^2 = 200$	
IV:	$(225-200)^2 + (50-32)^2 = 949$	$(225-200)^2 + (50-40)^2 = 725$
	$(225-200)^2 + (50-50)^2 = 625$	$(225-215)^2 + (50-50)^2 = 100$
	$(225-215)^2 + (50-35)^2 = 325$	

Το αποτέλεσμα είναι το εξής:

$$D^{(0)} = \frac{1508 + 164 + 1544 + 200 + 900 + 500 + 200 + 949 + 725 + 625 + 100 + 325}{12}$$
$$= 645$$

Το τέταρτο βήμα υποδηλώνει πως δεν υπάρχει σύγκλιση, καθώς $D^{(-1)} = \infty$, επομένως αυξάνουμε το k κατά 1 και υπολογίζουμε τις τέσσερεις νέες εγγραφές στο κωδικό-βιβλίο $C_i^{(1)}$:

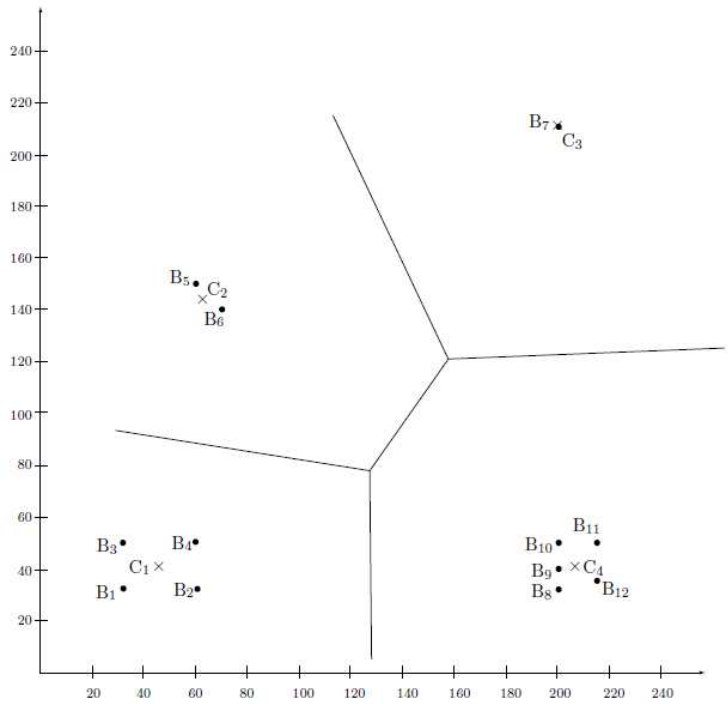
$$C_1^{(1)} = \frac{B_1 + B_2 + B_3 + B_4}{4} = (46, 41)$$

$$C_2^{(1)} = \frac{B_5 + B_6}{2} = (65, 145) \qquad (4.22)$$

$$C_3^{(1)} = B_7 = (200, 210)$$

$$C_4^{(1)} = \frac{B_8 + B_9 + B_{10} + B_{11} + B_{12}}{5} = (206, 41)$$

Τα οποία παρουσιάζονται στο Σχήμα 4.12.



Σχήμα 4.12 : Δώδεκα Σημεία και Τέσσερεις Εγγραφές $C_i^{(1)}$ στο Κωδικό-Βιβλίο

Κατά την επόμενη επανάληψη επιστρέφουμε στο βήμα 2 όπου έχουμε τα ίδια σημεία να συμμετέχουν στο σύνολο κάθε εγγραφής του κωδικό-βιβλίου (αυτό συμβαίνει επειδή τα σημεία είναι συγκεντρωμένα σε τέσσερεις διακριτές περιοχές αλλά στη γενική περίπτωση ένα σύνολο $P_i^{(k)}$ ενδέχεται να αποτελείται από διαφορετικά μπλοκ εικόνας σε κάθε επανάληψη k). Τις παραμορφώσεις που υπολογίζονται στο 3^{ο} βήμα συνοψίζει ο Πίνακας 4.4.

Πίνακας 4.4 : Υπολογισμός των 12 παραμορφώσεων για k=1

I:	$(46-32)^2+(41-32)^2=277$	$(46-60)^2+(41-32)^2=277$
	$(46-32)^2+(41-50)^2=277$	$(46-60)^2+(41-50)^2=277$
II:	$(65-60)^2+(145-150)^2=50$	$(65-70)^2+(145-140)^2=50$
III:	$(210-200)^2+(200-210)^2=200$	
IV:	$(206-200)^2+(41-32)^2=117$	$(206-200)^2+(41-40)^2=37$
	$(206-200)^2+(41-50)^2=117$	$(206-215)^2+(41-50)^2=162$
	$(206-215)^2+(41-35)^2=117$	

Η μέση παραμόρφωση $D_i^{(1)}$ είναι

$$D^{(1)}=\frac{277+277+277+2777+50+50+200+117+37+117+162+117}{12}=163.2$$

η οποία είναι σημαντικά μικρότερη σε σχέση με την αρχική 645. Εάν στο 4^{ο} βήμα δεν υπάρχει σύγκλιση ακολουθούν και άλλες επαναλήψεις προκειμένου να μειωθεί η παραμόρφωση βελτιώνοντας τις τέσσερεις εγγραφές του κωδικό-βιβλίου.

4.3.3 Αλγόριθμος Επιλογής Αρχικών Εγγραφών του Κωδικού-Βιβλίου

Παρά το γεγονός ότι ο αλγόριθμος LBG μειώνει τη μέση παραμόρφωση από επανάληψη σε επανάληψη, δεν εγγυάται ότι οι εγγραφές του κωδικό-βιβλίου συγκλίνουν στη βέλτιστη λύση. Υπάρχει το ενδεχόμενο σύγκλισης σε μία λιγότερο βέλτιστη λύση και από αυτή την άποψη η αποτελεσματικότητα του αλγορίθμου εξαρτάται από την αρχική επιλογή των εγγραφών του κωδικού-βιβλίου (δηλαδή το $C_i^{(0)}$). Η αρχική έκδοση του αλγορίθμου LBG προτείνει μία μέθοδο διαίρεσης, υπολογιστικά μη-δαπανηρή,

σύμφωνα με την οποία οι αρχικές εγγραφές $C_i^{(0)}$ του κωδικό-βιβλίου υπολογίζονται στα ακόλουθα βήματα.

Βήμα 1°: Ορίζουμε $k = 1$ και επιλέγουμε ένα κωδικό-βιβλίο με k εγγραφές (δηλαδή μία) $C_1^{(0)}$ η οποία είναι ο μέσος όρος όλων των μπλοκ B_m της εικόνας.

Βήμα 2°: Σε κάποια επανάληψη θα υπάρχουν k εγγραφές $C_i^{(0)}$ στο κωδικό-βιβλίο, όπου $i = 1,2, \dots, k$. Διαιρούμε κάθε εγγραφή (δηλαδή κάθε διάνυσμα) στις δύο όμοιες εγγραφές $C_i^{(0)} \pm e$, όπου e ένα σταθερό διάνυσμα. Ορίζουμε $k \leftarrow 2k$. (Εναλλακτικά, μπορούν να χρησιμοποιηθούν δύο διανύσματα $C_i^{(0)}$ και $C_i^{(0)} + e$, μία επιλογή που οδηγεί συνολικά σε μικρότερη παραμόρφωση).

Βήμα 3°: Εάν υπάρχουν αρκετές εγγραφές στο κωδικό-βιβλίο, σταμάτησε τη διαδικασία της διαίρεσης. Το τρέχων σύνολο των k εγγραφών του κωδικό-βιβλίου μπορεί να θεωρηθεί ως το αρχικό σύνολο $C_i^{(0)}$ του αλγορίθμου LBG. Εφόσον απαιτείται μεγαλύτερο πλήθος εγγραφών, εκτελείται ο αλγόριθμος με LBG με το τρέχων σύνολο των k εγγραφών, προκειμένου να συγκλίνει σε ένα καλύτερο σύνολο. Στη συνέχεια επανεκτελείται το βήμα 2.

Η διαδικασία αυτή ξεκινά με μία εγγραφή $C_1^{(0)}$ του κωδικό-βιβλίου η οποία και χρησιμοποιείται προκειμένου να δημιουργηθούν δύο εγγραφές, στη συνέχεια τέσσερεις, οκτώ κλπ. Ο αριθμός των εγγραφών είναι πάντοτε δύναμη του 2. Εάν απαιτείται διαφορετικός αριθμός τότε την τελευταία φορά που εκτελείται το βήμα 2, μπορούν να διαιρεθούν μόνο κάποιες εγγραφές. Για παράδειγμα εάν απαιτούνται 11 εγγραφές στο κωδικό-βιβλίο τότε η διαδικασία της διαίρεσης επαναλαμβάνεται έως ότου δημιουργηθούν 8 εγγραφές. Το βήμα 2 εκτελείται για να διαιρέσει μόνο τις τρείς από τις οκτώ εγγραφές, καταλήγοντας με τον τρόπο αυτό σε 11 εγγραφές.

Ένα ακόμα χαρακτηριστικό του αλγορίθμου LBG αξίζει να σημειωθεί. Στο 2° βήμα ο αλγόριθμος υποδεικνύει να ευρεθούν όλα τα μπλοκ B_m της εικόνας τα οποία βρίσκονται, βάσει κάποιας μετρικής, πλησιέστερα στο C_i από κάθε άλλο C_j και να θεωρηθούν ως ένα σύνολο $P_i^{(k)}$. Υπάρχει περίπτωση κανένα μπλοκ να μη βρίσκεται κοντά σε κάποια εγγραφή C_i, γεγονός που θα δημιουργήσει ένα κενό σύνολο $P_i^{(k)}$. Αυτό είναι ένα πρόβλημα δεδομένου ότι η μέση τιμή των μπλοκ της εικόνας τα οποία περιέχονται στο σύνολο $P_i^{(k)}$ χρησιμοποιούνται για να υπολογιστεί μία καλύτερη εγγραφή στο κωδικό-βιβλίο κατά την επόμενη επανάληψη. Μία καλή λύση δίνεται με τη διαγραφή της

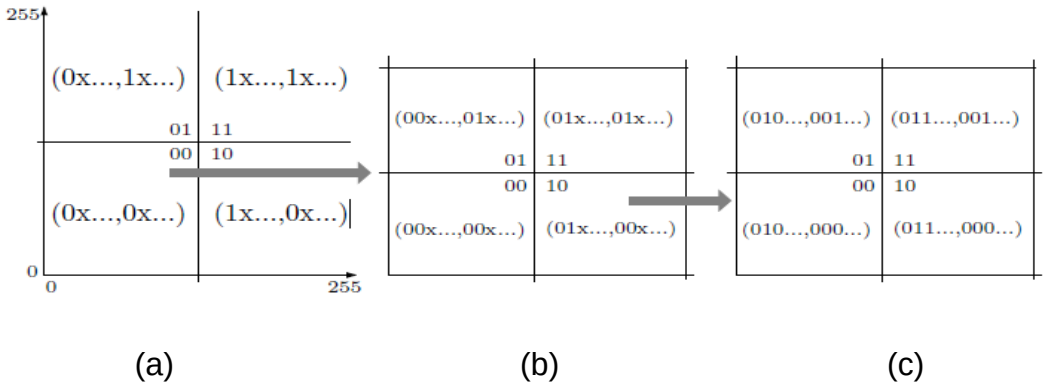
συγκεκριμένης εγγραφής C_i από το κωδικό-βιβλίο και η αντικατάσταση της με μία νέα η οποία θα επιλεγεί τυχαία από ένα μπλοκ της εικόνας που περιέχεται στο μεγαλύτερο σύνολο $P_j^{(k)}$.

Εάν το μπλοκ εικόνας είναι ένα n-διάστατο διάνυσμα, τότε η διαδικασία κατασκευής των συνόλων του βήματος 2 του αλγορίθμου LBG διαιρεί τον n-διάστατο χώρο σε περιοχές Voronoi με μία εγγραφή C_i στο κέντρο κάθε περιοχής. Στο Σχήμα 4.11 και Σχήμα 4.12 παρουσιάζονται οι περιοχές Voronoi κατά τις δύο πρώτες επαναλήψεις του προαναφερθέντος παραδείγματος. Σε κάθε επανάληψη οι εγγραφές του κωδικού-βιβλίου τροποποιούνται και επομένως μεταβάλλονται και οι περιοχές Voronoi.

4.3.4 Δενδροειδής Μορφή VQ

Ο αλγόριθμος LBG είναι υπολογιστικά δαπανηρός καθώς ο υπολογισμός των συνόλων του βήματος 2 απαιτεί πολλές συγκρίσεις. Κάθε μπλοκ B_m της εικόνας θα πρέπει να συγκριθεί με κάθε εγγραφή C_j του κωδικό-βιβλίου προκειμένου να βρεθεί η εγγραφή που ταιριάζει περισσότερο στην C_i. Επομένως ένας αλγόριθμος απαιτεί n βήματα ανά μπλοκ εικόνας, όπου n το πλήθος των εγγραφών του κωδικού-βιβλίου. Η παραλλαγή αυτή ονομάζεται δενδροειδής μορφή διανυσματικής κβάντισης (TSVQ). Η βασική ιδέα είναι η χρήση μεμονωμένων bit των εικονοστοιχείων που αποτελούν ένα μπλοκ της εικόνας και η εύρεση του μπλοκ αυτού στο χώρο. Στο παράδειγμα μας κάθε μπλοκ αποτελείται από δύο εικονοστοιχεία των 8 bit τα οποία θεωρούνται ως οι συντεταγμένες (x, y) του μπλοκ σε έναν δισδιάστατο χώρο. Γενικά ένα μπλοκ αποτελείται από k εικονοστοιχεία και μπορεί να θεωρηθεί ως ένα σημείο του k-διάστατου χώρου.

Διαιρούμε το χώρο σε τέσσερα τεταρτημόρια αριθμημένα ως $00, 01, 10$ και 11, όπως φαίνεται στο Σχήμα 4.13.



(a) (b) (c)

Σχήμα 4.13 : Αρίθμηση τεταρτημορίων

Το περισσότερο σημαντικό bit από τα x και y των συντεταγμένων ενός σημείου καθορίζουν το τεταρτημόριο. Επομένως, τα μπλοκ της εικόνας με δυαδικές τιμές $(0, \dots, 0 \dots)$ τοποθετούνται στο τεταρτημόριο 00, όλα τα μπλοκ της εικόνας της μορφής $(1 \dots, 0 \dots)$ τοποθετούνται στο τεταρτημόριο 10 και αντιστοίχως για τα υπόλοιπα. Στη συνέχεια διαιρείται κάθε τεταρτημόριο σε τέσσερα υπό-τεταρτημόρια και είναι προφανές πως πλέον το δεύτερο πιο σημαντικό bit καθορίζει το υπό-τεταρτημόριο. Στο Σχήμα 4.13 b παρουσιάζονται τα τέσσερα υπό-τεταρτημόρια του τεταρτημορίου 00 και τα δύο πρώτα πιο σημαντικά ψηφία τα οποία έχουν τοποθετηθεί σε αυτό. Ομοίως και στο Σχήμα 4.13 c παρουσιάζεται η διαίρεση των υπό-τεταρτημορίων με το τρίτο περισσότερο σημαντικό bit να προσδιορίζει τη θέση του.

Δεδομένου ότι τα σημεία (μπλοκ εικόνας) έχουν συντεταγμένες 8 bit μπορούν να καθοριστούν επακριβώς δημιουργώντας υπό-τεταρτημόρια όγδοης τάξης. Μπορεί πλέον να κατασκευαστεί ένα τετραδικό δένδρο βάθους οκτώ όπου κάθε φύλλο του σχετίζεται με μία εγγραφή στο κωδικό-βιβλίο. Δοθέντος ενός μπλοκ εικόνας $B_m = (x, y)$ χρησιμοποιούμε διαδοχικά ζεύγη bit (x, y) ώστε να προχωρήσουμε στο δένδρο. Όταν φτάσουμε σε φύλλο τοποθετούμε το B_m στο τμήμα της εγγραφής του κωδικό-βιβλίου που βρέθηκε στο φύλλο αυτό. Για παράδειγμα, έστω το μπλοκ εικόνας B_4. Οι συντεταγμένες του είναι $(60,50) = (00111100, 00110010)$. Τα ζεύγη (x, y) bit προς bit είναι $(0, 0), (0,0), (1, 1), (1, 1), (1, 0), (1, 0)$ και $(0,0)$. Τα ζεύγη αυτά καθορίζουν καθένα από τα οκτώ βήματα που θα ακολουθηθούν μέχρι το φύλλο του δένδρου. Όταν φτάσουμε στο φύλλο βρίσκουμε την εγγραφή C_1 στο κωδικό-βιβλίο και συμπεριλαμβάνουμε το μπλοκ B_4 στο τμήμα P_1.

Αξίζει να σημειωθεί πως η περιοχή ενός υπό-τεταρτημορίου όγδοης τάξης είναι $4^{-8} \approx 0.000015$ της περιοχής ολόκληρου του χώρου.

4.3.5 Προσαρμοστική Διανυσματική Κβάντιση

Η βασική μέθοδος διανυσματικής κβάντισης που περιγράφηκε στην προηγούμενη ενότητα χρησιμοποιεί σταθερό κωδικό-βιβλίο ή εκτελεί τον αλγόριθμο LBG (ή κάποιον αντίστοιχο) προκειμένου να το κατασκευάσει. Σε όλες αυτές τις περιπτώσεις το κωδικό-βιβλίο αποτελείται από εγγραφές σταθερού μήκους όμοιες με το μέγεθος του μπλοκ της εικόνας. Η προσαρμοστική μέθοδος με την οποία θα ασχοληθούμε σε αυτή την ενότητα, σύμφωνα με τους Constantinescu και Storer[27], χρησιμοποιεί μεταβλητού μήκους μπλοκ και εγγραφές κωδικό-βιβλίου. Το κωδικό-βιβλίο σε αυτή την περίπτωση ονομάζεται λεξικό. Η μέθοδος συνδυάζει το προσαρμοστικό λεξικό κατά το μονό

πέρασμα το οποίο χρησιμοποιούν διάφοροι αλγόριθμοι βασισμένοι σε λεξικά, με τις μετρικές απόστασης τις οποίες χρησιμοποιούν διάφορες μέθοδοι διανυσματικής κβάντισης προκειμένου να επιτύχουν καλές προσεγγίσεις των μπλοκ δεδομένων.

Σε κάθε βήμα της διαδικασίας της κωδικοποίησης ο κωδικοποιητής επιλέγει ένα μπλοκ της εικόνας (ορθογώνιο οποιουδήποτε μεγέθους) και το αντιστοιχεί σε κάποια εγγραφή του λεξικού, δίνοντας ως έξοδο έναν δείκτη στην εγγραφή αυτή και ενημερώνει το λεξικό προσθέτοντας μία ή περισσότερες εγγραφές σε αυτό. Οι νέες εγγραφές βασίζονται στο τρέχων μπλοκ εικόνας. Η περίπτωση όπου ένα λεξικό έχει γεμίσει θα πρέπει να ληφθεί υπόψη και επομένως θα συζητηθεί στη συνέχεια.

Πειραματισμός με τη μέθοδο αυτή οδήγησε σε λόγους συμπίεσης αντίστοιχους ή και καλύτερους από αυτούς που επιτυγχάνονταν με το JPEG. Παράλληλα η μέθοδος αυτή είναι γρήγορη, καθώς πραγματοποιείται σε ένα πέρασμα, απλή, αφού ξεκινά με άδειο λεξικό και επομένως δεν απαιτείται εκπαίδευση, και αξιόπιστη δεδομένου ότι το μέγεθος των απωλειών ελέγχεται εύκολα και με ακρίβεια χρησιμοποιώντας μία παράμετρο ανοχής. Ο αποκωδικοποιητής λειτουργεί παρόμοια με έναν αποκωδικοποιητή διανυσματικής κβάντισης, επομένως είναι απλός και γρήγορος.

4.3.5.1 **Βασική Βήματα Κωδικοποίησης και Αποκωδικοποίησης**

Ο κωδικοποιητής θα πρέπει να επιλέξει τα μπλοκ της εικόνας προσεκτικά, βεβαιώνοντας ότι καλύπτουν ολόκληρη την εικόνα χωρίς να υπάρχει μεγάλη επικάλυψη ανάμεσα σε διαφορετικά μπλοκ. Προκειμένου να επιλεγούν τα μπλοκ ο κωδικοποιητής επιλέγει ένα ή περισσότερα αναπτυσσόμενα σημεία σε κάθε επανάληψη τα οποία και χρησιμοποιεί ως αρχή (γωνία) του νέου μπλοκ για την επόμενη επανάληψη. Το μπλοκ ξεκινά έχοντας μικρό μέγεθος και μεγαλώνει (επεκτείνεται από το αρχικό του σημείο) από τον κωδικοποιητή μέχρις ότου να βρεθεί μία εγγραφή στο λεξικό με την οποία να ταιριάζει. Ένα αναπτυσσόμενο σημείο συμβολίζεται με το G και αποτελεί μία τριάδα (x, y, q) όπου x και y οι συντεταγμένες του σημείου και q εκφράζει τη γωνία του μπλοκ της εικόνας όπου βρίσκεται το σημείο αυτό. Θεωρούμε τιμές του q τις $0, 1, 2$ και 3 οι οποίες αναπαριστούν την επάνω αριστερή, επάνω δεξιά, κάτω αριστερή και κάτω δεξιά γωνία αντίστοιχα (επομένως το q είναι ένας αριθμός των 2 bit). Ένα μπλοκ B εικόνας το οποίο τοποθετείται στο αναπτυσσόμενο σημείο G σημειώνεται ως $B = (G, w, h)$, όπου w και h, το μήκος και το ύψος του μπλοκ.

Θα ακολουθήσουν τα βασικά βήματα τόσο του κωδικοποιητή όσο και του αποκωδικοποιητή τα οποία στη συνέχεια θα αναλυθούν λεπτομερέστερα. Τα βασικά βήματα του κωδικοποιητή είναι τα εξής:

Βήμα 1°: Αρχικοποίηση του λεξικού D με κάθε πιθανή τιμή των εικονοστοιχείων της εικόνας. Αρχικοποίηση του συνόλου των αναπτυσσόμενων σημείων (GPP) με ένα ή περισσότερα αναπτυσσόμενα σημεία χρησιμοποιώντας έναν από τους αλγορίθμους που θα ακολουθήσουν.

Βήμα 2°: Επανάληψη των βημάτων 3-7 μέχρι το GPP να είναι κενό.

Βήμα 3°: Χρήση ενός αναπτυσσόμενου αλγορίθμου ώστε να επιλεγεί ένα αναπτυσσόμενο σημείο G από το GPP.

Βήμα 4°: Επέκτεινε το μπλοκ B της εικόνας με επιλογή γωνίας G. Χρησιμοποίησε αλγόριθμο ταιριάσματος προκειμένου να ταιριαστεί το B, καθώς επεκτείνεται, σε μία εγγραφή του λεξικού με ευκρίνεια που έχει επιλέξει ο χρήστης.

Βήμα 5°: Όταν το B επιτύχει το μέγιστο μέγεθος για το οποίο υπάρχει ακόμα ταίριασμα με μία εγγραφή d του λεξικού, δίνεται ως έξοδος ένας δείκτης στο d. Το μέγεθος του δείκτη εξαρτάται από το πλήθος εγγραφών, δηλαδή το μέγεθος του D.

Βήμα 6°: Διαγραφή του G από το GPP και χρήση αλγορίθμου για την επιλογή νέου αναπτυσσόμενου σημείου (εάν υπάρχει) προκειμένου να προστεθεί στο GPP.

Βήμα 7°: Εάν το D είναι γεμάτο, τότε χρήση αλγορίθμου για τη διαγραφή μίας ή περισσοτέρων εγγραφών. Χρήση αλγορίθμου για την ανανέωση του λεξικού βάσει του B.

Η λειτουργία το κωδικοποιητή εξαρτάται, όπως είναι προφανές, από πολλούς αλγορίθμους. Αντίθετα από αποκωδικοποιητής είναι πολύ πιο απλός και γρήγορος. Τα βασικά βήματα του είναι τα εξής:

Βήμα 1°: Αρχικοποίηση του λεξικού D και του GPP όπως και στο 1° βήμα του κωδικοποιητή.

Βήμα 2°: Επανάληψη των βημάτων 3-5 έως ότου το GPP να είναι κενό.

Βήμα 3°: Χρήση του αναπτυσσόμενου αλγορίθμου που χρησιμοποιήθηκε από τον κωδικοποιητή ώστε να επιλεγεί ένα αναπτυσσόμενο σημείο από το GPP.

Βήμα 4°: Είσοδος ενός δείκτη από τη συμπιεσμένη ροή δεδομένων και χρήση του για ανάκτηση μίας εγγραφής d από το λεξικό στο σημείο και τη θέση που καθορίζει το G.

Βήμα 5°: Ενημέρωση του D και του GPP όπως στα βήματα 6-7 του κωδικοποιητή.

Στη συνέχεια θα δοθούν διάφοροι αλγόριθμοι οι οποίοι χρησιμοποιούνται καθώς και σχετικό παράδειγμα για βαθύτερη κατανόηση.

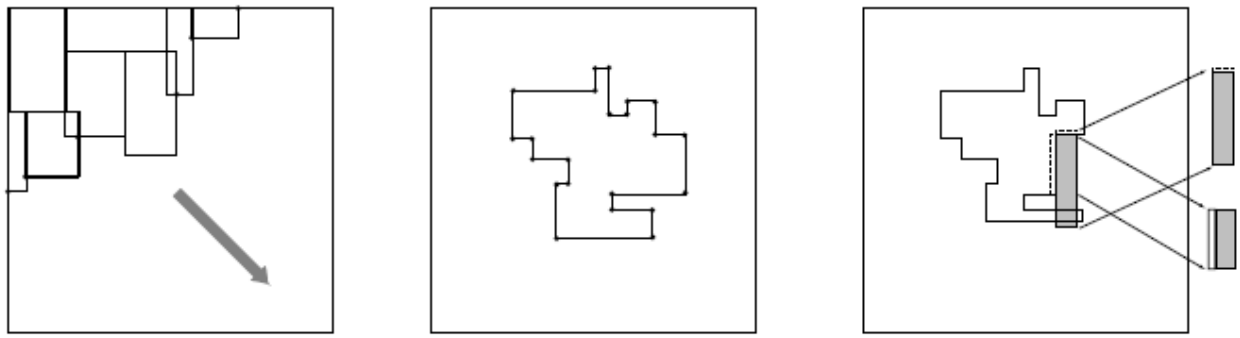
4.3.5.2 **Αλγόριθμος Επιλογής ενός Αναπτυσσόμενου Σημείου G από το GPP**

Οι απλούστερες μέθοδοι είναι οι LIFO (η οποία δεν οδηγεί συνήθως σε καλά αποτελέσματα) και FIFO (η οποία περιγράφεται και με παράδειγμα), όμως οι μέθοδοι κάλυψης οι οποίες θα ακολουθήσουν, ενδέχεται να αποδώσουν καλύτερα αφού καθορίζουν τον τρόπο με τον οποίο θα καλυφθεί η εικόνα με μπλοκ.

- Κυματοειδής Κάλυψη: Ο αλγόριθμος αυτός επιλέγει ανάμεσα από όλα τα αναπτυσσόμενα σημεία στο GPP, το σημείο $G = (x_s, y_s, 0)$ εκείνο το οποίο ικανοποιεί τη σχέση

$$x_s + y_s \leq x + y, \qquad \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ G(x, y, 0)\sigma\tau o\ GPP$$

Παρατηρούμε ότι το αναπτυσσόμενο σημείο που επιλέγεται έχει $q = 0$, επομένως τοποθετείται στην επάνω αριστερή γωνία του μπλοκ της εικόνας. Όταν χρησιμοποιείται ο αλγόριθμος αυτός, έχει νόημα να αρχικοποιηθεί το GPP με ένα μόνο σημείο, δηλαδή το επάνω αριστερό άκρο της εικόνας. Εφόσον γίνει αυτό, τότε ο αλγόριθμος κυματικής κάλυψης θα επιλέξει μπλοκ τα οποία καλύπτουν την εικόνα με μία κυμάτωση η οποία κινείται από την επάνω αριστερή έως την κάτω δεξιά γωνία της εικόνας (Σχήμα 4.14 a).



(a) (b) (c)

Σχήμα 4.14 : (a): Κυματοειδής Κάλυψη. (b): Κυκλική Κάλυψη. (c): Διαγώνια κάλυψη

- Κυκλική Κάλυψη: Επέλεξε το αναπτυσσόμενο σημείο G του οποίου η απόσταση από το κέντρο της εικόνας είναι ελάχιστη σε σχέση με όλα τα υπόλοιπα αναπτυσσόμενα σημεία G στο GPP. Ο αλγόριθμος αυτός καλύπτει την εικόνα με μπλοκ τα οποία εντοπίζονται σε μία κυκλικώς αναπτυσσόμενη περιοχή γύρω από το κέντρο (Σχήμα 4.14 b).

- Διαγώνια κάλυψη: Ο αλγόριθμος αυτός επιλέγει ανάμεσα από όλα τα αναπτυσσόμενα σημεία του GPP εκείνο το G για το οποίο ισχύει $G = (x_s, y_s, q)$ και ικανοποιεί τη σχέση

$$|x_s - y_s| \leq |x - y|, \qquad \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ G(x, y, p)\sigma\tau o\ GPP$$

Το αποτέλεσμα θα είναι μπλοκ τα οποία ξεκινούν γύρω από την κύρια διαγώνιο και απομακρύνονται από αυτή ως δύο κύματα τα οποία εκτίνονται παράλληλα σε αυτή.

4.3.5.3 Αλγόριθμος Ταιριάσματος ενός Μπλοκ B της Εικόνας (με αναπτυσσόμενο σημείο G ως γωνία) με μία Εγγραφή του Λεξικού.

Η καλύτερη προσέγγιση φαίνεται να είναι η έναρξη της διαδικασίας αντιστοίχισης με το μικρότερο δυνατό μπλοκ B (δηλαδή ένα εικονοστοιχείο) και στη συνέχεια η προσπάθεια αντιστοίχισης όλο και μεγαλύτερων μπλοκ με τις εγγραφές του λεξικού μέχρι το σημείο όπου η περεταίρω αύξηση του B δεν μπορεί να επιτύχει καλύτερη αντιστοίχιση με κάποια εγγραφή του λεξικού σύμφωνα με την παράμετρο ανοχής που έχει προσδιορίσει ο χρήστης. Οι ακόλουθες παράμετροι ελέγχουν την αντιστοίχιση:

1. Η μετρική απόστασης. Οποιαδήποτε από τις μετρικές (4.19), (4.20), (4.21) οι οποίες προτάθηκαν μπορεί να χρησιμοποιηθεί.
2. Η ανοχή. Η πραγματική παράμετρος t η οποία καθορίζεται από τον χρήστη, θέτει τη μέγιστη επιτρεπόμενη απόσταση ανάμεσα σε ένα μπλοκ και μία εγγραφή του λεξικού.
3. Ο τύπος κάλυψης. Δεδομένου ότι τα μπλοκ της εικόνας έχουν διαφορετικά μεγέθη και επεκτείνονται σε σχέση με το αναπτυσσόμενο σημείο προς διαφορετικές κατευθύνσεις, ενδέχεται να υπάρχουν και μεταξύ τους αλληλεπικαλύψεις (Σχήμα 4.14 a). Υπάρχει η δυνατότητα υπολογισμού της απόστασης χρησιμοποιώντας μόνο το τμήμα του μπλοκ το οποίο δεν αλληλεπικαλύπτεται με κανένα άλλο μπλοκ. Αυτό σημαίνει ότι το πρώτο μπλοκ το οποίο καλύπτει μία συγκεκριμένη περιοχή της εικόνας θα καθορίσει την ποιότητα της αποσυμπιεσμένης εικόνας στην περιοχή αυτή. Αυτό ονομάζεται πρώτη κάλυψη. Το αντίθετο είναι η τελευταία κάλυψη κατά την οποία η απόσταση, ανάμεσα σε ένα μπλοκ B της εικόνας και μίας εγγραφής του λεξικού, υπολογίζεται χρησιμοποιώντας ολόκληρο το B, ανεξαρτήτος επικαλύψεων. Υπάρχει επίσης η δυνατότητα για μέση κάλυψη σύμφωνα με την οποία η απόσταση υπολογίζεται για όλο το B αλλά σε μία περιοχή επικάλυψης, κάθε τιμή εικονοστοιχείου που χρησιμοποιείται για τον υπολογισμό της απόστασης θα δίνεται από τη μέση τιμή των εικονοστοιχείων κάθε μπλοκ της περιοχής αυτής.
4. Το μέγεθος του στοιχειώδους υπό-μπλοκ l. Πρόκειται για μία παράμετρο κανονικοποίησης η οποία αντισταθμίζει τα διαφορετικά μεγέθη μπλοκ. Είναι γνωστό ότι τα μπλοκ της εικόνας μπορούν να έχουν διαφορετικό μέγεθος. Ενδέχεται επομένως κάποια μεγάλα μπλοκ να ταιριάζουν με μία εγγραφή του λεξικού βάσει της παραμέτρου ανοχής, ενώ άλλες περιοχές εντός του μπλοκ να ταιριάζουν σε πολύ μικρό βαθμό ή άλλα σε πολύ μεγαλύτερο. Επομένως το μπλοκ κάθε εικόνας θα πρέπει να διαιρεθεί σε υπό-μπλοκ μεγέθους $l \times l$ (όπου το l καθορίζεται από το μέγεθος της εικόνας αλλά συνήθως έχει τιμές 4 ή 5) και κάθε εγγραφή στο λεξικό θα πρέπει ομοίως να διαιρεθεί. Ο αλγόριθμος που χρησιμοποιείται για την ανίχνευση του καλύτερου ταιριάσματος θα πρέπει να

υπολογίσει την απόσταση ανάμεσα στο κάθε υπό-μπλοκ της εικόνας και της αντίστοιχης υπό-εγγραφής του λεξικού. Το μέγιστο των αποστάσεων αυτών θα πρέπει να επιλεγεί ως η απόσταση ανάμεσα στο μπλοκ της εικόνας και της εγγραφής του λεξικού.

4.3.5.4 **Αλγόριθμος Ανανέωσης του GPP**

Μία καλή τακτική για την επιλογή των νέων αναπτυσσόμενων σημείων είναι η χρήση σημείων τα οποία βρίσκονται κοντά στο πλαίσιο της μερικώς κωδικοποιημένης εικόνας (Σχήμα 4.14 b). Αυτό έχει ως αποτέλεσμα τα νέα μπλοκ της εικόνας να είναι γειτονικά ως προς τα προηγούμενα. Η τακτική αυτή έχει νόημα καθότι ένα παλαιό μπλοκ έχει συνεισφέρει στις νέες εγγραφές ενός λεξικού και οι εγγραφές αυτές είναι οι καλύτερες υποψήφιες ώστε να ταιριάξουν σε έναν νέο γειτονικό μπλοκ καθώς συνήθως τα γειτονικά μπλοκ δεν διαφέρουν ιδιαίτερα. Αρχικά, όταν η μερικώς κωδικοποιημένη εικόνα είναι μικρή, αποτελούμενη κυρίως από μπλοκ του ενός εικονοστοιχείου, η τακτική αυτή προσθέτει δύο αναπτυσσόμενα σημεία στο GPP για κάθε μικρό μπλοκ της κωδικοποιημένης εικόνας. Αυτά τα σημεία είναι κάτω και δεξιά του μπλοκ της εικόνας. Μία άλλη καλή τοποθεσία για νέα αναπτυσσόμενα σημεία είναι κάθε νέα κοίλη γωνία η οποία δημιουργήθηκε προσθέτοντας ένα πρόσφατο μπλοκ στη μερικώς κωδικοποιημένη εικόνα (Σχήμα 4.14 a,b). Θα αναδειχθεί στη συνέχεια πως κάθε νέο μπλοκ το οποίο αναπτύσσεται από μία γωνία αυτού του είδους, συνεισφέρει δύο ή και τρείς εγγραφές στο λεξικό.

4.3.5.5 **Αλγόριθμος Ανανέωσης του Λεξικού**

Η ανανέωση του λεξικού βασίζεται σε δύο αρχές:

1. Κάθε αντιστοιχισμένο μπλοκ το οποίο έχει προστεθεί στην εικόνα μέχρι εκείνη τη στιγμή χρησιμοποιείται προκειμένου να προστεθούν μία ή περισσότερες εγγραφές στο λεξικό.
2. Οι νέες εγγραφές οι οποίες προστίθενται θα πρέπει να περιέχουν εικονοστοιχεία της εικόνας τα οποία έχουν κωδικοποιηθεί μέχρι εκείνη τη στιγμή ώστε ο αποκωδικοποιητής να είναι σε θέση να ανανεώνει το λεξικού του με τον ίδιο τρόπο.

Συμβολίζοντας το μπλοκ που μόλις αντιστοιχίστηκε με B, ένας απλός κανόνας είναι η προσθήκη στο λεξικό των μπλοκ τα οποία προέκυψαν από την πρόσθεση μίας στήλης εικονοστοιχείων αριστερά του B και μίας γραμμής εικονοστοιχείων στην κορυφή του B αντίστοιχα. Μερικές φορές όπως παρατηρείται και στο Σχήμα 4.14 c, η γραμμή και η στήλη μπορεί να προστεθούν σε τμήμα του B. Πειραματικά προέκυψε ότι με τη μέθοδο αυτή μπορεί να προστεθεί και ένα τρίτο μπλοκ στο λεξικό, το οποίο προκύπτει από την πρόσθεση μίας στήλης εικονοστοιχείων αριστερά του B και μίας γραμμής

εικονοστοιχείων στην κορυφή του B. Δεδομένου ότι τα μπλοκ B της εικόνας ξεκινούν μικρά και τετράγωνα, αυτή η τρίτη εγγραφή στο λεξικό είναι επίσης τετράγωνη. Η εισαγωγή αυτής της τρίτης εγγραφής βελτιώνει σημαντικά της απόδοση της συμπίεσης.

4.3.5.6 **Αλγόριθμος Διαγραφής Λεξικού**

Το λεξικό αναπτύσσεται διαρκώς όμως η διαδικασία αυτή δεν εξελίσσεται για πάντα. Όταν το λεξικό φτάσει ένα μέγιστο μέγεθος υπάρχουν οι εξής τρείς εναλλακτικές:

1. Διαγραφή της λιγότερο πρόσφατα χρησιμοποιημένης εγγραφής (LRU) (εκτός από τις αρχικές οι οποίες δε θα πρέπει να διαγραφούν).
2. Μην προσθέτεις άλλες εγγραφές στο λεξικό.
3. Διαγραφή ολόκληρου του λεξικού (εκτός των αρχικών εγγραφών) και έναρξη από την αρχή.

Δοκιμές έδειξαν ότι ο βασικός αλγόριθμος είναι αξιόπιστος και δεν εξαρτάται ιδιαίτερα από την επιλογή των παραπάνω αλγορίθμων εκτός από δύο επιλογές οι οποίες φαίνεται να βελτιώνουν σημαντικά την απόδοση:

1. Η κυματοειδής κάλυψη φαίνεται να προσφέρει καλύτερα αποτελέσματα από την κυκλική και τη διαγώνια.
2. Η οπτική αξιολόγηση πολλών αποσυμπιεσμένων εικόνων έδειξε ότι η απώλεια των δεδομένων είναι πιο εμφανής στο μάτι στις ομοιόμορφες περιοχές της εικόνας. Αυτό σημαίνει ότι η μέθοδος μπορεί να βελτιωθεί με τη χρήση μικρότερης παραμέτρου ανοχής σε περιοχές με λεία υφή. Μία απλή μετρική A της ομοιομορφίας είναι ο λόγος V/M, όπου V είναι η διασπορά των εικονοστοιχείων της περιοχής και M η μέση τιμή. Όσο μεγαλύτερο το A τόσο λιγότερο ομοιόμορφη είναι η περιοχή. Η παράμετρος ανοχής t καθορίζεται από το χρήστη και συνίσταται η χρήση μικρότερης παραμέτρου t σύμφωνα με τα ακόλουθα:

$$\text{Εάν} \begin{cases} A \leq 0.05, & \text{τότε } 0.4t \\ 0.05 < A \leq 0.1, & \text{τότε } 0.6t \\ A > 0.1, & \text{τότε } t \end{cases}$$

4.3.6 Παράδειγμα

Στο παράδειγμα[58] που ακολουθεί επιλέγεται εικόνα η οποία αποτελείται από 4-bit εικονοστοιχεία. Τα εννέα εικονοστοιχεία της επάνω αριστερής γωνίας της εικόνας παρουσιάζονται στο Σχήμα 4.15. Οι συντεταγμένες τις εικόνες ορίζονται από $(0,0)$ μέχρι το $(2,2)$. Οι πρώτες 16 εγγραφές του λεξικού D (θέσεις 0 έως 15) αρχικοποιούνται στις 16 πιθανές τιμές των εικονοστοιχείων. Το GPP αρχικοποιείται στην επάνω αριστερή

γωνία της εικόνας στο σημείο $(0,0)$. Θεωρούμε ότι το GPP χρησιμοποιείται ως ουρά (FIFO). Ακολουθούν οι λεπτομέρειες των πρώτων βημάτων.

	0	1	2	...
0	8	2	3	
1	4	5	6	
2	7	8	9	
⋮				

Σχήμα 4.15 : Επάνω Αριστερό Τμήμα Εικόνας

Βήμα 1°: Το σημείο $(0,0)$ βγαίνει από την GPP. Η τιμή του εικονοστοιχείου σε αυτή τη θέση είναι 8. Δεδομένου ότι το λεξικό περιέχει μόνο μεμονωμένα εικονοστοιχεία και όχι μπλοκ, το σημείο αυτό μπορεί να αντιστοιχιστεί μόνο με την εγγραφή του λεξικού που περιέχει το 8. Η εγγραφή αυτή τυγχάνει να βρίσκεται στη θέση 8 του λεξικού και επομένως ο κωδικοποιητής δίνει στην έξοδο του έναν δείκτη με τον αριθμό 8. Η αντιστοίχιση αυτή δεν έχει κοίλες γωνίες επομένως εισάγουμε στο GPP το σημείο στα δεξιά του ταιριασμένου μπλοκ, δηλαδή το $(0,1)$ καθώς και το σημείο κάτω από αυτό, δηλαδή το $(1,0)$.

Βήμα 2°: Το σημείο $(1,0)$ βγαίνει από την GPP. Η τιμή του εικονοστοιχείου στη θέση αυτή είναι το 4. Το λεξικό εξακολουθεί να περιέχει μόνο μεμονωμένα εικονοστοιχεία και όχι μεγαλύτερα μπλοκ, επομένως το σημείο αυτό μπορεί να αντιστοιχιστεί μόνο με την εγγραφή του λεξικού που περιέχει το 4. Η εγγραφή αυτή βρίσκεται στη θέση 4 του λεξικού και έτσι ο κωδικοποιητής δίνει ως έξοδο έναν δείκτη στη θέση 4. Η αντιστοίχιση αυτή δεν έχει κοίλες γωνίες επομένως εισάγουμε στο GPP το σημείο στα δεξιά του ταιριασμένου μπλοκ, δηλαδή το $(1,1)$ καθώς και το σημείο κάτω από αυτό, δηλαδή το $(2,0)$. Το GPP πλέον περιέχει (από το περισσότερο πρόσφατο στο λιγότερο πρόσφατο) τα σημεία $(1,1),(2,0)\ και\ (0,1)$. Το λεξικό ενημερώνεται με την προσθήκη του μπλοκ $\begin{bmatrix}8\\4\end{bmatrix}$ όπως φαίνεται στο Σχήμα 4.16 a.

$$0,1,2,\ldots,15,\boxed{\begin{matrix}8\\4\end{matrix}}$$

(a)

$$0,1,2,\ldots,15,\boxed{\begin{matrix}8\\4\end{matrix}},\boxed{8|2},\boxed{\begin{matrix}4\\7\end{matrix}},\boxed{\begin{matrix}2\\5\end{matrix}},\boxed{4|5}.$$

(b)

Σχήμα 4.16 : Εγγραφές με τις οποίες Ενημερώνεται το Λεξικό

Βήμα 3°: Το σημείο $(0,1)$ βγαίνει από την GPP. Η τιμή του εικονοστοιχείου στη θέση αυτή είναι 2. Το λεξικό περιλαμβάνει τα 16 μεμονωμένα εικονοστοιχεία και ένα μεγαλύτερο μπλοκ. Η καλύτερη αντιστοίχιση για το σημείο αυτό είναι η εγγραφή στο λεξικό που περιέχει το 2 και το οποίο βρίσκεται στη δεύτερη θέση συνεπώς ο κωδικοποιητής δίνει ως έξοδο έναν δείκτη στο 2. Η αντιστοίχιση αυτή δεν έχει κοίλες γωνίες επομένως εισάγουμε στο GPP το σημείο στα δεξιά του ταιριασμένου μπλοκ, δηλαδή το $(0,2)$ καθώς και το σημείο κάτω από αυτό, δηλαδή το $(1,1)$. Το GPP πλέον περιέχει (από το περισσότερο πρόσφατο στο λιγότερο πρόσφατο) τα σημεία $(10\ 2), (1,1)\ \kappa\alpha\iota\ (2,0)$. Το λεξικό ενημερώνεται με την εισαγωγή του μπλοκ $[8|2]$ (Σχήμα 4.16 b).

Βήμα 4°: Το σημείο $(2,0)$ βγαίνει από το GPP. Η τιμή του εικονοστοιχείου στη θέση αυτή είναι 7. Η καλύτερη αντιστοίχηση στο σημείο αυτό επιτυγχάνεται με την εγγραφή του λεξικού 7, επομένως ο κωδικοποιητής δίνει ως έξοδο έναν δείκτη στο 7. Η αντιστοίχηση δεν έχει κοίλες γωνίες συνεπώς εισάγουμε στο GPP το σημείο στα δεξιά του ταιριασμένου μπλοκ, δηλαδή το $(2,1)$ καθώς και το σημείο κάτω από αυτό, δηλαδή το $(3,0)$. Το GPP περιλαμβάνει πλέον τα σημεία $(2,1), (3,0), (0,2)$ και $(1,1)$. Το λεξικό ενημερώνεται εισάγοντας στη θέση 18 το μπλοκ $\begin{bmatrix}4\\7\end{bmatrix}$.

Βήμα 5°: Το σημείο $(1,1)$ βγαίνει από το GPP. Η τιμή του εικονοστοιχείου στη θέση αυτή είναι 5. Η καλύτερη αντιστοίχηση στο σημείο αυτό επιτυγχάνεται με την εγγραφή του λεξικού 5, επομένως ο κωδικοποιητής δίνει ως έξοδο έναν δείκτη στο 5. Η αντιστοίχηση δεν έχει κοίλες γωνίες συνεπώς εισάγουμε στο GPP το σημείο στα δεξιά του ταιριασμένου μπλοκ, δηλαδή το $(1,2)$ καθώς και το σημείο κάτω από αυτό, δηλαδή το $(2,1)$. Το GPP περιλαμβάνει πλέον τα σημεία $(1,2), (2,1), (3,0)$ και $(0,2)$. Το λεξικό ενημερώνεται εισάγοντας στις θέσεις 19 και 20 τα μπλοκ $\begin{bmatrix}2\\5\end{bmatrix}$ και $[4|5]$.

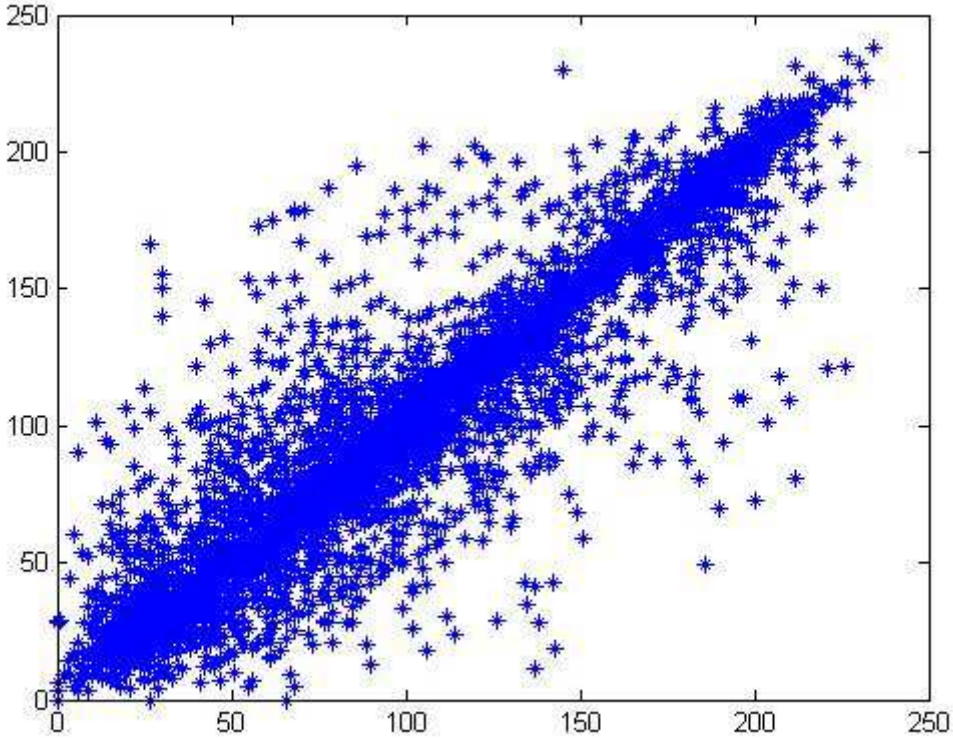
5. Μετασχηματισμοί Εικόνας

5.1 Η Βασική Ιδέα των Μετασχηματισμών

Η ιδέα ενός μαθηματικού μετασχηματισμού αποτελεί ένα ισχυρό εργαλείο σε πολλές επιστημονικές περιοχές, μεταξύ αυτών και στη συμπίεση εικόνας. Μια εικόνα μπορεί να συμπιεστεί μετασχηματίζοντας τις τιμές των εικονοστοιχείων (τα οποία συνήθως είναι συσχετισμένα μεταξύ τους) σε μια μορφή στην οποίο είναι ασυσχέτιστα. Η συμπίεση επιτυγχάνεται εάν οι νέες τιμές είναι μικρότερες, ως προς μέση τιμή, σε σχέση με τις αρχικές. Συμπίεση με απώλειες μπορεί να επιτευχθεί κβαντίζοντας τις μετασχηματισμένες τιμές. Ο αποκωδικοποιητής λαμβάνει ως είσοδο τις μετασχηματισμένες κβαντισμένες τιμές και ανακατασκευάζει την εικόνα (με ακρίβεια ή προσεγγιστικά) εφαρμόζοντας τον αντίστροφο μετασχηματισμό. Οι μετασχηματισμοί που θα συζητηθούν και που εν γένει χρησιμοποιούνται στη συμπίεση εικόνας είναι ορθογώνιοι.

Με τον όρο ασυσχέτιστες εννοούμε τις τιμές οι οποίες είναι ανεξάρτητες η μία από την άλλη. Αυτό έχει ως αποτέλεσμα να μπορούν να κωδικοποιηθούν ξεχωριστά, κάτι το οποίο βοηθά στη δημιουργία ενός στατιστικού μοντέλου. Μια εικόνα μπορεί να συμπιεστεί εφόσον η απεικόνιση της έχει πλεονασμό. Ο πλεονασμός στις εικόνες προέρχεται από τη συσχέτιση μεταξύ των εικονοστοιχείων που τις αποτελούν. Εάν μια εικόνα μετασχηματιστεί σε μια αναπαράσταση όπου τα εικονοστοιχεία της είναι ασυσχέτιστα, τότε ο πλεονασμός έχει εξαλειφθεί και η εικόνα έχει πλήρως συμπιεστεί.

Θα παρουσιαστεί ένα απλό παράδειγμα όπου θα υποθέσουμε πως σαρώνουμε μια εικόνα κατά γραμμές, ομαδοποιώντας τις τιμές των γειτονικών εικονοστοιχείων σε ζευγάρια. Επειδή τα εικονοστοιχεία είναι συσχετισμένα, τα δύο εικονοστοιχεία (x, y) ενός τυπικού ζευγαριού θα έχουν παρόμοιες τιμές. Θεωρούμε τα ζεύγη των εικονοστοιχείων ως σημεία του δισδιάστατου χώρου και τα σχεδιάζουμε. Γνωρίζουμε ότι όλα τα σημεία της μορφής (x, x) είναι τοποθετημένα στην ευθεία με κλίση 45°, $y = x$, συνεπώς περιμένουμε ότι τα σημεία θα είναι συγκεντρωμένα γύρω από αυτή τη γραμμή. Το Σχήμα 5.1 απεικονίζει την κατανομή των ζευγών αυτών σε μια τυπική εικόνα (Lena) όπου οι τιμές των εικονοστοιχείων βρίσκονται στο διάστημα [0, 255].



Σχήμα 5.1 : Ζεύγη των Τιμών Φωτεινοτήτων Γειτονικών Εικονοστοιχείων ως Συντεταγμένες ενός Δισδιάστατου Χώρου

Τα περισσότερα σημεία σχηματίζουν ένα σύννεφο γύρω από τη γραμμή αυτή και μόνο λίγα σημεία βρίσκονται μακριά από αυτή. Θα μετασχηματίσουμε την εικόνα περιστρέφοντας όλα τα σημεία κατά 45° προς την φορά δεικτών του ρολογιού γύρω από την αρχή των αξόνων. Με την περιστροφή αυτή η ευθεία των 45° θα συμπίπτει πλέων με τον άξονα χ[29]. Η περιστροφή αυτή επιτυγχάνεται με τη χρήση της ακόλουθης εξίσωσης:

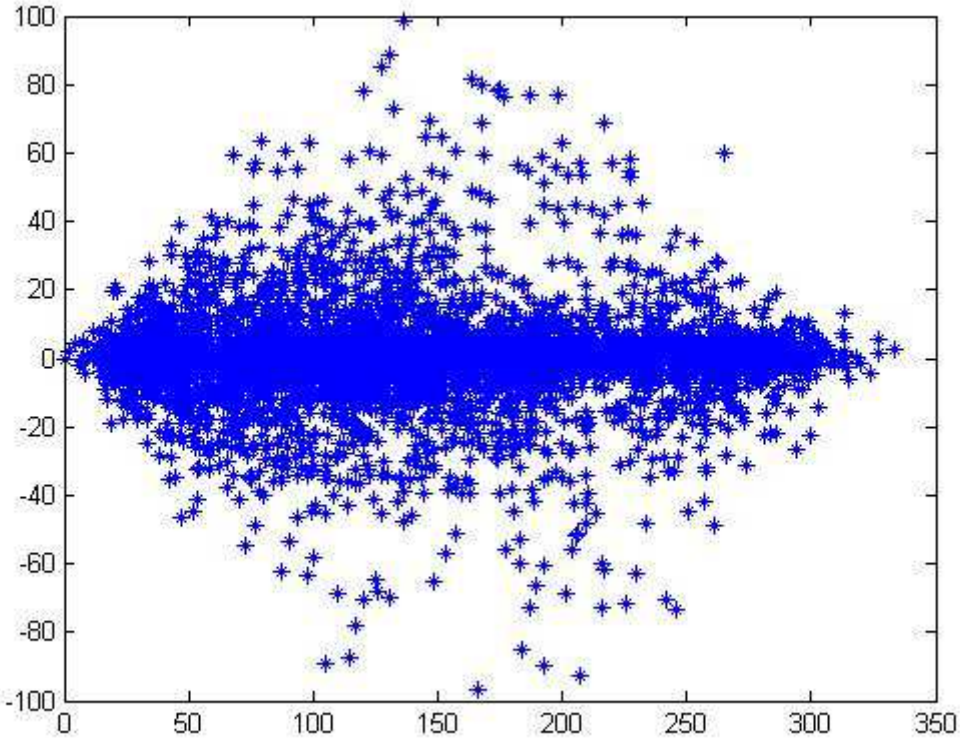
$$(x', y') = (x, y)\begin{pmatrix} cos45° & -sin45° \\ sin45° & cos45° \end{pmatrix} = (x, y)\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = (x, y)R \qquad (5.1)$$

όπου R ο ορθοκανονικός πίνακας περιστροφής (δηλαδή το εσωτερικό γινόμενο μιας γραμμής με τον εαυτό της είναι 1 και μεταξύ διαφορετικών γραμμών 0. Το ίδιο ισχύει και για τις στήλες). Ο αντίστροφος μετασχηματισμός είναι ο

$$(x, y) = (x', y')R^{-1} = (x', y')R^{T} = (x', y')\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \qquad (5.2)$$

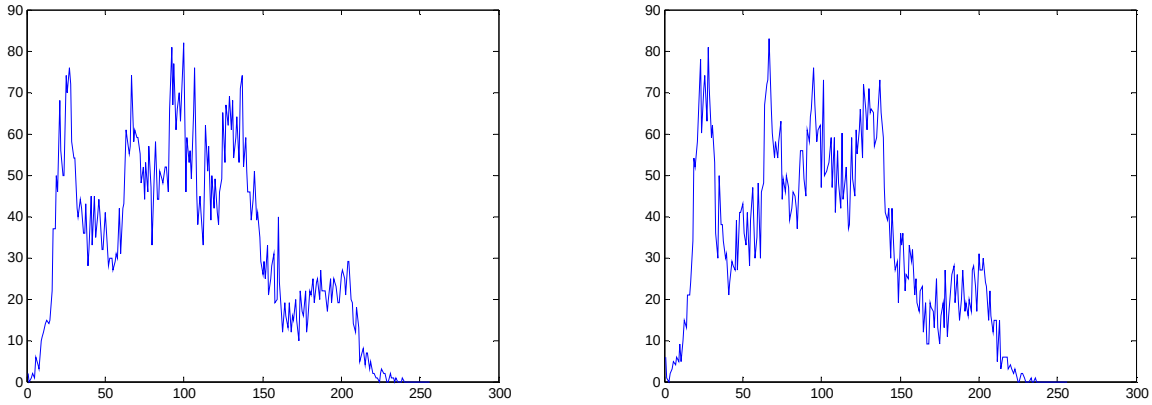
Παρατηρούμε ότι ο αντίστροφος ενός ορθοκανονικού πίνακα είναι ο ανάστροφος του.

Είναι προφανές πως τα περισσότερα σημεία στον άξονα y φαίνεται να έχουν τιμές οι οποίες είναι μηδέν ή βρίσκονται κοντά στο μηδέν ενώ οι τιμές στον άξονα x δεν έχουν αλλάξει ιδιαίτερα.



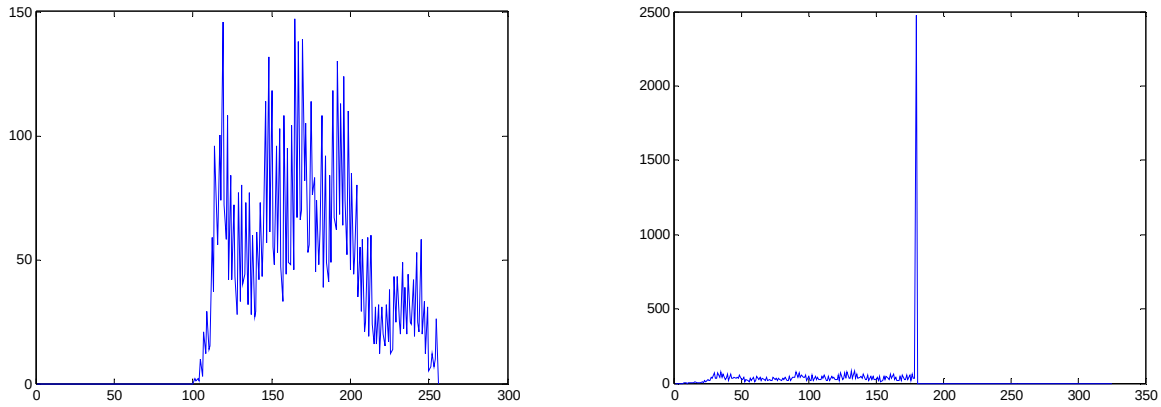
Σχήμα 5.2 : Περιστροφή των σημείων κατά 45 μοίρες

Στο Σχήμα 5.3 παρουσιάζεται η κατανομή των τιμών στους άξονες x και y της εικόνας της Lena μεγέθους 128×128, με 256 αποχρώσεις του γκρίζου, πριν τη στροφή των αξόνων. Είναι εμφανές πως οι δύο κατανομές δε διαφέρουν ιδιαίτερα. Στο Σχήμα 5.4 παρουσιάζεται η κατανομή μετά τη στροφή των αξόνων. Παρατηρούμε ότι η κατανομή στον άξονα x έχει παραμείνει σχεδόν ίδια ενώ ότι βρισκόταν στον άξονα y έχει πλέον τιμές κοντά στο μηδέν.



(a) (b)

Σχήμα 5.3 : Κατανομή των Τιμών στους Άξονες x και y αντίστοιχα



(a) (b)

Σχήμα 5.4 : Κατανομή των Τιμών στους Άξονες x και y αντίστοιχα μετά την στροφή των αξόνων

Εφόσον οι συντεταγμένες των σημείων είναι γνωστές πριν και μετά την περιστροφή, εύκολα μπορούμε να μετρήσουμε την μείωση στη συσχέτιση των τιμών. Μια απλή μετρική αποτελεί το άθροισμα $\sum_i x_i y_i$, γνωστή και ως ετεροσυσχέτιση των σημείων (x_i, y_i).

Παράδειγμα. Δίνονται πέντε σημεία $(5,5), (6,7), (12.1, 13.2), (23, 25)$ και $(32, 29)$ τα οποία τα στρέφουμε κατά 45° δεξιόστροφα και υπολογίζουμε την ετεροσυσχέτιση τους πριν και μετά την περιστροφή. Τα σημεία τα οποία προκύπτουν μετά την περιστροφή είναι τα $(7.071, 0), (9.19, 0.7071), (17.9, 0.78), (33.9, 1.41), (43.13, -2.12)$. Παρατηρούμε

πως όλες οι συντεταγμένες y είναι μικροί αριθμοί. Υπολογίζοντας την ετεροσυσχέτιση διαπιστώνουμε πως στην πρώτη περίπτωση, χωρίς την περιστροφή, η ετεροσυσχέτιση είναι 1729.72 και μετά την περιστροφή -23.0846, το οποίο αποτελεί μία σημαντική μείωση στη συσχέτιση.

Σε αυτό το σημείο η εικόνα μπορεί να συμπιεστεί γράφοντας σε μια ακολουθία δεδομένων τις μετασχηματισμένες τιμές των εικονοστοιχείων. Εάν η συμπίεση με απώλειες είναι αποδεκτή τότε όλες οι τιμές μπορούν να κβαντιστούν. Αυτό θα έχει ως αποτέλεσμα ακόμα μικρότερες τιμές. Υπάρχει επίσης η δυνατότητα να τοποθετηθούν σε αυτή την ακολουθία δεδομένων πρώτα οι τιμές του x (οι οποίες είναι οι μεγαλύτερες) και στη συνέχεια οι σημαντικά μικρότερες τιμές του y. Αυτό θα βοηθήσει τον κωδικοποιητή εντροπίας, όπως έχει ήδη αναφερθεί αλλά και όπως θα συζητηθεί αναλυτικότερα στη συνέχεια, αφού μετά την κβάντιση πολλοί συντελεστές θα έχουν τιμή μηδέν, γεγονός που θα επιτρέψει την καλύτερη συμπίεση των δεδομένων. Οι δύο αυτές ακολουθίες ονομάζονται διανύσματα συντελεστών του μετασχηματισμού.

Μπορεί να δειχθεί ότι η συνολική διακύμανση των τιμών των εικονοστοιχείων δεν αλλάζει με την περιστροφή αφού ο πίνακας περιστροφής είναι ορθοκανονικός. Συνεπώς, αφού η διακύμανση των y μετά την περιστροφή είναι πολύ μικρή, το μεγαλύτερο μέρος της διακύμανσης βρίσκεται τώρα συγκεντρωμένο στις τιμές των x. Η διακύμανση πολλές φορές ονομάζεται και ενέργεια της κατανομής των εικονοστοιχείων, επομένως, συμπεραίνουμε πως μετά την περιστροφή η ενέργεια συγκεντρώθηκε στις τιμές των x, επιτυγχάνοντας με αυτό τον τρόπο συμπίεση.

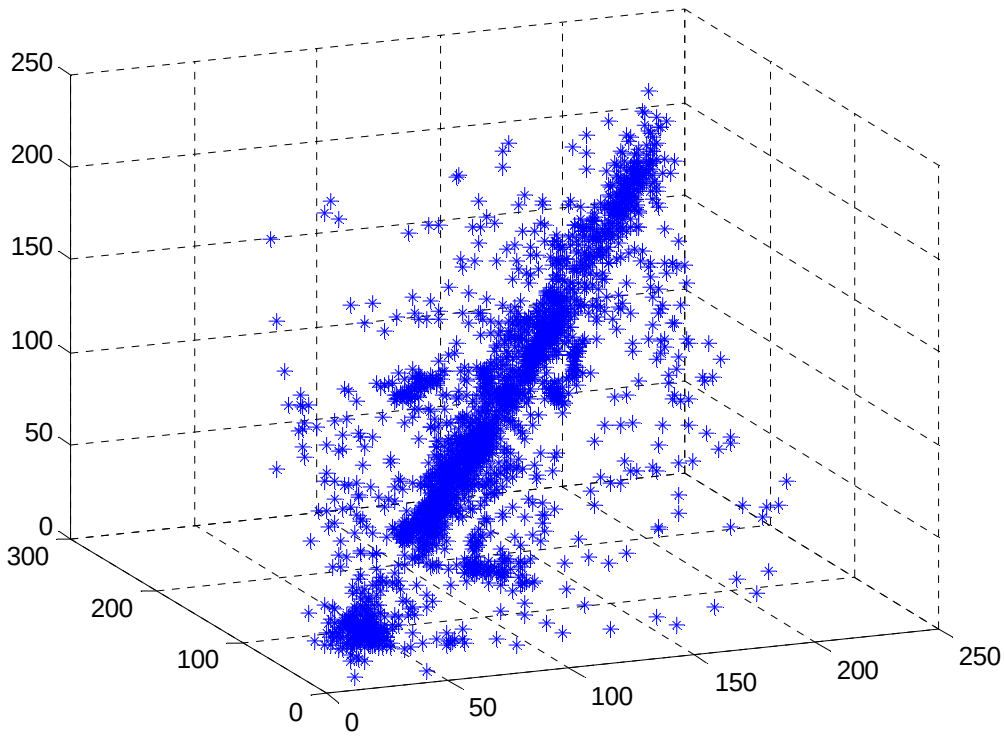
Συγκεντρώνοντας την ενέργεια στη μία από τις δύο διαστάσεις έχει και ένα επιπλέον πλεονέκτημα. Καθιστά την κβάντιση των τιμών αυτής της διάστασης πιο λεπτομερή και συνεπώς πιο αποδοτική σε σχέση με την κβάντιση των άλλων διαστάσεων. Αυτός ο τύπος κβάντισης επιτυγχάνει καλύτερη συμπίεση (με απώλειες).

Το ακόλουθο παράδειγμα παρουσιάζει τη δύναμη αυτού του βασικού μετασχηματισμού. Υποθέτουμε ότι ξεκινάμε από το σημείο $(4,5)$, των οποίων τόσο το x όσο και το y έχουν παρόμοιες τιμές. Χρησιμοποιώντας την εξίσωση (5.1) το μετασχηματισμού το σημείο αυτό μετασχηματίζεται σε $(4,5)R = (9,1)/\sqrt{2} \approx (6.36396, 0.7071)$. Η ενέργεια του σημείου αυτού και του μετασχηματισμού του είναι $4^2 + 5^2 = 41 = (9^2 + 1^2)/2$. Διαγράφοντας τον μικρότερο εκ των δύο συντελεστών (4) του αρχικού σημείου έχουμε ένα σφάλμα της τάξης του $\frac{4^2}{41} = 0.39$. Αντίθετα διαγράφοντας τον μικρότερο συντελεστή στην περίπτωση των μετασχηματισμένων σημείων (0.7071), το σφάλμα είναι της τάξης

του $\frac{0.7071^2}{41} = 0.012$. Ένας ακόμα τρόπος για να αντιληφθούμε τη διαφορά στο σφάλμα είναι να ανακατασκευάσουμε το αρχικό σημείο. Περνώντας το $\frac{1}{\sqrt{2}}(9,1)$ από τη συνάρτηση (5.2) του αντίστροφου μετασχηματισμού προφανώς ανακτάμε το αρχικό σημείο (4, 5). Μετά τη διαγραφή της τιμής του y, το σημείο που εισάγουμε στον μετασχηματισμό είναι το $\frac{1}{\sqrt{2}}(9,0)$, η ανακατασκευή του οποίου μας δίνει $(4.5, 4.5)$ και το οποίο προσεγγίζει πολύ καλά το αρχικό. Το σφάλμα μεταξύ του αρχικού και του ανακατασκευασμένου σημείου είναι και πάλι η ίδια πολύ μικρή τιμή:

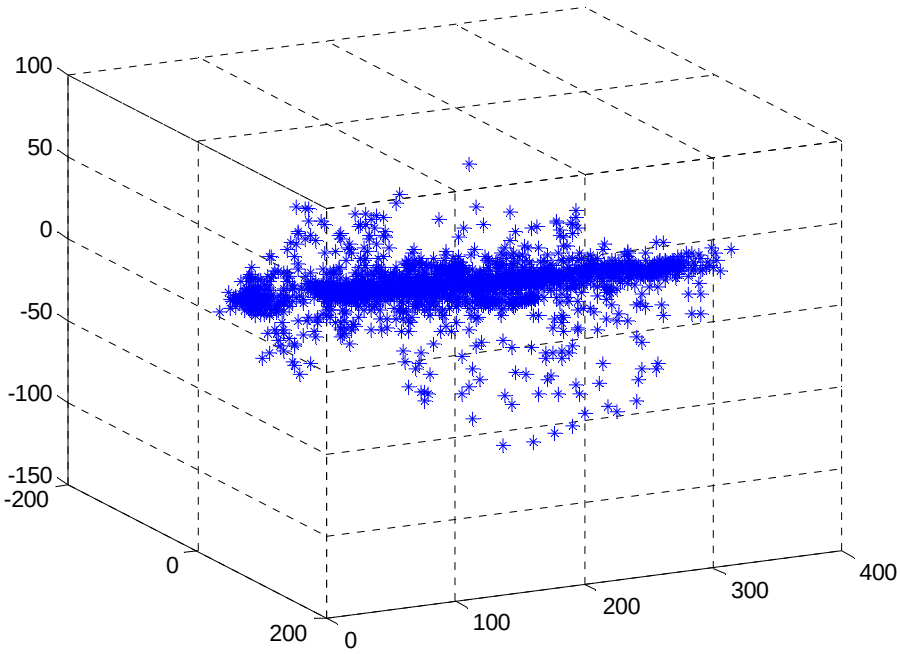
$$\frac{(4^2+5^2)-(4.5^2+4.5^2)}{4^2+5^2} = \frac{41-40.5}{41} = 0.0012$$

Αυτός ο απλός μετασχηματισμός μπορεί εύκολα να επεκταθεί σε οποιοδήποτε αριθμό διαστάσεων. Αντί για την επιλογή ζευγών γειτονικών εικονοστοιχείων μπορούμε να χρησιμοποιήσουμε τριάδες. Κάθε τριάδα αποτελεί ένα σημείο ενός τρισδιάστατου χώρου και τα σημεία αυτά είναι συγκεντρωμένα γύρω από μια ευθεία γραμμή η οποία σχηματίζει ίσες γωνίες με τους τρείς άξονες (Σχήμα 5.5).



Σχήμα 5.5 : Ζεύγη των Τιμών Φωτεινοτήτων Γειτονικών Εικονοστοιχείων ως Συντεταγμένες ενός Τρισδιάστατου Χώρου

Όταν αυτή η γραμμή περιστραφεί με τέτοιο τρόπο ώστε να συμπίπτει με τον άξονα x τότε οι συντεταγμένες y και z των μετασχηματισμένων σημείων γίνονται μικροί αριθμοί. Ο μετασχηματισμός πραγματοποιείται πολλαπλασιάζοντας κάθε σημείο με έναν 3×3 πίνακα περιστροφής ο οποίος θα είναι ορθοκανονικός. Στη συνέχεια τα μετασχηματισμένα σημεία χωρίζονται σε τρία διανύσματα συντελεστών εκ των οποίων τα δύο τελευταία αποτελούνται από μικρούς αριθμούς. Για μέγιστη συμπίεση, κάθε διάνυσμα συντελεστών πρέπει να κβαντίζεται ξεχωριστά.



Σχήμα 5.6 : Περιστροφή των σημείων κατά 45 μοίρες

Η περιστροφή στις τρείς διαστάσεις βασίζεται στη σχέση (5.1) και δίνεται από τους ακόλουθους πίνακες

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix} \qquad (5.3)$$

$$R_z(\theta) = \begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ομοίως με τη δισδιάστατη περίπτωση το αποτέλεσμα προκύπτει πολλαπλασιάζοντας τα τρισδιάστατα διανυσμάτων με τους πίνακες (5.3) διαδοχικά ώστε να επιτευχθεί η επιθυμητή περιστροφή. Στο συγκεκριμένο παράδειγμα, όπου $\theta = 45°$, τα διανύσματα πολλαπλασιάστηκαν με τους πίνακες R_x και R_z. Το αποτέλεσμα παρουσιάζεται στο Σχήμα 5.6.

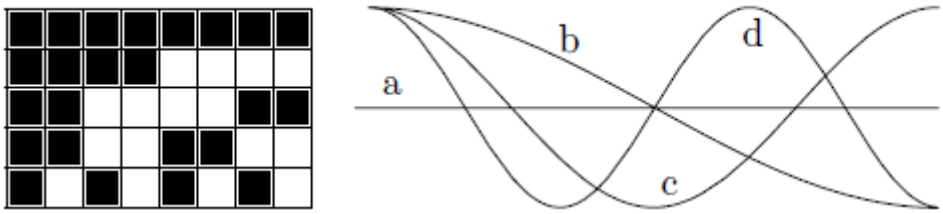
Η διαδικασία αυτή μπορεί να επεκταθεί και σε περισσότερες από τρείς διαστάσεις με μόνη διαφορά ότι δεν είναι δυνατό να παρασταθούν γραφικά οι χώροι περισσοτέρων από τριών διαστάσεων. Παρόλα αυτά από μαθηματικής άποψης η επέκταση είναι εύκολη. Κάποιες μέθοδοι συμπίεσης όπως η JPEG που θα εξετάσουμε και στη συνέχεια, διαιρεί την εικόνα σε μπλοκ εικονοστοιχείων, μεγέθους 8×8 έκαστο, στρέφοντας πρώτα κάθε γραμμή και στη συνέχεια κάθε στήλη. Αυτή η διπλή περιστροφή παράγει ένα σύνολο 64 μετασχηματισμένων τιμών εκ των οποίων η πρώτη ονομάζεται DC συνιστώσα και έχει μεγάλη τιμή ενώ οι υπόλοιπες 63 αποτελούν τις AC συνιστώσες οι οποίες κανονικά έχουν μικρές τιμές. Συνεπώς ο μετασχηματισμός αυτός συγκεντρώνει την ενέργεια του σήματος στην πρώτη από της 64 διαστάσεις. Το σύνολο των DC συνιστωσών και κάθε ένα από τα σύνολα των 63 AC συνιστωσών πρέπει να κβαντίζεται ξεχωριστά.

5.2 Ορθογώνιοι Μετασχηματισμοί

Οι μετασχηματισμοί εικόνας έχουν σχεδιαστεί ώστε να έχουν δύο ιδιότητες:

1. Να μειώνουν τον πλεονασμό της εικόνας μειώνοντας με τον τρόπο αυτό το μέγεθος των περισσοτέρων εικονοστοιχείων και
2. Να ανιχνεύουν τα λιγότερο σημαντικά σημεία της εικόνας εξαλείφοντας κάποιες από τις συχνότητες της εικόνας (κατά βάση τις υψηλές).

Συνεπώς σε αυτή την ενότητα θα εξετάσουμε την έννοια των συχνοτήτων σε μία εικόνα. Διαισθητικά συνδέουμε τη συχνότητα με ένα κύμα. Τα κύματα του νερού, τα ηχητικά κύματα όπως και τα ηλεκτρομαγνητικά κύματα έχουν συχνότητες. Όμως και οι εικόνες μπορούν επίσης να έχουν συχνότητες. Το Σχήμα 5.7 παρουσιάζει μία μικρή εικόνα 5×8 εικονοστοιχείων προκειμένου να αναδείξει αυτή την ιδέα. Η πρώτη γραμμή είναι ομοιόμορφη, συνεπώς μπορούμε να της αποδώσουμε μηδενική συχνότητα. Οι γραμμές που βρίσκονται από κάτω αυξάνουν συνεχώς συχνότητες με την έννοια ότι εναλλάσσονται συχνότερα οι τιμές του χρώματος κατά μήκος μιας γραμμής.



Σχήμα 5.7 : Συχνότητες Εικόνας

Οι συχνότητες σε μια εικόνα είναι σημαντικές για τον εξής λόγο: Οι χαμηλές συχνότητες αντιστοιχούν στη σημαντική πληροφορία της εικόνας ενώ οι υψηλές συχνότητες αντιστοιχούν στις λεπτομέρειες οι οποίες συνήθως είναι λιγότερο σημαντικές και ορατές από το μάτι του παρατηρητή. Επομένως, όταν ένας μετασχηματισμός διαχωρίζει τις επί μέρους συχνότητες της εικόνας, τότε τα εικονοστοιχεία τα οποία αντιστοιχούν στις υψηλές συχνότητες μπορούν να κβαντιστούν πιο έντονα ενώ τα εικονοστοιχεία που αντιστοιχούν σε χαμηλές συχνότητες να κβαντιστούν με μικρότερους συντελεστές ή ακόμα και καθόλου. Με αυτό τον τρόπο επιτυγχάνεται αποτελεσματική συμπίεση της εικόνας με απώλειες αλλά μόνο στην πληροφορία που σχετίζεται με ασήμαντες λεπτομέρειες της εικόνας.

Στην πράξη οι μετασχηματισμοί εικόνας πρέπει να είναι γρήγοροι και εύκολοι στην υλοποίηση. Αυτό συνιστά τη χρήση γραμμικών μετασχηματισμών. Στους μετασχηματισμούς αυτούς, κάθε μετασχηματισμένη τιμή (ή συντελεστής) c_i είναι ένα σταθμισμένο άθροισμα των τιμών των εικονοστοιχείων της εικόνας d_j , όπου κάθε στοιχείο πολλαπλασιάζεται από μια τιμή (βάρος) w_{ij} . Συνεπώς, $c_i = \sum_j d_j w_{ij}$, για $i, j = 1, 2, \dots, n$. Στην περίπτωση όπου $n = 4$, αυτό με συμβολισμό πινάκων εκφράζεται ως:

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}$$

Στη γενική περίπτωση το γράφουμε $C = WD$. Κάθε γραμμή του W ονομάζεται διάνυσμα βάσης. Οι μόνες τιμές που πρέπει να υπολογιστούν είναι οι τιμές βαρών w_{ij}. Οι αρχές που πρέπει να ακολουθηθούν είναι οι εξής:

1. Μείωση του πλεονασμού της πληροφορίας. Ο πρώτος συντελεστής c_1 μπορεί να είναι μεγάλος όμως οι υπόλοιποι συντελεστές c_2, c_3, ... πρέπει να είναι μικροί.
2. Διαχωρισμός συχνοτήτων. Ο πρώτος συντελεστής c_1 που προκύπτει από τον μετασχηματισμό θα πρέπει να αντιστοιχεί στη μηδενική συχνότητα

εικονοστοιχείων και οι υπόλοιποι συντελεστές να αντιστοιχούν σε όλο και μεγαλύτερες συχνότητες.

Κλειδί για τον καθορισμό των βαρών w_{ij} αποτελεί το γεγονός ότι τα δεδομένα μας d_j δεν είναι αυθαίρετοι αριθμοί αλλά τιμές εικονοστοιχείων, δηλαδή μη αρνητικές και συσχετισμένες.

Η βασική σχέση $c_i = \sum_j d_j w_{ij}$ συνιστά ότι ο πρώτος συντελεστής c_1 θα είναι μεγάλος όταν όλα τα βάρη w_{1j} είναι θετικά. Προκειμένου να κάνουμε τους υπόλοιπους συντελεστές c_i μικρούς, είναι εύκολο να ορίσουμε τα μισά βάρη να είναι θετικά και τα άλλα μισά αρνητικά. Μία απλή επιλογή είναι να αποδώσουμε τιμές +1 στα μισά βάρη και τιμές -1 στα άλλα μισά. Στην ακραία περίπτωση όπου όλα τα εικονοστοιχεία d_j είναι πανομοιότυπα θα έχουμε μηδενικούς συντελεστές c_i. Όταν οι τιμές των εικονοστοιχείων d_j είναι παρόμοιες τότε οι συντελεστές c_i θα είναι μικροί (θετικοί ή αρνητικοί).

Η επιλογή των βαρών που καθορίσαμε παραπάνω ικανοποιεί την πρώτη απαίτηση σχετικά με τη μείωση του πλεονασμού στο μετασχηματισμένο σήμα. Προκειμένου να ικανοποιηθεί και η δεύτερη απαίτηση, τα βάρη w_{ij} της γραμμής i θα πρέπει να παρουσιάζουν ολοένα και μεγαλύτερες συχνότητες καθώς μεγαλώνει το i. Τα βάρη w_{1j} θα πρέπει να έχουν μηδενική συχνότητα συνεπώς η τιμή τους θα είναι +1. Τα βάρη w_{2j} θα πρέπει να έχουν μία αλλαγή πρόσημου, δηλαδή +1 +1, … , +1, -1, -1, … , -1. Σύμφωνα με τη μαθηματική ορολογία των διανυσματικών χώρων θα χρησιμοποιούμε τον όρο διανύσματα βάσης για τις γραμμές των βαρών.

Επιπροσθέτως, για να επιτύχουμε διαχωρισμό των συχνοτήτων των εικονοστοιχείων d_j θα πρέπει τα διανύσματα βάσης να είναι ορθογώνια. Τα διανύσματα βάσης όπως αναφέραμε είναι οι γραμμές του πίνακα W και αυτός είναι και ο λόγος που ο πίνακας αυτός, κατά συνέπεια και ολόκληρος ο μετασχηματισμός, είναι ορθογώνιος. Οι θεωρήσεις αυτές ικανοποιούνται από τον ορθογώνιο πίνακα:

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \tag{5.4}$$

Το πρώτο διάνυσμα βάσης (η πρώτη γραμμή του W) αποτελείται μόνο από άσσους, συνεπώς η συχνότητα του είναι μηδέν. Κάθε ένα από τα υπόλοιπα διανύσματα έχει δύο βάρη $+1$ και δύο βάρη -1, επομένως θα παραχθούν μικρές μετασχηματισμένες τιμές και οι συχνότητες τους μεγαλώνουν από γραμμή σε γραμμή καθώς μεταβάλλεται και το

πλήθος των αλλαγών πρόσημου κατά μήκος μιας γραμμής. Ο πίνακας αυτός είναι παρόμοιος με τον μετασχηματισμό Walsh-Hadamard[33][34].

Προκειμένου να παρουσιαστεί πως ο πίνακας αυτός αναγνωρίζει τις συχνότητες σε ένα διάνυσμα δεδομένων, πολλαπλασιάζουμε τα ακόλουθα διανύσματα με τον πίνακα W:

$$W\begin{bmatrix}1\\0\\0\\1\end{bmatrix}=\begin{bmatrix}2\\0\\2\\0\end{bmatrix},\quad W\begin{bmatrix}0\\0.33\\-0.33\\-1\end{bmatrix}=\begin{bmatrix}0\\2.66\\0\\1.33\end{bmatrix},\quad W\begin{bmatrix}1\\0\\0\\0\end{bmatrix}=\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\quad W\begin{bmatrix}1\\-0.8\\1\\-0.8\end{bmatrix}=\begin{bmatrix}0.4\\0\\0\\3.6\end{bmatrix}$$

Τα αποτελέσματα αποκτούν νόημα όταν εντοπίσουμε τον τρόπο με τον οποίο καθορίστηκαν τα τέσσερα δοκιμαστικά διανύσματα.

$$(1,0,0,1)=0.5(1,1,1,1)+0.5(1,-1,-1,1),$$

$$(0,0.33,-0.33,-1)=0.66(1,1,-1,-1)+0.33(1,-1,1,-1),$$

$$(1,0,0,0)=0.25(1,1,1,1)+0.25(1,1,-1,-1)+0.25(1,-1,-1,1)+0.25(1,-1,1,-1),$$

$$(1,-0.8,1,-0.8)=0.1(1,1,1,1)+0.9(1,-1,1,-1)$$

Στο πρώτο διάνυσμα δεδομένων παρατηρούμε πως η δεύτερη και η τέταρτη γραμμή του μετασχηματισμένου διανύσματος έχουν τιμή μηδέν, συνεπώς τα διανύσματα συντελεστών που θα χρησιμοποιηθούν θα είναι το πρώτο και το τρίτο, δηλαδή τα (1, 1, 1, 1) και (1, -1, -1, 1). Οι συντελεστές προκύπτουν από το γεγονός ότι όσο μέρος του σήματος περιέχεται στην πρώτη συχνότητα (2) το ίδιο περιέχεται και στην τέταρτη συχνότητα. Επομένως αφού δύο μετασχηματισμένοι συντελεστές μοιράζονται ίσο μέρος από την πρώτη και την τέταρτη συχνότητα ο συντελεστής θα είναι 0.5

Αντίστοιχα στο τέταρτο διάνυσμα, οι μετασχηματισμένοι συντελεστές που έχουν μη μηδενική τιμή είναι ο πρώτος και ο τέταρτος άρα τα διανύσματα συντελεστών που χρησιμοποιούνται είναι οι $(1,1,1,1)$ και $(1,-1,1,-1)$. Η διαφορά του πρώτου συντελεστή από τον τέταρτο είναι $\frac{3.6}{0.4}=9$. Αυτό σημαίνει πως το διάνυσμα δεδομένων περιέχει 9 φορές περισσότερο σήμα στις ψηλές συχνότητες (τέταρτο διάνυσμα βαρών) από ότι στις χαμηλές συχνότητες (πρώτο διάνυσμα). Για το λόγο αυτό συντελεστής στο πρώτο διάνυσμα είναι το 0.1 ενώ στο τέταρτο το 0.9.

Είναι επίσης δυνατό να τροποποιήσουμε τον μετασχηματισμό αυτό προκειμένου να διατηρηθεί η ενέργεια του διανύσματος δεδομένων. Το μόνο που απαιτείται είναι ο πολλαπλασιασμός του πίνακα μετασχηματισμού W με έναν σταθερό παράγοντα ½ . Επομένως, το διανυσματικό γινόμενο $\left(\frac{w}{2}\right)(a,b\,,c,d)$ έχει την ίδια ενέργεια a^2+b^2+

$c^2 + d^2$ με το διάνυσμα δεδομένων (a, b, c, d). Ως παράδειγμα μπορεί να αναφερθεί το εσωτερικό γινόμενο του $W/2$ και του έντονα συσχετισμένου διανύσματος $(5, 6, 7, 8)$. Στη μετασχηματισμένη του μορφή οι συντελεστές που προκύπτουν είναι $(13, -2, 0, -1)$ όπου ο πρώτος συντελεστής έχει μεγάλη τιμή και οι υπόλοιποι έχουν μικρότερες τιμές από τις αρχικές τους. Η ενέργεια τόσο στο διάνυσμα $(5, 6, 7, 8)$ όσο και στο $(13, -2, 0, -1)$ είναι 174 όμως ενώ στο αρχικό διάνυσμα ο πρώτος συντελεστής αποτελεί μόνο το 14% της ενέργειας του σήματος, στο μετασχηματισμένο διάνυσμα αποτελεί το 97% της ενέργειας. Με αυτό τον τρόπο ο απλός αυτός μετασχηματισμός συγκεντρώνει τη συνολική ενέργεια του σήματος σε λίγους συντελεστές.

Ένα ακόμα πλεονέκτημα του W είναι ότι εκτελεί και τον αντίστροφο μετασχηματισμό. Το εσωτερικό γινόμενο $\left(\frac{W}{2}\right)(13, -2, 0, -1)^T$ ανακατασκευάζει τα αρχικά μας δεδομένα $(5, 6, 7, 8)$.

Μετά από αυτή την αναφορά στις ιδέες που εφαρμόζονται για τη συμπίεση της εικόνας ας δούμε μέσω ενός παραδείγματος τις δυνατότητες ενός τέτοιου μετασχηματισμού. Χρησιμοποιούμε τον πίνακα $W/2$ για να μετασχηματίσουμε το όχι και πολύ έντονα συσχετισμένο διάνυσμα δεδομένων $d = (4, 6, 5, 2)$. Το αποτέλεσμα είναι $t = (8.5, 1.5, -2.5, 0.5)$. Εύκολα μετασχηματίζουμε το διάνυσμα t πίσω στο d, όμως η διαδικασία αυτή από μόνη της δεν παρέχει κάποιας μορφής συμπίεση. Προκειμένου να επιτευχθεί συμπίεση κβαντίζουμε τους συντελεστές του t και όπως θα διαπιστώσουμε, ακόμα και με έντονη κβάντιση, είναι εφικτό να ανακατασκευάσουμε το αρχικό διάνυσμα και να είναι πολύ κοντά στις αρχικές τιμές του διανύσματος.

Αρχικά κβαντίζουμε το t στους ακεραίους $(9, 1, -3, 0)$ και εκτελούμε τον αντίστροφο μετασχηματισμό. Το ανακατασκευασμένο διάνυσμα είναι το $(3.5, 6{,}5, 5.5, 2.5)$. Ομοίως μπορούμε να διαγράψουμε τους δύο μικρότερους συντελεστές του διανύσματος t και να επιχειρήσουμε να ανακατασκευάσουμε το μετασχηματισμένο μας διάνυσμα από το διάνυσμα $(8.5, 0, -2.5, 0)$. Οι τιμές που παίρνουμε μετά την εφαρμογή του αντίστροφου μετασχηματισμού είναι $(3, 5.5, 5.5, 3)$, οι οποίες εξακολουθούν να βρίσκονται πολύ κοντά στις αρχικές τιμές του d. Συμπεραίνουμε λοιπόν ότι ακόμα και αυτός ο απλός, διαισθητικός μετασχηματισμός αποτελεί ένα ισχυρό εργαλείο για την εξάλειψη του πλεονασμού των δεδομένων. Πιο εξελιγμένοι μετασχηματισμοί παρέχουν ακόμα καλύτερα αποτελέσματα όπου είναι δυνατό οι συντελεστές που προκύπτουν από το μετασχηματισμό να κβαντιστούν έντονα και πάλι να είναι δυνατή η ανακατασκευή των δεδομένων σε πολύ υψηλό βαθμό.

5.3 Δισδιάστατοι Ορθογώνιοι Μετασχηματισμοί

Δοθέντος ενός μπλοκ δεδομένων υπό τη μορφή πίνακα μεγέθους 4×4

$$D = \begin{pmatrix} 5 & 6 & 7 & 4 \\ 6 & 5 & 7 & 5 \\ 7 & 7 & 6 & 6 \\ 8 & 8 & 8 & 8 \end{pmatrix},$$

όπου οι τιμές των στοιχείων κατά στήλες είναι ισχυρά συσχετισμένα, μπορεί να εφαρμοστεί ο απλός μετασχηματισμός που χρησιμοποιήσαμε ως τώρα στις στήλες του D. Το αποτέλεσμα είναι:

$$C' = WD = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} D = \begin{pmatrix} 26 & 26 & 28 & 23 \\ -4 & -4 & 0 & -5 \\ 0 & 2 & 2 & 1 \\ -2 & 0 & -2 & -3 \end{pmatrix}$$

Κάθε στήλη του πίνακα C' αποτελεί τον μετασχηματισμό κάθε στήλης του D. Παρατηρούμε ότι το πρώτο στοιχείο κάθε στήλης του C' υπερισχύει των υπολοίπων γεγονός που δηλώνει ότι οι τιμές των στοιχείων στις στήλες είναι συσχετισμένες. Ο υπολογισμός του πίνακα C' είναι το πρώτο από τα δύο βήματα προκειμένου να παραχθεί ο δισδιάστατος μετασχηματισμός του πίνακα D. Στο δεύτερο βήμα μετασχηματίζεται κάθε γραμμή του πίνακα C' και αυτό επιτυγχάνεται με τον πολλαπλασιασμό του πίνακα C' με τον ανάστροφο W^T. Δεδομένου ότι ο συγκεκριμένος πίνακας W που χρησιμοποιούμε είναι συμμετρικός καταλήγουμε στη σχέση:

$$C = C'W^T = WDW^T = WDW \tag{5.5}$$

ή στο παράδειγμά μας

$$C = \begin{pmatrix} 26 & 26 & 28 & 23 \\ -4 & -4 & 0 & -5 \\ 0 & 2 & 2 & 1 \\ -2 & 0 & -2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 103 & 1 & -5 & 5 \\ -13 & -3 & -5 & 5 \\ 5 & -1 & -3 & -1 \\ -7 & 3 & -3 & -1 \end{pmatrix}$$

Το επάνω αριστερό στοιχείο παρατηρούμε πως κυριαρχεί έναντι των υπολοίπων. Περιέχει το μεγαλύτερο μέρος τις ενέργειας του αρχικού πίνακα D. Τα στοιχεία που βρίσκονται επάνω και αριστερά του πίνακα είναι λίγο μεγαλύτερα από τα αρχικά στοιχεία του πίνακα D ενώ τα υπόλοιπα είναι μικρότερα. Ο δισδιάστατος αυτός μετασχηματισμός, μείωσε τη συσχέτιση τόσο στην οριζόντια όσο και στην κατακόρυφη διάσταση. Όπως συνέβη και στη μονοδιάστατη περίπτωση που μελετήσαμε στην ενότητα 5.2, μπορεί να επιτευχθεί εξαιρετική συμπίεση με κβάντιση των συντελεστών του πίνακα C, ιδίως όσων αντιστοιχούν στις ψηλότερες συχνότητες (δηλαδή όσους

συντελεστές βρίσκονται προς την κάτω δεξιά γωνία).

Μόλις παρουσιάστηκε η ουσία των δισδιάστατων ορθογώνιων μετασχηματισμών. Στη συνέχεια θα αναφερθούμε συνοπτικά στους μετασχηματισμούς:

1. Walsh-Hadamard (WHT), ο οποίος είναι γρήγορος στον υπολογισμό του καθώς απαιτεί μόνο προσθέσεις και αφαιρέσεις, όμως η απόδοση του από πλευράς συγκέντρωσης ενέργειας είναι χαμηλότερη από αυτή του μετασχηματισμού DCT.
2. Ο μετασχηματισμός Haar[30], ο οποίος υπολογίζεται με έναν απλό και γρήγορο αλγόριθμος. Αποτελεί τον απλούστερο μετασχηματισμό κυματιδίου και θα τον εξετάσουμε αναλυτικότερα στην πορεία αυτής της εργασίας.
3. Ο μετασχηματισμός Karhunen-Loeve[31][32] (KLT) ο οποίος θεωρητικά είναι ο καλύτερος από την άποψη ότι συγκεντρώνει την ενέργεια με τον πιο αποδοτικό τρόπο. Παρόλα αυτά, οι συντελεστές του δεν είναι σταθεροί καθώς εξαρτώνται από τα δεδομένα που πρόκειται να συμπιεστούν. Ο υπολογισμός αυτών των συντελεστών (της βάσης, δηλαδή, του μετασχηματισμού) είναι μια αργή διαδικασία καθώς και ο ίδιος ο υπολογισμός του μετασχηματισμού. Επίσης, εφόσον οι συντελεστές του μετασχηματισμού εξαρτώνται από τα δεδομένα εισόδου, πρέπει να σταλούν και αυτοί στο δέκτη προκειμένου να είναι δυνατή η αποκωδικοποίηση τους. Για τους λόγους αυτούς και επειδή ο μετασχηματισμός DCT αποδίδει εξίσου καλά, ο KLT δεν χρησιμοποιείται γενικά στην πράξη.
4. Ο μετασχηματισμός DCT που θα περιγραφεί αναλυτικότατα. Ο DCT αποδίδει εξίσου καλά με τον KLT υπό την έννοια συγκέντρωσης της ενέργειας αλλά χρησιμοποιεί σταθερή βάση, ανεξάρτητα από τα δεδομένα εισόδου. Υπάρχουν επίσης και γρήγοροι αλγόριθμοι για τον υπολογισμό του DCT. Η μέθοδος αυτή χρησιμοποιείται από τα πρότυπα JPEG και MPEG.

5.3.1 Walsh-Hadamard Transform

Όπως ήδη αναφέρθηκε, ο μετασχηματισμός αυτός έχει μικρή απόδοση συμπίεσης, συνεπώς δεν χρησιμοποιείται συχνά στην πράξη. Παρόλα αυτά είναι γρήγορος αφού υπολογίζεται μόνο με προσθέσεις, αφαιρέσεις και περιστασιακές διαιρέσεις με δυνάμεις του δύο οι οποίες επιτυγχάνονται με δεξιές ολισθήσεις.

Δοθέντος ενός μπλοκ από $N \times N$ εικονοστοιχεία P_{xy} (όπου το N πρέπει να είναι δύναμη του 2, $N = 2^n$), ο δισδιάστατος μετασχηματισμός WHT και ο αντίστροφος του ορίζονται από τις εξισώσεις (5.6) και (5.7):

$$H(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} p_{xy} g(x,y,\ u,v) => \tag{5.6}$$

$$H(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} p_{xy}(-1)\sum_{i=0}^{n-1}[b_i(x)p_i(x) + b_i(y)p_i(v)]$$

$$P_{xy} = \sum_{u=0}^{N-1}\sum_{u=0}^{N-1} H(u,v)h(x,y,u,v) =>$$
$$P_{xy} = \frac{1}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} H(u,v)(-1)\sum_{i=0}^{n-1}[b_i(x)p_i(u) + b_i(y)p_i(v)] \tag{5.7}$$

Όπου $H(u,v)$ είναι τα αποτελέσματα του μετασχηματισμού (δηλαδή οι συντελεστές WHT), η ποσότητα $b_i(u)$ είναι το bit i της δυαδικής αναπαράστασης του ακεραίου u και η $p_i(u)$ ορίζεται σε σχέση με την $b_j(u)$ από της εξισώσεις:

$$\begin{aligned} p_0(u) &= b_{n-1}(u), \\ p_1(u) &= b_{n-1}(u) + b_{n-2}(u), \\ p_2(u) &= b_{n-2}(u) + b_{n-3}(u), \\ &\vdots \\ p_{n-1}(u) &= b_1(u) + b_0(u), \end{aligned} \tag{5.8}$$

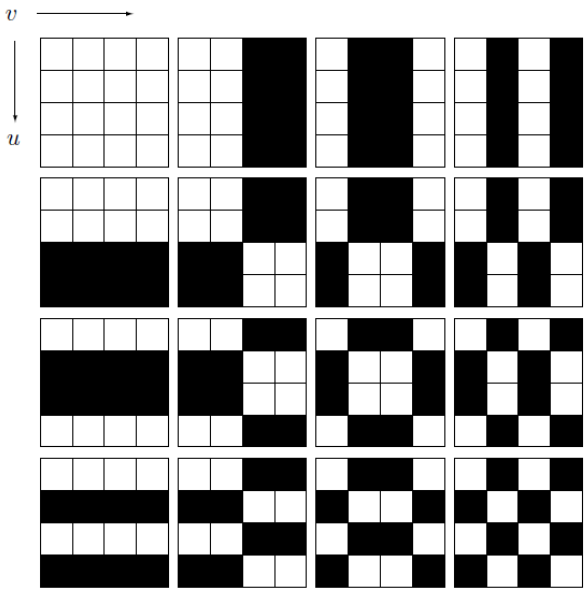
Υπενθυμίζεται ότι το n ορίστηκε παραπάνω από τη σχέση $N = 2^n$.

Ως παράδειγμα, θεωρούμε $u = 6 = 110_2$ οπότε $b_0(6) = 0, b_1(6) = 1$ και $b_2(6) = 1$.

Οι ποσότητες $g(x,y,u,v)$ και $h(x,y,u,v)$ ονομάζονται πυρήνες (Kernels) ή εικόνες βάσης του WHT. Οι πίνακες αυτού είναι πανομοιότυποι. Τα στοιχεία τους είναι τιμές $+1$ και -1 πολλαπλασιασμένες με τον παράγοντα $\frac{1}{N}$. Συνεπώς ο μετασχηματισμός WHT αποτελείται από τον πολλαπλασιασμό κάθε εικονοστοιχείου της εικόνας με τις τιμές $+1$ ή -1, αθροίζοντας και διαιρώντας το άθροισμα με το N. Δεδομένου ότι το $N = 2^n$ άρα δύναμη του δύο, η διαίρεση μπορεί να επιτευχθεί με διαδοχικές ολισθήσεις κατά n θέσεις προς τα δεξιά.

Οι πυρήνες WHT παρουσιάζονται με γραφική αναπαράσταση για $N = 4$ στο Σχήμα 5.8, όπου λευκό σημαίνει τιμή $+1$ και όπου μαύρο τιμή -1 (ο παράγοντας $\frac{1}{N}$ έχει παραληφθεί). Οι γραμμές και οι στήλες των μπλοκ του σχήματος σε αυτή την εικόνα αντιστοιχούν στις τιμές των u και v από 0 μέχρι 3 αντίστοιχα. Το πλήθος των αλλαγών

πρόσημου κατά μήκος μιας στήλης ή μιας γραμμής του πίνακα καθορίζουν τη συχνότητα της γραμμής η της στήλης. Οι γραμμές και οι στήλες στο σχήμα έχουν ταξινομηθεί κατά αύξουσα σειρά συχνότητας.



Σχήμα 5.8 : Ο Πυρήνας WHT για N=4

Η διαδικασία συμπίεσης εικόνας με τον μετασχηματισμό WHT είναι αντίστοιχη με τη διαδικασία συμπίεσης που χρησιμοποιεί ο DCT με τη διαφορά ότι χρησιμοποιούνται διαφορετικές εξισώσεις. Με τη χρήση λογισμικών όπως το Matalb είναι εύκολο να υπολογιστούν και να αναπαρασταθούν με γραφική απεικόνιση οι εικόνες βάσης του WHT για N=8.

5.3.2 Μετασχηματισμός Haar

Ο μετασχηματισμός Haar είναι βασισμένος στις συναρτήσεις Haar $h_k(x)$, οι οποίες ορίζονται στο $\chi \in [0,1]$ και για $\kappa = 0, 1, \dots, N-1$, όπου $N = 2^n$. Προτού αναφερθούμε στον μετασχηματισμό πρέπει να σημειωθεί ότι κάθε ακέραιος k μπορεί να εκφρασθεί ως άθροισμα $k = 2^p + p - 1$, όπου $0 \le p \le n-1, q = 0$ ή 1 για $p = 0$ και $1 \le q \le 2^p$ για $p \ne 0$. Με $N = 4 = 2^2$, για παράδειγμα ισχύει ότι $0 = 2^0 + 0 - 1$, $1 = 2^0 + 1 - 1$, $2 = 2^1 + 1 - 1$ και $3 = 2^1 + 2 - 1$. Η βάση Haar ορίζεται λοιπόν ως:

$$h_0(x) \stackrel{\text{def}}{=} h_{00}(x) = \frac{1}{\sqrt{N}} \; \gamma\iota\alpha \; 0 \le x \le 1, \tag{5.9}$$

και

$$h_k(x) \stackrel{\text{def}}{=} h_{pq}(x) = \frac{1}{\sqrt{N}} \begin{cases} 2^{\frac{p}{2}}, & \frac{q-1}{2^p} \le x < \frac{q-\frac{1}{2}}{2^p}, \\ -2^{\frac{p}{2}}, & \frac{q-1}{2^p} \le x < \frac{q}{2^p}, \\ 0, \textit{διαφορετικά για } x \in [0,1]. \end{cases} \tag{5.10}$$

Ο πίνακας μετασχηματισμού Haar A_N της τάξης $N \times N$ μπορεί πλέον να κατασκευαστεί. Ένα οποιοδήποτε στοιχείο i, j αυτού του πίνακα είναι μια βάση της συνάρτησης $h_i(j)$, όπου $i = 0, 1, \dots, N-1$ και $j = \frac{0}{N}, \frac{1}{N}, \dots, (N-1)/N$. Για παράδειγμα:

$$A_2 = \begin{pmatrix} h_0(0/2) & h_0(1/2) \\ h_1(0/2) & h_1(1/2) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Υπενθυμίζεται ότι $i = 1$ $\textit{συνεπάγεται}$ $p = 0$ $\textit{και}$ $q = 1$.

Ως παράδειγμα υπολογίζουμε τους πίνακες συντελεστών Haar A_4 $\textit{και}$ A_8 . Ο A_4 είναι ο 4×4 πίνακας:

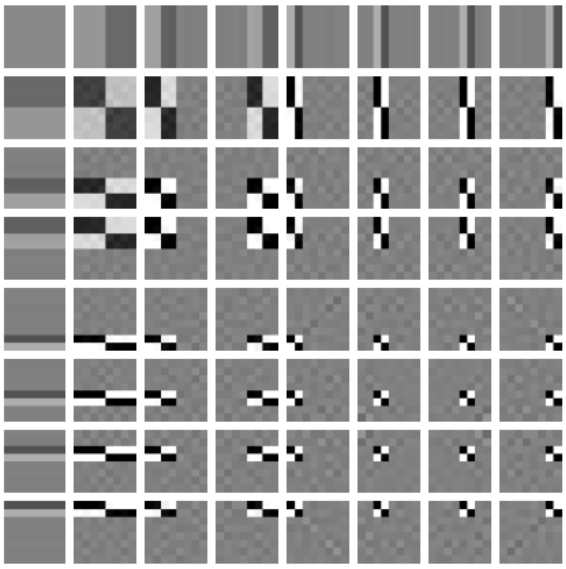
$$A_4 = \begin{pmatrix} h_0(0/4) & h_0(1/4) & h_0(2/4) & h_0(3/4) \\ h_1(0/4) & h_1(1/4) & h_1(2/4) & h_1(3/4) \\ h_2(0/4) & h_2(1/4) & h_2(2/4) & h_2(3/4) \\ h_3(0/4) & h_3(1/4) & h_3(2/4) & h_3(3/4) \end{pmatrix} = \frac{1}{\sqrt{4}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{pmatrix}$$

ομοίως, A_8 είναι ο πίνακας

$$A_8 = \frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{pmatrix}$$

Η εικόνα βάσης του μετασχηματισμού Haar παρουσιάζεται στο Σχήμα 5.9.

Δοθέντος ενός μπλοκ εικόνας της τάξης $N \times N$ όπου $N = 2^n$, ο μετασχηματισμός Haar του μπλοκ είναι το εσωτερικό γινόμενο $A_N P A_N$.



Σχήμα 5.9 : Η Εικόνα Βάσης του μετασχηματισμού Haar για $n = 8$

5.3.3 Μετασχηματισμός Karhunen-Loeve

Ο μετασχηματισμός Karhunen-Loeve προσφέρει την μέγιστη αποδοτικότητα από άποψη συγκέντρωσης της ενέργειας αλλά για τους λόγους που ήδη έχουν προαναφερθεί, έχει περισσότερο θεωρητική παρά πρακτική αξία. Δοθείσας μίας εικόνας, τη χωρίζουμε σε k μπλοκ με n εικονοστοιχεία το καθένα, όπου n συνήθως, αλλά όχι απαραίτητα, είναι το 64 και το k εξαρτάται από το μέγεθος της εικόνας. Θεωρούμε τα μπλοκ ως διανύσματα και τα συμβολίζουμε με $b^{(i)}$, για $i = 1, 2, \dots, k$. Η μέση τιμή του διανύσματος είναι $\bar{b} = \frac{\sum_i b^{(i)}}{k}$. Ένα νέο σύνολο διανυσμάτων ορίζεται ως $v^{(i)} = b^{(i)} - \bar{b}$ θέτοντας με αυτό τον τρόπο τη μέση τιμή $\frac{\sum_i v^{(i)}}{k}$ να είναι μηδέν. Συμβολίζουμε τον πίνακα $n \times n$ του KLT μετασχηματισμού τον οποίο και αναζητάμε με A. Το αποτέλεσμα του μετασχηματισμού του διανύσματος $v^{(i)}$ είναι το σταθμισμένο διάνυσμα $w^{(i)} = Av^{(i)}$. Η μέση τιμή του $w^{(i)}$ είναι επίσης μηδέν. Στη συνέχεια κατασκευάζουμε έναν πίνακα V του οποίου οι στήλες είναι τα διανύσματα $v^{(i)}$ και έναν πίνακα W του οποίου οι στήλες είναι τα διανύσματα βαρών $w^{(i)}$:

$$V = \left(v^{(1)}, v^{(2)}, \dots, v^{(k)}\right), \qquad W = \left(w^{(1)}, w^{(2)}, \dots, w^{(k)}\right)$$

Οι πίνακες V και W έχουν n γραμμές και k στήλες η κάθε μια. Από τον ορισμό του $w^{(i)}$ έχουμε $W = AV$

Τα n διανύσματα συντελεστών $c^{(j)}$ του μετασχηματισμού Karhunen-Loeve δίνονται από το

$$c^{(j)} = \left(w_j^{(1)}, w_j^{(2)}, \dots, w_j^{(k)}\right), \qquad j = 1,2,\dots,n$$

Συνεπώς το διάνυσμα $c^{(j)}$ αποτελείται από τα j στοιχεία όλων των διανυσμάτων βαρών $w^{(i)}$ για $i = 1,2,\dots,k$ ($c^{(j)}$ είναι η j^{η} συντεταγμένη των $w^{(i)}$ διανυσμάτων.

Στη συνέχεια εξετάζουμε τα στοιχεία του πίνακα που προκύπτουν από το γινόμενο WW^T (το οποίο θα είναι ένας πίνακας μεγέθους $n \times n$. Όπως είναι γνωστό, ένα στοιχείο του πίνακα που βρίσκεται στη γραμμή a και στη στήλη b είναι το άθροισμα των γινομένων:

$$(WW^T)_{ab} = \sum_{i=1}^{k} w_a^{(i)} w_b^{(i)} = \sum_{i=1}^{k} c_i^{(a)} c_i^{(b)} = c^{(a)} c^{(b)}, \text{για } a, b \in [1, n] \qquad (5.11)$$

Το γεγονός ότι ο μέση τιμή καθενός από τα $w^{(i)}$ είναι μηδέν συνεπάγεται ότι οποιοδήποτε στοιχείο της διαγωνίου $(WW^T)_{jj}$ του πίνακα που προέκυψε από το γινόμενο των πινάκων, είναι η διασπορά (μέχρι έναν παράγοντα k) του $j^{ου}$ στοιχείου των διανυσμάτων $w^{(i)}$. Αυτό ασφαλώς είναι η διασπορά του διανύσματος διασποράς $c^{(j)}$.

Για να αντιληφθούμε το λόγο για τον οποίο αυτό είναι αληθές σκεφτόμαστε την ακόλουθη περίπτωση. Η μέση τιμή του διανύσματος $w^{(i)}$ είναι μηδέν συνεπώς η εξίσωση (5.11) γίνεται:

$$(WW^T)_{jj} = \sum_{i=1}^{k} w_j^{(i)} w_j^{(i)} = \sum_{i=1}^{k} \left(w_j^{(i)} - 0\right)^2 = \sum_{i=1}^{k} \left(c_i^{(j)} - 0\right)^2 = k\ \Delta\iota\alpha\sigma\pi o\rho\acute{\alpha}\left(c^{(j)}\right)$$

Τα στοιχεία εκτός της διαγωνίου (WW^T) είναι η συν-διασπορά των διανυσμάτων $w^{(i)}$, όπως το στοιχείο $(WW^T)_{ab}$ είναι η συν-διασπορά των στοιχείων με συντεταγμένες a και b του διανύσματος $w^{(i)}$. Η εξίσωση (5.11) δείχνει πως αυτό είναι επίσης το γινόμενο των $c^{(a)} c^{(b)}$. Ένας από τους βασικότερους στόχους στον μετασχηματισμό μιας εικόνας είναι η αποσυσχέτιση των συντεταγμένων των διανυσμάτων και η θεωρία πληροφορίας λέει ότι δύο συντεταγμένες είναι ασυσχέτιστες εάν η συν-διασπορά τους είναι μηδέν (ο άλλος στόχος είναι η συγκέντρωση της ενέργειας αλλά οι δύο αυτοί στόχοι επιτυγχάνονται μαζί). Συνεπώς, στόχος είναι η εύρεση ενός πίνακα A τέτοιου ώστε το γινόμενο WW^T να είναι διαγώνιο. Από τον ορισμό το πίνακα Wέχουμε οτι:

$$WW^T = (AV)(AV)^T = A(VV^T)A^T$$

Ο πίνακας VV^T είναι συμμετρικός και τα στοιχεία του είναι οι τιμές συν-διασποράς των διανυσμάτων $v^{(i)}$.

$$(VV^T)_{ab} = \sum_{i=1}^{k} v_a^{(i)} v_b^{(i)}, \gamma\iota\alpha\ a, b \in [1, n]$$

Δεδομένου ότι ο (VV^T) είναι συμμετρικός, τα ιδιοδιανύσματα του είναι ορθογώνια. Κανονικοποιώντας τις τιμές των διανυσμάτων τα μετατρέπουμε σε ορθοκανονικά και τα θέτουμε ως γραμμές του πίνακα A. Αυτό έχει το ακόλουθο αποτέλεσμα:

$$WW^T = A(VV^T)A^T = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

Η επιλογή αυτού του πίνακα A έχει ως αποτέλεσμα τον διαγώνιο πίνακα WW^T του οποίου τα διαγώνια στοιχεία είναι τα ιδιοδιανύσματα του VV^T. Ο πίνακας A είναι ο πίνακας μετασχηματισμού Karhunen-Loeve του οποίου οι γραμμές είναι τα διανύσματα βάσης του KLT και οι ενέργειες (διασπορά) των μετασχηματισμένων διανυσμάτων είναι οι ιδιοτιμές $\lambda_1, \lambda_2, \dots, \lambda_n$ του VV^T.

Τα διανύσματα βάσης του KLT υπολογίζονται από τις αρχικές τιμές των εικονοστοιχείων της εικόνας και συνεπώς εξαρτώνται από τα δεδομένα. Σε μία πρακτική μέθοδο συμπίεσης, τα διανύσματα αυτά θα πρέπει να εμπεριέχονται στη συμπιεσμένη ροή δεδομένων προκειμένου να είναι σε θέση ο αποκωδικοποιητής να εκτελέσει τον αντίστροφο μετασχηματισμό κατά τη διαδικασία αποσυμπίεσης. Το γεγονός αυτό, μαζί με ότι δεν έχει βρεθεί ακόμα γρήγορος αλγόριθμος για τον υπολογισμό του KLT καθιστούν τον μετασχηματισμό αυτό λιγότερο κατάλληλο για πρακτικές εφαρμογές.

5.4 Ο Διακριτός Μετασχηματισμός Συνημίτονου

Ο σημαντικός αυτός μετασχηματισμός (DCT) προτάθηκε από τον Ahmed[35] το 1974 και έκτοτε μελετήθηκε και χρησιμοποιήθηκε εκτεταμένα. Εξαιτίας της σημαντικότητας του στη συμπίεση δεδομένων ο μετασχηματισμός DCT θα μελετηθεί λεπτομερώς. Αρχικά θα εξεταστούν οι μαθηματικές εκφράσεις για τον DCT σε μία διάσταση και στη συνέχεια στις δύο διαστάσεις χωρίς όμως να παρουσιαστεί το θεωρητικό υπόβαθρό και πολλές αναλύσεις. Η χρήση του μετασχηματισμού και τα πλεονεκτήματα του θα παρουσιαστούν παρακάτω μέσα από παραδείγματα. Στη συνέχεια θα μελετηθεί η

θεωρία του DCT και θα συζητηθούν οι δυο εκδοχές του ως περιστροφή και ως βάση ενός διανυσματικού χώρου. Τέλος θα παρουσιαστούν τρόποι επιτάχυνσης της διαδικασίας τόσο με χρήση αλγορίθμων όσο και υλοποίησης με χρήση υλικού καθώς και άλλες σημαντικές πληροφορίες.

Ο μετασχηματισμός DCT στη μια διάσταση δίνεται από την εξίσωση:

$$G_f = \sqrt{\frac{2}{n}}\, C_f \sum_{t=0}^{n-1} P_t \cos\left[\frac{(2t+1)f\pi}{2n}\right] \tag{5.12}$$

όπου

$$C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0, \\ 1\,, & f > 0, \end{cases} \quad \gamma\iota\alpha\ f = 0,1,\dots,n-1$$

Η είσοδος είναι ένα σύνολο από n τιμές δεδομένων P_t (εικονοστοιχεία, ηχητικά δείγματα ή άλλου τύπου δείγματα και η έξοδος είναι ένα σύνολο από n μετασχηματισμένους DCT συντελεστές (ή βάρη) G_f. Ο πρώτος συντελεστής G_0 ονομάζεται ο DC συντελεστής και οι υπόλοιποι ονομάζονται συντελεστές AC (όροι οι οποίοι είναι δανεισμένοι από την επιστήμη της ηλεκτρολογίας και αφορούν το συνεχές και το εναλλασσόμενο ρεύμα (Direct Current, Alternating Current). Σημειώνεται ότι οι συντελεστές είναι πραγματικοί αριθμοί παρά το γεγονός ότι στην είσοδο έχουμε ακεραίους. Ομοίως, οι συντελεστές μπορεί να είναι θετικοί ή αρνητικοί αριθμοί ενώ στην είσοδο έχουμε μόνο θετικούς αριθμούς. Ο υπολογισμός είναι απλός αλλά αργός (όμως υπάρχουν και γρήγορες μέθοδοι υπολογισμού του). Ο αποκωδικοποιητής λαμβάνει ως είσοδο του συντελεστές DCT σε σύνολα των n στοιχείων και χρησιμοποιεί τον αντίστροφο μετασχηματισμό DCT (IDCT) προκειμένου να ανακατασκευάσει τις αρχικές τιμές δεδομένων (επίσης σε σύνολα των n στοιχείων). Ο αντίστροφος μετασχηματισμός DCT στη μία διάσταση δίνεται από τη σχέση:

$$P_t = \sqrt{\frac{2}{n}} \sum_{j=0}^{n-1} C_j G_j \cos\left[\frac{(2t+1)j\pi}{2n}\right], \gamma\iota\alpha\ t = 0,1,\dots,n-1 \tag{5.13}$$

Το σημαντικό χαρακτηριστικό του μετασχηματισμού DCT το οποίο και κάνει ιδανική τη χρήση του στη συμπίεση δεδομένων είναι το γεγονός ότι δέχεται ως είσοδο συσχετισμένα δεδομένα και συγκεντρώνει την ενέργεια τους σε λίγους συντελεστές. Εάν τα δεδομένα εισόδου είναι ισχυρά συσχετισμένες ποσότητες, τότε οι περισσότεροι

από τους n μετασχηματισμένους συντελεστές θα είναι μηδέν ή μικροί αριθμοί και μόνο λίγοι θα έχουν μεγάλες τιμές (συνήθως οι πρώτοι). Όπως θα δούμε και στη συνέχεια, οι πρώτοι συντελεστές περιέχουν τη σημαντικότερη πληροφορία της εικόνας (χαμηλές συχνότητες) και οι τελευταίοι τη λιγότερη σημαντική (υψηλές συχνότητες). Η συμπίεση των δεδομένων με χρήση του DCT επιτυγχάνεται με την κβάντιση των συντελεστών. Οι συντελεστές με μικρές τιμές κβαντίζονται ισχυρά (πιθανότατα καταλήγουν σε μηδενική τιμή) ενώ οι μεγάλοι συντελεστές κβαντίζονται στρογγυλοποιώντας το αποτέλεσμα στον πλησιέστερο ακέραιο. Μετά την κβάντιση, οι συντελεστές (ή οι κωδικοί μεταβλητού μήκους οι οποίοι τους ανατίθενται) εγγράφονται σε στη συμπιεσμένη ακολουθία δεδομένων. Η αποσυμπίεση ολοκληρώνεται με την εφαρμογή του αντίστροφου μετασχηματισμού DCT στους κβαντισμένους συντελεστές. Τα αποτελέσματα δεν είναι πανομοιότυπα με τα αρχικά αλλά δε διαφέρουν και ιδιαίτερα.

Σε πρακτικές εφαρμογές τα δεδομένα που πρόκειται να συμπιεστούν διαμερίζονται σε σύνολα των n στοιχείων και κάθε σύνολο μετασχηματίζεται και κβαντίζεται ξεχωριστά. Η επιλογή της τιμής του n είναι κρίσιμη. Μικρές τιμές του n όπως 3, 4 ή 6 έχουν ως αποτέλεσμα τη δημιουργία μεγάλου πλήθους συνόλων. Αυτό συνεπάγεται ότι ένα μικρό σύνολο θα μετασχηματιστεί σε ένα μικρό πλήθος συντελεστών όπου η ενέργεια θα συγκεντρωθεί σε λίγους μόνο συντελεστές, όμως σε αυτή την περίπτωση υπάρχουν ήδη λίγοι συντελεστές λόγω της διαμέρισης που επιλέχθηκε βάσει του n. Συνεπώς δεν θα υπάρχουν αρκετοί μικροί συντελεστές για να κβαντιστούν. Μεγάλες τιμές του n έχουν ως αποτέλεσμα τη δημιουργία λίγων συνόλων με πολλά στοιχεία το καθένα. Το πρόβλημα σε αυτή την περίπτωση είναι ότι συνήθως τα δεδομένα σε ένα μεγάλο σύνολο δεν είναι συσχετισμένα και αυτό θα προκαλέσει τη δημιουργία συνόλων τα οποία θα έχουν μεγάλο μέρος των συντελεστών με μεγάλες τιμές. Εμπειρικά έχει προκύψει ότι $n = 8$ είναι μια καλή τιμή και οι περισσότερες μέθοδοι συμπίεσης με DCT χρησιμοποιούν αυτή την τιμή του n.

Το παράδειγμα που ακολουθεί θα παρουσιάσει τη δύναμη του μετασχηματισμού DCT στη μία διάσταση. Τα δεδομένα εισόδου είναι συσχετισμένα με τιμές $p = (12, 10, 8, 10, 12, 10, 8, 11)$. Εφαρμόζοντας τον μονοδιάστατο μετασχηματισμό DCT βρίσκουμε τους ακόλουθους οκτώ συντελεστές:

$$28.6375, 0.571202, 0.46194, 1.757, 3.18198, -1.72956, 0.191342, -0.308709$$

Οι τιμές αυτές μπορούν να δοθούν στον αντίστροφο DCT και να μετασχηματιστούν με μεγάλη ακρίβεια στις αρχικές τιμές (εκτός από μικρά σφάλματα τα οποία να προκύψουν

ανάλογα με την ακρίβεια της εκάστοτε μηχανής στην οποία εκτελείται ο μετασχηματισμός). Στόχος όμως είναι η συμπίεση των δεδομένων κβαντίζοντας τους συντελεστές. Αρχικά κβαντίζουμε τις τιμές σε:

$$28.6, 0.6, 0.5, 1.8, 3.2, -1.8, 0.2, -0.3$$

και εφαρμόζουμε τον αντίστροφο DCT για να ανακατασκευάσουμε τις τιμές εισόδου οι οποίες είναι:

$$12.0254, 10.0233, 7.96054, 9.93097, 12.0164, 9.99321, 7.94354, 10.9989$$

Στη συνέχεια κβαντίζουμε τους συντελεστές ακόμα περισσότερο σε:

$$28, 1, 1, 2, 3, -1, 0, 0$$

και με τον αντίστροφο DCT έχουμε:

$$12.1883, 10.2315, 7.74931, 9.29863, 11.7876, 9.54549, 7.82865, 10.6557$$

Τέλος κβαντίζουμε τους συντελεστές σε:

$$28, 0, 0, 2, 3, -2, 0, 0$$

και με τον αντίστροφο ανακτάμε την ακολουθία:

$$11.236, 9.62443, 7.66286, 9.57302, 12.3471, 10.0146, 8.05304, 10.6842$$

όπου η μεγαλύτερη διαφορά ανάμεσα στην αρχική τιμή (12) και την ανακατασκευασμένη (11.236) είναι 0.764 (ή 6.4% του 12).

Φαίνεται εκπληκτικό πώς τα οκτώ αρχικά στοιχεία έχουν ανακατασκευαστεί με τόσο μεγάλη ακρίβεια από μόλις τέσσερεις συντελεστές. Η εξήγηση βρίσκεται στο γεγονός ότι ο αντίστροφος DCT δέχεται ως είσοδο και τους οκτώ συντελεστές, επομένως γνωρίζει και τις θέσεις των μη μηδενικών συντελεστών και όχι μόνο τις τιμές τους. Επίσης οι πρώτοι μόνο συντελεστές (οι οποίοι έχουν μεγάλες τιμές) περιέχουν τη σημαντική πληροφορία των αρχικών δεδομένων. Οι συντελεστές με μικρές τιμές οι οποίοι και κβαντίζονται ισχυρά, περιέχουν τη λιγότερο σημαντική πληροφορία (στην περίπτωση της εικόνας, περιέχουν μόνο τις λεπτομέρειες της). Βεβαίως, σημαντική παράμετρο αποτελεί η συσχέτιση που υπάρχει στα αρχικά δείγματα εισόδου.

Το παράδειγμα που ακολουθεί παρουσιάζει την απόδοση του DCT όταν αυτός εφαρμοσθεί σε δεδομένα τα οποία είναι ασυσχέτιστα. Έστω ότι δίνονται τα ασυσχέτιστα δεδομένα:

$$-12, 24, -181, 209, 57.8, 3, -184, -250$$

Ο μετασχηματισμός DCT παράγει τους ακόλουθους συντελεστές:

$$-117.803, 166.823, -240.83, 126.887, 121.198, 9.02198, -109.496, -185.206$$

Όταν οι συντελεστές αυτοί κβαντιστούν σε:

$$-120, 170, -240, 125, 120, 9, -110, -185$$

Το αποτέλεσμα του IDCT θα είναι:

$$-12.1249, 25.4974, -179.852, 208.237, 55.5898, 0.364874, -185.42, -251.701$$

Όπου η μέγιστη διαφορά ανάμεσα στο 3 και το 0.364874 είναι 2.63513, δηλαδή το 88% του 3. Προφανώς ακόμα και με αυτή την καθόλου ισχυρή κβάντιση, η ανακατασκευή δεν είναι τόσο καλή σε σχέση με αυτή που επετεύχθη όταν τα δείγματα εισόδου ήταν συσχετισμένα.

Ένας μετασχηματισμός σχετικός με τον DCT είναι ο μετασχηματισμός Fourier ο οποίος έχει και διακριτή έκδοση, τον διακριτό μετασχηματισμό Fourier (DFT). Ο DFT έχει και αυτός σημαντικές εφαρμογές όμως δεν αποδίδει ικανοποιητικά στη συμπίεση δεδομένων καθότι θεωρεί ότι τα δείγματα προς μετασχηματισμό είναι περιοδικά.

Το ακόλουθο παράδειγμα παρουσιάζει τη διαφορά στην απόδοση ανάμεσα στον DCT και τον DFT. Υποθέτουμε την ισχυρά συσχετισμένη ακολουθία με τα οκτώ δείγματα $(8, 16, 24, 32, 40, 48, 56, 64)$. Την γραφική της αναπαράσταση παρατηρούμε και στο Σχήμα 5.10 a. Εφαρμόζοντας τον DCT στην ακολουθία αυτή λαμβάνουμε τους συντελεστές $(100, -52, 0, -5, 0, -2, 0, 0.4)$. Όταν αυτοί κβαντιστούν σε $(100, -52, 0, -5, 0, 0, 0, 0)$ και ανασυσταθούν με τον αντίστροφο DCT έχουμε την ανακατασκευασμένη ακολουθία $(8, 15, 24, 32, 40, 48, 57, 63)$ η οποία είναι σχεδόν πανομοιότυπη με την αρχική. Εάν εφαρμόσουμε τον DFT στα ίδια δεδομένα, οι συντελεστές που θα προκύψουν θα είναι $(36, 10, 10, 6, 6, 4, 4, 4)$. Εάν κβαντίσουμε τους συντελεστές σε $(36, 10, 10, 6, 0, 0, 0, 0)$ και επιχειρήσουμε να κάνουμε ανασύσταση της εισόδου, ο αντίστροφος μετασχηματισμός DFT θα μας δώσει $(24, 12, 20, 32, 40, 51, 59, 48)$. Η έξοδος εμφανίζεται στο Σχήμα 5.10 b και παρουσιάζει την τάση του μετασχηματισμού Fourier να παράγει περιοδικό αποτέλεσμα.

(a)

(b)

Σχήμα 5.10 : (a) Μονοδιάστατα Δεδομένα Εισόδου και (b) ο Αντίστροφος Μετασχηματισμός Fourier μετά την Κβάντιση

Ο μονοδιάστατος μετασχηματισμός DCT μπορεί να χρησιμοποιηθεί για τη συμπίεση μονοδιάστατων δεδομένων όπως δείγματα ηχητικού σήματος. Η παρούσα εργασία ωστόσο μελετά τη συμπίεση εικόνας η οποία είναι βασισμένη στη συσχέτιση των εικονοστοιχείων και στις δύο διαστάσεις (ένα εικονοστοιχείο είναι συνήθως συσχετισμένο με τους γείτονες του και προς τις δύο διαστάσεις και όχι μόνο κατά γραμμές). Για το λόγο αυτό πρακτικά στη συμπίεση εικόνας χρησιμοποιείται ο δισδιάστατος μετασχηματισμός DCT[36]. Η μέθοδος αυτή του DCT εφαρμόζεται σε μικρά μπλοκ δειγμάτων της εικόνας. Υπολογίζεται εφαρμόζοντας τον DCT στη μία διάσταση κατά γραμμές και στη συνέχεια κατά στήλες στο ήδη υπάρχων αποτέλεσμα. Εξαιτίας του χωρικού τρόπου με τον οποίο υπολογίζεται ο DCT στις δύο διαστάσεις λέμε ότι ο μετασχηματισμός είναι διαχωρίσιμος στις δύο διαστάσεις. Επειδή εφαρμόζεται σε μπλοκ της εικόνας, ονομάζεται μπλοκ μετασχηματισμός και ορίζεται ως:

$$G_{ij} = \frac{2}{\sqrt{mn}} C_i C_j \sum_{x=0}^{n-1} \sum_{y=0}^{m-1} p_{xy} \cos\left[\frac{(2y+1)j\pi}{2m}\right] \cos\left[\frac{(2x+1)i\pi}{2n}\right] \qquad (5.14)$$

για $0 \leq i \leq n-1$ και $0 \leq j \leq m-1$ και με C_i, C_j όπως ορίστηκε στην εξίσωση (5.12)

$$C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0, \\ 1\,, & f > 0, \end{cases} \quad \gamma\iota\alpha\ f = 0,1,\dots,n-1$$

Ο πρώτος συντελεστής G_{00} ονομάζεται, όπως και στην περίπτωση του μονοδιάστατου μετασχηματισμού, συντελεστής DC ενώ οι υπόλοιποι, συντελεστές AC.

Η εικόνα χωρίζεται σε μπλοκ μεγέθους $n \times m$ εικονοστοιχείων p_{xy} (συνήθως με $m = n = 8$), και η εξίσωση (5.14) χρησιμοποιείται προκειμένου να παράγει το μπλοκ $n \times m$ συντελεστών DCT G_{ij} για κάθε μπλοκ εικονοστοιχείων. Στη συνέχεια οι συντελεστές κβαντίζονται, κάτι που συνεπάγεται συμπίεση με απώλειες αλλά πολύ αποδοτική. Ο αποκωδικοποιητής ανακατασκευάζει το μπλοκ των κβαντισμένων

δεδομένων με τον αντίστροφο δισδιάστατο DCT, ο ορισμός του οποίου είναι:

$$p_{xy} = \frac{2}{\sqrt{mn}} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} C_i C_j G_{ij} \cos\left[\frac{(2x+1)i\pi}{2n}\right] \cos\left[\frac{(2y+1)j\pi}{2m}\right] \tag{5.15}$$

όπου

$$C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0 \\ 1\,, & f > 0, \end{cases}$$

Για $0 \leq x \leq n-1$ και $0 \leq y \leq m-1$. Σε αυτό το σημείο παρουσιάζεται μια μέθοδος για τη συμπίεση μίας ολόκληρης εικόνας με τον μετασχηματισμό DCT με τη χρήση των ακόλουθων βημάτων:

1. Η εικόνα διαιρείται σε k μπλοκ μεγέθους 8×8 εικονοστοιχείων το καθένα. Τα εικονοστοιχεία συμβολίζονται με p_{xy}. Εάν το πλήθος των γραμμών ή στηλών της εικόνας δεν διαιρείται ακριβώς με το 8 τότε η τελευταία γραμμή ή στήλη αντιγράφεται όσες φορές χρειαστεί.
2. Η εξίσωση του δισδιάστατου μετασχηματισμού DCT εφαρμόζεται σε κάθε μπλοκ B_i. Το αποτέλεσμα είναι ένα μπλοκ, το οποίο θα ονομάζουμε διάνυσμα $W^{(i)}$ και το οποίο αποτελείται από 64 μετασχηματισμένους συντελεστές $w_j^{(i)}$ (όπου $j = 0,1, \dots, 63$). Τα k διανύσματα $W^{(i)}$ γίνονται οι γραμμές του πίνακα W.

$$W = \begin{bmatrix} w_0^{(1)} & w_0^{(1)} & \cdots & w_{63}^{(1)} \\ w_0^{(2)} & w_1^{(2)} & \cdots & w_{63}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ w_0^{(k)} & w_1^{(k)} & \cdots & w_{63}^{(k)} \end{bmatrix}$$

3. Οι 64 στήλες του W συμβολίζονται ως $C^{(0)}, C^{(1)}, \dots, C^{(63)}$. Τα k στοιχεία του $C^{(j)}$ είναι τα $(w_j^{(1)}, w_j^{(2)}, \dots, w_j^{(k)})$. Το πρώτο διάνυσμα συντελεστών $C^{(0)}$ αποτελείται από τους k συντελεστές DC.
4. Κάθε διάνυσμα $C^{(j)}$ κβαντίζεται ξεχωριστά για να δημιουργηθεί το διάνυσμα $Q^{(j)}$ κβαντισμένων συντελεστών (το JPEG διαχειρίζεται τη διαδικασία αυτή λίγο διαφορετικά αλλά θα το εξετάσουμε λεπτομερέστερα στη συνέχεια). Τα στοιχεία του $Q^{(j)}$ εγγράφονται στη συνέχεια στη συμπιεσμένη ακολουθία δεδομένων. Στην πράξη, κωδικοί μεταβλητού μήκους ανατίθενται στα στοιχεία αυτά και όχι τα ίδια τα στοιχεία συνεπώς οι κωδικοί μεταβλητού μήκους εγγράφονται τελικά στην ακολουθία δεδομένων. Σε κάποιες περιπτώσεις όπως συμβαίνει και στην περίπτωση του JPEG, κώδικες μεταβλητού μήκους ανατίθενται στο πλήθος μηδενικών συντελεστών προκειμένου να αυξήσουν την αποδοτικότητα συμπίεσης.

Στην πράξη ο μετασχηματισμός DCT χρησιμοποιείται για συμπίεση με απώλειες. Για συμπίεση χωρίς απώλειες (όπου οι συντελεστές δεν κβαντίζονται) ο μετασχηματισμός DCT είναι αναποτελεσματικός αλλά παρόλα αυτά μπορεί ακόμα να χρησιμοποιηθεί, τουλάχιστον θεωρητικά αφού (1) οι περισσότεροι από τους συντελεστές είναι μικροί αριθμοί και (2) υπάρχουν συχνά ακολουθίες με διαδοχικά μηδενικά. Παρόλα αυτά, οι μικροί συντελεστές είναι πραγματικοί αριθμοί και όχι ακέραιοι, συνεπώς δεν είναι ξεκάθαρο το πώς θα μπορούσαν να γραφούν με πλήρη αριθμητική ακρίβεια στη συμπιεσμένη ακολουθία δεδομένων και να μπορεί να επιτευχθεί ακόμα κάποια συμπίεση. Άλλες μέθοδοι συμπίεσης ταιριάζουν καλύτερα στη συμπίεση εικόνας. Ο αποκωδικοποιητής λαμβάνει τα 64 κβαντισμένα διανύσματα συντελεστών $Q^{(j)}$ με k στοιχεία το καθένα, τα αποθηκεύει ως στήλες σε έναν πίνακα και θεωρεί τις k γραμμές του πίνακα ως διανύσματα βαρών $W^{(i)}$ των 64 στοιχείων το καθένα. Να σημειωθεί ότι αυτά τα διανύσματα $W^{(i)}$ δεν είναι απολύτως όμοια με τα αρχικά $W^{(i)}$ εξαιτίας της κβάντισης. Στη συνέχεια εφαρμόζεται ο αντίστροφος μετασχηματισμός DCT σε κάθε διάνυσμα βάρους προκειμένου να ανακατασκευαστούν (προσεγγιστικά) οι τιμές των 64 εικονοστοιχείων του μπλοκ B_i. Και σε αυτή την περίπτωση το JPEG διαχειρίζεται διαφορετικά τη διαδικασία.

Θα παρουσιαστεί η επίδοση του DCT στις δύο διαστάσεις εφαρμόζοντας τη μέθοδο σε δύο μπλοκ τιμών 8×8. Το πρώτο μπλοκ περιλαμβάνει τιμές ισχυρά συσχετισμένες στο εύρος $[8, 12]$ (Πίνακας 5.1 a) και το δεύτερο μπλοκ έχει τυχαίες τιμές στο ίδιο πάλι εύρος (Πίνακας 5.1 b).

Πίνακας 5.1 : Τα Αρχικά Δεδομένα του Πρώτου (a) και του Δεύτερου (b) Μπλοκ

12	10	8	10	12	10	8	11
11	12	10	8	10	12	10	8
8	11	12	10	8	10	12	10
10	8	11	12	10	8	10	12
12	10	8	11	12	10	8	10
10	12	10	8	11	12	10	8
8	10	12	10	8	11	12	10
10	8	10	12	10	8	11	12

(a)

8	10	9	11	11	9	9	12
11	8	12	8	11	10	11	10
9	11	9	10	12	9	9	8
9	12	10	8	8	9	8	9
12	8	9	9	12	10	8	11
8	11	10	12	9	12	12	10
10	10	12	10	12	10	10	12
12	9	11	11	9	8	8	12

(b)

Πίνακας 5.2 : Οι DCT Συντελεστές του Πρώτου (a) και του Δεύτερου (b) Μπλοκ

81	0	0	0	0	0	0	0
0	1.57	0.61	1.90	0.38	-1.81	0.20	-0.32
0	-0.61	0.71	0.35	0	0.07	0	0.02
0	1.90	-0.35	4.76	0.77	-3.39	0.25	-0.54
0	-0.38	0	-0.77	8.00	0.51	0	0.07
0	-1.81	-0.07	-3.39	-0.51	1.5	0.56	0. 5
0	-0.20	0	-0.25	0	0.56	-0.71	0.29
0	-0.32	-0.02	-0.54	-0.07	0.25	-0.29	-0.90

(a)

79.12	0.98	0.64	-1.51	-0.62	-0.86	1.22	0.32
0.15	-1.64	-0.09	1.23	0.10	3.29	1.08	-2.97
-1.26	-0.29	-3.27	1.69	-0.51	1.13	1.52	1.33
-1.27	-0.25	-0.67	-0.15	1.63	-1.94	0.47	-1.30
-2.12	-0.67	-0.07	-0.79	0.13	-1.40	0.16	-0.15
-2.68	1.08	-1.99	-1.93	-1.77	-0.35	0	-0.80
1.20	2.10	-0.98	0.87	-1.55	-0.59	-0.98	2.76
-2.24	0.55	0.29	0.75	-2.40	-0.05	0.06	1.14

(b)

Το πρώτο μπλοκ έχει ως αποτέλεσμα έναν DC συντελεστή με μεγάλη τιμή και μικρούς συντελεστές AC. Όταν οι συντελεστές κβαντίζονται το αποτέλεσμα παρουσιάζει ο Πίνακας 5.3. Όπως παρατηρούμε το αποτέλεσμα είναι παραπλήσιο με το αρχικό. Αντίθετα οι συντελεστές του δεύτερου μπλοκ έχουν μόλις ένα μηδενικό. Αφού κβαντιστούν και ανακατασκευαστούν παρατηρούμε πως πολλοί από αυτούς είναι αρκετά διαφορετικοί σε σχέση με τους αρχικούς.

Πίνακας 5.3 : Οι Κβαντισμένοι Συντελεστές του Πρώτου (a) και του Δεύτερου (b) Μπλοκ

81	0	0	0	0	0	0	0
0	2	1	2	0	-2	0	0
0	-1	1	0	0	0	0	0
0	2	0	5	1	-3	0	-1
0	0	0	-1	8	1	0	0
0	-2	0	-3	-1	2	1	0
0	0	0	0	0	-1	-1	0
0	0	0	-1	0	0	0	-1

(a)

79	1	1	-2	-1	-1	1	0
0	-2	0	1	0	3	1	-3
-1	0	-3	2	0	1	2	1
-1	0	-1	0	2	-2	0	-10
-20	-1	0	-1	0	-10	0	0
-3	1	-2	-2	-2	0	0	-1
1	2	-1	1	-2	-1	-1	3
-2	1	0	1	-2	0	0	1

(b)

Εύκολα μπορεί να παρατηρηθεί ο λόγος για τον οποίο οι 64 τιμές ο περιέχει ο Πίνακας 5.1 a είναι συσχετισμένοι. Οι 8 τιμές στην πρώτη γραμμή είναι παραπλήσιοι (οι διαφορές ανάμεσα τους είναι 2 ή 3). Κάθε άλλη γραμμή προκύπτει με μια αριστερή ολίσθηση της προηγούμενης γραμμής.

Πίνακας 5.4 : Τα ανακατασκευασμένα Δείγματα του Πρώτου (a) και του Δεύτερου (b) Μπλοκ

12.29	10.26	7.92	9.93	11.51	9.94	8.81	10.97
10.90	12.06	10.07	7.68	10.30	11.64	10.17	8.18
7.83	11.39	12.19	9.62	8.28	10.10	11.64	9.94
10.15	7.74	11.16	11.96	9.90	8.28	10.30	11.51
12.21	10.08	8.15	11.38	11.96	9.62	7.68	9.93
10.09	12.10	9.30	8.15	11.16	12.19	10.07	7.92
7.87	9.50	12.10	10.08	7.74	11.39	12.06	10.26
9.66	7.87	10.09	12.21	10.15	7.83	10.90	12.29

(a)

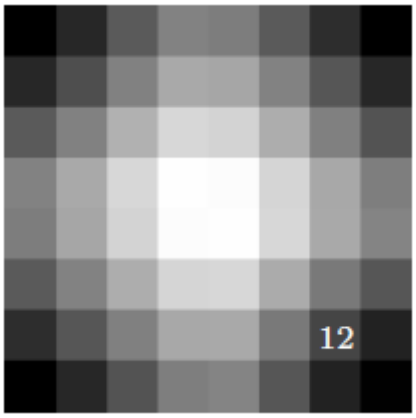
7.59	9.23	8.33	11.88	7.12	12.47	6.98	8.56
12.09	7.97	9.3	11.52	9.28	11.62	10.98	12.39
11.02	10.06	13.81	6.5	10.82	8.28	13.02	7.54
8.46	10.22	11.16	9.57	8.45	7.77	10.28	11.89
9.71	11.93	8.04	9.59	8.04	9.7	8.59	12.14
10.27	13.58	9.21	11.83	9.99	10.66	7.84	11.27
8.34	10.32	10.53	9.9	8.31	9.34	7.47	8.93
10.61	9.04	13.66	6.04	13.47	7.65	10.97	8.89

(b)

Το επόμενο παράδειγμα θα αναδείξει τη διαφορά στις επιδόσεις του μετασχηματισμού DCT ανάμεσα σε μια εικόνα συνεχούς τόνου και μίας εικόνας διακριτού τόνου. Θα εξετασθεί πρώτα το ισχυρά συσχετισμένο πρότυπο που παρουσιάζει ο Πίνακας 5.5.

Πίνακας 5.5 : Εικόνα Συνεχούς Τόνου Μεγέθους 8×8

0	10	20	30	30	20	10	0
10	20	30	40	40	30	20	10
20	30	40	50	50	40	30	20
30	40	50	60	60	50	40	30
30	40	50	60	60	50	40	3
20	30	40	50	50	40	30	20
10	20	30	40	40	30	12	10
0	10	20	30	30	20	10	0



Πρόκειται για ένα εξιδανικευμένο παράδειγμα μιας εικόνας συνεχούς τόνου δεδομένου ότι τα γειτονικά εικονοστοιχεία διαφέρουν μεταξύ τους κατά ένα σταθερό μέγεθος εκτός

από το υπογραμμισμένο εικονοστοιχείο στη θέση (7, 7). Τους 64 συντελεστές DCT αυτού του προτύπου εμφανίζει ο Πίνακας 5.6.

Πίνακας 5.6 : Οι DCT Συντελεστές της Εικόνας του Πίνακας 5.5

239	1.19	-89.76	-0.28	1.00	-1.39	-5.03	-0.79
1.18	-1.39	0.64	0.32	-1.18	1.63	-1.54	0.92
-89.76	0.64	-0.29	-0.15	0.54	-0.75	0.71	-0.43
-0.28	0.32	-0.15	-0.08	0.28	-0.38	0.36	-0.22
1.00	-1.18	0.54	0.28	-1.00	1.39	-1.31	0.79
-1.39	1.63	-0.75	-0.38	1.39	-1.92	1.81	-1.09
-5.03	-1.54	0.71	0.36	-1.31	1.81	-1.71	1.03
-0.79	0.92	-0.43	-0.22	0.79	-1.09	1.03	-0.62

Είναι σαφές πως υπάρχουν ελάχιστοι συντελεστές οι οποίοι υπερισχύουν. Ο Πίνακας 5.7 παρουσιάζει τους συντελεστές μετά από ισχυρή κβάντιση και όπως παρατηρούμε μόνο τέσσερεις μη μηδενικοί συντελεστές έχουν προκύψει.

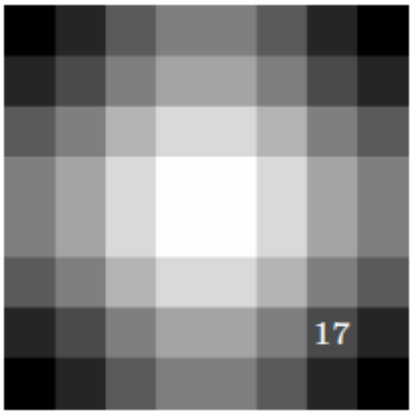
Πίνακας 5.7 : Οι Κβαντισμένοι DCT

239	1	-90	0	0	0	0	0
0	0	0	0	0	0	0	0
-90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Το αποτέλεσμα που προκύπτει από την εφαρμογή του αντίστροφου μετασχηματισμού στις κβαντισμένες τιμές παρουσιάζει ο Πίνακας 5.8. Είναι προφανές πως οι τέσσερεις συντελεστές ανακατασκεύασαν το αρχικό πρότυπο σε μεγάλο βαθμό. Η μόνη εμφανής διαφορά είναι στο εικονοστοιχείο (7, 7) το οποίο άλλαξε τιμή από 12 σε 17.55.

Πίνακας 5.8 : Ανακατασκευασμένη Εικόνα με τον Αντίστροφο DCT στους Κβαντισμένους Συντελεστές του Πίνακας 5.7

0.65	9.23	21.36	29.91	29.84	21.17	8.94	0.30
9.26	17.85	29.97	38.52	38.45	29.78	17.55	9.91
21.44	30.02	42.15	50.70	50.63	41.95	29.73	21.09
30.05	38.63	50.76	59.31	59.24	50.56	38.34	29.70
30.05	38.63	50.76	59.31	59.24	50.56	38.34	29.70
21.44	30.02	42.15	50.70	50.63	41.95	29.73	21.09
9.26	17.85	29.97	38.52	38.45	29.78	17.55	8.91
0.65	9.23	21.36	29.91	29.84	21.17	8.94	0.30



Η ίδια διαδικασία εφαρμόστηκε και στο πρότυπο με το σχήμα Υ το οποίο αποτελεί τυπικό παράδειγμα εικόνας διακριτού τόνου (Πίνακας 5.9).

Πίνακας 5.9 : Εικόνα Διακριτού Τόνου Μεγέθους 8×8

0	10	0	0	0	0	0	10
0	0	10	0	0	0	10	0
0	0	0	10	0	10	0	0
0	0	0	0	10	0	0	0
0	0	0	0	10	0	0	0
0	0	0	0	10	0	0	0
0	0	0	0	10	0	0	0
0	0	0	0	10	0	0	0

Η κβάντιση που παρουσιάζει ο Πίνακας 5.11 είναι ελαφριά. Οι συντελεστές απλώς έχουν στρογγυλοποιηθεί στον πλησιέστερο ακέραιο.

Πίνακας 5.10 : Οι DCT Συντελεστές της Εικόνας που Παρουσιάζει ο Πίνακας 5.9

13.75	-3.11	-8.17	2.46	3.75	-6.86	-3.38	6.59
4.19	-0.29	6.86	-6.85	-7.13	4.48	1.69	-7.28
1.63	0.19	6.40	-4.81	-2.99	-1.11	-0.88	-0.94
-0.61	0.54	5.12	-2.31	1.30	-6.04	-2.78	3.05
-1.25	0.52	2.99	-0.20	3.75	-7.39	-2.59	1.16
-0.41	0.18	0.65	1.03	3.87	-5.19	-0.71	-4.76
0.68	-0.15	-0.88	1.28	2.59	-1.92	1.10	-9.05
0.83	-0.21	-0.99	0.82	1.13	-0.08	1.31	-7.21

Εύκολα συμπεραίνουμε παρατηρώντας τις τιμές που περιέχει ο Πίνακας 5.12 ότι η ανακατασκευή δεν είναι τόσο καλή όσο στην προηγούμενη περίπτωση. Ποσότητες οι οποίες θα έπρεπε να είχαν τιμή 10 βρίσκονται ανάμεσα στο 8.96 και το 10.11. Ποσότητες που θα έπρεπε να είχαν τιμή μηδέν έχουν τιμές μέχρι 0.86. Συνεπώς ο μετασχηματισμός DCT αποδίδει πολύ καλά σε εικόνες συνεχούς τόνου αλλά πολύ λιγότερο όταν εφαρμοστεί σε εικόνα διακριτού τόνου.

Πίνακας 5.11 : Οι Κβαντισμένοι DCT Συντελεστές της Εικόνας που Παρουσιάζει ο Πίνακας 5.9

13.75	-3	-8	2	3	-6	-3	6
4	0	6	-6	-7	4	1	-7
1	0	6	-4	-2	-1	0	0
0	0	5	-2	1	-6	-2	3
-1	0	2	0	3	-7	-2	1
0	0	0	1	3	-5	0	-4
0	0	0	1	2	-1	1	-9
0	0	0	0	1	0	1	-7

Πίνακας 5.12 : Ανακατασκευασμένη Εικόνα με τον Αντίστροφο DCT στους Κβαντισμένους Συντελεστές του Πίνακας 5.11

-0.13	8.96	0.55	-0.27	0.27	0.86	0.15	9.22
0.32	0.22	9.10	0.40	0.84	-0.11	9.36	-0.14
0	0.62	-0.20	9.71	-1.30	8.57	0.28	-0.33
-0.58	0.44	0.78	0.71	10.11	1.14	0.44	-0.49
-0.39	0.67	0.07	0.38	8.82	0.09	0.28	0.41
0.34	0.11	0.26	0.18	8.93	0.41	0.47	0.37
0.09	-0.32	0.78	-0.20	9.78	0.05	-0.09	0.49
0.16	-0.83	0.09	0.12	9.15	-0.11	-0.08	0.01

5.4.1 Ο Μετασχηματισμός DCT ως βάση

Μέχρι στιγμής η αναφορά στον DCT επικεντρώθηκε σε ότι αφορά τον τρόπο με τον οποίο χρησιμοποιείται ο μετασχηματισμός για να συμπιέσει μονοδιάστατα ή δισδιάστατα δεδομένα. Στόχος αυτής της ενότητας είναι να αναδείξει το λόγο για τον οποίο ο μετασχηματισμός DCT συμπεριφέρεται με αυτό τον τρόπο καθώς και τον τρόπο που προέκυψαν οι εξισώσεις (5.12) και (5.14) του μετασχηματισμού[37]. Το ζήτημα αυτό θα προσεγγιστεί με δύο διαφορετικούς τρόπους. Η πρώτη επεξήγηση θα πραγματοποιηθεί θεωρώντας τον DCT ως μια ειδική βάση ενός n-διάστατου διανυσματικού χώρου. Θα δειχθεί ότι ο μετασχηματισμός ενός διανύσματος δεδομένων p με τον DCT, είναι ισοδύναμος με την αναπαράσταση του σε αυτή την ειδική βάση η οποία απομονώνει τις διάφορες συχνότητες που περιέχονται στο διάνυσμα αυτό. Συνεπώς οι συντελεστές DCT που προκύπτουν από το μετασχηματισμό ενός διανύσματος p ουσιαστικά υποδεικνύουν τις διάφορες συχνότητες που περιέχονται στο διάνυσμα. Οι χαμηλές συχνότητες περιέχουν τη σημαντική πληροφορία του p ενώ οι υψηλές συχνότητες αντιστοιχούν στις λεπτομέρειες των δεδομένων του p και επομένως είναι λιγότερο σημαντικές. Αυτός είναι και ο λόγος που οι συντελεστές αυτοί κβαντίζονται ισχυρότερα.

Η δεύτερη επεξήγηση του DCT είναι ως μετασχηματισμός περιστροφής όπως αναφέραμε και νωρίτερα σε προηγούμενη ενότητα για $n = 2$ (δισδιάστατα σημεία). Αυτή η προσέγγιση θεωρεί τον DCT ως έναν πίνακα περιστροφής ο οποίος περιστρέφει ένα

n-διάστατο σημείο με πανομοιότυπες συντεταγμένες $(\chi, \chi, \dots, \chi)$ από την αρχή των αξόνων ως προς τον άξονα x, όπου οι συντεταγμένες γίνονται $(\alpha, \varepsilon_2, \dots, \varepsilon_n)$ όπου τα ε_i είναι μικροί αριθμοί ή μηδενικά. Και οι δύο προσεγγίσεις παρουσιάζονται για την περίπτωση $n = 3$ καθότι αυτός είναι ο μεγαλύτερος αριθμός διαστάσεων που μπορεί να αναπαρασταθεί γραφικά.

Για την ειδική περίπτωση όπου $n = 3$ η συνάρτηση του DCT γίνεται:

$$G_f = \sqrt{\frac{2}{3}} C_f \sum_{t=0}^{2} p_t \cos\left[\frac{(2t+1)f\pi}{6}\right], \qquad \gamma\iota\alpha\ f = 0,1,2.$$

Αγνοώντας προσωρινά τους παράγοντες κανονικοποίησης $\sqrt{2/3}$ και C_f, αυτό μπορεί να γραφτεί με σημειολογία πίνακα ως:

$$\begin{bmatrix} G_0 \\ G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} cos0 & cos0 & cos0 \\ cos\frac{\pi}{6} & cos\frac{3\pi}{6} & cos\frac{5\pi}{6} \\ cos2\frac{\pi}{6} & cos2\frac{3\pi}{6} & cos2\frac{5\pi}{6} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = DP$$

Επομένως, ο μετασχηματισμός DCT των τριών τιμών $P = (p_0, p_1, p_2)$ είναι το εσωτερικό γινόμενο του πίνακα DCT, D και του διανύσματος P. Συνεπώς μπορούμε να θεωρήσουμε τον DCT ως το εσωτερικό γινόμενο του πίνακα μετασχηματισμού D με το διάνυσμα δεδομένων. Ο πίνακας μετασχηματισμού κατασκευάζεται ως εξής: Επιλέγονται τρείς γωνίες $\frac{\pi}{6}, \frac{3\pi}{6}\ \kappa\alpha\iota \frac{5\pi}{6}$ και υπολογίζονται τα τρία διανύσματα βάσης $\cos(f\theta)$ για $f = 0,1\ \kappa\alpha\iota\ 2$ και για τις τρείς γωνίες. Τα αποτελέσματα παρουσιάζει ο Πίνακας 5.13 :

Πίνακας 5.13 : Ο Πίνακας Μετασχηματισμού DCT για $n = 3$

$\boldsymbol{\theta}$	0.5236	1.5708	2.618
$\mathbf{cos}\ (\mathbf{0}\boldsymbol{\theta})$	1	1	1
$\mathbf{cos}\ (\mathbf{1}\boldsymbol{\theta})$	0.866	0	-0.866
$\mathbf{cos}\ (\mathbf{2}\boldsymbol{\theta})$	0.5	-1	0.5

Λόγω της επιλογής των συγκεκριμένων τριών γωνιών, τα διανύσματα αυτά είναι ορθογώνια αλλά όχι ορθοκανονικά. Τα μεγέθη τους είναι $\sqrt{3}, \sqrt{1.5}\ \kappa\alpha\iota\ \sqrt{1.5}$ αντίστοιχα. Η

κανονικοποίηση των διανυσμάτων αυτών δημιουργεί τα εξής τρία διανύσματα:

$$v_1 = (0.5774, 0.5774, 05774),$$

$$v_2 = (0.7071, 0, -0.7071)\ \kappa\alpha\iota\ \ v_3 = (0.4082, -0.8165, 0.4082).$$

Εάν τα τοποθετήσουμε κατά γραμμές παράγουμε τον ακόλουθο πίνακα:

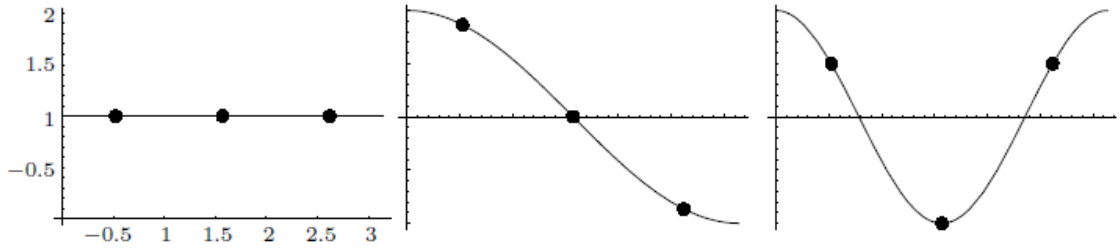
$$M = \begin{bmatrix} 0.5774 & 0.5774 & 0.5774 \\ 0.7071 & 0 & -0.7071 \\ 0.4082 & -0.8165 & 0.4082 \end{bmatrix} \tag{5.16}$$

Η εξίσωση του μετασχηματισμού DCT αναφέρει τον τρόπο κανονικοποίησης των διανυσμάτων, πολλαπλασιάζοντας καθένα με την τιμή $\sqrt{2/3}$ και στη συνέχεια το πρώτο με την τιμή $1/\sqrt{2}$. Παρατηρούμε ότι μετά την κανονικοποίηση, οι στήλες του πίνακα M είναι επίσης ορθοκανονικές, επομένως ο M είναι πλέον ένας ορθοκανονικός πίνακας.

Η μέθοδος υπολογισμού του πίνακα DCT για οποιαδήποτε τιμή του n είναι η ακόλουθη:

1. Επιλέγουμε τις n γωνίες $\theta_j = (j + 0.5)\pi/n$ για $j = 0, \dots, n-1$. Εάν διαιρέσουμε το διάστημα $[0, \pi]$ σε n ίσα τμήματα, οι γωνίες αυτές είναι τα κεντροειδή των τμημάτων.
2. Υπολογίζουμε τα n διανύσματα v_k για $k = 0, 2, \dots, n-1$, καθένα με τα n στοιχεία $\cos(k\theta_j)$.
3. Κανονικοποιούμε καθένα από τα n διανύσματα και τα τοποθετούμε ως n γραμμές ενός πίνακα.

Οι γωνίες που επιλέγονται για τον DCT είναι $\theta_j = (j + 0.5)\pi/n$, συνεπώς τα στοιχεία κάθε διανύσματος v_k είναι $cos[k(j + 0{,}5)\pi/n]$ ή $\cos[k(2j + 1)\pi/2n]$. Στη συνέχεια θα αναφερθούν τρείς τρόποι με τους οποίους μπορούν οι γωνίες αυτές να επιλεχθούν. Η επιλογή των συγκεκριμένων γωνιών έχει δύο πλεονεκτήματα. Αφενός τα διανύσματα που προκύπτουν είναι ορθογώνια και αφετέρου για αυξανόμενες τιμές του k, τα n διανύσματα v_k περιέχουν αυξανόμενες συχνότητες (Σχήμα 5.11). Για $n = 3$, η πρώτη γραμμή του πίνακα M αντιστοιχεί σε μηδενική συχνότητα, η μεσαία γραμμή (της οποίας τα στοιχεία φθίνουν μονότονα) αντιστοιχούν στη χαμηλή συχνότητα και η τελευταία γραμμή όπου τα στοιχεία εναλλάσσονται (επάνω, κάτω και τέλος επάνω) αναπαριστούν την υψηλή συχνότητα. Δοθέντος ενός τρισδιάστατου διανύσματος $v = (v_1, v_2, v_3)$, το γινόμενο Mv^T είναι μια τριάδα, τα στοιχεία της οποίας υποδεικνύουν τις τιμές των διαφόρων συχνοτήτων οι οποίες εμφανίζονται στο διάνυσμα v. Ουσιαστικά πρόκειται για συντελεστές συχνότητας. Τα ακόλουθα (ακραία) παραδείγματα παρουσιάζουν τη βασική ιδέα:



Σχήμα 5.11 : Αυξανόμενες Συχνότητες

Το πρώτο παράδειγμα είναι το $v = (u, u, u)$. Τα τρία στοιχεία του v είναι όμοια, επομένως αντιστοιχούν στη μηδενική συχνότητα. Το γινόμενο Mv^T παράγει τους συντελεστές συχνότητες $(1.7322u, 0, 0)$, που πρακτικά σημαίνει ότι δεν υπάρχουν υψηλές συχνότητες. Το δεύτερο παράδειγμα αφορά το διάνυσμα $v = (u, 0, -u)$. Τα τρία στοιχεία του v μεταβάλλονται αργά από το u έως το $-u$, επομένως αυτό το διάνυσμα περιέχει χαμηλές συχνότητες. Το γινόμενο Mv^T παράγει τους συντελεστές $(0, 1.4142u, 0)$, επιβεβαιώνοντας το αποτέλεσμα αυτό. Το τελευταίο παράδειγμα αφορά το διάνυσμα $v = (u, -u, u)$. Τα τρία στοιχεία του διανύσματος μεταβάλλονται από το u στο $-u$ και τέλος πάλι στο u, επομένως σε αυτό εμπεριέχονται υψηλές συχνότητες. Το γινόμενο Mv^T παράγει τους συντελεστές $(0, 0, 1.6329u)$το οποίο και πάλι υποδεικνύει τη σωστή συχνότητα.

Τα παραδείγματα αυτά δεν είναι ιδιαιτέρως ρεαλιστικά καθώς τα διανύσματα τα οποία δοκιμάστηκαν είναι μικρού μήκους, απλά και περιέχουν μόνο μια συχνότητα το καθένα. Τα περισσότερα διανύσματα είναι πιο σύνθετα και περιέχουν περισσότερες από μία συχνότητες το καθένα, κάτι που καθιστά τη μέθοδο αυτή εξαιρετικά χρήσιμη. Ένα απλό παράδειγμα ενός διανύσματος που περιέχει δύο συχνότητες είναι το $v = (1, 0.33, -0.34)$. Το αποτέλεσμα του γινομένου Mv^T παράγει τους συντελεστές $(0.572, 0.948, 0)$ τα οποία υποδεικνύουν μια μεγάλη χαμηλή συχνότητα, μικρή μηδενική συχνότητα και καθόλου υψηλή συχνότητα. Αυτό προφανώς είναι λογικό καθώς μπορούμε να θεωρήσουμε το διάνυσμα που ελέγχουμε ως το άθροισμα $0.33(1, 1, 1) + 0.67(1, 0, -1)$. Ένα αντίστοιχο παράδειγμα αποτελεί το άθροισμα $0.9(-1, 1, -1) + 0.1(1, 1, 1) = (-0.8, 1, -0.8)$ το οποίο όταν πολλαπλασιαστεί με τον πίνακα M δίνει $(-0.346, 0, -1.469)$. Από την άλλη πλευρά, ένα διάνυσμα το οποίο περιέχει τυχαία στοιχεία όπως $(1, 0, 0.33)$ συνήθως περιέχουν ίσα ποσά των τριών συχνοτήτων και συνεπώς παράγονται μεγάλοι συντελεστές για κάθε συχνότητα. Το γινόμενο $M(1, 0, 0.33)$ παράγει $(0.77, 0.47, 0.54)$ αφού το $(1, 0, 0.33)$ είναι το άθροισμα:

$$0.33(1,1,1) + 0.33(1,0,-1) + 0.33(1,-1,1)$$

Σημειώνεται ότι εάν $Mv = c$ τότε $M^T c = M^{-1} c = v$. Συνεπώς το αρχικό διάνυσμα v μπορεί να ανακατασκευαστεί από τους συντελεστές συχνοτήτων. Ο αντίστροφος M^{-1} του M είναι επίσης ο ανάστροφος M^T αφού ο M είναι ορθοκανονικός.

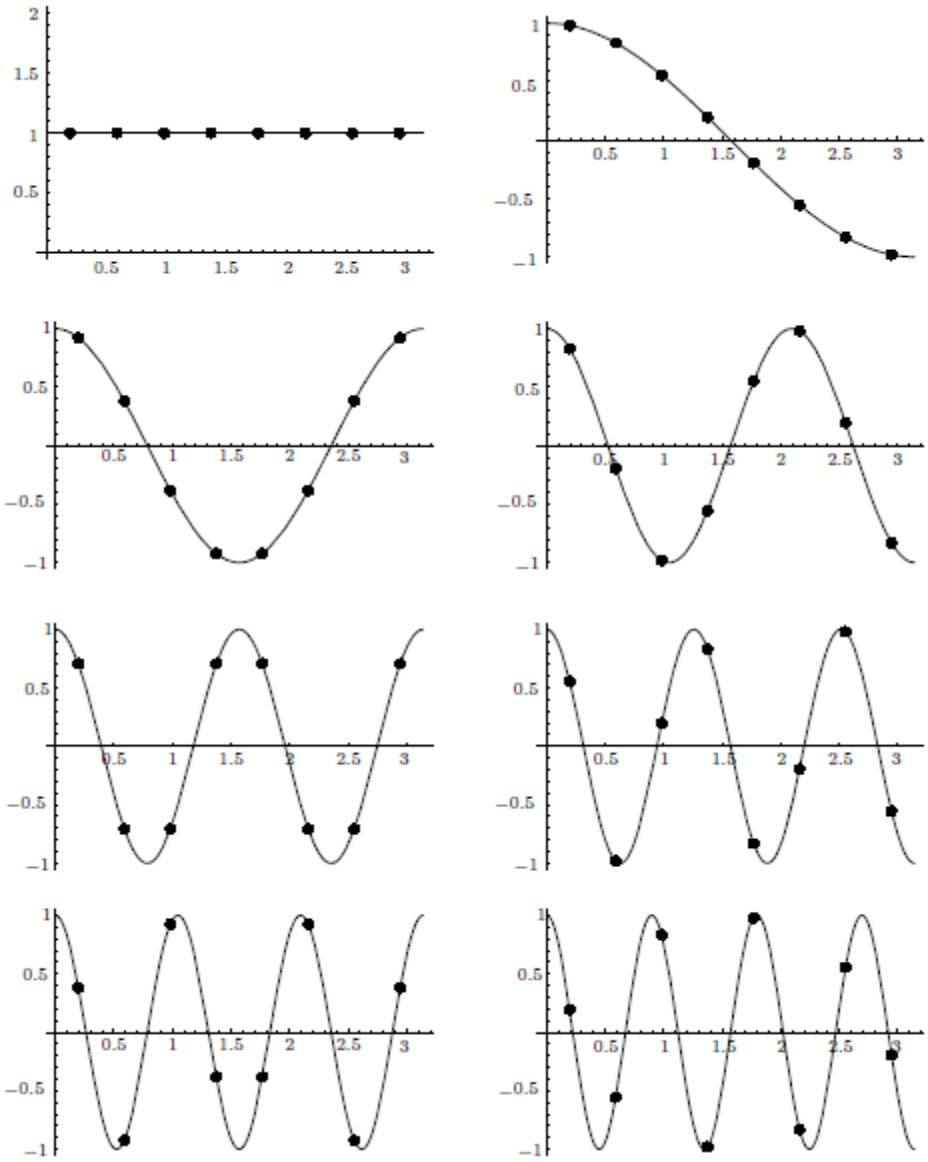
Ένα τρισδιάστατο διάνυσμα μπορεί να έχει μόνο τρείς συχνότητες. Μηδενική, μεσαία και υψηλή. Αντίστοιχα, ένα n-διάστατο διάνυσμα μπορεί να έχει n διαφορετικές συχνότητες τις οποίες η μέθοδος αυτή μπορεί να αναγνωρίσει. Αρχικά θα επικεντρωθούμε στην περίπτωση του μονοδιάστατου μετασχηματισμού και για $n = 8$. Στο Σχήμα 5.12 παρουσιάζονται οκτώ συνημιτονοειδείς κυματομορφές της μορφής $\cos(f\theta_j)$ για $0 \leq \theta_j \leq \pi$, με συχνότητες $f = 0,1,\ldots,7$. Κάθε κυματομορφή έχει δειγματοληπτηθεί σε οκτώ σημεία

$$\theta_j = \frac{\pi}{16}, \frac{3\pi}{16}, \frac{5\pi}{16}, \frac{7\pi}{16}, \frac{9\pi}{16}, \frac{11\pi}{16}, \frac{13\pi}{16}, \frac{15\pi}{16} \tag{5.17}$$

προκειμένου να σχηματισθεί ένα διάνυσμα βάσης v_f καθώς και τα οκτώ διανύσματα v_f για $f = 0,1,\ldots,7$ (δηλαδή συνολικά 64 συντελεστές) τα οποία παρουσιάζει ο Πίνακας 5.14.

Πίνακας 5.14 : Ο μη-Κανομικοποιημένος πίνακας DCT στη μία Διάσταση για $n = 8$

Θ	*0.196*	*0.589*	*0.982*	*1.374*	*1.767*	*2.160*	*2.553*	*2.945*
Cos 0θ	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Cos 1θ	*0.981*	*0.831*	*0.556*	*0.195*	*-0.195*	*-0.556*	*-0.831*	*-0.981*
Cos 2θ	*0.924*	*0.383*	*-0383*	*-0.924*	*-0.924*	*-0.383*	*0.383*	*0.924*
Cos 3θ	*0.831*	*-0.195*	*-0.981*	*-0.556*	*0.556*	*0.981*	*0.195*	*-0.831*
Cos 4θ	*0.707*	*-0.707*	*-0.707*	*0.707*	*0.707*	*-0.707*	*-0.707*	*0.707*
Cos 5θ	*0.556*	*-0.981*	*0.195*	*0.831*	*-0.831*	*-0.195*	*0.981*	*-0.556*
Cos 6θ	*0.383*	*-0.924*	*0.924*	*-0.383*	*-0.383*	*0.924*	*-0.924*	*0.383*
Cos 7θ	*0.195*	*-0.556*	*0.831*	*-0.981*	*0.981*	*-0.831*	*0.556*	*-0.195*



Σχήμα 5.12 : Οι Γωνίες και οι Τιμές του Συνημίτονου για τον Μετασχηματισμό DCT 8-Σημείων

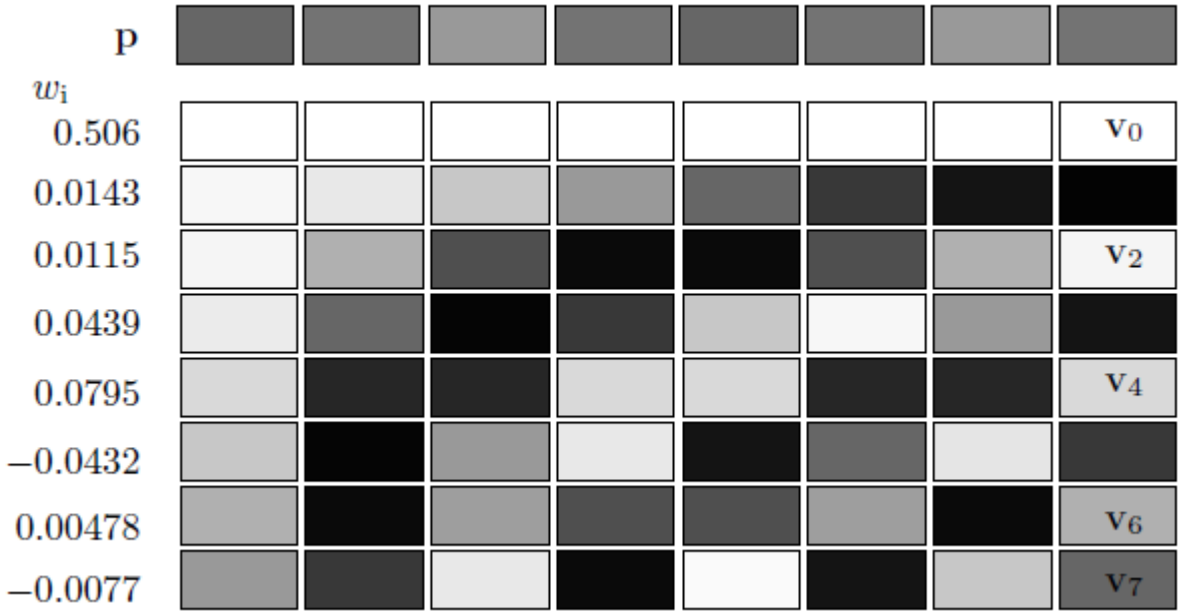
Ουσιαστικά αποτελούν έναν πίνακα βάσης του μετασχηματισμού DCT. Παρατηρούμε τη διαφορά με τον πίνακα W. Λόγω της συγκεκριμένης επιλογής των οκτώ δειγμάτων, τα v_i είναι ορθογώνια. Αυτό είναι εύκολο να ελεγχθεί με κάποιο μαθηματικό λογισμικό αλλά στη συνέχεια θα δούμε έναν πιο εκλεπτυσμένο τρόπο για να αποδείξουμε τη συγκεκριμένη ιδιότητα. Μετά την κανονικοποίηση των διανυσμάτων, τα v_i μπορούν να θεωρηθούν είτε ως ένας πίνακας μετασχηματισμού 8×8 (συγκεκριμένα, ένας πίνακας περιστροφής, δεδομένου ότι είναι ορθοκανονικός) είτε ως ένα σύνολο οκτώ ορθογώνιων διανυσμάτων τα οποία αποτελούν μια βάση του διανυσματικού χώρου. Κάθε διάνυσμα p σε αυτό το χώρο μπορεί να εκφραστεί ως γραμμικός συνδυασμός των v_i. Ως παράδειγμα μπορεί να θεωρηθεί το διάνυσμα p με τους οκτώ ισχυρά συσχετισμένους

αριθμούς $p = (0.6, 0.5, 0.4, 0.5, 0.6, 0.5, 0.4, 0.55)$ το οποίο μπορεί να εκφρασθεί ως γραμμικός συνδυασμός $p = \sum w_i v_i$ των οκτώ συναρτήσεων βάσης v_i. Η λύση του συστήματος των οκτώ εξισώσεων οδηγεί στα ακόλουθα οκτώ βάρη:

$$w_0 = 0.506, w_1 = 0.0143, w_2 = 0.0115, w_3 = 0.0439,$$

$$w_4 = 0.0795, w_5 = -0.0432, w_6 = 0.00478, w_7 = -0.0077$$

Το βάρος w_0 δεν είναι πολύ διαφορετικό από τα στοιχεία του p όμως οι υπόλοιποι επτά συντελεστές είναι εμφανώς μικρότεροι. Αυτός είναι ο τρόπος με τον οποίο ο DCT (ή και άλλοι ορθογώνιοι μετασχηματισμοί) οδηγεί σε συμπίεση. Τα οκτώ αυτά βάρη μπορούν να κβαντιστούν και στη συνέχεια να εγγραφούν στη συμπιεσμένη ακολουθία δεδομένων στην οποία θα καταλαμβάνουν σημαντικά λιγότερο χώρο σε σχέση με τους οκτώ αρχικούς αριθμούς. Το Σχήμα 5.13 αναπαριστά με γραφικό τρόπο τον γραμμικό αυτό συνδυασμό.



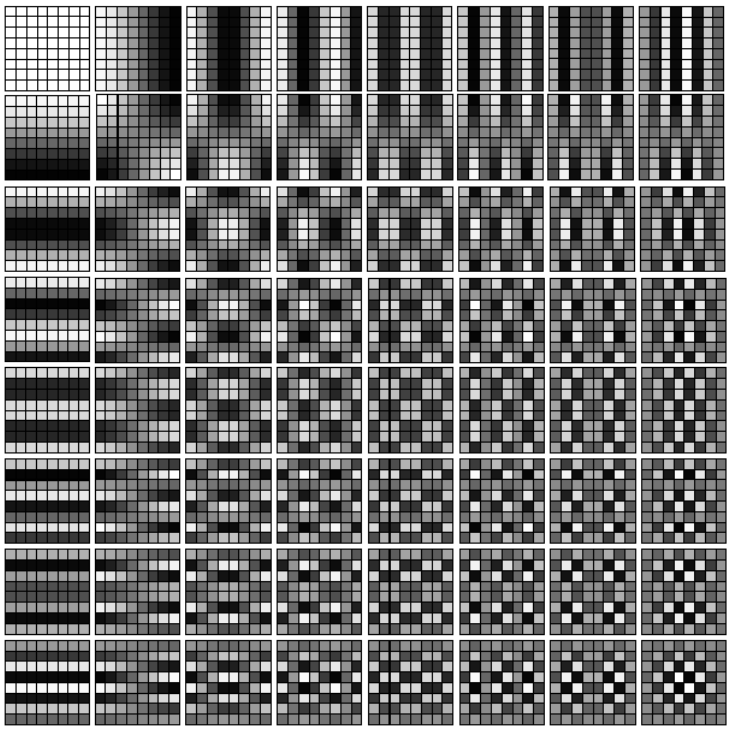
Σχήμα 5.13 : Γραφική Αναπαράσταση του Μονοδιάστατου DCT

Καθένα από τα οκτώ v_i το οποίο παρουσιάζεται ως μια γραμμή από οκτώ ορθογώνια σε κλίμακα του γκρι (εικόνα βάσης) όπου η τιμή +1 εμφανίζεται με λευκό χρώμα και η τιμή -1 με μαύρο. Τα οκτώ στοιχεία του διανύσματος p απεικονίζονται και αυτά στην πρώτη γραμμή του σχήματος με τον ίδιο τρόπο.

Συνοψίζοντας, ερμηνεύσαμε τον μονοδιάστατο μετασχηματισμό DCT ως ένα σύνολο εικόνων βάσης οι οποίες έχουν όλο και υψηλότερες συχνότητες. Δοθέντος ενός διανύσματος δεδομένων, ο DCT διαχωρίζει τις συχνότητες των δεδομένων και το

αναπαριστά ως γραμμικό συνδυασμό (ή τον σταθμισμένο μέσο όρο) των εικόνων βάσης. Τα βάρη είναι οι συντελεστές DCT.

Η ερμηνεία αυτή μπορεί να επεκταθεί και στις δύο διαστάσεις. Εφαρμόζοντας τη συνάρτηση (5.14) στην περίπτωση όπου $n = 8$ για να δημιουργήσουμε 64 μικρές εικόνες βάσης των 8 x 8 εικονοστοιχείων η κάθε μία. Οι 64 αυτές εικόνες χρησιμοποιούνται στη συνέχεια ως βάση ενός 64-διάστατου διανυσματικού χώρου. Κάθε εικόνα B μεγέθους 8 x 8 θα μπορεί να εκφρασθεί ως γραμμικός συνδυασμός των εικόνων βάσης και τα 64 βάρη του γραμμικού αυτού συνδυασμού θα είναι οι συντελεστές DCT της εικόνας B.



Σχήμα 5.14 : Οι 64 Εικόνες Βάσης του Δισδιάστατου Μετασχηματισμού DCT

5.4.2 Ο DCT ως Μετασχηματισμός Περιστροφής

Η δεύτερη ερμηνεία του πίνακα M είναι η περιστροφή. Ήδη γνωρίζουμε ότι το γινόμενο $M(u, u, u)$ έχει ως αποτέλεσμα $(1.7322u, 0{,}0)$ και μπορεί να θεωρηθεί ως η περιστροφή του σημείου (u, u, u) στο σημείο $(1.7322u, 0{,}0)$. Το πρώτο στοιχείο τοποθετείται στη γραμμή που σχηματίζει ίσες γωνίες με τις συντεταγμένες των τριών αξόνων και το τελευταίο στοιχείο βρίσκεται στον άξονα x. Εάν αναγάγουμε τη θεώρηση αυτή σε μία εικόνα, τότε η περιστροφή έχει πολλή απλή έννοια. Μπορούμε να φανταστούμε τρία γειτονικά στοιχεία μιας εικόνας. Συνήθως έχουν παρόμοιες τιμές φωτεινότητας. Αρχικά θα θεωρήσουμε πως είναι απολύτως όμοια. Όταν τα τρία αυτά εικονοστοιχεία

θεωρηθούν ως ένα σημείο του τρισδιάστατου χώρου τότε το σημείο αυτό θα βρίσκεται στην ευθεία $x = y = z$. Η περιστροφή της γραμμής αυτής ως προς τον άξονα x έχει ως αποτέλεσμα το σημείο που βρισκόταν στον άξονα x να μην έχει αλλάξει ιδιαίτερα αλλά οι συντεταγμένες $y\ και\ z$ να μηδενιστούν. Με αυτό τον τρόπο η περιστροφή οδηγεί σε συμπίεση. Στην πράξη τρία γειτονικά εικονοστοιχεία $p_1, p_2\ και\ p_3$ είναι όμοια αλλά όχι απόλυτα. Αυτό σημαίνει ότι το σημείο (p_1, p_2, p_3) είναι λίγο εκτός της γραμμής $x = y = z$. Μετά την περιστροφή, το σημείο θα βρίσκεται κοντά στον άξονα x και οι συντεταγμένες y, z θα είναι μικροί αριθμοί.

Η θεώρηση αυτή του πίνακα M είναι λογική δεδομένου ότι ο πίνακας M είναι ορθοκανονικός και κάθε ορθοκανονικός πίνακας είναι πίνακας περιστροφής. Παρόλα αυτά, η ορίζουσα ενός πίνακα περιστροφής πρέπει να έχει τιμή 1. Στην περίπτωση μας όμως η ορίζουσα του πίνακα έχει τιμή -1. Ένας ορθοκανονικός πίνακας του οποίου η ορίζουσα είναι -1 εκτελεί περιστροφή η οποία συνοδεύεται και από ανάκλαση. Για να γίνει καλύτερα αυτό αντιληπτό εφαρμόζουμε τη μέθοδο QR αποδόμησης ενός πίνακα προκειμένου να αποδομήσουμε τον πίνακα M στο γινόμενο $T1\ X\ T2\ X\ T3\ X\ T4$ όπου

$$T1 = \begin{bmatrix} 0.8165 & 0 & -0.5774 \\ 0 & 1 & 0 \\ 0.5774 & 0 & 0.8165 \end{bmatrix}, T2 = \begin{bmatrix} 0.7071 & -0.7071 & 0 \\ 0.7071 & 0.7071 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$T3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, T4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Κάθε πίνακας $T1$, $T2$ και $T3$ εκτελεί μια περιστροφή στον κάθε άξονα του τρισδιάστατου χώρου. Ο πίνακας $T4$ είναι η ανάκλαση στον άξονα z. Ο μετασχηματισμός $M(1,1,1)$ μπορεί πλέον να γραφεί ως $T1\ X\ T2\ X\ T3\ X\ T4\ X\ (1,1,1)$ όπου ο $T4$ ανακλά το σημείο $(1,1,1)$ στο $(1,1,-1)$, ο $T3$ περιστρέφει το $(1,1,-1)$ 90° ως προς τον άξονα x στο σημείο $(1,-1,-1)$, το οποίο περιστρέφεται από τον $T2$ κατά 45° ως προς τον άξονα z στο σημείο $(1.4142, 0, -1)$ και το οποίο περιστρέφεται από τον $T1$ κατά 35.26° γύρω από τον άξονα $(1.7321,0,0)$.

Για κάθε n η ιδέα είναι η ίδια. Ξεκινάμε με ένα διάνυσμα από n γειτονικά εικονοστοιχεία τα οποία και θεωρούμε ως συντεταγμένες ενός σημείου του n-διάστατου χώρου. Εφόσον τα εικονοστοιχεία είναι παρόμοια, το σημείο αυτό βρίσκεται κοντά στην ευθεία η οποία δημιουργεί ίσες γωνίες με τους άξονες όλων των συνιστωσών. Η εφαρμογή του μετασχηματισμού DCT σε μια διάσταση περιστρέφει το σημείο αυτό φέρνοντας το πιο κοντά στον άξονα x, όπου η πρώτη συντεταγμένη δεν έχει μεταβληθεί ιδιαίτερα αλλά οι υπόλοιπες $n - 1$ είναι πλέον μικροί αριθμοί. Αυτός είναι ο τρόπος με τον οποίο ο

μονοδιάστατος μετασχηματισμός DCT εκτελεί μια περιστροφή στο n-διάστατο χώρο. Η περιστροφή μπορεί να διαχωριστεί σε μία ανάκλαση ακολουθούμενη από $n-1$ περιστροφές.

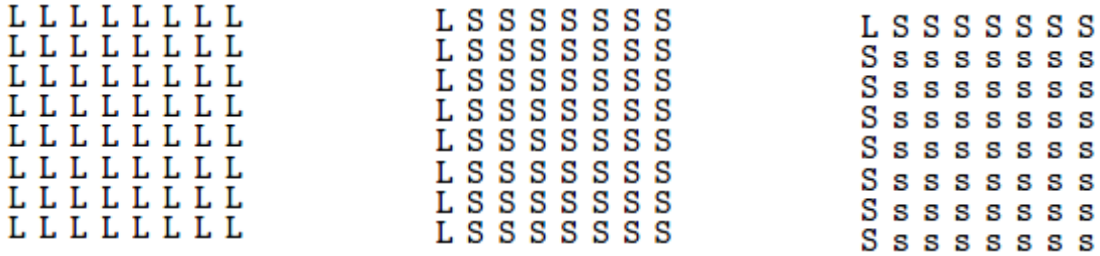
Ο DCT στις δύο διαστάσεις ερμηνεύεται με τον ίδιο τρόπο αλλά ως διπλή περιστροφή. Σε αυτή τη θεώρηση ξεκινάμε με ένα μπλοκ των $n \times n$ εικονοστοιχείων (στο Σχήμα 5.15 σημειώνονται με L). Αρχικά θεωρούμε κάθε γραμμή αυτού του μπλοκ ως ένα σημείο $(p_{x,0}, p_{x,1}, \dots, p_{x,n-1})$ στον n-διάστατο χώρο και στρέφουμε το σημείο αυτό σύμφωνα με το εσωτερικό άθροισμα της εξίσωσης (5.14)

$$G1_{x,j} = \sqrt{\frac{2}{n}} C_j \sum_{y=0}^{n-1} p_{xy} \cos\left(\frac{(2y+1)j\pi}{2n}\right)$$

Αυτό έχει ως αποτέλεσμα ένα μπλοκ $G1_{x,j}$ από $n \times n$ συντελεστές όπου το πρώτο στοιχείο κάθε γραμμής κυριαρχεί (σημειώνονται με L στο Σχήμα 5.15 b) ενώ τα υπόλοιπα στοιχεία είναι μικρά (σημειώνονται με S). Το εξωτερικό άθροισμα της εξίσωσης (5.14) είναι:

$$G_{ij} = \sqrt{\frac{2}{n}} C_i \sum_{x=0}^{n-1} G1_{x,j} \cos\left(\frac{(2x+1)i\pi}{2n}\right)$$

Εδώ οι στήλες του $G1_{x,j}$ θεωρούνται σημεία του n-διάστατου χώρου και περιστρέφονται. Το αποτέλεσμα είναι ένας συντελεστής με μεγάλη τιμή στην επάνω αριστερή γωνία του μπλοκ και n^2-1 μικροί συντελεστές οπουδήποτε αλλού (S και s στο σχήμα). Σύμφωνα με την ερμηνεία αυτή, ο δισδιάστατος μετασχηματισμός DCT θεωρείται ως δύο ξεχωριστές περιστροφές στις n διαστάσεις. Η πρώτη περιστρέφει κάθε μια από τις n γραμμές ενώ η δεύτερη κάθε μια από τις n στήλες. Αξίζει να σημειωθεί ότι οι 2ν περιστροφές στις n διαστάσεις είναι μια ταχύτερη διαδικασία από μια περιστροφή στις n^2 διαστάσεις αφού στη δεύτερη απαιτούνται $n^2 \times n^2$ πίνακες περιστροφής.



(a) (b) (c)

Σχήμα 5.15 : Ο Δισδιάστατος Μετασχηματισμός DCT ως Διπλή Περιστροφή

5.4.3 Οι τέσσερεις τύποι DCT

Υπάρχουν τέσσερεις τρόποι για να επιλέξει κανείς n ισαπέχουσες γωνίες που δημιουργούν ορθογώνια διανύσματα συνημίτονων. Αντιστοιχούν (μετά την κανονικοποίηση τους) σε τέσσερεις διακριτούς μετασχηματισμούς συνημίτονου, γνωστούς ως DCT-1 έως DCT-4. Ο πιο χρήσιμος είναι ο DCT-2 ο οποίος συνήθως αναφέρεται ως ο μετασχηματισμός DCT. Οι επόμενες εξισώσεις είναι οι ορισμοί των τεσσάρων αυτών τύπων DCT. Τις γωνίες των DCT-1 και DCT-2 παρουσιάζει ο Πίνακας 5.15 και Πίνακας 5.16 αντίστοιχα. Σημειώνεται ότι ο DCT-3 είναι ο ανάστροφος του DCT-2 και ο DCT-4 είναι η μετατοπισμένη εκδοχή του DCT-1. Παρατηρούμε ότι ο DCT-1 έχει $n+1$ διανύσματα αποτελούμενα από $n+1$ συνημίτονα το καθένα. Σε καθένα από τους τέσσερεις τύπους, τα n (ή τα $n+1$) διανύσματα DCT είναι ορθογώνια και γίνονται ορθοκανονικά αφού πολλαπλασιαστούν με τον σωστό παράγοντα.

$$DCT_{k,j}^{1} = \sqrt{\frac{2}{n}} C_k C_j \cos\left[\frac{kj\pi}{n}\right], \qquad k,j = 0,1,\dots,n$$

$$DCT_{k,j}^{2} = \sqrt{\frac{2}{n}} C_k \cos\left[\frac{k\left(j+\frac{1}{2}\right)\pi}{n}\right], \qquad k,j = 0,1,\dots,n-1$$

$$DCT_{k,j}^{3} = \sqrt{\frac{2}{n}} C_j \cos\left[\frac{(k+\frac{1}{2})j\pi}{n}\right], \qquad k,j = 0,1,\dots,n-1 \tag{5.18}$$

$$DCT_{k,j}^{4} = \sqrt{\frac{2}{n}} \cos\left[\frac{\left(k+\frac{1}{2}\right)\left(j+\frac{1}{2}\right)\pi}{n}\right], \qquad k,j = 0,1,\dots,n-1$$

όπου ο παράγοντας κανονικοποίησης C_x ορίζεται ως

$$C_x = \begin{cases} \frac{1}{\sqrt{2}}, & \gamma\iota\alpha\ x = 0\ \acute{\eta}\ x = n \\ 1, & \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\acute{\alpha} \end{cases}$$

Πίνακας 5.15 : Οι Τιμές των Επιλεχθέντων Γωνιών του Μετασχηματισμού DCT-1 $(\mathbf{9} \times \mathbf{9})$

K	Scale									
0	$\frac{1}{2\sqrt{2}}$	0	0	0	0	0	0	0	0	0
1	$\frac{1}{2}$	0	$\frac{\pi}{8}$	$\frac{2\pi}{8}$	$\frac{3\pi}{8}$	$\frac{4\pi}{8}$	$\frac{5\pi}{8}$	$\frac{6\pi}{8}$	$\frac{7\pi}{8}$	$\frac{8\pi}{8}$
2	$\frac{1}{2}$	0	$\frac{2\pi}{8}$	$\frac{4\pi}{8}$	$\frac{6\pi}{8}$	$\frac{8\pi}{8}$	$\frac{10\pi}{8}$	$\frac{12\pi}{8}$	$\frac{14\pi}{8}$	$\frac{16\pi}{8}$
3	$\frac{1}{2}$	0	$\frac{3\pi}{8}$	$\frac{6\pi}{8}$	$\frac{9\pi}{8}$	$\frac{12\pi}{8}$	$\frac{15\pi}{8}$	$\frac{18\pi}{8}$	$\frac{21\pi}{8}$	$\frac{24\pi}{8}$
4	$\frac{1}{2}$	0	$\frac{4\pi}{8}$	$\frac{8\pi}{8}$	$\frac{12\pi}{8}$	$\frac{16\pi}{8}$	$\frac{20\pi}{8}$	$\frac{24\pi}{8}$	$\frac{28\pi}{8}$	$\frac{32\pi}{8}$
5	$\frac{1}{2}$	0	$\frac{5\pi}{8}$	$\frac{10\pi}{8}$	$\frac{15\pi}{8}$	$\frac{20\pi}{8}$	$\frac{25\pi}{8}$	$\frac{30\pi}{8}$	$\frac{35\pi}{8}$	$\frac{40\pi}{8}$
6	$\frac{1}{2}$	0	$\frac{6\pi}{8}$	$\frac{12\pi}{8}$	$\frac{18\pi}{8}$	$\frac{24\pi}{8}$	$\frac{30\pi}{8}$	$\frac{36\pi}{8}$	$\frac{42\pi}{8}$	$\frac{48\pi}{8}$
7	$\frac{1}{2}$	0	$\frac{7\pi}{8}$	$\frac{14\pi}{8}$	$\frac{21\pi}{8}$	$\frac{28\pi}{8}$	$\frac{35\pi}{8}$	$\frac{42\pi}{8}$	$\frac{49\pi}{8}$	$\frac{56\pi}{8}$
8	$\frac{1}{2\sqrt{2}}$	0	$\frac{8\pi}{8}$	$\frac{16\pi}{8}$	$\frac{24\pi}{8}$	$\frac{32\pi}{8}$	$\frac{40\pi}{8}$	$\frac{48\pi}{8}$	$\frac{56\pi}{8}$	$\frac{64\pi}{8}$

Η ορθογωνιότητα μπορεί να αποδειχθεί είτε άμεσα, πολλαπλασιάζοντας τα ζεύγη των διαφόρων διανυσμάτων είτε έμμεσα. Στην περίπτωση αυτή αποδεικνύεται ότι τα διανύσματα του DCT είναι ορθογώνια εάν τα διανύσματα αυτά είναι τα ιδιοδιανύσματα συγκεκριμένων συμμετρικών πινάκων. Στην περίπτωση του DCT-2, για παράδειγμα, ο συμμετρικός πίνακας είναι ο

$$A = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \vdots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

Για $n = 3$ ο πίνακας γίνεται

$$A_3 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Με ιδιοδιανύσματα $(0.5774, 0.5774, 0.5774), (0.7071, 0, -0.7071)$,

$(0.4082, -0.8165, 0.4082)$ και ιδιοτιμές $0, 1\ \kappa\alpha\iota\ 3$ αντίστοιχα. Υπενθυμίζουμε ότι τα ιδιοδιανύσματα αυτά είναι οι γραμμές του πίνακα M τις εξίσωσης (5.16).

Πίνακας 5.16 : Οι Τιμές των Επιλεχθέντων Γωνιών του Μετασχηματισμού DCT-2 (8×8)

K	Scale	Γωνίες DCT-2 (8×8)							
0	$\frac{1}{2\sqrt{2}}$	0	0	0	0	0	0	0	0
1	$\frac{1}{2}$	$\frac{\pi}{16}$	$\frac{3\pi}{16}$	$\frac{5\pi}{16}$	$\frac{7\pi}{16}$	$\frac{9\pi}{16}$	$\frac{11\pi}{16}$	$\frac{13\pi}{16}$	$\frac{15\pi}{16}$
2	$\frac{1}{2}$	$\frac{2\pi}{16}$	$\frac{6\pi}{16}$	$\frac{10\pi}{16}$	$\frac{14\pi}{16}$	$\frac{18\pi}{16}$	$\frac{22\pi}{16}$	$\frac{26\pi}{16}$	$\frac{30\pi}{16}$
3	$\frac{1}{2}$	$\frac{3\pi}{16}$	$\frac{9\pi}{16}$	$\frac{15\pi}{16}$	$\frac{21\pi}{16}$	$\frac{27\pi}{16}$	$\frac{33\pi}{16}$	$\frac{39\pi}{16}$	$\frac{45\pi}{16}$
4	$\frac{1}{2}$	$\frac{4\pi}{16}$	$\frac{12\pi}{16}$	$\frac{20\pi}{16}$	$\frac{28\pi}{16}$	$\frac{36\pi}{16}$	$\frac{44\pi}{16}$	$\frac{52\pi}{16}$	$\frac{60\pi}{16}$
5	$\frac{1}{2}$	$\frac{5\pi}{16}$	$\frac{15\pi}{16}$	$\frac{25\pi}{16}$	$\frac{35\pi}{16}$	$\frac{45\pi}{16}$	$\frac{55\pi}{16}$	$\frac{65\pi}{16}$	$\frac{75\pi}{16}$
6	$\frac{1}{2}$	$\frac{6\pi}{16}$	$\frac{18\pi}{16}$	$\frac{30\pi}{16}$	$\frac{42\pi}{16}$	$\frac{54\pi}{16}$	$\frac{66\pi}{16}$	$\frac{78\pi}{16}$	$\frac{90\pi}{16}$
7	$\frac{1}{2}$	$\frac{7\pi}{16}$	$\frac{21\pi}{16}$	$\frac{35\pi}{16}$	$\frac{49\pi}{16}$	$\frac{63\pi}{16}$	$\frac{77\pi}{16}$	$\frac{91\pi}{16}$	$\frac{105\pi}{16}$

6. Το Πρότυπο Συμπίεσης JPEG

Στόχος του προτύπου JPEG[38][39] ήταν η ανάπτυξη μιας μεθόδου συμπίεσης εικόνας η οποία να πληροί τις ακόλουθες απαιτήσεις:

1. Να επιτυγχάνει την καλύτερη δυνατή συμπίεση σε σχέση με τον επιθυμητό ρυθμό και την πιστότητα της εικόνας, για πλήθος διαφορετικών ποιοτήτων εικόνας, ειδικά στο εύρος όπου η οπτική ποιότητα της εικόνας θεωρείτε από τον παρατηρητή από πολύ καλή έως άριστη. Παράλληλα ο κωδικοποιητής θα πρέπει να είναι παραμετροποιήσημος ούτως ώστε να μπορεί να καθορίζεται από την εφαρμογή η επιθυμητή ποιότητα της συμπιεσμένης εικόνας.
2. Να μπορεί να εφαρμοστεί πρακτικά σε κάθε είδος ψηφιακής εικόνας χωρίς να περιορίζεται από τον τύπο της πηγαίας εικόνας (όπως για παράδειγμα το μέγεθος της ή το πλήθος χρωμάτων της) και χωρίς περιορισμούς που αφορούν την πολυπλοκότητα της σκηνής, τις στατιστικές της ιδιότητες ή το εύρος των χρωμάτων που την αποτελούν.
3. Να έχει χαμηλή υπολογιστική πολυπλοκότητα προκειμένου να καταστούν εφικτές υλοποιήσεις λογισμικού με καλές επιδόσεις για ένα πλήθος διαφορετικών επεξεργαστών καθώς και υλοποιήσεις με χρήση υλικού για εφαρμογές υψηλών απαιτήσεων χωρίς όμως τη σημαντική αύξηση του κόστους.
4. Να περιλαμβάνει τους ακόλουθους τρόπους κωδικοποίησης:
    - Ακολουθιακή κωδικοποίηση: Κάθε στοιχείο της εικόνας κωδικοποιείται με ένα σάρωμα από τα αριστερά προς τα δεξιά και από πάνω προς τα κάτω.
    - Προοδευτική κωδικοποίηση: Η εικόνα κωδικοποιείται με πολλαπλά περάσματα για εφαρμογές όπου ο χρόνος μετάδοσης είναι μεγάλος και ο παρατηρητής προτιμά να παρατηρεί την εικόνα να ανακτάται και να δομείται με πολλά περάσματα ξεκινώντας από εικόνα χαμηλής ευκρίνειας για να καταλήξει στην τελική καθαρή της μορφή.
    - Συμπίεση χωρίς απώλειες: Η εικόνα κωδικοποιείται με τέτοιο τρόπο ώστε να εγγυάται την ακριβή ανασύσταση κάθε δείγματος της αρχικής εικόνας (παρόλα αυτά η συμπίεση που επιτυγχάνεται σε αυτή την περίπτωση είναι μικρή σε σχέση με τις μεθόδους συμπίεσης με απώλειες.
    - Ιεραρχική κωδικοποίηση: Η εικόνα κωδικοποιείται σε πολλές αναλύσεις και επομένως οι εκδόσεις της σε χαμηλές αναλύσεις μπορούν να προσπελαστούν χωρίς να απαιτείται πρώτα αποσυμπίεση της εικόνας στην πλήρη ανάλυση της.

Το πρότυπο περιλαμβάνει τις τέσσερεις διαφορετικούς μεθόδους κωδικοποίησης που αναφέρθηκαν παραπάνω. Για κάθε μια, ένας ή και περισσότεροι κωδικό-αποκωδικοποιητές (codecs) έχουν καθοριστεί. Οι διαφορετικοί κωδικό-αποκωδικοποιητές οι οποίοι υλοποιούν μια μέθοδο διαφέρουν σχετικά με την ακρίβεια των δειγμάτων της πηγαίας εικόνας που μπορούν να διαχειριστούν ή του κωδικοποιητή εντροπίας τον οποίο χρησιμοποιούν. Παρά το γεγονός ότι αναφερόμαστε σε κωδικό-αποκωδικοποιητές δεν υπάρχει κάποια απαίτηση για την υλοποίηση τόσο του

κωδικοποιητή όσο και του αποκωδικοποιητή σε ένα σύστημα. Πολλές εφαρμογές και συστήματα υλοποιούν μόνο τον κωδικοποιητή ή μόνο τον αποκωδικοποιητή.

Οι τέσσερεις αυτές μέθοδοι λειτουργίας και οι διάφοροι κωδικό-αποκωδικοποιητές προέκυψαν από την ανάγκη του προτύπου JPEG να είναι όσο το δυνατόν πιο γενικευμένο καθώς και από την ποικιλία των τύπων εικόνας σε διαφορετικές εφαρμογές. Σχεδόν καμία υλοποίηση δεν υποστηρίζει κάθε εργαλείο του προτύπου. Αντίθετα, οι περισσότερες υλοποιήσεις αφορούν τον Βασικό Ακολουθιακό Κωδικό-αποκωδικοποιητή.

Ο Βασικός Ακολουθιακός κωδικό-αποκωδικοποιητής ενσωματώνει μία πλούσια και εξελιγμένη μέθοδο συμπίεσης η οποία κρίνεται επαρκής για ένα πλήθος εφαρμογών. Τα βήματα επεξεργασίας τα οποία υφίσταται το σήμα της πηγαίας εικόνας ακολουθούν στην επόμενη ενότητα.

6.1 Βασικά Βήματα JPEG

Τα βασικά βήματα επεξεργασίας του σήματος της εικόνας παρουσιάζονται στο Σχήμα 6.1 και Σχήμα 6.2 και αποτελούν την καρδιά της μεθόδου λειτουργίας η οποία χρησιμοποιεί τον μετασχηματισμό DCT. Τα σχήματα αυτά παρουσιάζουν την υποπερίπτωση όπου έχουμε εικόνα με αποχρώσεις του γκρι, δηλαδή με ένα μόνο κανάλι και όχι έγχρωμη όπου χρησιμοποιούνται τρία κανάλια (RGB). Η μεθοδολογία βεβαίως είναι πανομοιότυπη και για τις έγχρωμες εικόνες με τη διαφορά ότι ο μετασχηματισμός εφαρμόζεται σε κάθε ένα από τα τρία κανάλια της εικόνας ξεχωριστά.

Σχήμα 6.1 : Μπλοκ Διάγραμμα Κωδικοποιητή JPEG Βασισμένο στον Μετασχηματισμό DCT

Σχήμα 6.2 : Μπλοκ Διάγραμμα Αποκωδικοποιητή JPEG Βασισμένο στον Μετασχηματισμό DCT

6.1.1 Ο Μετασχηματισμός DCT στο JPEG

Στην είσοδο του κωδικοποιητή τα δείγματα της πηγαίας εικόνας ομαδοποιούνται σε μπλοκ μεγέθους 8×8. Τα δείγματα αυτά μετατοπίζουν το εύρος τιμών τους από μη προσημασμένους ακεραίους $[0, 2^P - 1]$ σε προσημασμένους $[-2^{P-1}, 2^{P-1} - 1]$ και προχωρούν στην εφαρμογή του μετασχηματισμού DCT. Αντίστοιχα, στην έξοδο του αποκωδικοποιητή ο αντίστροφος μετασχηματισμός (IDCT) παράγει ένα μπλοκ μεγέθους 8×8 με τα ανακατασκευασμένα δείγματα της εικόνας. Οι ακόλουθες εξισώσεις αποτελούν τις μαθηματικές σχέσεις του μετασχηματισμού 8×8 DCT (6.1) και 8×8 IDCT (6.2).

$$G_{ij} = \frac{1}{4} C_i C_j \sum_{\chi=0}^{7} \sum_{y=0}^{7} p_{xy} \, cos\left(\frac{(2x+1)i\pi}{16}\right) \cos\left(\frac{(2y+1)j\pi}{16}\right) \tag{6.1}$$

$$p_{xy} = \frac{1}{4} \sum_{i=0}^{7} \sum_{j=0}^{7} C_i C_j G_{ij} \, cos\left(\frac{(2x+1)i\pi}{16}\right) \cos\left(\frac{(2y+1)j\pi}{16}\right) \tag{6.2}$$

$$\text{όπου:} \begin{cases} C_f = \frac{1}{\sqrt{2}} \; \gamma\iota\alpha \; f = 0 \\ C_f = 1 \; \gamma\iota\alpha \; f \neq 0 \end{cases}$$

Ο μετασχηματισμός DCT συζητήθηκε εκτενέστατα στην ενότητα 5.4. Για λόγους πληρότητας της περιγραφής του προτύπου JPEG υπενθυμίζεται πως ο μετασχηματισμός DCT αποτελεί έναν αναλυτή των αρμονικών συνιστωσών του σήματος ενώ ο αντίστροφος DCT το συνθέτη των αρμονικών. Συνεπώς, κάθε ένα από τα μπλοκ μεγέθους 8×8 αποτελεί ένα διακριτό σήμα εξήντα-τεσσάρων σημείων το οποίο ουσιαστικά είναι μια συνάρτηση των δύο χωρικών διαστάσεων της εικόνας (x, y). Ο μετασχηματισμός λαμβάνει το σήμα αυτό ως είσοδο και προχωρά στην αποδόμηση

του σε εξήντα-τέσσερα σήματα τα οποία αποτελούν μια ορθογώνια βάση. Κάθε ένας συντελεστής περιέχει μία από τις 64 μοναδικές δισδιάστατες (2D) "χωρικές συναρτήσεις" που αποτελούν το φάσμα του σήματος εισόδου. Η έξοδος του μετασχηματισμού DCT παράγει ένα σύνολο με το πλάτος των 64 σημάτων βάσης το οποίο ονομάζεται σύνολο των συντελεστών DCT και οι τιμές τους καθορίζονται μονοσήμαντα από το σήμα εισόδου 64 σημείων.

Οι συντελεστές DCT μπορούν επομένως να θεωρηθούν ως το σχετικό ποσό των δισδιάστατων χωρικών συχνοτήτων που περιέχονται στο σήμα εισόδου 64 σημείων. Επειδή οι τιμές των δειγμάτων συνήθως μεταβάλλονται με αργούς ρυθμούς από σημείο σε σημείο κατά μήκος της εικόνας, το βήμα επεξεργασίας κατά το οποίο εφαρμόζεται ο μετασχηματισμός DCT θέτει ουσιαστικά τη βάση για την επίτευξη συμπίεσης των δεδομένων συγκεντρώνοντας την περισσότερη από την ενέργεια του σήματος στις χαμηλότερες χωρικές συχνότητες. Σε ένα τυπικό μπλοκ δειγμάτων 8×8 από μια τυπική εικόνα, οι περισσότερες χωρικές συχνότητες έχουν μηδενικό ή σχεδόν μηδενικό πλάτος και δεν χρειάζεται να κωδικοποιηθούν.

Στον αποκωδικοποιητή ο αντίστροφος μετασχηματισμός αντιστρέφει το προηγούμενα βήμα επεξεργασίας. Λαμβάνει τους 64 συντελεστές DCT (οι οποίοι σε αυτό το σημείο έχουν ήδη κβαντιστεί) και ανακατασκευάζει τα 64 σημεία του σήματος της εικόνας αθροίζοντας τα σήματα βάσης. Από μαθηματικής άποψης ο μετασχηματισμός DCT αποτελεί μια αντιστοίχηση ένα προς ένα για διανύσματα 64 σημείων ανάμεσα στην εικόνα και το πεδίο των συχνοτήτων της. Εάν ο μετασχηματισμός DCT και ο αντίστροφός του υπολογίζονταν με άριστη ακρίβεια και εάν οι συντελεστές DCT δεν κβαντίζονταν όπως θα δούμε στη συνέχεια, τότε το αρχικό σήμα της εικόνας θα μπορούσε να ανακτηθεί με απόλυτη ακρίβεια. Πρέπει να διευκρινιστεί ότι ο DCT δεν είναι ένας μετασχηματισμός που εισάγει απώλειες στα δείγματα της πηγαίας εικόνας. Απλώς μετασχηματίζει τα δείγματα της σε ένα άλλο πεδίο (συχνοτήτων) στο οποίο μπορούν να κωδικοποιηθούν αποδοτικότερα.

Κάποιες ιδιότητες του μετασχηματισμού DCT και του αντιστρόφου του αναδεικνύουν στην πράξη τι ακριβώς θα πρέπει να καθορίζεται στο πρότυπο JPEG. Μία θεμελιώδης ιδιότητα είναι ότι τόσο ο μετασχηματισμός DCT όσο και ο αντίστροφος του, χρησιμοποιούν εξισώσεις οι οποίες εμπεριέχουν υπερβατικές συναρτήσεις. Συνεπώς, δεν υπάρχει πρακτική υλοποίηση η οποία να μπορεί να παρέχει απόλυτη ακρίβεια. Λόγω της σημαντικότητας του μετασχηματισμού DCT σε διάφορες εφαρμογές και της σχέσης του με τον διακριτό μετασχηματισμό Fourier, πλήθος διαφορετικών αλγορίθμων

έχουν επινοηθεί για τον υπολογισμό του. Η έρευνα στους γρήγορους αλγορίθμους DCT έδειξε ότι δεν υπάρχει βέλτιστος αλγόριθμος για όλες τις υλοποιήσεις. Ένας αλγόριθμος που μπορεί να είναι πολύ αποδοτικός όταν υλοποιηθεί με λογισμικό για ένα πλήθος επεξεργαστών είναι σχεδόν αδύνατο να είναι εξίσου αποδοτικός εάν υλοποιηθεί σε κάποιον προγραμματιζόμενο επεξεργαστή ψηφιακού σήματος ή σε ένα κύκλωμα VLSI.

Ακόμα και υπό το πρίσμα της πεπερασμένης ακρίβειας στην είσοδο και την έξοδο του DCT, διαφορετικές σχεδιάσεις σε υλικό οι οποίες όμως υλοποιούν τον ίδιο αλγόριθμο διαφέρουν έστω και σε ένα μικρό βαθμό στην ακρίβεια που επιτυγχάνουν στην αναπαράσταση των όρων του συνημίτονου ή σε κάποια ενδιάμεσα αποτελέσματα ή ακόμα και στον τρόπο που αθροίζουν ή στρογγυλοποιούν τις τιμές. Αυτό αναμφισβήτητα επιφέρει μια απόκλιση από τα ιδανικά αποτελέσματα σε κάθε περίπτωση. Προκειμένου το πρότυπο JPEG να αφήσει κάποιο χώρο για καινοτομία σε όσους το υλοποιούν δεν προτείνει κάποιον συγκεκριμένο αλγόριθμο προς υλοποίηση. Για να μην οδηγήσει το γεγονός αυτό σε ασυμβατότητα μεταξύ των διαφόρων κωδικοποιητών και αποκωδικοποιητών, πραγματοποιείται μια σειρά από δοκιμές προκειμένου να εξασφαλιστεί μια ελάχιστη εγγύηση ακρίβειας που θα οδηγεί στην επιθυμητή ποιότητα.

Για κάθε μέθοδο λειτουργίας του προτύπου JPEG έχουν προταθεί ξεχωριστοί κωδικοποιητές για εικόνες με 8 bit ανά δείγμα και 12 bit ανά δείγμα της πηγαίας εικόνας. Οι κωδικό-αποκωδικοποιητές 12 bit χρησιμοποιούνται σε ειδικές περιπτώσεις όπως για εφαρμογές σε ιατρικές εικόνες ή άλλου τύπου όπου η μεγαλύτερη ακρίβεια είναι επιβεβλημένη. Εικόνες με διαφορετικό πλήθος bit ανά δείγμα μπορούν επίσης να υλοποιηθούν αλλά κάτι τέτοιο θα πραγματοποιηθεί εκτός του προτύπου JPEG αφού δεν υπάρχει συγκεκριμένη οδηγία μέσα σε αυτό. Για παράδειγμα, θα είναι ευθύνη της εφαρμογής να αποφασίσει τον τρόπο με τον οποίο θα προσαρμόσει την πληροφορία μιας εικόνας με 6 bit ανά δείγμα σε έναν κωδικοποιητή 8 bit ή τον τρόπο που θα τα ανακτήσει στην έξοδο του αποκωδικοποιητή ή πώς να κωδικοποιήσει την απαραίτητη σχετική πληροφορία.

6.1.2 Κβάντιση

Μετά την έξοδο του μετασχηματισμού, κάθε ένας από τους 64 συντελεστές κβαντίζεται ομοιόμορφα με τη χρήση ενός πίνακα 64 στοιχείων ο οποίος καθορίζεται από την εφαρμογή (χρήστη) και δίνεται ως είσοδος στον κωδικοποιητή. Κάθε στοιχείο του μπορεί να είναι ένας ακέραιος αριθμός από το 1 έως το 255 και το οποίο καθορίζει το

μέγεθος της κβάντισης για κάθε συντελεστή DCT. Στόχος της κβάντισης είναι να συμπιέσει περεταίρω τα δεδομένα αναπαριστώντας τους συντελεστές DCT με όση ακρίβεια είναι απαραίτητη για την επίτευξη της επιθυμητής ποιότητας της εικόνας. Από μια άλλη σκοπιά, στόχος αυτού του βήματος είναι να αγνοήσει την πληροφορία που δεν επηρεάζει αισθητά την οπτική ποιότητα της εικόνας. Η διαδικασία της κβάντισης είναι μια αντιστοίχηση πολλά προς ένα και επομένως είναι εξ ορισμού διαδικασία με απώλειες όπως συζητήθηκε εκτενώς στο κεφάλαιο 4. Αποτελεί τη βασική πηγή που εισάγει απώλειες σε όλους τους κωδικοποιητές που βασίζονται στον μετασχηματισμό DCT.

Η κβάντιση ορίζεται ως η διαίρεση κάθε συντελεστή DCT με τον αντίστοιχο αριθμό που αποτελεί το μέγεθος κβάντισης και στη συνέχεια η στρογγυλοποίηση του αποτελέσματος στον πλησιέστερο ακέραιο:

$$F_{ij}^{Q} = Integer\ Round\left(\frac{F_{ij}}{Q_{ij}}\right) \tag{6.3}$$

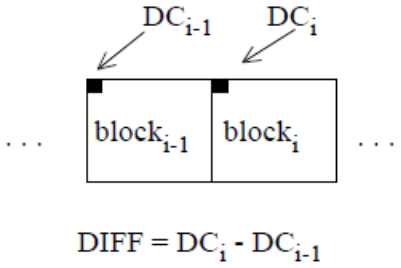
Η τιμή της εξόδου κανονικοποιείται από το μέγεθος κβάντισης. Η αποκβάντιση είναι η αντίστροφη συνάρτηση η οποία σε αυτή την περίπτωση σημαίνει ότι αναιρείται η κανονικοποίηση πολλαπλασιάζοντας με το μέγεθος κβάντησης. Η διαδικασία αυτή επιστρέφει ως αποτέλεσμα μια κατάλληλη αναπαράσταση των συντελεστώ για είσοδο τους στον αντίστροφο μετασχηματισμό DCT.

$$F_{ij}^{Q'} = F_{ij}^{Q} * Q_{ij} \tag{6.4}$$

Όταν στόχος είναι η μέγιστη συμπίεση της εικόνας με τις ελάχιστες οπτικές απώλειες, κάθε μέγεθος κβάντισης θα έπρεπε ιδανικά να επιλεγεί ως το κατώφλι εκείνο όπου η διαφορά στην ποιότητα της εικόνας μόλις που αρχίζει και γίνεται εμφανής για μια συγκεκριμένη συνημιτονοειδή συνάρτηση βάσης. Τα κατώφλια αυτά αποτελούν επίσης συναρτήσεις των χαρακτηριστικών της πηγαίας εικόνας, των χαρακτηριστικών προβολής καθώς και της απόστασης λήψης του αντικειμένου. Για τις εφαρμογές όπου αυτά τα κατώφλια μπορούν να οριστούν αρκετά καλά, ψυχο-οπτικά πειράματα έχουν διεξαχθεί προκειμένου να καθοριστούν οι ιδανικές τιμές κατωφλίων. Τα πειράματα αυτά οδήγησαν στη δημιουργία κάποιων πινάκων κβάντισης οι οποίοι όμως εμφανίζονται στο πρότυπο μόνο με πληροφοριακό χαρακτήρα και δεν απαιτείται η χρήση τους σύμφωνα με το πρότυπο JPEG.

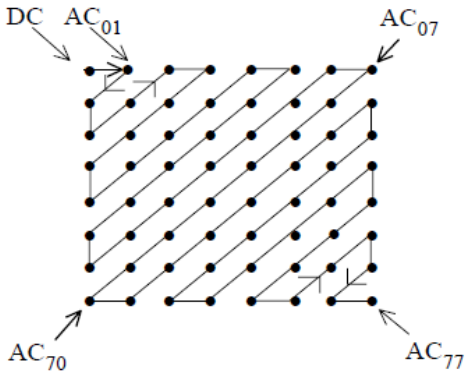
6.1.3 Κωδικοποίηση της DC Συνιστώσας και Σάρωση zig-zag

Μετά το βήμα της κβάντισης, η DC συνιστώσα διαχειρίζεται διαφορετικά από τις υπόλοιπες 63 AC συνιστώσες. Ο συντελεστής DC ουσιαστικά αποτελεί μια μετρική της μέσης τιμής των 64 δειγμάτων της εικόνας τα οποία βρίσκονται στο εκάστοτε μπλοκ το οποίο δίνεται ως είσοδος στον κωδικοποιητή. Επειδή συνήθως υπάρχει ισχυρή συσχέτιση ανάμεσα στους DC συντελεστές των γειτονικών 8×8 μπλοκ, ο κβαντισμένος DC συντελεστής κωδικοποιείται ως η διαφορά του με τον προηγούμενο, κατά σειρά σάρωσης, συντελεστή DC όπως αυτό εμφανίζεται στο Σχήμα 6.3.



Σχήμα 6.3 : Διαφορική Κωδικοποίηση των Συντελεστών DC

Αυτή η ειδική μεταχείριση συνήθως αποδίδει σημαντικά αφού οι συντελεστές DC όπως ήδη γνωρίζουμε συχνά περιέχουν το μεγαλύτερος μέρος της ενέργειας του σήματος της εικόνας. Τέλος όλοι οι κβαντισμένοι συντελεστές ταξινομούνται σε μια ακολουθία αφού σαρωθούν με το σχήμα zig-zag. Η σάρωση αυτή βοηθά τον κωδικοποιητή εντροπίας τοποθετώντας τους συντελεστές των χαμηλών συχνοτήτων (οι οποίοι συνήθως είναι μη μηδενικοί) πριν τους συντελεστές υψηλών συχνοτήτων (μεγάλο μέρος των οποίων συνήθως είναι μηδέν).



Σχήμα 6.4 : Ακολουθία Σάρωσης Zig-Zag

6.1.4 Κωδικοποίηση Εντροπίας

Το τελευταίο βήμα επεξεργασίας ενός κωδικοποιητή που είναι βασισμένος στον μετασχηματισμό DCT είναι η κωδικοποίηση εντροπίας. Το βήμα αυτό επιτυγχάνει επιπλέον συμπίεση χωρίς απώλειες κωδικοποιώντας τους κβαντισμένους συντελεστές DCT σε μια πιο συμπυκνωμένη μορφή και η οποία επιτυγχάνεται εκμεταλλευόμενη τα στατιστικά χαρακτηριστικά των κβαντισμένων συντελεστών. Το πρότυπο JPEG καθορίζει δύο μεθόδους κωδικοποίησης εντροπίας. Τη μέθοδο Huffman και τη μέθοδο αριθμητικής κωδικοποίηση, μέθοδοι που συζητήθηκαν στο κεφάλαιο 3. Ο Βασικός ακολουθιακός κωδικοποιητής χρησιμοποιεί τη μέθοδο Huffman αλλά κωδικό-αποκωδικοποιητές και με τις δύο μεθόδους κωδικοποίησης προσδιορίζονται για όλους τους τρόπους λειτουργίας στο JPEG.

Είναι χρήσιμο να σκεφτούμε την κωδικοποίηση εντροπίας ως μια διαδικασία δύο βημάτων. Το πρώτο βήμα μετατρέπει την ακολουθία zig-zag των κβαντισμένων συντελεστών σε μια ενδιάμεση ακολουθία συμβόλων. Το δεύτερο βήμα μετατρέπει τα σύμβολα αυτά σε μια ροή δεδομένων στην οποία τα σύμβολα πλέον δεν έχουν κάποια εξωτερικά αναγνωρίσιμα όρια. Ο τύπος και ο ορισμός των ενδιάμεσων συμβόλων εξαρτάται τόσο από τον τρόπο λειτουργίας του DCT κωδικοποιητή όσο και από τη μέθοδο κωδικοποίησης εντροπίας.

Η κωδικοποίηση Huffman απαιτεί ένα ή περισσότερα σύνολα πινάκων κώδικα Huffman τα οποία καθορίζονται από την εφαρμογή. Οι ίδιοι πίνακες που χρησιμοποιούνται για τη συμπίεση της εικόνας, χρησιμοποιούνται και για την αποσυμπίεση της. Οι πίνακες Huffman μπορεί να είναι είτε προκαθορισμένοι και να χρησιμοποιούνται ως προεπιλεγμένοι από την εφαρμογή είτε να υπολογίζονται αποκλειστικά για μια δοθείσα εικόνα αφού προηγηθεί ένα αρχικό πέρασμα προκειμένου να συγκεντρωθούν τα απαραίτητα στατιστικά πριν τη συμπίεση. Αυτές οι αποφάσεις αφορούν τις εφαρμογές που χρησιμοποιούν το πρότυπο JPEG και δεν καθορίζεται κανένας πίνακας Huffman από το ίδιο το πρότυπο. Η περιγραφή της κωδικοποίησης εντροπίας με τη μέθοδο Huffman στον Βασικό Ακολουθιακό κωδικοποιητή θα περιγραφεί σε επόμενη ενότητα.

Αντίθετα, η μέθοδος αριθμητικής κωδικοποίησης, όπως προσδιορίζει το πρότυπο JPEG, δεν απαιτεί την ύπαρξη κανενός πίνακα που θα δοθεί εξωτερικά ως είσοδος στον κωδικοποιητή αφού έχει τη δυνατότητα να προσαρμόζεται στα στατιστικά της εικόνας καθώς κωδικοποιείται η εικόνα. (Υπάρχει η δυνατότητα να δοθούν στην είσοδο του κωδικοποιητή κάποιοι πίνακες της στατιστικής κατάστασης της εικόνας για ελαφρώς καλύτερα αποτελέσματα αλλά δεν υπάρχει καμία τέτοια απαίτηση.) Η αριθμητική

κωδικοποίηση προσφέρει 5-10% καλύτερη συμπίεση σε σχέση με την κωδικοποίηση με τη μέθοδο του Huffman για ένα μεγάλο πλήθος εικόνων που δοκιμάστηκαν. Παρόλα αυτά, η συγκεκριμένη μέθοδος κρίθηκε δυσκολότερη στην υλοποίηση της, ιδίως σε υλοποιήσεις με υλικό όπου απαιτείται υψηλή ταχύτητα, σε σχέση με τον κωδικοποιητή Huffman και για το λόγο αυτό υιοθετήθηκε από τον Βασικό Ακολουθιακό κωδικοποιητή JPEG. Εάν η μοναδική διαφορά ανάμεσα σε δύο κωδικο-αποκωδικοποιητές JPEG ήταν η μέθοδος εντροπίας τότε η δια-κωδικοποίηση μεταξύ των δύο αυτών μεθόδων θα μπορούσε να επιτευχθεί αποκωδικοποιώντας με τη μια μέθοδο και επανα-κωδικοποιώντας με τη δεύτερη.

6.2 Συμπίεση και Ποιότητα Εικόνας

Σε έγχρωμες εικόνες με σκηνές μέτριας πολυπλοκότητας, όλοι οι τρόποι λειτουργίας του JPEG που βασίζονται στον μετασχηματισμό DCT επιτυγχάνουν τα ακόλουθα επίπεδα ποιότητας της εικόνας για τα αναφερόμενα εύρη συμπίεσης. Τα επίπεδα αυτά αποτελούν απλώς έναν οδηγό. Η ποιότητα της εικόνας και η συμπίεση που μπορεί να επιτευχθεί ενδέχεται να διαφέρει σημαντικά από εικόνα σε εικόνα αναλόγως των χαρακτηριστικών της και των περιεχομένων της σκηνής. Οι τιμές που ακολουθούν αναφέρονται σε bit/pixel, δηλαδή το πλήθος των bits της συμπιεσμένης εικόνας διαιρεμένα με το πλήθος των εικονοστοιχείων της.

- 0.25-0.5 bit/pixel: Μέτρια προς καλή ποιότητα, ικανοποιητική για κάποιες εφαρμογές
- 0.5-0.75 bit/pixel: Καλή προς πολύ καλή ποιότητα, ικανοποιητική για αρκετές εφαρμογές
- 0.7-0.75 bit/pixel: Άριστη ποιότητα, ικανοποιητική για τις περισσότερες εφαρμογές.
- 1,5-2.0 bit/pixel: Συνήθως είναι αδύνατο να ξεχωρίσει κανείς τη διαφορά της από την πρωτότυπη εικόνα, ικανοποιητική για τις περισσότερες απαιτητικές εφαρμογές.

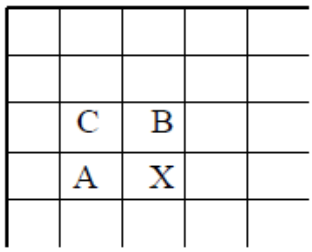
6.3 Βήματα Επεξεργασίας στη Συμπίεση με Πρόγνωση για Συμπίεση χωρίς Απώλειες

Το πρότυπο JPEG προκειμένου να πληροί τις απαιτήσεις της συμπίεσης χωρίς απώλειες επέλεξε να υιοθετήσει μια απλή μέθοδο πρόγνωσης η οποία είναι εντελώς ανεξάρτητη από την επεξεργασία με χρήση του μετασχηματισμού DCT που περιγράφηκε στην προηγούμενη ενότητα. Η επιλογή της συγκεκριμένης μεθόδου δεν ήταν αποτέλεσμα μιας αυστηρά ανταγωνιστικής αξιολόγησης όπως στην περίπτωση της προηγούμενης μεθόδου με απώλειες. Παρόλα αυτά, η συμπίεση JPEG χωρίς

απώλειες επιτυγχάνει αποτελέσματα τα οποία υπό τη σκοπιά της απλότητας της είναι εξαιρετικά κοντά στα καλύτερα δυνατά αποτελέσματα που μπορούν να επιτευχθούν για συμπίεση εικόνας χωρίς απώλειες. Στο Σχήμα 6.5 παρουσιάζονται τα βασικά βήματα επεξεργασίας της εικόνας.

Σχήμα 6.5 : Μπλοκ Διάγραμμα Κωδικοποιητή JPEG Χωρίς Απώλειες

Ένας προγνώστης συνδυάζει τις τιμές μέχρι και τριών γειτονικών δειγμάτων (A,B και C) προκειμένου να σχηματιστεί μια πρόγνωση του δείγματος X το οποίο επιθυμούμε να κωδικοποιήσουμε όπως παρουσιάζεται στο Σχήμα 6.6.



Σχήμα 6.6 : Πρόγνωση του Δείγματος X από τα Γειτονικά Δείγματα A, B και C

Η πρόγνωση αυτή στη συνέχεια αφαιρείται από την πραγματική τιμή του δείγματος X και η διαφορά τους είναι αυτή που τελικά κωδικοποιείται χωρίς απώλειες από τις μεθόδους κωδικοποίησης εντροπίας (Huffman ή αριθμητική κωδικοποίηση). Σύμφωνα με τα όσα παρουσιάζει ο Πίνακας 6.1 οι επιλογές 1 ,2 και 3 σχηματίζουν πρόγνωση μίας μόνο διάστασης ενώ οι 4, 5, 6 και 7 σχηματίζουν πρόγνωση από δείγματα που βρίσκονται στις δύο διαστάσεις της εικόνας. Ο κωδικοποιητής εντροπίας στη συνέχεια είναι αντίστοιχος με αυτόν που χρησιμοποιήθηκε στην περίπτωση κωδικοποίησης των συντελεστών DC με απώλειες (Huffman, αριθμητική κωδικοποίηση). Τα αποτελέσματα της συμπίεσης χωρίς απώλειες μειώνουν το μέγεθος δεδομένων της εικόνας περίπου κατά ½ σε έγχρωμες εικόνες με σκηνές μέτριας πολυπλοκότητας.

Πίνακας 6.1 : Τρόπος Υπολογισμού της Πρόγνωσης Ανάλογα με την Τιμή Επιλογής

Τιμή Επιλογής	Πρόγνωση
0	Καμία πρόγνωση
1	A
2	B
3	C
4	A+B-C
5	A+((B-C)/2)
6	B+((A-C)/2)
7	(A+B)/2

6.4 Στάδια Κωδικοποίησης Εντροπίας

Όπως αναφέρθηκε σε προηγούμενη ενότητα τα στάδια του κωδικοποιητή εντροπίας είναι δύο. Αρχικά μετατρέπονται οι συντελεστές DCT σε μια ακολουθία συμβόλων ενδιάμεσης αναπαράστασης και στη συνέχεια τα σύμβολα αυτά κωδικοποιούνται με έναν κώδικα μεταβλητού μήκους.

6.4.1 Ενδιάμεση Αναπαράσταση Συμβόλων

Κατά την ενδιάμεση αναπαράσταση των συμβόλων κάθε μη μηδενικός συντελεστής AC αναπαρίσταται σε συνδυασμό με το τρέχον μήκος (“runlength” coding) των μηδενικών συντελεστών AC που προηγούνται του μη μηδενικού συντελεστή κατά το σάρωμα τους με το σχήμα zig-zag. Επομένως ο συνδυασμός του τρέχοντος μήκους μηδενικών των AC συντελεστών (RUNLENGTH) με την τιμή του μη μηδενικού AC συντελεστή αναπαρίστανται ως το ζεύγος των συμβόλων:

Σύμβολο-1 (RUNLENGTH, SIZE)	Σύμβολο-2 (AMPLITUDE)

Όπου το Σύμβολο-1 αναπαριστά δύο τμήματα πληροφορίας, το RUNLENGTH και το

SIZE το οποίο είναι το πλήθος των bit που απαιτείται για την κωδικοποίηση του Συμβόλου-2. Το Σύμβολο-2 δεν είναι τίποτα άλλο παρά το μέγεθος του εκάστοτε μη-μηδενικού AC συντελεστή (επομένως προσημασμένο). Το πρώτο τμήμα του Συμβόλου-1, RUNLENGTH, αναπαριστά όπως αναφέραμε το πλήθος των μηδενικών συντελεστών AC που προηγούνται ενός μη μηδενικού συντελεστή AC. Το εύρος τιμών που λαμβάνει είναι από 0 έως 15. Άμεσα γίνεται εμφανές ότι το πλήθος των διαδοχικών μηδενικών ενδέχεται να είναι μεγαλύτερο του 15. Προκειμένου να μπορέσει να υπάρξει επέκταση στο επιθυμητό εύρος, το δεύτερο τμήμα του Συμβούλου-1 το οποίο είναι το μέγεθος των bit που απαιτείται για την κωδικοποίηση του Συμβόλου-2 λαμβάνει την τιμή μηδέν. Αυτό σημαίνει ότι μετά το πρώτο Σύμβολο-1 θα ακολουθήσει και δεύτερο στη σειρά σύμβολο-1. Οπότε για παράδειγμα εάν το πλήθος το μηδενικών ήταν 20 αυτό θα μπορούσε να αναπαρασταθεί ως (15,0) (5,X) όπου X το πλήθος των bit που θα κωδικοποιήσουν τον επόμενο μη μηδενικό συντελεστή του οποίου το πλάτος θα ακολουθήσει ως Σύμβολο-2. Θα πρέπει να σημειωθεί ότι μπορούν να υπάρξουν μέχρι τρείς διαδοχικές επεκτάσεις πριν το τερματικό σύμβολο. Όταν κωδικοποιηθεί και ο τελευταίος μη μηδενικός συντελεστής AC τότε ακολουθεί ένα ειδικό Σύμβολο-1 με τιμή (0, 0) το οποίο υποδεικνύει τον τερματισμό του μπλοκ (End Of Block) και μπορεί να θεωρηθεί ως χαρακτήρας διαφυγής του τρέχοντος μπλοκ.

Συνοψίζοντας, για κάθε μπλοκ δειγμάτων 8×8, η ακολουθία που προκύπτει από το σάρωμα zig-zag των 63 κβαντισμένων AC συντελεστών αναπαριστάται ως μια ακολουθία ζευγών Συμβόλων-1 και Συμβόλων-2 όπου σε κάθε ζεύγος μπορεί να περιλαμβάνονται περισσότερα του ενός Συμβόλου-1 εφόσον χρειαστεί επέκταση ή μόνο ένα Σύμβολο-1 (0, 0) στην περίπτωση τερματισμού του μπλοκ.

Το πιθανό εύρος των κβαντισμένων AC συντελεστών καθορίζει το εύρος τιμών της πληροφορίας σχετικά με το μήκος όσο και το μέγεθος τα οποία πρέπει να αναπαρασταθούν. Η αριθμητική ανάλυση της εξίσωσης του 8×8 μετασχηματισμού DCT δείχνει ότι εάν το σήμα εισόδου 64 σημείων περιέχει ακεραίους των N-bit, τότε η έξοδος του μετασχηματισμού DCT (δηλαδή οι συντελεστές) ενδέχεται να είναι μεγαλύτεροι το πολύ κατά 3 bit. Αυτό είναι επίσης και το μεγαλύτερο μέγεθος ενός κβαντισμένου συντελεστή DCT όταν το μέγεθος του βήματος κβάντισης είναι μονάδα.

Ο Βασικός Ακολυθιακός Κωδικοποιητής έχει ως δείγματα εισόδου 8 bit ακεραίους στο εύρος $[-2^7, 2^7 - 1]$, επομένως τα μεγέθη των κβαντισμένων AC συντελεστών καλύπτονται από ακεραίους στο εύρος $[-2^{10}, 2^{10} - 1]$. Η κωδικοποίηση με προσημασμένους ακεραίους χρησιμοποιεί Σύμβολα-2 με κωδικούς για τις τιμές, μήκους

από 1 έως 10 bit (συνεπώς και το SIZE αναπαρίσταται επίσης με τιμές από 1 έως 10) και το RUNLENGTH αναπαρίσταται με τιμές από 0 έως 15 όπως εξηγήσαμε νωρίτερα. Για τους συντελεστές AC, τη δομή των ενδιάμεσων συμβόλων 1 και 2 παρουσιάζει ο Πίνακας 6.2.

Πίνακας 6.2 : (a) Δομή του Συμβόλου-1 κατά την Κωδικοποίηση Huffman στον Βασικό Κωδικοποιητή. (b) Δομή του Συμβόλου-2 κατά την Κωδικοποίηση στον Βασικό Κωδικοποιητή

	SIZE					
	0	1	2	. . .	9	10
RUN LENGTH	0 EOB					
	. X					
	. X		RUN-SIZE values			
	. X					
	15 ZRL					

(a)

SIZE	AMPLITUDE
1	-1,1
2	-3,-2,2,3
3	-7..-4,4..7
4	-15..-8,8..15
5	-31..-16,16..31
6	-63..-32,32..63
7	-127..-64,64..127
8	-255..-128,128..255
9	-511..-256,256..511
10	-1023..-512,512..1023

(b)

Η ενδιάμεση αναπαράσταση του διαφορικού DC συντελεστή ενός μπλοκ δειγμάτων 8×8 δομείται με αντίστοιχο τρόπο. Όμως το Σύμβολο-1 αναπαριστά μόνο την πληροφορία σχετικά με το SIZE και το Σύμβολο-2 ασφαλώς την τιμή του AC συντελεστή. Επειδή ο συντελεστής DC είναι κωδικοποιημένος διαφορικά, καλύπτεται από τις διπλάσιες ακέραιες τιμές, $[-2^{11}, 2^{11}-1]$, σε σχέση με αυτές που απαιτούνται για τους AC συντελεστές, επομένως μια επιπλέον στάθμη πρέπει να προστεθεί στο τέλος των τιμών που παρουσιάζει ο Πίνακας 6.2 b για τους DC συντελεστές. Συνεπώς το Σύμβολο-1 για τους DC συντελεστές λαμβάνει τιμές από 1 έως 11.

6.4.2 Κωδικοποίηση Εντροπίας Μεταβλητού Μήκους

Μετά την κβάντιση των συντελεστών ενός μπλοκ δεδομένων 8×8 ακολουθεί η ενδιάμεση αναπαράσταση τους όπως αυτή περιγράφηκε στην προηγούμενη ενότητα. Το τελευταίο βήμα είναι απόδοση, στα σύμβολα αυτά, κωδικών μεταβλητού μήκους. Για κάθε μπλοκ 8×8, η αναπαράσταση των Συμβόλων 1 και 2 της DC συνιστώσας κωδικοποιούνται και παράγονται πρώτοι.

Τόσο για τους DC όσο και για τους AC συντελεστές, κάθε Σύμβολο-1 κωδικοποιείται με κώδικα μεταβλητού μήκους (VLC) από το σύνολο πινάκων Huffman το οποίο έχει ανατεθεί στο 8×8 μπλοκ της εικόνας. Κάθε Σύμβολο-2 κωδικοποιείται με έναν ακέραιο

μεταβλητού μήκους (VLI) κώδικα του όποιου το μήκος σε bit παρουσιάζει ο Πίνακας 6.2 b. Τόσο οι VLIs όσο και οι VLCs είναι κώδικες μεταβλητού μήκους όμως οι VLIs δεν είναι κώδικες Huffman. Μια σημαντική διαφοροποίηση βρίσκεται στο γεγονός ότι το μήκος του VLC (κώδικα Huffman) δεν είναι γνωστό μέχρι τη στιγμή αποκωδικοποίησης του αλλά το μήκος του VLI βρίσκεται αποθηκευμένο στον προηγούμενο VLC.

6.5 Παράδειγμα στον Βασικό Ακολουθιακό Κωδικοποιητή

Σε αυτό το σημείο θα ακολουθήσει ένα παράδειγμα του Βασικού Συμπιεστή προκειμένου να κωδικοποιηθεί ένα μπλοκ δειγμάτων 8×8. Ο Πίνακας 6.3 παρουσιάζει ένα μπλοκ 8×8 με δείγματα των 8 bit από την πηγαία εικόνα. Οι μικρές μεταβολές της φωτεινότητας από δείγμα σε δείγμα φανερώνουν την επικράτηση των χαμηλών χωρικών συχνοτήτων.

Πίνακας 6.3 : Δείγματα Πηγαίας Εικόνας

139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

Αφού αφαιρέσουμε την τιμή 128 από κάθε δείγμα για την απαιτούμενη ολίσθηση της στάθμης, το 8×8 μπλοκ δίνεται ως είσοδος στη συνάρτηση μετασχηματισμού του DCT. Ο Πίνακας 6.4 εμφανίζει το αποτέλεσμα του μετασχηματισμού, δηλαδή τους συντελεστές DCT. Εκτός από μερικούς από τους συντελεστές χαμηλών συχνοτήτων, τα πλάτη των υπολοίπων είναι αρκετά μικρά.

Πίνακας 6.4 : Οι Συντελεστές DCT που Προέκυψαν από το Μετασχηματισμό της Πηγαίας Εικόνας

235.6	-1	-12.1	-5.2	2.1	-1.7	-2.7	1.3
-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2
-10.9	-9.3	-1.6	1.5	-0.2	-0.9	-0.6	-0.1
-7.1	-1.9	0.2	1.5	0.9	-0.1	0	0.3
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
1.8	-0.2	1.6	-0.3	-0.8	1.5	1	-1
-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

Ο Πίνακας 6.5 αποτελεί τον πίνακα κβάντισης για τα στοιχεία φωτεινότητας (αποχρώσεις του γκρι) ο οποίος προτείνεται μέσα στο πρότυπο JPEG.

Πίνακας 6.5 : Προτεινόμενος Πίνακας Κβάντισης

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Ο Πίνακας 6.6 εμφανίζει τους κβαντισμένους συντελεστές DCT, κανονικοποιημένους

βάσει των τιμών που παρουσιάζει ο Πίνακας 6.5.

Πίνακας 6.6 : Κανονικοποιημένοι Συντελεστές Κβάντισης

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Στον αποκωδικοποιητή αυτοί οι αριθμοί αποκανονικοποιούται αφού πολλαπλασιαστούν πάλι με τις τιμές που περιέχει ο Πίνακας 6.5 και δίνονται ως είσοδος στον αντίστροφο μετασχηματισμό DCT (Πίνακας 6.7).

Πίνακας 6.7 : Αποκανονικοποιημένοι Συντελεστές Κβάντισης

240	0	-10	0	0	0	0	0
-24	-12	0	0	0	0	0	0
-14	-13	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Τελικά ο Πίνακας 6.8 παρουσιάζει το ανακατασκευασμένο σήμα με τις τιμές των δειγμάτων του μπλοκ της εικόνας οι οποίες όπως μπορούμε να παρατηρήσουμε είναι

παραπλήσιες με αυτές του μπλοκ της πηγαίας εικόνας (Πίνακας 6.3).

Πίνακας 6.8 : Ανακατασκευασμένα Δείγματα Εικόνας

144	146	149	152	154	156	156	156
148	150	152	154	156	156	156	156
155	156	157	158	158	157	156	155
160	161	161	162	161	159	157	155
163	163	164	163	162	160	158	156
163	164	164	164	162	160	158	157
160	161	162	162	162	161	159	158
158	159	161	161	162	161	159	158

Οι κανονικοποιημένοι συντελεστές (Πίνακας 6.6) θα πρέπει στη συνέχεια να κωδικοποιηθούν με κώδικα Huffman πριν την αποστολή τους στον αποκωδικοποιητή. Ο πρώτος αριθμός που θα πρέπει να κωδικοποιηθεί είναι αυτός της DC συνιστώσας και ο οποίος θα κωδικοποιηθεί διαφορικά. Εάν υποθέσουμε ότι η κβαντισμένη DC συνιστώσα του προηγούμενου μπλοκ είχε για παράδειγμα τιμή 12, τότε η διαφορά της τρέχουσας DC συνιστώσας από την προηγούμενη είναι $+3$. Συνεπώς η ενδιάμεση αναπαράσταση θα είναι (2)(3), δηλαδή $SIZE = 2$ και τιμή συντελεστή $= 3$.

Στη συνέχεια κωδικοποιείται οι κβαντισμένοι AC συντελεστές. Εάν ακολουθηθεί η σάρωση των συντελεστών με το σχήμα zig-zag, τότε ο πρώτος μη μηδενικός συντελεστής έχει τιμή -2 και πριν από αυτόν υπάρχει ένας μηδενικός συντελεστής. Αυτό σημαίνει ότι η ενδιάμεση αναπαράσταση θα είναι (1,2)(-2). Στη συνέχεια υπάρχουν τρείς διαδοχικοί μη-μηδενικοί συντελεστές, πλάτους -1. Αυτό σημαίνει ότι κάθε ένας από αυτούς δεν έχει κανένα μηδενικό που να προηγείται, συνεπώς η ενδιάμεση αναπαράσταση θα είναι (0,1)(-1). Ο τελευταίος μη μηδενικός συντελεστής έχει τιμή -1 και προηγούνται αυτού δύο μηδενικά. Επομένως η ενδιάμεση αναπαράσταση θα είναι (2,1)(-1). Επειδή αυτός είναι και ο τελευταίος μη μηδενικός συντελεστής, το τελευταίο σύμβολο θα σηματοδοτήσει το τέλος του 8×8 μπλοκ με την ειδική τιμή (0,0).

Άρα, η ακολουθία με τα ενδιάμεσα σύμβολα για το 8×8 μπλοκ του παραδείγματος θα

είναι:

$$(2)(3),(1,2)(-2),(0,1)(-1),(0,1)(-1),(0,1)(-1),(2,1)(-1),(0,0)$$

Το τελευταίο βήμα είναι η ανάθεση των κωδικών σε κάθε σύμβολο. Στο παράδειγμα αυτό, χρησιμοποιούνται οι VLCs (κώδικες Huffman) από το πληροφοριακό παράρτημα του προτύπου JPEG. Η τιμή της κωδικοποιημένης διαφορικής DC συνιστώσας για το παράδειγμα αυτό είναι:

$$(2) => 011$$

Οι συντελεστές AC κωδικοποιημένοι με VLC, για το συγκεκριμένο παράδειγμα είναι

$$(0,0) => 1010$$

$$(0,1) => 00$$

$$(1,2) => 11011$$

$$(2,1) => 11100$$

Οι VLIs που καθορίζονται στο ίδιο παράρτημα αναπαρίστανται με το συμπλήρωμα ως προς δύο και είναι:

$$(3) => 11$$

$$(-2) => 01$$

$$(-1) => 0$$

Τελικά, η ακολουθία από bit που θα χρησιμοποιηθεί για την κωδικοποίηση του συγκεκριμένου μπλοκ 8×8 είναι η ακόλουθη: 0111111011010000000001110001010

Παρατηρούμε ότι απαιτούνται μόλις 31bit για να αναπαραστήσουν τους 64 συντελεστές το οποίο πρακτικά σημαίνει ότι επιτυγχάνεται συμπίεση με λιγότερο από 0.5bit/δείγμα.

7. Μετασχηματισμός Κυματιδίου

7.1 Ο Συνεχής Μετασχηματισμός Κυματιδίου και ο Αντίστροφος του

Ο συνεχής μετασχηματισμός κυματιδίου[45] μιας συνάρτησης $f(t)$ εμπλέκει το μητρικό κυματίδιο $\psi(t)$. Το μητρικό κυματίδιο μπορεί να είναι οποιαδήποτε συνεχής συνάρτηση, πραγματική ή φανταστική, η οποία ικανοποιεί τις δύο ακόλουθες ιδιότητες:

1. Η συνολική επιφάνεια κάτω από την καμπύλη της συνάρτησης να είναι μηδέν, δηλαδή:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0$$

2. Η συνολική επιφάνεια της $|\psi(t)|^2$ να είναι πεπερασμένη

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$

Η συνθήκη αυτή ασφαλώς συνεπάγεται ότι το ολοκλήρωμα του τετραγώνου του κυματιδίου πρέπει να υπάρχει.

Η πρώτη ιδιότητα συνιστά μια συνάρτηση η οποία ταλαντώνεται άνω και κάτω του άξονα του χρόνου. Μια συνάρτηση αυτής της μορφής τείνει να είναι κυματοειδής. Η δεύτερη ιδιότητα συνεπάγεται ότι η ενέργεια της συνάρτησης αυτής είναι πεπερασμένη και συνεπώς η συνάρτηση αυτή περιορίζεται σε ένα πεπερασμένο διάστημα και είναι μηδέν ή σχεδόν μηδέν έξω από το διάστημα αυτό. Άπειρος αριθμός συναρτήσεων ικανοποιεί τις συνθήκες αυτές και κάποιες από αυτές έχουν μελετηθεί και χρησιμοποιηθεί ευρύτατα σε μετασχηματισμούς κυματιδίων. Στο Σχήμα 7.1 a παρουσιάζεται το κυματίδιο Morlet το οποίο ορίζεται ως:

$$\psi(t) = e^{-t^2} \cos\left(\pi t \sqrt{\frac{2}{ln2}}\right) \approx e^{-t^2} \cos(2.885\pi t)$$

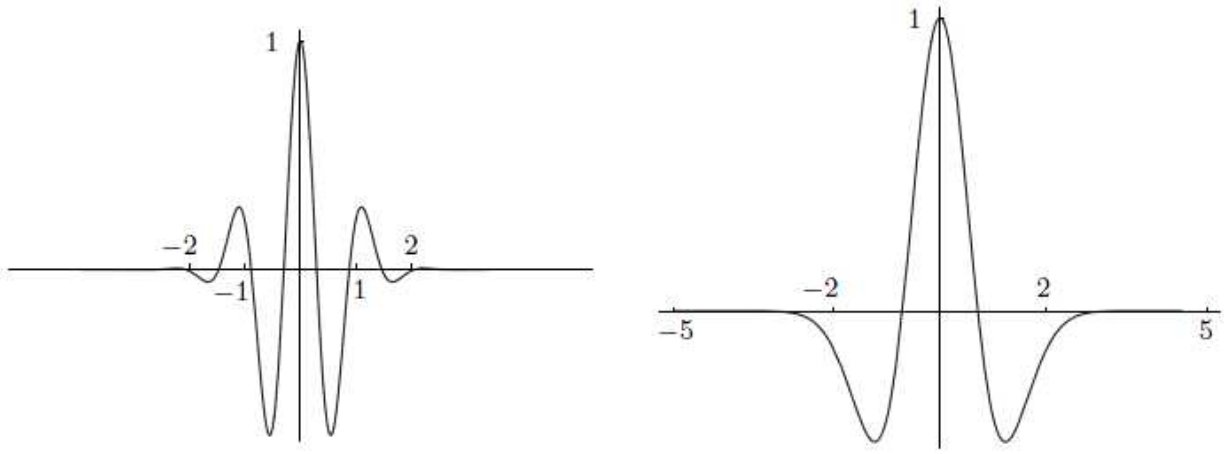
Αυτή είναι μια συνημιτονοειδής καμπύλη της οποίας οι ταλαντώσεις επηρεάζονται από έναν εκθετικό ή Gaussian παράγοντα. Περισσότερο από το 99% τις ενέργειας της είναι συγκεντρωμένη στο διάστημα $-2.5 \leq t \leq 2.5$. Το Σχήμα 7.1 παρουσιάζει το κυματίδιο γνωστό ως Mexican hat το οποίο ορίζεται ως:

$$\psi(t) = (1 - 2t^2)e^{-t^2}$$

Αυτή είναι και η δεύτερη παράγωγος της (αρνητικής) Gaussian συνάρτησης $-0.5e^{-t^2}$. Από τη στιγμή που ένα κυματίδιο $\psi(t)$ έχει επιλεγεί, ο συνεχής

μετασχηματισμός κυματιδίου μιας τετραγωνικά ολοκληρώσιμης συνάρτησης $f(t)$ δίνεται από τη σχέση:

$$W(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a}\right) dt \tag{7.1}$$



(a) (b)

Σχήμα 7.1 : (a) Το Κυματίδιο Τύπου Morlet. (b) Το Κυματίδιο Τύπου Mecixan Hat

Ο μετασχηματισμός είναι μια συνάρτηση των δύο πραγματικών παραμέτρων α και b. Το σύμβολο * υποδηλώνει το μιγαδικό συζυγή. Εάν ορίσουμε μια συνάρτηση

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t-b}{a}\right)$$

μπορούμε να γράψουμε τη συνάρτηση (7.1) ως

$$W(\alpha,b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt \tag{7.2}$$

Από μαθηματικής άποψης, ο μετασχηματισμός αυτός είναι το εσωτερικό γινόμενο των δυο συναρτήσεων $f(t)$ και $\psi_{a,b}(t)$. Η ποσότητα $1/\sqrt{|\alpha|}$ είναι ένας παράγοντας κανονικοποίησης ο οποίος διασφαλίζει ότι η ενέργεια της $\psi(\tau)$ παραμένει ανεξάρτητη των α και b. Για παράδειγμα

$$\int_{-\infty}^{\infty} \left|\psi_{a,b}(t)\right|^2 dt = \int_{-\infty}^{\infty} |\psi(t)|^2 \, dt$$

Για κάθε α, η $\psi_{a,b}(t)$ είναι ένα αντίγραφο της $\psi_{a,0}$ μετατοπισμένη b μονάδες κατά μήκος του άξονα του χρόνου. Συνεπώς το b είναι μια παράμετρος μετατόπισης. Θέτοντας το $b = 0$ έχουμε ότι:

$$\psi_{a,0}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t}{a}\right)$$

το οποίο συνεπάγεται ότι το α είναι μια παράμετρος κλιμάκωσης (ή διαστολής). Τιμές του $\alpha > 1$ διαστέλλουν το κυματίδιο ενώ τιμές $0 < a < 1$ το συστέλλουν.

Δεδομένου ότι τα κυματίδια χρησιμοποιούνται για να μετασχηματίσουν μια συνάρτηση, απαραίτητος είναι και ο αντίστροφος μετασχηματισμός τους. Συμβολίζουμε με $\Psi(\omega)$ τον μετασχηματισμό Fourier της συνάρτησης $\psi(t)$:

$$\Psi(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-i\omega t} dt$$

Εάν το $W(a,b)$ είναι ο συνεχής μετασχηματισμός κυματιδίου μιας συνάρτησης $f(t)$ με ένα κυματίδιο $\psi(t)$, τότε ο αντίστροφος συνεχής μετασχηματισμός κυματιδίου ορίζεται ως:

$$f(t) = \frac{1}{C}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{|a|^2} W(a,b)\psi_{a,b}(t)\, da\, db, \tag{7.3}$$

όπου το C ορίζεται ως

$$C = \int_{\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega$$

Ο αντίστροφος μετασχηματισμός κυματιδίου υπάρχει εφόσον το C είναι θετικό και πεπερασμένο. Δεδομένου ότι το C είναι ορισμένο συναρτήσει του Ψ το οποίο με τη σειρά του είναι ορισμένο συναρτήσει του κυματιδίου $\psi(t)$, η απαίτηση για να είναι το C θετικό και πεπερασμένο επιβάλλει έναν ακόμα περιορισμό γνωστό ως “συνθήκη παραδοχής” της επιλογής του κυματιδίου.

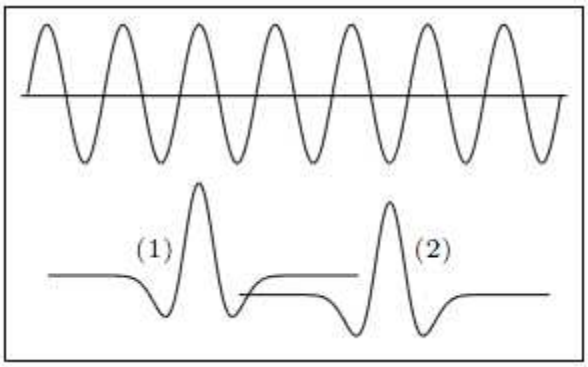
Ο καλύτερος τρόπος για να σκεφτούμε τον συνεχή μετασχηματισμό κυματιδίου είναι ως έναν πίνακα αριθμών ο οποίος προκύπτει από το εσωτερικό γινόμενο των συναρτήσεων $f(t)$ και $\psi_{a,b}(t)$. Οι γραμμές του πίνακα αντιστοιχούν στις τιμές του a και οι στήλες δεικτοδοτούνται από το b. Το εσωτερικό γινόμενο δύο συναρτήσεων $f(t)$ και $g(t)$ ορίζεται ως:

$$\langle f(t), g(t) \rangle = \int_{\infty}^{\infty} f(t) g^*(t)\, dt$$

συνεπώς ο συνεχής μετασχηματισμός κυματιδίου είναι ο

$$\langle f(t), \psi_{a,b}(t) \rangle = \int_{\infty}^{\infty} f(t) \psi_{a,b}(t)\, dt$$

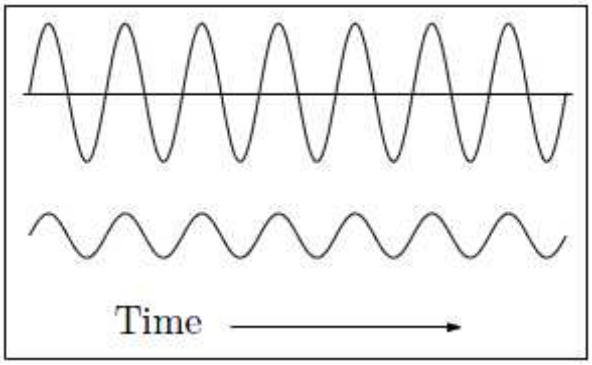
Μετά από την εισαγωγή αυτή, μπορούμε σε αυτό το σημείο να δώσουμε μια πιο διαισθητική ερμηνεία του συνεχούς μετασχηματισμού κυματιδίου ξεκινώντας με ένα απλό παράδειγμα: του συνεχούς μετασχηματισμού κυματιδίου μιας ημιτονοειδούς κυματομορφής όπου ως κυματίδιο έχει επιλεγεί το κυματίδιο τύπου Mexiacan Hat. Το Σχήμα 7.2 παρουσιάζει την ημιτονοειδή κυματομορφή με δύο αντίγραφα του κυματιδίου. Το πρώτο αντίγραφο έχει τοποθετηθεί στο σημείο όπου η ημιτονοειδής κυματομορφή παρουσιάζει μέγιστο.



Σχήμα 7.2 : Ημιτονοειδής Κυματομορφή με Δύο Αντίγραφα του Κυματιδίου Mexican Hat

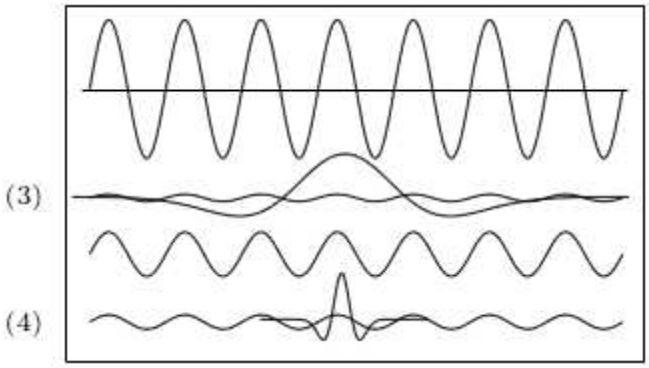
Στο σημείο αυτό παρατηρούμε μια καλή αντιστοιχία ανάμεσα στην αναλυθείσα συνάρτηση (ημίτονο) και το κυματίδιο. Το κυματίδιο αναπαράγει τις ιδιότητες της ημιτονοειδούς κυματομορφής. Αυτό έχει ως αποτέλεσμα το εσωτερικό γινόμενο του ημιτόνου σε συνδυασμό με το κυματίδιο να παράγει έναν μεγάλο θετικό αριθμό. Αντίθετα, το δεύτερο αντίγραφο έχει τοποθετηθεί στο σημείο όπου η ημιτονοειδής κυματομορφή παρουσιάζει ελάχιστο. Παρατηρούμε ότι στο σημείο αυτό το ημίτονο και το κυματίδιο είναι σχεδόν αντίθετα (αντικατοπτρισμός). Δηλαδή στα σημεία που το ημίτονο είναι θετικό, το κυματίδιο είναι αρνητικό και αντίστροφα. Το γινόμενο αυτή τη φορά είναι αρνητικό και ο μετασχηματισμός κυματιδίου (το εσωτερικό τους γινόμενο) είναι ένας μεγάλος αρνητικός αριθμός. Ανάμεσα σε αυτά τα δύο σημεία, ο μετασχηματισμός κυματιδίου μεταβαίνει από τα θετικά προς το μηδέν και στη συνέχεια στα αρνητικά.

Καθώς το κυματίδιο μετατοπίζεται κατά μήκος του ημιτονοειδούς κύματος, από τα αριστερά προς τα δεξιά, ο συνεχής μετασχηματισμός κυματιδίου μεταβάλλεται από θετικές σε αρνητικές τιμές παράγοντας το μικρό κύμα όπως αυτό φαίνεται στο Σχήμα 7.3. Το σχήμα αυτό είναι αποτέλεσμα της στενής συσχέτισης ανάμεσα στη συνάρτηση που αναλύεται (ημιτονοειδής) και το κυματίδιο που χρησιμοποιείται. Όπως παρατηρούμε έχουν παρόμοιο σχήμα και παρόμοιες συχνότητες.



Σχήμα 7.3 : Παραχθέν Κύμα από τη Μετατόπιση του Κυματιδίου

Μπορούμε πλέον να επεκτείνουμε την ανάλυση, καλύπτοντας και διαφορετικές συχνότητες. Αυτό επιτυγχάνεται με την κλιμάκωση του κυματιδίου. Στο Σχήμα 7.4 παρουσιάζεται το αποτέλεσμα, όταν το κυματίδιο διαστέλλεται ώστε να καλύπτει αρκετές περιόδους της ημιτονοειδούς κυματομορφής.

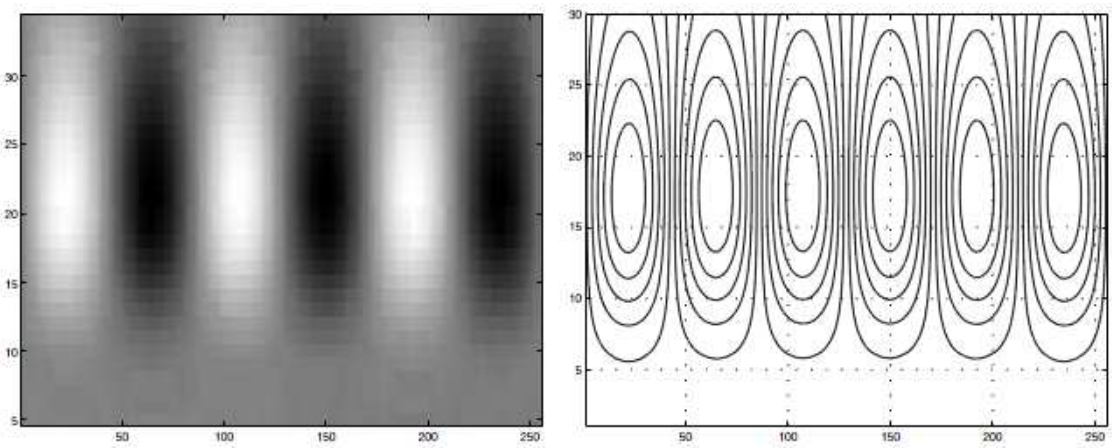


Σχήμα 7.4 : Διαστολή Κυματιδίου

Η μετατόπιση του κυματιδίου από τα αριστερά προς τα δεξιά δεν επηρεάζει ιδιαίτερα την ταύτιση ανάμεσα στο ημίτονο και το κυματίδιο, πράγμα που έχει ως αποτέλεσμα, ο συνεχής μετασχηματισμός κυματιδίου να διακυμαίνεται ελάχιστα. Όσο πιο διεσταλμένο είναι ένα κυματίδιο, τόσο ο μετασχηματισμός κυματιδίου τείνει σε μία σταθερά. Παρατηρούμε πώς το πλάτος του κυματιδίου έχει μειωθεί και ως εκ τούτου η περιοχή

του με αποτέλεσμα να παράγει μία μικρή σταθερά. Ομοίως και στο τέταρτο μέρος του σχήματος όπου το κυματίδιο έχει συρρικνωθεί με αποτέλεσμα να είναι πολύ στενότερο από έναν κύκλο του ημιτόνου. Δεδομένου ότι το κυματίδιο είναι τόσο στενό, το εσωτερικό γινόμενο του με την ημιτονοειδή κυματομορφή έχει πάντα ως αποτέλεσμα έναν μικρό αριθμό, θετικό ή αρνητικό, ανεξαρτήτως της ακρίβειας της θέσης που έχει τοποθετηθεί το κυματίδιο σε σχέση με την ημιτονοειδή κυματομορφή. Παρατηρούμε και πάλι ότι η μετατόπιση του κυματίδιου δεν επηρεάζει ιδιαίτερα την ταύτιση με την ημιτονοειδή κυματομορφή έχοντας ως αποτέλεσμα, ο συνεχής μετασχηματισμός κυματιδίου να τείνει σε μια σταθερά.

Το αποτέλεσμα της μετατόπισης του κυματιδίου, της κλιμάκωσης του και η επανάληψη της ίδιας διαδικασίας ξανά και ξανά εμφανίζεται στο Σχήμα 7.5 a. Αυτή είναι η πυκνότητα της συνάρτησης μετασχηματισμού $W(a,b)$ όπου ο οριζόντιος άξονας αντιστοιχεί στις τιμές του b (μετατόπιση) και ο κατακόρυφος άξονας αντιστοιχεί στις τιμές του α (κλιμάκωση). Το Σχήμα 7.5 b παρουσιάζει το περίγραμμα του προηγούμενου σχήματος. Τα διαγράμματα αυτά δείχνουν ότι υπάρχει πολύ καλή ταύτιση ανάμεσα στη συνάρτηση και το κυματίδιο σε συγκεκριμένη συχνότητα (τη συχνότητα του ημιτονοειδούς κύματος). Σε άλλες συχνότητες η ταύτιση φθίνει με αποτέλεσμα ο μετασχηματισμός να τείνει όλο και περισσότερο σε μια σταθερά.



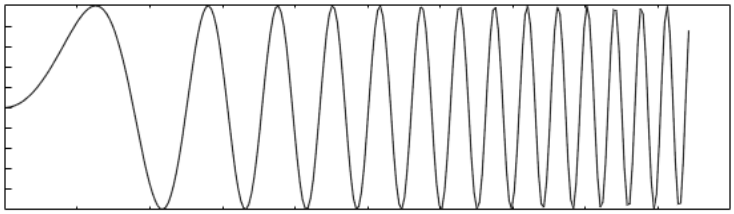
(a) (b)

Σχήμα 7.5 : (a) Ο Συνεχής Μετασχηματισμός Κυματιδίου μίας Ημιτονοειδούς Κυματομορφής και (b) το Περίγραμμα του

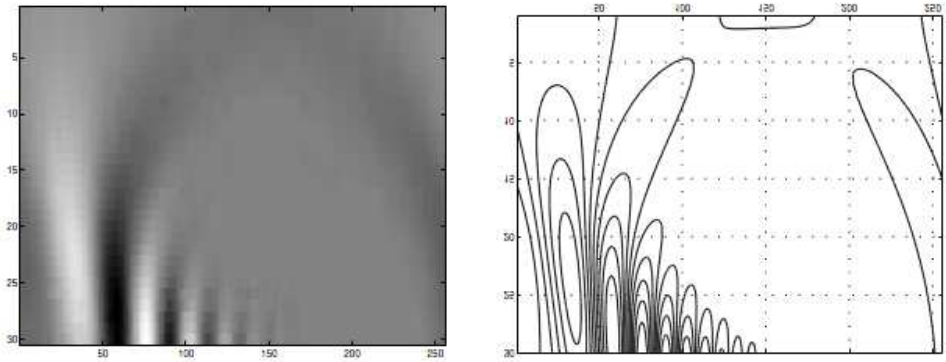
Αυτός είναι ο τρόπος με τον οποίο ο συνεχής μετασχηματισμός κυματιδίου πραγματοποιεί ανάλυση χρόνου – συχνότητας μιας συνάρτησης $f(t)$. Το αποτέλεσμα είναι μια συνάρτηση δύο μεταβλητών $W(a,b)$ η οποία φανερώνει τη συσχέτιση ανάμεσα

στην $f(t)$ και το κυματίδιο σε διαφορετικές συχνότητες στις διάφορες χρονικές στιγμές. Προφανώς η ποιότητα της συσχέτισης εξαρτάται από την επιλογή του κυματιδίου. Εάν το κυματίδιο είναι πολύ διαφορετικό σε σχέση με την $f(t)$ σε κάθε συχνότητα, κάθε χρονική στιγμή, τότε οι τιμές της $W(a,b)$ θα είναι όλες μικρές και δε θα παρουσιάζουν μεγάλες μεταβολές. Συνεπώς, όταν χρησιμοποιούνται κυματίδια στη συμπίεση εικόνας, διαφορετικά κυματίδια θα επιλεγούν για τους διάφορους τύπους εικόνας (συνεχούς τόνου, διακριτού κλπ) αλλά η ακριβής επιλογή του κατάλληλου κυματιδίου αποτελεί θέμα αρκετής έρευνας.

Στη συνέχεια ακολουθεί ένα παράδειγμα το οποίο είναι ελαφρώς πιο πολύπλοκο και οδηγεί σε καλύτερη κατανόηση του συνεχούς μετασχηματισμού κυματιδίων. Η συνάρτηση η οποία θα αναλυθεί είναι μια ημιτονοειδής κυματομορφή με αυξανόμενη συχνότητα. Δίνεται από τη συνάρτηση $f(t) = \sin(t^2)$, και παρουσιάζεται στο Σχήμα 7.6. Το κυματίδιο που θα χρησιμοποιηθεί είναι το Mexican Hat. Ο μετασχηματισμός κυματιδίου φαίνεται στο Σχήμα 7.7 a και b.



Σχήμα 7.6 : Ημιτονοειδής Κυματομορφή Αυξανόμενης Συχνότητας



(a) (b)

Σχήμα 7.7 : (a) Ο Συνεχής Μετασχηματισμός Κυματιδίου της Ημιτονοειδούς Κυματομορφής Αυξανόμενης Συχνότητας και (b) το Περίγραμμα του

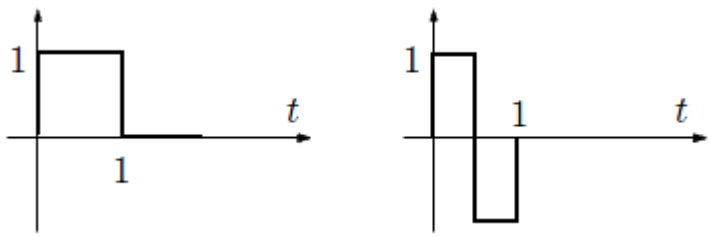
Μπορούμε να συγκρίνουμε την κυματιδιακή ανάλυση με την παρατήρηση ενός

αντικείμενου, αρχικά, από απόσταση, στη συνέχεια από πιο κοντά, μετά με έναν μεγεθυντικό φακό και τέλος με έναν μικροσκόπιο. Όταν παρατηρούμε το αντικείμενο από απόσταση, βλέπουμε το γενικό του σχήμα αλλά όχι τις μικρές λεπτομέρειες. Όταν κοιτάζουμε από ένα μικροσκόπιο, βλέπουμε μικρές λεπτομέρειες αλλά όχι το γενικό σχήμα. Αυτός είναι και ο λόγος που πραγματοποιούμε την ανάλυση σε διαφορετικές κλίμακες. Αλλάζοντας την κλίμακα του κυματιδίου λαμβάνουμε νέα πληροφορία σε σχέση με τη συνάρτηση που αναλύουμε.

7.2 Ο Μετασχηματισμός Κυματιδίου Haar

Στις πραγματικές εφαρμογές, η πληροφορία η οποία μετασχηματίζεται με χρήση κυματιδίων αφορά ψηφιακό ήχο και εικόνες. Η πληροφορία αυτή είναι διακριτή και αποτελείται από μεμονωμένους αριθμούς. Αυτός είναι και ο λόγος για τον οποίο στην πράξη χρησιμοποιούμε τον διακριτό και όχι τον συνεχή μετασχηματισμό κυματιδίου. Τον διακριτό μετασχηματισμό κυματιδίου θα τον περιγράψουμε στη συνέχεια αλλά πρώτα θα παρουσιαστεί ένα απλό παράδειγμα μετασχηματισμού αυτής της μορφής με την ονομασία “Μετασχηματισμός Κυματιδίου Haar”. Η χρήση του μετασχηματισμού κυματιδίου Haar για συμπίεση εικόνας θα περιγραφεί σε πρακτικό επίπεδο. Αρχικά θα περιγραφεί η εφαρμογή του για συμπίεση εικόνας στην κλίμακα του γκρι και στη συνέχεια η περιγραφή θα επεκταθεί και στις έγχρωμες εικόνες.

Ο μετασχηματισμός Haar χρησιμοποιεί τη συνάρτηση κλιμάκωσης $\Phi(t)$ και το κυματίδιο $\psi(t)$, τα οποία εμφανίζονται στο Σχήμα 7.8, προκειμένου να αναπαριστούν ένα μεγάλο πλήθος συναρτήσεων $f(t)$.



Σχήμα 7.8 : Κυματίδιο Haar

Η αναπαράσταση δίνεται από το άπειρο άθροισμα:

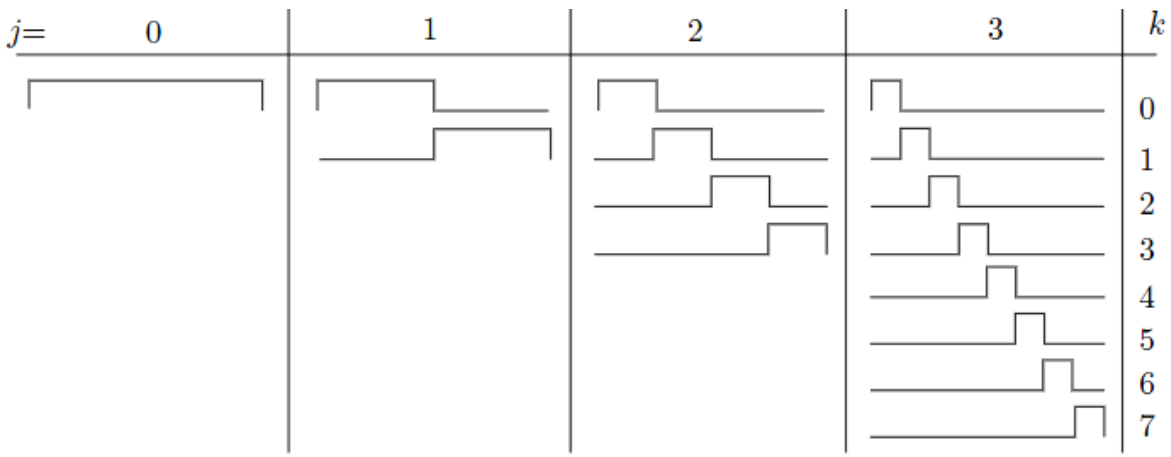
$$f(t) = \sum_{k=-\infty}^{\infty} c_k \Phi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi\left(2^j t - k\right), \tag{7.4}$$

όπου c_k και $d_{j,k}$ είναι οι συντελεστές που πρέπει να υπολογιστούν.

Η βασική συνάρτηση κλιμάκωσης $\Phi(t)$ είναι ένας μοναδιαίος παλμός

$$\Phi(t) := \begin{cases} 1 \; \gamma\iota\alpha \; 0 \leq t < 1 \\ 0 \; \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\acute{\alpha} \end{cases}$$

Η συνάρτηση $\Phi(t-k)$ είναι ένα αντίγραφο της $\Phi(t)$, μετατοπισμένο k μονάδες προς τα δεξιά. Αντίστοιχα, η $\Phi(2t-k)$ είναι ένα αντίγραφο της $\Phi(t-k)$ συμπιεσμένο στο χρόνο κατά έναν παράγοντα 2 σε σχέση με την $\Phi(t-k)$. Τα μετατοπισμένα αντίγραφα χρησιμοποιούνται για να προσεγγίσουν την $f(t)$ σε υψηλότερες αναλύσεις. Το Σχήμα 7.9 απεικονίζει τις συναρτήσεις $\Phi(2^j t-k)$ για $j = 0, 1, 2 \; \kappa\alpha\iota \; 3$ και για $k = 0, 1, \ldots, 7$.



Σχήμα 7.9 : Μετατοπισμένα Αντίγραφα το Κυματιδίου Harr

Το βασικό κυματίδιο Haar είναι η βηματική συνάρτηση

$$\psi(t) := \begin{cases} 1 \; \gamma\iota\alpha \; 0 \leq t < \frac{1}{2} \\ -1 \; \gamma\iota\alpha \frac{1}{2} \leq t < 1 \\ 0 \; \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\acute{\alpha} \end{cases}$$

Από αυτό μπορούμε να διαπιστώσουμε ότι ο γενικός μετασχηματισμός κυματιδίου Haar $\psi(2^j t-k)$ είναι ένα αντίγραφο της $\psi(t)$ μετατοπισμένο κατά k μονάδες προς τα δεξιά και με μια κλιμάκωση της διάρκειας του παλμού κατά $1/2^j$της διάρκειας του.

Στο Σχήμα 7.10 παρουσιάζονται τα τέσσερα κυματίδια Haar $\psi(2^j t-k)$ για $k = 0,1,2 \; \kappa\alpha\iota \; 3$.

Σχήμα 7.10 : Κυματίδια Haar για $k = 0, 1, 2, 3$

Τόσο η $\Phi(2^j t - k)$ όσο και η $\psi(2^j t - k)$ είναι μη μηδενικές σε ένα διάστημα μήκους $1/2^j$. Το διάστημα αυτό είναι η στήριξη τους (support). Δεδομένου ότι το διάστημα αυτό τείνει να είναι πολύ μικρό, μπορούμε να πούμε ότι οι συναρτήσεις αυτές έχουν συγκεντρωμένη στήριξη.

Η παρουσίαση του βασικού μετασχηματισμού θα γίνει με ένα απλό παράδειγμα της βηματικής συνάρτησης

$$f(t) = \begin{cases} 5, & 0 \le t < \frac{1}{2} \\ 3, & \frac{1}{2} \le t < 1 \end{cases}$$

Εύκολα φαίνεται ότι η $f(t) = 4\Phi(t) + \psi(t)$. Λέμε ότι τα αρχικά βήματα $(5,3)$ έχουν μετασχηματιστεί σε μια χαμηλή ανάλυση με μέση τιμή 4 και σε μια υψηλή ανάλυση με λεπτομέρεια 1. Με χρήση συμβολισμού πινάκων, αυτό μπορεί να εκφραστεί ως $(5,3)A_2 = (4,1)$ όπου το A_2 είναι ο δεύτερης τάξης μετασχηματισμός Haar.

Μια εικόνα (με αποχρώσεις του γκρι) είναι ένας δισδιάστατος πίνακας με τις τιμές των εικονοστοιχείων. Για να παρουσιαστεί ο τρόπος με τον οποίο ο μετασχηματισμός Haar χρησιμοποιείται για τη συμπίεση εικόνας, ξεκινάμε με μια γραμμή από τιμές εικονοστοιχείων, για παράδειγμα έναν μονοδιάστατο πίνακα με n τιμές. Για λόγους απλότητας θεωρούμε (χωρίς βλάβη της γενικότητας) ότι το n είναι δύναμη του 2. Εφόσον το n έχει διαφορετική τιμή συμπληρώνεται με μηδενικά τα οποία κατά την αποσυμπίεση αφαιρούνται. Θεωρούμε τον πίνακα με τις οκτώ τιμές $(1, 2, 3, 4, 5, 6, 7, 8)$. Αρχικά υπολογίζουμε τις τέσσερεις μέσες τιμές

$$\frac{1+2}{2}=\frac{3}{2},\qquad \frac{3+4}{2}=\frac{7}{2},\qquad \frac{5+6}{2}=\frac{11}{2}\qquad \kappa\alpha\iota\qquad \frac{7+8}{2}=\frac{15}{2}$$

Προφανώς δεν είναι δυνατό να ανακατασκευαστούν οι οκτώ αρχικές τιμές από τους τέσσερεις αυτούς μέσους όρους, για το λόγο αυτό υπολογίζουμε και τις τέσσερεις διαφορές τους

$$\frac{1-2}{2}=-\frac{1}{2},\frac{3-4}{2}=-\frac{1}{2},\frac{5-6}{2}=-\frac{1}{2}\ \kappa\alpha\iota\frac{7-8}{2}=-\frac{1}{2}$$

Οι συντελεστές αυτοί ονομάζονται συντελεστές λεπτομέρειας. Μπορούμε να θεωρήσουμε τις τιμές των μέσω όρων ως τις τιμές που περιέχουν τη βασική αναπαράσταση της αρχικής εικόνας και τις διαφορές που αποτελούν τους συντελεστές λεπτομέρειας ως την πληροφορία που απαιτείται για την ανακατασκευή της αρχικής εικόνας. Εφόσον τα εικονοστοιχεία της εικόνας είναι συσχετισμένα, η βασική αναπαράσταση θα μοιάζει σε μεγάλο βαθμό με την αρχική εικόνα καθώς οι τιμές των συντελεστών λεπτομέρειας θα έχουν μικρές τιμές. Η ανάλυση αυτή μας εξηγεί το λόγο που η συμπίεση εικόνας με κυματίδια Haar χρησιμοποιεί μέσες τιμές και διαφορές (συντελεστές λεπτομέρειας).

Είναι προφανές ότι ο πίνακας

$$\left(\frac{3}{2},\quad \frac{7}{2},\quad \frac{11}{2},\quad \frac{15}{2},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2}\right)$$

ο οποίος αποτελείται από τις τέσσερεις μέσες τιμές και τις τέσσερεις διαφορές μπορεί να χρησιμοποιηθεί για να ανακατασκευάσει τις οκτώ αρχικές τιμές της εικόνας. Ο πίνακας αυτός έχει οκτώ τιμές όμως οι τελευταίες τέσσερεις, δηλαδή οι διαφορές, τείνουν να είναι μικροί αριθμοί, γεγονός που βοηθά στη συμπίεση. Το γεγονός αυτό μας ενθαρρύνει ώστε να επαναλάβουμε τη διαδικασία στις τέσσερεις μέσες τιμές που είναι και οι μεγάλοι αριθμοί του πίνακα. Μετασχηματίζονται πλέον σε δύο μέσες τιμές και δύο διαφορές οι οποίες είναι

$$\left(\frac{10}{4},\quad \frac{26}{4},\quad -\frac{4}{4},\quad -\frac{4}{4},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2}\right)$$

Η επόμενη και τελευταία επανάληψη μετασχηματίζει τα δύο πρώτα στοιχεία του νέου πίνακα σε μία μέση τιμή (η οποία ουσιαστικά είναι η μέση τιμή των οκτώ στοιχείων του αρχικού πίνακα) και μία διαφορά:

$$\left(\frac{36}{8},\quad -\frac{16}{8},\quad -\frac{4}{4},\quad -\frac{4}{4},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2},\quad -\frac{1}{2}\right)$$

Ο τελευταίος πίνακας αποτελεί τον μετασχηματισμό κυματιδίου Haar του αρχικού

πίνακα.

Λόγω των διαφορών, ο μετασχηματισμός κυματιδίου τείνει να έχει μικρότερες τιμές σε σχέση με τις αρχικές τιμές των εικονοστοιχείων, συνεπώς είναι ευκολότερη η συμπίεση με χρήση RLE σε συνδυασμό με κωδικοποίηση Huffman. Συμπίεση με απώλειες μπορεί να επιτευχθεί με κβάντιση των μικρότερων συντελεστών ή ακόμα και με ολοκληρωτική διαγραφή τους (μηδενίζοντας τους).

Στο σημείο αυτό θα ήταν χρήσιμο να αναφερθούμε στην πολυπλοκότητα του συγκεκριμένου μετασχηματισμού, δηλαδή στο πλήθος των απαιτούμενων πράξεων σε σχέση με μέγεθος του προβλήματος εισόδου. Στο παράδειγμα μας χρειάστηκαν $8 + 4 + 2 = 14$ αριθμητικές πράξεις (προσθέσεις και αφαιρέσεις), αριθμός που μπορεί να εκφραστεί ως $2(8 - 1)$. Στη γενική περίπτωση υποθέτουμε ότι ξεκινάμε με $N = 2^n$ δεδομένα εισόδου. Στην πρώτη επανάληψη απαιτούνται 2^n πράξεις, στη δεύτερη 2^{n-1} κλπ, μέχρι την τελευταία επανάληψη όπου απαιτούνται $2^{n-(n-1)} = 2^1$ πράξεις. Συνεπώς, ο συνολικός αριθμός πράξεων είναι

$$\sum_{i=1}^{n} 2^i = \sum_{i=0}^{n} 2^i - 1 = \frac{1 - 2^{n+1}}{1 - 2} - 1 = 2^{n+1} - 2 = 2(2^n - 1) = 2(N - 1)$$

Ο μετασχηματισμός κυματιδίου Haar για N δεδομένα εισόδου μπορεί να εκτελεστεί σε $2(N - 1)$ πράξεις, επομένως η πολυπλοκότητα του είναι $O(N)$, αποτέλεσμα εξαιρετικό.

Είναι χρήσιμο να συνδεθεί με κάθε επανάληψη, μια ποσότητα που ονομάζεται ανάλυση και αναφέρεται στο πλήθος των μέσων τιμών που απομένων να υπολογιστούν μετά το τέλος κάθε επανάληψης. Η ανάλυση στο τέλος κάθε μιας από τις τρείς επαναλήψεις είναι $4(= 2^2), 2(= 2^1)\ \kappa\alpha\iota\ 1(= 2^0)$. Στη συνέχεια θα δείξουμε ότι κάθε τιμή του μετασχηματισμού κυματιδίου πρέπει να κανονικοποιηθεί. Αυτό επιτυγχάνεται με τη διαίρεση κάθε τιμής του μετασχηματισμού με την τετραγωνική ρίζα της τιμής ανάλυσης τη δεδομένη χρονική στιγμή (ορθοκανονικός μετασχηματισμός Haar). Επομένως στο παράδειγμα μας ο μετασχηματισμός κυματιδίου τελικά γίνεται:

$$\left(\frac{\frac{36}{8}}{\sqrt{2^0}}, -\frac{\frac{16}{8}}{\sqrt{2^0}}, -\frac{\frac{4}{4}}{\sqrt{2^1}}, -\frac{\frac{4}{4}}{\sqrt{2^1}}, -\frac{\frac{1}{2}}{\sqrt{2^2}}, -\frac{\frac{1}{2}}{\sqrt{2^2}}, -\frac{\frac{1}{2}}{\sqrt{2^2}}, -\frac{\frac{1}{2}}{\sqrt{2^2}} \right)$$

Εάν χρησιμοποιηθεί ο κανονικοποιημένος μετασχηματισμός κυματίδιου, αποδεικνύεται ότι εάν αγνοήσουμε τις μικρότερες διαφορές (συντελεστές ανάλυσης), αυτό αποτελεί την καλύτερη επιλογή στη συμπίεση κυματιδίου με απώλειες αφού προκαλεί την

ελάχιστη απώλεια πληροφορίας της εικόνας.

Οι δύο ρουτίνες στο Σχήμα 7.11 παρουσιάζουν τον τρόπο με τον οποίο υπολογίζεται ο κανονικοποιημένος μετασχηματισμός κυματιδίου ενός πίνακα n στοιχείων (όπου n είναι δύναμη του 2). Η ανακατασκευή του αρχικού πίνακα από τις κανονικοποιημένες τιμές του μετασχηματισμού κυματιδίου υλοποιείται από τις ρουτίνες στο Σχήμα 7.12.

```
function a=NWTcalc(a,n)
    a=a/sqrt(n);
    j=n;

    while j >= 2
        NWTstep(a,j);
        j=j/2;
    end
end

function a=NWTstep(a,j)

    for i=1:j/2
        b(i)=(a(2*i-1)+a(2*i))/sqrt(2);
        b(j/2+i)=(a(2*i-1)-a(2*i))/sqrt(2);
    end
    a=b;
end
```

Σχήμα 7.11 : Υπολογισμός του Κανονικοποιημένου Μετασχηματισμού Κυματιδίου Haar

```
function a=NWTreconst(a,n)
    j=2;

    while j < n
        NWTstep(a,j);
        j=2*j;
    end
    a=a*sqrt(n);
end

function a=NWTRstep(a,j)

    for i=1:j/2
        b(2*i-1)=(a(i)+a(j/2+i))/sqrt(2);
        b(2*i)=(a(i)-a(j/2+i))/sqrt(2);
    end
    a=b;
end
```

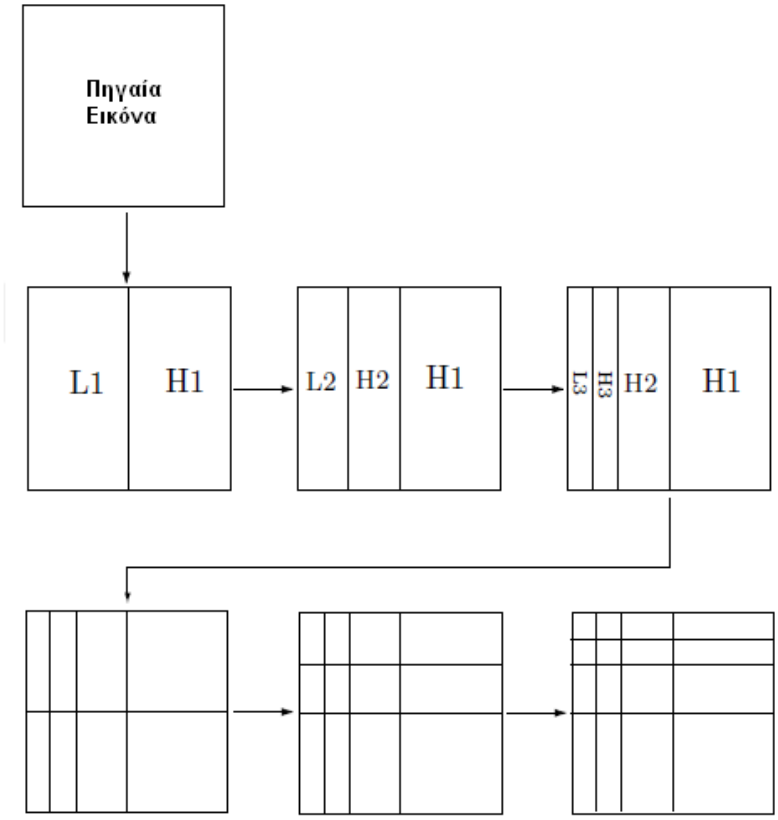
Σχήμα 7.12 : Ανακατασκευή από τους Κανονικοποιημένους Συντελεστές του Μετασχηματισμού Κυματιδίου Haar

7.3 Εφαρμογή του Μετασχηματισμού Κυματιδίου Haar

7.3.1 Ο Δισδιάστατος Μετασχηματισμός Haar στη Συμπίεση Εικόνας

Αφού επεξηγήθηκε αναλυτικά η ιδέα του μετασχηματισμού κυματιδίου είναι εύκολο να γενικευτεί στις δύο διαστάσεις[47]. Αυτό μπορεί να επιτευχθεί με αρκετούς τρόπους όπως θα φανεί στη συνέχεια. Αρχικά θα εξετασθούν δύο μέθοδοι. Η μέθοδος της τυπικής αποδόμηση[46] και η μέθοδος αποδόμηση πυραμίδας.

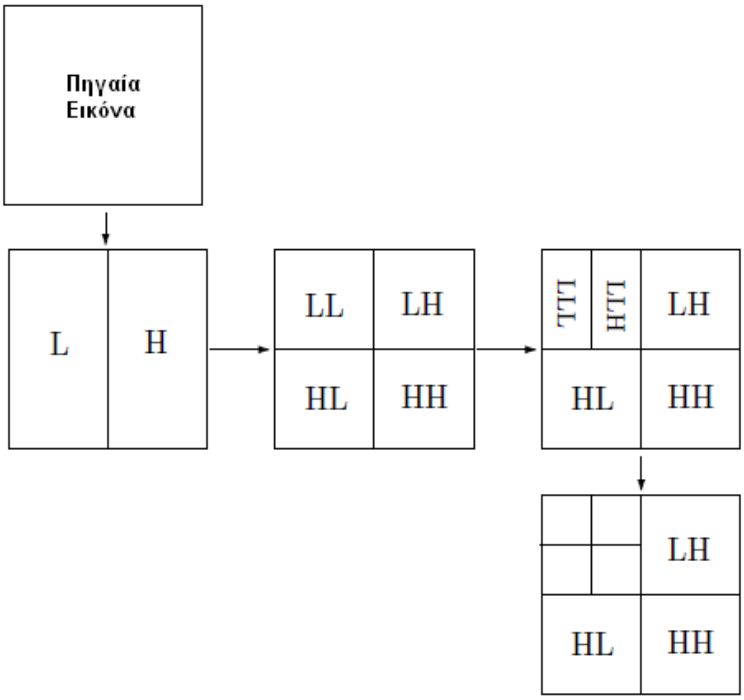
Στο Σχήμα 7.13 παρουσιάζεται η πρώτη μέθοδος, σύμφωνα με την οποία ο υπολογισμός μετασχηματισμού κυματιδίου πραγματοποιείται ως εξής: Σε κάθε γραμμή της εικόνας υπολογίζεται ο μετασχηματισμός κυματιδίου. Αυτό έχει ως αποτέλεσμα μια μετασχηματισμένη εικόνα όπου ή πρώτη στήλη περιέχει τις μέσες τιμές ενώ όλες οι υπόλοιπες στήλες περιέχουν τις τιμές των διαφορών. Ο αλγόριθμος στη συνέχεια υπολογίζει τον μετασχηματισμό κυματιδίου σε κάθε στήλη. Αυτό έχει σαν αποτέλεσμα τη δημιουργία μίας μόνο μέσης τιμής στην επάνω αριστερή γωνία με τα υπόλοιπα στοιχεία της πρώτης γραμμής να περιέχουν τις μέσες τιμές των διαφορών και κάθε άλλο εικονοστοιχείο να έχει μετασχηματιστεί σε μια διαφορά.



Σχήμα 7.13 : Ο Τυπικός Μετασχηματισμός Κυματιδίου και η Αποδόμηση του

Σύμφωνα με τη δεύτερη μέθοδο, ο υπολογισμός του μετασχηματισμού κυματιδίου της

εικόνας επιτυγχάνεται με εναλλαγές ανάμεσα σε γραμμές και στήλες. Στο πρώτο βήμα υπολογίζονται οι μέσες τιμές και οι διαφορές για κάθε γραμμή (μόνο για μία επανάληψη και όχι ολόκληρος ο μετασχηματισμός). Αυτό δημιουργεί μέσες τιμές στο αριστερό μισό τμήμα της εικόνας και διαφορές στο δεξί μισό της τμήμα. Στο δεύτερο βήμα υπολογίζονται οι μέσες τιμές και οι διαφορές για κάθε στήλη αυτή τη φορά. Αυτό έχει ως αποτέλεσμα τη δημιουργία μέσων τιμών στο επάνω αριστερό τέταρτο (τετράγωνο) της εικόνας και διαφορές σε όλη την υπόλοιπη. Τα βήματα 3 και 4 δρουν στις γραμμές και τις στήλες του επάνω αριστερού τετάρτου της εικόνας και με αυτό τον τρόπο συγκεντρώνονται οι μέσες τιμές στο επάνω αριστερό υπό-τέταρτο. Τα δύο αυτά βήματα επαναλαμβάνονται συνεχώς σε όλο και μικρότερα υπό-τετράγωνα της εικόνας μέχρις ότου να μείνει μόνο μία μέση τιμή και όλες οι υπόλοιπες τιμές εικονοστοιχείων να έχουν μετασχηματιστεί σε διαφορές. Η διαδικασία αυτή παρουσιάζεται στο Σχήμα 7.14.



Σχήμα 7.14 : Ο Μετασχηματισμός Κυματιδίου Πυραμίδας

Οι μετασχηματισμοί που περιγράφηκαν σε προηγούμενες ενότητες ήταν ορθογώνιοι. Μετασχημάτιζαν τις αρχικές τιμές των εικονοστοιχείων σε ένα μικρό πλήθος μεγάλων αριθμών και ένα μεγάλο πλήθος μικρών αριθμών. Αντίθετα, μετασχηματισμοί κυματιδίων όπως ο μετασχηματισμός Haar είναι μετασχηματισμοί υποζώνης[40]. Διαμερίζουν την εικόνα σε περιοχές όπου μια περιοχή περιέχει μεγάλους αριθμούς (μέσες τιμές στην περίπτωση του μετασχηματισμού Haar) ενώ οι άλλες περιοχές έχουν

μικρούς αριθμούς (διαφορές). Ωστόσο, οι περιοχές αυτές οι οποίες ονομάζονται υποζώνες είναι κάτι περισσότερο από απλά σύνολα μεγάλων και μικρών αριθμών. Αναπαριστούν διαφορετικά γεωμετρικά χαρακτηριστικά της εικόνας. Για να επεξηγηθεί αυτό εξετάζουμε μία μικρή, σχεδόν ομοιόμορφη εικόνα με μία κάθετη και μια οριζόντια γραμμή. Ο Πίνακας 7.1 a δείχνει μια εικόνα 8×8 με τις τιμές των εικονοστοιχείων ίσες με 12 εκτός από μια κατακόρυφη γραμμή με τιμές εικονοστοιχείων 14 και μια οριζόντια γραμμή με τιμές εικονοστοιχείων 16.

Πίνακας 7.1 : Μία Εικόνα Μεγέθους 8×8 και η Αποδόμησή της σε Υποζώνες

12	12	12	12	14	12	12	12
12	12	12	12	14	12	12	12
12	12	12	12	14	12	12	12
12	12	12	12	14	12	12	12
12	12	12	12	14	12	12	12
16	16	16	16	14	16	16	16
12	12	12	12	14	12	12	12
12	12	12	12	14	12	12	12

(a)

12	12	13	12	0	0	2	0
12	12	13	12	0	0	2	0
12	12	13	12	0	0	2	0
12	12	13	12	0	0	2	0
12	12	13	12	0	0	2	0
16	16	15	16	0	0	-2	0
12	12	13	12	0	0	2	0
12	12	13	12	0	0	2	0

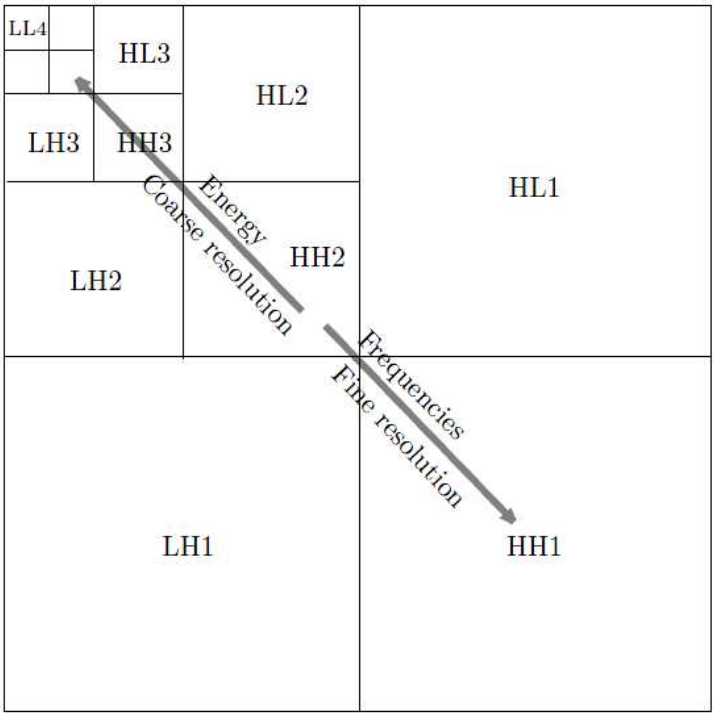
(b)

12	12	13	12	0	0	0	0
12	12	13	12	0	0	2	0
14	14	14	14	0	0	0	0
12	12	13	12	0	0	2	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-4	-4	-2	-4	0	0	4	0
0	0	0	0	0	0	0	0

(c)

Ο Πίνακας 7.1 b παρουσιάζει τα αποτελέσματα της εφαρμογής του μετασχηματισμού Haar για μία μόνο επανάληψη στις γραμμές της εικόνας. Το δεξί μισό τμήμα της εικόνας (οι διαφορές) είναι σχεδόν μηδέν, αντικατοπτρίζοντας την ομοιόμορφη δομή της εικόνας. Ωστόσο, τα ίχνη της κατακόρυφης γραμμής μπορούν εύκολα να παρατηρηθούν (τα στοιχεία με τιμή 2 υποδηλώνουν την αρνητική διαφορά). Ο Πίνακας 7.1 c παρουσιάζει τα αποτελέσματα από την εφαρμογή του μετασχηματισμού για μία μόνο επανάληψη στις τιμές των στηλών που εμπεριέχει ο Πίνακας 7.1 b. Η επάνω δεξιά υποζώνη πλέον περιέχει τα ίχνη της κατακόρυφης γραμμής, ενώ η κάτω αριστερή υποζώνη εμφανίζει τα ίχνη της οριζόντιας γραμμής. Οι υποζώνες αυτές συμβολίζονται με HL και LH αντίστοιχα. Η κάτω δεξιά υποζώνη, σημειώνεται με HH και αντικατοπτρίζει τα διαγώνια χαρακτηριστικά της εικόνας (τα οποία δεν υπάρχουν στην εικόνα μας). Η πλέον ενδιαφέρουσα είναι η επάνω αριστερή υποζώνη, που συμβολίζεται με LL η οποία αποτελείται αποκλειστικά από μέσες τιμές. Η υποζώνη αυτή αποτελεί το ένα τέταρτο της αρχικής εικόνας και περιέχει τα ίχνη τόσο των οριζόντιων όσο και των κατακόρυφων γραμμών.

Στο Σχήμα 7.15 εμφανίζονται τέσσερα επίπεδα από υποζώνες, όπου το επίπεδο 1 περιέχει τα χαρακτηριστικά λεπτομέρειας της εικόνας (συντελεστές κυματιδίου υψηλής συχνότητας) και το κορυφαίο επίπεδο, το επίπεδο 4, το οποίο αποτελείται από τα βασικά χαρακτηριστικά της εικόνας (συντελεστές χαμηλής συχνότητας). Είναι προφανές ότι τα χαμηλότερα επίπεδα μπορούν να κβαντιστούν ισχυρότερα χωρίς ιδιαίτερες απώλειες σε ότι αφορά τη σημαντική πληροφορία της εικόνας, ενώ τα υψηλότερα επίπεδα θα πρέπει να κβαντιστούν λιγότερο. Η τμηματοποίηση της εικόνας σε υποζώνες αποτελεί τη βάση όλων των μεθόδων συμπίεσης που χρησιμοποιούν μετασχηματισμό κυματιδίων.

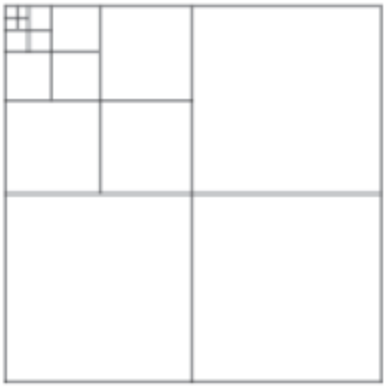


Σχήμα 7.15 : Επίπεδα και Υποζώνες Κατά την Αποδόμηση του Κυματιδίου

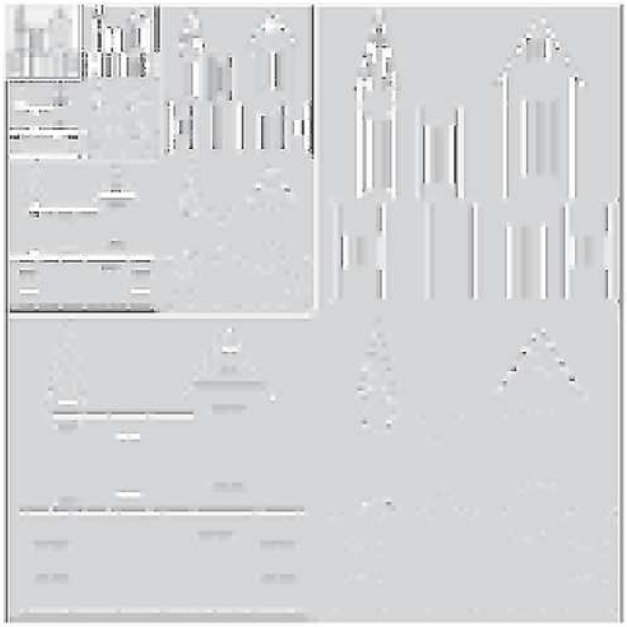
Το Σχήμα 7.16 παρουσιάζει το τυπικό αποτέλεσμα του μετασχηματισμού κυματιδίου με χρήση της μεθόδου πυραμίδας. Στο Σχήμα 7.16 a εμφανίζεται η αρχική εικόνα ενώ στο Σχήμα 7.16 b είναι η γενική μορφή αποδόμησης με τη μέθοδο πυραμίδας. Προκειμένου να παρουσιαστεί ο τρόπος με τον οποίο δρα ο μετασχηματισμός πυραμίδας, η εικόνα αυτή αποτελείται μόνο από οριζόντιες, κάθετες και διαγώνιες γραμμές. Τα τέσσερα τέταρτα στο Σχήμα 7.16 c παρουσιάζουν τέσσερεις μικρότερες εκδοχές της εικόνας. Η επάνω αριστερή υποζώνη η οποία περιέχει τις μέσες τιμές, είναι παρόμοια με την αρχική εικόνα, ενώ οι υπόλοιπες τρείς υποζώνες της εικόνας παρουσιάζουν τις λεπτομέρειες της. Λόγο του τρόπου που σχεδιάστηκε ο μετασχηματισμός πυραμίδας, η επάνω δεξιά υποζώνη περιέχει τις κατακόρυφες λεπτομέρειες, η κάτω αριστερή

υποζώνη τις οριζόντιες λεπτομέρειες και η κάτω δεξιά τις λεπτομέρειες των διαγώνιων γραμμών. Στο Σχήμα 7.16 c παρουσιάζονται τα αποτελέσματα που προκύπτουν εφαρμόζοντας τον μετασχηματισμό επανειλημμένα. Η εικόνα έχει μετασχηματιστεί σε υποζώνες οριζόντιων, κάθετων και διαγώνιων λεπτομερειών, ενώ το επάνω αριστερό υπό-τέταρτο το οποίο περιλαμβάνει τις μέσες τιμές έχει συρρικνωθεί σε ένα μόνο εικονοστοιχείο.

(a)



(b)



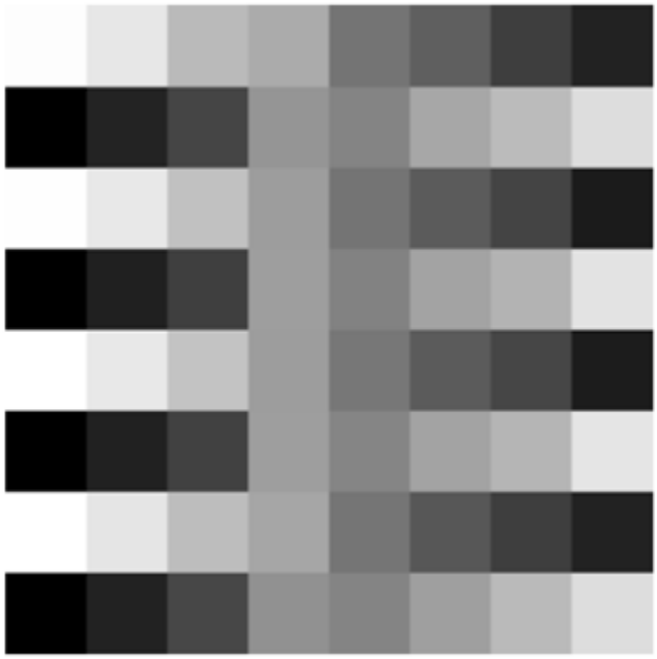
(c)

Σχήμα 7.16 : Παράδειγμα του Μετασχηματισμού Κυματιδίου Πυραμίδας

7.3.2 Προσέγγιση του Μετασχηματισμού Haar με τη Χρήση Πινάκων

Η βασική αρχή του μετασχηματισμού Haar είναι ο υπολογισμός των μέσων τιμών και

των διαφορών. Αποδεικνύεται ότι αυτό μπορεί να επιτευχθεί και με πολλαπλασιασμούς πινάκων. Ως παράδειγμα μπορεί να αναφερθεί ο μετασχηματισμός της εικόνας στο Σχήμα 7.17.



255	224	192	159	127	95	63	32
0	32	64	159	127	159	191	223
255	224	192	159	127	95	63	32
0	32	64	159	127	159	191	223
255	224	192	159	127	95	63	32
0	32	64	159	127	159	191	223
255	224	192	159	127	95	63	32
0	32	64	159	127	159	191	223

Σχήμα 7.17 : Παράδειγμα Αρχικής Εικόνας

Εύκολα μπορεί να κατασκευαστεί ένας πίνακας ο οποίος όταν πολλαπλασιαστεί με το διάνυσμα αυτό να μας δώσει τις τέσσερεις μέσες τιμές και τις τέσσερεις διαφορές. Αυτό επιτυγχάνει ο πίνακας A_1 ο οποίος όταν πολλαπλασιαστεί με τις τιμές των εικονοστοιχείων της πρώτης γραμμής που παρουσιάζει το Σχήμα 7.17, δημιουργεί το διάνυσμα $(239.5, 185.5\ 111.0, 47.5, 15.5, 16.5, 16.0, 15.5)$. Ομοίως, οι πίνακες A_2 και A_3 εκτελούν τα βήματα 2 και 3, αντίστοιχα, του μετασχηματισμού. Το αποτέλεσμα παρουσιάζεται στην εξίσωση (7.6):

$$A_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \quad A_1 \begin{pmatrix} 255 \\ 224 \\ 192 \\ 159 \\ 127 \\ 95 \\ 63 \\ 32 \end{pmatrix} = \begin{pmatrix} 239.5 \\ 175.5 \\ 111.0 \\ 47.5 \\ 15.5 \\ 16.5 \\ 16.0 \\ 15.5 \end{pmatrix} \tag{7.5}$$

$$A_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A_2 \begin{pmatrix} 239.5 \\ 175.5 \\ 111.0 \\ 47.5 \\ 15.5 \\ 16.5 \\ 16.0 \\ 15.5 \end{pmatrix} = \begin{pmatrix} 207.5 \\ 79.25 \\ 32.0 \\ 31.75 \\ 15.5 \\ 16.5 \\ 16.0 \\ 15.5 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 207.5 \\ 79.25 \\ 32.0 \\ 31.75 \\ 15.5 \\ 16.5 \\ 16.0 \\ 15.5 \end{pmatrix} = \begin{pmatrix} 143.375 \\ 64.125 \\ 32 \\ 31.75 \\ 15.5 \\ 16.5 \\ 16 \\ 15.5 \end{pmatrix} \tag{7.6}$$

Επομένως αντί για να υπολογιστούν οι μέσες τιμές και οι διαφορές, μπορούν απλώς να κατασκευαστούν οι πίνακες A_1, A_2 $και$ A_3, οι οποίοι πολλαπλασιαζόμενοι μεταξύ τους δημιουργούν τον πίνακα $W = A_1 A_2 A_3$. Από τον πολλαπλασιασμό του πίνακα W με κάθε στήλη της εικόνας I, που περιέχει τις τιμές των εικονοστοιχείων, προκύπτει:

$$WI = W\begin{pmatrix}255\\224\\192\\159\\127\\95\\63\\32\end{pmatrix} = \begin{pmatrix}\frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8}\\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8}\\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4}\\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}255\\224\\192\\159\\127\\95\\63\\32\end{pmatrix} = \begin{pmatrix}143.375\\64.125\\32\\31.75\\15.5\\16.5\\16\\15.5\end{pmatrix}$$

Ασφαλώς, η διαδικασία δεν έχει ολοκληρωθεί ακόμα. Προκειμένου να ολοκληρωθεί ο μετασχηματισμός, θα πρέπει ο πίνακας W να εφαρμοστεί και στις γραμμές του γινομένου WI. Αυτό επιτυγχάνεται εφαρμόζοντας τον πίνακα W στις στήλες του ανάστροφου $(WI)^T$ και στη συνέχεια αναστρέφοντας το αποτέλεσμα. Συνεπώς, ο ολοκληρωμένος μετασχηματισμός είναι:

$$I_{tr} = (W(WI)^T)^T = WIW^T$$

Ο αντίστροφος μετασχηματισμός είναι:

$$W^{-1}(W^{-1}I_{tr}^T)^T = W^{-1}(I_{tr}(W^{-1})^T)$$

Σε αυτό το σημείο ο κανονικοποιημένος μετασχηματισμός Haar γίνεται ιδιαιτέρως σημαντικός. Αντί για τον υπολογισμό των μέσων τιμών $[\tau\eta\varsigma\ \mu o\rho\varphi\acute{\eta}\varsigma\ (d_i + d_{i+1})/2]$ και των διαφορών $[(d_i - d_{i+1})/2]$, κρίνεται προτιμότερη η χρήση των ποσοτήτων $(d_i + d_{i+1})/\sqrt{2}$ και $(d_i - d_{i+1})/\sqrt{2}$. Αυτό οδηγεί σε έναν ορθοκανονικό πίνακα W και όπως είναι γνωστό ο αντίστροφος ενός ορθοκανονικού πίνακα είναι ο ανάστροφος του. Επομένως, ο αντίστροφος μετασχηματισμός μπορεί να γραφεί ως:

$$W^T I_{tr} W$$

Μεταξύ του μετασχηματισμού Haar και του αντιστρόφου του θα υπάρξει κβάντιση των μετασχηματισμένων συντελεστών ή ακόμα και διαγραφή τους. Εναλλακτικά, ο πίνακας I_{tr} μπορεί να συμπιεστεί με χρήση RLE και κώδικα Huffman.

7.4 Τράπεζες Φίλτρων

7.4.1 Η Έννοια και ο Ρόλος των Φίλτρων στο Μετασχηματισμό Υποζώνης

Η προσέγγιση του μετασχηματισμού Haar με χρήση πινάκων χρησιμοποιήθηκε για να αποτελέσει μια εισαγωγή στην ιδέα των τραπεζών φίλτρων[48]. Στην ενότητα αυτή θα εξετάσουμε τον μετασχηματισμό Haar ως ένα σύνολο φίλτρων αποτελούμενο από δύο φίλτρα, ένα βαθυπερατό και ένα υψηπερατό. Θα επεξηγηθούν οι όροι των φίλτρων, βαθυπερατό και υψηπερατό και να αναδειχθεί η ιδέα των τραπεζών φίλτρων η οποία οδηγεί με φυσικό τρόπο στον μετασχηματισμό υποζώνης. Ο μετασχηματισμός Haar είναι ο απλούστερος μετασχηματισμός κυματιδίου συνεπώς θα χρησιμοποιηθεί για να εισαχθούν κάποιες νέες ιδέες και έννοιες. Ωστόσο, η χρήση του μετασχηματισμού αυτού ως τράπεζα φίλτρου ενδέχεται να μην είναι ιδιαίτερα αποδοτική. Οι περισσότερες πρακτικές εφαρμογές που χρησιμοποιούν φίλτρα κυματιδίων, κάνουν χρήση πιο εξελιγμένων συνόλων από συντελεστές φίλτρων, όμως όλες είναι βασισμένες στην ιδέα των φίλτρων και των τραπεζών φίλτρων.

Ένα φίλτρο είναι ένας γραμμικός συνδυασμός των συντελεστών φίλτρου $h(0), h(1), h(2)$ ... και οποιουδήποτε διανύσματος εισόδου x ώστε να παραχθεί το διάνυσμα εξόδου y σύμφωνα με τη σχέση

$$y(n) = \sum_k h(k)x(n-k) = h * x$$

όπου το σύμβολο $*$ δηλώνει τη συνέλιξη. Παρατηρούμε ότι τα όρια του αθροίσματος δεν έχουν δηλωθεί άμεσα καθώς εξαρτώνται από το μήκος των διανυσμάτων $x\ και\ h$. Δεδομένου ότι η ανεξάρτητη μεταβλητή t είναι ο χρόνος, μας διευκολύνει να υποθέσουμε ότι οι είσοδοι (συνεπώς και οι έξοδοι) δημιουργούνται οποιαδήποτε χρονική στιγμή $t = \cdots, -2, -1, 0, 1, 2, \ldots$ και επομένως χρησιμοποιούμε τον συμβολισμό

$$x = (\ldots, a, b, c, d, e, \ldots)$$

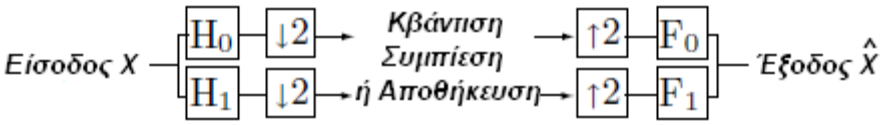
όπου η κεντρική τιμή c αποτελεί την είσοδο τη χρονική στιγμή μηδέν $[c = x(0)]$, οι τιμές d και e είναι οι τιμές εισόδου τις χρονικές στιγμές 1 και 2 αντίστοιχα και $b = x(-1) και\ a = x(-2)$. Στην πράξη οι είσοδοι είναι πάντοτε πεπερασμένοι, επομένως το άπειρο διάνυσμα x θα έχει πεπερασμένο αριθμό μη μηδενικών στοιχείων.

Μπορούμε να κατανοήσουμε καλύτερα τη συμπεριφορά ενός γραμμικού φίλτρου θεωρώντας το απλό σήμα εισόδου $x = (\ldots, 0, 0, 1, 0, 0, \ldots)$. Το σήμα αυτό είναι μηδέν κάθε χρονική στιγμή εκτός από την $t = 0$ όπου έχει τιμή 1. Το σήμα αυτό ονομάζεται

σήμα μοναδιαίου παλμού. Παρά το γεγονός ότι δεν έχουν προσδιοριστεί τα όρια του αθροίσματος, είναι προφανές ότι για κάθε n υπάρχει μόνο μία μη μηδενική τιμή στο άθροισμα, επομένως $y(n) = h(n)x(0) = h(n)$. Άρα η έξοδος $y(n) = h(n)$ τη χρονική στιγμή n είναι η απόκριση τη χρονική στιγμή n στο σήμα μοναδιαίου παλμού $x(0) = 1$. Δεδομένου ότι το πλήθος των συντελεστών του φίλτρου $h(i)$ είναι πεπερασμένο, το φίλτρο είναι ένα φίλτρο πεπερασμένης απόκρισης ή FIR φίλτρο.

Το Σχήμα 7.18 παρουσιάζει τη βασική ιδέα της τράπεζας φίλτρων. Η τράπεζα ανάλυσης αποτελείται από δύο φίλτρα. Ένα βαθυπερατό φίλτρο H_0 και ένα υψηπερατό H_1. Το βαθυπερατό φίλτρο χρησιμοποιεί τη διαδικασία της συνέλιξης για να αφαιρέσει τις υψηλές συχνότητες από το σήμα εισόδου x, διατηρώντας μόνο τις χαμηλές συχνότητες. Το υψηπερατό φίλτρο πραγματοποιεί το ακριβώς αντίθετο. Η χρήση και των δύο φίλτρων διαχωρίζει το σήμα εισόδου σε ζώνες συχνοτήτων.

Η είσοδος x μπορεί να είναι ένα μονοδιάστατο σήμα (διάνυσμα πραγματικών αριθμών) ή ένα δισδιάστατο σήμα ή εικόνα. Τα στοιχεία $x(n)$ του σήματος x εισάγονται στα φίλτρα ένα προς ένα και κάθε φίλτρο υπολογίζει και δίνει στην έξοδο του έναν αριθμό $y(n)$ ως απόκριση του $x(n)$. Ο αριθμός των αποκρίσεων θα είναι ο διπλάσιος του αριθμού των εισόδων (από τη στιγμή που υπάρχουν δύο φίλτρα), γεγονός πολύ κακό εάν αναλογιστεί κανείς ότι ενδιαφερόμαστε να συμπιέσουμε τα δεδομένα. Για να διορθωθεί αυτή η ανεπιθύμητη κατάσταση κάθε φίλτρο ακολουθείται από έναν υποδειγματολήπτη ο οποίος αγνοεί τα περιττά δείγματα εξόδου. Η διαδικασία αυτή υποδηλώνεται στο σχήμα με το σύμβολο " $\downarrow 2$". Μετά την υποδειγματοληψία, το πλήθος των αριθμών στην έξοδο, ισούται με το πλήθος των αριθμών στην είσοδο. Αξίζει να σημειωθεί ότι η τράπεζα φίλτρων που μόλις περιγράφηκε, ακολουθούμενη από τους υποδειγματολήπτες, εκτελεί τους ίδιους ακριβώς υπολογισμούς με τον πίνακα $W = A_1A_2A_3$ που περιγράφηκε στην προηγούμενη ενότητα. Οι τράπεζες φίλτρων αποτελούν απλώς έναν πιο γενικό τρόπο θεώρησης του μετασχηματισμού Haar και των μετασχηματισμών κυματιδίου εν γένει.



Σχήμα 7.18 : Φίλτρο δύο Καναλιών

Ο λόγος που χρησιμοποιούμε τράπεζες από φίλτρα αντί για ένα μοναδικό φίλτρο

έγκειται στο γεγονός ότι η χρήση περισσοτέρων, του ενός, φίλτρων μαζί σε συνδιασμό με υποδειγματοληψία, μπορεί να επιτύχει συμπεριφορά η οποία είναι αδύνατο να επιτευχθεί με ένα μόνο φίλτρο. Το σημαντικότερο χαρακτηριστικό της τράπεζας φίλτρων είναι η ικανότητα τους να ανακτούν το σήμα εισόδου από τις εξόδους $H_0x \; και \; H_1x$ παρά το γεγονός ότι έχει προηγηθεί υποδειγματοληψία.

Η διαδικασία της υποδειγματοληψίας δεν είναι αμετάβλητη στο χρόνο. Μετά την υποδειγματοληψία, στην έξοδο του φίλτρου εμφανίζονται οι τιμές των άρτιων δειγμάτων $y(0), y(2), y(4)$ ... , όμως εάν καθυστερήσουμε τις εισόδους κατά μία χρονική μονάδα, οι τιμές στην έξοδο θα είναι οι $y(-1), y(1), y(3)$ ... οι οποίες είναι διαφορετικές και ανεξάρτητες των τιμών της αρχικής εξόδου. Οι δύο αυτές ακολουθίες του σήματος αποτελούν τις δύο φάσεις του διανύσματος y.

Οι έξοδοι από την τράπεζα ανάλυσης ονομάζονται συντελεστές υποζώνης. Εφόσον η συμπίεση με απώλειες είναι αποδεκτή, μπορούν να κβαντιστούν και να συμπιεστούν με χρήση RLE, Huffman, αριθμητικής κωδικοποίησης ή οποιασδήποτε άλλης μεθόδου. Τελικά, οι συντελεστές αυτοί δίνονται ως είσοδοι στην τράπεζα σύνθεσης όπου πρώτα υπερδειγματοληπτούνται (με την εισαγωγή μηδενικών στα θέση των περιττών δειγμάτων που αγνοήθηκαν κατά τη διαδικασία ανάλυσης) και στη συνέχεια περνούν μέσα από τα αντίστροφα φίλτρα $F_0 \; και \; F_1$ όπου και συνδυάζονται ώστε να σχηματίσουν το τελικό διάνυσμα εξόδου $\hat{x}$. Η έξοδος κάθε φίλτρου ανάλυσης μετά την υποδειγματοληψία είναι:

$$(\downarrow y) = (\dots, y(-4), y(-2), y(0), y(2), y(4), \dots)$$

Η διαδικασία της υπερδειγματοληψίας εισάγει μηδενικά στη θέση των περιττών συντελεστών που αγνοήθηκαν κατά την ανάλυση με αποτέλεσμα, το διάνυσμα εξόδου να έχει τη μορφή:

$$(\uparrow y) = (\dots, y(-4), 0, y(-2), 0, y(0), 0, y(2), 0, y(4), 0, \dots)$$

Η διαδικασία της υποδειγματοληψίας προκαλεί απώλεια δεδομένων. Η υπερδειγματολήπτηση από μόνη της δεν αρκεί για να την αντισταθμίσει αφού στη θέση των τιμών που αγνοούνται εισάγει απλώς μηδενικά. Προκειμένου να επιτευχθεί συμπίεση χωρίς απώλειες, τα φίλτρα θα πρέπει να έχουν σχεδιαστεί με κατάλληλο τρόπο ώστε να αντισταθμίζουν την απώλεια αυτή των δεδομένων. Ένα χαρακτηριστικό το οποίο χρησιμοποιείται πολύ συχνά στη σχεδίαση φίλτρων είναι η ορθογωνιότητα. Το Σχήμα 7.19 παρουσιάζει ένα σύνολο ορθογώνιων φίλτρων μεγέθους 4. Τα φίλτρα του συνόλου αυτού είναι ορθογώνια καθώς το εσωτερικό τους γινόμενο είναι μηδέν.

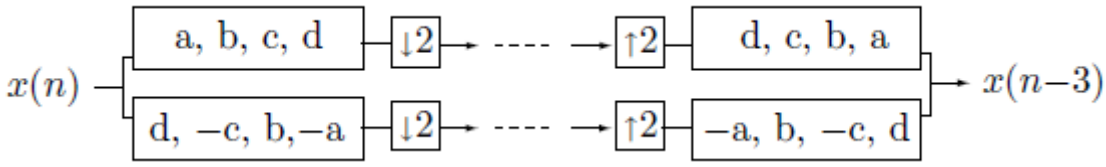
$$(a, b, c, d)(d, -c, b, -a) = 0$$

Παρατηρούμε την ομοιότητα των H_0 $\kappa\alpha\iota$ F_0 καθώς και των $H_1 \kappa\alpha\iota$ F_1. Η επιλογή των τιμών αυτών είναι ένα μεγάλο ζήτημα. Βασικές μέθοδοι και κανόνες σχετικά με την επιλογή τους θα αναφερθούν στη συνέχεια. Τυπικό παράδειγμα αποτελούν οι ακόλουθοι συντελεστές:

$$c_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \approx 0.48296, \qquad c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \approx 0.8365$$

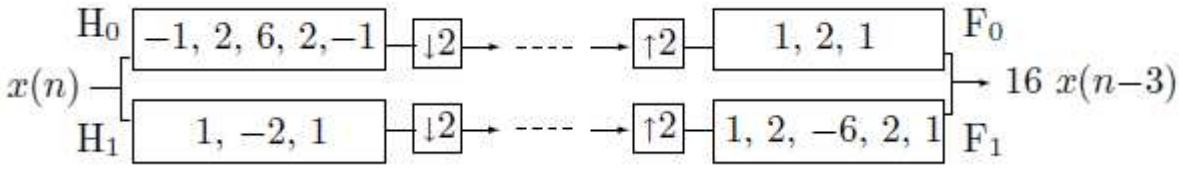
$$c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \approx 0.2241, \qquad c_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \approx -0.1294$$

Η προσομοίωση της διαδικασίας που πραγματοποιεί το φίλτρο, δείχνει ότι η ανακατασκευασμένη είσοδος είναι πανομοιότυπη με την αρχική είσοδο αλλά καθυστερημένη κατά τρείς χρονικές μονάδες.



Σχήμα 7.19 : Ορθογώνια Τράπεζα Φίλτρων με τέσσερεις Συντελεστές

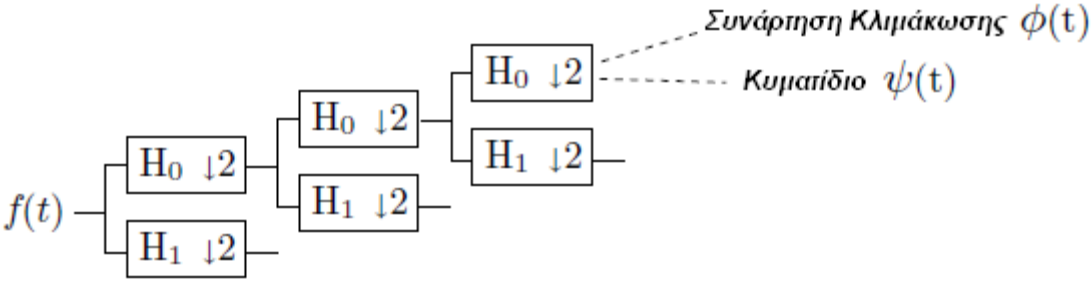
Μια τράπεζα φίλτρων μπορεί επίσης να είναι αμφιορθογώνια, ένας λιγότερο περιορισμένος τύπος φίλτρου σύμφωνα με τον οποίο χρησιμοποιούνται δύο διαφορετικές βάσεις οι οποίες είναι κάθετες μεταξύ τους αλλά ως σύνολο δεν αποτελούν ορθογώνια βάση. Το Σχήμα 7.20 παρουσιάζει ένα παράδειγμα φίλτρων αυτού του τύπου τα οποία μπορούν να ανακατασκευάσουν το σήμα επακριβώς. Παρατηρούμε την ομοιότητα ανάμεσα στα H_0 $\kappa\alpha\iota$ F_1 καθώς και ανάμεσα στα $H_1 \kappa\alpha\iota$ F_0.



Σχήμα 7.20 : Ημιορθογώνιο Φίλτρο με Τέλεια Ανακατασκευή

Όπως αντιλαμβανόμαστε από τη περιγραφή που είχαμε σε προηγούμενες ενότητες για τον μετασχηματισμό Haar, οι έξοδοι του βαθυπερατού φίλτρου H_0 θα περάσουν από το

φίλτρο ανάλυσης αρκετές φορές παράγοντας όλο και μικρότερες ακολουθίες. Η αναδρομική αυτή διαδικασία μπορεί να παρασταθεί με μία δενδροειδή μορφή όπως εμφανίζεται στο Σχήμα 7.21



Σχήμα 7.21 : Συνάρτηση Κλιμάκωσης και Κυματιδίου ως Όρια ενός Λογαριθμικού Δένδρου

Δεδομένου ότι κάθε κόμβος του δένδρου παράγει τον μισό αριθμό των δειγμάτων που προέκυψαν απο την προηγούμενη έξοδο, το δένδρο ονομάζεται λογαριθμικό δένδρο. Το Σχήμα 7.21 δείχνει πως η συνάρτηση κλιμάκωσης $\Phi(t)$ και το κυματίδιο $\psi(t)$ λαμβάνονται στο τέλος του δένδρου. Αυτή είναι η σύνδεση ανάμεσα στο διακριτό (με χρήση τράπεζας φίλτρων) και τον συνεχή μετασχηματισμό κυματιδίου.

Καθώς ανεβαίνουμε προς τη ρίζα του λογαριθμικού δένδρου, από το επίπεδο i στο επίπεδο $i+1$, υπολογίζονται οι νέες μέσες τιμές από τις νέες, υψηλότερης ανάλυσης συναρτήσεις κλιμάκωσης $\Phi(2^i t - k)$ και οι νέες λεπτομέρειες από τα νέα κυματίδια $\psi(2^i t - k)$.

$$\Sigma\acute{\eta}\mu\alpha\ \sigma\tau o\ \epsilon\pi\acute{\iota}\pi\epsilon\delta o\ i+1 = \begin{cases} B\alpha\sigma\iota\kappa\acute{o}\ \sigma\acute{\eta}\mu\alpha\ \sigma\tau o\ \epsilon\pi\acute{\iota}\pi\epsilon\delta o\ i\ (\mu\acute{\epsilon}\sigma\epsilon\varsigma\ \tau\iota\mu\acute{\epsilon}\varsigma) \\ + \\ \Sigma\acute{\eta}\mu\alpha\ \mu\epsilon\ \lambda\epsilon\pi\tau o\mu\acute{\epsilon}\rho\epsilon\iota\epsilon\varsigma\ \sigma\tau o\ \epsilon\pi\acute{\iota}\pi\epsilon\delta o\ i\ (\delta\iota\alpha\varphi o\rho\acute{\epsilon}\varsigma) \end{cases}$$

Κάθε επίπεδο του δένδρου αντιστοιχεί στη διπλάσια συχνότητα (ή στη διπλάσια ανάλυση) του προηγούμενου επιπέδου. Για τον λόγο αυτό το λογαριθμικό δένδρο ονομάζεται και δένδρο πολλαπλής ανάλυσης (multi-resolution tree). Διαδοχικά περάσματα μέσα από τα φίλτρα αυτά οδηγούν σε διαχωρισμό των υψηλών και των χαμηλών συχνοτήτων.

7.4.2 Φίλτρα στο Πεδίο της Συχνότητας

Προκειμένου να κατανοηθούν καλύτερα οι έννοιες του βαθυπερατού και του υψηπερατού φίλτρου θα πρέπει να εργαστούμε στο πεδίο των συχνοτήτων, όπου η συνέλιξη δύο διανυσμάτων αντικαθίσταται από τον πολλαπλασιασμό των μετασχηματισμών Fourier των διανυσμάτων αυτών. Το διάνυσμα $x(n)$ βρίσκεται στο

πεδίο του χρόνου, επομένως το ισοδύναμο του στο πεδίο των συχνοτήτων δίνεται από τον διακριτό μετασχηματισμό Fourier:

$$X(\omega) \stackrel{\text{def}}{=} X\left(e^{i\omega}\right) = \sum_{-\infty}^{\infty} x(n)e^{-in\omega}, \tag{7.7}$$

το οποίο συχνά γράφεται στο πεδίο z ως:

$$X(z) = \sum_{-\infty}^{\infty} x(n)z^{-n} \tag{7.8}$$

όπου $z \stackrel{\text{def}}{=} e^{i\omega}$. Η συνέλιξη με το h στο πεδίο του χρόνου γίνεται πλέον πολλαπλασιασμός, με τη συνάρτηση $H(\omega) = \sum h(n)e^{-in\omega}$, στο πεδίο των συχνοτήτων, επομένως μπορούμε να εκφράσουμε την έξοδο στο πεδίο των συχνοτήτων με:

$$Y\left(e^{i\omega}\right) = H\left(e^{i\omega}\right)X\left(e^{i\omega}\right), \tag{7.9}$$

ή σε συνεπτυγμένη μορφή $Y(\omega) = H(\omega)X(\omega)$ ή στο πεδίο του z $Y(z) = H(z)X(z)$. Όταν όλες οι είσοδοι είναι $X(\omega) = 1$, τότε η έξοδος στη συχνότητα ω είναι $Y(\omega) = H(\omega)$.

Σε αυτό το σημείο γίνεται αντιληπτή ή λειτουργία του βαθυπερατού φίλτρου Haar, υπολογίζοντας τη μέση τιμή δύο διαδοχικών τιμών εισόδου και παράγοντας στην έξοδο:

$$y(n) = \frac{1}{2}x(n) + \frac{1}{2}x(n-1) \tag{7.10}$$

Αυτή είναι η συνέλιξη με δύο μόνο όρους, $k = 0\ \kappa\alpha\iota\ k = 1$. Οι συντελεστές του φίλτρου είναι $h(0) = h(1) = 1/2$ και την έξοδο την ονομάζουμε κινητό μέσο όρο καθώς η $y(n)$ εξαρτάται από την τρέχουσα τιμή της εισόδου και την προηγούμενή της. Εάν η είσοδος είναι ο μοναδιαίος παλμός $x = (\dots,0,0,1,0,0,\dots)$, τότε η έξοδος είναι $y(0) = y(1) = 1/2$ ή $y = \left(\dots,0,0,\frac{1}{2},\frac{1}{2},0,\dots\right)$. Οι έξοδοι θα είναι απλώς οι συντελεστές του φίλτρου όπως αναφέρθηκε νωρίτερα.

Το φίλτρο υπολογισμού της μέσης τιμής μπορεί να θεωρηθεί ως ένας συνδυασμός του τελεστή ταυτότητας και του τελεστή καθυστέρησης. Η έξοδος που προκύπτει από τον τελεστή ταυτότητας ισούται με την τρέχουσα τιμή της εισόδου ενώ η έξοδος που παράγεται από τον τελεστή καθυστέρησης είναι η τιμή της εισόδου μια χρονική μονάδα νωρίτερα. Επομένως,

$$\Phi\acute{\iota}\lambda\tau\rho o\ \mu\acute{\varepsilon}\sigma\eta\varsigma\ \tau\iota\mu\acute{\eta}\varsigma = \frac{1}{2}(\tau\iota\mu\acute{\eta}\ \varepsilon\iota\sigma\acute{o}\delta o\upsilon) + \frac{1}{2}(\tau\iota\mu\acute{\eta}\ \varepsilon\iota\sigma\acute{o}\delta o\upsilon\ \kappa\alpha\theta\upsilon\sigma\tau\varepsilon\rho\eta\mu\acute{\varepsilon}\nu\eta\ \kappa\alpha\tau\acute{\alpha}\ 1\ \mu o\nu.)$$

Με σημειολογία πινάκων

$$\begin{pmatrix} \dots \\ y(-1) \\ y(0) \\ y(1) \\ \dots \end{pmatrix} = \begin{pmatrix} \dots & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ & \frac{1}{2} & \frac{1}{2} & & \\ & & \frac{1}{2} & \frac{1}{2} & \\ & & & \frac{1}{2} & \frac{1}{2} \\ & & & & \dots \end{pmatrix} \begin{pmatrix} \dots \\ x(-1) \\ x(0) \\ x(1) \\ \dots \end{pmatrix}$$

Οι τιμές του ½ στην κύρια διαγώνιο είναι αντίγραφα των βαρών της τρέχουσας τιμής της εισόδου. Όλες οι τιμές αυτές ουσιαστικά είναι ίσες με τον συντελεστή $h(0)$ του φίλτρου Haar. Οι τιμές του ½ στη διαγώνιο κάτω από την κύρια, είναι αντίγραφα των βαρών των τιμών της εισόδου την προηγούμενη χρονική στιγμή. Οι συντελεστές αυτοί είναι ίσοι με τον συντελεστή $h(1)$ του φίλτρου Haar. Επομένως, ο πίνακας είναι ένας σταθερός διαγώνιος πίνακας. Ένα φίλτρο κυματιδίου το οποίο διέθετε και συντελεστή $h(2)$ θα αντιστοιχούσε σε έναν πίνακα όπου ο συντελεστής φίλτρου θα εμφανιζόταν στη δεύτερη διαγώνιο κάτω από την κύρια διαγώνιο. Ο κανόνας πολλαπλασιασμού των πινάκων παράγει τη γνωστή μας συνέλιξη:

$$y(n) = h(0)x(n) + h(1)x(n-1) + h(2)x(n-2) + \cdots = \sum_k h(k)x(n-k)$$

Παρατηρούμε ότι ο πίνακας είναι κάτω τριγωνικός. Η άνω διαγώνιος η οποία θα αντιστοιχούσε στους συντελεστές φίλτρου $h(-1), h(-2), \dots$, είναι μηδέν. Όλοι οι συντελεστές φίλτρου με αρνητικούς δείκτες πρέπει να έχουν τιμή μηδέν αφού σε διαφορετική περίπτωση θα οδηγούμασταν σε εξόδους οι οποίες θα προηγούνταν των τιμών της εισόδου (μη αιτιατό).

Υπενθυμίζεται ότι ένα απλό FIR φίλτρο με $N+1$ συντελεστές $h(0), h(1), \dots, h(N)$ ($N+1$ taps) έχει $h(i) = 0$ για κάθε αρνητικό i και για $i > N$. Όταν αυτό εκφραστεί με όρους πινάκων, τότε έχουμε έναν πίνακα κάτω τριγωνικό και σταθερό. Τα φίλτρα αυτά είναι σημαντικά και χρησιμοποιούνται ευρέως στην πράξη.

Προκειμένου να επεξηγήσουμε την απόκριση συχνότητας ενός φίλτρου, επιλέγουμε ένα διάνυσμα εισόδου της μορφής

$$x(n) = e^{jn\omega} = \cos(n\omega) + jsin(n\omega), \gamma\iota\alpha - \infty < n < \infty$$

Αυτή είναι μια μιγαδική συνάρτηση όπου το πραγματικό και το φανταστικό της μέρος

είναι ένα συνημίτονο και ένα ημίτονο αντίστοιχα, και τα δύο με συχνότητα ω. Υπενθυμίζεται ότι ο μετασχηματισμός Fourier ενός παλμού περιέχει όλες τις συχνότητες ενώ ο μετασχηματισμός Fourier ενός συνημίτονου, έχει μόνο μια συχνότητα. Η μικρότερη συχνότητα είναι $\omega = 0$ για την οποία το διάνυσμα $x = (\dots,1,1,1,1,1,\dots)$. Η υψηλότερη συχνότητα είναι $\omega = \pi$, όπου το ίδιο διάνυσμα γίνεται $x = (\dots,1,-1,1,-1,1,\dots)$. Το ειδικό χαρακτηριστικό της εισόδου αυτής είναι ότι το διάνυσμα εξόδου $y(n)$ είναι πολλαπλάσιο της εισόδου. Για την κινητή μέση τιμή, η έξοδος (απόκριση φίλτρου) είναι:

$$y(n) = \frac{1}{2}x(n) + \frac{1}{2}x(n-1) = \frac{1}{2}e^{jn\omega} + \frac{1}{2}e^{j(n-1)\omega} = \left(\frac{1}{2} + \frac{1}{2}e^{-j\omega}\right)e^{jn\omega} = H(\omega)x(n)$$

όπου $H(\omega) = (\frac{1}{2} + \frac{1}{2}e^{-j\omega})$ είναι συνάρτηση απόκρισης συχνότητας του φίλτρου. Δεδομένου ότι $H(0) = \frac{1}{2} + \frac{1}{2} = 1$, παρατηρούμε ότι η είσοδος $x = (\dots,1,1,1,1,1,\dots)$ μετασχηματίζεται στον εαυτό της. Επίσης η $H(\omega)$ για μικρές τιμές του ω παράγει εξόδους οι οποίες είναι σε μεγάλο βαθμό όμοιες με την είσοδο. Το φίλτρο αυτό επιτρέπει στις χαμηλές συχνότητες να περνούν από όπου και προκύπτει το όνομα βαθυπερατό φίλτρο. Για $\omega = \pi$ η είσοδος είναι $x = (\dots,1,-1,1,-1,1,\dots)$ και η έξοδος αποτελείται μόνο από μηδενικά (αφού η μέση τιμή του 1 και του -1 είναι 0). Αυτό το βαθυπερατό φίλτρο εξομαλύνει τις περιοχές υψηλών συχνοτήτων του σήματος εισόδου.

Μπορούμε να γράψουμε:

$$H(\omega) = \left(cos\frac{\omega}{2}\right)e^{j\omega/2}$$

Εάν σχεδιάσουμε το μέγεθος $|H(\omega)| = \cos\left(\frac{\omega}{2}\right)$ του $H(\omega)$ (Σχήμα 7.22 a) παρατηρούμε αμέσως ότι παρουσιάζει μέγιστο όταν $\omega = 0$ (η χαμηλότερη συχνότητα) και δύο ελάχιστα όταν $\omega = \pm\pi$ (υψηλότερες συχνότητες).

Το υψηπερατό φίλτρο χρησιμοποιεί διαφορές για να αναδείξει τις υψηλές συχνότητες του σήματος εισόδου αφαιρώντας τα ομαλά τμήματα του (χαμηλές συχνότητες). Στην περίπτωση του μετασχηματισμού Haar, το υψηπερατό φίλτρο υπολογίζεται ως

$$y(n) = \frac{1}{2}x(n) - \frac{1}{2}x(n-1) = h * x$$

όπου οι συντελεστές του φίλτρου είναι $h(0) = \frac{1}{2}\kappa\alpha\iota\ h(1) = -\frac{1}{2}$ ή $h = \left(\dots,0,0,\frac{1}{2},-\frac{1}{2},0,\dots\right)$

Με σημειολογία πίνακα αυτό μπορεί να γραφεί ως:

$$\begin{pmatrix} \dots \\ y(-1) \\ y(0) \\ y(1) \\ \dots \end{pmatrix} = \begin{pmatrix} \dots & & & & \\ -\frac{1}{2} & \frac{1}{2} & & & \\ & -\frac{1}{2} & \frac{1}{2} & & \\ & & -\frac{1}{2} & \frac{1}{2} & \\ & & & \dots & \end{pmatrix} \begin{pmatrix} \dots \\ x(-1) \\ x(0) \\ x(1) \\ \dots \end{pmatrix}$$

Η κύρια διαγώνιος περιέχει τα αντίγραφα του $h(0)$ ενώ η διαγώνιος από κάτω περιέχει τα αντίγραφα $h(1)$. Χρησιμοποιώντας τους τελεστές ταυτότητας και καθυστέρησης έχουμε:

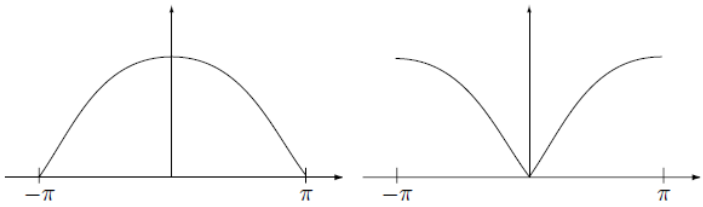
$$\textit{Φίλτρο μέσης τιμής} = \frac{1}{2}(\textit{τιμή εισόδου}) - \frac{1}{2}(\textit{τιμή εισόδου καθυστερημένη κατά } 1\ \textit{μον.})$$

Ξανά, επιλέγοντας ως είσοδο την $x(n) = e^{jn\omega}$, εύκολα φαίνεται ότι:

$$y(n) = \frac{1}{2}e^{jn\omega} - \frac{1}{2}e^{j(n-1)\omega} = \left(\frac{1}{2} - \frac{1}{2}e^{-j\omega}\right)e^{-j\omega/2} = \sin\left(\frac{\omega}{2}\right)je^{-j\omega/2}$$

Το μέγεθος $|H_1(\omega)| = \left|\sin\left(\frac{\omega}{2}\right)\right|$ παρουσιάζεται στο Σχήμα 7.22 b και είναι προφανές ότι παρουσιάζει ελάχιστο για μηδενική συχνότητα και δύο μέγιστα στις υψηλές συχνότητες.

Μια σημαντική ιδιότητα της τράπεζας φίλτρων είναι ότι κανένα από τα φίλτρα που τις αποτελούν δεν είναι αντιστρέψιμο, όμως η τράπεζα ως οντότητα πρέπει να σχεδιάζεται με τέτοιο τρόπο ώστε το σήμα εισόδου να μπορεί να ανακατασκευαστεί τέλεια από την έξοδο παρά την απώλεια πληροφορίας που προκαλείται από την υποδειγματοληψία. Εύκολα μπορούμε να διακρίνουμε για παράδειγμα πως το σταθερό σήμα $x = (\dots, 1, 1, 1, 1, 1, \dots)$ μετασχηματίζεται από το υψηπερατό φίλτρο H_1 σε ένα διάνυσμα εξόδου το οποίο αποτελείται μόνο από μηδενικά. Προφανώς, δεν μπορεί να υπάρξει αντίστροφο φίλτρο H_1^{-1} το οποίο να είναι σε θέση να ανακατασκευάσει το αρχικό σήμα εισόδου από ένα μηδενικό διάνυσμα. Στην καλύτερη των περιπτώσεων, ένας αντίστροφος μετασχηματισμός θα μπορούσε να χρησιμοποιήσει το μηδενικό διάνυσμα προκειμένου να ανακατασκευάσει ένα άλλο μηδενικό διάνυσμα. Το διάνυσμα $x = (\dots, 1, -1, 1, -1, 1, \dots)$ που αποτελείται από εναλλασσόμενες τιμές μετασχηματίζεται από το βαθυπερατό φίλτρο H_0 σε ένα διάνυσμα που αποτελείται μόνο από μηδενικά.

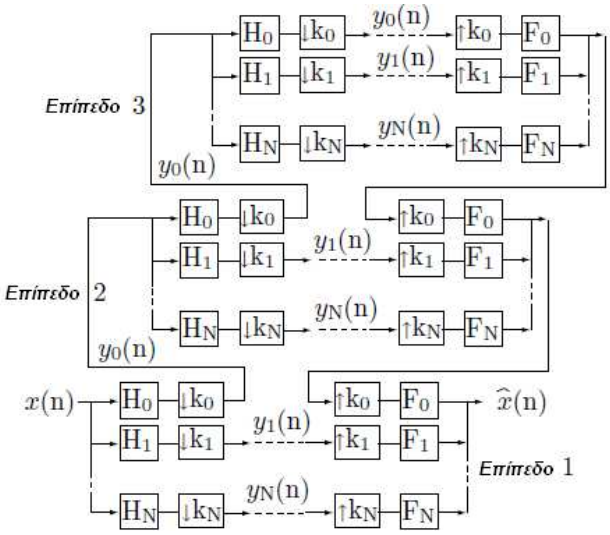


(a) (b)

Σχήμα 7.22 : Τα πλάτη ενός Βαθυπερατού (a) και ενός Υψηπερατού Φίλτρου

Συνοψίζοντας, όσα μελετήθηκαν σε αυτή την ενότητα σχετικά με τις τράπεζες φίλτρων θα πρέπει να συγκριθούν με όσα μελετήθηκαν στην ενότητα με τους ορθογώνιους μετασχηματισμούς εικόνας. Ενώ και στις δύο περιγράφηκαν μετασχηματισμοί, διαφέρουν στην προσέγγιση καθώς μελετούν διαφορετικές κλάσεις μετασχηματισμών. Στην πρώτη κλάση είδαμε μετασχηματισμούς βασισμένους σε σύνολα συναρτήσεων ορθογώνιων βάσεων οι οποίοι υπολογίζονται ως το εσωτερικό γινόμενο του σήματος εισόδου με τις συναρτήσεις βάσεις. Το αποτέλεσμα είναι ένα σύνολο μετασχηματισμένων συντελεστών οι οποίοι συμπιέζονται είτε χωρίς απώλειες με τη μέθοδο RLE ή κάποιον κωδικοποιητή εντροπίας είτε με απώλειες όπου σε αυτή την περίπτωση κβαντίζονται οι συντελεστές και ακολουθεί κωδικοποίηση εντροπίας. Στη δεύτερη κλάση ανήκουν οι μετασχηματισμοί υποζώνης[41][42]. Ένας διαφορετικός τύπος μετασχηματισμού ο οποίος υπολογίζεται από τη συνέλιξη του σήματος εισόδου με ένα σύνολο ζωνοπερατών φίλτρων και στη συνέχεια υποδειγματοληπτώντας το αποτέλεσμα τους. Κάθε υποδειγματοληπτημένο σύνολο μετασχηματισμένων συντελεστών είναι μια υποζώνη η οποία κωδικοποιεί ένα συγκεκριμένο εύρος συχνοτήτων της εισόδου. Η ανακατασκευή επιτυγχάνεται με υπερδειγματοληψία ακολουθούμενη από τον υπολογισμό του αντίστροφου μετασχηματισμού και προσθέτοντας τα σύνολα του αποτελέσματος που προέρχονται από τις εξόδους των αντίστροφων φίλτρων. Το βασικό πλεονέκτημα των μετασχηματισμών υποζώνης είναι ο διαχωρισμός των διαφορετικών συχνοτήτων του σήματος εισόδου δίνοντας με τον τρόπο αυτό στον χρήστη τη δυνατότητα να ελέγχει με μεγάλη ακρίβεια την απώλεια δεδομένων για κάθε εύρος συχνοτήτων. Στην πράξη ένας μετασχηματισμός αυτής της μορφής αποδομεί την εικόνα σε πολλές υποζώνες οι οποίες αντιστοιχούν σε διαφορετικές συχνότητες της εικόνας και επομένως κάθε υποζώνη μπορεί να κβαντιστεί διαφορετικά. Το βασικό μειονέκτημα αυτής της μεθόδου μετασχηματισμού είναι η εισαγωγή παραμορφώσεων στην ανακατασκευασμένη εικόνα όπως το aliasing (δημιουργία προεξοχών σε λείες επιφάνειες ή καμπύλες) και το φαινόμενο του “ringing”

λόγω της υποδειγματοληψίας. Αυτός είναι και ο λόγος για τον οποίο ο μετασχηματισμός Haar δεν λειτουργεί ικανοποιητικά και αρκετή έρευνα έχει πραγματοποιηθεί στο πεδίο αυτό ώστε να βρεθούν καλύτερα σύνολα φίλτρων.



Σχήμα 7.23 : Μια Γενική Τράπεζα Φίλτρων

Στο Σχήμα 7.23 παρουσιάζεται η γενική δομή μιας τράπεζας φίλτρων υποζώνης με N ζωνοπερατά φίλτρα και τρείς βαθμίδες. Παρατηρούμε πώς η έξοδος του βαθυπερατού φίλτρου H_0 κάθε βαθμίδας δίνεται ως είσοδος στην επόμενη βαθμίδα για περεταίρω αποδόμηση καθώς και πώς η συνδυασμένη έξοδος της τράπεζας σύνθεσης μιας βαθμίδας, οδηγείται στο κορυφαίο αντίστροφο φίλτρο της τράπεζας σύνθεσης της προηγούμενης βαθμίδας.

7.4.3 Εύρεση των συντελεστών των φίλτρων

Αφού μελετήθηκε η βασική λειτουργία των τραπεζών φίλτρων, εκκρεμεί ένα σημαντικό ερώτημα σχετικά με τον τρόπο που υπολογίζουμε τους συντελεστές των φίλτρων. Δε θα εξετάσουμε αναλυτικά το συγκεκριμένο θέμα αλλά θα αναφερθούμε επιγραμματικά στους κανόνες και τις μεθόδους που χρησιμοποιούνται για την εύρεση των συντελεστών των τραπεζών φίλτρων.

Δοθέντος ενός συνόλου που αποτελείται από δύο κανονικά και δύο αντίστροφα φίλτρα των N συντελεστών, H_0, H_1και F_0, F_1 (με άρτιο N), συμβολίζουμε τους συντελεστές τους με:

$$h_0 = \big(h_0(0), h_0(1), \dots, h_0(N-1)\big), \qquad h_1 = \big(h_1(0), h_1(1), \dots, h_1(N-1)\big),$$

$$f_0 = \big(f_0(0), f_0(1), \dots, f_0(N-1)\big), \qquad f_1 = \big(f_1(0), f_1(1), \dots, f_1(N-1)\big)$$

Τα τέσσερα διανύσματα h_0, h_1, f_0 $και$ f_1 είναι οι κρουστικές αποκρίσεις των τεσσάρων φίλτρων. Το απλούστερο σύνολο συνθηκών το οποίο πρέπει να ικανοποιούν οι ποσότητες αυτές είναι το εξής:

1. *Κανονικοποίηση: Το διάνυσμα h_0 πρέπει να είναι κανονικοποιημένο (δηλαδή το άθροισμα όλων των τιμών του διανύσματος να ισούται με τη μονάδα).*
2. *Ορθογωνιότητα: Για κάθε ακέραιο i που ικανοποιεί την ανίσωση $1 \leq i < \frac{N}{2}$, το διάνυσμα που σχηματίζεται από τα πρώτα $2i$ στοιχεία του h_0 πρέπει να είναι ορθογώνιο στο διάνυσμα που σχηματίζεται από τα τελευταία $2i$ στοιχεία του h_0.*
3. *Το διάνυσμα f_0 να είναι το ανεστραμμένο h_0.*
4. *Το διάνυσμα h_1 είναι αντίγραφο της f_0 όπου τα πρόσημα των στοιχείων με περιττό αριθμό (πρώτο, τρίτο κλπ) έχουν αντιστραφεί.*
5. *Το διάνυσμα f_1 είναι αντίγραφο του h_0 όπου τα πρόσημα των στοιχείων με άρτιο αριθμό έχουν αντιστραφεί.*

Σε ένα φίλτρο δύο συντελεστών, ο πρώτος κανόνας συνεπάγεται

$$h_0^2(0) + h_0^2(1) = 1 \tag{7.11}$$

Ο δεύτερος κανόνας δεν μπορεί να εφαρμοστεί αφού $N = 2$, επομένως $i < N/2$ συνεπάγεται $i < 1$. Οι κανόνες τρία έως πέντε συνεπάγονται:

$$f_0 = \big(h_0(1), h_0(0)\big), \qquad h_1 = \big(-h_0(1), h_0(0)\big), \qquad f_1 = \big(h_0(0), -h_0(1)\big)$$

Όλα εξαρτώνται από τις τιμές των $h_0(0)$ και $h_0(1)$, όμως η εξίσωση $h_0^2(0) + h_0^2(1) = 1$ από μόνη της δεν είναι αρκετή για να τις καθορίσει. Ωστόσο, δεν είναι δύσκολο να δούμε ότι η επιλογή $h_0(0) = h_0(1) = 1/\sqrt{2}$ ικανοποιεί την εξίσωση.

Σε ένα φίλτρο τεσσάρων συντελεστών ο πρώτος και ο δεύτερος κανόνας συνεπάγονται:

$$h_0^2(0) + h_0^2(1) + h_0^2(2) + h_0^2(3) = 1, \qquad h_0(0)h_0(2) + h_0(1)h_0(3) = 0 \tag{7.12}$$

Από τους κανόνες τρία έως πέντε προκύπτει:

$$f_0 = \big(h_0(3), h_0(2), h_0(1), h_0(0)\big),$$

$$h_1 = \big(-h_0(3), h_0(2), -h_0(1), h_0(0)\big)$$

$$f_1 = \big(h_0(0), -h_0(1), h_0(2), -h_0(3)\big)$$

Προφανώς και πάλι η εξίσωση (7.12) δεν αρκεί για να προσδιοριστούν οι συντελεστές αλλά χρειάζονται επιπλέων θεωρήσεις προκειμένου να υπολογιστούν οι τέσσερεις συντελεστές τους οποίους βλέπουμε παρακάτω:

$$c_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \approx 0.48296, \qquad c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \approx 0.8365$$

$$c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \approx 0.2241, \qquad c_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \approx -0.1294$$

Η διαδικασία υπολογισμού των συντελεστών φίλτρων είναι εκτός της σκοπιάς αυτής της εργασίας και επομένως δε μελετηθεί περεταίρω ο τρόπος υπολογισμού τους.

7.5 Ο Διακριτός Μετασχηματισμός Κυματιδίου

Η πληροφορία η οποία δημιουργείται και αναλύεται καθημερινά στις εφαρμογές του πραγματικού κόσμου είναι διακριτή. Συνήθως πρόκειται για μεμονωμένους αριθμούς παρά συνεχείς συναρτήσεις. Αυτός είναι και ο λόγος που στην πράξη χρησιμοποιείται ο διακριτός μετασχηματισμός κυματιδίου[44]. Υπενθυμίζεται ότι ο συνεχής μετασχηματισμός κυματιδίου είναι το ολοκλήρωμα του γινομένου $f(t)\psi^*(\frac{t-b}{a})$, όπου a ο παράγοντας κλιμάκωσης και b η μετατόπιση στο χρόνο οι οποίοι μπορεί να είναι οποιοιδήποτε πραγματικοί αριθμοί. Ο αντίστοιχος υπολογισμός της διακριτής περίπτωσης εμπεριέχει τη διαδικασία της συνέλιξης, όμως η εμπειρία έχει δείξει ότι η ποιότητα των μετασχηματισμών αυτού του τύπου εξαρτώνται κυρίως από δύο παραμέτρους. Την επιλογή των συντελεστών κλιμάκωσης καθώς και την επιλογή του κυματιδίου.

Στην πράξη ο διακριτός μετασχηματισμός κυματιδίου υπολογίζεται με συντελεστές κλιμάκωσης οι οποίοι είναι αρνητικές τιμές του δύο και με μετατοπίσεις στο χρόνο οι οποίες είναι θετικές δυνάμεις του 2. Τα κυματίδια τα οποία χρησιμοποιούνται είναι τα κυματίδια εκείνα τα οποία σχηματίζουν ορθοκανονική βάση.

Ο απλούστερος τρόπος για να περιγράψει κανείς το διακριτό μετασχηματισμό κυματιδίου είναι με τη χρήση πολλαπλασιασμού πινάκων. Ο μετασχηματισμός Haar εξαρτάται από δύο συντελεστές φίλτρου c_0 και c_1 μαζί με την τιμή $\frac{1}{\sqrt{2}} \approx 0.7071$. Ο μικρότερος πίνακας μετασχηματισμού που μπορεί να κατασκευαστεί στην περίπτωση αυτή είναι ο $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}/\sqrt{2}$. Αυτός ο πίνακας μετασχηματισμού μεγέθους 2×2 δημιουργεί δύο συντελεστές μετασχηματισμού, έναν με τη μέση τιμή και έναν με τη διαφορά. Παρατηρούμε ότι δεν πρόκειται ακριβώς για τη μέση τιμή αφού χρησιμοποιείται ο παράγοντας $\sqrt{2}$ και όχι ο 2. Γενικά ο διακριτός μετασχηματισμός κυματιδίου μπορεί να χρησιμοποιήσει οποιοδήποτε σύνολο φίλτρων κυματιδίου αλλά ο υπολογισμός του πραγματοποιείται με τον ίδιο ακριβώς τρόπο ανεξάρτητα από το ποιο

φίλτρο θα χρησιμοποιηθεί.

Ένας από τους δημοφιλέστερους τύπους κυματιδίων είναι το Daubechies D4[43]. Όπως υποδηλώνει και το όνομα του, βασίζεται σε φίλτρο τεσσάρων συντελεστών c_0, c_1, c_2 και c_3, οι τιμές των οποίων εμφανίζονται στην εξίσωση (7.5). Ο πίνακας μετασχηματισμού είναι ο

$$W = \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & 0 & 0 & \cdots & 0 \\ c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & c_0 & c_1 & c_2 & c_3 & \cdots & 0 \\ 0 & 0 & c_3 & -c_2 & c_1 & -c_0 & \cdots & 0 \\ \vdots & \vdots & & & & \ddots & & \\ 0 & 0 & \cdots & 0 & c_0 & c_1 & c_2 & c_3 \\ 0 & 0 & \cdots & 0 & c_3 & -c_2 & c_1 & -c_0 \\ c_2 & c_3 & 0 & \cdots & 0 & 0 & c_0 & c_1 \\ c_1 & -c_0 & 0 & \cdots & 0 & 0 & c_3 & -c_2 \end{pmatrix}$$

Όταν ο πίνακας αυτός εφαρμοστεί σε ένα διάνυσμα στήλης με στοιχεία $(x_1, x_2, \dots, x_n)$, η πρώτη γραμμή δημιουργεί το σταθμισμένο άθροισμα $s_1 = c_0x_1 + c_1x_2 + c_2x_3 + c_3x_4$, η τρίτη γραμμή δημιουργεί το σταθμισμένο άθροισμα $s_2 = c_0x_3 + c_1x_4 + c_2x_5 + c_3x_6$ και ομοίως οι υπόλοιπες περιττές γραμμές δημιουργούν τα αντίστοιχα σταθμισμένα αθροίσματα s_i. Τα αθροίσματα αυτά είναι αποτέλεσμα της συνέλιξης των διανυσμάτων δεδομένων x_i με τους τέσσερεις συντελεστές φίλτρου. Στη γλώσσα των κυματιδίων κάθε ένα από αυτά τα αθροίσματα ονομάζεται συντελεστής λείανσης και όλοι μαζί αποτελούν ένα H φίλτρο λείανσης.

Αντίστοιχα, η δεύτερη γραμμή του πίνακα δημιουργεί την ποσότητα $d_1 = c_3x_1 - c_2x_2 + c_1x_3 - c_0x_4$, καθώς επίσης και όλες οι υπόλοιπες άρτιες γραμμές δημιουργούν αντίστοιχες συνελίξεις. Κάθε συντελεστής d_i ονομάζεται συντελεστής λεπτομέρειας και όλοι μαζί αποτελούν ένα G φίλτρο. Το G δεν είναι ένα φίλτρο λείανσης. Στην πράξη οι συντελεστές του φίλτρου έχουν επιλεγεί με τέτοιο τρόπο ώστε να προκύπτουν μικρές τιμές όταν τα δεδομένα του διανύσματος x_i είναι συσχετισμένα. Τα φίλτρα H και G μαζί ονομάζονται QMF (Quadrature Mirror Filters).

Ο διακριτός μετασχηματισμός κυματιδίου μίας εικόνας μπορεί να θεωρηθεί ως ένα πέρασμα της αρχικής εικόνας από ένα QMF το οποίο αποτελείται από ένα βαθυπερατό (H) και ένα υψηπερατό (G) φίλτρο. Εάν ο W είναι ένα πίνακας μεγέθους $n \times n$, δημιουργεί $n/2$ συντελεστές λείανσης s_i και $n/2$ συντελεστές λεπτομέρειας d_i. Ο ανεστραμμένος πίνακας είναι ο ακόλουθος:

$$W^T = \begin{pmatrix} c_0 & c_3 & 0 & 0 & \cdots & & & & c_2 & c_1 \\ c_1 & -c_2 & 0 & 0 & \cdots & & & & c_3 & -c_0 \\ c_2 & c_1 & c_0 & c_3 & \cdots & & & & 0 & 0 \\ c_3 & -c_0 & c_1 & -c_2 & \cdots & & & & 0 & 0 \\ & & & & \ddots & & & & & \\ & & & & c_2 & c_1 & c_0 & c_3 & 0 & 0 \\ & & & & c_3 & -c_0 & c_1 & -c_2 & 0 & 0 \\ & & & & & & c_2 & c_1 & c_0 & c_3 \\ & & & & & & c_3 & -c_0 & c_1 & -c_2 \end{pmatrix}$$

Αποδεικνύεται ότι προκειμένου ο πίνακας W να είναι ορθοκανονικός, οι τέσσερεις συντελεστές θα πρέπει να ικανοποιούν τις σχέσεις $c_0^2 + c_1^2 + c_2^2 + c_3^2 = 1$ και $c_2 c_0 + c_3 c_1 = 0$. Οι δύο εξισώσεις που ακολουθούν χρησιμοποιούνται ώστε να υπολογιστούν οι τέσσερεις συντελεστές του φίλτρου:

$$c_3 - c_2 + c_1 - c_0 = 0, \qquad 0c_1 - 1c_2 + 2c_1 - 3c_0 = 0$$

Οι λύση των εξισώσεων είναι η ακόλουθη:

$$c_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \approx 0.48296, \qquad c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \approx 0.8365$$
$$c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \approx 0.2241, \qquad c_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \approx -0.1294 \tag{7.13}$$

Η χρήση του πίνακα μετασχηματισμού W είναι θεωρητικά απλή αλλά όχι ιδιαιτέρως πρακτική αφού ο W θα πρέπει να έχει το ίδιο μέγεθος με την εικόνα, η οποία συνήθως είναι μεγάλη. Παρόλα αυτά, εάν κάποιος παρατηρήσει τη δομή του πίνακα W θα αντιληφθεί πως πρόκειται για έναν απολύτως ομοιόμορφο πίνακα με αποτέλεσμα να χρειαζόμαστε στην πραγματικότητα μόνο τους τέσσερεις συντελεστές για να υπολογίσουμε το μετασχηματισμό.

Οι δημοφιλέστεροι τύποι φίλτρων με τα χαρακτηριστικά τους είναι οι ακόλουθοι:

- Η οικογένεια φίλτρων Daubechies, μεγιστοποιεί τη λείανση του πατρικού κυματιδίου (συνάρτηση κλιμάκωσης) μεγιστοποιώντας τον ρυθμό αποδόμησης του μετασχηματισμού Fourier.
- Το κυματίδιο Haar μπορεί να θεωρηθεί ως ένα Daubechies φίλτρο 2^{ης} τάξης. Πρόκειται για το παλαιότερο φίλτρο. Είναι απλό στη χρήση του αλλά δεν παράγει τα καλύτερα αποτελέσματα καθώς δεν είναι συνεχές.
- Το φίλτρο Beylkin τοποθετεί τις ρίζες της συνάρτησης απόκρισης συχνότητας κοντά στη συχνότητα Nyquist στον πραγματικό άξονα.
- Το φίλτρο Caifman τάξης p (όπου p θετικός ακέραιος) το οποίο δίνει τόσο στο μητρικό όσο και στο πατρικό κυματίδιο $2p$ μηδενισμούς.

- Συμμετρικά φίλτρα είναι τα πιο συμμετρικά συμπαγή κυματίδια με το μέγιστο αριθμό μηδενισμών.
- Το φίλτρο Vaidynathan δεν ικανοποιεί καμία συνθήκη σε σχέση με τους μηδενισμούς όμως παρέχει ακριβή ανακατασκευή. Το φίλτρο αυτό είναι εξαιρετικά χρήσιμο στη συμπίεση ομιλίας.

7.6 Διάφοροι τύποι Αποδόμησης Εικόνας

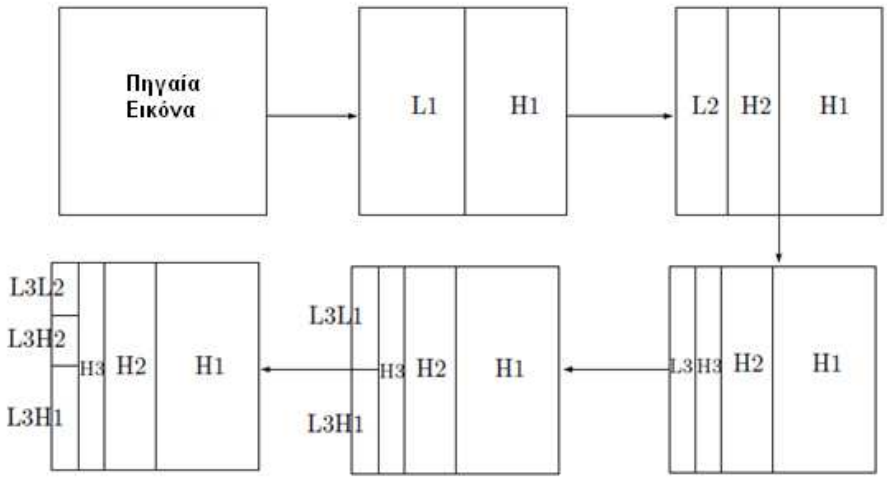
Στην ενότητα 7.3 μελετήθηκαν δύο τρόποι για την εφαρμογή του διακριτού μετασχηματισμού κυματιδίου σε μία εικόνα προκειμένου να διαχωριστεί σε πλήθος υποζωνών. Στη ενότητα αυτή θα εξετασθούν επτά τρόποι οι οποίοι πραγματοποιούν την ίδια διαδικασία, καθένας όμως με τη χρήση διαφορετικού αλγορίθμου και παράγοντας τελικά διαφορετικά αποτελέσματα σε ότι αφορά τις υποζώνες και τη συγκέντρωση της ενέργειας[49].

Αρχικά είναι σημαντικό να καταστεί σαφές ότι τα φίλτρα κυματιδίων και η μέθοδος αποδόμησης είναι ανεξάρτητα. Κατά τον διακριτό μετασχηματισμό κυματιδίου σε μία εικόνα μπορεί να χρησιμοποιηθεί οποιοδήποτε σύνολο φίλτρων κυματιδίου καθώς επίσης και να υπάρξει αποδόμηση της εικόνας με οποιοδήποτε μέθοδο. Ο μόνος περιορισμός σχετίζεται με την ύπαρξη αρκετών στοιχείων σε μία υποζώνη ώστε να καλυφθούν όλοι οι συντελεστές του φίλτρου. Για παράδειγμα εάν υποθέσουμε ότι χρησιμοποιούμε ένα φίλτρο τύπου Daubechies 12 συντελεστών και το μέγεθος υποζώνης είναι δύναμη του δύο, τότε η μικρότερη υποζώνη που μπορεί να παραχθεί θα έχει μέγεθος 8×8. Αυτό συμβαίνει γιατί μια υποζώνη μεγέθους 16×16 είναι η μικρότερη η οποία μπορεί να πολλαπλασιαστεί με τους 12 συντελεστές του φίλτρου. Εάν μια υποζώνη αυτού του μεγέθους αποδομηθεί, το αποτέλεσμα θα είναι μια υποζώνη μεγέθους 8×8 η οποία όμως δεν είναι δυνατόν να πολλαπλασιαστεί με τους 12 συντελεστές του φίλτρου ώστε να αποδομηθεί περαιτέρω.

1. Η Πυραμίδα του Laplace: Το βασικό χαρακτηριστικό της μεθόδου είναι η προοδευτική μετάδοση της εικόνας. Κατά την ανακατασκευή της εικόνας, ο χρήστης διακρίνει μικρές, θολές εικόνες οι οποίες μεγαλώνουν και γίνονται όλο και καθαρότερες. Η πυραμίδα του Laplace δημιουργείται αφαιρώντας μια υπέρ-δειγματοληπτημένη χαμηλοπερατή έκδοση της εικόνας από την αρχική εικόνα. Η εικόνα διαχωρίζεται σε μία Gaussian πυραμίδα (τις χαμηλοπερατές υποζώνες) και μία Laplacian πυραμίδα η οποία αποτελείται από συντελεστές λεπτομέρειας (τις υψηπερατές υποζώνες). Μόνο η Laplacian πυραμίδα χρειάζεται για την ανακατασκευή της εικόνας. Η μετασχηματισμένη εικόνα είναι μεγαλύτερη από την αρχική και αυτή είναι η βασική διαφορά της μεθόδου από την μέθοδο αποδόμησης της πυραμίδας στην οποία η οποία θα αναφερθεί στη συνέχεια.

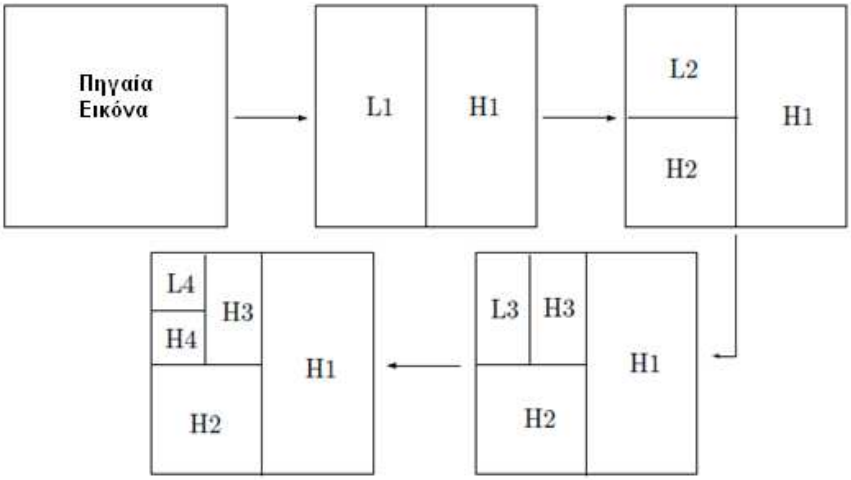
2. Μέθοδος Αποδόμησης Γραμμής: Η μέθοδος αυτή αποτελεί μία απλουστευμένη εκδοχή της τυπικής αποδόμησης κυματιδίου. Ο μετασχηματισμός κυματιδίου εφαρμόζεται σε κάθε γραμμή της εικόνας, γεγονός που έχει ως αποτέλεσμα τους συντελεστές λείανσης στα αριστερά (υποζώνη L_1) και τους συντελεστές λεπτομέρειας στα δεξιά (υποζώνη H_1). Η υποζώνη L_1 στη συνεχεία διαχωρίζεται σε L_2 και H_2 και η διαδικασία επαναλαμβάνεται έως ότου όλος ο πίνακας συντελεστών να αποτελείται μόνο από συντελεστές λεπτομέρειας, εκτός από τους συντελεστές της αριστερής στήλης οι οποίοι θα είναι συντελεστές λείανσης. Στη συνέχεια ο μετασχηματισμός κυματιδίου εφαρμόζεται αναδρομικά στην αριστερή στήλη με αποτέλεσμα να παραμείνει τελικά μόνο ένας συντελεστής λείανσης στην επάνω αριστερή γωνία του πίνακα. Το τελευταίο βήμα μπορεί να παραληφθεί εάν η μέθοδος συμπίεσης η οποία χρησιμοποιείται απαιτεί οι γραμμές τις εικόνας να συμπιέζονται ξεχωριστά. Η τεχνική αυτή εκμεταλλεύεται τη συσχέτιση μεταξύ των στοιχείων μόνο κατά γραμμές προκειμένου να υπολογιστούν οι συντελεστές μετασχηματισμού. Επίσης η παράληψη ενός συντελεστή που βρίσκεται στην αριστερή στήλη μπορεί να επηρεάσει σημαντικά την ποιότητα της ανακατασκευασμένης εικόνας στη γραμμή αυτή.

Η υλοποίηση αυτής της μεθόδου είναι απλή και η εκτέλεση της περίπου δύο φορές γρηγορότερη από την μέθοδο τυπικής αποδόμησης. Ο τύπος αποδόμησης αυτής της μορφής παρουσιάζεται στο Σχήμα 7.24. Είναι δυνατό η αποδόμηση να πραγματοποιηθεί στις στήλες της εικόνας αντί για τις γραμμές. Ιδανικά ο μετασχηματισμός πρέπει να εφαρμόζεται προς την κατεύθυνση που υπάρχει μεγαλύτερη συσχέτιση των εικονοστοιχείων. Η εμπειρία έχει δείξει ότι συνήθως στις φυσικές εικόνες, η κατεύθυνση αυτή είναι η οριζόντια, για το λόγο αυτό στην πράξη η αποδόμηση που προτιμάται είναι κατά γραμμές.



Σχήμα 7.24 : Μέθοδος Αποδόμησης Γραμμής

3. Quincunx: Παρόμοια με την πυραμίδα του Laplace, η αποδόμηση Quincunx προβαίνει σε μια αποδόμηση επίπεδο προς επίπεδο, αποδομώντας την υποζώνη L_i του επιπέδου i σε υποζώνες H_{i+1} και L_{i+1} του επιπέδου $i+1$. Το Σχήμα 7.25 παρουσιάζει τη μορφή αυτή αποδόμησης. Τη μέθοδο αυτή αρχικά την αποκαλούσαν μη τυπική μέθοδο αποδόμησης. Πρόκειται για μία αποδοτική μέθοδο και υπολογιστικά απλή. Κατά μέσο όρο επιτυγχάνει τέσσερεις φορές μεγαλύτερη συγκέντρωση ενέργειας από τη μέθοδο αποδόμησης γραμμής. Η μέθοδος Quincunx δημιουργεί λιγότερες υποζώνες από τις περισσότερες μεθόδους αποδόμησης με κυματίδια γεγονός που ενδεχομένως να οδηγήσει σε ανακατασκευή εικόνων χαμηλότερης οπτικής ποιότητας. Στην πράξη δεν χρησιμοποιείται συχνά όμως σύμφωνα με πειραματικά αποτελέσματα η μέθοδος αυτή επιτυγχάνει εξαιρετικά αποτελέσματα και θα μπορούσε να χρησιμοποιηθεί σε πλήθος πρακτικών εφαρμογών.

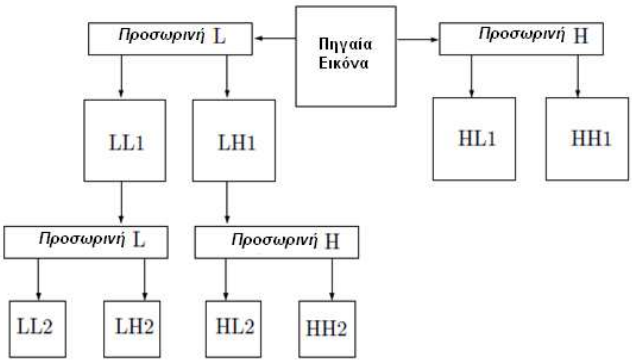


Σχήμα 7.25 : Μέθοδος Αποδόμησης Quincunx

4. Πυραμίδα: Η αποδόμηση πυραμίδας αποτελεί μακράν τη συνηθέστερη μέθοδο αποδόμησης εικόνας με χρήση μετασχηματισμού κυματιδίων. Η εφαρμογή τους έχει ως αποτέλεσμα τη δημιουργία υποζωνών με οριζόντιες, κατακόρυφες και διαγώνιες λεπτομέρειες εικόνων όπως παρουσιάζεται στο Σχήμα 7.26. Οι τρείς υποζώνες σε κάθε επίπεδο περιέχουν τα οριζόντια, κάθετα και διαγώνια χαρακτηριστικά της εικόνας σε συγκεκριμένη κλίμακα και κάθε κλίμακα διαιρείται με μία οκτάβα στο πεδίο συχνοτήτων (διαίρεση της συχνότητας με το 2).
   Η αποδόμηση πυραμίδας συμπεραίνεται πως αποτελεί μια πολύ αποδοτική μέθοδο για τη μεταφορά σημαντικής οπτικής πληροφορίας στους συντελεστές λεπτομέρειας. Η υπολογιστική πολυπλοκότητα είναι 30% μεγαλύτερη από της μεθόδου Quincunx όμως η ανακατασκευή της εικόνας φαίνεται να είναι ακριβέστερη ως προς την αρχική. Η συγκεκριμένη μέθοδος είναι τόσο δημοφιλής εξαιτίας της συμμετρίας της, της απλής μαθηματικής της περιγραφής και της ευρείας χρήσης σε δημοσιεύσεις με μεγάλη επιρροή στην επιστημονική

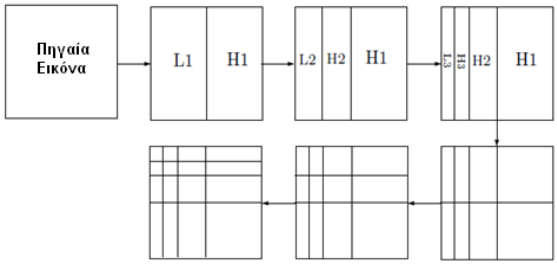
κοινότητα.

Στο Σχήμα 7.26 παρουσιάζεται η αποδόμηση πυραμίδας. Είναι προφανές πως το πρώτο βήμα είναι όμοιο με αυτό της μεθόδου αποδόμησης Quincunx. Παρόλα αυτά, ενώ η μέθοδος Quincunx αφήνει την υποζώνη υψηλής συχνότητας ανέγγιχτη, η μέθοδος πυραμίδας την αναλύει σε δύο ζώνες. Από την άλλη, η αποδόμηση πυραμίδας έχει μεγαλύτερο υπολογιστικό κόστος προκειμένου να αναλύσει την ασύμμετρη ζώνη υψηλής συχνότητας σε δύο συμμετρικές συχνοτικές ζώνες, μία υψηλή και μία χαμηλή.



Σχήμα 7.26 : Μέθοδος Αποδόμησης Πυραμίδας

5. Τυπική Μέθοδος Αποδόμησης: Το πρώτο βήμα κατά την τυπική μέθοδο αποδόμησης είναι η εφαρμογή του διακριτού φίλτρου κυματιδίου σε κάθε γραμμή της εικόνας ώστε να προκύψουν οι υποζώνες L_1 και H_1. Αυτό επαναλαμβάνεται στην L_1 ώστε να προκύψουν οι L_2 και H_2 και ούτω καθεξής k φορές. Η διαδικασία αυτή ακολουθείται από ένα δεύτερο βήμα όπου αντίστοιχος υπολογισμός πραγματοποιείται και για τις στήλες. Εφόσον το $k = 1$, η αποδόμηση εναλλάσσεται μεταξύ γραμμών και στηλών, όμως το k ενδεχομένως να είναι μεγαλύτερο. Το τελικό αποτέλεσμα είναι η ύπαρξη συντελεστών λείανσης στην επάνω αριστερή γωνία του πίνακα συντελεστών. Η μέθοδος αυτή είναι παρόμοια με την μέθοδο αποδόμησης γραμμής. Ένα σημαντικό χαρακτηριστικό της τυπικής αποδόμησης είναι το γεγονός ότι όταν οι συντελεστές κβαντιστούν ενδέχεται να επηρεαστούν οι μακριές και λεπτές ορθογώνιες περιοχές της ανακατασκευασμένης εικόνας. Συνεπώς ισχυρή κβάντιση ενδεχομένως να δημιουργήσει οπτικές ατέλειες στα οριζόντια ορθογώνια της ανακατασκευασμένης εικόνας. Η τυπική αποδόμηση έχει τη δεύτερη καλύτερη ποιότητα ανακατασκευής μεταξύ των επτά μεθόδων που περιγράφονται. Ο λόγος για τον οποίο παρατηρείται βελτίωση σε σχέση με τη μέθοδο πυραμίδας πιθανώς να οφείλεται στο γεγονός ότι η υψηλότερη κατευθυντική ανάλυση μπορεί να καλύψει αποδοτικότερα μεγαλύτερες, ομοιόμορφες περιοχές. Πρέπει να σημειωθεί ότι η μέθοδος αυτή έχει υπολογιστικά μεγαλύτερο κόστος από τη μέθοδο πυραμίδας. Η μέθοδος παρουσιάζεται στο Σχήμα 7.27



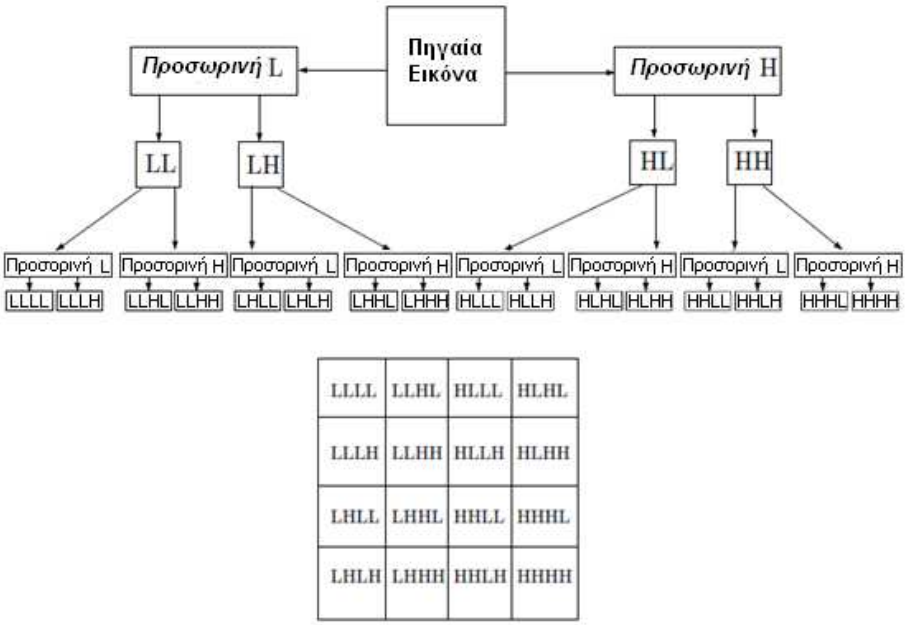
Σχήμα 7.27 : Τυπική Μέθοδος Αποδόμισης

6. Ομοιόμορφη Αποδόμηση: Η μέθοδος ομοιόμορφης αποδόμησης ονομάζεται επίσης και μετασχηματισμός πακέτου κυματιδίου και παρουσιάζεται στο Σχήμα 7.28. Στην περίπτωση όπου ο χρήστης επιλέγει να υπολογίσει όλα τα επίπεδα μετασχηματισμού, η ομοιόμορφη αποδόμηση γίνεται όμοια με τον διακριτό μετασχηματισμό Fourier σύμφωνα με τον οποίο κάθε συντελεστής αναπαριστά μία φασματική συχνότητα σε ολόκληρη την εικόνα. Σε μια τέτοια περίπτωση, η αφαίρεση ενός συντελεστή της μετασχηματισμένης εικόνας επηρεάζει ολόκληρη την ανακατασκευασμένη εικόνα.

   Το υπολογιστικό κόστος της ομοιόμορφης αποδόμησης είναι πολύ υψηλό καθώς υπολογίζει n^2 συντελεστές για κάθε επίπεδο αποδόμησης, όπου n είναι το πλήθος των γραμμών της (τετραγωνικής) εικόνας. Παρά το γεγονός ότι η ποιότητα της ανακατασκευασμένης εικόνας είναι κατά μέσο όρο υψηλή, η οπτική ποιότητα της εικόνας υποβαθμίζεται γρηγορότερα σε ότι αφορά την κβάντιση, σε σχέση με άλλες μεθόδους αποδόμησης. Ο λόγος που συμβαίνει αυτό είναι ο ίδιος με ότι συμβαίνει στο μετασχηματισμό Fourier. Από τη στιγμή που ένας συντελεστής έχει καθολική επίδραση, η αφαίρεση ή ισχυρή κβάντιση του έχει ως αποτέλεσμα τη θολερότητα της ανακατασκευασμένης εικόνας σε κάποιο βαθμό. Συμπερασματικά, αναφέρεται πως η αυξημένη υπολογιστική πολυπλοκότητα της μεθόδου δε συνεπάγεται και αντίστοιχη αύξηση της οπτικής ποιότητας της ανακατασκευασμένης εικόνας.

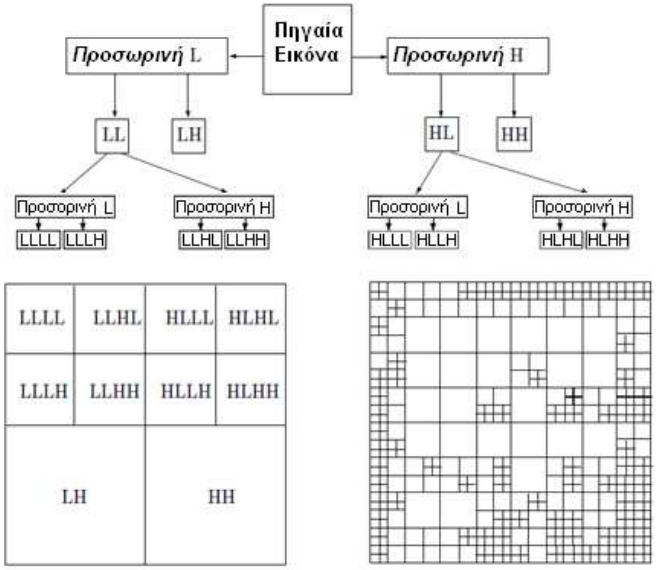
   Μία ειδική περίπτωση της ομοιόμορφης αποδόμησης είναι η πλήρης αποδόμηση κυματιδίου. Θεωρούμε την αρχική εικόνα ως I_0 με μέγεθος $2^l \times 2^l$. Αφού εφαρμοστεί ο διακριτός μετασχηματισμός κυματιδίου προκύπτει η εικόνα I_1 η οποία έχει διαχωριστεί σε τέσσερεις υποζώνες. Ο ίδιος διακριτός μετασχηματισμός κυματιδίου εφαρμόζεται αναδρομικά σε κάθε μία από τις τέσσερεις υποζώνες. Η διαδικασία αυτή έχει ως αποτέλεσμα τον πίνακα I_2 ο οποίος αποτελείται από 16 υποζώνες. Αφού η διαδικασία αυτή επαναληφθεί r φορές, το αποτέλεσμα είναι ένας πίνακας συντελεστών αποτελούμενος από $2^r \times 2^r$ υποζώνες, κάθε μία μεγέθους $2^{l-r} \times 2^{l-r}$. Η επάνω-αριστερή υποζώνη

περιέχει τους συντελεστές λείανσης (ανάλογα με το κυματίδιο που έχει χρησιμοποιηθεί ενδέχεται να μοιάζει με μία μικρή εκδοχή της αρχικής εικόνας) ενώ οι υπόλοιπες υποζώνες περιέχουν τους συντελεστές λεπτομέρειας. Κάθε υποζώνη αντιστοιχεί σε μία ζώνη συχνοτήτων ενώ κάθε συντελεστής αντιστοιχεί σε μία τοπική χωρική περιοχή. Αυξάνοντας το βάθος αναδρομής r υπάρχει η δυνατότητα αύξησης της συχνοτικής ανάλυσης σε βάρος της χωρικής ανάλυσης. Ο τύπος αυτός αποδόμησης δεν χρησιμοποιείται ιδιαίτερα στον χώρο της συμπίεσης εικόνας.



Σχήμα 7.28 : Ομοιόμορφη Μέθοδος Αποδόμησης

7. Προσαρμοστική Αποδόμηση Πακέτου Κυματιδίων: Η μέθοδος ομοιόμορφης αποδόμησης έχει μεγάλο υπολογιστικό κόστος και η προσαρμοστική αποδόμηση πακέτου κυματιδίων ενδέχεται να έχει ακόμα μεγαλύτερο. Κεντρική ιδέα της μεθόδου αποτελεί η παράλειψη των υποζωνών εκείνων που δε συνεισφέρουν σημαντικά από την άποψη της συγκέντρωσης ενέργειας. Το αποτέλεσμα είναι ένας πίνακας συντελεστών αποτελούμενος από υποζώνες διαφορετικών μεγεθών. Στο Σχήμα 7.29 παρουσιάζεται αυτή ακριβώς η μέθοδος αποδόμησης.



Σχήμα 7.29 : Προσαρμοστική Μέθοδος Αποδόμησης Πακέτου Κυματιδίων

Το βασικό πρόβλημα αποδόμησης αυτής της μορφής σχετίζεται με την εύρεση ενός αποδοτικού αλγορίθμου ο οποίος θα ανιχνεύει ποιες διαιρέσεις υποζώνης θα μπορούσαν να παραληφθούν. Ένας τέτοιος αλγόριθμος θα ήταν πολύ αποδοτικός. Αντίθετα ένας αλγόριθμος ο οποίος διαιρεί κάθε υποζώνη, θα εκτελέσει πλήθος υπολογισμών οι οποίοι δεν είναι απαραίτητοι, καταλήγοντας σε έναν πίνακα όπου κάθε συντελεστής θα είναι και μια υποζώνη όπου ουσιαστικά οδηγούμαστε στη μέθοδο ομοιόμορφης αποδόμησης. Η μέθοδος αυτή παρέχει την υψηλότερη ποιότητα αναπαραγωγής σε σχέση με τις υπόλοιπες μεθόδους, γεγονός που δικαιολογεί το αυξημένο υπολογιστικό κόστος σε ορισμένες εξειδικευμένες εφαρμογές. Παρόλα αυτά η ποιότητα δεν είναι πολύ καλύτερη σε σχέση με αυτή που επιτυγχάνεται από άλλες μεθόδους όπως την τυπική μέθοδο, την μέθοδο πυραμίδας ή τη μέθοδο Quincunx.

7.7 Ο Ακέραιος Μετασχηματισμός Κυματιδίου

Ο διακριτός μετασχηματισμός κυματιδίου είναι απλός όμως έχει ένα σημαντικό μειονέκτημα. Δεν χρησιμοποιεί ακέραιους συντελεστές φίλτρου, με αποτέλεσμα να δημιουργούνται δεκαδικοί συντελεστές μετασχηματισμού. Υπάρχουν αρκετοί τρόποι προκειμένου να τροποποιηθεί ο βασικός διακριτός μετασχηματισμός κυματιδίου ούτως ώστε να δημιουργούνται ακέραιοι συντελεστές μετασχηματισμού. Στην ενότητα αυτή θα περιγραφεί ένας απλός ακέραιος μετασχηματισμός κυματιδίου ο οποίος μπορεί να χρησιμοποιηθεί με κάθε μέθοδο αποδόμησης που περιγράφηκε νωρίτερα. Ο μετασχηματισμός είναι αντιστρέψιμος, συνεπώς κάθε εικόνα μπορεί να

ανακατασκευαστεί πλήρως από τους ακέραιους συντελεστές μετασχηματισμού. Ο μετασχηματισμός αυτός μπορεί να χρησιμοποιηθεί είτε για συμπίεση εικόνας με απώλειες (με κβάντιση των συντελεστών) είτε χωρίς απώλειες. Η αρχή του μετασχηματισμού είναι απλή και θα παρουσιαστεί για τη μονοδιάστατη περίπτωση. Δοθέντος ενός διανύσματος δεδομένων N ακεραίων x_i, όπου $i = 0,1, \dots, N-1$, ορίζεται $k = N/2$ και προκύπτει το διάνυσμα μετασχηματισμού y_i υπολογίζοντας τα άρτια και τα περιττά στοιχεία ξεχωριστά. Αρχικά εξετάζεται η περίπτωση όπου το N είναι άρτιος. Τα $N/2$ περιττά στοιχεία y_{2i+1}, όπου $i = 0,1, \dots, k-1$ υπολογίζονται ως διαφορές των x_i . Οι συντελεστές αυτοί χαρακτηρίζουν τις υψηλές συχνότητες, δηλαδή αποτελούν τους συντελεστές λεπτομέρειας του μετασχηματισμού. Κάθε ένα από τα στοιχεία y_{2i}, όπου $0 \le i \le k-1$ υπολογίζεται ως ο σταθμισμένος μέσος όρος των πέντε στοιχείων x_i . Αυτοί οι $N/2$ συντελεστές χαρακτηρίζουν τις χαμηλές συχνότητες της εικόνας και συνήθως μετασχηματίζονται ξανά σε $N/4$ συντελεστές χαμηλών συχνοτήτων και $N/4$ συντελεστές υψηλών συχνοτήτων.

Ο βασικός κανόνας για τον μετασχηματισμό των περιττών συντελεστών είναι:

$$y_{2i+1} = -\frac{1}{2}x_{2i} + x_{2i+1} - \frac{1}{2}x_{2i+2}$$

εκτός του τελευταίου συντελεστή, όπου $i = k-1$ ο οποίος υπολογίζεται ως η απλή διαφορά:

$$y_{2k-1} = x_{2k-1} - x_{2k-2}$$

Ο μετασχηματισμός συνεπώς συνοψίζεται ως:

$$y_{2i+1} = \begin{cases} x_{2i+1} - \dfrac{x_{2i} + x_{2i+2}}{2}, & \gamma\iota\alpha\ i = 0,1, \dots, k-2 \\ x_{2i+1} - x_{2i}\ , & \gamma\iota\alpha\ i = k-1 \end{cases} \tag{7.14}$$

Οι άρτιοι συντελεστές υπολογίζονται ως ο σταθμισμένος μέσος όρος:

$$y_{2i} = -\frac{1}{8}x_{2i-2} + \frac{1}{4}x_{2i-1} + \frac{3}{4}x_{2i} + \frac{1}{4}x_{2i+1} - \frac{1}{8}x_{2i+2}$$

εκτός του πρώτου συντελεστή όπου $i = 0$ ο οποίος υπολογίζεται ως:

$$y_0 = \frac{3}{4}x_0 + \frac{1}{2}x_1 - \frac{1}{4}x_2$$

Στην πράξη ο υπολογισμός αυτός πραγματοποιείται υπολογίζοντας κάθε άρτιο συντελεστή y_{2i} με τη βοήθεια του x_{2i} και των δύο περιττών συντελεστών y_{2i-1} και y_{2i+1}. Συνοψίζοντας έχουμε:

$$y_{2i} = \begin{cases} x_{2i} + \dfrac{y_{2i+1}}{2}, & \gamma\iota\alpha\ i = 0 \\ x_{2i} + \dfrac{y_{2i-1} + y_{2i+1}}{4}, & \gamma\iota\alpha\ i = 1,2,\dots,k-1 \end{cases} \tag{7.15}$$

Ο αντίστροφος μετασχηματισμός γίνεται εύκολα αντιληπτός και χρησιμοποιεί τους συντελεστές y_i για να υπολογίσει τα δεδομένα z_i τα οποία θα είναι ίδια με το αρχικό διάνυσμα x_i.

Για τα άρτια στοιχεία ισχύει ότι:

$$z_{2i} = \begin{cases} y_{2i} - \dfrac{y_{2i+1}}{2}, & \gamma\iota\alpha\ i = 0 \\ y_{2i} - \dfrac{y_{2i-1} + y_{2i+1}}{4}, & \gamma\iota\alpha\ i = 1,2,\dots,k-1 \end{cases} \tag{7.16}$$

Για τα περιττά ισχύει ότι:

$$z_{2i+1} = \begin{cases} y_{2i+1} + \dfrac{z_{2i} + z_{2i+2}}{2}, & \gamma\iota\alpha\ i = 0,1,\dots,k-2 \\ y_{2i+1} + z_{2i}, & \gamma\iota\alpha\ i = k-1 \end{cases} \tag{7.17}$$

Παρατηρώντας τις εξισώσεις συμπεραίνουμε πως οι συντελεστές που προκύπτουν γενικά δεν είναι ακέραιοι καθώς υπάρχουν διαιρέσεις με το 2 και το 4. Το ίδιο ισχύει και για την ανακατασκευή των δεδομένων. Το ενδιαφέρον στον μετασχηματισμό αυτό το οποίο αποτελεί και το βασικό του χαρακτηριστικό είναι η χρήση της στρογγυλοποίησης προς τα κάτω (truncation). Η στρογγυλοποίηση προς τα κάτω σημειώνεται με τα σύμβολα $\lfloor \kappa\alpha\iota \rfloor$ η οποία χρησιμοποιείται για να παραχθούν οι συντελεστές μετασχηματισμού y_i καθώς και η ανακατασκευή των δεδομένων z_i. Συνεπώς, η τελική μορφή των εξισώσεων είναι η ακόλουθη:

$$
\begin{aligned}
y_{2i+1} &= \begin{cases} x_{2i+1} - \left\lfloor \dfrac{x_{2i} + x_{2i+2}}{2} \right\rfloor, & \gamma\iota\alpha\ i = 0,1,\dots,k-2 \\ x_{2i+1} - x_{2i}\ , & \gamma\iota\alpha\ i = k-1 \end{cases} \\
y_{2i} &= \begin{cases} x_{2i} + \left\lfloor \dfrac{y_{2i+1}}{2} \right\rfloor, & \gamma\iota\alpha\ i = 0 \\ x_{2i} + \left\lfloor \dfrac{y_{2i-1} + y_{2i+1}}{4} \right\rfloor, & \gamma\iota\alpha\ i = 1,2,\dots,k-1 \end{cases} \\
z_{2i} &= \begin{cases} y_{2i} - \left\lfloor \dfrac{y_{2i+1}}{2} \right\rfloor, & \gamma\iota\alpha\ i = 0 \\ y_{2i} - \left\lfloor \dfrac{y_{2i-1} + y_{2i+1}}{4} \right\rfloor, & \gamma\iota\alpha\ i = 1,2,\dots,k-1 \end{cases} \\
z_{2i+1} &= \begin{cases} y_{2i+1} + \left\lfloor \dfrac{z_{2i} + z_{2i+2}}{2} \right\rfloor, & \gamma\iota\alpha\ i = 0,1,\dots,k-2 \\ y_{2i+1} + z_{2i}, & \gamma\iota\alpha\ i = k-1 \end{cases}
\end{aligned}
\tag{7.18}
$$

Εξαιτίας της στρογγυλοποίησης προς τα κάτω, τμήμα της πληροφορίας χάνεται κατά τον υπολογισμό του y_i. Παρόλα αυτά η στρογγυλοποίηση προς τα κάτω χρησιμοποιείται και κατά τον υπολογισμό του z_i το οποίο αποκαθιστά την απώλεια της πληροφορίας. Συνεπώς ο ακέραιος μετασχηματισμός κυματιδίου είναι ένας πραγματικά αντιστρέψιμος μετασχηματισμός ο οποίος ανακατασκευάζει τα αρχικά δείγματα επακριβώς.

8. Το Πρότυπο Συμπίεσης JPEG2000

Μετά την ολοκλήρωση του πολύ επιτυχημένου προτύπου συμπίεσης JPEG η επιτροπή φωτογραφίας προχώρησε στη δημιουργία ενός νέου προτύπου συμπίεσης, γνωστό ως JPEG 2000. Το πρότυπο JPEG 2000[50] υποστηρίζει συμπίεση με και χωρίς απώλειες για εικόνες ενός ή και περισσοτέρων καναλιών (δηλαδή εικόνες με αποχρώσεις του γκρι ή έγχρωμες). Εκτός από τις βασικές λειτουργίες συμπίεσης, υποστηρίζεται ένα πλήθος άλλων δυνατοτήτων, συμπεριλαμβανομένου:

1) προοδευτική αποκατάσταση της εικόνας βάσει ευκρίνειας ή ανάλυσης
2) Κωδικοποίηση της εικόνας βάσει περιοχής ενδιαφέροντος, δηλαδή κωδικοποιώντας διάφορα τμήματα της εικόνας με διαφορετική ευκρίνεια.
3) Τυχαία προσπέλαση σε συγκεκριμένες περιοχές της εικόνας χωρίς να απαιτείται αποκωδικοποίηση ολόκληρης της εικόνας.
4) Ευέλικτο τύπο δεδομένων με προβλέψεις σχετικά με την αδιαφάνεια της πληροφορίας και της ακολουθίας των εικόνων.
5) Ανοχή στα σφάλματα.

Λόγω της εξαιρετικής απόδοσης στην κωδικοποίηση και των επιπλέων πολύ καλών δυνατοτήτων που παρέχονται από τον κωδικοποιητή, το JPEG 2000 έχει ένα πολύ μεγάλο εύρος εφαρμογών στο οποίο μπορεί να αξιοποιηθεί. Κάποιες από τις περιοχές αυτές είναι η αρχειοθέτηση των εικόνων, το διαδίκτυο, η αναζήτηση στο δίκτυο, οι απεικονίσεις εγγράφων, η ψηφιακή φωτογραφία, οι ιατρικές φωτογραφίες και η τηλεπισκόπηση.

Οι λόγοι για τους οποίους η επιτροπή JPEG θέλησε να προτείνει ένα νέο πρότυπο ήταν αφενός για την λύση κάποιων γνωστών αδυναμιών του προτύπου JPEG και αφετέρου για να προσφέρει δυνατότητες οι οποίες δεν υπήρχαν στο πρότυπο JPEG. Συνεπώς, το νέο πρότυπο έπρεπε να επιτρέπει την αποδοτική συμπίεση με ή χωρίς απώλειες μέσα από ένα ενιαίο πλαίσιο εργασίας, να παρέχει εξαιρετική ποιότητα εικόνας τόσο με αντικειμενικά όσο και με υποκειμενικά κριτήρια σε χαμηλούς ρυθμούς δεδομένων, να υποστηρίζει επιπλέον δυνατότητες για κλιμάκωση ρυθμού και ανάλυσης, κωδικοποίηση ανάλογα με την περιοχή ενδιαφέροντος και έναν πιο ευέλικτο τύπο δεδομένων. Τέλος θα έπρεπε να αποφευχθεί η υπερβολική πολυπλοκότητα τόσο σε υπολογιστικό όσο και σε χωρητικό επίπεδο (μνήμη).

Στο κεφάλαιο αυτό θα εξεταστούν οι λεπτομέρειες του JPEG 2000. Βασικός στόχος του θα είναι η γενική κατανόηση του προτύπου καθώς και των καινοτομιών που ενσωματώνει.

8.1 Στοιχεία Χρώματος

Μία έγχρωμη εικόνα αποτελείται από τρία στοιχεία χρώματος. Το πρώτο βήμα του κωδικοποιητή JEPG 2000 είναι ο μετασχηματισμός των στοιχείων αυτών είτε με έναν αντιστρέψιμο είτε με έναν μη αντιστρέψιμο μετασχηματισμό. Κάθε μετασχηματισμένο στοιχείο συμπιέζεται στη συνέχεια ξεχωριστά. Εάν οι τιμές των εικονοστοιχείων είναι μη προσημασμένοι αριθμοί (στη συνήθη περίπτωση), τότε του μετασχηματισμού προηγείται μια ολίσθηση των τιμών από το διάστημα $[0,2^s-1]$ (όπου s το βάθος χρώματος) στο προσημασμένο διάστημα $[-2^{s-1},2^{s-1}-1]$ αφαιρώντας 2^{s-1} από κάθε τιμή. Έτσι για παράδειγμα εάν θεωρήσουμε την περίπτωση $s=4$ τότε υπάρχουν $2^4=16$ πιθανοί συνδυασμοί τιμών των εικονοστοιχείων οι οποίες θα ολισθήσουν από το διάστημα $[0,15]$ στο διάστημα $[-8,+7]$ αφαιρώντας την τιμή $2^{4-1}=8$ από κάθε τιμή.

Ο αντιστρέψιμος μετασχηματισμός είναι ένας μετασχηματισμός αποσυσχέτισης. Μπορεί να χρησιμοποιηθεί μόνο με τον ακέραιο μετασχηματισμό κυματιδίου (ο οποίος είναι αντιστρέψιμος). Σημειώνοντας τις τιμές φωτεινότητας των εικονοστοιχείων του στοιχείου της εικόνας i (μετά την πιθανή ολίσθηση) με $I_i(x,y)$ για $i=0,1\ \kappa\alpha\iota\ 2$, ο αντιστρέψιμος μετασχηματισμός παράγει τις νέες τιμές $Y_i(x,y)$ σύμφωνα με τις σχέσεις:

$$Y_0(x,y)=\left\lfloor\frac{I_0(x,y)+2I_1(x,y)+I_2(x,y)}{4}\right\rfloor,$$

$$Y_1(x,y)=I_2(x,y)-I_1(x,y),$$

$$Y_2(x,y)=I_0(x,y)-I_1(x,y).$$

Σημειώνεται ότι οι τιμές των στοιχείων Y_1 και Y_2 (αλλά όχι του Y_0) απαιτούν ένα bit περισσότερο σε σχέση με τις αρχικές τιμές I_i.

Ο μη αντιστρέψιμος μετασχηματισμός είναι και αυτός ένας μετασχηματισμός αποσυσχέτισης. Χρησιμοποιείται μόνο με το μετασχηματισμό κυματιδίου κινητής υποδιαστολής (ο οποίος είναι μη αντιστρέψιμος). Ο μη αντιστρέψιμος μετασχηματισμός υπολογίζεται σύμφωνα με τις σχέσεις:

$$Y_0(x,y)=0.299I_0(x,y)+0.587I_1(x,y)+0.144I_2(x,y),$$

$$Y_1=-0.16875I_0(x,y)-0.33126I_1(x,y)+0.5I_2(x,y),$$

$$Y_2(x,y)=0.5I_0(x,y)-0.41869I_1(x,y)-0.08131I_2(x,y).$$

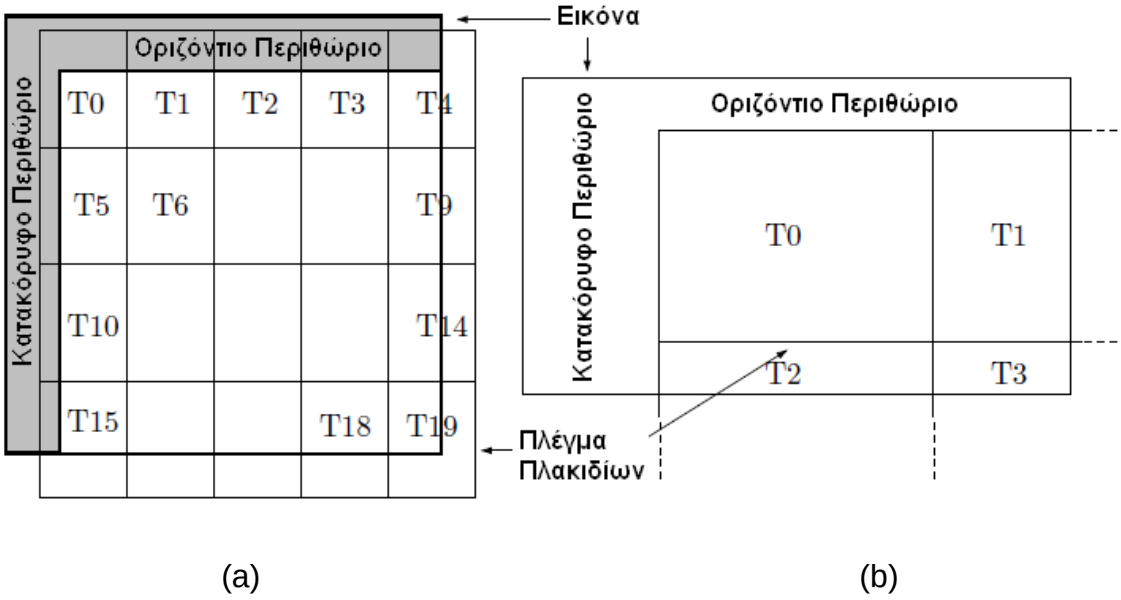
Εάν τα αρχικά χρωματικά στοιχεία της εικόνας είναι το κόκκινο (R), το πράσινο (G) και το μπλε (B) τότε ο μη αντιστρέψιμος μετασχηματισμός είναι παρόμοιος με την

χρωματική αναπαράσταση YC_bC_R.

8.2 Πλακίδια

Κάθε μετασχηματισμένο (αντιστρέψιμο ή μη) χρωματικό στοιχείο της εικόνας, διαμερίζεται σε ορθογώνια, μη αλληλεπικαλυπτόμενα πλακίδια. Δεδομένου ότι τα διάφορα χρωματικά στοιχεία της εικόνας μπορεί να έχουν διαφορετικές αναλύσεις μπορεί να χρησιμοποιηθεί διαφορετικό μέγεθος πλακιδίου για κάθε ένα από αυτά. Τα πλακίδια μπορεί να έχουν οποιοδήποτε μέγεθος, από πολύ μικρό μέχρι όσο και το μέγεθος ολόκληρης της εικόνας. Όλα τα πλακίδια κάθε χρωματικού στοιχείου έχουν το ίδιο μέγεθος εκτός αυτών που βρίσκονται στο περιθώριο της εικόνας. Κάθε πλακίδιο συμπιέζεται ξεχωριστά.

Στο Σχήμα 8.1 παρουσιάζεται ο διαμερισμός της εικόνας σε πλακίδια. Το πρότυπο JPEG 2000 επιτρέπει μία κάθετη από την κορυφή και μία οριζόντια από τα αριστερά τιμή περιθωρίου (offset). Το αρχικό σημείο του πλέγματος μπορεί να τοποθετηθεί οπουδήποτε προσδιορίζουν οι δύο αυτές τιμές περιθωρίου. Όπως παρατηρείται, όλα τα πλακίδια έχουν ίδιο μέγεθος εκτός αυτών που βρίσκονται στην περιφέρεια του πλέγματος το οποίο καθορίζουν οι τιμές περιθωρίου όπως τα πλακίδια T0, T4 και T19 τα οποία έχουν περικοπεί.



Σχήμα 8.1 : Πλακίδια στο Πρότυπο JPEG 2000

Ο βασικός λόγος για τον οποίο εισήχθη η έννοια των πλακιδίων στο πρότυπο είναι για να παρέχεται η δυνατότητα στο χρήστη να αποκωδικοποιεί συγκεκριμένες περιοχές ενδιαφέροντος της εικόνας (ROIs). Ο αποκωδικοποιητής μπορεί να αναγνωρίσει τα

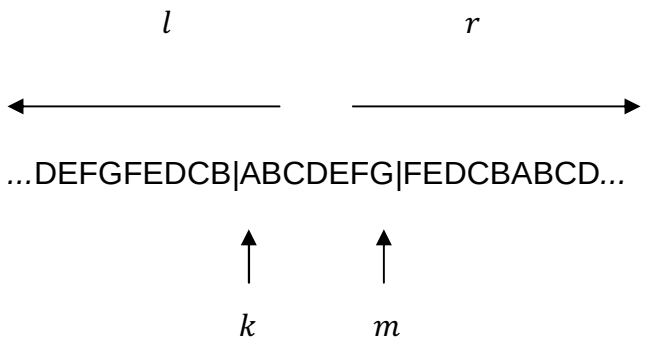
εικονοστοιχεία που ανήκουν σε ένα πλακίδιο.

8.3 Μετασχηματισμός Κυματιδίου

Δύο μετασχηματισμοί κυματιδίου έχουν καθοριστεί από το πρότυπο. Αυτοί είναι ο μη αντιστρέψιμος μετασχηματισμός κυματιδίου $(9,7)$ κινητής υποδιαστολής και ο ακέραιος μετασχηματισμός κυματιδίου $(5,3)$ ο οποίος είναι αντιστρέψιμος. Και οι δύο μετασχηματισμοί επιτρέπουν προοδευτική μετάδοση αλλά μόνο ο ακέραιος μετασχηματισμός παράγει συμπίεση χωρίς απώλειες.

Συμβολίζουμε μια γραμμή με τα εικονοστοιχεία ενός πλακιδίου με $P_k, P_{k+1}, \dots, P_m$. Εξαιτίας της φύσης των μετασχηματισμών κυματιδίου που χρησιμοποιούνται στο πρότυπο JPEG 2000, συχνά, ενδέχεται να απαιτείται η χρήση εικονοστοιχείων με δείκτες μικρότερους του k και μεγαλύτερους του m. Συνεπώς προτού υπολογιστεί ο μετασχηματισμός κυματιδίου για κάποιο πλακίδιο, τα εικονοστοιχεία του πλακιδίου αυτού θα πρέπει να επεκταθούν. Το πρότυπο JPEG 2000 καθορίζει μία απλή μέθοδο επέκτασης που ονομάζεται περιοδική συμμετρική επέκταση, η οποία παρουσιάζεται στο Σχήμα 8.2. Το σχήμα παρουσιάζει μια γραμμή αποτελούμενη από επτά σύμβολα “ABCDEFG” και τον τρόπο με τον οποίο ανακλώνται προς τα αριστερά και τα δεξιά προκειμένου να επεκτείνουν την γραμμή κατά l και r σύμβολα αντίστοιχα. Ο

Πίνακας 8.1 εμφανίζει τις ελάχιστες τιμές l και r ως συνάρτηση των τιμών k και m καθώς και του μετασχηματισμού που χρησιμοποιείται.



Σχήμα 8.2 : Επέκταση μίας Γραμμής Εικονοστοιχείων

Πίνακας 8.1 : Ελάχιστες τιμές l και r ως Συνάρτηση των Τιμών k και m

k	l (5,3)	l (9,7)	m	r (5,3)	r (9,7)
Άρτια	2	4	Άρτιο	2	4
Περιττά	1	3	Περιττό	1	3

Θεωρούμε τώρα μια γραμμή επεκταμένων εικονοστοιχείων ενός πλακιδίου $P_k, P_{k+1}, \dots, P_m$. Δεδομένου ότι πλέον τα εικονοστοιχεία έχουν επεκταθεί, τιμές δεικτών μικρότερες του k και μεγαλύτερες του m μπορούν να χρησιμοποιηθούν. Ο ακέραιος μετασχηματισμός κυματιδίου $(5,3)$ υπολογίζει τους συντελεστές κυματιδίου $C(i)$ υπολογίζοντας αρχικά τις περιττές τιμές $C(2i+1)$ τις οποίες χρησιμοποιεί στη συνέχεια για να υπολογίσει τις άρτιες τιμές $C(2i)$. Οι υπολογισμοί είναι οι ακόλουθοι:

$$C(2i+1) = P(2i+1) - \left\lfloor \frac{P(2i) + P(2i+2)}{2} \right\rfloor, \gamma\iota\alpha\ k-1 \le 2i+1 < m+1,$$

$$C(2i) = P(2i) + \left\lfloor \frac{C(2i-1) + C(2i+1) + 2}{4} \right\rfloor, \gamma\iota\alpha\ \kappa \le 2i < m+1$$

Ο μετασχηματισμός κυματιδίου κινητής υποδιαστολής $(9,7)$ υπολογίζεται εκτελώντας αρχικά μια “άρση” τεσσάρων βημάτων, ακολουθούμενη από μία “κλιμάκωση” δύο βημάτων στις τιμές των επεκταμένων εικονοστοιχείων P_k έως P_m. Κάθε βήμα εφαρμόζεται σε όλα τα εικονοστοιχεία του πλακιδίου προτού ξεκινήσει το επόμενο. Το πρώτο βήμα εφαρμόζεται σε κάθε εικονοστοιχείο με δείκτη i όπου το i ικανοποιεί τη συνθήκη $k-3 \le 2i+1 < m+3$. Το δεύτερο βήμα εκτελείται για κάθε i όπου $k-2 \le 2i < m+2$. Το τρίτο βήμα για $k-1 \le 2i+1 < m+1$. Το τέταρτο για $k \le 2i < m$. Το πέμπτο βήμα πραγματοποιείται για $k \le 2i+1 < m$. Το έκτο και τελευταίο βήμα εκτελείται για $k \le 2i < m$. Οι υπολογισμοί είναι οι ακόλουθοι:

$$C(2i+1) = P(2i+1) + a[P(2i) + P(2i+2)], \qquad B\acute{\eta}\mu\alpha\ 1^o$$

$$C(2i) = P(2i) + \beta[C(2i-1) + C(2i+1)], \qquad B\acute{\eta}\mu\alpha\ 2^o$$

$$C(2i+1) = C(2i+1) + \gamma[C(2i) + C(2i+2)], \qquad B\acute{\eta}\mu\alpha\ 3^o$$

$$C(2i) = C(2i) + \delta[C(2i-1) + C(2i+1)], \qquad B\acute{\eta}\mu\alpha\ 4^o$$

$$C(2i+1) = -K \times C(2i+1), \qquad Βήμα\ 5^o$$

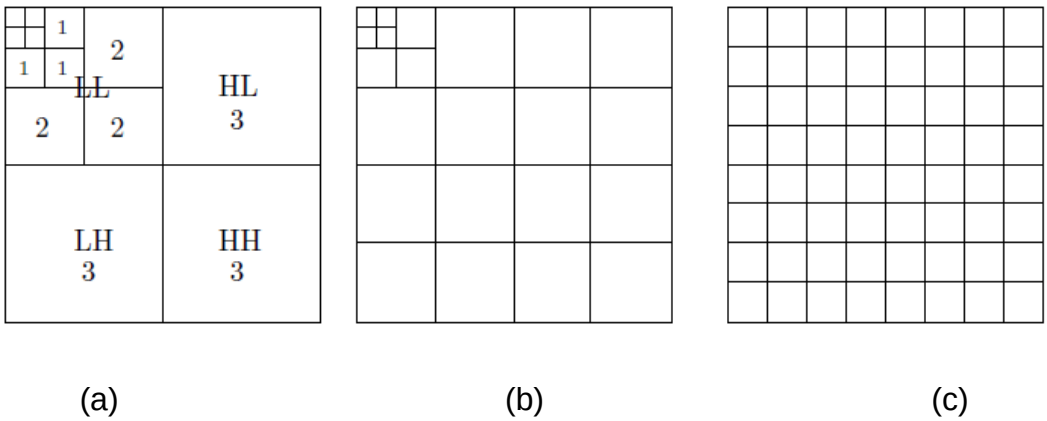
$$C(2i) = \frac{1}{K} \times C(2i), \qquad Βήμα\ 6^o$$

Όπου οι πέντε σταθερές (συντελεστές του φίλτρου) που χρησιμοποιούνται από το JPEG 2000 είναι οι:

$$\alpha = -1.586134342, \beta = -0.052980118, \gamma = 0.882911075, \delta = 0.443506852\ και$$

$$\mathrm{K} = 1.230174105$$

Οι μονοδιάστατοι αυτοί μετασχηματισμοί κυματιδίου εφαρμόζονται L φορές, (όπου L παράμετρος η οποία καθορίζεται είτε από τον χρήστη είτε από τον κωδικοποιητή) και εναλλάσσονται μεταξύ των γραμμών και των στηλών προκειμένου να σχηματιστούν L επίπεδα υποζωνών. Η εικόνα με ανάλυση $L-1$ είναι η αρχική εικόνα (στο Σχήμα 8.3 a η ανάλυση είναι 3) ενώ με ανάλυση 0 είναι η υποζώνη με τη χαμηλότερη συχνότητα. Οι υποζώνες μπορούν να οργανωθούν με έναν από τους τρείς τρόπους που παρουσιάζονται στο Σχήμα 8.3.



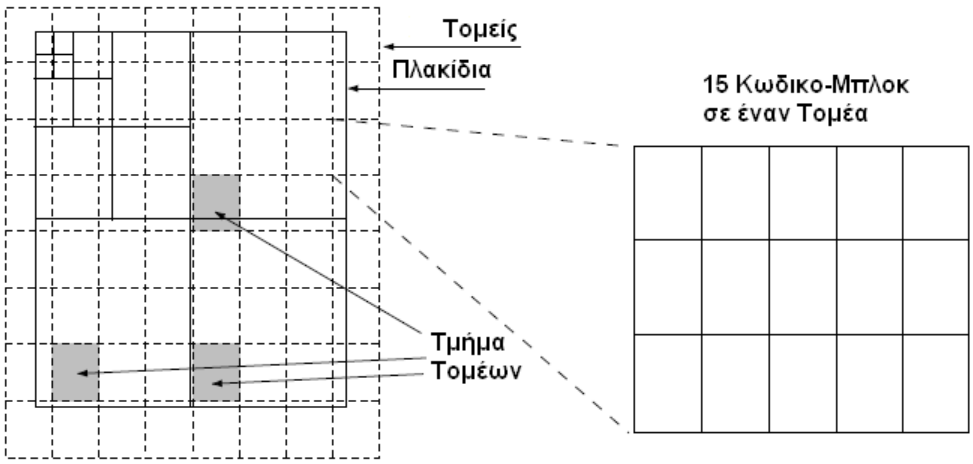
Σχήμα 8.3 : Οργάνωση Υποζωνών στο JEPG 2000

8.4 Κβάντιση

Κάθε υποζώνη δύναται να έχει διαφορετικό βήμα κβάντισης. Κάθε συντελεστής του μετασχηματισμού σε μια υποζώνη διαιρείται με ένα βήμα κβάντισης και το αποτέλεσμα στρογγυλοποιείται. Το μέγεθος του βήματος κβάντισης μπορεί να υπολογιστεί επαναληπτικά προκειμένου να επιτευχθεί συγκεκριμένος ρυθμός ή ποιότητα εικόνας ενώ μπορεί και να προεπιλεγεί από τον χρήστη. Εφόσον ο χρήστης επιθυμεί συμπίεση χωρίς απώλειες το βήμα κβάντισης λαμβάνει την τιμή 1.

8.5 Τομείς και Κωδικά-Μπλοκ

Θεωρούμε ένα πλακίδιο ενός χρωματικού στοιχείου. Οι αρχικές τιμές των εικονοστοιχείων μετασχηματίζονται με μετασχηματισμό κυματιδίων. Αυτό έχει ως αποτέλεσμα τη δημιουργία υποζωνών L επιπέδων ανάλυσης. Στο Σχήμα 8.4 a παρουσιάζεται ένα πλακίδιο και τέσσερα επίπεδα ανάλυσης. Υπάρχουν τρεις υποζώνες σε κάθε επίπεδο ανάλυσης εκτός του χαμηλότερου επιπέδου. Το συνολικό μέγεθος όλων των υποζωνών ισούται με το μέγεθος του πλακιδίου. Ένα πλέγμα ορθογωνίων παραλληλογράμμων, γνωστά ως τομείς, παρουσιάζονται τώρα σε ολόκληρη την εικόνα όπως φαίνεται στο Σχήμα 8.4 b.



(a) (b)

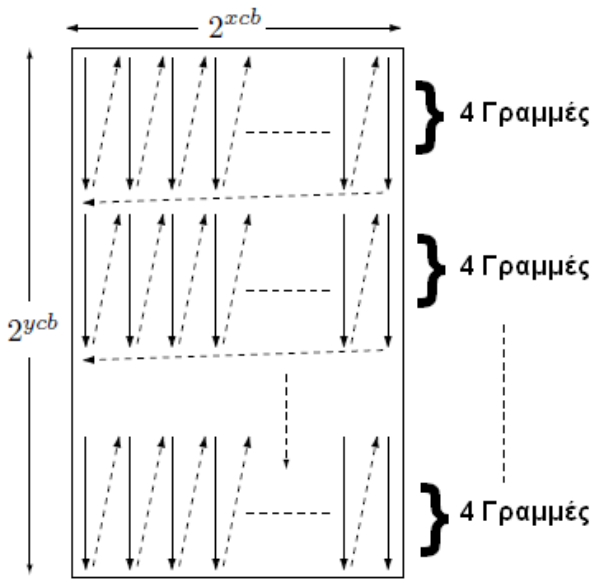
Σχήμα 8.4 : Υποζώνες, Τομείς και Κωδικά-Μπλοκ

Το αρχικό σημείο του πλέγματος τομέων τοποθετείται στην επάνω αριστερή γωνία της εικόνας ενώ οι διαστάσεις ενός τομέα είναι δυνάμεις του 2. Παρατηρούμε ότι τα όρια κάθε υποζώνης στη γενική περίπτωση δεν είναι όμοια με τα όρια των τομέων. Εάν θεωρήσουμε τις τρείς υποζώνες μιας συγκεκριμένης ανάλυσης και επιλέξουμε τρείς τομείς τοποθετημένους στην ίδια περιοχή της κάθε υποζώνης (περιοχές με γκρι στο Σχήμα 8.4 a) τότε οι τρείς αυτοί τομείς αποτελούν ένα τμήμα τομέων. Το πλέγμα των τομέων διαιρείται ακόμα περισσότερο σε ένα πλέγμα από κωδικά-μπλοκ. Τα κωδικά-μπλοκ αποτελούν τις στοιχειώδεις μονάδες οι οποίες κωδικοποιούνται με αριθμητικούς κώδικες. Στο ακόλουθο σχήμα παρατηρείται ένας τομέας ο οποίος έχει διαιρεθεί σε 15 κωδικά-μπλοκ. Συνεπώς ένα τμήμα τομέων στο συγκεκριμένο παράδειγμα αποτελείται από 45 κωδικά-μπλοκ. Ένα κωδικό-μπλοκ είναι και αυτό ορθογώνιο παραλληλόγραμμο μεγέθους 2^{xcb} (μήκος) $\times$ 2^{ycb} (ύψος), όπου $2 \leq xcb, ycb \leq 10$, και $xcb + ycb \leq 12$.

Όπως γίνεται αντιληπτό, τα πλακίδια, οι τομείς και τα κωδικά-μπλοκ μπορούν να θεωρηθούν ως μεγάλα, μεσαία και μικρά τμήματα μιας εικόνας. Ο διαμερισμός της εικόνας σε μικρότερα τμήματα παρουσιάζει ένα πλήθος πλεονεκτημάτων όπως δυνατότητα δημιουργίας αποδοτικότερων υλοποιήσεων σε ότι αφορά τη μνήμη, αποδοτικότερη μετάδοση καθώς και εύκολη πρόσβαση σε διάφορα σημεία της κωδικοποιημένης ακολουθίας. Βεβαίως, απλούστερες υλοποιήσεις του κωδικοποιητή JPEG 2000 αγνοούν τον διαμερισμό σε τόσο μικρά τμήματα, δημιουργώντας ένα μόνο πλακίδιο με έναν τομέα και ένα κωδικό-μπλοκ. Αντίθετα, εξελιγμένοι κωδικοποιητές ενδέχεται να καταλήξουν σε ένα μεγάλο πλήθος από κωδικά-μπλοκ προκειμένου να δώσουν τη δυνατότητα στον αποκωδικοποιητή να πραγματοποιήσει προοδευτική αποσυμπίεση, γρήγορη μετάδοση, μεγέθυνση της εικόνας και άλλες λειτουργίες σε τμήματα της εικόνας.

8.6 Κωδικοποίηση Εντροπίας

Οι συντελεστές κυματιδίου ενός κωδικό-μπλοκ κωδικοποιούνται με αριθμητικούς κώδικες ανά επίπεδο bit (bit-plane). Η κωδικοποίηση πραγματοποιείται από το σημαντικότερο επίπεδο bit (το οποίο αποτελείται από τα σημαντικότερα bit των συντελεστών) προς το λιγότερο σημαντικό επίπεδο bit. Κάθε επίπεδο bit σαρώνεται όπως φαίνεται στο Σχήμα 8.5. Αρχικά καθορίζεται ένα πλαίσιο για κάθε bit (context) και εκτιμάται η πιθανότητα εμφάνισης του βάσει του επιλεχθέντος πλαισίου. Το bit και η πιθανότητα εμφάνισης του στέλνονται στη συνέχεια στον αριθμητικό κωδικοποιητή.



Σχήμα 8.5 : Λωρίδες σε ένα Κωδικό-Μπλοκ

Πολλές μέθοδοι συμπίεσης λειτουργούν διαφορετικά. Μια τυπική μέθοδος συμπίεσης εικόνας ενδέχεται να χρησιμοποιεί γειτονικά εικονοστοιχεία ως πλαίσιο του προς κωδικοποίηση εικονοστοιχείου και στη συνέχεια να κωδικοποιεί ολόκληρη την τιμή του εικονοστοιχείου βάσει του πλαισίου και όχι ενός μεμονωμένου bit. Σε ένα πλαίσιο αυτής της μορφής θα πρέπει να χρησιμοποιηθούν εικονοστοιχεία τα οποία θα είναι ήδη γνωστά στον αποκωδικοποιητή (συνήθως εικονοστοιχεία των προηγούμενων γραμμών και στηλών). Το JPEG 2000 εργάζεται διαφορετικά στο συγκεκριμένο θέμα. Κωδικοποιεί μεμονωμένα bit (για το λόγο αυτό χρησιμοποιεί κωδικοποιητή MQ, ο οποίος κωδικοποιεί bit και όχι αριθμούς) χρησιμοποιώντας συμμετρικά πλαίσια. Το πλαίσιο ενός bit υπολογίζεται από τους οκτώ πλησιέστερους γείτονες. Δεδομένου όμως ότι ο αποκωδικοποιητής δεν γνωρίζει τις τιμές όλων των γειτονικών στοιχείων τη στιγμή της αποκωδικοποίησης, δεν είναι δυνατό να χρησιμοποιηθούν οι τιμές τους στο πλαίσιο. Αντί αυτού χρησιμοποιείται η “σημαντικότητα” κάθε γείτονα. Κάθε συντελεστής κυματιδίου διαθέτει μία μεταβλητή ενός bit η οποία εκφράζει τη σημαντικότητα του. Η μεταβλητή αυτή ονομάζεται “κατάσταση σημαντικότητας” του συντελεστή. Όταν ξεκινά η κωδικοποίηση ενός κωδικού-μπλοκ, όλοι οι συντελεστές κυματιδίου θεωρούνται ασήμαντοι και έτσι η κατάσταση σημαντικότητας είναι μηδενική.

Κάποια από τα πιο σημαντικά επίπεδα bit ενδέχεται να έχουν τιμή μηδέν. Το πλήθος των επιπέδων αυτών αποθηκεύεται στην κωδικοποιημένη ροή δεδομένων μετάδοσης και χρησιμοποιείται από τον αποκωδικοποιητή. Η κωδικοποίηση ξεκινά από το πρώτο επίπεδο bit με μη-μηδενική τιμή. Η κωδικοποίηση αυτού του επιπέδου bit πραγματοποιείται σε ένα πέρασμα. Καθένα από τα λιγότερο σημαντικά επίπεδα bit κωδικοποιείται σε τρία περάσματα γνωστά ως “πέρασμα διάδοσης της σημαντικότητας”, “πέρασμα βελτίωσης του μεγέθους” και “πέρασμα εκκαθάρισης”. Κάθε πέρασμα διαιρεί το επίπεδο bit σε λωρίδες ύψους τεσσάρων γραμμών κάθε μία (Σχήμα 8.5). Κάθε λωρίδα σαρώνεται στήλη προς στήλη από τα αριστερά προς τα δεξιά. Κάθε bit του επιπέδου κωδικοποιείται σε ένα από τα τρία περάσματα.

Το πρώτο πέρασμα κωδικοποίησης (διάδοση σημαντικότητας) ενός επιπέδου bit, κωδικοποιεί όλα τα bit τα οποία ανήκουν σε συντελεστές κυματιδίου και ικανοποιούν τις δύο ακόλουθες συνθήκες. Πρώτον, ο συντελεστής να είναι μη-σημαντικός και δεύτερον, τουλάχιστον ένας από τους οκτώ πλησιέστερους γείτονες να είναι σημαντικός. Εάν ένα bit κωδικοποιηθεί σε αυτό το πέρασμα και εάν το bit έχει τιμή 1, τότε ο συντελεστής κυματιδίου σημειώνεται ως σημαντικός θέτοντας στην κατάσταση σημαντικότητας την τιμή 1.

Είναι προφανές πως για να κωδικοποιηθούν κάποια bit σε αυτό το βήμα θα πρέπει κάποιοι συντελεστές κυματιδίου να χαρακτηριστούν σημαντικοί προτού ξεκινήσει το πέρασμα. Αυτός είναι και ο λόγος για τον οποίο το πρώτο επίπεδο bit κωδικοποιείται σε ένα πέρασμα, το πέρασμα εκκαθάρισης. Στο πέρασμα αυτό κωδικοποιούνται όλα τα bit του επιπέδου. Εάν ένα bit έχει τιμή 1 τότε ο συντελεστής του θεωρείται σημαντικός.

Το δεύτερο πέρασμα (βελτίωσης του μεγέθους) κωδικοποιεί όλα τα bit των συντελεστών κυματιδίου οι οποίοι σημειώθηκαν ως σημαντικοί σε προηγούμενο επίπεδο. Επομένως, εφόσον ένας συντελεστής γίνεται σημαντικός τότε όλα τα λιγότερο σημαντικά bit του συντελεστή θα κωδικοποιηθούν ένα προς ένα, καθένα στο δεύτερο πέρασμα του επόμενου επιπέδου bit από το οποίο ο συντελεστής σημειώθηκε ως σημαντικός.

Το τρίτο και τελευταίο πέρασμα (εκκαθάρισης) ενός επιπέδου bit κωδικοποιεί όλα τα bit τα οποία δεν κωδικοποιήθηκαν στα προηγούμενα δύο περάσματα. Εάν θεωρήσουμε ένα συντελεστή C στο επίπεδο bit B τότε εάν ο C είναι μη-σημαντικός, κάθε bit στο B δε θα κωδικοποιηθεί στο δεύτερο πέρασμα. Εάν οκτώ πλησιέστεροι γείτονες του C είναι μη-σημαντικοί τότε το bit του C στο επίπεδο B δεν θα κωδικοποιηθεί ούτε στο πρώτο επίπεδο. Ένα τέτοιο bit κωδικοποιείται στο τρίτο πέρασμα. Εφόσον το bit τυγχάνει να έχει τιμή 1, ο C θα γίνει σημαντικός.

Οι συντελεστές κυματιδίου είναι προσημασμένοι ακέραιοι και διαθέτουν bit πρόσημου. Στο JPEG 2000 οι αριθμοί αναπαρίστανται με τη μέθοδο “πρόσημου-τιμής”. Το bit πρόσημου έχει τιμή 0 όταν ο αριθμός είναι θετικός (ή μηδέν) και 1 όταν είναι αρνητικός. Η τιμή είναι ίδια ανεξαρτήτως πρόσημου.

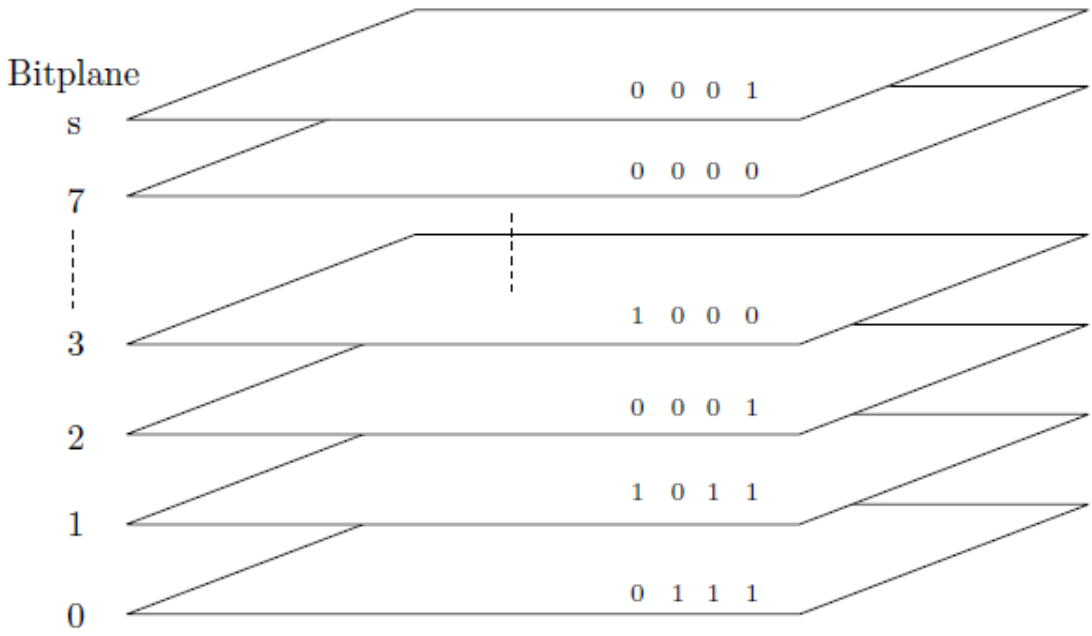
Η μέθοδος κωδικοποίησης των τριών περασμάτων παρουσιάζεται με τη χρήση ενός αριθμητικού παραδείγματος. Θεωρούμε τους τέσσερεις συντελεστές με τιμές:

$$10 = 0|\mathbf{0000}1010,\ \ 1 = 0|\mathbf{0000}0001,\ \ 3 = 0|\mathbf{0000}0011\ \ \kappa\alpha\iota\ \ -7 = 1|\mathbf{0000}0111$$

Υπάρχουν οκτώ επίπεδα bit αριθμημένα από το 7 μέχρι το 0 από τα αριστερά (περισσότερο σημαντικά) προς τα δεξιά (λιγότερο σημαντικά). Τα πρώτα τέσσερα επίπεδα 7 έως 4 αποτελούνται μόνο από μηδενικά, συνεπώς η κωδικοποίηση ξεκινά από το επίπεδο 3 (το πρόσημα αρχικά δε λαμβάνονται υπόψη) όπως φαίνεται στο Σχήμα 8.6. Σε αυτό το επίπεδο (το πρώτο μη-μηδενικό) θα εφαρμοστεί μόνο το τρίτο πέρασμα (εκκαθάρισης). Ένα bit για κάθε έναν από τους τέσσερεις συντελεστής κωδικοποιείται στο πέρασμα αυτό. Το bit αυτό για τον συντελεστή 10 έχει τιμή 1 επομένως ο συντελεστής αυτός σημειώνεται ως σημαντικός και επομένως τα υπόλοιπα bit στα επίπεδα 2, 1 και 0 θα κωδικοποιηθούν στο δεύτερο πέρασμα. Επίσης

κωδικοποιείται και το bit πρόσημου μετά από το bit αυτό. Στη συνέχεια κωδικοποιείται το δεύτερο επίπεδο bit. Ο συντελεστής 10 είναι σημαντικός επομένως το bit σε αυτό το επίπεδο (το οποίο έχει τιμή μηδέν) θα κωδικοποιηθεί στο δεύτερο πέρασμα. Ο συντελεστής 1 είναι μη-σημαντικός όμως ένας από τους γείτονες του (το 10) είναι σημαντικός, επομένως το bit του συντελεστή 1 στο δεύτερο επίπεδο bit (το οποίο έχει τιμή μηδέν) κωδικοποιείται στο πρώτο πέρασμα. Τα bit των συντελεστών 3 και -7 σε αυτό το επίπεδο bit (ένα μηδενικό και ένας άσσος) θα κωδικοποιηθούν στο τρίτο πέρασμα. Το bit πρόσημου του συντελεστή -7 κωδικοποιείται μετά το bit 1 του συντελεστή αυτού. Επίσης ο συντελεστής -7 χαρακτηρίζεται σημαντικός.

Στη συνέχεια κωδικοποιείται το επίπεδο 1. Το bit του επιπέδου αυτού για τον συντελεστή 10 κωδικοποιείται στο δεύτερο πέρασμα. Το bit του συντελεστή κωδικοποιείται στο πρώτο πέρασμα, όμοια με το επίπεδο 2. Το bit του συντελεστή 3 κωδικοποιείται στο πρώτο πέρασμα δεδομένου ότι ένας εκ των πλησιέστερων γειτόνων, ο -7, είναι πλέον σημαντικός. Το bit αυτό έχει τιμή 1 οπότε ο συντελεστής 3 γίνεται πλέον σημαντικός. Το bit αυτό είναι ο πρώτος άσσος του συντελεστή 3 συνεπώς το πρόσημο κωδικοποιείται μετά από το bit αυτό.



Σχήμα 8.6 : Επίπεδα bit των Τεσσάρων Συντελεστών ενός Κωδικό-Μπλοκ

Στο τελευταίο επίπεδο κωδικοποιείται το bit του συντελεστή 10 στο δεύτερο πέρασμα. Το bit του συντελεστή 1 κωδικοποιείται στο πρώτο πέρασμα δεδομένου ότι ο συντελεστής είναι ακόμα μη-σημαντικός και εξακολουθεί να έχει σημαντικούς πλησιέστερους γείτονες. Το bit αυτό έχει τιμή 1, επομένως ο συντελεστής 1 γίνεται

σημαντικός (γεγονός που δεν είναι ιδιαιτέρως χρήσιμο για τη συνέχεια καθώς δεν υπάρχουν άλλα επίπεδα προς κωδικοποίηση). Όπως και στις προηγούμενες περιπτώσεις θα ακολουθήσει η κωδικοποίηση του bit πρόσημου του συντελεστή 1. Τα bit των συντελεστών 3 και -7 κωδικοποιούνται στο δεύτερο πέρασμα αφού οι συντελεστές αυτοί είναι σημαντικοί.

Το πλαίσιο ενός bit καθορίζεται με διαφορετικό τρόπο σε κάθε πέρασμα. Εδώ εξετάσθηκε ο τρόπος με τον οποίο καθορίστηκε στο πέρασμα διάδοσης της σημαντικότητας. Οι οκτώ πλησιέστεροι γείτονες ενός συντελεστή κυματιδίου *X* χρησιμοποιούνται για τον καθορισμό του πλαισίου που θα χρησιμοποιηθεί για την κωδικοποίηση των bit του συντελεστή *X*. Ένα εκ των εννέα διαθέσιμων πλαισίων επιλέγεται και χρησιμοποιείται προκειμένου να υπολογιστεί η πιθανότητα του bit προς κωδικοποίηση. Αξίζει να σημειωθεί για μία ακόμη φορά πως δεν εξετάζονται οι αριθμητικές τιμές των γειτόνων για να καθοριστούν τα πλαίσια αλλά η κατάσταση σημαντικότητας τους. Στο Σχήμα 8.7 παρουσιάζονται τα ονόματα που δόθηκαν στην κατάσταση σημαντικότητας των γειτόνων. Ο Πίνακας 8.2 παρουσιάζει τα κριτήρια τα οποία χρησιμοποιούνται για την επιλογή ενός εκ των εννέα διαθέσιμων πλαισίων. Παρατηρούμε ότι τα κριτήρια αυτά εξαρτώνται από την υποζώνη η οποία κωδικοποιείται (HH, HL, LH ή LL). Το πλαίσιο 0 επιλέγεται όταν ο συντελεστής *X* δεν έχει σημαντικούς πλησιέστερους γείτονες. Το πλαίσιο 8 επιλέγεται όταν και οι οκτώ γείτονες του συντελεστή *X* είναι σημαντικοί.

D_0	V_0	D_1
H_0	X	H_1
D_2	V_1	D_3

Σχήμα 8.7 : Οι 8 Γείτονες

Το πρότυπο JPEG 2000 καθορίζει αντίστοιχους κανόνες επιλογής πλαισίου για κάθε ένα bit τόσο του περάσματος βελτίωσης του μεγέθους όσο και του περάσματος εκκαθάρισης αλλά και της επιλογής πλαισίου για το bit πρόσημου. Ο καθορισμός του περάσματος εκκαθάρισης είναι όμοιος με αυτόν του περάσματος σημαντικότητας με τη διαφορά ότι η κωδικοποίηση του τρέχοντος μήκους μηδενικών χρησιμοποιείται εφόσον

και τα τέσσερα bit της στήλης μίας λωρίδας είναι μη-σημαντικά και το κάθε ένα έχει μη σημαντικούς γείτονες. Στην περίπτωση αυτή ένα bit κωδικοποιείται υποδηλώνοντας την ύπαρξη ή όχι μίας μηδενικής στήλης. Εάν όχι, τότε στη στήλη υπάρχει ένα πλήθος διαδοχικών μηδενικών και το πλήθος τους κωδικοποιείται. Προφανώς τα διαδοχικά μηδενικά θα ακολουθεί κάποιος άσσος. Αυτός ο άσσος δεν κωδικοποιείται καθώς είναι απλό να συναχθεί από τον αποκωδικοποιητή. Από τη στιγμή που ένα πλαίσιο έχει καθοριστεί για ένα bit, χρησιμοποιείται για την εκτίμηση της πιθανότητας του bit προς κωδικοποίηση.

Πίνακας 8.2 : Τα εννέα Πλαίσια του Περάσματος Διάδοσης Σημαντικότητας

Υποζώνες LL και LH			Υποζώνη HL			Υποζώνη HH		Πλαίσιο
$\sum H_i$	$\sum V_i$	$\sum D_i$	$\sum H_i$	$\sum V_i$	$\sum D_i$	$\sum (H_i + V_i)$	$\sum H_i$	
2				2			≥ 3	8
1	≥ 1		≥ 1	1		≥ 1	2	7
1	0	≥ 1	0	1	≥ 1	0	2	6
1	0	0	0	1	0	≥ 2	1	5
0	2		2	0		1	1	4
0	1		1	0		0	1	3
0	0	≥ 2	0	0	≥ 2	≥ 2	0	2
0	0	1	0	0	1	1	0	1
0	0	0	0	0	0	0	0	0

8.7 Πακέτα

Αφού κωδικοποιηθούν όλα τα bit όλων των συντελεστών όλων των κωδικών-μπλοκ ενός τμήματος τομέων σε μία συμπυκνωμένη ροή δεδομένων, μία κεφαλίδα προστίθεται στη ροή αυτή μετατρέποντας τη σε πακέτο. Στο Σχήμα 8.4 εμφανίζεται ένα τμήμα τομέων το οποίο αποτελείται από τρείς τομείς, καθένας διαιρεμένος σε 15 κωδικά-μπλοκ. Η κωδικοποίηση επομένως του τμήματος αυτού δημιουργεί ένα πακέτο με 45 κωδικοποιημένα κωδικό-μπλοκ. Η κεφαλίδα περιέχει όλες τις πληροφορίες που είναι απαραίτητες για την αποκωδικοποίηση του πακέτου. Εάν όλα τα κωδικά-μπλοκ ενός τμήματος τομέων είναι μηδενικά τότε το σώμα του πακέτου θα είναι κενό. Υπενθυμίζεται

ότι ένα τμήμα τομέων αντιστοιχεί στην ίδια χωρική περιοχή και στις τρείς υποζώνες. Συνεπώς, ένα πακέτο μπορεί να θεωρηθεί ως μία ποιοτική αύξηση κατά ένα επίπεδο ανάλυσης σε ένα συγκεκριμένο χωρικό σημείο.

8.8 Επίπεδα

Επίπεδο είναι ένα σύνολο πακέτων. Αποτελείται από ένα πακέτο για κάθε τμήμα τομέων κάθε επιπέδου ανάλυσης. Επομένως, ένα επίπεδο μπορεί να θεωρηθεί ως η αύξηση της ποιότητας ολόκληρης της εικόνας σε πλήρη ανάλυση.

8.9 Προοδευτική Μετάδοση

Η προοδευτική μετάδοση αποτελεί ένα από τα σημαντικότερα χαρακτηριστικά του JPEG 2000. Το πρότυπο παρέχει τέσσερεις τρόπους προοδευτικής μετάδοσης και αποκωδικοποίησης της εικόνας βάσει της ανάλυσης, της ποιότητας, της χωρικής περιοχής και του στοιχείου. Η προοδευτική μετάδοση επιτυγχάνεται απλά με την αποθήκευση των πακέτων με συγκεκριμένη σειρά στη ροή δεδομένων.

Όταν μια εικόνα κωδικοποιείται, τα πακέτα τοποθετούνται στη ροή δεδομένων με συγκεκριμένη σειρά, ανάλογα με τον τύπο προοδευτικής μετάδοσης που επιθυμεί ο χρήστης. Εφόσον ένας χρήστης ή εφαρμογή απαιτεί έναν διαφορετικό τύπο προόδου (δηλαδή διαφορετική σειρά πακέτων) θα είναι εύκολο διαβάζοντας τη ροή των δεδομένων να αναγνωρίσει τα πακέτα και να τα ταξινομήσει με την επιθυμητή σειρά. Η διαδικασία αυτή είναι γνωστή ως συντακτική ανάλυση (parsing) και θεωρείται απλή λόγω του πλήθους δεικτών (markers) που έχουν τοποθετηθεί μέσα στη ροή δεδομένων. Υπάρχουν διάφοροι τύποι δεικτών και καθένας χρησιμοποιείται για διαφορετικό σκοπό. Κάποιοι δείκτες προσδιορίζουν τον τύπο της προοδευτικής μετάδοσης ενώ άλλοι περιέχουν το μήκος όλων των πακέτων.

Τυπικό παράδειγμα συντακτικής ανάλυσης αποτελεί η εκτύπωση μιας έγχρωμης εικόνας σε έναν ασπρόμαυρο εκτυπωτή. Σε μια περίπτωση σαν και αυτή, δεν υπάρχει λόγος αποστολής της πληροφορίας σχετικά με το χρώμα στον εκτυπωτή. Ένας αναλυτής μπορεί χρησιμοποιώντας τους δείκτες να αναγνωρίσει όλα τα πακέτα που περιέχουν τη χρωματική πληροφορία και να τα απορρίψει. Άλλο παράδειγμα είναι η μείωση του μεγέθους της ροής δεδομένων (αυξάνοντας προφανώς τις απώλειες σε ότι αφορά την ποιότητα της εικόνας). Ο αναλυτής αναγνωρίζει και απορρίπτει τα επίπεδα εκείνα που συνεισφέρουν λιγότερο στην ποιότητα της εικόνας. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να επιτευχθεί το επιθυμητό μέγεθος. Ο συντακτικός αναλυτής

μπορεί να είναι μέρος ενός εξυπηρετητή εικόνων. Ο πελάτης αποστέλλει ένα αίτημα σε έναν τέτοιο εξυπηρετητή με το όνομα της εικόνας και τις επιθυμητές της ιδιότητες. Ο εξυπηρετητής εκτελεί τη συντακτική ανάλυση και αποκτά την εικόνα με τις επιθυμητές ιδιότητες την οποία και αποστέλλει στον πελάτη.

8.10 Περιοχές Ενδιαφέροντος (ROIs)

Ο πελάτης πολλές φορές θέλει να αποκωδικοποιήσει συγκεκριμένο μέρος της εικόνας, δηλαδή μία περιοχή ενδιαφέροντος. Ο αναλυτής θα πρέπει να είναι σε θέση να αναγνωρίζει τα τμήματα της ροής δεδομένων τα οποία αντιστοιχούν στη συγκεκριμένη περιοχή και να αποστέλλει μόνο τα τμήματα αυτά. Μία περιοχή ενδιαφέροντος μπορεί να καθοριστεί κατά τη συμπίεση της εικόνας. Στην περίπτωση αυτή ο κωδικοποιητής προσδιορίζει τα κωδικά-μπλοκ τα οποία σχετίζονται με τη συγκεκριμένη περιοχή και τα τοποθετεί στην αρχή της ροής δεδομένων. Σε πολλές περιπτώσεις όμως, μία περιοχή ενδιαφέροντος ζητείται από τον αναλυτή αφού έχει συμπιεστεί η εικόνα. Σε μια τέτοια περίπτωση, ο αναλυτής μπορεί να χρησιμοποιήσει τα πλακίδια προκειμένου να προσδιορίσει την περιοχή ενδιαφέροντος. Κάθε περιοχή ενδιαφέροντος θα περιέχεται σε μερικά πλακίδια (ίσως και μόνο ένα) και έτσι ο αναλυτής θα αναγνωρίσει τα πλακίδια αυτά και θα αποστείλει στον πελάτη μία ροή δεδομένων η οποία θα αποτελείται μόνο από τα πλακίδια αυτά. Αυτή είναι μία σημαντική εφαρμογή των πλακιδίων. Μικρά πλακίδια είναι δυνατό να καθορίσουν μικρές περιοχές ενδιαφέροντος αλλά έχουν ως αποτέλεσμα χαμηλής ποιότητας συμπίεση. Η εμπειρία έχει δείξει ότι πλακίδια μεγέθους 256×256 έχουν αμελητέα επίδραση στην απόδοση της συμπίεσης και συνήθως είναι αρκετά μικρά ώστε να παρέχεται ευελιξία στην επιλογή μικρών περιοχών ενδιαφέροντος.

Εφόσον κάθε πλακίδιο συμπιέζεται ξεχωριστά, είναι εύκολο να βρεθούν οι πληροφορίες που σχετίζονται με κάθε πλακίδιο εντός της ροής δεδομένων αφού κάθε πλακίδιο φέρει κεφαλίδα και δείκτες. Κάθε συντακτική ανάλυση η οποία μπορεί να πραγματοποιηθεί σε ολόκληρη την εικόνα μπορεί επίσης να πραγματοποιηθεί και σε κάθε πλακίδιο ξεχωριστά. Άλλα πλακίδια μπορεί να αγνοηθούν ή να μεταδοθούν σε χαμηλότερη ποιότητα. Εναλλακτικά, ο συντακτικός αναλυτής μπορεί να διαχειριστεί μικρές περιοχές ενδιαφέροντος, εκμαιεύοντας από τη ροή δεδομένων την πληροφορία για κάθε ξεχωριστό κωδικό-μπλοκ. Ο αναλυτής σε αυτή την περίπτωση θα πρέπει να καθορίσει ποια κωδικό-μπλοκ περιέχουν ένα συγκεκριμένο εικονοστοιχείο, να βρεθούν τα πακέτα τα οποία περιέχουν τα κωδικό-μπλοκ αυτά, να αποκωδικοποιηθούν οι κεφαλίδες του πακέτου, να χρησιμοποιηθεί η κεφαλίδα για την εύρεση της πληροφορίας σχετικά με το

κωδικό-μπλοκ μέσα σε κάθε πακέτο και τέλος να αποσταλεί η απαραίτητη πληροφορία στον αποκωδικοποιητή.

Συνοψίζοντας, μετρήσεις και δοκιμές με τον κωδικοποιητή JPEG 2000 έχουν δείξει ότι επιτυγχάνεται καλύτερη συμπίεση σε σχέση με τον κωδικοποιητή JPEG, ιδίως σε εικόνες όπου συμπιέζονται με πολύ χαμηλούς ρυθμούς ή όπου υπάρχει απαίτηση για πολύ υψηλή ποιότητα. Σε συμπίεση χωρίς απώλειες, το JPEG 2000 προσφέρει μόνο μικρές βελτιώσεις έναντι του JPEG.

9. Μέθοδοι Συμπίεσης Τρισδιάστατων Εικόνων

Μέχρι στιγμής έχουν μελετηθεί βασικές μέθοδοι συμπίεσης δεδομένων όπως η κωδικοποίηση RLE, οι κώδικες εντροπίας Huffman και οι αριθμητικοί κώδικες. Επίσης αναλύθηκαν σε βάθος οι δύο επικρατέστεροι μετασχηματισμοί εικόνας καθώς και τα πρότυπα συμπίεσης JPEG και JPEG 2000 τα οποία χρησιμοποιούν τους μετασχηματισμούς DCT και κυματιδίων αντίστοιχα. Ασφαλώς οι κωδικοποιητές που υλοποιούν τα πρότυπα JPEG (http://www.ijg.org/) και JPEG 2000 (http://www.ece.uvic.ca/~frodo/jasper/) μπορούν να συμπιέσουν ολοκληρωτικές εικόνες με αρκετά καλά αποτελέσματα. Συγκριτικές μετρήσεις σχετικά με την απόδοση των κωδικοποιητών αυτών θα πραγματοποιηθούν στην τελευταία ενότητα της εργασίας. Η ιδιαιτερότητα που παρουσιάζουν οι εικόνες ολοκληρωτικής φωτογράφισης έγκειται στο γεγονός ότι κάθε εικόνα αποτελείται από πλήθος στοιχειωδών εικόνων οι οποίες είναι πολύ έντονα συσχετισμένες μεταξύ τους όπως παρατηρούμε και στην Εικόνα 9.1.

Εικόνα 9.1 : Εικόνα Ολοκληρωτικής Φωτογράφισης

Το γεγονός αυτό σε συνδυασμό με την ανάγκη για εύρεση αποδοτικότερων μεθόδων

συμπίεσης λόγω των μεγάλων διαστάσεων των ολοκληρωτικών εικόνων οδήγησαν στην ανάπτυξη μεθόδων που εκμεταλλεύονται αποδοτικότερα τη συσχέτιση των στοιχειωδών εικόνων[52][53]. Αυτές οι μέθοδοι θα αποτελέσουν το αντικείμενο της παρούσας ενότητας.

9.1 Συμπίεση Ολοκληρωτικής Εικόνας ως Βίντεο-Ροή με Χρήση του Προτύπου H.264

Αν και σκοπός την παρούσας εργασίας δεν είναι η εκτενής μελέτη αλγορίθμων συμπίεσης βίντεο, σε αυτό το σημείο θα πραγματοποιηθεί μία σύντομη αναφορά στις βασικές αρχές συμπίεσης βίντεο[59]. Η εισαγωγή αυτή θα βοηθήσει ώστε να γίνουν αντιληπτοί οι λόγοι για τους οποίους θεωρείται ότι η μέθοδος αυτή μπορεί να συμπιέσει αποδοτικότερα ολοκληρωτικές εικόνες.

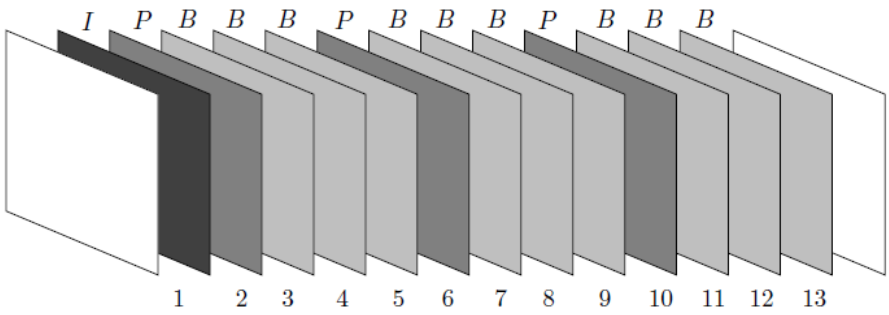
9.1.1 Βασικές Αρχές Συμπίεσης Βίντεο

Η συμπίεση βίντεο βασίζεται σε δύο βασικές αρχές. Η πρώτη αφορά τη χωρική συσχέτιση η οποία υπάρχει σε κάθε καρέ του βίντεο. Η δεύτερη σχετίζεται με το γεγονός ότι πολύ συχνά, ένα καρέ του βίντεο είναι όμοιο σε μεγάλο βαθμό με τους γείτονες του, δηλαδή τα αμέσως προηγούμενα και τα αμέσως επόμενα καρέ. Αυτή ονομάζεται χρονική συσχέτιση. Μία τυπική μέθοδος συμπίεσης βίντεο ξεκινά κωδικοποιώντας το πρώτο καρέ με τη χρήση κάποιας μεθόδου συμπίεσης εικόνας. Συγκεκριμένα ο κωδικοποιητής H.264[61] χρησιμοποιεί τον μετασχηματισμό DCT με μία μέθοδο η οποία αποτελεί μία παραλλαγή του προτύπου JPEG. Στη συνέχεια κάθε καρέ που ακολουθεί μπορεί να οριστεί μέσω των διαφορών ανάμεσα στο τρέχων καρέ και τα αμέσως προηγούμενα του ώστε να κωδικοποιηθούν μόνο οι διαφορές τους. Εάν το τρέχων καρέ διαφέρει σημαντικά σε σχέση με τα προηγούμενα τότε μπορεί να κωδικοποιηθεί και ξεχωριστά χωρίς τη χρήση κάποιου προηγούμενου καρέ. Ένα καρέ το οποίο κωδικοποιείται ανεξάρτητο από τα υπόλοιπα ονομάζεται intra frame ενώ κάθε καρέ που κωδικοποιείται μέσω των διαφορών του σε σχέση με ένα άλλο καρέ ονομάζεται inter frame.

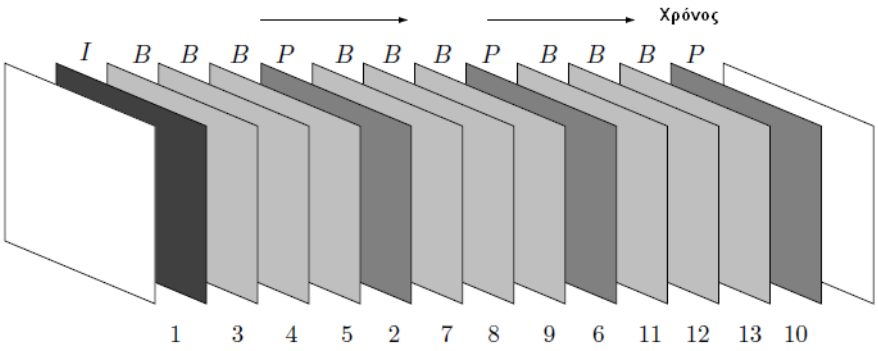
Η συμπίεση βίντεο στη γενική περίπτωση είναι συμπίεση με απώλειες. Η κωδικοποίηση ενός καρέ F_i σε σχέση με κάποιο προηγούμενο καρέ F_{i-1} εισάγει παραμορφώσεις. Ομοίως, η κωδικοποίηση του καρέ F_{i+1}σε σχέση με το F θα αυξήσει την παραμόρφωση ακόμα περισσότερο. Ακόμα και στην περίπτωση όπου επιθυμούμε συμπίεση βίντεο χωρίς απώλειες, υπάρχει το ενδεχόμενο ένα καρέ να χάσει κάποια bit κατά τη μετάδοση. Αυτό θα έχει ως αποτέλεσμα όλα τα επόμενα καρέ τα οποία εξαρτώνται από

αυτό να αποκωδικοποιηθούν λανθασμένα συσσωρεύοντας τα σφάλματα όσο περνούν τα καρέ. Για το λόγο αυτό ανά διαστήματα αποστέλλεται εκ νέου ένα intra frame και δεν μπορούμε σε καμία περίπτωση να αρκεστούμε μόνο στο αρχικό. Κάθε intra frame συμβολίζεται με το γράμμα I ενώ κάθε inter frame με το γράμμα P.

Ο κωδικοποιητής H.264 δεν υποστηρίζει την κωδικοποίηση ενός καρέ μόνο σε σχέση με τα προηγούμενα του και αλλά και από τα αμέσως επόμενα. Τα καρέ τα οποία κωδικοποιούνται τόσο σε σχέση με τα προηγούμενα όσο και σε σχέση με τα επόμενα συμβολίζονται με το γράμμα B. Προφανώς για να αποκωδικοποιηθεί ένα B καρέ θα πρέπει ο αποκωδικοποιητής να γνωρίζει και το αμέσως επόμενο καρέ της ακολουθίας διαφορετικά δε θα είναι σε θέση να το αποκωδικοποιήσει. Δεδομένου ότι τα B καρέ χρησιμοποιούνται πολύ συχνά και αποτελούν μεγάλο μέρος των καρέ σε ένα βίντεο κωδικοποιημένο με H.264, αντιλαμβανόμαστε πως σε πολλές περιπτώσεις η σειρά με την οποία αποστέλλονται τα δεδομένα στο αποκωδικοποιητή είναι διαφορετική από τη σειρά με την οποία θα προβληθούν, για παράδειγμα, από μία συσκευή αποκωδικοποίησης ψηφιακού βίντεο. Συνεπώς μπορεί τα καρέ που θέλουμε να κωδικοποιήσουμε να επιθυμούμε να εμφανιστούν με τη σειρά 1, 2, 3 όμως εφόσον ο κωδικοποιητής αποφασίσει ότι είναι αποδοτικότερο το καρέ 2 να κωδικοποιηθεί ως B καρέ, τότε θα κωδικοποιήσει και θα εγγράψει στη συμπιεσμένη ροή δεδομένων τα καρέ με τη σειρά 1, 3, 2 και με τη σειρά αυτή θα τα αποκωδικοποιήσει ο αποκωδικοποιητής ασχέτως εάν αυτή δεν είναι η επιθυμητή σειρά προβολής τους. Τα Σχήμα 9.1 και Σχήμα 9.2 απεικονίζουν ακριβώς αυτή τη διαδικασία. Βεβαίως ο σειρά προβολής θα πρέπει να γνωστοποιηθεί στο δέκτη διαφορετικά η σειρά προβολής θα είναι όμοια με τη σειρά κωδικοποίησης.



Σχήμα 9.1 : Σειρά Κωδικοποίησης των Καρέ



Σχήμα 9.2 : Σειρά Προβολής των Καρέ

Τα τρία βασικότερα στάδια στη συμπίεση ενός P ή B καρέ είναι (1) ο τεμαχισμός του καρέ σε μάκρο-μπλοκ, (2) η εκτίμηση κίνησης (motion estimation) και (3) η αντιστάθμιση κίνησης (motion compensation)

9.1.2 Τεμαχισμός Εικόνας σε Μακρο-Μπλοκ

Ο κωδικοποιητής H.264 επιχειρεί να κωδικοποιήσει την εικόνα τεμαχίζοντας τη σε μπλοκ (μακρο-μπλοκ) μεγέθους από 16×16 έως 4×4 επιλέγοντας για κάθε μακρο-μπλοκ το μέγεθος που οδηγεί συνολικά στην καλύτερη απόδοση (Σχήμα 9.4). Κάθε μακρο-μπλοκ κωδικοποιείται ξεχωριστά. Συνεπώς σε κάθε καρέ στη γενική περίπτωση θα υπάρχουν μακρο-μπλοκ με μεγέθη που ποικίλουν από 16×16 έως 4×4. Τα μεγάλα μακρο-μπλοκ είναι κατάλληλα για λείες και ομοιόμορφες περιοχές της εικόνας ενώ τα μικρότερα αποδίδουν καλύτερα τις περιοχές με λεπτομέρειες. Μειονέκτημα των μικρών μακρο-μπλοκ αποτελεί το μεγάλο πλήθος διανυσμάτων κίνησης τα οποία πρέπει να αποσταλούν. Ο ρόλος των διανυσμάτων κίνησης θα επεξηγηθεί στο στάδιο εκτίμησης της κίνησης.

9.1.3 Εκτίμηση Κίνησης

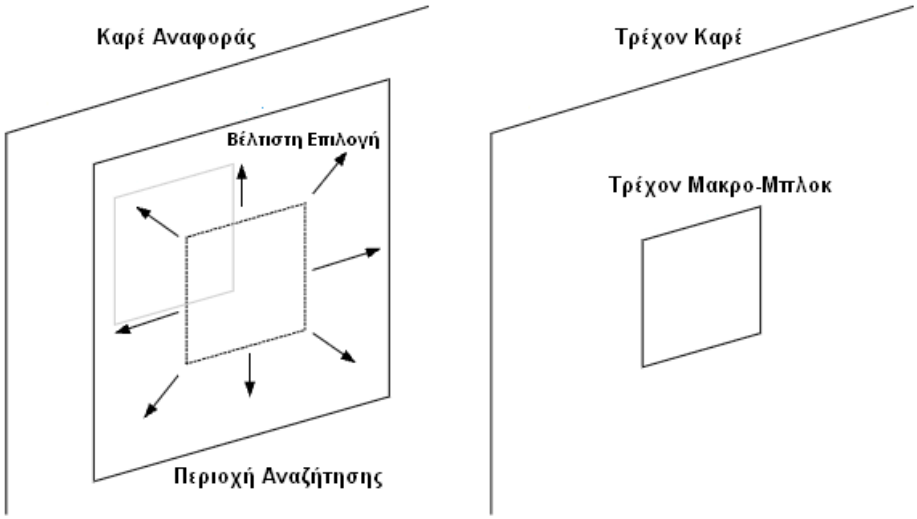
Με τον όρο εκτίμηση κίνησης, αναφερόμαστε στη διαδικασία κατά την οποία για κάθε μακρο-μπλοκ του τρέχοντος καρέ, αναζητείται ένα όμοιο μακρο-μπλοκ σε προηγούμενο ή σε επόμενο καρέ. Οι μετρικές οι οποίες χρησιμοποιούνται ως κριτήριο ομοιότητας συνήθως είναι η μέση απόλυτη διαφορά (MAD) ή το μέσο τετραγωνική σφάλμα (MSE) οι σχέσεις των οποίων ορίζονται ως:

$$MAD = \frac{1}{b^2}\sum_{i=1}^{b}\sum_{j=1}^{b}\left|B_{ij} - C_{ij}\right| \tag{9.1}$$

$$MSE = \frac{1}{b^2}\sum_{i=1}^{b}\sum_{j=1}^{b}\left(B_{ij} - C_{ij}\right)^2 \tag{9.2}$$

όπου C το μακρο-μπλοκ του τρέχοντος καρέ και B του προηγούμενου ή επόμενου.

Το εύρος της αναζήτησης ορίζεται από το χρήστη, ενώ οι τιμές των MAD και MSD οι οποίες θεωρούνται ή όχι αποδεκτές υπολογίζονται από τον κωδικοποιητή αναλόγως με την ποιότητα και το ρυθμό δεδομένων που επιθυμεί ο χρήστης. Πρόκειται για την πλέον χρονοβόρα διαδικασία και για το λόγο αυτό πλήθος αλγορίθμων έχει προταθεί ώστε να αποφεύγεται η εξαντλητική αναζήτηση με προσεγγιστικές μεθόδους.



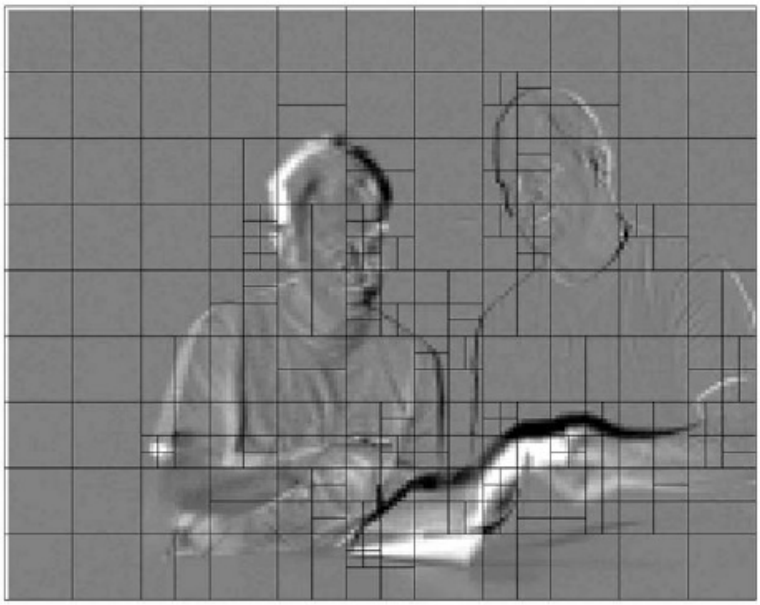
Σχήμα 9.3 : Διαδικασία Εκτίμησης Κίνησης

Αφού ο αλγόριθμος ανιχνεύσει το μακρο-μπλοκ με το μικρότερο MAD ή MSD παράγονται στη συνέχεια τα διανύσματα κίνησης. Το διάνυσμα κίνησης είναι ένα προσημασμένο ζεύγος συντεταγμένων (x, y) το οποίο προσδιορίζει τη σχετική θέση στην οποία βρέθηκε το ομοιότερο μακρο-μπλοκ. Η διαδικασία σε κάθε μακρο-μπλοκ επαναλαμβάνεται για μπλοκ μεγέθους 16×16 έως 4×4 και στο τέλος ο κωδικοποιητής αφού προσθέσει τα bit που απαιτούνται για την κωδικοποίηση και των διανυσμάτων κίνησης (προφανώς για μικρά μακρο-μπλοκ απαιτούνται περισσότερα διανύσματα κίνησης) και συνυπολογίσει το σφάλμα που υπεισέρχεται σε κάθε περίπτωση, λαμβάνει τη βέλτιστη απόφαση σχετικά με το μέγεθος του μακρο-μπλοκ που τελικά θα χρησιμοποιηθεί.

9.1.4 Αντιστάθμιση Κίνησης

Από τη στιγμή που το καρέ αναφοράς είναι γνωστό και έχει υπολογιστεί το διάνυσμα

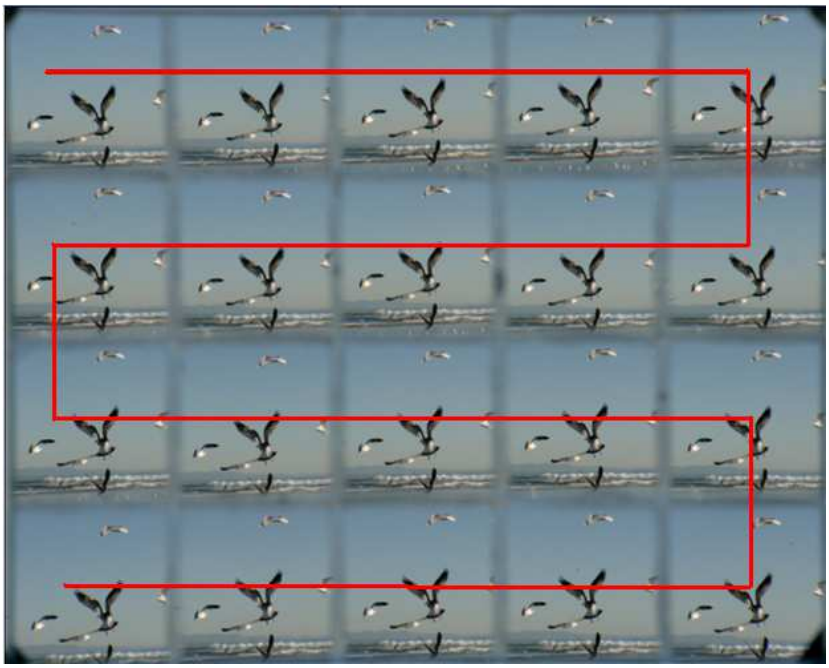
κίνησης, ο κωδικοποιητής αφαιρεί από το μακρο-μπλοκ του τρέχοντος καρέ το μακρο-μπλοκ του καρέ αναφοράς που υποδεικνύει το διάνυσμα κίνησης και κωδικοποιεί τη διαφορά τους (residual) η οποία στην ιδανική περίπτωση θα είναι ένα μπλοκ γεμάτο μηδενικά. Στην πράξη η διαφορά αυτή θα είναι συνήθως ένας μικρός αριθμός. Η κωδικοποιημένη διαφορά αποστέλλεται μαζί με τα διανύσματα κίνησης ώστε να είναι εφικτή η αποκωδικοποίηση της εικόνας στον δέκτη μαζί με κάποια bit που υποδηλώνουν το μέγεθος του μακρο-μπλοκ. Ο Η.264 έχει τη δυνατότητα να κωδικοποιήσει κάποια μακρο-μπλοκ ανεξάρτητα χωρίς τη χρήση καρέ αναφοράς εάν το μέτρο του σφάλματος είναι υψηλό και δεν ικανοποιεί τις απαιτήσεις σε ποιότητα. Επίσης αξίζει να σημειωθεί πως επειδή τα διανύσματα κίνησης γειτονικών μακρο-μπλοκ είναι συσχετισμένα δεν αποστέλλονται τα ίδια τα διανύσματα αλλά η διαφορά τους.



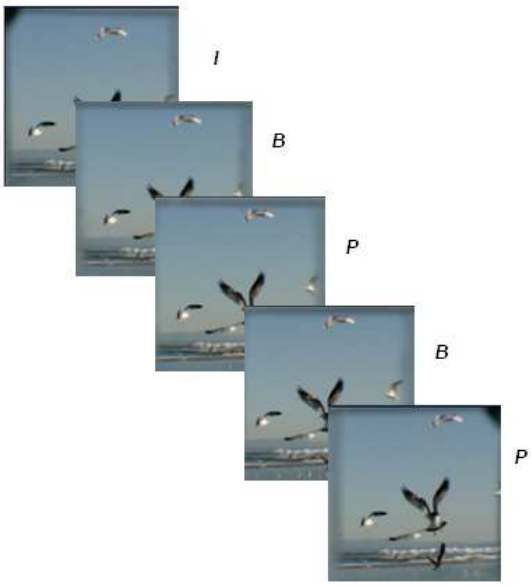
Σχήμα 9.4 : Τεμαχισμός Εικόνας σε Μκρο-Μπλοκ Διαφορετικών Μεγεθών και Αντιστάθμιση τους

9.1.5 Ανάπτυξη Μεθόδου

Η βασική ιδέα πίσω από τη συμπίεση της ολοκληρωτικής εικόνας ως βίντεο-ροή βρίσκεται στην εκμετάλλευση της συσχέτισης μεταξύ των στοιχειωδών εικόνων. Εάν θεωρήσουμε κάθε στοιχειώδη εικόνα ως ένα καρέ και τοποθετήσουμε γειτονικές στοιχειώδεις εικόνες με κάποιο τρόπο σε μία βίντεο-ροή τότε ο κωδικοποιητής Η.264 θα μπορέσει να εκμεταλλευτεί τη συσχέτιση τους και να κωδικοποιήσει ολόκληρη την εικόνα με λίγα I καρέ μεγέθους όσο μία στοιχειώδης εικόνα και όλα τα υπόλοιπα με χρήση P και B καρέ.



Εικόνα 9.2 : Παράδειγμα Σάρωσης Ολοκληρωτικής Εικόνας με το σχήμα των Παράλληλων Γραμμών

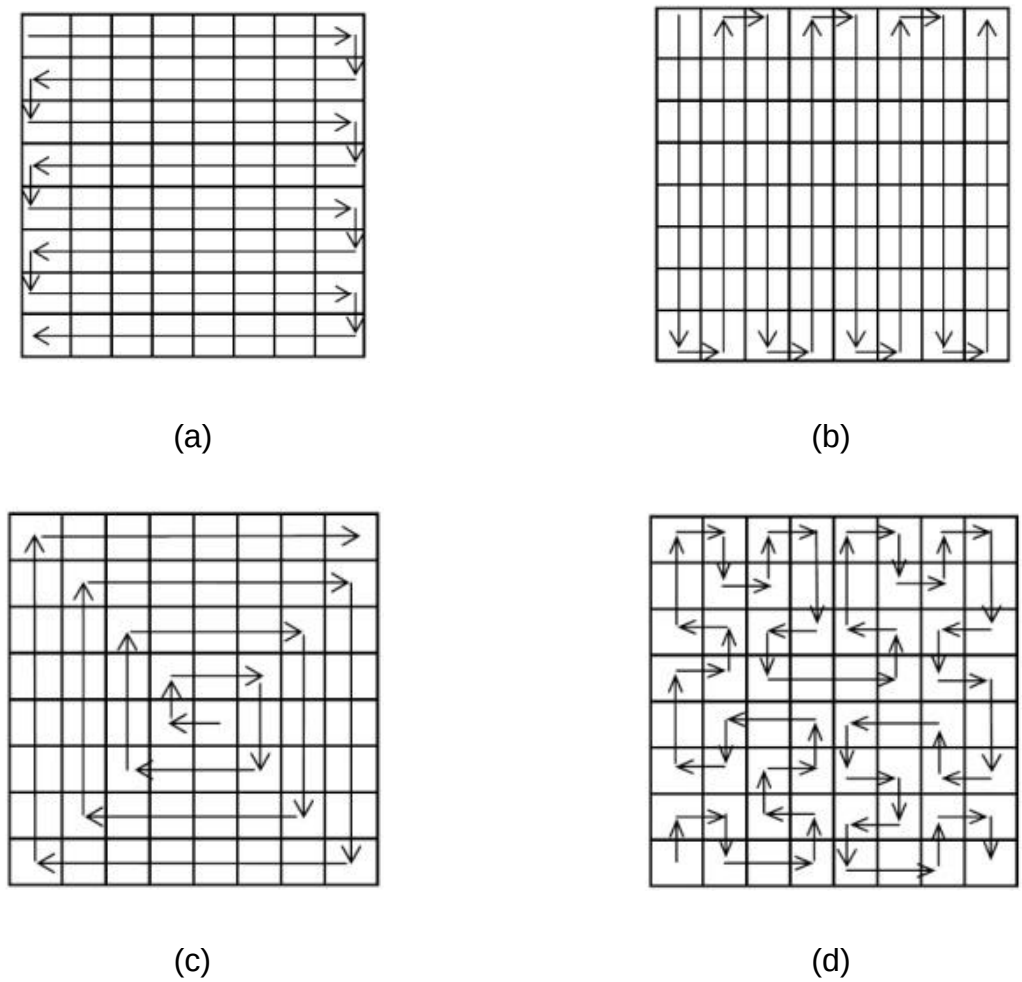


Εικόνα 9.3 : Στοιχειώδεις Εικόνες ως Καρέ Βίντεο-Ροής Μετά το Σάρωμα που Απεικονίζεται στην Εικόνα 9.2. Στην εικόνα Παρουσιάζονται τα 5 Πρώτα Καρέ.

Ο τρόπος με τον οποίο η εικόνα μπορεί να διασπαστεί σε μία ακολουθία από καρέ όπου κάθε καρέ αντιπροσωπεύει μία στοιχειώδη εικόνα είναι ο ακόλουθος. Διαιρώντας το μέγεθος της εικόνας με το πλήθος των μικρο-φακών σε κάθε διάσταση υπολογίζεται το μέγεθος κάθε στοιχειώδους εικόνας. Στη συνέχεια, η εξαγωγή των στοιχειωδών εικόνων

μία προς μία από την ολοκληρωτική εικόνα και η εγγραφή τους σε μία ακολουθία βίντεο επιτυγχάνεται εύκολα με τη χρήση του υπολογιστικού πακέτου Matlab.

Η σειρά με την οποία θα επιλεγούν οι στοιχειώδεις εικόνες να εγγραφούν στη βίντεο-ροή μπορεί να είναι οποιαδήποτε όμως επιλέγεται ένα σχήμα σάρωσης[56] το οποίο θα συγκεντρώσει όσο το δυνατόν καλύτερα τις γειτονικές στοιχειώδεις εικόνες αφού αυτές θα είναι εντονότερα συσχετισμένες μεταξύ τους. Τα γνωστότερα σχήματα σάρωσης είναι το παράλληλο οριζόντιο, παράλληλο κάθετο, σπιράλ και Hilbert και παρουσιάζονται στο Σχήμα 9.5.



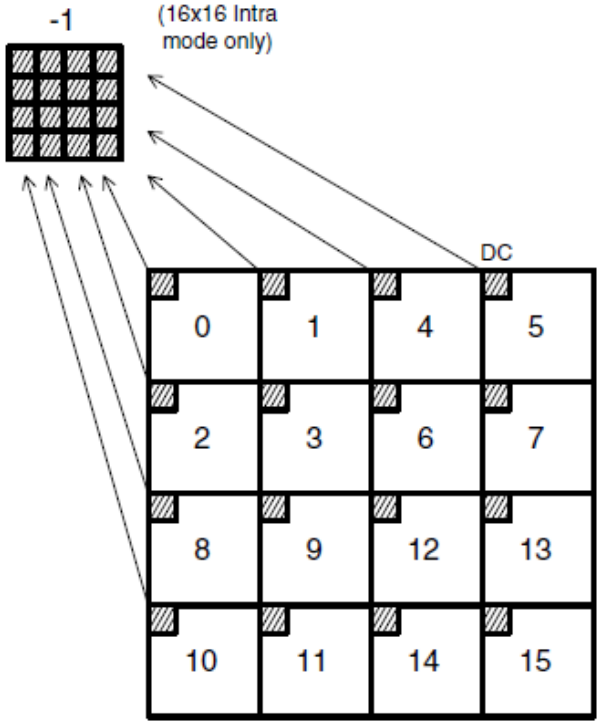
Σχήμα 9.5 : Δημοφιλή Σχήματα Σάρωσης των Στοιχειωδών Εικόνων Ολοκληρωτικής Εικόνας (a) Παράλληλο Οριζόντιο, (b) Παράλληλο Κάθετο, (c) Σπιράλ και (d) Hilbert

9.1.6 Κωδικοποίηση Ολοκληρωτικής Εικόνας ως Καρέ Τύπου I (Intra)

Αναφέρθηκε στην αρχή της υποενότητας αυτής πως το πρώτο καρέ σε μία βίντεο-ροή κωδικοποιείται με μία τεχνική η οποία αποτελεί παραλλαγή του προτύπου συμπίεσης

JPEG. Όπως θα αποδειχθεί στην επόμενη ενότητα της εργασίας, η κωδικοποίηση των δεδομένων με τον συγκεκριμένο κωδικοποιητή είναι εξαιρετικά αποδοτική, συνεπώς έχει ιδιαίτερη αξία η αναφορά στον τρόπο με τον οποίο ο κωδικοποιητής H.264 Intra βελτιστοποιεί την απόδοση του JPEG.

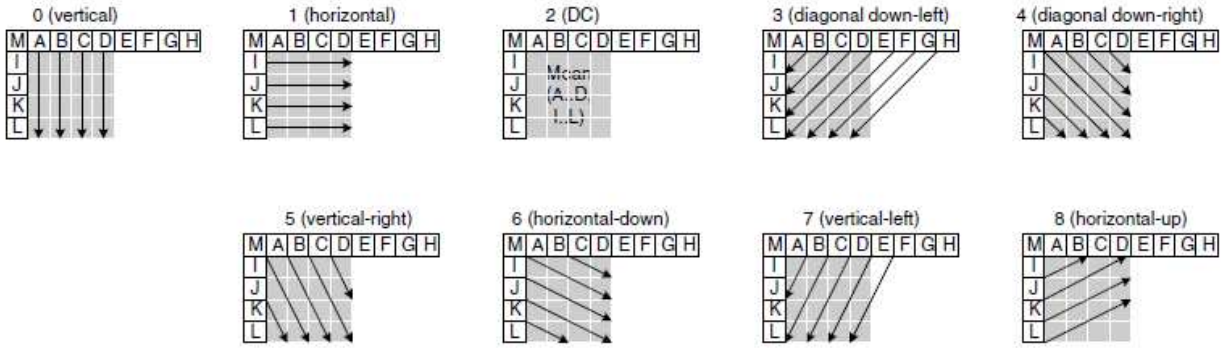
Στην πράξη, ο H.264 συνδυάζει τη μέθοδο συμπίεσης χωρίς απώλειες του JPEG με τη μέθοδο συμπίεσης με απώλειες. Αρχικά για κάθε μπλοκ μεγέθους 16×16 ο κωδικοποιητής δημιουργεί μία πρόγνωση βάσει των γειτονικών μπλοκ τα οποία έχουν κωδικοποιηθεί και ανακατασκευαστεί νωρίτερα. Αξίζει να σημειωθεί πως την ίδια διαδικασία επαναλαμβάνει για μπλοκ μεγέθους 4×4 επιλέγοντας τελικά το μέγεθος εκείνο που οδηγεί στη βέλτιστη λύση. Στη συνέχεια το προγνωσθέν μπλοκ αφαιρείται από το τρέχον και η διαφορά τους μετασχηματίζεται με τον μετασχηματισμό DCT 4×4. Τέλος οι (4×4) DC συνιστώσες που περιέχονται στο 16×16 μπλοκ, μετασχηματίζονται με τον μετασχηματισμό Hadamard όπως παρουσιάζει το .



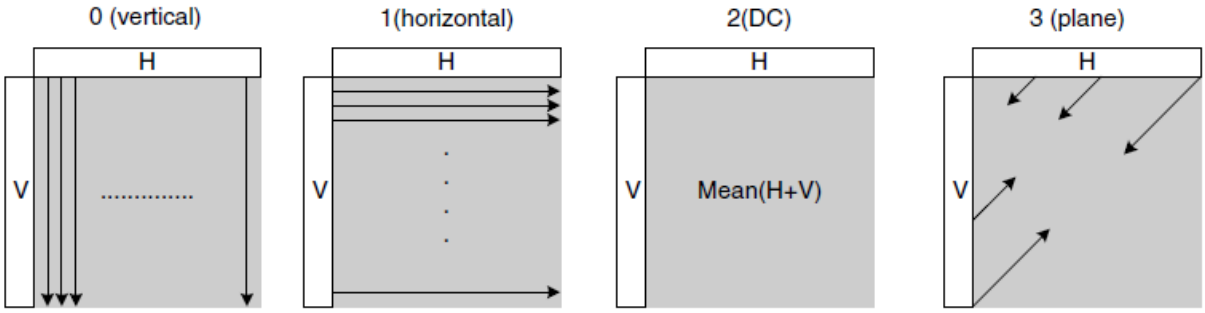
Σχήμα 9.6 : Σάρωση των DC Συνιστωσών των 4×4 Μπλοκ για Εφαρμογή του Μετασχηματισμού Hadamard

Λόγω του συνδυασμού των δύο μεθόδων καθώς και του πλήθους επιλογής σχημάτων πρόγνωσης ο κωδικοποιητής H.264 Intra προσφέρει εξαιρετική απόδοση η οποία όπως θα αποδειχθεί και από τις μετρήσεις του ακόλουθου κεφαλαίου, δεν ξεπερνά μόνο το

JPEG αλλά ακόμα και το JPEG 2000. Παρακάτω, παρουσιάζονται τα διαθέσιμα σχήματα πρόγνωσης όταν ο κωδικοποιητής χρησιμοποιεί μπλοκ μεγέθους 16×16 και 4×4.



Σχήμα 9.7 : Σχήματα πρόγνωσης του κωδικοποιητή H.264 σε Μπλοκ Μεγέθους 4×4



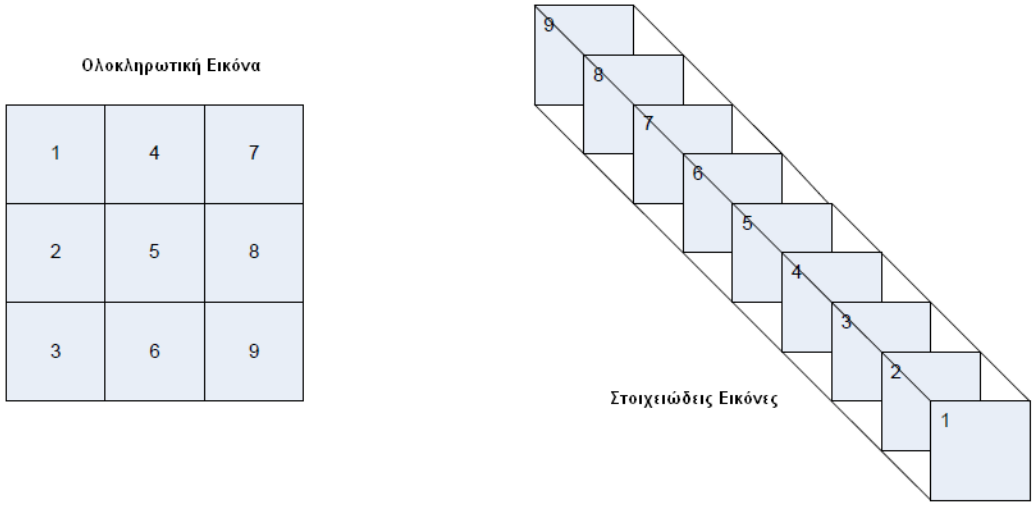
Σχήμα 9.8 : Σχήματα πρόγνωσης του κωδικοποιητή H.264 σε Μπλοκ Μεγέθους 16×16

9.2 Υβριδική Μέθοδος Συμπίεσης με τη Χρήση των Μετασχηματισμών Κυματιδίου και DCT

Η υβριδική μέθοδος συμπίεσης προτάθηκε από τον Javidi[57] στοχεύοντας στην αξιοποίηση των πλεονεκτημάτων των μετασχηματισμών DCT και κυματιδίου προκειμένου να συμπιεστούν αποδοτικότερα οι ολοκληρωτικές εικόνες. Η κεντρική ιδέα της μεθόδου μοιάζει αρκετά με την μέθοδο συμπίεσης της ολοκληρωτικής εικόνας ως βίντεο-ροή που συζητήθηκε σε προηγούμενη ενότητα. Κοινό σημείο και των δύο προσεγγίσεων είναι η διάσπαση της αρχικής εικόνας σε στοιχειώδεις εικόνες και η εκμετάλλευση της συσχέτισης τους. Η προηγούμενη μέθοδος εκμεταλλεύτηκε τη συσχέτιση με χρήση τεχνικών βίντεο, δηλαδή πρακτικά στο πεδίο του χρόνου. Η υβριδική μέθοδος επιχειρεί να εκμεταλλευτεί τη συσχέτιση στο πεδίο της συχνότητας εφαρμόζοντας τον μετασχηματισμό DCT στις ζώνες συχνοτήτων που προκύπτουν από την εφαρμογή του μετασχηματισμού κυματιδίου σε κάθε στοιχειώδη εικόνα δημιουργώντας με αυτό τον τρόπο έναν τετραδιάστατο μετασχηματισμό (4D). Στη συνέχεια θα παρουσιαστεί αναλυτικά η υλοποίηση της μεθόδου.

9.2.1 Τεμαχισμός Ολοκληρωτικής Εικόνας

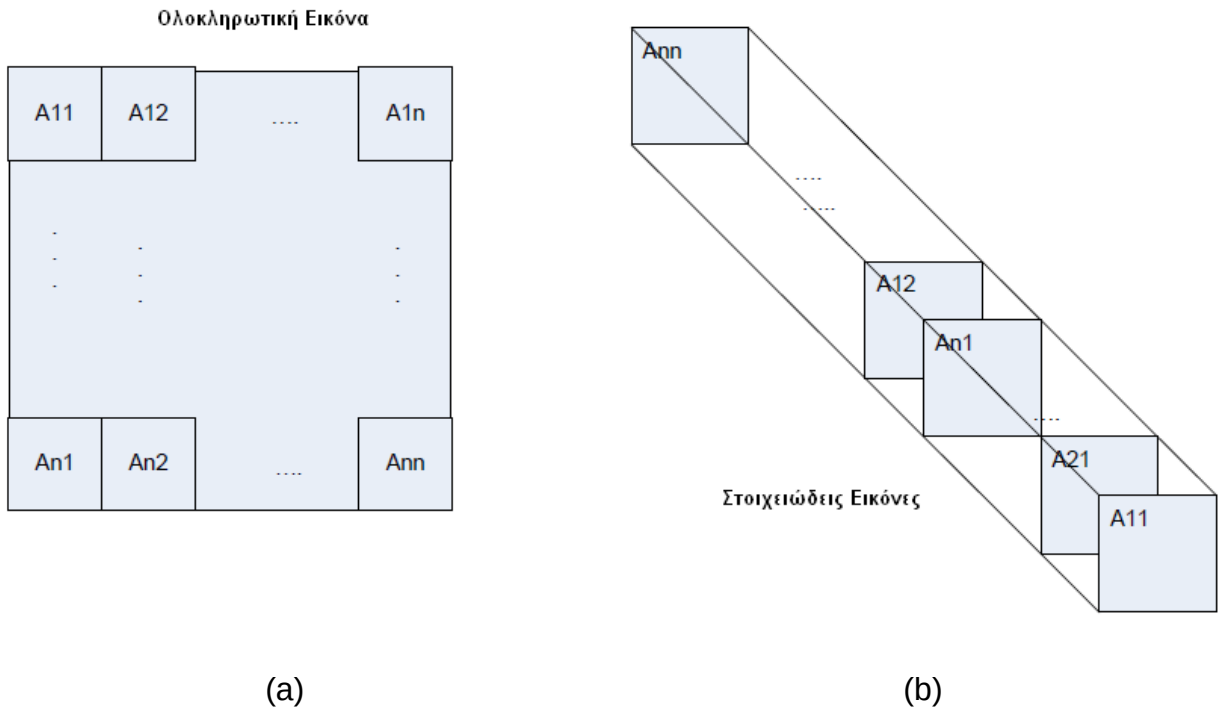
Το πρώτο στάδιο υλοποίησης της μεθόδου όπως έχει ήδη προαναφερθεί αποτελείται από την τμηματοποίηση της αρχικής εικόνας στις επιμέρους στοιχειώδεις εικόνες και αποθήκευση τους σε κατάλληλες δομές δεδομένων. Η γραφική απεικόνιση της τμηματοποίησης παρουσιάζεται στο Σχήμα 9.9.



(a) (b)

Σχήμα 9.9 : (a) Ολοκληρωτική Εικόνα και (b) Τμηματοποίηση της σε Στοιχειώδεις Εικόνες

Ως παράδειγμα θεωρείται μία ολοκληρωτική εικόνα αποτελούμενη από εννέα στοιχειώδεις εικόνες και η δομή στην οποία αποθηκεύονται είναι ένας τρισδιάστατος πίνακας με τις δύο διαστάσεις να δεικτοδοτούν το εικονοστοιχείο της στοιχειώδους εικόνας και την τρίτη διάσταση τη στοιχειώδη εικόνα. Το Σχήμα 9.10 παρουσιάζει τη διαδικασία και την αντίστοιχη δομή δεδομένων στη γενική περίπτωση

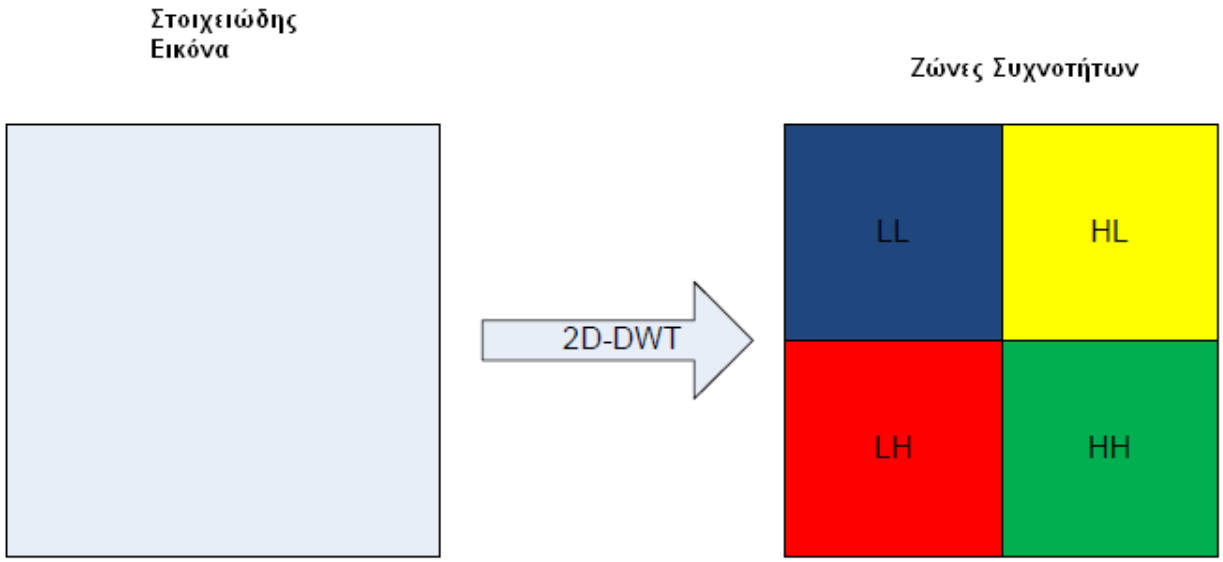


Σχήμα 9.10 : (a) Ολοκληρωτική Εικόνα και (b) Τμηματοποίηση της σε Στοιχειώδεις Εικόνες στη Γενική Περίπτωση

9.2.2 Εφαρμογή του Μετασχηματισμού Κυματιδίου

Το επόμενο στάδιο της μεθόδου υποδεικνύει την εφαρμογή του μετασχηματισμού κυματιδίου ενός επιπέδου σε κάθε στοιχειώδη εικόνα. Ο μετασχηματισμός κυματιδίου ενός επιπέδου θα δημιουργήσει στην έξοδο του τέσσερεις ζώνες συχνοτήτων για κάθε στοιχειώδη εικόνα. Ένα ερώτημα που τίθεται είναι για ποιο λόγο εφαρμόζεται ο μετασχηματισμός κυματιδίου και όχι ο DCT. Όπως είχαμε μελετήσει σε προηγούμενη ενότητα, ο μετασχηματισμός DCT είναι εξαιρετικά αποτελεσματικός στη συγκέντρωση της ενέργειας σε λίγους συντελεστές, αποσυσχετίζοντας τα δεδομένα, όμως πρόκειται για έναν μπλοκ μετασχηματισμό. Αυτό έχει ως αποτέλεσμα τη δημιουργία εμφανών ατελειών στις ακμές των μπλοκ μεγέθους 8×8 της ανακατασκευασμένης εικόνας στις περιπτώσεις όπου υπάρχουν λεπτομέρειες (υψηλότερες συχνότητες) είτε ο ρυθμός δεδομένων είναι σχετικά χαμηλός λόγω της κβάντισης. Αντίθετα η κβάντιση στο μετασχηματισμό κυματιδίου επιφέρει ομοιόμορφη απώλεια πληροφορίας σε ολόκληρη

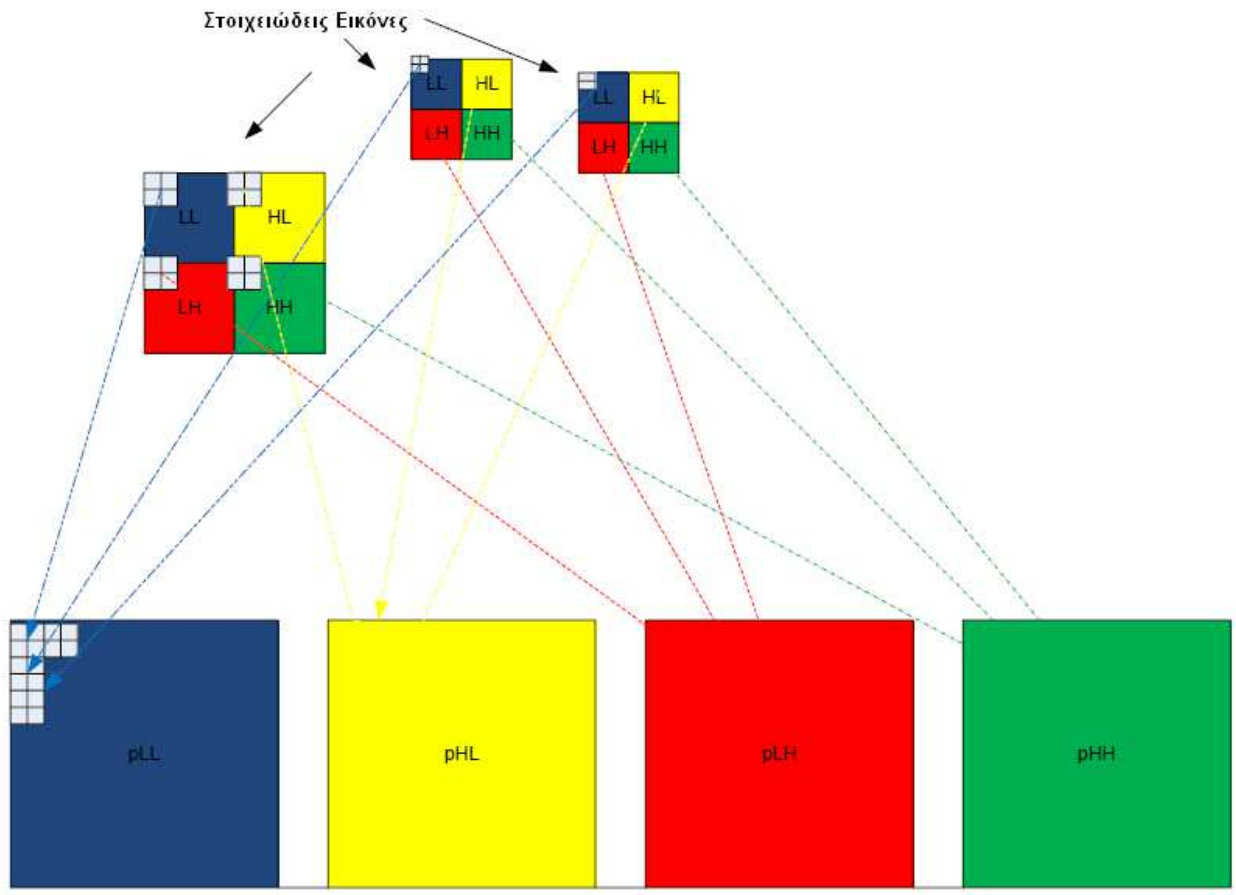
την εικόνα και συνεπώς δεν εμφανίζονται μπλοκ που δημιουργούν έντονες ακμές με ατέλειες. Επίσης ο μετασχηματισμός κυματιδίου παράγει ζώνες συχνοτήτων συνεπώς υπάρχουν ακόμα χαρακτηριστικά προς εκμετάλλευση από τον μετασχηματισμό DCT σε κάθε ζώνη (ειδικά στη ζώνη χαμηλών συχνοτήτων) και όχι ασυσχέτιστους συντελεστές που παράγει ο μετασχηματισμός DCT. Η διαδικασία παρουσιάζεται στο Σχήμα 9.11 και επαναλαμβάνεται για κάθε στοιχειώδη εικόνα.



Σχήμα 9.11 : Εφαρμογή Μετασχηματισμού Κυματιδίου σε κάθε Στοιχειώδη Εικόνα

9.2.3 Διαχωρισμός Πακέτων

Η διαδικασία διαχωρισμού πακέτων αναφέρεται στην ομαδοποίηση των συντελεστών των γειτονικών στοιχειωδών εικόνων που αντιστοιχούν στις ίδιες ζώνες συχνοτήτων. Αυτό θα έχει ως αποτέλεσμα τη δημιουργία τεσσάρων πινάκων όπου ο κάθε πίνακας θα αποτελείται από το σύνολο των συντελεστών των στοιχειωδών εικόνων για μία συγκεκριμένη ζώνη συχνοτήτων. Ο τρόπος με τον οποίο θα τοποθετηθούν τα πακέτα σε κάθε πίνακα θα πρέπει να είναι κατάλληλος ώστε γειτονικές στοιχειώδεις εικόνες θα τοποθετηθούν γειτονικά και στο πεδίο των ζωνών συχνοτήτων. Τόσο το μέγεθος των πακέτων όσο και ο τρόπος με τον οποίο θα τοποθετηθούν στον πίνακα που αντιπροσωπεύει κάθε ζώνη συχνοτήτων αποτελούν σημαντικές παραμέτρους στην απόδοση της διαδικασίας[62]. Το Σχήμα 9.12 περιγράφει χαρακτηριστικά την διαδικασία διαχωρισμού των πακέτων.



Σχήμα 9.12 : Διαδικασία Διαχωρισμού Πακέτου

9.2.4 Εφαρμογή του Μετασχηματισμού DCT

Έχοντας κατασκευάσει του τέσσερεις πίνακες που αντιστοιχούν στις τέσσερεις ζώνες συχνοτήτων του μετασχηματισμού κυματιδίου, το τελευταίο στάδιο της μεθόδου πριν το κβάντιση και την κωδικοποίηση εντροπίας είναι η εφαρμογή του μετασχηματισμού DCT σε κάθε έναν από τους τέσσερεις πίνακες. Το μέγεθος του μπλοκ δεν είναι απαραίτητο να οριστεί όμοιο με αυτό που χρησιμοποιείται στο JPEG καθώς εξαρτάται άμεσα από την επιλογή του μεγέθους των πακέτων και του τρόπου τοποθέτησης τους κατά το βήμα διαχωρισμού.

9.2.5 Ολοκλήρωση της Μεθόδου

Η μέθοδος ολοκληρώνεται με τις τυπικές τεχνικές συμπίεσης που παρουσιάζονται στο JPEG. Συγκεκριμένα κάθε πίνακας κβαντίζεται με έναν πίνακα κβάντισης. Αξίζει να σημειωθεί ότι οι πίνακες με τις υψηλότερες συχνότητες μπορούν να κβαντιστούν εντονότερα σε σχέση για παράδειγμα με τον πίνακα pLL που περιέχει τις χαμηλές συχνότητες. Οι DC συντελεστές κωδικοποιούνται διαφορικά σε σχέση με τους DC συντελεστές γειτονικών μπλοκ εκτός από τον πρώτο που κωδικοποιείται ανεξάρτητα.

Ακολουθεί η κωδικοποίηση των AC συντελεστών με τη μέθοδο τρέχοντος μήκους μηδενικών αφού πρώτα σαρωθούν με το σχήμα zig-zag όπως και στο πρότυπο συμπίεσης JPEG.

9.3 Μέθοδος Προ-Επεξεργασίας Ολοκληρωτικών Εικόνων

9.3.1 Εισαγωγή στη Μέθοδο

Μία μέθοδος με σκοπό την αποδοτικότερη κωδικοποίηση ολοκληρωτικών εικόνων προτάθηκε από τον Olsson[54][55]. Βάση της μεθόδου αποτελεί η τμηματοποίηση της εικόνας στις επιμέρους στοιχειώδεις εικόνες και η εφαρμογή ενός σταδίου προ-επξεργασίας σε αυτές. Αντίστοιχη λογική ακολουθήθηκε και στα πρώτα βήματα της υβριδικής μεθόδου συμπίεσης που παρουσιάστηκε στην προηγούμενη ενότητα. Ειδοποιό διαφορά των δύο μεθόδων αποτελεί η εφαρμογή της προ-επεξεργασίας σε διαφορετικά πεδία. Η υβριδική μέθοδος μετά την τμηματοποίηση της εικόνας στις στοιχειώδεις εικόνες εργάζεται σχεδόν εξολοκλήρου στο πεδίο των συχνοτήτων. Ο Olsson πρότεινε μία λιγότερο εκλεπτυσμένη μέθοδο σύμφωνα με την οποία τα εικονοστοιχεία που βρίσκονται στην ίδια θέση διαφορετικών στοιχειωδών εικόνων, θα ομαδοποιούνται σε μπλοκ. Στη συνέχεια τα μπλοκ αυτά θα επανατοποθετηθούν βάσει κάποιου σχήματος δημιουργώντας μία εικόνα μεγέθους όμοιου με της αρχικής.

Όπως γίνεται άμεσα αντιληπτό, η μέθοδος εφαρμόζεται στο πεδίο του χώρου με σκοπό την ομαδοποίηση όσο το δυνατόν περισσότερων εικονοστοιχείων με όμοιες τιμές φωτεινότητας. Αυτό θα έχει ως αποτέλεσμα, οι νέες εικόνες που θα προκύψουν από το χωρικό μετασχηματισμό να συμπιεστούν αποδοτικότερα όταν περαστούν από τους κωδικοποιητές JPEG και JPEG 2000 αφού η συσχέτιση μεταξύ των γειτονικών εικονοστοιχείων της νέας εικόνας θα είναι ισχυρότερη σε σχέση με την αρχική. Η μέθοδος προτείνει επίσης και τη χρήση του κωδικοποιητή H.264 για την κωδικοποίηση των διαδοχικών μετασχηματισμένων μπλοκ μετά την εγγραφή του κάθε μπλοκ ως καρέ με τη χρήση κάποιου σχήματος σάρωσης. Αδυναμία της μεθόδου ενδεχομένως να αποτελεί η φύση της ολοκληρωτικής εικόνας καθώς η απόδοση συνήθως εξαρτάται από το πλήθος των στοιχειωδών εικόνων.

9.3.2 Ανάπτυξη της Μεθόδου

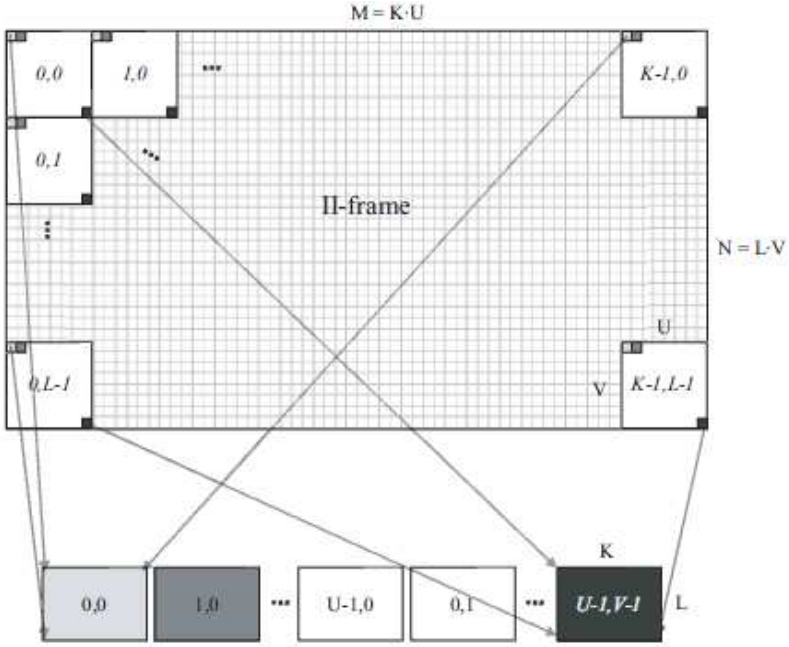
Συμβολίζουμε ολοκληρωτική εικόνα μεγέθους $M \times N$ εικονοστοιχείων ως $II(m,n)$, όπου $m = 0,1,\dots,M-1$ και $n = 0,1,\dots,N-1$ αντιστοιχούν στην οριζόντια και κάθετη θέση ενός εικονοστοιχείου αντίστοιχα. Συνεπώς το $II(m_0,n_0) = [ii_R, ii_G, ii_B]^T$ αντιστοιχεί στα

τρία χρώματα RGB του εικονοστοιχείου στην $m_0^{\iota o \sigma \tau \acute{\eta}}$ στήλη της $n_0^{\iota o \sigma \tau \acute{\eta} \varsigma}$ γραμμής.

Ένα πλήθος από στοιχειώδεις εικόνες αποτελεί την ολοκληρωτική εικόνα όπως παρουσιάζεται στο Σχήμα 9.13. Η αντιστοίχηση ανάμεσα στα εικονοστοιχεία της ολοκληρωτικής εικόνας και η κατανομή τους στις επιμέρους στοιχειώδεις εικόνες δίνεται από την ακόλουθη σχέση

$$EI_{k,l}(u,v) = II(k \cdot U + u, l \cdot V + u)$$

όπου k, l η στοιχειώδης εικόνα που βρίσκεται στη γραμμή k και τη στήλη l ενώ όπου $u = 0,1,\dots,U-1$ και $v = 0,1,\dots,V-1$ είναι η οριζόντια και κατακόρυφη θέση του εικονοστοιχείο εντός της στοιχειώδους εικόνας. Επομένως κάθε μία από τις $K \times L$ στοιχειώδεις εικόνες αποτελείται από $U \times V$ εικονοστοιχεία.



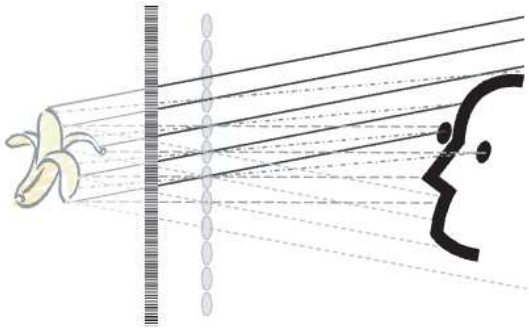
Σχήμα 9.13 : Ολοκληρωτική Εικόνα και η Τμηματοποίηση της σε Υπο-Εικόνες

Προκειμένου να βελτιωθεί η απόδοση της εκτίμησης του βάθους, εισάγεται η ιδέα των υπο-εικόνων (SI). Σύμφωνα με αυτή εξάγονται και συνδυάζονται τα εικονοστοιχεία μίας ολοκληρωτικής εικόνας τα οποία έχουν την ίδια θέση στον οριζόντιο και κατακόρυφο άξονα κάθε στοιχειώδους εικόνας. Η υπο-εικόνα ορίζεται ως:

$$SI_{u,v}(k,l) = II(k \cdot U + u, l \cdot V + u)$$

όπου $k = 0,1,\dots,K-1$ και $l = 0,1,\dots,L-1$ η θέση του εικονοστοιχείου στον οριζόντιο και κατακόρυφο άξονα αντίστοιχα σε μία υπο-εικόνα. Συνεπώς, κάθε μία από τις $U \times V$ υπο-εικόνες έχει ανάλυση $K \times L$ εικονοστοιχείων. Στο Σχήμα 9.14 οι ακτίνες οι οποίες

τέμνονται από τις παράλληλες γραμμές του ίδιου τύπου ανήκουν στην ίδια υπο-εικόνα.



Σχήμα 9.14 : Οθόνη Ολοκληρωτικής Απεικόνισης με Συστοιχία Μικρο-Φακών. Οι Ακτίνες οι οποίες Τέμνονται από τις Παράλληλες γραμμές του Ίδιου Τύπου Ανήκουν στην Ίδια Υπο-Εικόνα

Ένας ακόμα χωρικός μετασχηματισμός, ο οποίος είχε προταθεί από τον Javidi για άλλες εφαρμογές, μετασχηματίζει την εικόνα στο πεδίο των ακτινών και ορίζεται ομοίως με τον προηγούμενο μετασχηματισμό σύμφωνα με τη σχέση:

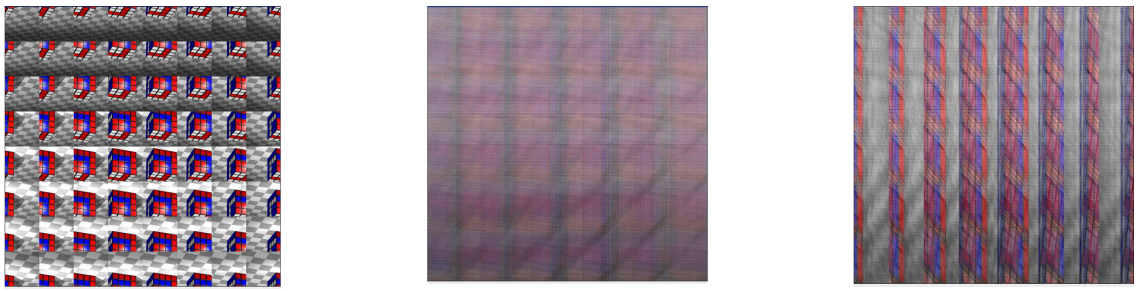
$$RI_{u,l}(k,v) = II(k \cdot U + v, l \cdot V + u)$$

Οι προτεινόμενες μέθοδοι για τη συμπίεση των μετασχηματισμένων εικόνων είναι:

a) Η εφαρμογή των κωδικοποιητών JPEG και JPEG 2000 στις εικόνες SI

b) Η εφαρμογή των κωδικοποιητών JPEG και JPEG 2000 στις εικόνες RI

c) Η σάρωση των SIs και RIs με κάποιο σχήμα (paraline) και η εγγραφή τους ως καρέ σε μία ακολουθία βίντεο προκειμένου να περαστούν από τον κωδικοποιητή H.264.



(a) (b) (c)

Εικόνα 9.4 : (a) Παράδειγμα Ολοκληρωτικής Εικόνας (b) Ο Μετασχηματισμός της σε Υπο-Εικόνες και (c) Ο Μετασχηματισμός της στο Πεδίο των Ακτίνων

10. Μετρήσεις Απόδοσης Κωδικοποιητών

Στην ενότητα 9 περιγράφηκαν οι υπάρχουσες τεχνικές συμπίεσης ολοκληρωτικών εικόνων. Σε αυτή την ενότητα θα παρατεθεί ένα πλήθος διαγραμμάτων τα οποία θα αναδεικνύουν την τεχνική συμπίεσης ολοκληρωτικών εικόνων η οποία παρέχει την καλύτερη δυνατή συμπίεση.

Η αξιολόγηση της ποιότητας των εικόνων θα πραγματοποιηθεί με τη χρήση της αντικειμενικής μετρικής του λόγου του σήματος προς το θόρυβο (PSNR) η οποία δίνεται από τη σχέση (10.1)

$$PSNR = 10 \cdot \log_2 \left(\frac{255^2}{MSE} \right) \tag{10.1}$$

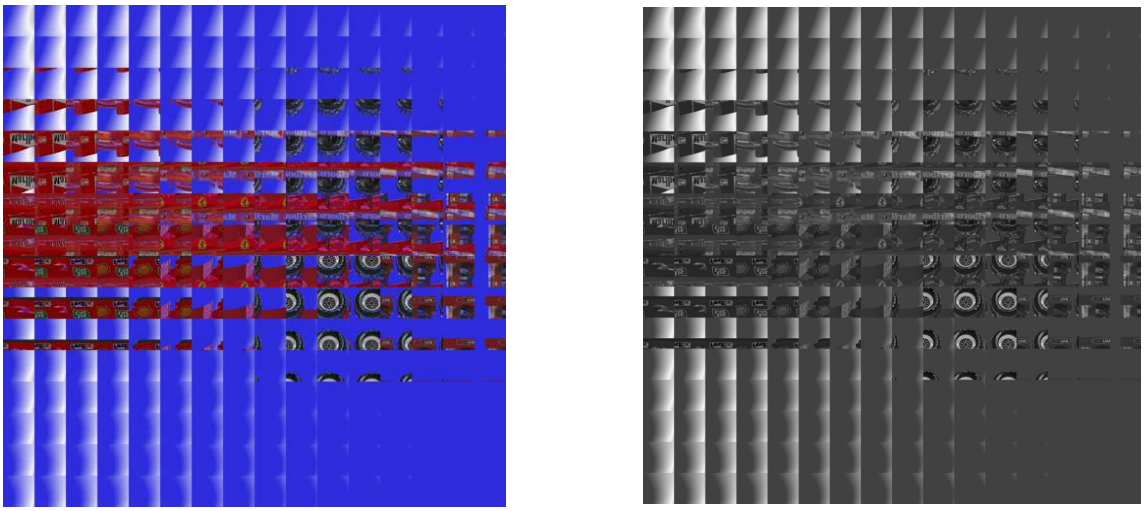
όπου το MSE δίνεται από τη σχέση (9.2) και το B συμβολίζει την αρχική εικόνα ενώ το C την ανακατασκευασμένη μετά από συμπίεση. Το μέγεθος του PSNR μετράται σε db ενώ τα διαγράμματα που θα ακολουθήσουν θα παραστήσουν γραφικά το πλήθος των bit που απαιτείται ανά κωδικοποιημένο εικονοστοιχείο (bpp) σε σχέση με την ποιότητα (PSNR) που επιτυγχάνεται.

Προκειμένου να πραγματοποιηθούν οι μετρήσεις, δημιουργήθηκαν σε πρόγραμμα προσομοίωσης και χρησιμοποιήθηκαν, τρείς βασικές εικόνες μεγέθους 2048×2048 εικονοστοιχείων, οι οποίες αποτελούνται από 16×16 στοιχειώδεις εικόνες. Όλες οι μετρήσεις πραγματοποιήθηκαν σε εικόνες 256 αποχρώσεων του γκρι αφού προηγήθηκε μετατροπή τους από έγχρωμες.

10.1 Συμπίεση Τεχνιτών Ολοκληρωτικών Εικόνων Δημιουργημένων με Προγράμματα Προσομοίωσης

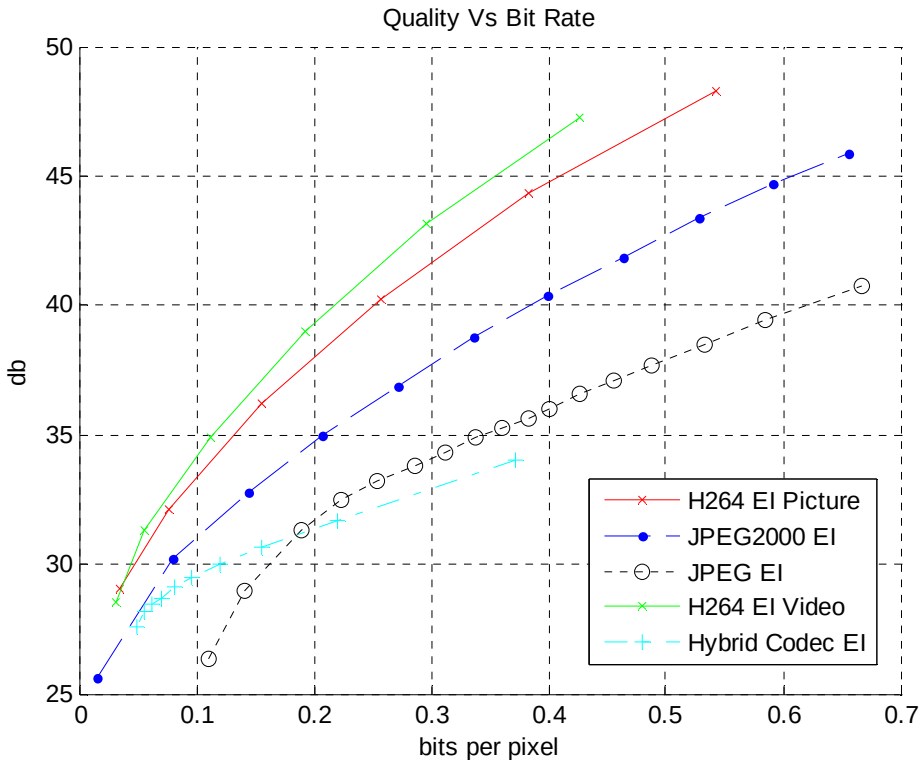
10.1.1 Συμπίεση της Εικόνας Ferrari

Η πρώτη εικόνα η οποία κωδικοποιήθηκε είναι μία ολοκληρωτική φωτογραφία ενός αγωνιστικού αυτοκινήτου της Ferrari. Η εικόνα συμπιέστηκε με τη χρήση των κωδικοποιητών JPEG EI, JPEG 2000 EI, H.264 ως εικόνα (H.264 EI Picture) και H.264 ως βίντεο-ροή η οποία προέκυψε από την τμηματοποίηση της εικόνας σε στοιχειώδεις εικόνες και η σάρωση τους με το σχήμα παράλληλων γραμμών (H.264 EI Video) όπως περιγράφηκε στην ενότητα 9.1. Τέλος χρησιμοποιήθηκε και ο υβριδικός κωδικοποιητής της ενότητας 9.2.



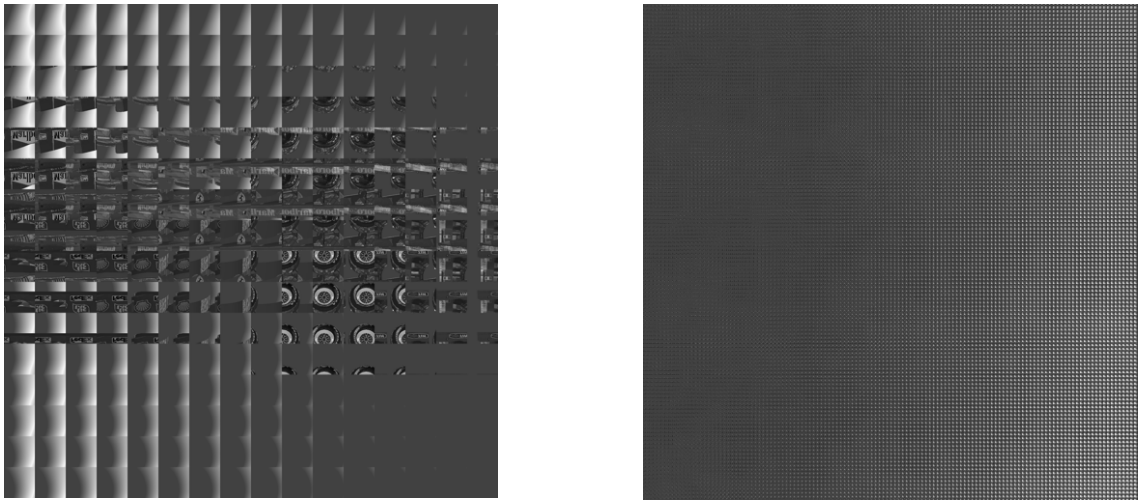
(a) (b)

Σχήμα 10.1 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα της Ferrari και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι



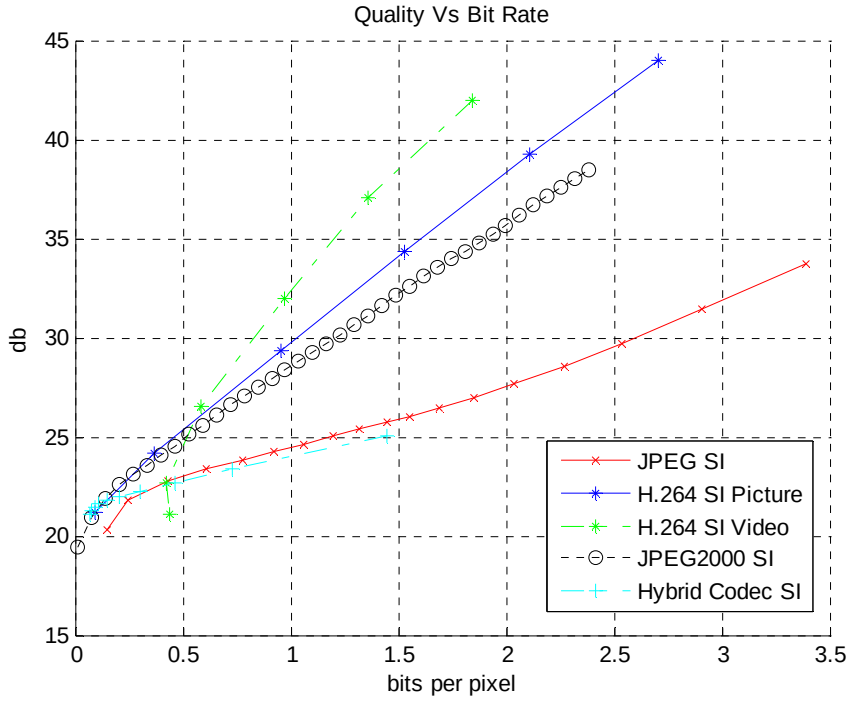
Σχήμα 10.2 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Ferrari η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

Η συμπίεση της εικόνας είναι εξαιρετική αφού ο κωδικοποιητής H.264 επιτυγχάνει με χαμηλούς ρυθμούς, πολύ υψηλή ποιότητα όπως παρατηρείται στο Σχήμα 10.2.

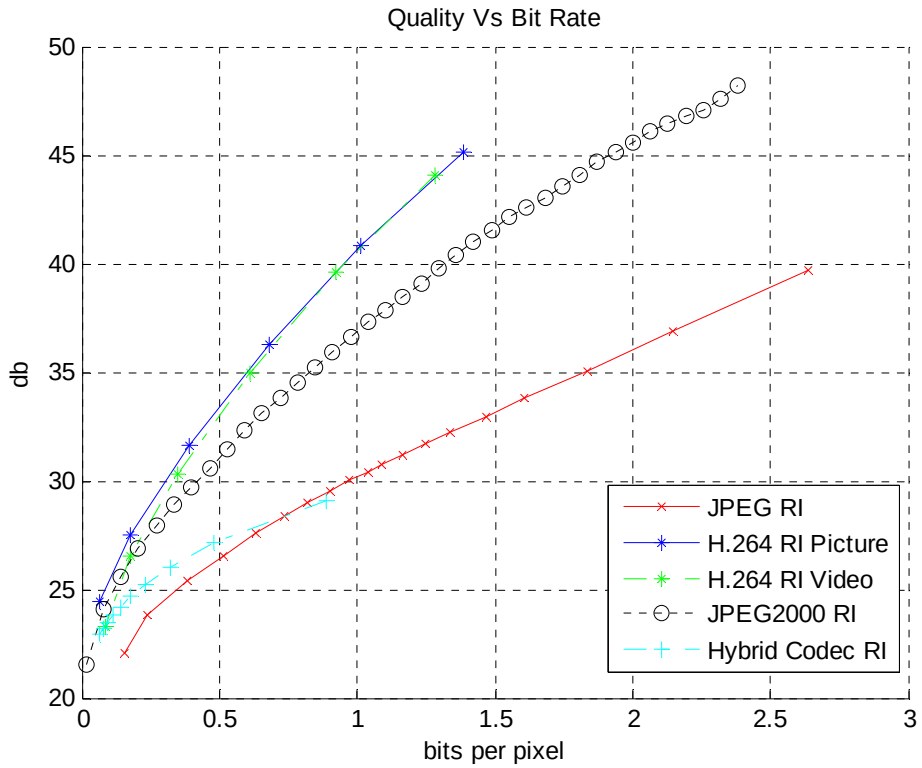


(a) (b)

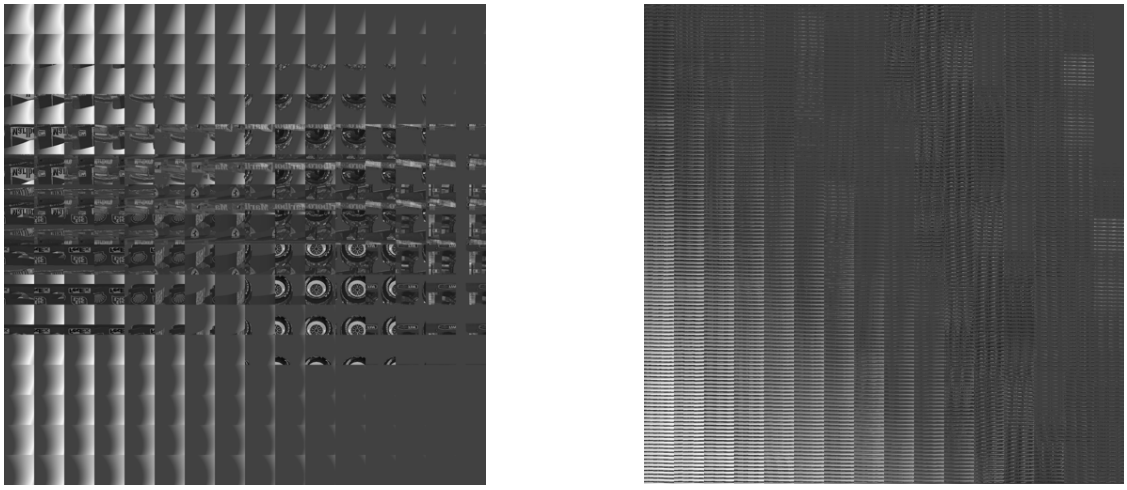
Σχήμα 10.3 : (a) Η Αρχική Ολοκληρωτική Εικόνα της Ferrari και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI)



Σχήμα 10.4 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα της Ferrari



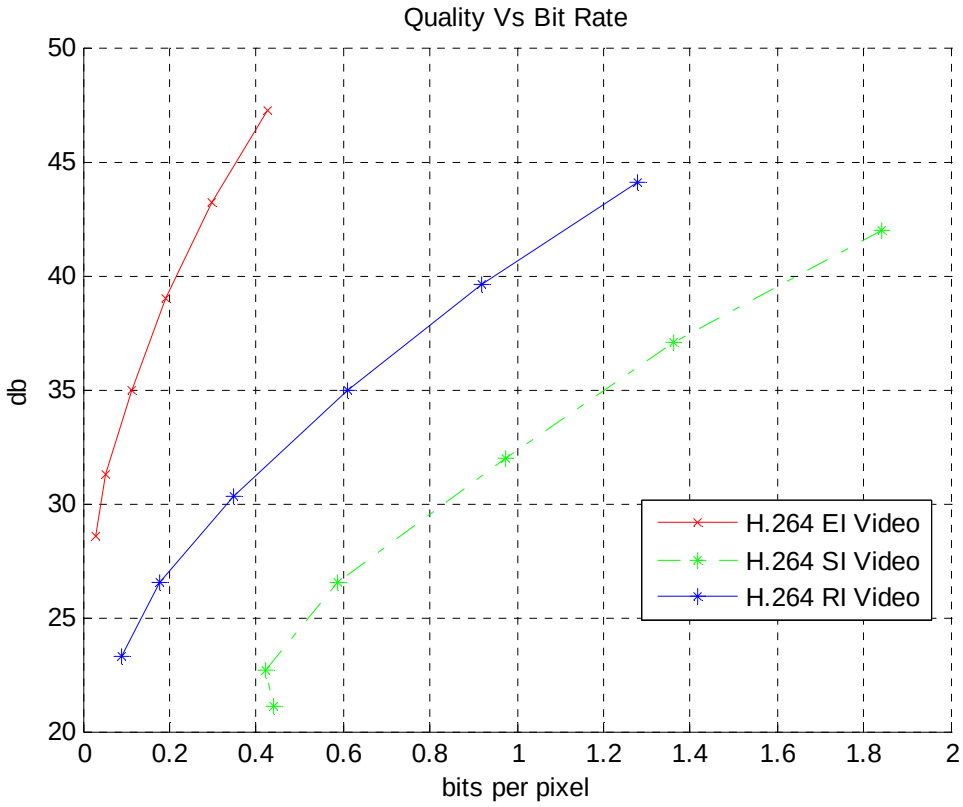
Σχήμα 10.5 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα της Ferrari



(a) (b)

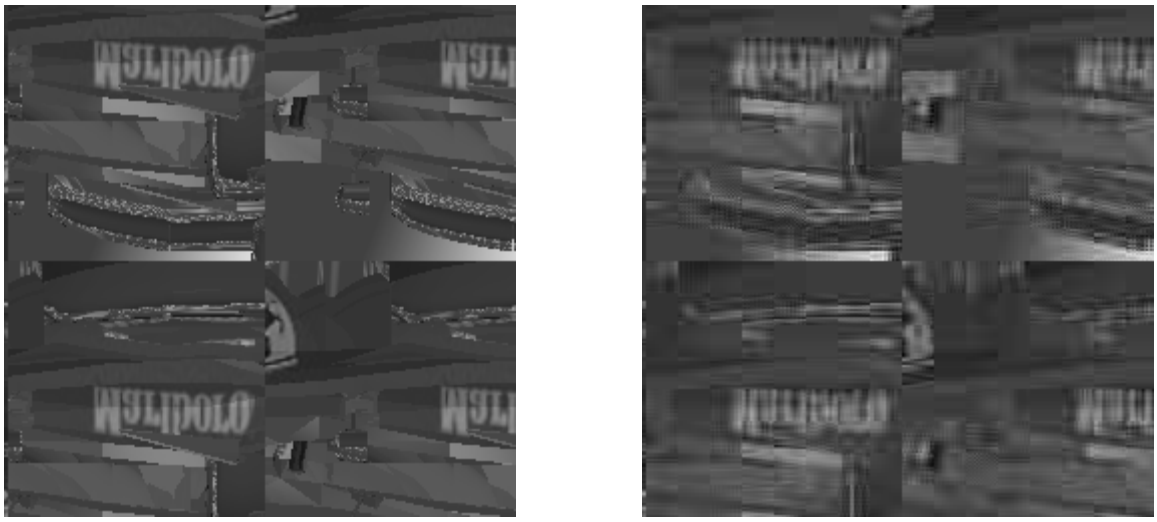
Σχήμα 10.6 : (a) Η Αρχική Ολοκληρωτική Εικόνα της Ferrari και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI)

Από το πρώτο παράδειγμα με τη χρήση των 5 κωδικοποιητών με των τριών διαφορετικών μεθόδων προ-επεξεργασίας (EI, SI και RI) είναι προφανές, παρατηρώντας τις γραφικές παραστάσεις, πως ο κωδικοποιητής συμπίεσης βίντεο H.264 υπερέχει έναντι των υπολοίπων. Αυτό συμβαίνει όχι μόνο στην περίπτωση χρήσης του σχήματος EI αλλά και των SI, RI (Σχήμα 10.7)



Σχήμα 10.7 : Οι καλύτερες Επιδόσεις και στα Τρία Σχήματα Προ-Επεξεργασίας Επιτυγχάνονται από τον H.264 όταν Κωδικοποιηθεί η Ολοκληρωτική Εικόνα ως Βίντεο-Ροή. Το Σχήμα Προ-Επεξεργασίας EI Αποδίδει Καλύτερα και με Σημαντική Διαφορά Έναντι των Υπολοίπων

Αξιοσημείωτο είναι το γεγονός ότι ο υβριδικός κωδικοποιητής αποδίδει στο σχήμα προ-επξεργασίας EI καλύτερα από το JPEG μόνο στους χαμηλούς ρυθμούς δεδομένων $(0.1 - 0.15\ bpp)$ ενώ σε υψηλότερους ρυθμούς αποδίδει λιγότερο από αυτό και προφανώς λιγότερο από κάθε άλλο κωδικοποιητή. Ο συνδυασμός αυτών των δύο παρατηρήσεων καθιστά των υβριδικό κωδικοποιητή εν μέρει ακατάλληλο για τη συμπίεση ολοκληρωτικών εικόνων αφού σε ρυθμούς δεδομένων της τάξης των $0.15bpp$ η ποιότητα είναι αρκετά χαμηλή ώστε να μπορέσει να ανασυσταθεί σωστά η τρισδιάστατη εικόνα όπως παρατηρούμε στο Σχήμα 10.8.



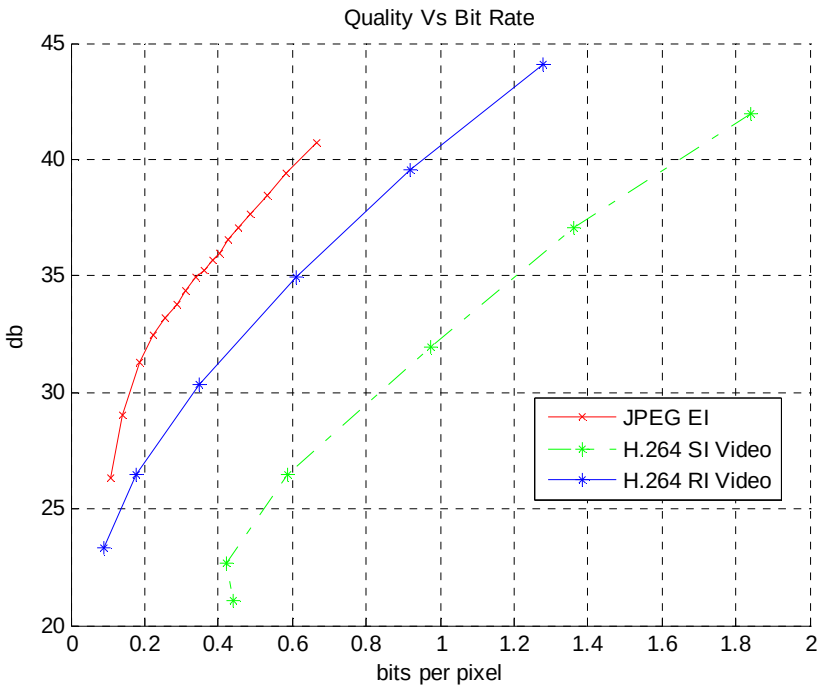
(a) (b)

Σχήμα 10.8 : 2×2 Στοιχειώδεις Εικόνες της (a) Αρχικής Ολοκληρωτικής Εικόνας και (b) της Ανασυσταμένης Εικόνας μετά την Κωδικοποίηση με τον Υβριδικό Κωδικοποιητή με Ρυθμό $0.15bpp$ στα 30.65db

Μία εξίσου σημαντική παρατήρηση σχετίζεται την απόδοση των σχημάτων προ-επεξεργασίας SI και RI. Όπως παρουσιάζεται στο Σχήμα 10.9, ακόμα και η χειρότερη επίδοση η οποία επιτυγχάνεται στο σχήμα EI (JPEG ή Υβριδικός κωδικοποιητής) είναι καλύτερη ακόμα και από την καλύτερη επίδοση των σχημάτων SI και RI την οποία επιτυγχάνει ο κωδικοποιητής H.264.

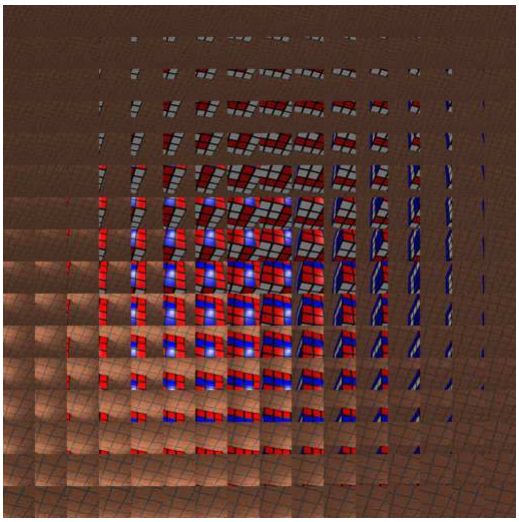
Η διαπίστωση αυτή μας οδηγεί στο συμπέρασμα πως τα σχήματα προ-επεξεργασίας SI και RI ουσιαστικά δεν καταφέρνουν να επιτύχουν το σκοπό για τον οποίο χρησιμοποιήθηκαν, δηλαδή να ομαδοποιήσουν τις όμοιες τιμές φωτεινότητας των εικονοστοιχείων ώστε να μεγιστοποιηθεί η συσχέτιση των γειτονικών εικονοστοιχείων, μετατρέποντας στην πράξη τις αρχικές εικόνες σε μία εικόνα χωρίς σαφείς δομές. Χαρακτηριστικό είναι το παράδειγμα στο Σχήμα 10.5 όπου η καμπύλη της μετασχηματισμένης RI εικόνας με τον κωδικοποιητή H.264 ως εικόνα συμπίπτει με την καμπύλη της κωδικοποιημένης εικόνας ως βίντεο-ροή (H.264 RI Video) γεγονός που φανερώνει την αδυναμία του κωδικοποιητή να εντοπίσει συσχετισμένα καρέ, δηλαδή συσχετισμένα μπλοκ που προέκυψαν μετά το σχήμα προ-επεξεργασίας RI.

Σε κάθε περίπτωση αυτό αποτελεί ένα πρώτο συμπέρασμα και όχι κάποιο πόρισμα. Ασφαλή συμπεράσματα θα εξαχθούν στο τέλος της εργασίας αφού πρώτα διερευνηθεί η συμπεριφορά των σχημάτων SI και RI και στις υπόλοιπες εικόνες.

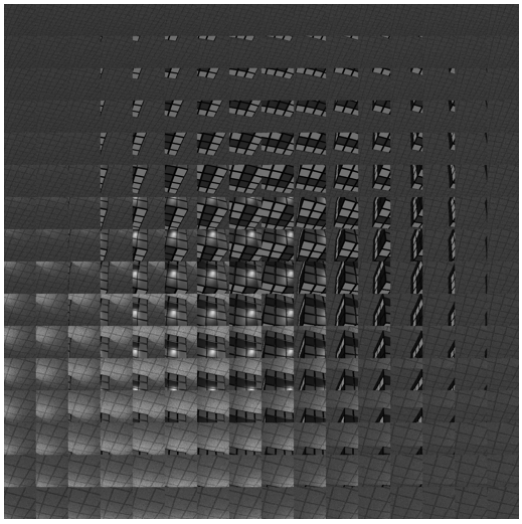


Σχήμα 10.9 : Η Χειρότερη Επίδοση στο σχήμα EI (Jpeg) Αποδίδει Καλύτερα από την Καλύτερη Επίδοση των Σχημάτων Προ-Επεξεργασίας SI και RI

10.1.2 Συμπίεση της Εικόνας Cube

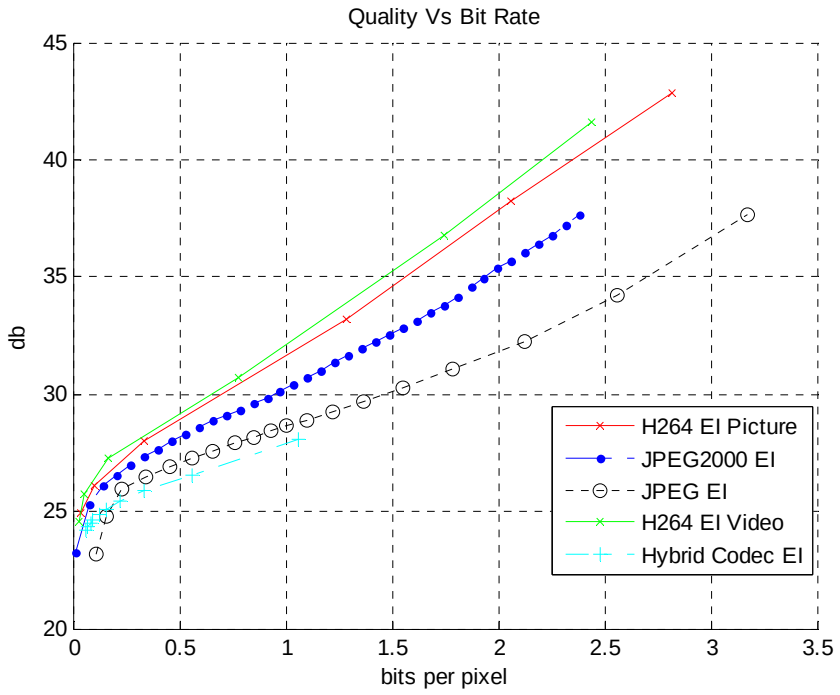


(a)



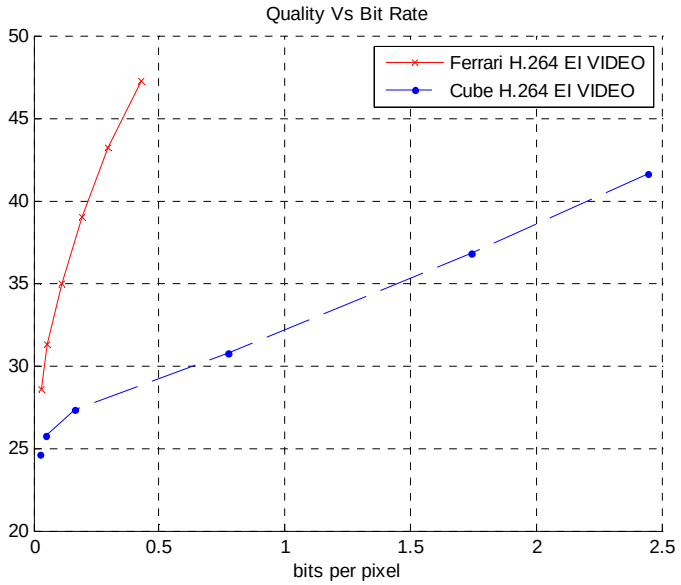
(b)

Σχήμα 10.10 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα Cube και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι

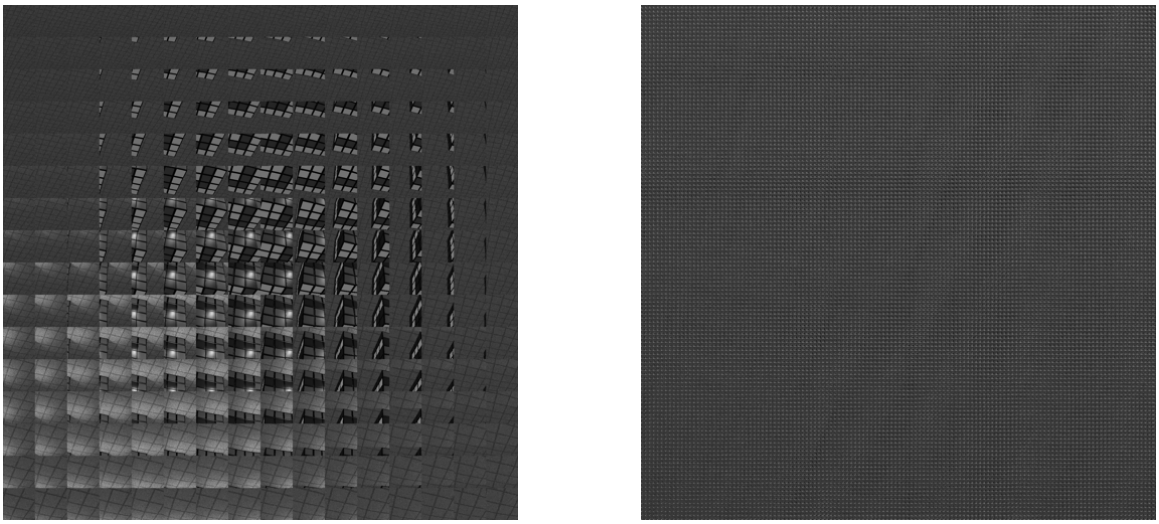


Σχήμα 10.11 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Cube η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

Η σκηνή με τον κύβο όπως παρατηρούμε δε συμπιέζεται το ίδιο αποδοτικά με τη σκηνή της Ferrari. Χαρακτηριστική είναι η καμπύλη του H.264 EI Video ο οποίος επιτυγχάνει την καλύτερη απόδοση. Για τη συμπίεση της εικόνας με ποιότητα κοντά στα $35db$ απαιτείται 1.5 bit σε αντίθεση με τη σκηνή της Ferrari όπου για την ίδια ποιότητα χρειάστηκε μόλις 0.1 bit από τον ίδιο κωδικοποιητή.



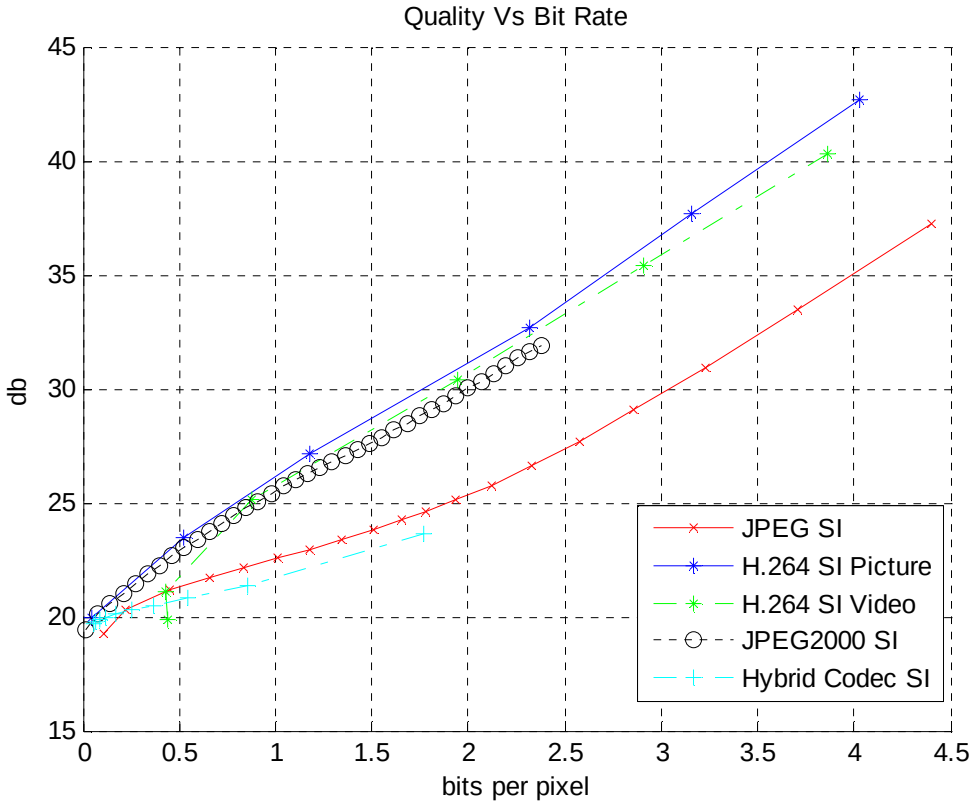
Σχήμα 10.12 : Αποδοτικότητα Συμπίεσης στην Εικόνα με τη Σκηνή της Ferrari και του Κύβου



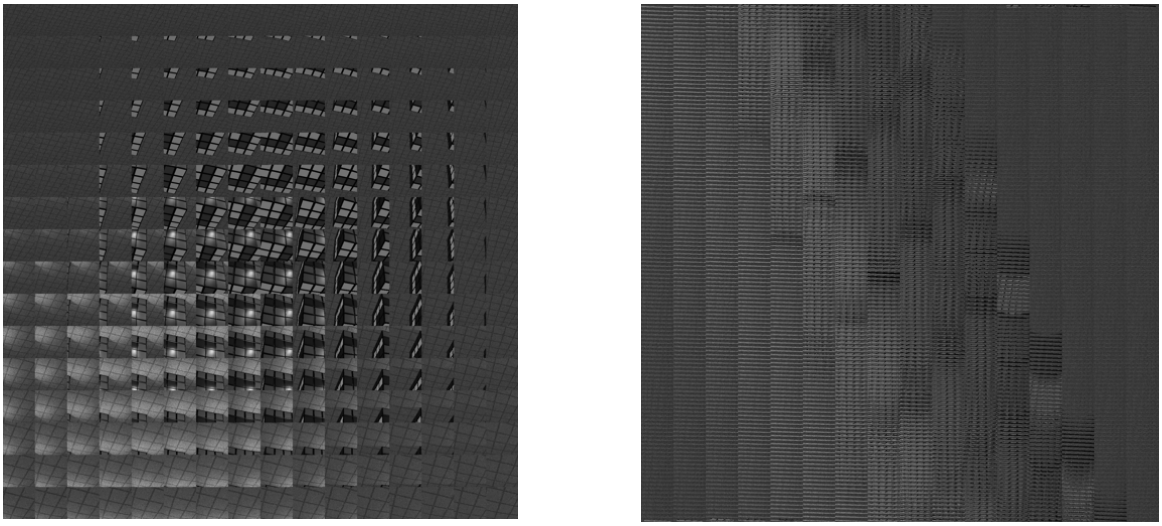
(a) (b)

Σχήμα 10.13 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cube και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI)

Στο Σχήμα 10.14 και Σχήμα 10.16 που ακολουθούν παρουσιάζεται η απόδοση των κωδικοποιητών όταν εφαρμοστούν στις προ-επεξεργασμένες εικόνες με τα σχήματα SI και RI.

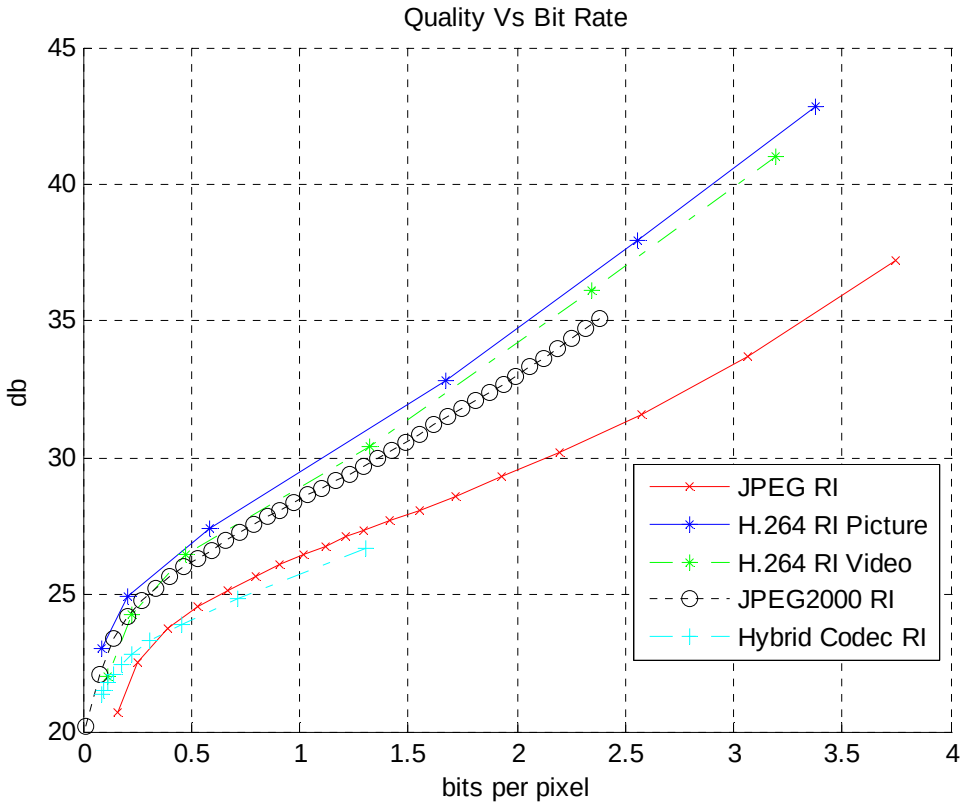


Σχήμα 10.14 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα



(a) (b)

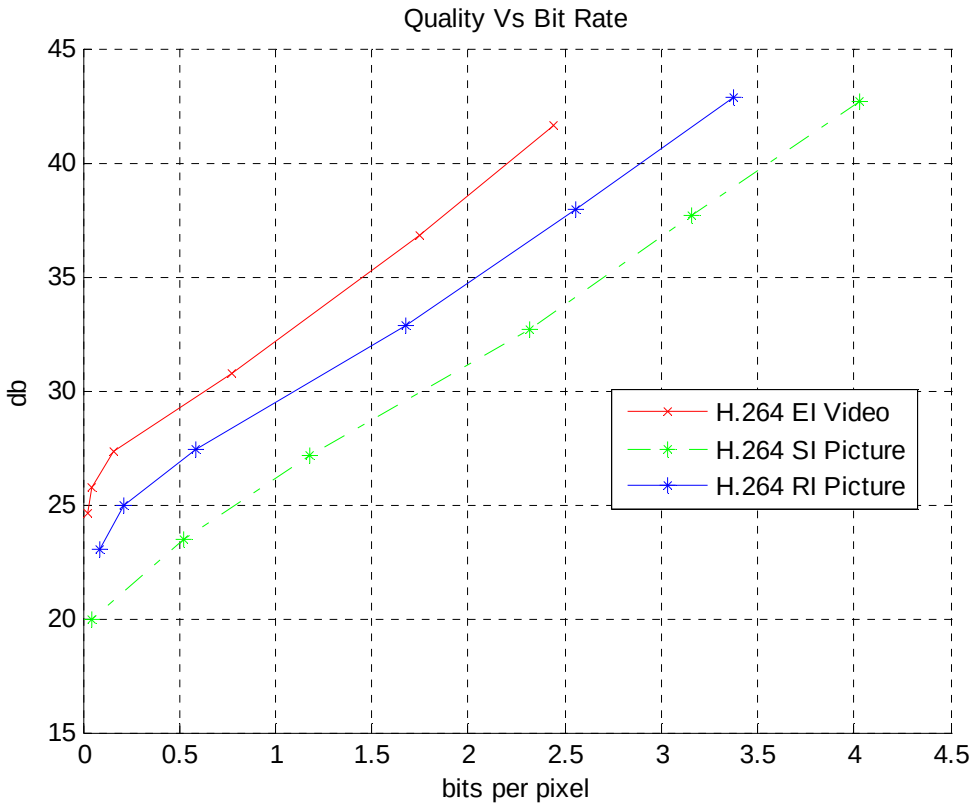
Σχήμα 10.15 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cube και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI)



Σχήμα 10.16 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα

Παρατηρώντας τις καμπύλες που προέκυψαν από την κωδικοποίηση της εικόνας του κύβου με όλες τις μεθόδους και τα σχήματα συμπίεσης, συμπεραίνουμε πως οι βασικές παρατηρήσεις που σημειώθηκαν στην προηγούμενη υποενότητα κατά την κωδικοποίηση της σκηνής με τη Ferrari, ισχύουν και εδώ. Συγκεκριμένα, ο κωδικοποιητής H.264 αποδίδει καλύτερα σε οποιαδήποτε μέθοδο προ-επεξεργασίας ενώ το σχήμα EI κωδικοποιεί αποδοτικότερα την εικόνα όπως παρουσιάζεται στο Σχήμα 10.17. Ο Υβριδικός κωδικοποιητής όπως και στην προηγούμενη περίπτωση αποδίδει στο σχήμα EI λίγο καλύτερα από το JPEG στους χαμηλούς ρυθμούς, ενώ από τα τρία σχήματα προ-επεξεργασίας, πιο αδύνατο παρουσιάζεται το σχήμα SI.

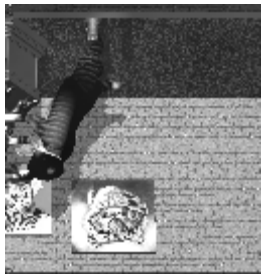
Χαρακτηριστικό της εικόνας αποτελούν οι παραπλήσιες καμπύλες του H.264 όταν η ολοκληρωτική εικόνα κωδικοποιεί τόσο ως βίντεο-ροή όσο και ως ένα καρέ, γεγονός που φανερώνει την αδυναμία εύρεσης αποδοτικής συσχέτισης μεταξύ των διαδοχικών καρέ. Ενδιαφέρον παρουσιάζει και η απόδοση του Jpeg 2000 στο σχήμα προ-επεξεργασίας SI όπου ο κωδικοποιητής επιτυγχάνει αντίστοιχη απόδοση με τον H264.



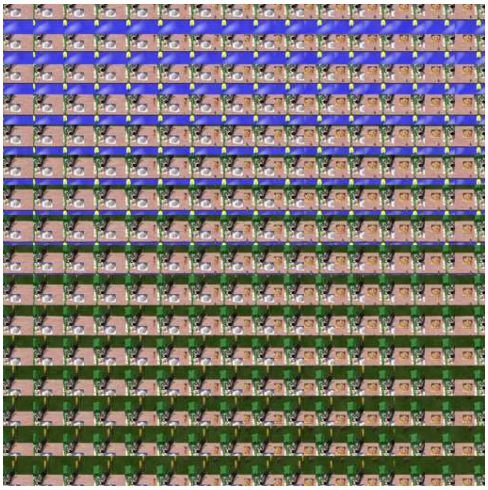
Σχήμα 10.17 : Σύγκριση των Τριών Καλύτερων Μεθόδων Συμπίεσης (H.264) σε Κάθε Σχήμα Προ-Επεξεργασίας

10.1.3 Συμπίεση της Εικόνας Cat

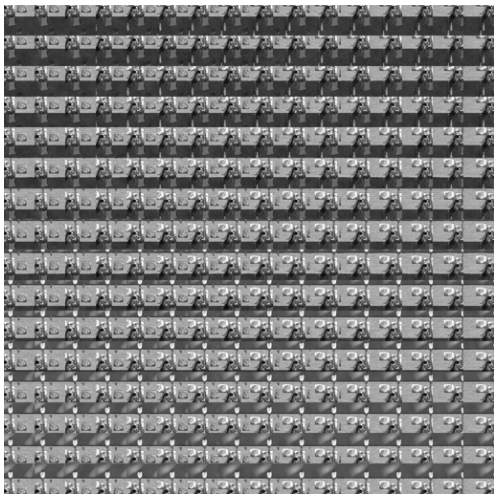
Η μετρήσεις με τις ολοκληρωτικές εικόνες που δημιουργήθηκαν με προγράμματα προσομοίωσης, ολοκληρώνονται με την ακόλουθη σκηνή η οποία εμφανίζει μία γάτα (Σχήμα 10.18).



Σχήμα 10.18 : Στοιχειώδης Εικόνα Ολοκληρωτικής Φωτογραφίας που Απεικονίζει μία Γάτα



(a)

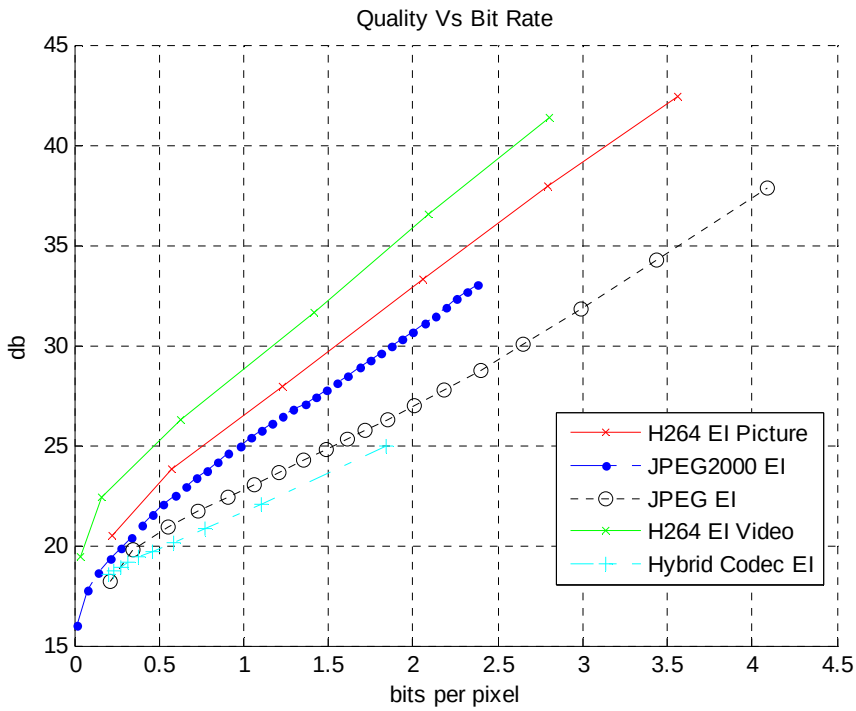


(b)

Σχήμα 10.19 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Εικόνα Cat και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι

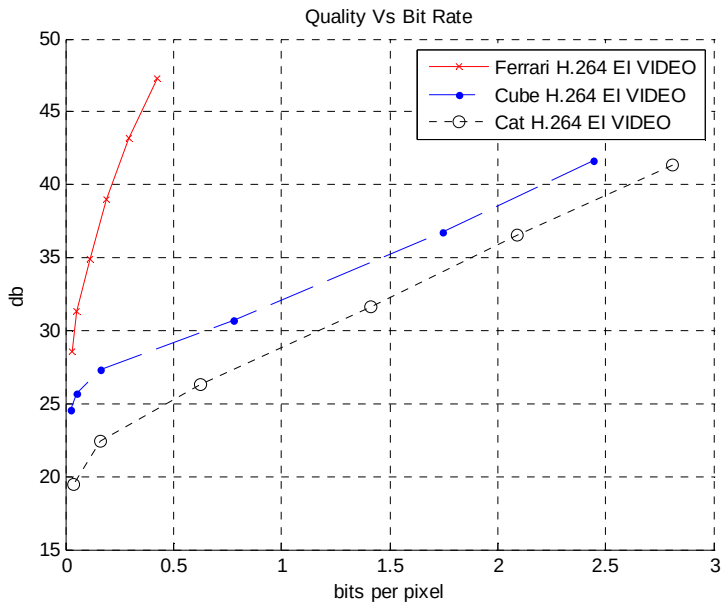
Μετά την ολοκλήρωση των μετρήσεων με τη συγκεκριμένη εικόνα θα είμαστε σε θέση να εξάγουμε γενικότερα συμπεράσματα σε σχέση με την αξιολόγηση των κωδικοποιητών αλλά και των σχημάτων προ-επεξεργασίας.

Στο Σχήμα 10.20 παρουσιάζεται η απόδοση των κωδικοποιητών με το σχήμα EI.



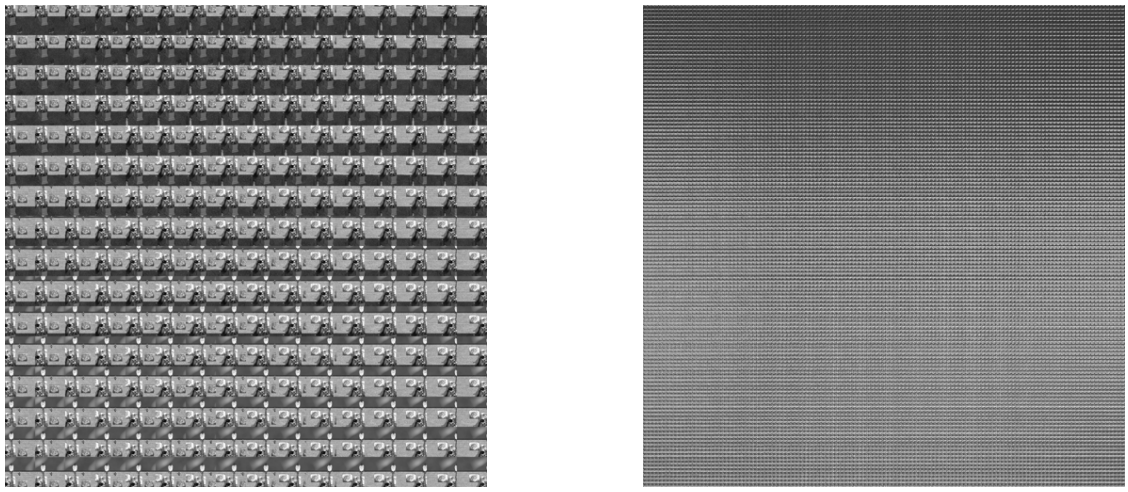
Σχήμα 10.20 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Cat η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

Ο κωδικοποιητής H.264 Video αποδίδει καλύτερα από κάθε άλλον και σε αυτή τη σκηνή. Σημειώνεται ότι η εικόνα με τη γάτα κωδικοποιείται λιγότερο αποδοτικά τόσο σε σχέση με τη σκηνή της Ferrari όσο και με τη σκηνή του κύβου (Σχήμα 10.21).



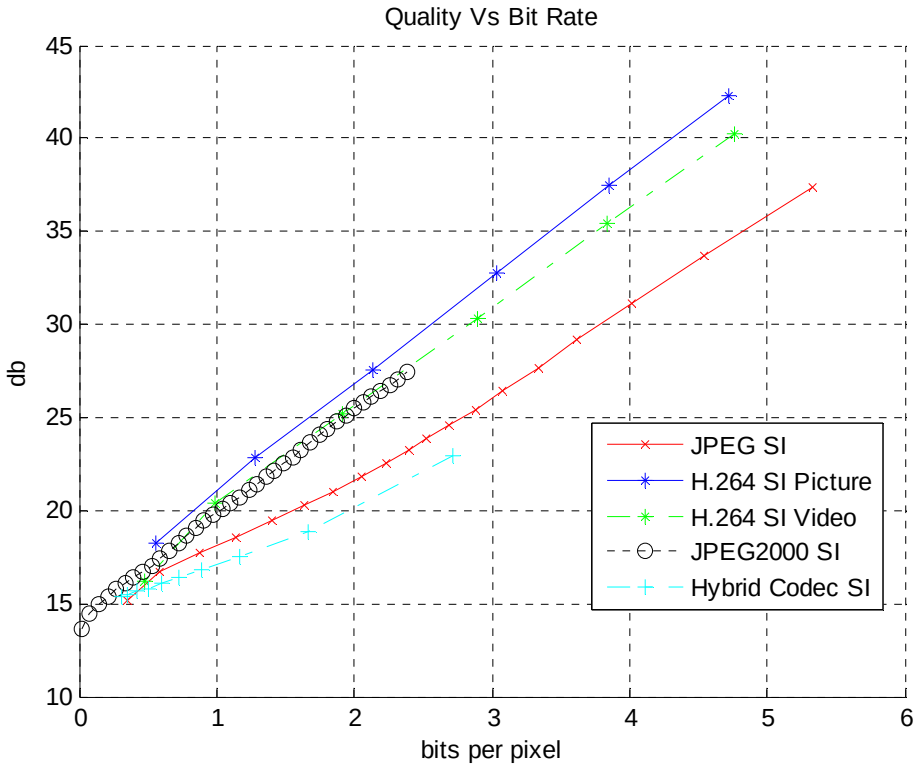
Σχήμα 10.21 : Αποδοτικότητα Συμπίεσης της Εικόνας της Ferrari του Κύβου και της Γάτας

Στη συνέχεια ακολουθούν μετρήσεις με τα σχήματα προ-επεξεργασίας SI και RI.

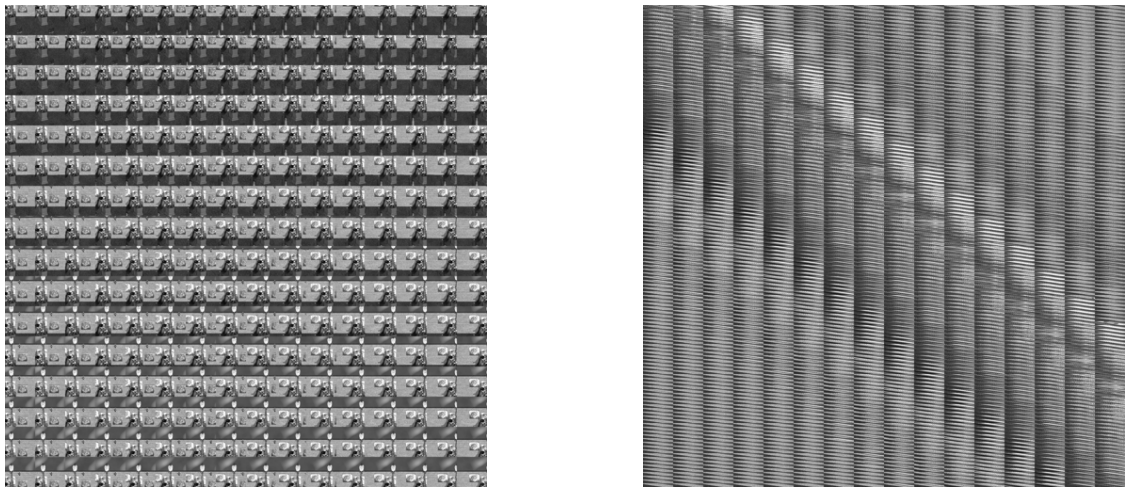


(a) (b)

Σχήμα 10.22 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cat και (b) ο Μετασχηματισμός της σε ΥποΕικόνες (SI)

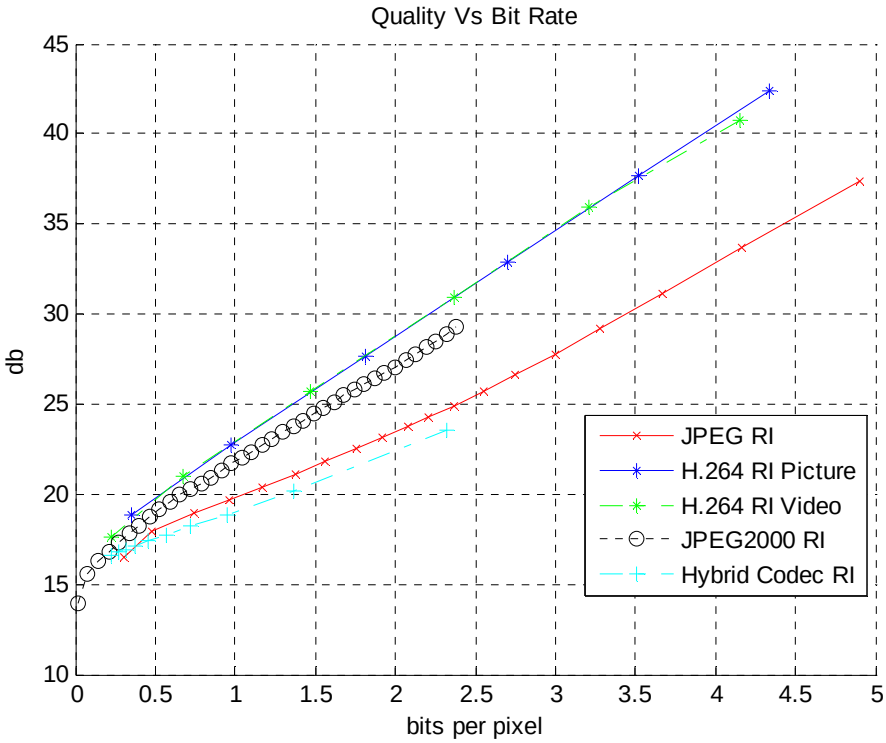


Σχήμα 10.23 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη SI Εικόνα Cat



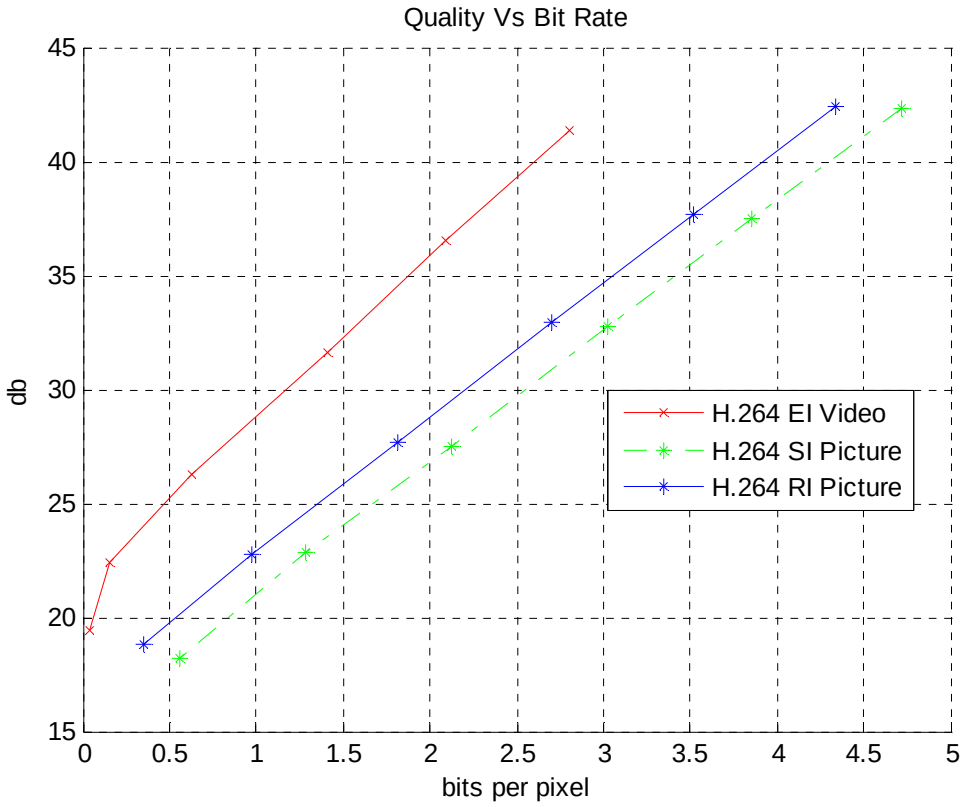
(a) (b)

Σχήμα 10.24 : (a) Η Αρχική Ολοκληρωτική Εικόνα Cat και (b) ο Μετασχηματισμός της στο Πεδίο των Ακτινών (RI)



Σχήμα 10.25 : Η Απόδοση των Πέντε Κωδικοποιητών στη Μετασχηματισμένη RI Εικόνα

Στο Σχήμα 10.26 παρουσιάζονται οι καλύτερες επιδόσεις των κωδικοποιητών για κάθε σχήμα προ-επεξεργασίας



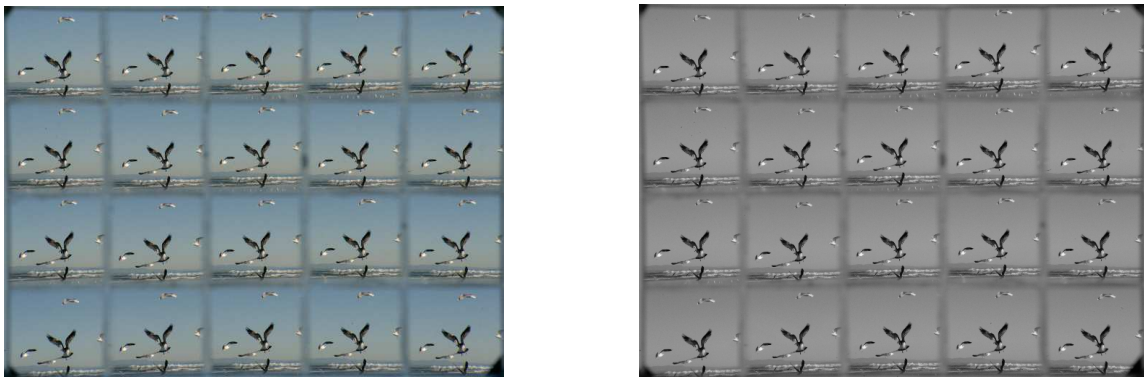
Σχήμα 10.26 : Οι καλύτερες Επιδόσεις για Κάθε Σχήμα Προ-Επεξεργασίας

Οι παρατηρήσεις και σε αυτή τη σκηνή συμπίπτουν απόλυτα με τις δύο προηγηθείσες σκηνές. Τα συμπεράσματα, συγκεντρωτικά θα παρατεθούν στο τελευταίο κεφάλαιο της παρούσας εργασίας.

10.2 Συμπίεση Πραγματικών Ολοκληρωτικών Εικόνων

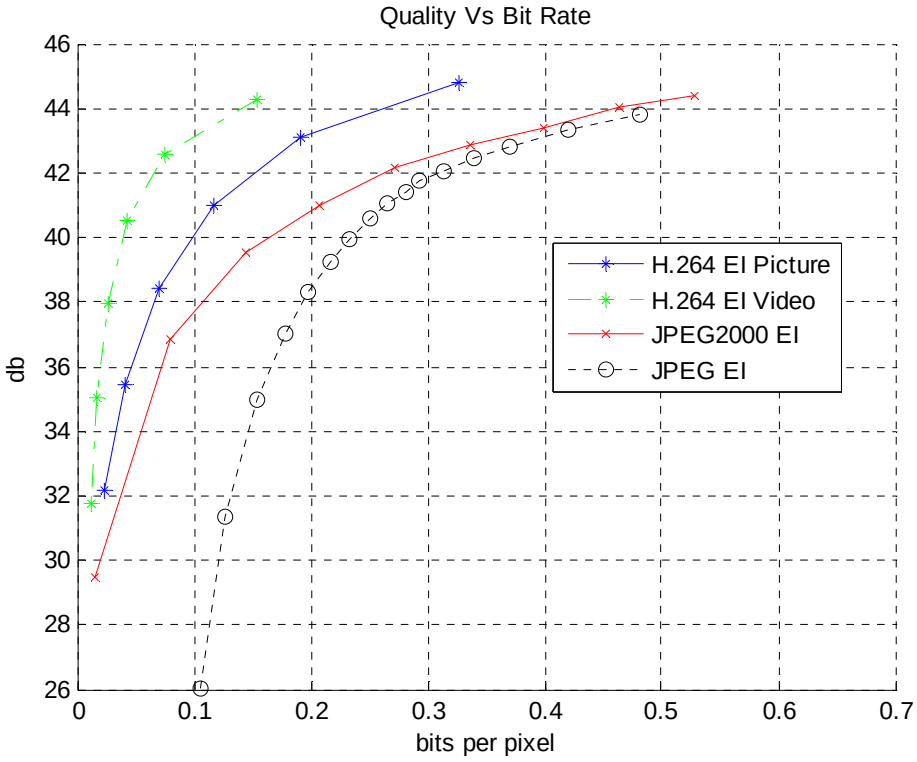
Στην παρούσα υποενότητα θα παρουσιαστούν οι μετρήσεις σχετικά με την απόδοση των κωδικοποιητών που εξετάστηκαν μέχρι στιγμής αλλά σε πραγματικές ολοκληρωτικές φωτογραφίες. Οι φωτογραφίες αυτές ελήφθησαν με τη χρήση μίας τροποποιημένης φωτογραφικής μηχανής η οποία δημιουργήθηκε από ανθρώπους της Adobe, του πανεπιστημίου της Ουάσινγκτον και του πανεπιστημίου Κολούμπια στα πλαίσια της έρευνας τους[63]. Δεδομένου ότι τα σχήματα SI και RI δεν απέφεραν τα προσδοκώμενα αποτελέσματα όπως επίσης και η μέθοδος του υβριδικού κωδικοποιητή, σε αυτή την υποενότητα θα συμπιεστούν οι εικόνες μόνο με το σχήμα EI χρησιμοποιώντας τους τρείς δημοφιλέστερους κωδικοποιητές (H.264 Video / Picture, Jpeg 2000 και Jpeg).

10.2.1 Συμπίεση της Εικόνας Birds



(a) (b)

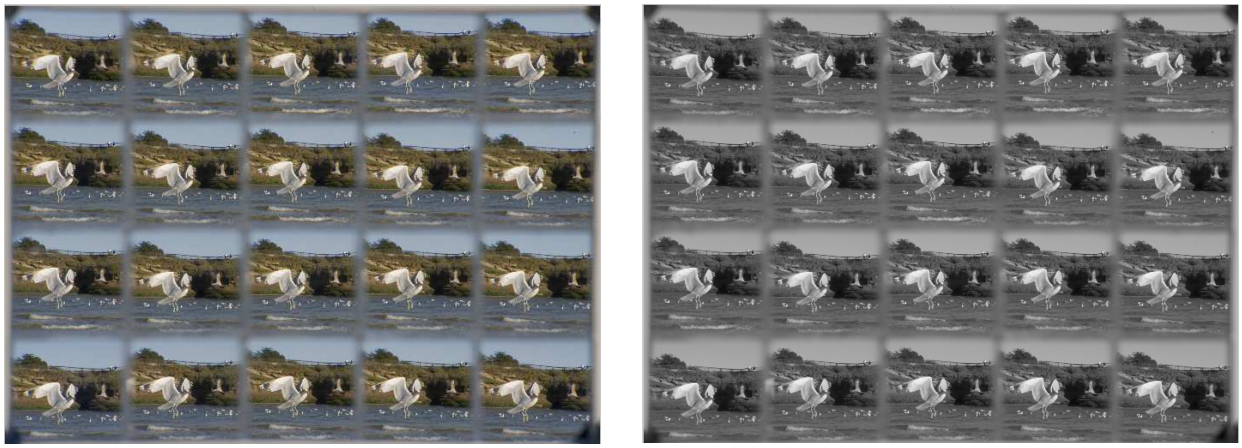
Εικόνα 10.1 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Birds και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι



Σχήμα 10.27 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Birds η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

Τα αποτελέσματα τα οποία παρουσιάζονται στο Σχήμα 10.27 δεν αποτελούν έκπληξη καθώς επιβεβαιώνουν τα πρώτα συμπεράσματα τα οποία άρχισαν να διαφαίνονται από την προηγούμενη υποενότητα. Για μία ακόμα φορά ο κωδικοποιητής H.264 επιτυγχάνει εξαιρετικά αποτελέσματα τόσο κωδικοποιώντας την εικόνα ως ένα μεμονωμένο καρέ όσο και ως βίντεο-ροή αποτελούμενη από γειτονικές στοιχειώδεις εικόνες. Χαρακτηριστική είναι η διαφορά απόδοσης ανάμεσα στον κωδικοποιητή Jpeg και τον H.264 αφού ο πρώτος με $0.1\ bit$ ανά εικονοστοιχείο επιτυγχάνει ποιότητα της τάξης των 26db ενώ για τον ίδιο ρυθμό ο H.264 επιτυγχάνει PSNR μεγαλύτερο από 40db.

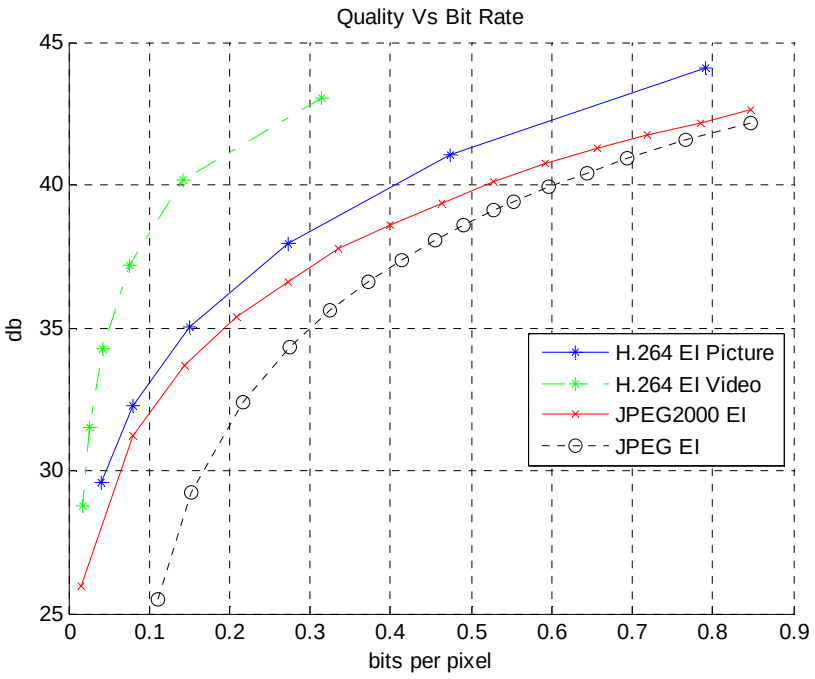
10.2.2 Συμπίεση της Εικόνας Seagull



(a) (b)

Εικόνα 10.2 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Seagull και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι

Σχετικά με την Εικόνα 10.2 ισχύουν οι ίδιες ακριβώς παρατηρήσεις που ίσχυαν και στις προηγούμενες εικόνες. Μία ενδιαφέρουσα παρατήρηση σε σχέση με την απόδοση των κωδικοποιητών στις δύο φωτογραφίες είναι ότι παρά το γεγονός ότι η λήψη και των δύο φωτογραφιών πραγματοποιήθηκε με τον ίδια τροποποιημένη μηχανή, στην οποία έχει τοποθετηθεί η ίδια συστοιχία μικρο-φακών, ο κωδικοποιητής H.264 ως βίντεο-ροή είναι πολύ αποδοτικότερες στην Εικόνα 10.2 σε σχέση με την Εικόνα 10.1. Η εξήγηση σε αυτό μπορεί να δοθεί παρατηρώντας την απόδοση του H.264 όταν συμπιέζει την εικόνα ως ένα μεμονωμένο καρέ. Σύμφωνα με το Σχήμα 10.27, η απόδοση του κωδικοποιητή H.264 Picture κυμαίνεται εξ αρχής σε πολύ υψηλά επίπεδα, συνεπώς δεν αφήνει πολλά περιθώρια στη διαδικασία εκτίμησης κίνησης του H.264 Video. Αντίθετα η Εικόνα 10.2 δε συμπιέζεται εξ αρχής το ίδιο αποδοτικά, γεγονός που οφείλεται κατά βάση στις υψηλότερες συχνότητες που εμφανίζονται σε αυτή.

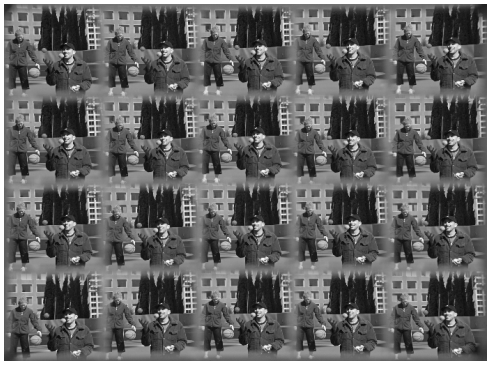


Σχήμα 10.28 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Seagull η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

10.2.3 Συμπίεση της Εικόνας Sports

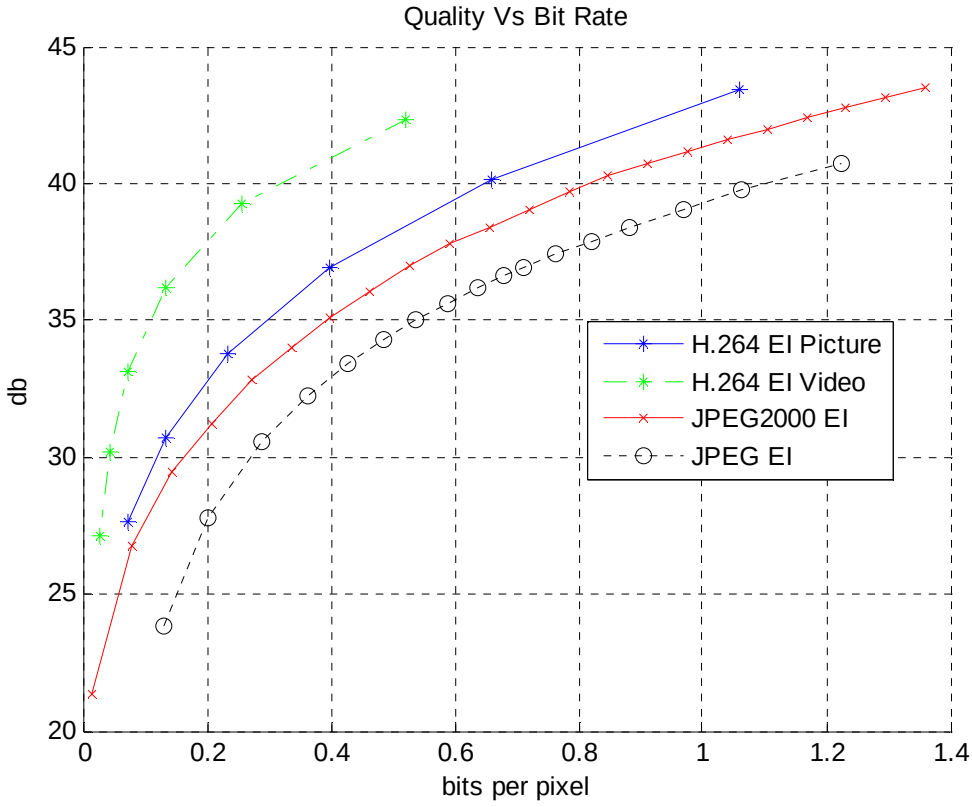
Η τελευταία ολοκληρωτική εικόνα η οποία θα συμπιεστεί με του κωδικοποιητές, H.264, JPEG 2000 και JPEG είναι η Εικόνα 10.3.

(a)



(b)

Εικόνα 10.3 : (a) Η Αρχική Έγχρωμη Ολοκληρωτική Φωτογραφία Sports και (b) η Μετατροπή της σε Εικόνα 256 Αποχρώσεων του Γκρι



Σχήμα 10.29 : Η Απόδοση των Πέντε Κωδικοποιητών στην Αρχική Εικόνα Sport η οποία Αποτελείται από 256 Στοιχειώδεις Εικόνες

Εκτός από το γεγονός ότι η εικόνα είναι πιο σύνθετη από τις δύο προηγούμενες με αποτέλεσμα να κωδικοποιείται λιγότερο αποδοτικά σε σχέση με τις εικόνες Birds και Seagull, δεν παρατηρείται καμία άλλη διαφορά σε ότι αφορά την ταξινόμηση των κωδικοποιητών βάσει της απόδοσης τους.

Στην επόμενη και τελευταία ενότητα της παρούσας εργασίας θα γίνει μία προσπάθεια ώστε να διεξαχθούν κάποια συμπεράσματα τόσο σε σχέση με τους κωδικοποιητές όσο και με τις εικόνες. Πιο συγκεκριμένα, θα μελετηθούν τα χαρακτηριστικά εκείνα τα οποία βοηθούν τους κωδικοποιητές να συμπιέσουν αποδοτικότερα κάποιες εικόνες σε σχέση με άλλες αλλά και ένα πλήθος στατιστικών μεγεθών τα οποία ενδέχεται να μπορούν να χρησιμοποιηθούν προκειμένου να προσδιοριστούν ποσοτικά κάποια χαρακτηριστικά των εικόνων. Με τον τρόπο αυτό, πιθανώς να καταστεί δυνατή η βελτιστοποίηση των κωδικοποιητών όταν κωδικοποιούν ολοκληρωτικές εικόνες ενώ παράλληλα να υπάρξει ένα φράγμα σε σχέση με τις επιδόσεις που μπορούν να επιτευχθούν βάσει των στατιστικών μεγεθών της κάθε εικόνας.

11. Συμπεράσματα

Στην προηγούμενη ενότητα, αναδείχθηκε μετά από ένα πλήθος μετρήσεων ότι ο κωδικοποιητής H.264 επιτυγχάνει τις καλύτερες επιδόσεις στη συμπίεση ολοκληρωτικών εικόνων, ιδίως όταν η εικόνα διαιρείται στις επιμέρους στοιχειώδεις εικόνες και κάθε στοιχειώδης εικόνα εγγράφεται ως ένα καρέ μίας βίντεο-ροής. Στην περίπτωση αυτή ο κωδικοποιητής δεν αφαιρεί μόνο τη συσχέτιση εντός της κάθε στοιχειώδους εικόνας αλλά και των στοιχειωδών εικόνων μεταξύ τους. Αυτό έχει ως αποτέλεσμα τη βέλτιστη συμπίεση της ολοκληρωτικής εικόνας, γεγονός που επιβεβαιώθηκε από όλες τις μετρήσεις. Στη μοναδική περίπτωση όπου ο κωδικοποιητής H.264 Video δεν κατάφερε να επιτύχει τις βέλτιστες επιδόσεις αλλά αντιθέτως υστερούσε έναντι του H.264 Picture ήταν στην περίπτωση των σχημάτων προ-επεξεργασίας SI και RI όπου τα σχήματα αυτά μετέτρεψαν την εικόνα σε ένα σήμα χωρίς δομές και συσχέτιση.

Αξίζει να σημειωθεί ότι η σειρά κατάταξης των κωδικοποιητών ως προς τις επιδόσεις τους είναι αντιστρόφως ανάλογη του χρόνου κωδικοποίησης. Ο JPEG ο οποίος υστερεί έναντι των υπολοίπων κωδικοποιητών, είναι ταχύτατος και απλούστατος στην υλοποίηση. Αντίθετα, ο κωδικοποιητής H.264 Video για να συμπιέσει μία ολοκληρωτική εικόνα μπορεί να χρειαστεί ακόμα και ώρες προκειμένου να τις κωδικοποιήσει. Ο H.264 Picture αποτελεί ίσως τη χρυσή τομή ανάμεσα στην απόδοση και το χρόνο συμπίεσης. Σε εφαρμογές όπου η ταχύτητα είναι η σημαντικότερη παράμετρος, ο κωδικοποιητής JPEG 2000 μπορεί να χρησιμοποιηθεί και να επιτύχει επιδόσεις μεταξύ του JPEG και του H.264 Picture οι οποίες τείνουν πιο κοντά στον H.264. Αντίθετα, όταν τον πρώτο λόγο έχει η απόδοση, τότε ο κωδικοποιητής H.264 Video, συνήθως επιτυγχάνει τα καλύτερα αποτελέσματα.

Ένα ενδιαφέρον ερώτημα το οποίο γεννάται, σχετίζεται με την επιλογή του κατάλληλου κωδικοποιητή ανάλογα με την εικόνα εισόδου. Χαρακτηριστικό παράδειγμα αποτελεί η εικόνα με τον κύβο (10.1.2). Παρατηρώντας το Σχήμα 10.10 διαπιστώνουμε πως η διαφορά ανάμεσα στο αποτελέσματα του H.264 Video και H.264 Picture είναι πολύ μικρή, η διαφορά τους όμως σε ότι αφορά το χρόνο κωδικοποιήσεις είναι εξαιρετικά μεγάλη αφού στην πρώτη περίπτωση η κωδικοποίηση ολοκληρώνεται σε λεπτά ή ώρες, ενώ στη δεύτερα σε μόλις μερικά δευτερόλεπτα. Προφανώς σε μία αντίστοιχη περίπτωση δεν υπάρχει κανένα όφελος από τη χρήση του H.264 Video και τη δαπάνη ενός μεγάλου χρονικού διαστήματος για αποτελέσματα τα οποία μπορούν να επιτευχθούν σε πολύ λιγότερο χρόνο.

Στη συνέχεια της εργασίας θα παρασταθούν τόσο γραφικά όσο και αριθμητικά κάποια στατιστικά μεγέθη των εικόνων προκειμένου να διεξαχθούν κάποια χρήσιμα συμπεράσματα τα οποία θα μπορούσαν να αξιοποιηθούν για αποδοτικότερη κωδικοποίηση.

11.1 Μετρικές και Στατιστικά Μεγέθη

Αρχικά θα οριστούν τα στατιστικά μεγέθη και οι μετρικές που θα χρησιμοποιηθούν προκειμένου να καταλήξουμε στα επιμέρους συμπεράσματα.

11.1.1 Τυπική Απόκλιση

Το μέγεθος της τυπικής απόκλισης αποτελεί ένα ευρέως χρησιμοποιούμενο μέτρο της μεταβλητότητας ή της ποικιλομορφίας το οποίο χρησιμοποιείται στη στατιστική και τη θεωρία πληροφορίας. Εκφράζει τη διασπορά που υπάρχει στα δεδομένα από το μέσο όρο των δειγμάτων. Από μία χαμηλή τυπική απόκλιση συμπεραίνουμε ότι τα δεδομένα δεν απέχουν πολύ από το μέσο όρο επομένως ενδέχεται να είναι ισχυρά συσχετισμένα μεταξύ τους. Αντίθετα υψηλές τυπικές αποκλίσεις φανερώνουν την ύπαρξη δειγμάτων που κυμαίνεται σε ένα μεγάλο εύρος τιμών. Η τυπική απόκλιση δίνεται από τη σχέση (11.1)

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^2} \tag{11.1}$$

όπου x_i η τιμή κάθε δείγματος φωτεινότητας ενώ $\tilde{x}$ η μέση τιμή των δειγμάτων φωτεινότητας x_i η οποία δίνεται από τη σχέση:

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

11.1.2 Ετεροσυσχέτιση (Cross-Correlation)

Η ετεροσυσχέτιση αποτελεί ένα μέτρο ομοιότητας μεταξύ δύο σημάτων. Πρόκειται ουσιαστικά για την τιμή που προκύπτει από τη συνέλιξη των σημάτων. Συνήθως χρησιμοποιείται σε εφαρμογές αναγνώρισης προτύπων όπου τα δεδομένα της ετεροσυσχέτισης, εμφανίζουν μέγιστη τιμή στο σημείο όπου το πρότυπο συμπίπτει ή είναι όμοιο με κάποιο τμήμα του αρχικού σήματος. Από τα μέγιστα τα οποία θα προκύψουν είναι δυνατό να καθοριστεί και η ακριβής θέση του προτύπου εντός του

αρχικού δείγματος.

$$R(f,t) = \frac{1}{n-1}\sum_{x,y}\frac{\left(f(x,y)-\tilde{f}\right)(t(x,y)-\tilde{t})}{\sigma_f\sigma_t} \tag{11.2}$$

11.1.3 Αμοιβαία Πληροφορία

Η αμοιβαία πληροφορία είναι ένα μέγεθος πολύ σημαντικό στη θεωρία πληροφορίας. Συνήθως στην επεξεργασία εικόνας χρησιμοποιείται με τον ίδιο τρόπου που χρησιμοποιείται και η ετεροσυσχέτιση. Για δύο εικόνες A και B, η αμοιβαία πληροφορία μπορεί να εκφραστεί ως

$$I(A,B) = H(B) - H(B|A) \tag{11.3}$$

όπου $H(B)$ είναι η εντροπία της εικόνας αποχρώσεων του γκρι B σύμφωνα με τον ορισμό του Shannon. Ο συμβολισμός $H(B=b|A=a)$ υποδηλώνει τη δεσμευμένη εντροπία η οποία εκφράζει την πιθανότητα να εμφανιστεί στην εικόνα B ένα εικονοστοιχείο με τιμή B δεδομένου ότι στην εικόνα A υπάρχει ένα εικονοστοιχείο με τιμή a. Θεωρώντας την εντροπία ως μέτρο της αβεβαιότητας τότε η σχέση (11.3) μεταφράζεται ως το μέγεθος της αβεβαιότητας της εικόνας B μείων την αβεβαιότητα της εικόνας B όταν το A είναι γνωστό. Επομένως η αμοιβαία πληροφορία είναι το μέγεθος κατά το οποίο μειώνεται η αβεβαιότητα σε σχέση με το B όταν το A είναι γνωστό, δηλαδή την ποσότητα πληροφορίας όπου το A περιέχει για το B.

11.1.4 Μέση Απόλυτη Διαφορά εντός της Στοιχειώδους Εικόνας (IntraMAD)

Ένα μέγεθος το οποίο είναι σε θέση να εκφράσει τη συσχέτιση μεταξύ των εικονοστοιχείων είναι η μέση απόλυτη διαφορά εντός της στοιχειώδους εικόνας. Όσο μικρότερη η τιμή της, τόσο πιο συσχετισμένα είναι τα δείγματα και επομένως τόσο καλύτερα θα αποδώσουν οι τεχνικές συμπίεσης, είτε πρόκειται για μπλοκ μετασχηματισμό (DCT) είτε για μετασχηματισμό υποζώνης (wavelet). Για τον υπολογισμό του $IntraMAD$ χρησιμοποιήθηκαν τα εικονοστοιχεία $r+1$ και το $c+1$ (όπου r γραμμές και c στήλες της στοιχειώδους εικόνας) δηλαδή για κάθε εικονοστοιχείο χρησιμοποιήθηκε το εικονοστοιχείο που βρίσκεται δεξιά του και το εικονοστοιχείο που βρίσκεται ακριβώς από κάτω του όπως παρουσιάζεται στο Σχήμα 11.1 για μία στοιχειώδη εικόνα 4×4.

(1,1)	(1.2)		
(2,1)			

Σχήμα 11.1 : Τα Γειτονικά Εικονοστοιχεία τα οποία Χρησιμοποιούνται για να Προσδιοριστεί η Συσχέτιση τους με το Εικονοστοιχείο (1,1)

Με μαθηματικό φορμαλισμό η μετρική $IntraMAD$ δίνεται από τη σχέση (11.4)

$$IntraMAD(i,j) = \frac{1}{2 \cdot m \cdot n} \sum_{r=1}^{m-1} \sum_{c=1}^{n-1} \left(\left| f_{ij}(r,c) - f_{i,j}(r,c+1) \right| \right) + \frac{1}{2 \cdot m \cdot n} \sum_{r=1}^{m-1} \sum_{c=1}^{n-1} \left(\left| f_{ij}(r,c) - f_{i,j}(r+1,c) \right| \right) \tag{11.4}$$

11.1.5 Μέση Απόλυτη Διαφορά μεταξύ των Στοιχειωδών Εικόνων (InterMAD)

Αντίστοιχα με την απόλυτη διαφορά εντός της στοιχειώδους εικόνας ορίζεται και η απόλυτη διαφορά μεταξύ των στοιχειωδών εικόνων. Η διαφορά τους έγκειται στο γεγονός ότι πλέον δε χρησιμοποιούνται τα γειτονικά εικονοστοιχεία από την ίδια στοιχειώδη εικόνα αλλά τα ίδια εικονοστοιχεία από τις γειτονικές στοιχειώδεις εικόνες. Το μέτρο $InterMAD$ δίνεται από τη σχέση (11.5)

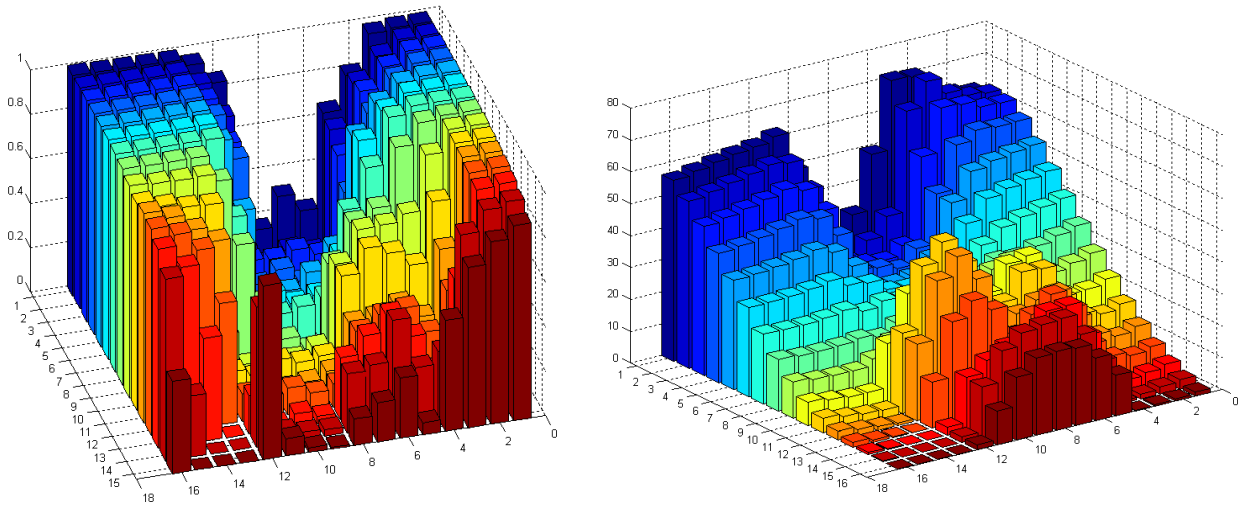
$$InterMAD(i,j) = \frac{1}{2 \cdot m \cdot n} \sum_{r=1}^{m} \sum_{c=1}^{n} \left(\left| f_{ij}(r,c) - f_{i,j+1}(r,c) \right| \right) + \frac{1}{2 \cdot m \cdot n} \sum_{r=1}^{m} \sum_{c=1}^{n} \left(\left| f_{ij}(r,c) - f_{i+1,j}(r,c) \right| \right) \tag{11.5}$$

Στις υποενότητες που θα ακολουθήσουν θα παρατεθούν τόσο τα διαγράμματα όσο και οι τιμές των μεγεθών αυτών με σκοπό τη διεξαγωγή συμπερασμάτων. Παράλληλα θα υπολογιστούν και άλλα μέτρα όπως οι μέσες τιμές και οι τυπικές αποκλίσεις των μεγεθών αυτών στις επιμέρους στοιχειώδεις εικόνες και θα αναζητηθούν μέθοδοι εφαρμογής τους. Τέλος θα πραγματοποιηθεί και μία σύντομη μελέτη σχετικά με τη στατιστική κατανομή των μετασχηματισμών σημάτων, δηλαδή στο πεδίο της συχνότητας.

11.2 Γραφική Αναπαράσταση των Μέτρων και των Στατιστικών Μεγεθών των Τεχνητά Κατασκευασμένων Ολοκληρωτικών Εικόνων

11.2.1 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Ferrari

Η εικόνα της Ferrari, εάν ανατρέξει κανείς στο κεφάλαιο των μετρήσεων, θα διαπιστώσει πως ήταν η εικόνα η οποία συμπιέστηκε αποδοτικότερα, τουλάχιστον σε σχέση με τις τεχνητές εικόνες που κατασκευάστηκαν με προγράμματα προσομοίωσης. Στο Σχήμα 11.2 (a) παρατηρούμε εξαιρετικά υψηλές τιμές κανονικοποιημένης ετεροσυσχέτισης σε πολλές από τις στοιχειώδεις εικόνες οι οποίες αποτελούν την ολοκληρωτική εικόνα της Ferrari. Το γεγονός αυτό υποδηλώνει πως ακόμα και εάν η εικόνα δεν συμπιεστεί αποδοτικά από τους μπλοκ μετασχηματισμούς ή τους μετασχηματισμούς υποζώνης, ο εκτιμητής κίνησης του κωδικοποιητή H.264 Video θα εκμεταλλευτεί στο έπακρον τη συσχέτιση μεταξύ των διαδοχικών στοιχειωδών εικόνων, επιτυγχάνοντας υψηλότατους λόγους συμπίεσης όπως συνέβη και στην εν λόγω εικόνα.

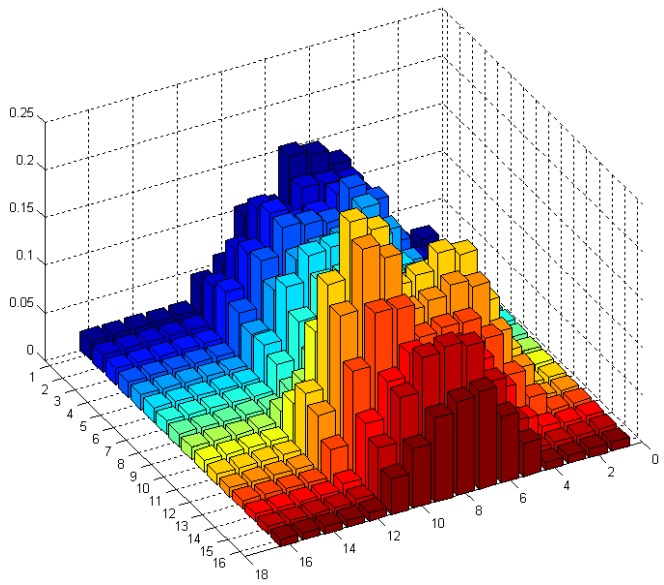


(a) (b)

Σχήμα 11.2 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Ferrari

Στο Σχήμα 11.2 (b) παρατηρούμε τις διασπορές των στοιχειωδών εικόνων. Στη γενική περίπτωση θα μπορούσαμε να πούμε πως όταν μία εικόνα έχει σχετικά μεγάλη διασπορά είναι πιθανό να εμπεριέχει περισσότερη λεπτομέρεια, συνεπώς και υψηλότερες συχνότητες, αλλά και σημαντικότερο πληροφοριακό περιεχόμενο. Το πληροφοριακό περιεχόμενο μίας πηγής γνωρίζουμε πως είναι αντιστρόφως ανάλογο της πιθανότητας εμφάνισης των συμβόλων που παράγει η πηγή καθώς σύμβολο το οποίο εμφανίζεται σπανιότερα φέρει περισσότερη πληροφορία. Αυτός είναι και ο λόγος

όπου η δομή του γραφήματος στο Σχήμα 11.4 (a) το οποίο εκφράζει την εντροπία ανά στοιχειώδη εικόνα, μοιάζει σε μεγάλο βαθμό με το γράφημα της διασποράς του σχήματος Σχήμα 11.2 (b). Υψηλή εντροπία σημαίνει περισσότερο ασυσχέτιστα σύμβολα, επομένως κάθε στοιχειώδης εικόνα μπορεί να θεωρηθεί ως μία πηγή με σύμβολα τις τιμές φωτεινότητας των εικονοστοιχείων. Είναι γεγονός ότι η εντροπία των εικόνων στο πεδίο του "χρόνου" δε μας προσφέρει σημαντική πληροφορία, καθώς οι κωδικοποιητές εντροπίας εφαρμόζονται στα κβαντισμένα σύμβολα που προκύπτουν μετά από την εφαρμογή των μετασχηματισμών συνεπώς το μέγεθος αυτό στο πεδίο του χρόνου θα θεωρηθεί ισοδύναμο της τυπικής απόκλισης. Προκειμένου να αντιληφθούμε πόσο διαφορετική είναι η εντροπία στο πεδίο του "χρόνου" από το πεδίο της συχνότητας και πιο συγκεκριμένα στο πεδίο του μετασχηματισμού DCT παρατίθεται το Σχήμα 11.3.

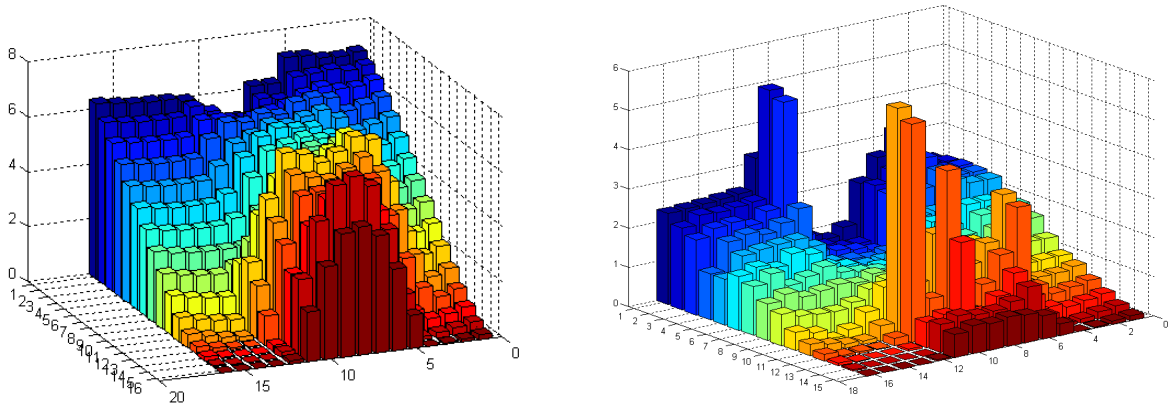


Σχήμα 11.3 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG

Είναι εμφανές πως η δομή των γραφημάτων στο Σχήμα 11.3 και Σχήμα 11.4 (a), δηλαδή της εντροπίας στα δύο διαφορετικά πεδία είναι εντελώς διαφορετική.

Στο Σχήμα 11.4 (b) παρατηρούμε την αμοιβαία πληροφορία μεταξύ των γειτονικών στοιχειωδών εικόνων. Το γράφημα αυτού του μεγέθους θα περιμέναμε να μοιάζει περισσότερο με το γράφημα της ετεροσυσχέτισης δεδομένου ότι τα δύο μεγέθη αν και διαφέρουν, χρησιμοποιούνται σε παρόμοιες εφαρμογές. Η δομή του γραφήματος δε φαίνεται να οδηγεί σε κάποιο χρήσιμο συμπέρασμα. Αναλυτικότερα οι μετρήσεις θα παρατεθούν σε σχετικό πίνακα ο οποίος θα ακολουθήσει της παρούσας υποενότητας

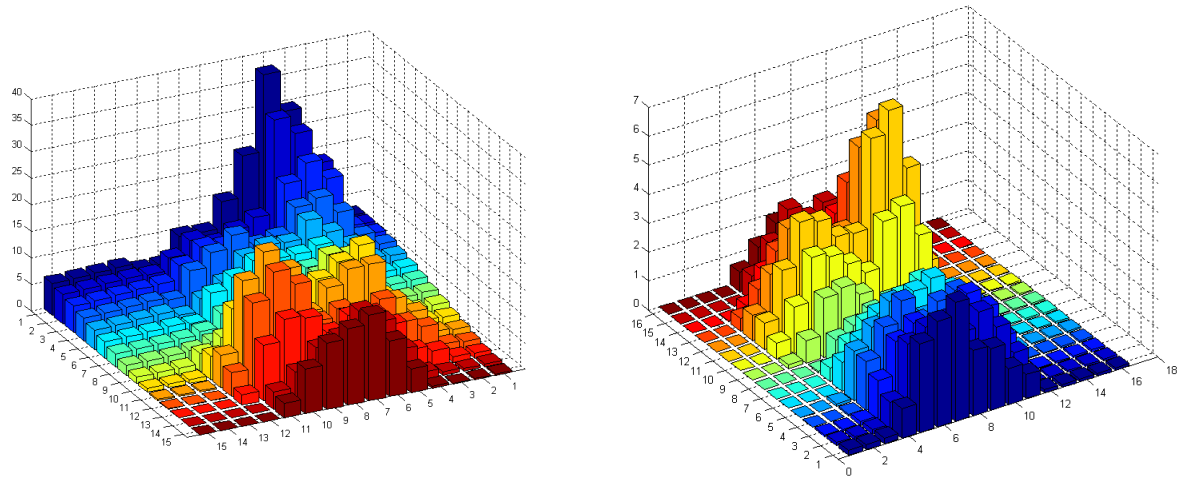
και θα συγκεντρώνει τις μετρήσεις των μεγεθών και των τριών τεχνιτών ολοκληρωτικών εικόνων.



(a) (b)

Σχήμα 11.4 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του "Χρόνου" και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Ferrari

Αντίθετα, τόσο το Σχήμα 11.5 (a) όσο και το Σχήμα 11.5 (b) φαίνεται ότι αποτελούν αξιόπιστα μέτρα. Αναλυτικότερα, η δομή του γραφήματος της απόλυτης διαφοράς μεταξύ των στοιχειωδών εικόνων μοιάζει σημαντικά με το αντίστροφο του γραφήματος της ετεροσυσχέτισης το οποίο όπως προαναφέραμε αποτελεί ένα μέτρο απόδοσης του εκτιμητή κίνησης στον κωδικοποιητή H.264 Video για ολοκληρωτικές εικόνες.



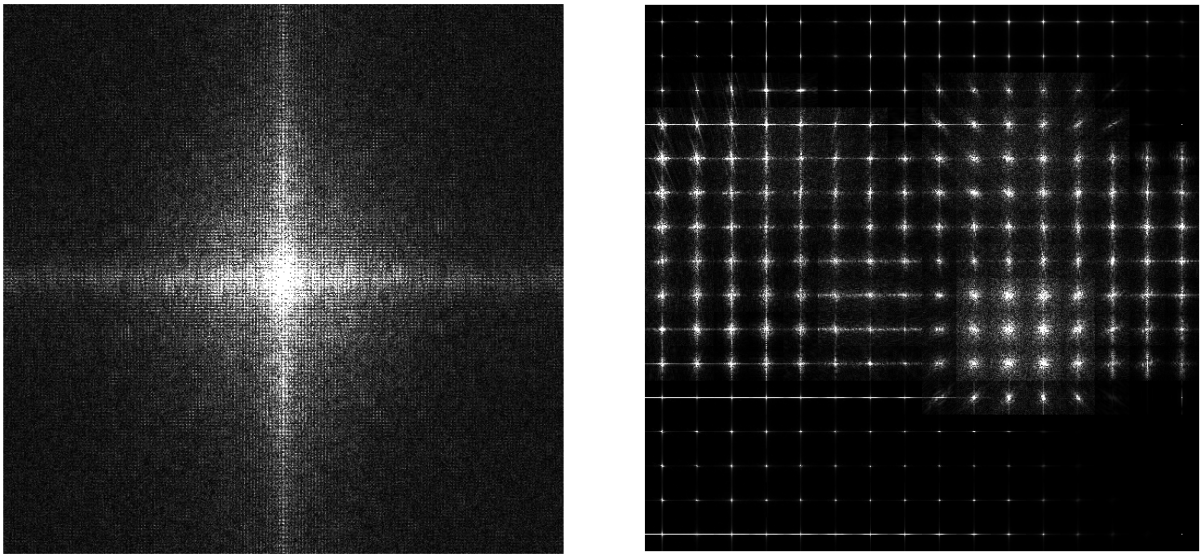
(a) (b)

Σχήμα 11.5 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Ferrari

Η δομή του γραφήματος της μετρικής της απόλυτης διαφοράς εντός της κάθε

στοιχειώδους εικόνας, παρατηρούμε πως μοιάζει εξαιρετικά με τη δομή της εντροπίας των στοιχειωδών εικόνων στο πεδίο του μετασχηματισμού DCT.

Ενδιαφέρον παρουσιάζει το Σχήμα 11.6 το οποίο απεικονίζει το μετασχηματισμό Fourier όταν αυτός εφαρμόζεται σε ολόκληρη την εικόνα και όταν εφαρμόζεται στις επιμέρους στοιχειώδεις εικόνες. Τα γραφήματα στο Σχήμα 11.6 (a) δίνει την εξήγηση σχετικά με λόγο για τον οποίο η εικόνα της Ferrari συμπιέζεται τόσο αποδοτικά από όλους τους κωδικοποιητές. Ο λόγος είναι ότι στην εικόνα δεν υπάρχουν υψηλές συχνότητες, οι οποίες εμφανίζονται με λευκό ή γκρι χρώμα στα άκρα της εικόνας, με αποτέλεσμα η ισχυρή κβάντιση που εφαρμόζεται από όλους τους κωδικοποιητές στους συντελεστές ή τις υποζώνες των υψηλών συχνοτήτων να μην επηρεάζουν πρακτικά την ποιότητα της ανακατασκευασμένης εικόνας. Αντίθετα, όλη η ενέργεια έχει συγκεντρωθεί στο κέντρο του γραφήματος στο Σχήμα 11.6 (a) το οποίο αναπαριστά τις χαμηλές συχνότητες. Στο Σχήμα 11.6 (b) παρατηρούμε πως λίγες στοιχειώδεις εικόνες εμπεριέχουν υψηλές συχνότητας ενώ οι περισσότερες και κυρίως όσες βρίσκονται στην περιφέρεια της εικόνας έχουν συγκεντρωμένη την ενέργεια σε ελάχιστες συνιστώσες που αντιστοιχούν στις χαμηλές συχνότητες.



(a) (b)

Σχήμα 11.6 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Ferrari και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

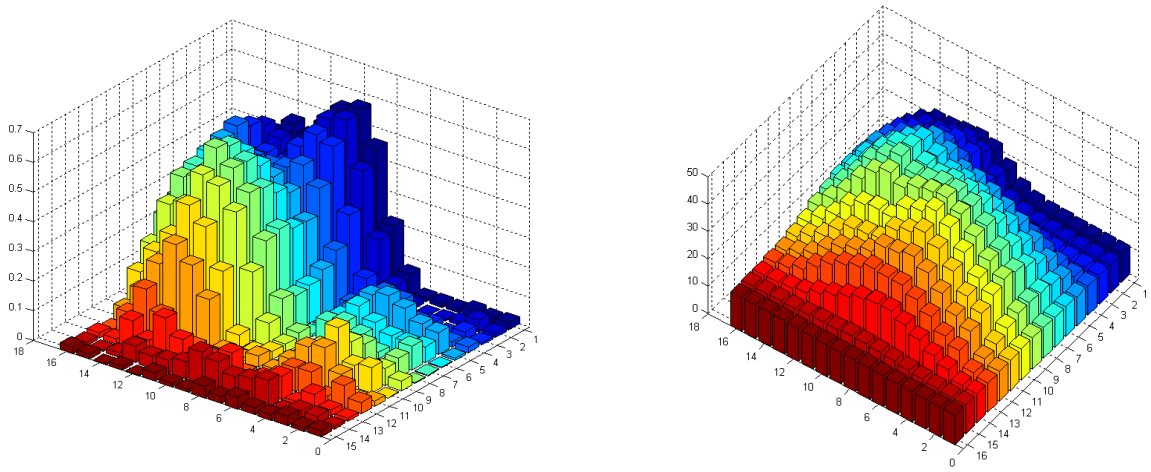
Όπως συζητήθηκε και σε προηγούμενα κεφάλαια, ο μετασχηματισμός DCT είναι αποδοτικότερος του μετασχηματισμού Fourier στη συμπίεση εικόνας. Ο λόγος για τον οποίο χρησιμοποιείται ο μετασχηματισμός Fourier είναι γιατί η γραφική του

αναπαράσταση γίνεται πιο αντιληπτή από το ανθρώπινο μάτι. Η απεικόνιση του μετασχηματισμού DCT μίας εικόνας μεγέθους 2048×2048 δε θα είχε ιδιαίτερο νόημα ώστε να κατανοήσουμε την κατανομή των συχνοτήτων, και αυτό, διότι κατά την υπο-δειγματοληψία των σχημάτων τόσο κατά την εκτύπωση όσο και κατά τη μετατροπή τους από έναν τύπο εικόνας σε άλλο αλλοιωνόταν όλη η ουσία του μετασχηματισμού DCT.

11.2.2 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Cube

Εάν ανατρέξουμε στη γραφική παράσταση που εμφανίζεται στο Σχήμα 10.12 θα θυμηθούμε πως η εικόνα που αναπαριστά τον κύβο, συμπιεζόταν λιγότερο αποδοτικά σε σχέση με την εικόνα της Ferrari. Τα γραφήματα των σχημάτων που ακολουθούν επιβεβαιώνουν πως από στατιστική άποψης, πρόκειται για δύο εντελώς διαφορετικές εικόνες.

Ξεκινώντας με το Σχήμα 11.7 (a) παρατηρούμε πως η ετετοσυσχέτιση μεταξύ των γειτονικών στοιχειωδών εικόνων είναι πολύ μικρή. Αντίθετα με την εικόνα της Ferrari όπου η μέση τιμή ετεροσυσχέτισης των στοιχειωδών εικόνων βρισκόταν στο 0.5916, η εικόνα του κύβου βρίσκεται στο 0.1606. Αυτό εξηγεί το γεγονός ότι η εικόνα του κύβου είτε συμπιεστεί ως εικόνα είτε ως βίντεο-ροη, η επίδοση του κωδικοποιητή H.264 παραμένει σχεδόν ίδια (Σχήμα 10.11).

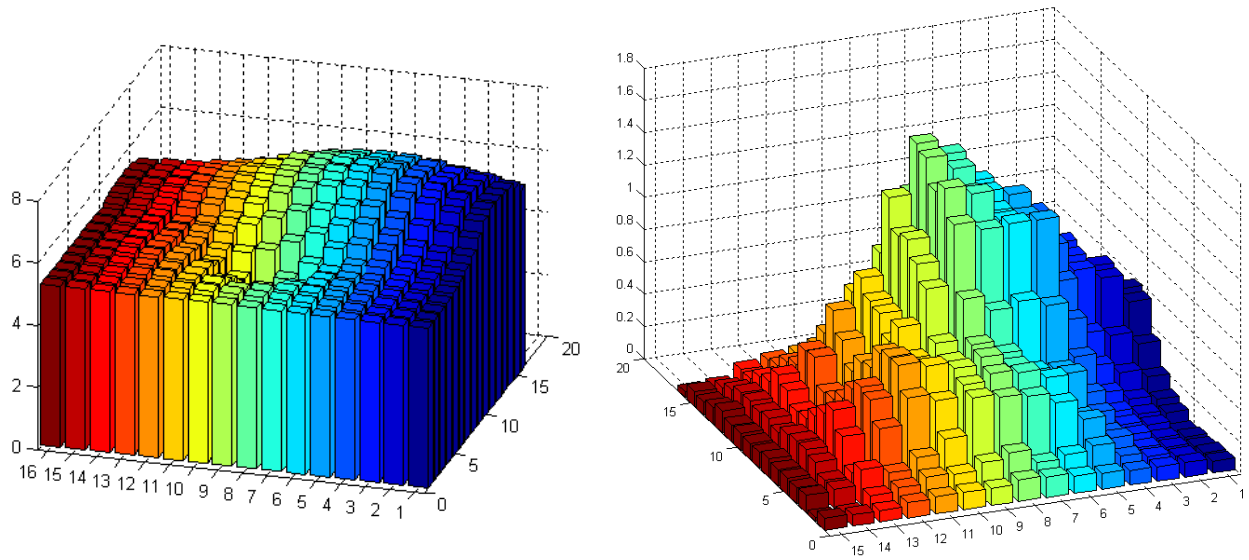


(a) (b)

Σχήμα 11.7 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube

Η δομή της τυπικής απόκλισης στο Σχήμα 11.7 (β) είναι πολύ διαφορετική από το αντίστοιχο γράφημα της Ferrari. Αυτό είναι αναμενόμενο καθώς εάν παρατηρήσουμε τις εικόνες της Ferrari και του κύβου θα διαπιστώσουμε πως στην μεν πρώτη, το θέμα

περιορίζεται στις στοιχειώδεις εικόνες γύρω από το κέντρο της εικόνας ενώ στο δε κύβο, το θέμα εκτίνεται σχεδόν σε όλες τις στοιχειώδεις εικόνες ενώ μόνο οι περιφερειακές περιέχουν το φόντο. Η υψηλή διασπορά στις στοιχειώδεις εικόνες που βρίσκονται στις πρώτες στήλες της εικόνας Ferrari, δεν οφείλονται στην παρουσία θέματος αλλά στον έντονο φωτισμό στο συγκεκριμένο σημείο της εικόνας.



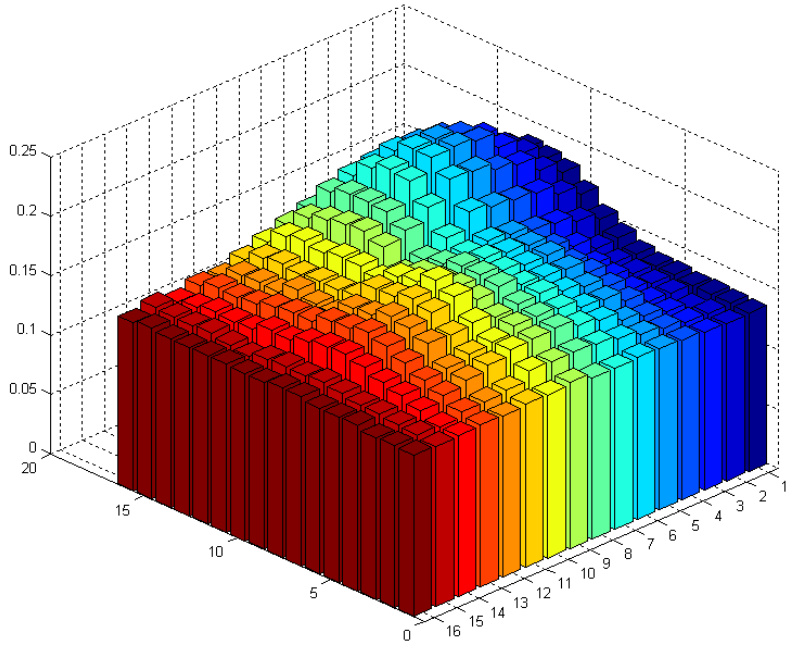
(a) (b)

Σχήμα 11.8 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του "Χρόνου" και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube

Στο Σχήμα 11.8 (a) παρατηρούμε πως η δομή του γραφήματος είναι σε κάποιο βαθμό όμοια με το γράφημα της τυπικής απόκλισης (όπως συνέβαινε και στην περίπτωση της Ferrari) αν και η κατανομή δεν είναι τόσο κοντά στην κανονική κατανομή την οποία παρουσιάζει η τυπική απόκλιση. Ενδιαφέρον παρουσιάζει το Σχήμα 11.8 (b) το οποίο απεικονίζει το μέγεθος της αμοιβαίας πληροφορίας ανά στοιχειώδη εικόνα. Αντίθετα με ότι συνέβαινε στην εικόνα της Ferrari, εδώ η δομή του γραφήματος της αμοιβαίας πληροφορίας είναι όμοια με αυτή της ετεροσυσχέτισης. Κρίνοντας εκ του αποτελέσματος και δεδομένων και των αποτελεσμάτων που θα ακολουθήσουν στη συνέχεια συμπεραίνουμε πως στη γενική περίπτωση τα μεγέθη της ετεροσυσχέτισης και της αμοιβαίας πληροφορίας μας παρέχουν την ίδια πληροφορία, με το μέγεθος της ετεροσυσχέτισης να εμφανίζεται πιο αξιόπιστο για τη συγκεκριμένη εφαρμογή.

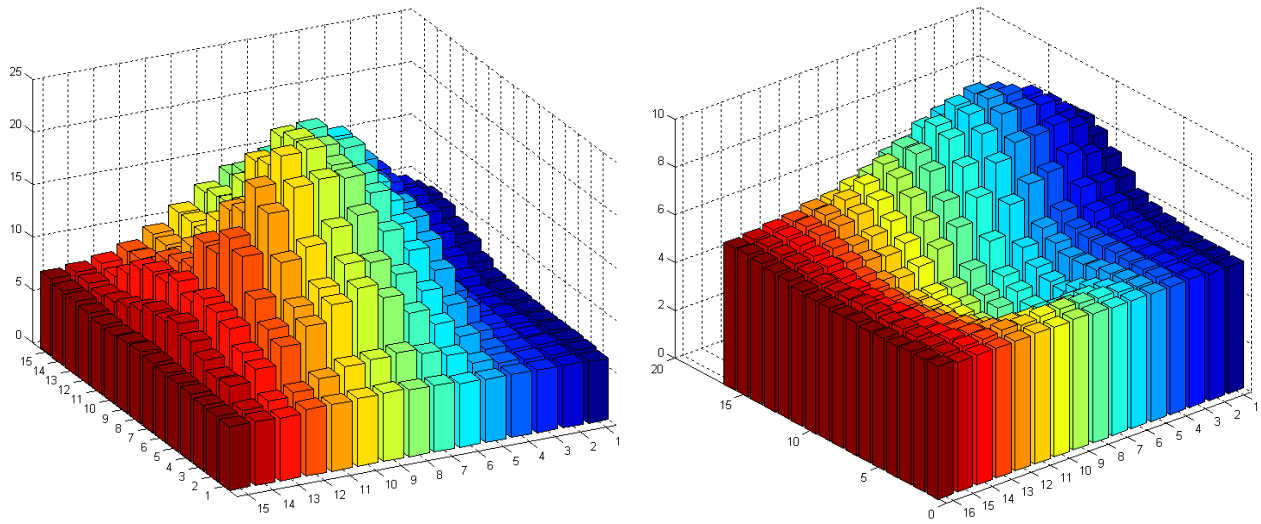
Στο Σχήμα 11.9 παρουσιάζεται ο λόγος συμπίεσης καθενός εκ των 256 στοιχειωδών εικόνων με το πρότυπο JPEG. Η δομή του γραφήματος μοιάζει εξαιρετικά με το

γράφημα της διασποράς αλλά και της εντροπίας στο πεδίο του χρόνου. Υπενθυμίζουμε ότι η εντροπία στο πεδίο του χρόνου δε σχετίζεται σε καμία περίπτωση με την απόδοση της συμπίεσης καθώς ο κωδικοποιητής εντροπίας εφαρμόζεται στους μετασχηματισμένους και κβαντισμένους συντελεστές οι οποίοι στη γενική περίπτωση δεν έχουν τη στατιστική της αρχικής εικόνας.



Σχήμα 11.9 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG

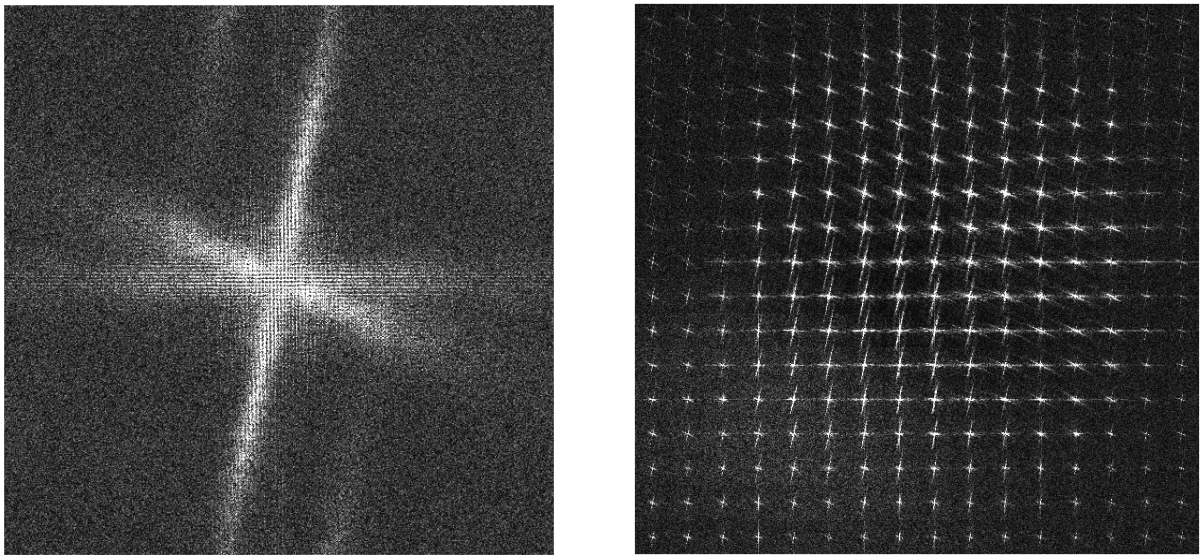
Όπως συνέβη και στην περίπτωση της Ferrari, το μέγεθος της απόλυτης διαφοράς εντός της κάθε στοιχειώδους εικόνας της ολοκληρωτικής εικόνας όπως απεικονίζεται στο Σχήμα 11.10 (b) χαρακτηρίζει πλήρως την μορφή του γραφήματος του λόγου συμπίεσης που επιτυγχάνει το JPEG. Αντίθετα, το μέτρο της απόλυτης διαφοράς μεταξύ των στοιχειωδών εικόνων της ολοκληρωτικής εικόνας δεν έχει την αντίστροφη δομή της ετεροσυσχέτισης όπως θα περιμέναμε, συνεπώς τα μεγέθη τόσο της απόλυτης διαφοράς μεταξύ των στοιχειωδών εικόνων όσο και της αμοιβαίας πληροφορίας παρατηρούμε πως δεν είναι απόλυτα συνεπή με αποτέλεσμα να θεωρήσουμε πως τα μέτρα της ετεροσυσχέτισης και της απόλυτης διαφοράς εντός της κάθε στοιχειώδους εικόνας είναι πιο αξιόπιστα. Αναλυτικά οι αριθμητικές τιμές των μεγεθών παρουσιάζει ο Πίνακας 11.1.



(a) (b)

Σχήμα 11.10 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Cube

Τέλος, το Σχήμα 11.11 αποτελεί τη γραφική αναπαράσταση των συντελεστών του μετασχηματισμού Fourier της εικόνας Cube όταν εφαρμόζεται σε ολόκληρη την εικόνα και όταν εφαρμόζεται ανά στοιχειώδη εικόνα. Οι υψηλές συχνότητες είναι εμφανείς τόσο στο (a) όσο και (b) μέρος του σχήματος. Εάν κοιτάξουμε πιο προσεκτικά το Σχήμα 11.11 (b) και το συγκρίνουμε με το Σχήμα 11.9 θα διαπιστώσουμε πως παρά το γεγονός ότι οι στοιχειώδεις εικόνες στο κέντρο του σχήματος έχουν μεγαλύτερες συνιστώσες στις χαμηλές συχνότητες, συμπιέζονται πιο αποδοτικά από τις στοιχειώδεις εικόνες στο κάτω αριστερό μέρος της εικόνας οι οποίες ναι μεν έχουν μικρότερους συντελεστές στις χαμηλές συχνότητες αλλά σημαντικό μέρος τους περιέχει και πολύ υψηλές συχνότητες.

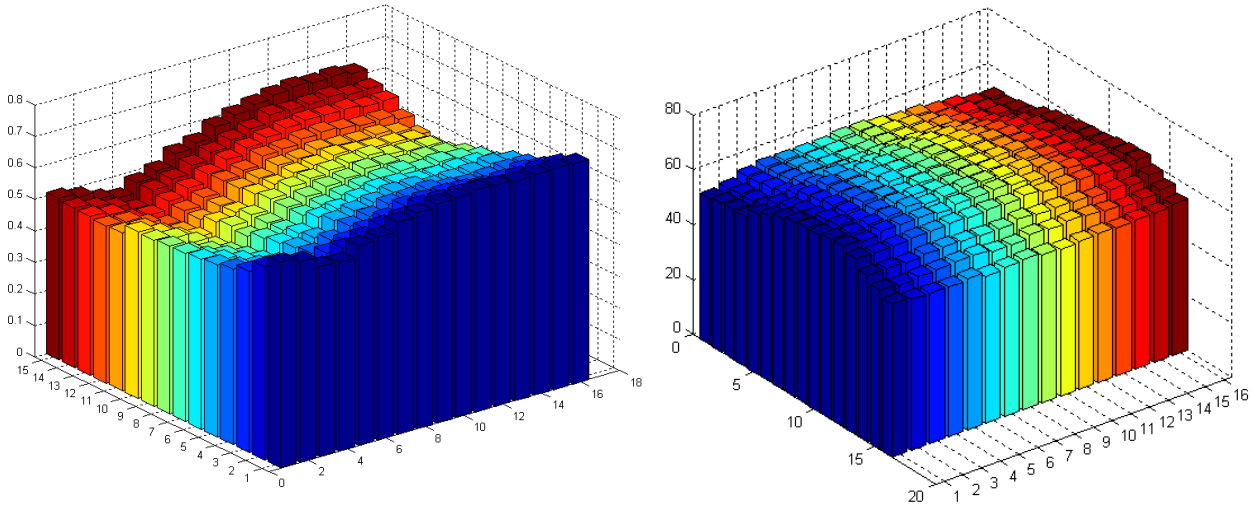


(a) (b)

Σχήμα 11.11 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Cube και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

11.2.3 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Εικόνας Cat

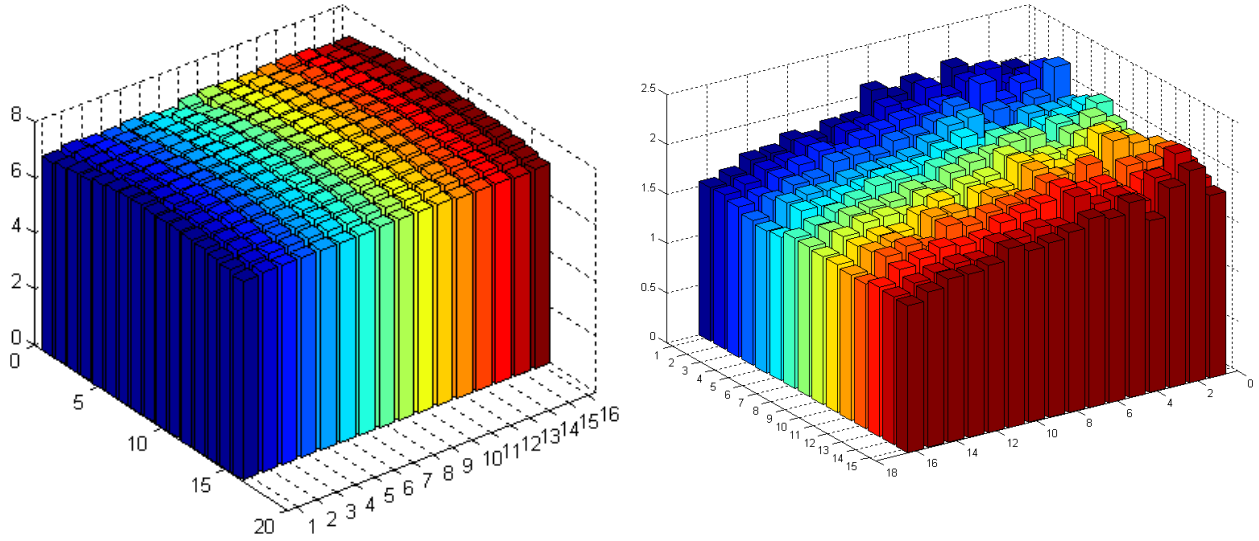
Η τελευταία εικόνα η οποία θα εξετασθεί είναι η εικόνα Cat η οποία κωδικοποιήθηκε λιγότερο αποδοτικά σε σχέση με οποιαδήποτε άλλη εικόνα. Όλα μας τα συμπεράσματα μέχρι και το σημείο αυτό ισχύουν και σε αυτή την περίπτωση.



(a) (b)

Σχήμα 11.12 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cat

Η ικανοποιητική ετεροσυσχέτιση με τιμή 0.5716 υποδηλώνει πως ο κωδικοποιητής H.264 Video θα προσφέρει σημαντικά καλύτερα αποτελέσματα σε σχέση με οποιαδήποτε άλλη μέθοδο η οποία δεν εκμεταλλεύεται τη συσχέτιση μεταξύ των στοιχειωδών εικόνων.

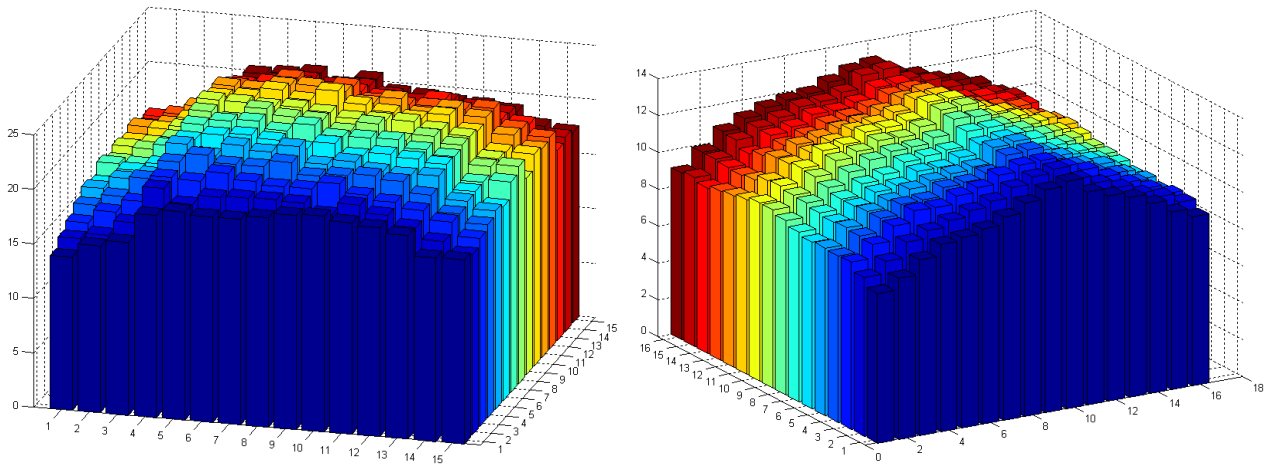


(a) (b)

Σχήμα 11.13 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του "Χρόνου" και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Εικόνας Cube

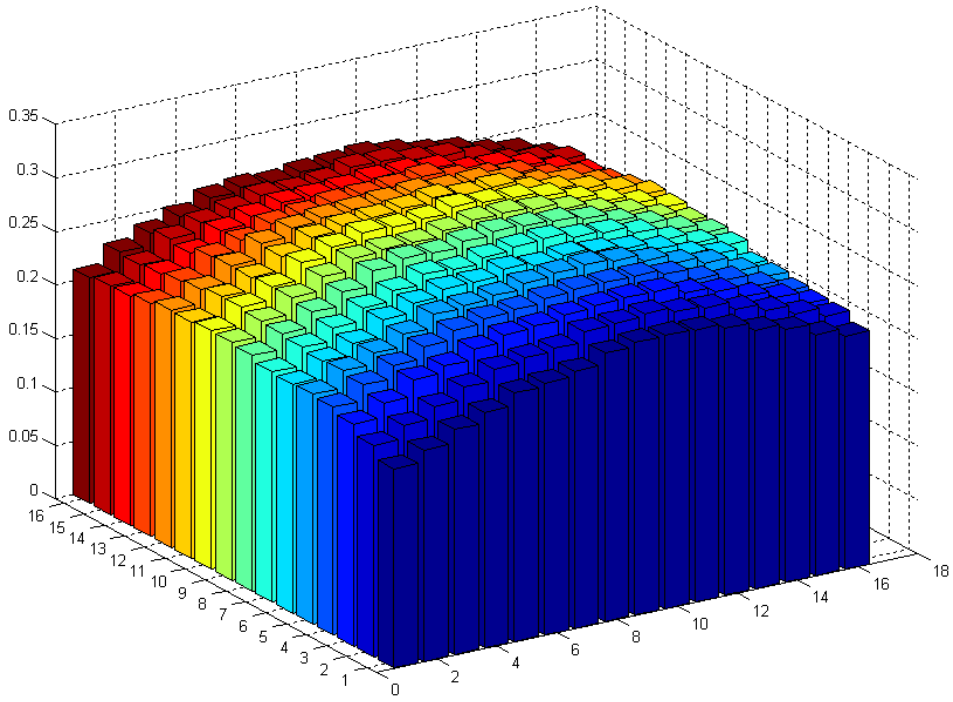
Η φτωχή σε διακυμάνσεις κατανομή των μεγεθών της εικόνας Cat, και κυρίως της τυπικής απόκλισης, φανερώνουν πως δεν θα έχουμε στη διάθεση μας πολλά στοιχεία προς εκμετάλλευση ώστε να συμπιέσουμε αποδοτικότερα την εικόνα. Συνεπώς ότι θα ισχύσει για κάθε στοιχειώδη εικόνα θα ισχύσει και για όλες τις υπόλοιπες. Ασφαλώς όπως προαναφέρθηκε περιμένουμε μία βελτιστοποίηση της κωδικοποίησης από τη σημαντική συσχέτιση μεταξύ των στοιχειωδών εικόνων, από τον κωδικοποιητή H.264.

Τη διαπίστωση αυτή μπορούμε να την επιβεβαιώσουμε παρατηρώντας το Σχήμα 11.15. Η δομή του σχήματος είναι όμοια με τη δομή του μεγέθους της απόλυτης διαφοράς εντός της κάθε στοιχειώδους εικόνας της ολοκληρωτικής εικόνας αλλά και της τυπικής απόκλισης.



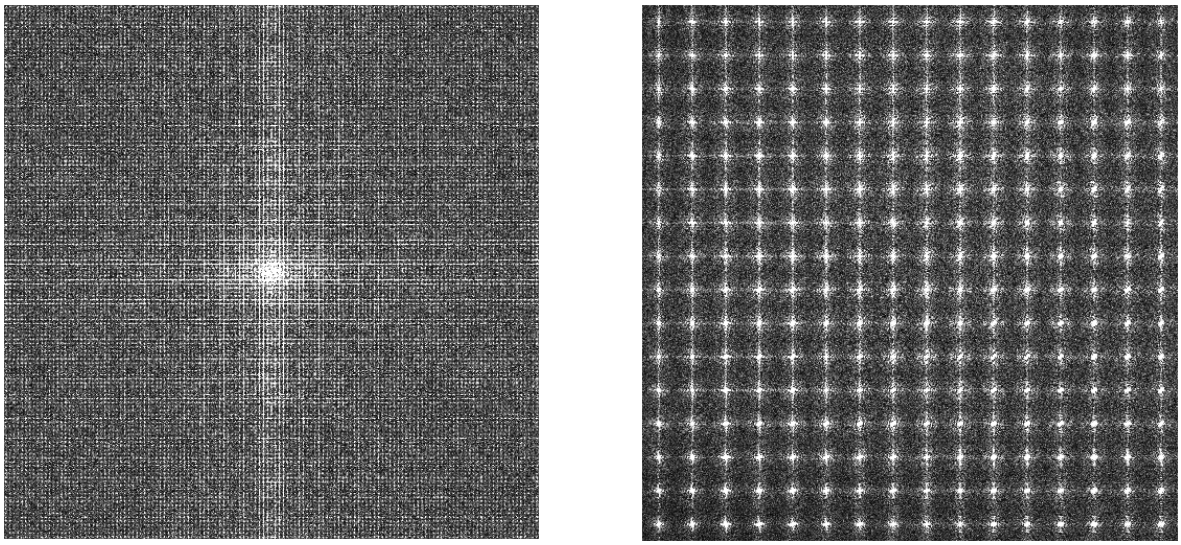
(a) (b)

Σχήμα 11.14 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Εικόνας Cat



Σχήμα 11.15 : Ο Λόγος του Αρχικού Μεγέθους της κάθε Στοιχειώδους Εικόνας με το Μέγεθος που Επετεύχθη μετά την Κωδικοποίηση JPEG

Όπως αντιλαμβανόμαστε, μία εικόνα σαν και αυτή, εάν είχε χαμηλές συχνότητες, τότε όλες οι στοιχειώδεις εικόνες θα συμπιέζονταν αποδοτικά οδηγώντας συνολικά σε μία πολύ καλή κωδικοποίηση της ολοκληρωτικής εικόνας. Αντίθετα, η εικόνα Cat, εμπεριέχει πολύ υψηλές συχνότητες, γεγονός που παρατηρούμε στο Σχήμα 11.16.



(a) (b)

Σχήμα 11.16 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Cat και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

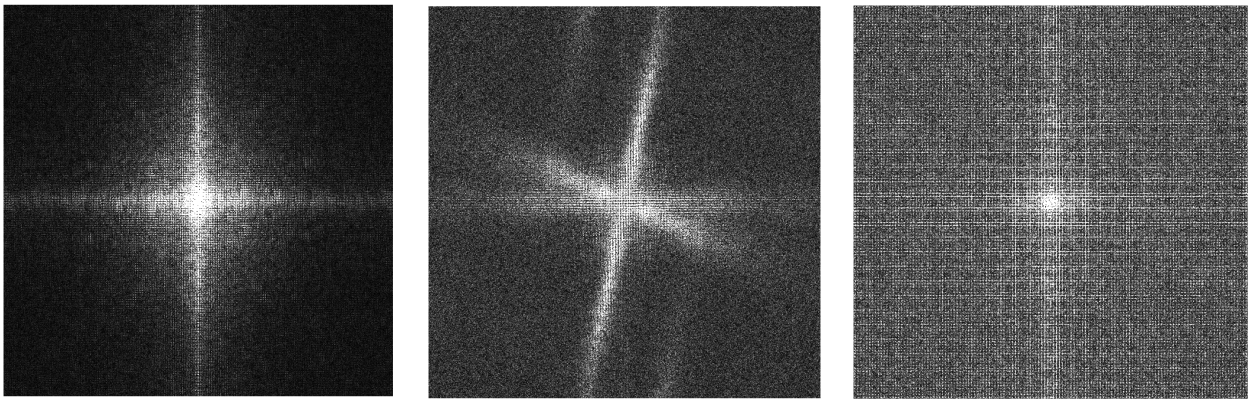
Κατά την πραγματοποίηση των μετρήσεων και στην προσπάθεια διεξαγωγής συμπερασμάτων επιχειρήθηκε η χρήση διαφόρων μεγεθών τα οποία χρησιμοποιούνται για να προσδιορίσουν την υφή μίας εικόνας προκειμένου να περιγραφεί με ένα μονόμετρο μέγεθος κατά πόσο μία εικόνα περιέχει υψηλές ή χαμηλές συχνότητες. Μεταξύ των μεγεθών αυτών είναι η τυπική απόκλιση, η ασυμμετρία (skewness) και η κύρτωση (kurtosis) των απόλυτων τιμών των συντελεστών του μετασχηματισμού Fourier. Δυστυχώς κανένα μέτρο από αυτά δεν κατέστη δυνατό να περιγράψει με μία τιμή το μέγεθος της κατανομής των συχνοτήτων και στις έξι εικόνες που παρουσιάστηκαν σε αυτή την εργασία (3 κατασκευασμένες και 3 φυσικές). Σε αυτές τις περιπτώσεις απαιτείται κάποιος ταξινομητής στον οποία θα δίδονται όλες οι τιμές ως διάνυσμα και αφού εκπαιδευθεί θα μπορεί στη συνέχει να μας παρέχει μία εκτίμηση σε σχέση με το εάν μία εικόνα περιέχει περισσότερες υψηλές ή χαμηλές συχνότητες και πώς αυτό μπορεί να κατηγοριοποιηθεί στο εάν θα υπάρχει μία αποδοτική ή όχι κωδικοποίηση.

Ο Πίνακας 11.1 παραθέτει τις αριθμητικές τιμές των μεγεθών και για τις τρείς εικόνες της υποενότητας.

Πίνακας 11.1 : Οι Αριθμητικές Τιμές των Μεγεθών Που Χρησιμοποιήθηκαν για την Περιγραφή των Ολοκληρωτικών Εικόνων της υποενότητας 10.1

	Εικόνα		
	Ferrari	Cube	Cat
Ετεροσυσχέτιση Μεταξύ των Γειτονικών Στοιχειωδών Εικόνων	0.5916	0.1606	0.5716
Τυπική Απόκλιση	39.6188	28.6954	62.8188
Μέση (μεταξύ των EI) Τυπική Απόκλιση	27.8437	23.6222	61.0251
Τυπική Απόκλιση των Τυπικών Αποκλίσεων	18.3389	10.7682	4.0116
Εντροπία	4.8346	6.4026	7.5770
Μέση (μεταξύ των EI) Εντροπία	3.9821	5.7053	7.3829
Τυπική Απόκλιση των Εντροπιών (των EI)	2.1446	0.6804	0.1551
Αμοιβαία Πληροφορία	1.2864	0.4815	1.8503
Inter-MAD	8.1587	11.9008	19.8181
Intra-MAD	1.2038	6.0370	11.3662

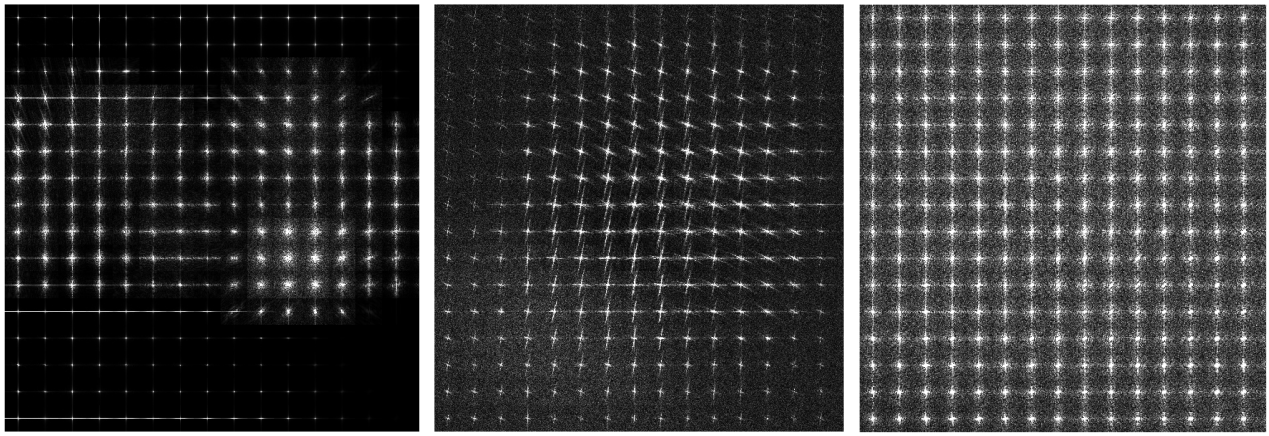
Η παράθεση των γραφικών απεικονίσεων του μετασχηματισμού Fourier των τριών εικόνων, όταν τον εφαρμόσουμε σε κάθε εικόνα παρουσιάζεται στο Σχήμα 11.17



(a) (b) (c)

Σχήμα 11.17 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων

Η παράθεση των γραφικών απεικονίσεων του μετασχηματισμού Fourier των τριών εικόνων, όταν τον εφαρμόσουμε σε κάθε στοιχειώδη εικόνα παρουσιάζεται στο Σχήμα 11.18



(a) (b) (c)

Σχήμα 11.18 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων ανά Στοιχειώδη Εικόνα

11.2.4 Συμπεράσματα που Απορρέουν από τις Μετρήσεις

Η τυπική απόκλιση σε συνδυασμό με την απόλυτη διαφορά εντός της κάθε στοιχειώδους εικόνας μίας ολοκληρωτικής εικόνας μπορούν να προσεγγίσουν ικανοποιητικά τις επιδόσεις ενός κωδικοποιητή συμπίεσης εικόνας τόσο σε ότι αφορά τη δομή του λόγου συμπίεσης όσο και της ικανότητας συμπίεσης (για παράδειγμα το γινόμενο των δύο μεγεθών). Το μέγεθος της διασποράς όπως διαπιστώσαμε στην περίπτωση της εικόνας Ferrari μπορεί να είναι παραπλανητικό καθώς στοιχειώδεις εικόνες οι οποίες με έντονο φωτισμό έχουν υψηλή διασπορά όμως λόγο της κατανομής της φωτεινότητας η οποία είναι συγκεντρωμένη σε συγκεκριμένο σημείο συμπιέζονται μέσω των μετασχηματισμών πολύ αποδοτικά. Αντίθετα, το μέτρο της απόλυτης διαφοράς εντός της κάθε στοιχειώδους εικόνας δεν επηρεάζεται από το φαινόμενο αυτό.

Σε ότι αφορά τα περιθώρια για βελτιστοποίηση την απόδοσης της συμπίεσης από τον κωδικοποιητή H.264 Video, όπως διαπιστώσαμε, μπορούμε να τα συμπεράνουμε χρησιμοποιώντας το μέγεθος της ετεροσυσχέτισης των γειτονικών στοιχειωδών εικόνων.

Ενδιαφέρον αποτέλεσμα αποτελεί η τυπική απόκλιση των τυπικών αποκλίσεων (Πίνακας 11.1) των στοιχειωδών εικόνων. Το μέτρο αυτό είναι δυνατό να μας πληροφορήσει σχετικά με την ύπαρξη ή μη, στοιχειωδών εικόνων οι οποίες εμπεριέχουν θέμα ή φόντο. Μία εικόνα με μικρή την τυπική απόκλιση των τυπικών αποκλίσεων συμπεραίνουμε πως σχεδόν σε κάθε στοιχειώδη εικόνα υπάρχει ένα μέρος του θέματος (Cat και Cube). Αντίθετα, εάν η τιμή του μεγέθους αυτού είναι μεγάλη τότε ενδέχεται στην εικόνα να υπάρχει πλήθος στοιχειωδών εικόνων με φόντο.

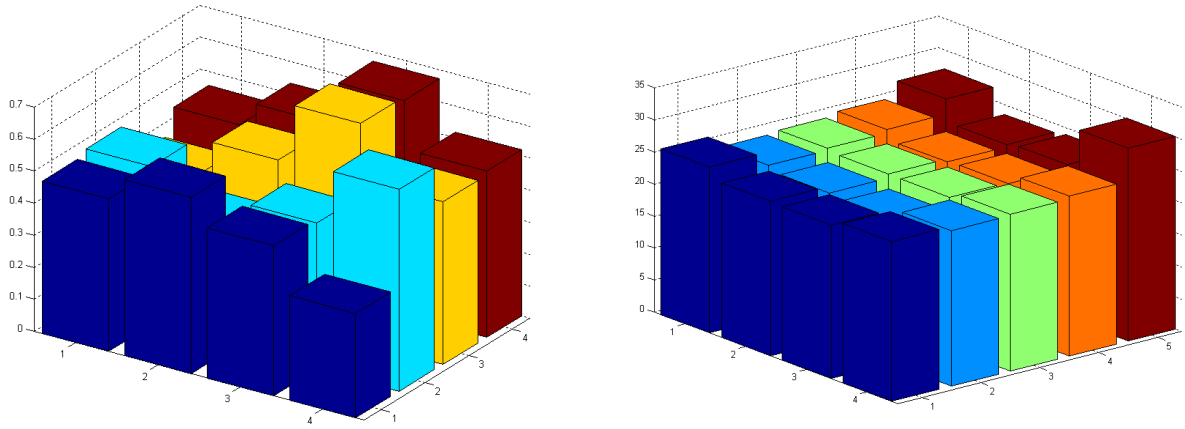
Αυτό μας γεννά την ιδέα ότι ένας εξυπνότερος κωδικοποιητής JPEG που γνωρίζει πως κωδικοποιεί ολοκληρωτικές εικόνες θα εξετάζει την τιμή της τυπικής απόκλισης των τυπικών αποκλίσεων εντός της κάθε στοιχειώδους εικόνας και θα αποφασίζει δυναμικά την τιμή κβάντισης της συγκεκριμένης στοιχειώδους εικόνας από ένα πλήθος N επιλογών τιμών κβάντισης. Έτσι, στοιχειώδεις εικόνες με έντονο φόντο (χαμηλές συχνότητες) θα κβαντίζονται ισχυρότερα ενώ όσες περιέχουν θέμα θα κβαντίζονται λιγότερο επιθετικά. Με τον τρόπο αυτό οι εικόνες θα μπορούν να συμπιεστούν αποδοτικότερα ενώ τα οφέλη από την ποιότητα που αντιλαμβάνεται ο χρήστης ενδέχεται να είναι ακόμα σημαντικότερα.

Τα αποτελέσματα από την τροποποίηση αυτή στο JPEG θα παρουσιαστούν στην τελευταία υποενότητα της εργασίας.

11.3 Γραφική Αναπαράσταση των Μέτρων και των Στατιστικών Μεγεθών των Φυσικών Ολοκληρωτικών Φωτογραφιών

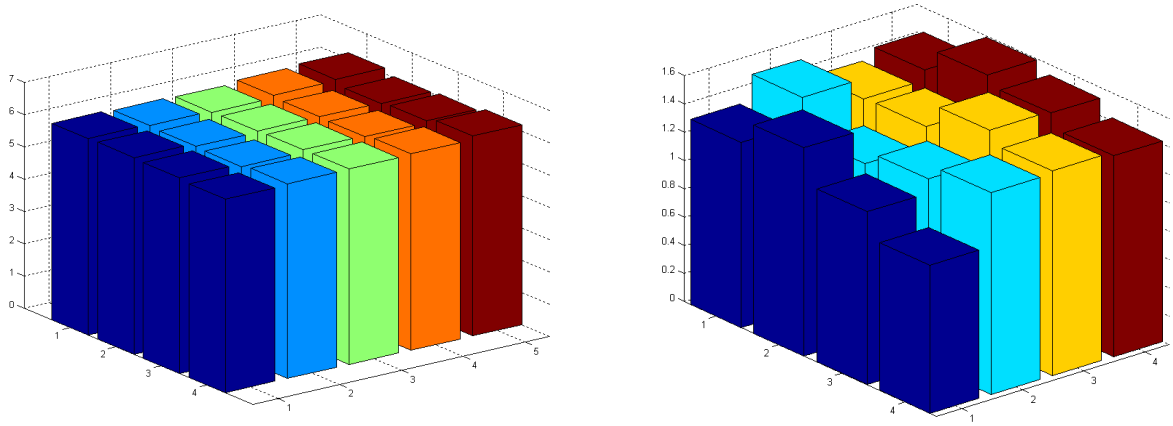
Για την πληρότητα της εργασίας και προς επιβεβαίωση των μέχρι στιγμής συμπερασμάτων και διαπιστώσεων θα ακολουθήσουν τα γραφήματα των μεγεθών στα όποια πραγματοποιήθηκε εκτενής αναφορά στην προηγούμενη υποενότητα για τις πραγματικές εικόνες ολοκληρωτικής φωτογράφισης..

11.3.1 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Birds



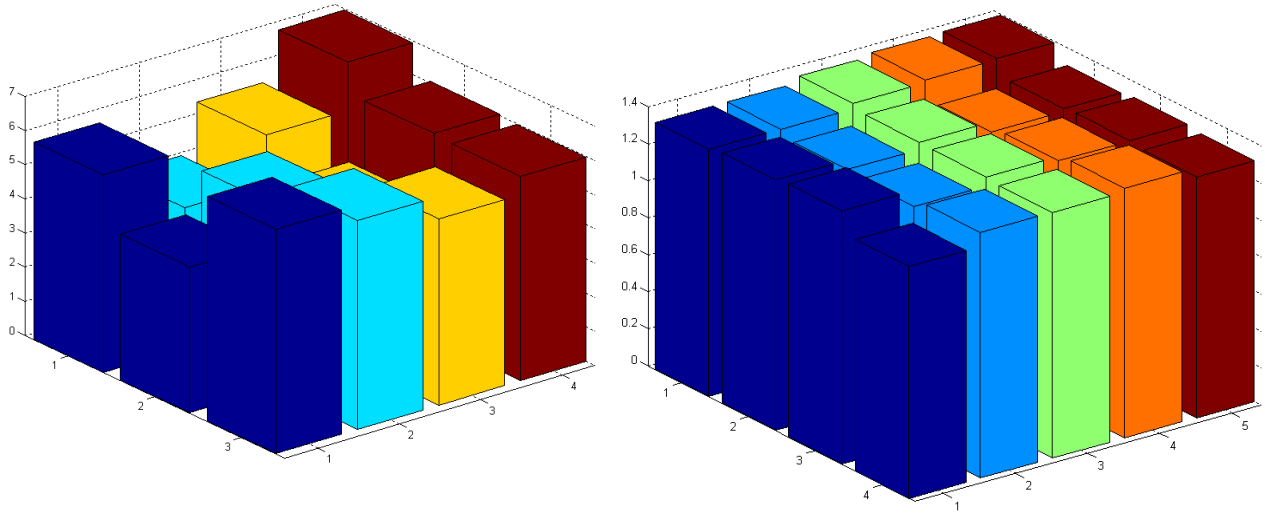
(a) (b)

Σχήμα 11.19 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Birds



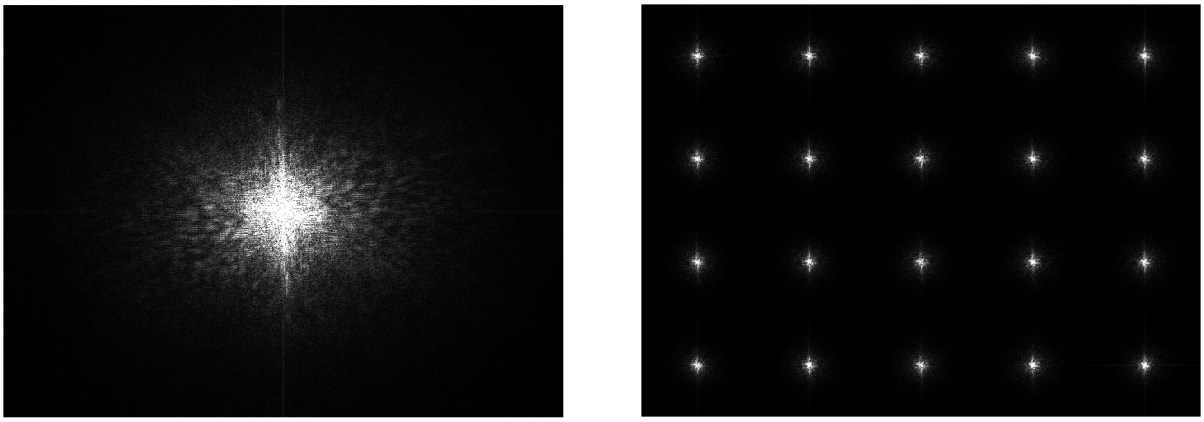
(a) (b)

Σχήμα 11.20 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Birds



(a) (b)

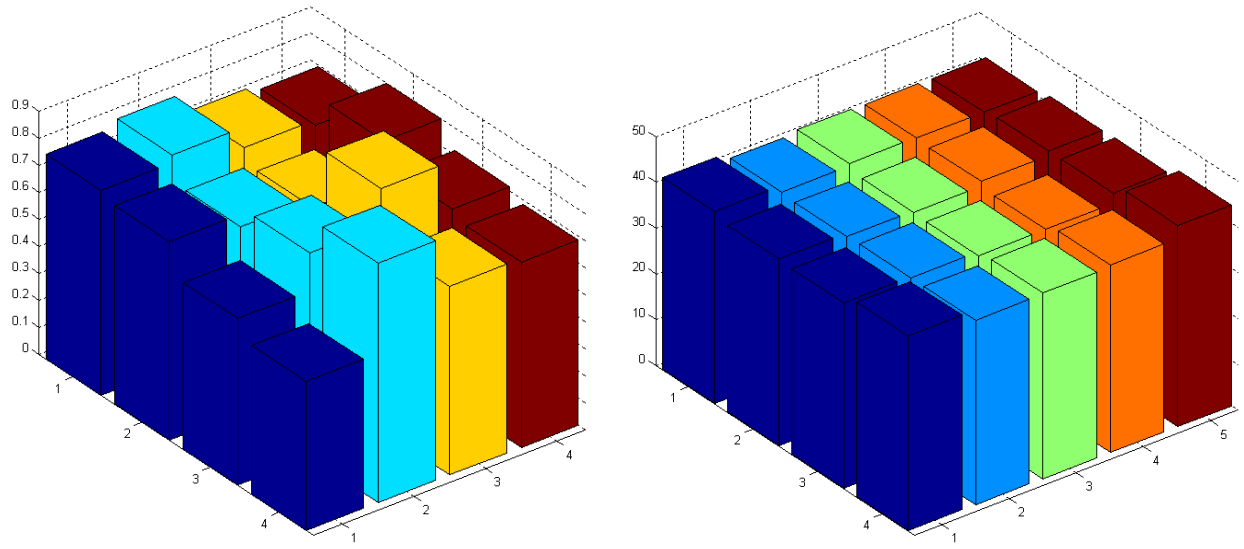
Σχήμα 11.21 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Birds



(a) (b)

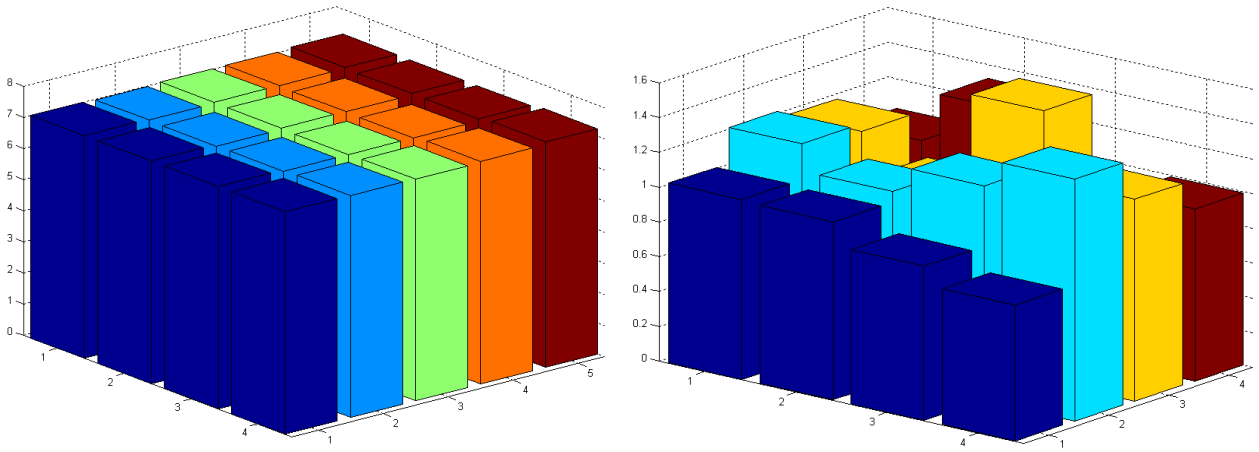
Σχήμα 11.22 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Birds και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

11.3.2 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Seagull



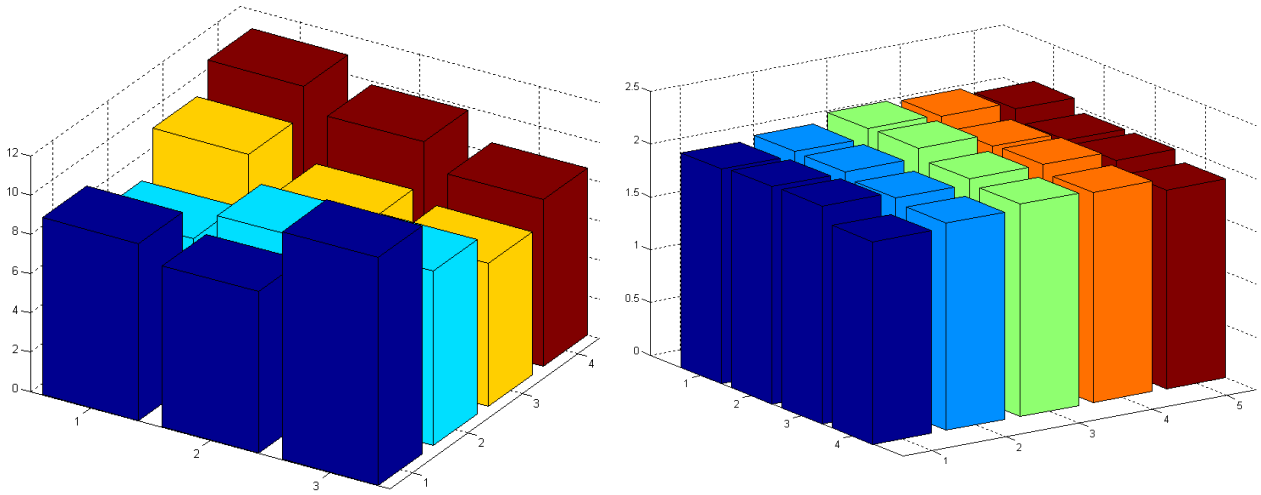
(a) (b)

Σχήμα 11.23 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Seagull



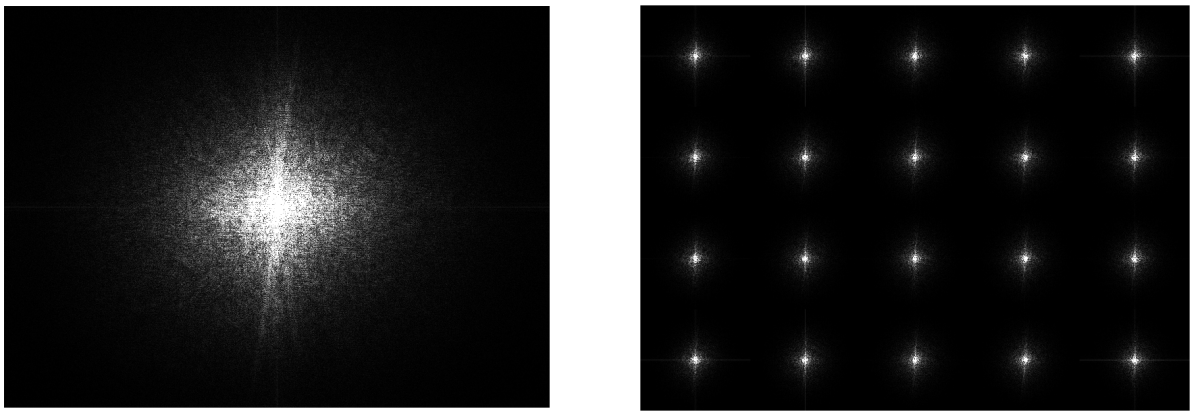
(a) (b)

Σχήμα 11.24 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του "Χρόνου" και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Seagull



(a) (b)

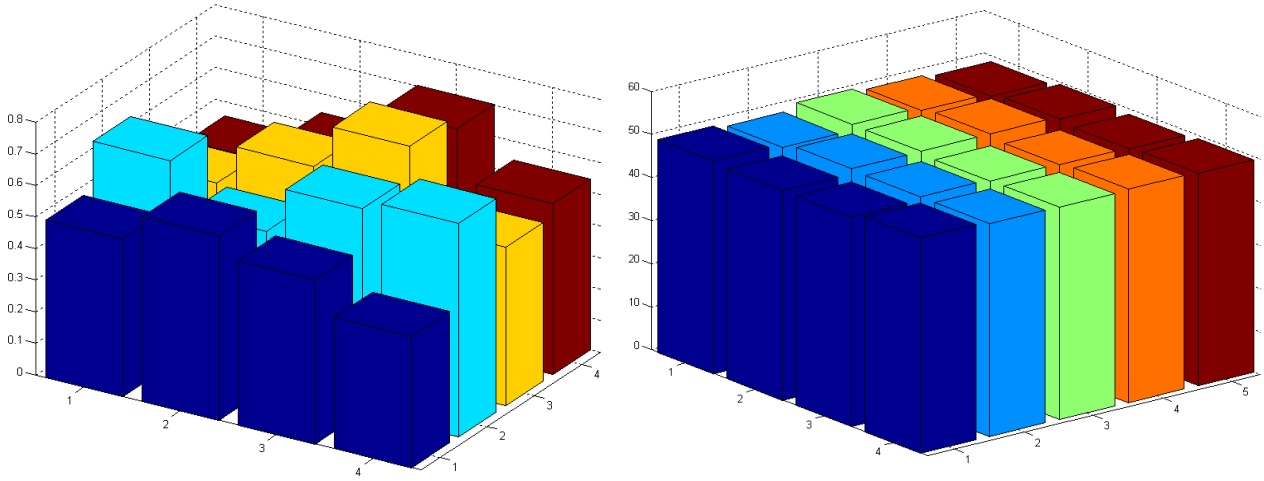
Σχήμα 11.25 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Seagull



(a) (b)

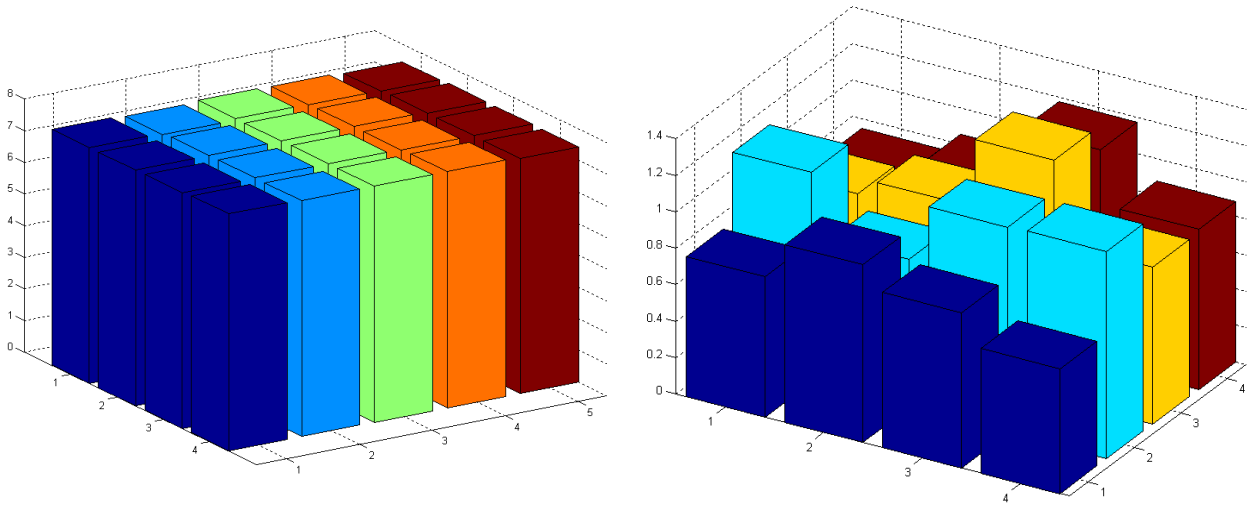
Σχήμα 11.26 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Seagull και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

11.3.3 Συμπεράσματα Σχετικά με την Κωδικοποίηση της Φωτογραφίας Sport



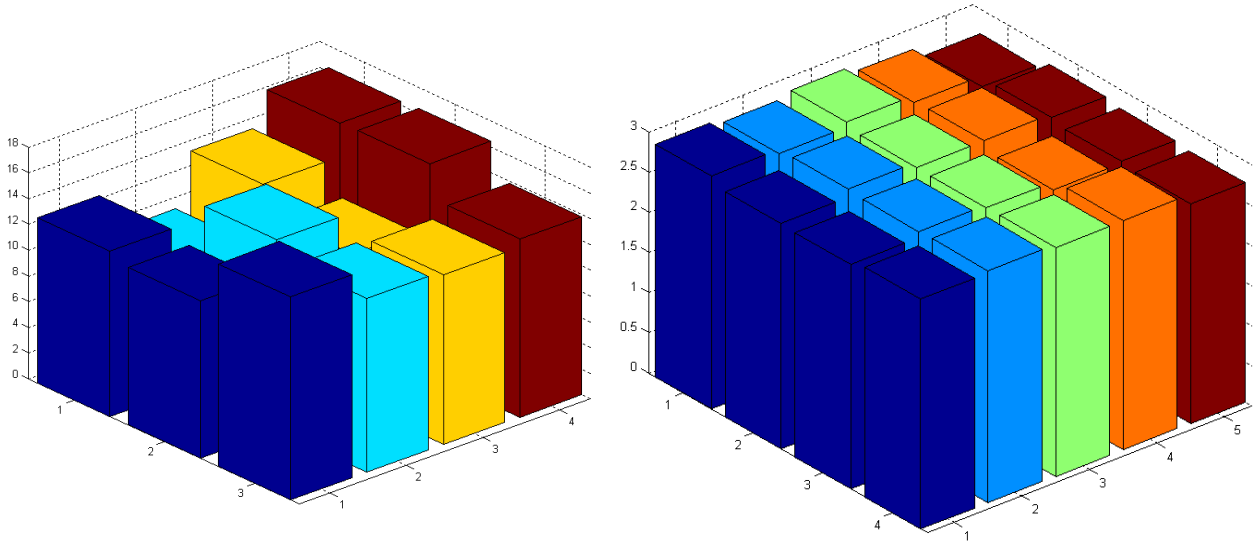
(a) (b)

Σχήμα 11.27 : Οι Τιμές της (a) Κανονικοποιημένης Ετεροσυσχέτισης και της (b) Τυπικής Απόκλισης ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Sport



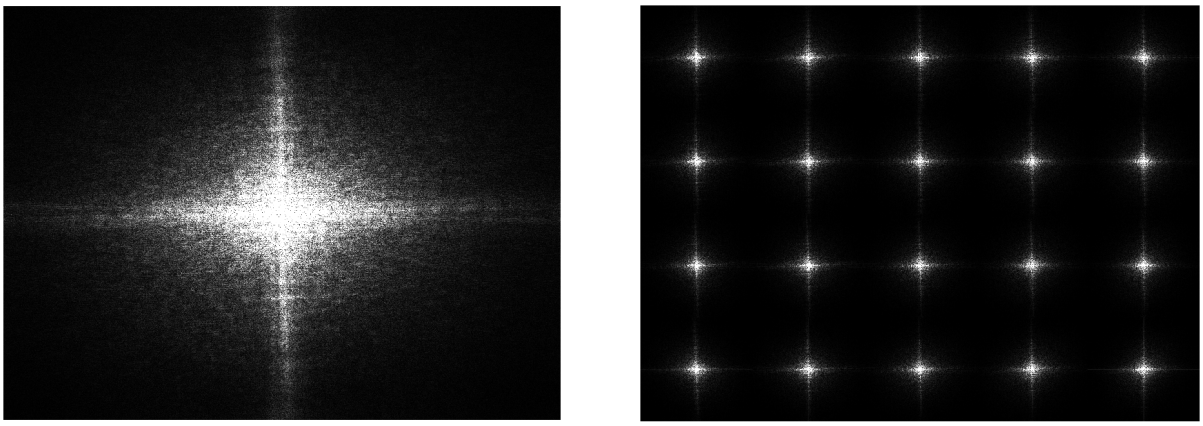
(a) (b)

Σχήμα 11.28 : Οι Τιμές της (a) Εντροπίας στο Πεδίο του “Χρόνου” και της (b) Αμοιβαίας Πληροφορίας ανά Στοιχειώδη Εικόνα της Ολοκληρωτικής Φωτογραφίας Sport



(a) (b)

Σχήμα 11.29 : Οι Τιμές της (a) Απόλυτης Διαφοράς Μεταξύ των Στοιχειωδών Εικόνων και της (b) Απόλυτης Διαφοράς Εντός της κάθε Στοιχειώδους Εικόνας της Ολοκληρωτικής Φωτογραφίας Sport

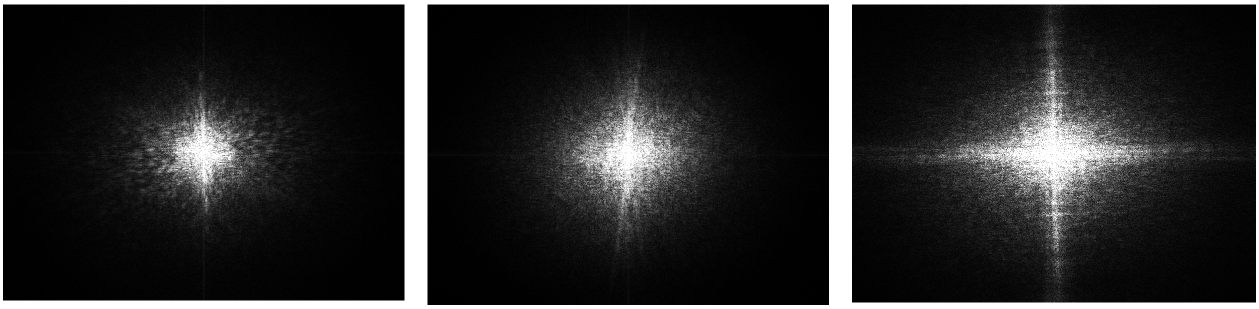


(a) (b)

Σχήμα 11.30 : Γραφική Αναπαράσταση του Μετασχηματισμού Fourier (a) στην Ολοκληρωτική Εικόνα Sport και (b) στις Επιμέρους Στοιχειώδεις Εικόνες

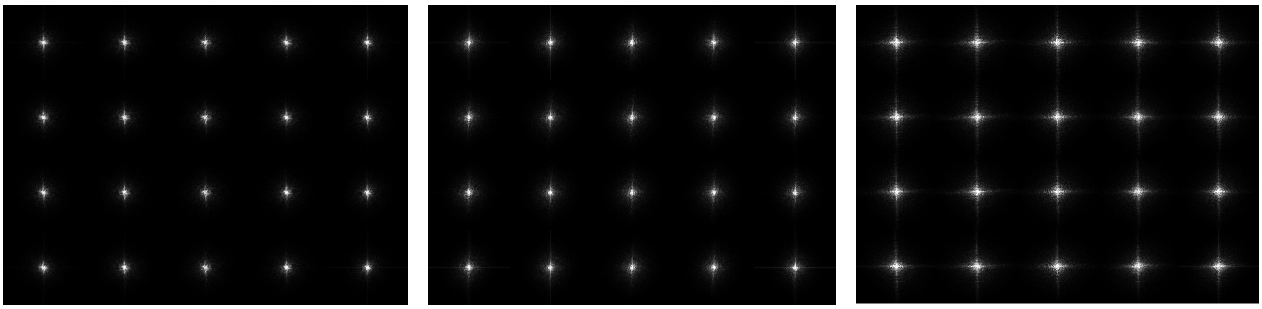
Πίνακας 11.2 : Οι Αριθμητικές Τιμές των Μεγεθών Που Χρησιμοποιήθηκαν για την Περιγραφή των Ολοκληρωτικών Φωτογραφιών

	Εικόνα		
	Birds	Seagull	Sport
Ετεροσυσχέτιση Μεταξύ των Γειτονικών Στοιχειωδών Εικόνων	0.5017	0.7275	0.5665
Τυπική Απόκλιση	24.5696	40.9697	49.4346
Μέση (μεταξύ των EI) Τυπική Απόκλιση	24.4336	40.9437	49.3542
Τυπική Απόκλιση των Τυπικών Αποκλίσεων	1.6739	1.0621	0.4918
Εντροπία	6.1283	7.1660	7.4749
Μέση (μεταξύ των EI) Εντροπία	6.0609	7.1373	7.4546
Τυπική Απόκλιση των Εντροπιών (των EI)	0.0558	0.0380	0.0179
Αμοιβαία Πληροφορία	1.3853	1.1169	0.9456
Inter-MAD	5.5487	8.9810	13.6589
Intra-MAD	1.3296	2.0332	2.8301



(a) (b) (c)

Σχήμα 11.31 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων



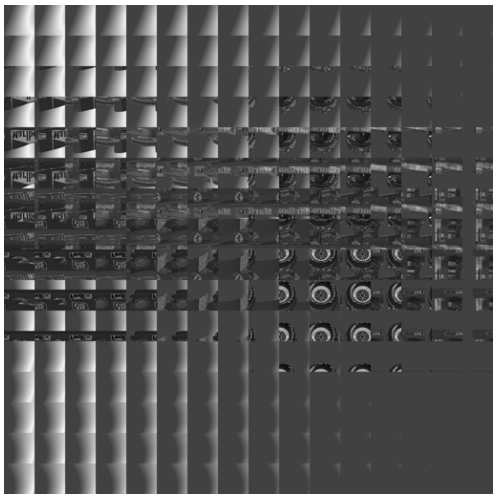
(a) (b) (c)

Σχήμα 11.32 : Η Γραφική Απεικόνιση του μετασχηματισμού Fourier των Τριών Εικόνων ανά Στοιχειώδη Εικόνα

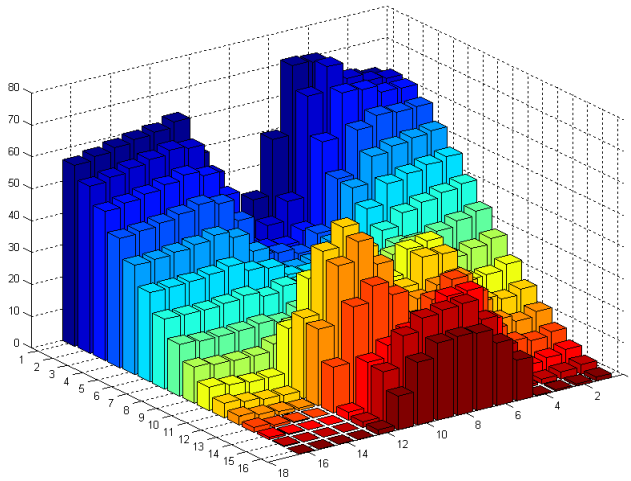
11.4 Προτεινόμενη Μέθοδος Βάσει των Συμπερασμάτων

Όπως αναφέρθηκε σε προηγούμενη ενότητα, οι επιδόσεις των κωδικοποιητών όσων αφορά τους λόγους συμπίεσης, είναι αντιστρόφως ανάλογοι του χρόνου εκτέλεσης τους. Συνεπώς σε εφαρμογές όπου ο χρόνος εκτέλεσης είναι η κρισιμότερη παράμετρος, είναι πιθανό, ο κωδικοποιητής JPEG να αποτελεί τη μόνη διαθέσιμη επιλογή. Βάσει των συμπερασμάτων της υποενότητας 11.2.4 θα προταθεί μία τροποποίηση του κωδικοποιητή JPEG, η οποία ενδέχεται να συμπιέσει αποδοτικότερα τις ολοκληρωτικές εικόνες όταν η στατιστική των εικόνων το επιτρέπει.

Ειδικότερα, παρατηρήσαμε πως η τυπική απόκλιση των τυπικών αποκλίσεων αποτελεί ένα μέγεθος το οποίο μπορεί να μας πληροφορήσει σχετικά με την ύπαρξη ή μη, έντονου φόντου. Εάν θυμηθούμε ξανά την ολοκληρωτική εικόνα της Ferrari θα διαπιστώσουμε πως πράγματι υπάρχει έντονη παρουσία φόντου ενώ το θέμα περιορίζεται στις κεντρικές στοιχειώδεις εικόνες.



(a)



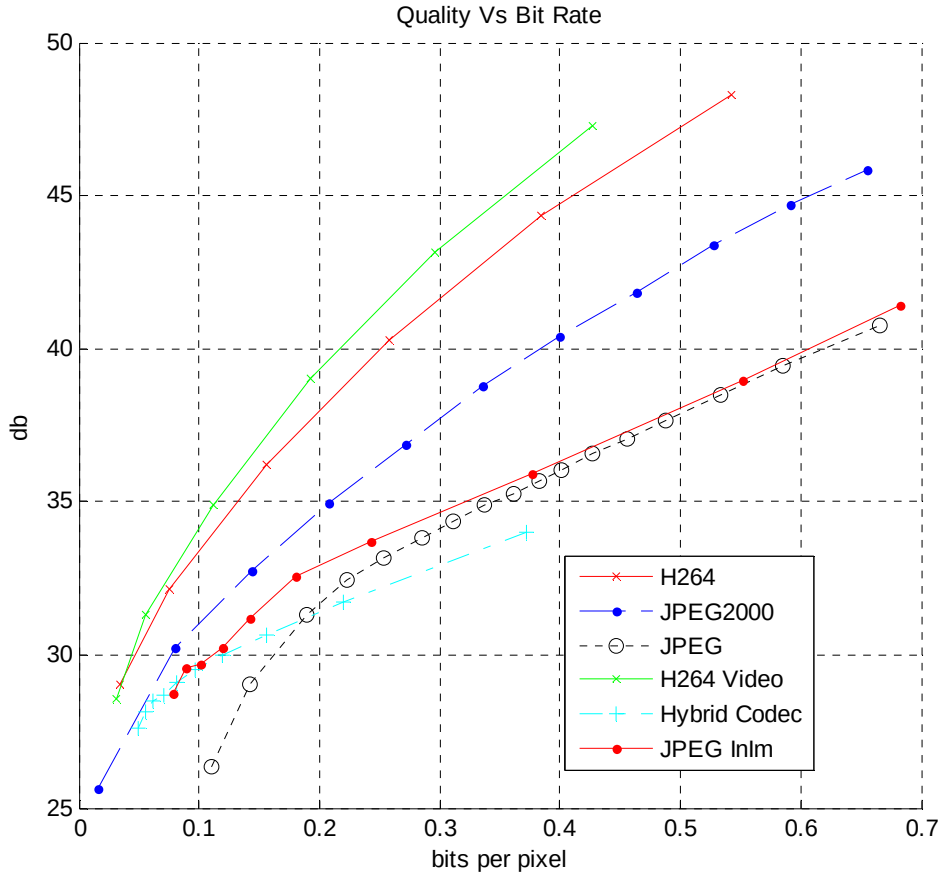
(b)

Σχήμα 11.33 : Η Ολοκληρωτική Εικόνα της Ferrari και η Γραφική Αναπαράσταση της Τυπικής Απόκλισης της Κάθε Στοιχειώδους Εικόνας με Μέση Τιμή Διασποράς 18.3

Προκειμένου να διαπιστωθεί κατά πόσο μία υλοποίηση που εκμεταλλεύεται τα στατιστικά μεγέθη της εικόνας μπορεί να αποδώσει καλύτερα στην περίπτωση των ολοκληρωτικών εικόνων, δημιουργήθηκε ένας συμπιεστής κατά το πρότυπο JPEG ο οποίος κωδικοποιεί ξεχωριστά κάθε στοιχειώδη εικόνα αφού αρχικά με ένα πέρασμα συγκεντρώσει τα στατιστικά στοιχεία τα οποία απαιτούνται. Στη συγκεκριμένη υλοποίηση το στατιστικό στοιχείο το οποίο εξετάζεται είναι η τυπική απόκλιση των τυπικών αποκλίσεων.

Ο κωδικοποιητής αφού συγκεντρώσει τις τυπικές αποκλίσεις στη συνέχεια υπολογίζει τη διασπορά τους. Εάν η τιμή αυτή υπερβαίνει την τιμή ενός κατωφλίου τότε ο κωδικοποιητής κβαντίζει ισχυρότερα τις στοιχειώδεις εικόνες με χαμηλή τυπική απόκλιση και λιγότερο έντονα τις στοιχειώδεις εικόνες με υψηλή τυπική απόκλιση. Το πλήθος επιλογών τιμών κβάντισης αποτελεί μία παράμετρο που καθορίζει ο χρήστης και ονομάζεται παράθυρο τιμών κβάντισης. Το παράθυρο μεγέθους 3 κρίθηκε κατά τις δοκιμές ικανοποιητικό και χρησιμοποιήθηκε για τις μετρήσεις.

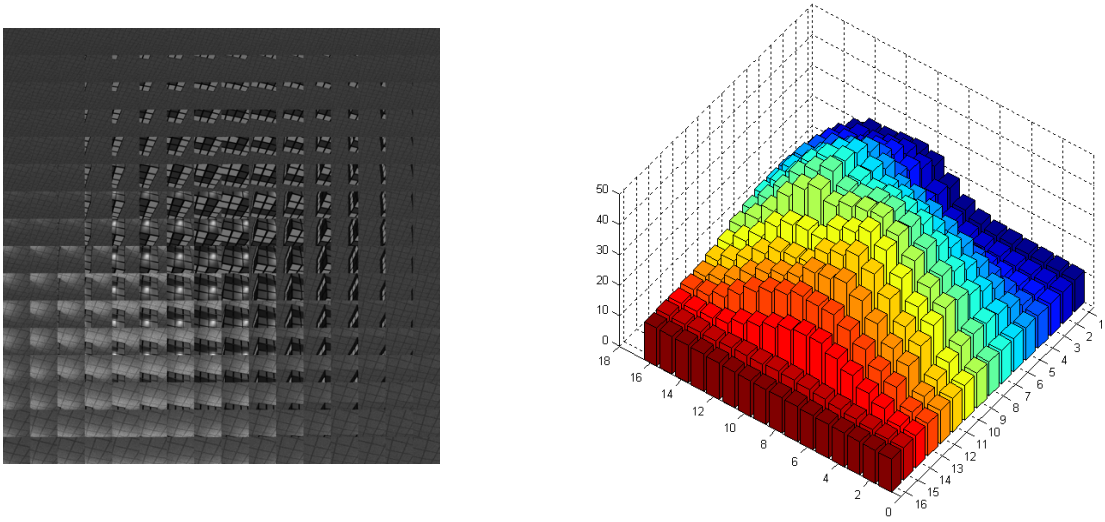
Στην περίπτωση της εικόνας της Ferrari το αποτέλεσμα του κωδικοποιητή παρουσιάζεται στο Σχήμα 11.34.



Σχήμα 11.34 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Ferrari σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Τυπικής Απόκλισης)

Όπως παρατηρούμε η προτεινόμενη τροποποίηση απέδωσε ικανοποιητικά καθώς στους χαμηλούς ρυθμούς όπου όμως η ποιότητα είναι ικανοποιητική, ο κωδικοποιητής ξεπέρασε τόσο το JPEG όσο και τον υβριδικό κωδικοποιητή. Αξίζει να σημειωθεί πως στα bpp συνυπολογίστηκε η κεφαλίδα του JPEG, προφανώς μία φορά και όχι για κάθε στοιχειώδη εικόνα που κωδικοποιήθηκε, ενώ η πρώτη DC συνιστώσα κάθε μίας στοιχειώδους εικόνας κωδικοποιήθηκε διαφορικά με την πρώτη DC συνιστώσα της αμέσως προηγούμενης στοιχειώδους εικόνας.

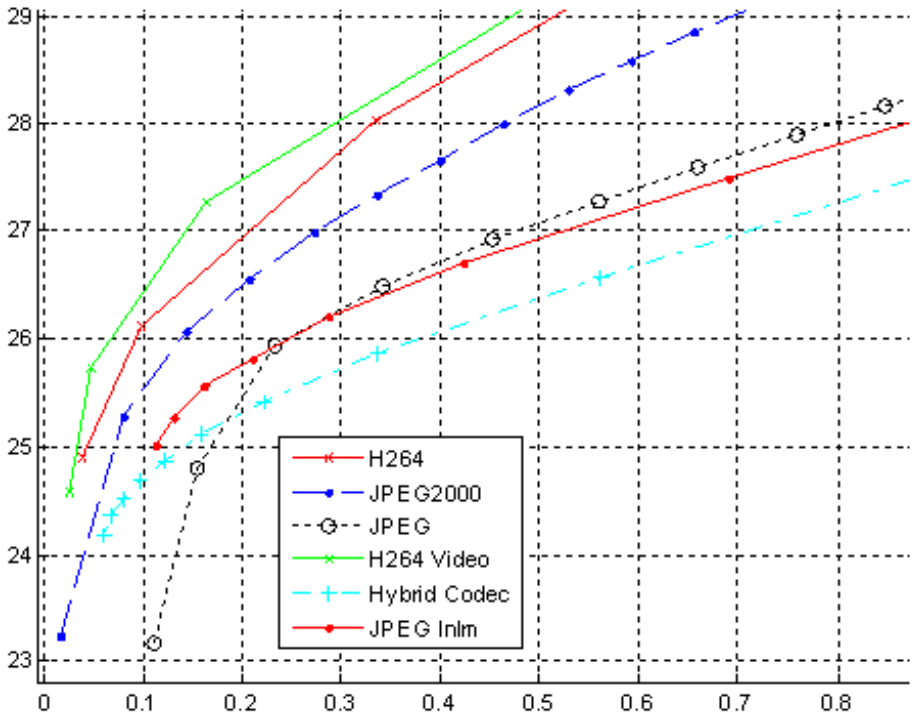
Αντίθετα, στην εικόνα Cube παρατηρείται ένας μικρός αριθμός στοιχειωδών εικόνων χωρίς θέμα. Ασφαλώς θα πρέπει να σημειωθεί πως στη συγκεκριμένη εικόνα, ακόμα και στοιχειώδεις εικόνες χωρίς θέμα παρουσιάζουν σχετικά υψηλή διασπορά. Αυτό οφείλεται στον έντονο φωτισμό και την ιδιόμορφη υφή της εικόνας (πλακίδια).



(a) (b)

Σχήμα 11.35 : Η Ολοκληρωτική Εικόνα Cube και η Γραφική Αναπαράσταση της Τυπικής Απόκλισης της Κάθε Στοιχειώδους Εικόνας με Μέση Τιμή Διασποράς 10.7

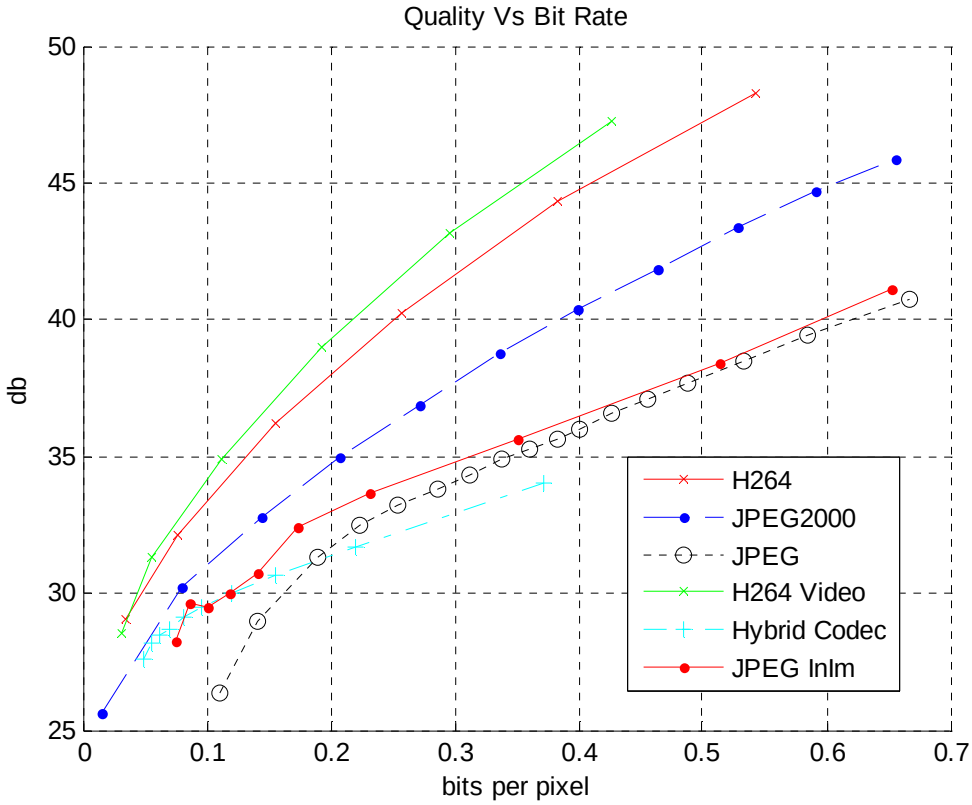
Όπως συμπεραίνουμε παρατηρώντας το Σχήμα 11.36, ο προτεινόμενος κωδικοποιητής αποδίδει χειρότερα από το JPEG, όχι όμως και για ρυθμούς μικρότερους των 0.2 bpp όπου και πάλι το τροποποιημένο JPEG υπερέχει.



Σχήμα 11.36 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Cube σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Τυπικής Απόκλισης)

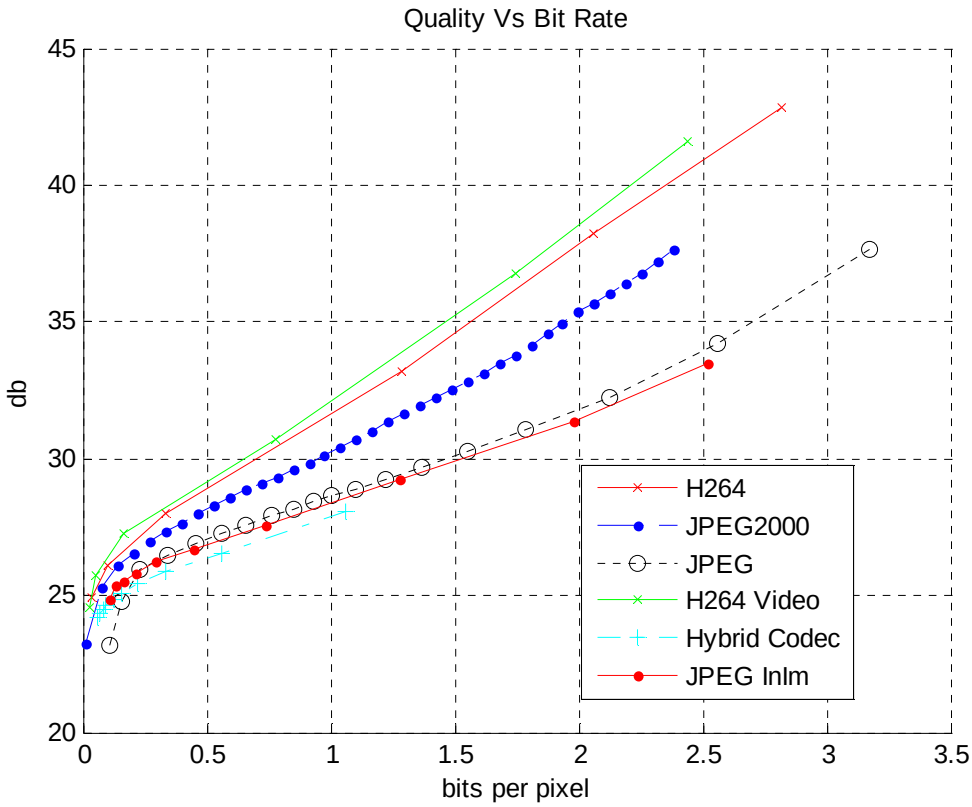
Ασφαλώς, στην εικόνα Cat η οποία δεν έχει καθόλου φόντο, καθώς κάθε στοιχειώδης εικόνα περιέχει το θέμα (γάτα), η μέθοδος δεν έχει καμία πρακτική αξία.

Στα συμπεράσματα του παρόντος κεφαλαίου αναδείχθηκε και η αξία της μετρικής Intra-MAD σχετικά με το κατά πόσο μία εικόνα μπορεί να κωδικοποιηθεί αποδοτικά. Προκειμένου να εξεταστεί εάν η συγκεκριμένη μετρική μπορεί να χρησιμοποιηθεί στη μέθοδο μας, αντικαταστάθηκε η χρήση της τιμής των τυπικών αποκλίσεων με την τιμή του μέσου απόλυτου σφάλματος. Το αποτέλεσμα στην εικόνα της Ferrari εμφανίζεται στο Σχήμα 11.37



Σχήμα 11.37 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Ferrari σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Intra-MAD)

Τα αποτελέσματα στην περίπτωση χρήσης του μέτρου Intra-MAD εμφανίζονται ελαφρώς καλύτερα σε σχέση με της μέσης τυπικής απόκλισης. Τη διαπίστωση αυτή βεβαιώνει και το Σχήμα 11.38 το οποίο παρουσιάζει την απόδοση του τροποποιημένου JPEG στην εικόνα Cube η οποία όμως και πάλι επιτυγχάνει χαμηλότερες επιδόσεις σε σχέση με το JPEG.



Σχήμα 11.38 : Η Απόδοση των Πέντε Κωδικοποιητών στην Εικόνα Cube σε Σχέση με τον Προτεινόμενο Κωδικοποιητή (Μέτρο Intra-MAD)

Ο συνδυασμός τόσο του μέτρου Intra-MAD όσο και του μέτρου της διασποράς των τυπικών αποκλίσεων ενδέχεται να οδηγήσει σε ακόμα καλύτερα αποτελέσματα. Όπως παρατηρήσαμε σημαντική παράμετρο αποτελεί η επιλογή της τιμής κατωφλίου, προκειμένου να χρησιμοποιηθεί ο τροποποιημένος κωδικοποιητής διαφορετικά θα υπάρξει μείωση της απόδοσης σε σχέση με το JPEG προς τους μεσαίους και υψηλούς ρυθμούς. Σε χαμηλούς ρυθμούς πάντοτε υπάρχει κέρδος ή έστω όμοια επίδοση με το JPEG.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	**Ελληνικός Όρος**
Integral Image	Ολοκληρωτική Εικόνα
Elemental Image	Στοιχειώδης Εικόνα
Transformation	Μετασχηματισμός
Pixel	Εικονοστοιχείο
Edges	Ακμές
Resolution	Ανάλυση
Stereoscopic Parallax	Στερεοσκοπική Παράλλαξη
Microlense	Μικρο-Φακός
Pinhole	Μικρή οπή
Focal Length	Εστιακή Απόσταση
Blurring	Θολερότητα
Image Reconstruction	Ανακατασκευή Εικόνας
Depth of Focus	Εστιακό Βάθος
Entropy	Εντροπία
Distortion	Παραμόρφωση
Quantization	Κβάντιση
Encoder	Κωδικοποιητής
Decoder	Αποκωδικοποιητής
Codeword	Κωδική Λέξη
Codebook	Κωδικό-Βιβλίο
Distribution	Κατανομή
Kernel	Πυρήνας
Wavelet	Κυματίδιο
Filter Banks	Τράπεζες Φίλτρων
Detail Coefficients	Συντελεστές Λεπτομέριας
Decomposition	Αποδόμηση
Sub-band	Υπο-ζώνη
Low-pass Filter	Βαθυπερατό Φίλτρο
High-pass Filter	Υψηπερατό Φίλτρο
Band-pass Filter	Ζωνοπερατό Φίλτρο
Multi-Resolution Analysis	Ανάλυση Πολλαπλής Διακριτικής Ικανότητας
Taps / Filter Coefficients	Συντελεστές Φίλτρου
Biorthogonal	Αμφιορθογώνιο
Normalization	Κανονικοποιήση
Truncation	Στρογγυλοποίηση προς τα κάτω
Tiles	Πλακίδια
Precincts	Τομείς
Layers	Επίπεδα
Video Frame	Καρέ
Motion Estimation	Εκτίμηση Κίνησης
Motion Compensation	Εξισορρόπηση Κίνησης
Video Sequence	Βίντεο-ροή
Grayscale Image	Εικόνα Αποχρώσεων του Γκρι
Redundancy	Πλεονασμός

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

II	Integral Image
EI	Elemental Image
CCD	Charge-Coupled Device
RLE	Run Length Encoding
JPEG	Joint Photographic Experts Group
DCT	Discrete Cosine Transform
IDCT	Inverse Discrete Cosine Transform
NBC	Natural Binary Code
VLC	Variable Length Codes
MSE	Mean Square Error
BPP	Bits Per Pixel
PSNR	Peak Signal to Noise Ratio
RMSE	Root Mean Square Error
PDF	Probability Density Function
DSAD	Digitally Sampled Analog Data
TSVQ	Tree-Structured Vector Quantization
GPP	Pool of Growing Points
LIFO	Last In First Out
FIFO	First In Last Out
AC	Alternate Current
DC	Direct Current
KLT	Karhunen-Loeve Transform
WHT	Walsh-Hadamard Transform
MPEG	Moving Pictures Experts Group
DFT	Discrete Fourier Transform
Codec	Encoder/Decoder
RGB	Red Green Blue
VLSI	Very Large Scale Integration
VLI	Variable Length Integers
FIR	Finite Impulse Response
QMF	Quadrature Mirror Filter
DWT	Discrete Wavelet Transform

ROI	Region Of Interest
MAD	Mean of Absolute Difference
SI	Sub-Images
RI	Ray-Space Image
Cos	Cosine
Sin	Sine

ΑΝΑΦΟΡΕΣ

[1] Javidi B, Okano F, eds (2002) Three Dimensional Television, Video, and Display Technologies, Springer Berlin.
[2] Arai J, Kawai H, Okano F (2007) Micro-lens arrays for integral imaging system. Appl Opt 45:9066–9078.
[3] Νικόλαος Σγούρος , "Συμβολή στην Ανάλυση και Κωδικοποίηση Συστοιχίας Εικόνων Τρισδιάστατης απεικόνισης", Διαδοκτορική διατριβή, ΕΚΠΑ, 2007.
[4] Jang J-S, Jin F, Javidi B (2003) Three-dimensional integral imaging with large depth-of-focus by use of real and virtual fields. Opt Lett 28:1421–1423.
[5] Jang J-S, Javidi B (2004) Depth and lateral size control of three-dimensional images in projection integral imaging. Opt Express 12:3778–3790.
[6] Jang J-S, Javidi B (2003) Large depth-of-focus time-multiplexed three dimensional integral imaging by use of lenslets with non-uniform focal lens and aperture sizes. Opt Lett 28:1925–1926.
[7] Lippmann G (1908) Epreuves reversibles. Photographies integrales. C R Acad Sci 146:446–451.
[8] Okano F, Hoshino H, Arai J, et al. (1997) Real-time pickup method for a three-dimensional image based on integral photography.Appl Opt 36:1958–1603.
[9] Shannon, C. E. A mathematical theory of communication, Bell Syst. Tech. J., 27, 379-423 (Part I), 1948; 623-656 (Part II), 1948.
[10] Berger, T. Rate Distortion Theory, Englewood Cliffs, NJ, Prentice-Hall, 1971.
[11] Capon, J. (1959) "A Probabilistic Model for Run-length Coding of Pictures," IEEE Transactions on Information Theory, 5(4):157–163, December.
[12] Golomb, Solomon W. (1966) "Run-Length Encodings," IEEE Transactions on Information Theory IT-12(3):399–401.
[13] Huffman, David (1952) "A Method for the Construction of Minimum Redundancy Codes," Proceedings of the IRE 40(9):1098–1101.
[14] Langdon, G. G., Jr. and J. Rissanen, Compression of black-white images with arithmetic coding, IEEE Trans. Commun., COM-29(6), 858-867, 1981.
[15] Langdon, G. G., Jr., An introduction to arithmetic coding, IBM J. Res. Dev., 28(2), 135-149, 1984.
[16] Gersho, A, Quantization, IEEE Commun. Mag., pp. 6-29, September, 1977.
[17] Li, W. and Y.-Q. Zhang, Vector-based signal processing and qunatization for image and video compression,Proc. IEEE, 83(2), 317-335, 1995.
[18] Max, J. Quantizing for minimum distortion, IRE Trans. Inf. Theory, it-6, 7-12, 1960.
[19] Lloyd, S. P. Least squares quantization in PCM, Inst. Mathematical Statistics Meet., Atlantic City, NJ, September 1957; IEEE Trans. Inf. Theory, pp. 129-136, March 1982.
[20] Paez, M. D. and T. H. Glisson, Minimum mean squared error qunatization in speech PCM and DPCM Systems, IEEE Trans. Commun., 225-230, April 1972.
[21] IEEE Trans. Commun., 225-230, April 1972.
[22] Smith, B. Instantaneous companding of quantized signals, Bell Syst. Tech. J., 36, 653-709, 1957
[23] Linde, Y., A. Buzo, and R. M. Gray (1980) "An Algorithm for Vector Quantization Design," IEEE Transactions on Communications, COM-28:84–95, January.
[24] Gersho, Allen, and Robert M. Gray (1992) Vector Quantization and Signal Compression, Boston, MA, Kluwer Academic Publishers.
[25] Nasrabadi, N. M. and R. A. King, Image Coding using Vector Quantization: A Review, IEEE Trans. Commun., COM-36(8), 957-971, 1988.
[26] Laroia, R. and N. Farvardin, A structured fixed rate vector quantizer derived from a variable length scaler quantizer: I & II, IEEE Trans. Inform. Theory, vol. 39, 851-876, 1993.
[27] Constantinescu, C., and J. A. Storer (1994b) "Improved Techniques for Single-Pass Adaptive Vector Quantization," Proceedings of the IEEE, 82(6)933–939, June.
[28] Strang, G., Introduction to Linear Algebra, Wellesley-Cambridge Press, 2nd ed., 1998.
[29] Barnsley, M. F., and Sloan, A. D. (1988) "A Better Way to Compress Images," Byte Magazine pp. 215–222 January.
[30] Mulcahy, Colm (1997) "Image Compression Using the Haar Wavelet Transform," Spelman College Science and Mathematics Journal, 1(1):22–31, April.
[31] Karhunen, K., On Linear Methods in Probability Theory, Ann. Acad. Sci. Fennicae, Ser. A137. (Translated into English.) The Rand Corp., Santa Monica, CA, 1960.
[32] Loeve, M., Fonctions Aléatoires de Second Ordre, in P. Levy, Processus Stochastique et Mouvement Brownien, Hermann, Paris.
[33] Hadamard, J. Resolution d'une question relative aux determinants, Bull. Sci. Math., Ser. 2, 17, Part I, 240-246,1893.
[34] Walsh, J. L. A closed set of normal orthogonal functions, Am. J. Math., 45(1), 5-24, 1923.
[35] Ahmed, N., T. Natarajan, and R. K. Rao (1974) "Discrete Cosine Transform," IEEE Transactions on

Computers C-23:90–93.
[36] Feig, E., and E. Linzer (1990) “Discrete Cosine Transform Algorithms for Image Data Compression,” in Proceedings Electronic Imaging ’90 East, pages 84–87, Boston , MA.
[37] Rao, K. R., and P. Yip (1990) Discrete Cosine Transform—Algorithms, Advantages, Applications, London, Academic Press.
[38] Pennebaker, William B., and Joan L. Mitchell (1992) JPEG Still Image Data Compression Standard, New York, Van Nostrand Reinhold.
[39] Wallace, Gregory K. (1991) “The JPEG Still Image Compression Standard,” Communications of the ACM 34(4):30–44, April.
[40] Ramstad, T. A., et al (1995) Subband Compression of Images: Principles and Examples, Amsterdam, Elsevier Science Publishers.
[41] Simoncelli, Eero P., and Edward. H. Adelson (1990) “Subband Transforms,” in John Woods, editor, Subband Coding, Boston, Kluwer Academic Press, 143–192.
[42] Vetterli, M., and J. Kovacevic (1995) Wavelets and Subband Coding, Englewood Cliffs, NJ, Prentice-Hall.
[43] Daubechies, Ingrid (1988) “Orthonormal Bases of Compactly Supported Wavelets,” Communications on Pure and Applied Mathematics, 41:909–996.
[44] DeVore, R., et al. (1992) “Image Compression Through Wavelet Transform Coding,” IEEE Transactions on Information Theory 38(2):719–746, March.
[45] Lewalle, Jacques (1995) “Tutorial on Continuous Wavelet Analysis of Experimental Data”.
[46] Mallat, Stephane (1989) “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 11(7):674–693, July.
[47] Mulcahy, Colm (1997) “Image Compression Using the Haar Wavelet Transform,” Spelman College Science and Mathematics Journal, 1(1):22–31, April.
[48] Strang, Gilbert, and Truong Nguyen (1996) Wavelets and Filter Banks, Wellesley, MA, Wellesley-Cambridge Press.
[49] Stromme, Oyvind, and Douglas R. McGregor (1997) “Comparison of Fidelity of Reproduction of Images After Lossy Compression Using Standard and Nonstandard Wavelet Decompositions,” in Proceedings of The First European Conference on Signal Analysis and Prediction (ECSAP 97), Prague, June.
[50] Taubman, David S. and Michael W. Marcellin (2002) JPEG 2000, Image Compression Fundamentals, Standards and Practice, Norwell, MA, Kluwer Academic.
[51] ISO/IEC (2000), International Standard IS 15444-1 “Information Technology - JPEG 2000 Image Coding System.” This is the FDC (final committee draft) version 1.0, 16 March 2000.
[52] N. P. Sgouros, A. G. Andreou, M.S. Sangriotis, P. G. Papageorgas, D. M. Maroulis, and N. G. Theofanous, “Compression of IP images for autostereo-scopic imaging applications”, in Proceedings of ISPA03, vol. 1, September 2003, pp.223-227.
[53] S. Yeom, A. Stern, and B. Javidi, “Compression of 3D color integral images”, Optics Express, vol. 12, no. 8, pp. 1632-1642, April 2004.
[54] R. Olsson, M. Sjostrom, and Y. Xu, “A combined pre-processing and H.264-compression scheme for 3D integral images” in Proceedings of ICIP2006. Atlanta (GA), USA:IEEE, October 2006, pp.513-516.
[55] R. Olsson, M. Sjostrom, and Y. Xu, “Evaluation of combined pre-processing and H.264-compression scheme for 3D integral images” in Proceedings of Electronic Imaging – VCIP. San Jose (CA), USA: IS&T/SPIE, January 2007.
[56] Nicholas Sgouros,Ioannis Kontaxakis, and Manolis Sangriotis, « Effect of different traversal schemes in integral image coding» , Optical Society of America, 2008
[57] E. Elharar, Adrian Stern, Ofer Hadar, Bahram Javidi, “A Hybrid Compression Method for Integral Images Using Discrete Wavelet Transform and Discrete Cosine Transform”, journal of display technology, vol. 3, no. 3, September 2007
[58] Salomon, D. Data Compression - The Complete Reference. New York: Springer-Verlag, 2004
[59] Yun Q. S., Huifang S. Image and Video Compression for Multimedia Engineering – Fundamentals, Algorithms and Standards. Florida: CRC Press, 2000
[60] Javidi B., Okano F., Jung-Young S. 3D Dimensional Imaging, Visualization and Display. New York: Springer, 2009.
[61] Ian E. G. Richardson. H.264 and MPEG-4 Video Compression. The Robert Gordon University, Aberdeen, UK: WILEY, 2003
[62] Δημήτριος Κ. Σιδεράκης, “Υβριδική Μέθοδος Κωδικοποίησης Ολοκληρωτικής Εικόνας”, Πτυχιακή Εργασία, ΕΚΠΑ, 2011.
[63] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala. Spatio-Angular Resolution Tradeoff in Integral Photography. Proceedings of Eurographics Symposium on Rendering, 2006.