



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ  
ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

## Συμπίεση Ψηφιακών Εικόνων με βάση τη Θεωρία των Fractals

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

*Βασίλειος Δ. Γερασόπουλος*

Επιβλέπουσα: Κωνσταντίνα Νικήτα  
*Αν. Καθηγήτρια ΕΜΠ*

Αθήνα, Μάιος 2005





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ  
ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

## Συμπύεση Ψηφιακών Εικόνων με βάση τη Θεωρία των Fractals

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

*Βασίλειος Δ. Γερασόπουλος*

**Επιβλέποντες:** Κωνσταντίνα Νικήτα  
*Αν. Καθηγήτρια ΕΜΠ*

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή τη 10<sup>η</sup> Μαΐου 2005

.....  
Κ. Νικήτα  
*Αν. Καθηγήτρια Ε.Μ.Π.*

.....  
Ν. Ουζούνογλου  
*Καθηγητής Ε.Μ.Π.*

.....  
Δ. Κουτσούρης  
*Καθηγητής Ε.Μ.Π.*

Αθήνα, Μάιος 2005

.....  
Βασίλειος Δ. Γερασόπουλος  
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Βασίλειος Δ. Γερασόπουλος  
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου







## Περίληψη

Αντικείμενο της εργασίας αυτής είναι η μελέτη της συμπίεσης εικόνων κλιμακίας γκριζου με μεθόδους βασισμένες στη θεωρία των fractal και ειδικότερα στη θεωρία των **Partitioned Iterated Function Systems (PIFS)**.

Αρχικά κάνουμε μια σύντομη αναφορά στην ψηφιακή επεξεργασία της εικόνας και υπενθυμίζουμε τις βασικές αρχές της συμπίεσης εικόνας.

Το δεύτερο Κεφάλαιο παρουσιάζει το βασικό μαθηματικό υπόβαθρο της θεωρίας των fractal και των Συστημάτων Επαναληπτικών Συναρτήσεων (**Iterated Function Systems, IFS**). Γίνεται η εισαγωγή στις έννοιες του μετασχηματισμού τύπου affine, των IFS και στην θεωρία των PIFS, η οποία αποτελεί τη βάση για την κατασκευή ενός συστήματος συμπίεσης με fractal. Στο τέλος του κεφαλαίου παρουσιάζεται ο βασικός fractal αλγόριθμος κωδικοποίησης και αποκωδικοποίησης σε μορφή ψευδοκώδικα.

Τα Κεφάλαια 3, 4, 5 και 6 που ακολουθούν, αφορούν πιο λεπτά ζητήματα που σχετίζονται με την τεχνική PIFS και τη βελτίωση της επίδοσής της.

Ειδικότερα, στο κεφάλαιο 3 μελετώνται πολλές πτυχές της διαμέρισης της εικόνας και αποδεικνύεται ότι σε κάποιες περιπτώσεις ο χωρισμός της εικόνας σε αλληλοεπικαλυπτόμενες περιοχές υπερτερεί της διαμέρισης.

Το Κεφάλαιο 4 διερευνά τις παραμέτρους μεταβολής της έντασης του γκρι (την αντίθεση και τη φωτεινότητα) που είναι συνάμα και οι συντελεστές του συστολικού μετασχηματισμού (στον οποίο στηρίζονται τα PIFS) και παρουσιάζει μεθόδους αποδοτικής αποθήκευσής τους χωρίς την απώλεια ποιότητας.

Στα Κεφάλαια 5 και 6, παρουσιάζονται οι μέθοδοι που έχουν αναπτυχθεί τα τελευταία χρόνια για την επιτάχυνση βελτίωση της ποιότητας της συμπίεσης με τη θεωρία των fractal.

Στο τέλος της εργασίας, κάνουμε έναν απολογισμό για την επίδοση όχι μόνο της τυπικής αλλά και όλων των fractal τεχνικών συμπίεσης προτείνονται στη διεθνή βιβλιογραφία.

### Λέξεις κλειδιά

Ψηφιακή επεξεργασία εικόνας, τεχνικές συμπίεσης, συστολικός μετασχηματισμός, μετασχηματισμός τύπου affine, διαμέριση εικόνας, αντίθεση, φωτεινότητα, gray scale, fractal, IFS, PIFS



## **Abstract**

The scope of this thesis is the use of fractals and **Partitioned Iterated Function Systems (PIFS)** in particular, for gray scale image compression.

In the beginning, we briefly examine digital image processing and re-visit the principles of digital image compression.

Chapter 2 presents a basic mathematical background for fractal and **Iterated Function Systems (IFS)**. Affine transformations, IFS and PIFS, which forms the basis for building a basic fractal compressor, are introduced along with related examples. At the end of the chapter we present the algorithm for the baseline fractal encoder and decoder.

Chapters 3, 4, 5 and 6 deal with the issues concerning the PIFS scheme and the improvement of its performance.

Chapter 3 studies many aspects of image partitioning and demonstrates that in some cases, there is an advantage in using image overlapped covering rather than partitioning.

Chapter 4 explores the color intensity parameters (the contrast scaling and luminance offset), which are also the coefficients of the contractive mapping on which PIFS is based. In this chapter the goal is to study the ways of efficient storing of these two parameters without any loss in image quality.

The methods for accelerating the process of the baseline fractal image encoder and improving its quality are presented in Chapters 5 and 6.

At the end of the thesis, we evaluate the performance of the conventional fractal compressor and of the techniques that have appeared in the literature, aiming at its improvement.

## **Key Words**

Digital image processing, gray scale images, fractal, encoder, decoder, contractive mapping, affine transformation, image compression, partitioning, overlapped covering, contrast scaling, luminance offset, IFS, PIFS

# ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>Εισαγωγή</b>	13
<b>Κεφάλαιο 1: Η επεξεργασία της εικόνας</b>	15
1.1 Η εικόνα ως κατανομή στο επίπεδο	15
1.2 Η εικόνα ως σύνολο διακριτών εικονοστοιχείων	18
1.3 Η Διανυσματική Κβαντοποίηση	22
1.3.1 Σφάλμα Κβαντοποίησης	22
1.3.2 Βέλτιστη συνάρτηση Κβαντοποίησης	23
1.3.3 Συνθήκη του Πλησιέστερου Γείτονα	24
1.3.4 Συνθήκη Κέντρου Μάζας	24
1.3.5 Η κατασκευή του βέλτιστου κωδικού βιβλίου	25
1.3.6 Η συμπίεση σύμφωνα με τη Διανυσματική Κβαντοποίηση	27
1.4 Η Διανυσματική Κβαντοποίηση με αφαίρεση μέσης τιμής	31
Βιβλιογραφία	35
<b>Κεφάλαιο 2: Η συμπίεση εικόνας με βάση τη θεωρία των fractal</b>	36
2.1 Ένα μονοδιάστατο παράδειγμα	37
2.2 Ο Συστολικός Μετασχηματισμός	38
2.3 Συστήματα Επαναληπτικής Εφαρμογής Συναρτήσεων	43
2.4 Συστήματα Επαναληπτικής Εφαρμογής Συναρτήσεων με Διαίρεση της εικόνας	48
Βιβλιογραφία	60
<b>Κεφάλαιο 3: Η διαμέριση της εικόνας</b>	61
3.1 Χωρισμός της εικόνας σε επικαλυπτόμενες περιοχές	62
3.2 Ιεραρχικές μέθοδοι διαμέρισης	66
3.2.1 Η διαμέριση με βάση το τετραεδρικό δένδρο	66
3.2.2 Οριζόντια-Κάθετη διαίρεση	72
3.3 Διαίρεση-και-Συγχώνευση	74
3.3.1 Τριγωνοποίηση Delaunay	74
3.3.2 Ευρετική Αναζήτηση	76
3.3.3 Ο Εξελικτικός Υπολογισμός	77
3.4 Διαμέριση και κωδικοποίηση με βάση τα χαρακτηριστικά της εικόνας	79
3.4.1 Επιλογή του χώρου Αναζήτησης	81
3.4.2 Επιλογή του τρόπου εκτίμησης της έντασης για τις τύπου D περιοχές	85
Βιβλιογραφία	87
<b>Κεφάλαιο 4: Φωτεινότητα και αντίθεση-Γενικότεροι μετασχηματισμοί έντασης</b>	88

4.1 Η συνιστώσα της φωτεινότητας	88
4.1.1 Διακριτοποίηση μέσης τιμής	91
4.1.2 Πίνακες Διακριτοποίησης	99
4.2 Διακριτοποίηση της αντίθεσης	101
4.2.1 Διακριτές τιμές αντίθεσης	105
4.3 Βέλτιστη δέσμευση ψηφίων για τη διακριτοποίηση των fractal παραμέτρων	105
4.4 Ο μετασχηματισμός των τετραγωνικών περιοχών	115
Βιβλιογραφία	121
<b>Κεφάλαιο 5: Τεχνικές μείωσης της πολυπλοκότητας</b>	122
5.1 Μία μέθοδος που βασίζεται στις προβολές για τον υπολογισμό του σφάλματος ελαχίστων τετραγώνων	123
5.2 Χαρακτηριστικά διανύσματα	126
5.3 Μέθοδοι ταξινόμησης	135
5.3.1 Η προσέγγιση του Jacquín	135
5.3.2 Ταξινόμηση με βάση την ένταση και τη διακύμανσή της	136
5.3.3 Ταξινόμηση με βάση το αρχέτυπο	140
5.3.4 Αμετάβλητες αποκλίσεις	141
5.4 Μέθοδοι ομαδοποίησης	142
5.4.1 Ο Αλγόριθμος της Επέκτασης	144
5.4.2 Η ομαδοποίηση Heckbert	146
5.4.3 Ευρετική μέθοδος ομαδοποίησης	149
5.4.4 Ομαδοποίηση με βάση την απόσταση	154
5.4.5 Ο Υβριδικός Αλγόριθμος των Saure-Hamzaoui	162
5.4.5.1 Ομοιόμορφη διαίρεση	162
5.4.5.2 Διαίρεση με βάση το τετραεδρικό δένδρο	167
5.5 Συναρτησιακές μέθοδοι	171
5.5.1 Ταχεία Κωδικοποίηση Βασισμένη σε Προσαρμοστική Αναζήτηση	174
5.5.2 Το κριτήριο των Lai, Lam και Siu	181
5.6 Μέθοδοι βασισμένες σε κατασκευη δένδρου	185
5.7 Προσεγγίσεις πολλαπλής ανάλυσης	187
5.8 Ταχεία αναζήτηση μέσω ταχείας συνέλιξης	189
5.9 Συμπύεση χωρίς αναζήτηση	191
Βιβλιογραφία	194
<b>Κεφάλαιο 6: Πιο Προχωρημένα Θέματα – Συμπεράσματα</b>	196
6.1 Ο Διακριτός Μετασχηματισμός Συνημιτόνου	196
6.1.1 Ο υπολογισμός του DCT των ισομετρικών μετασχηματισμών	196
6.1.2 Το σφάλμα ελαχίστων τετραγώνων στο πεδίο της συχνότητας	199
6.2 Απαλλαγμένη από επαναλήψεις αποκωδικοποίηση, βασισμένη	205

στον αποδοτικό σχεδιασμό του συνόλου των τύπου $D$ περιοχών	
6.2.1 Ο τρόπος απαλλαγής από τις επαναλήψεις	205
6.2.1.1 Κωδικοποίηση	206
6.2.1.2 Αποκωδικοποίηση	209
6.2.2 Αποδοτικός σχεδιασμός του χώρου των τύπου $D$ περιοχών	211
6.2.2.1 Σχεδιασμός με βάση τον αλγόριθμο LBG	211
6.2.2.2 Μέθοδος του μέσου όρου των τετραγωνικών περιοχών	212
6.3 Συμπεράσματα	216
Βιβλιογραφία	220
<b>Παραρτήματα</b>	221
1. Παράρτημα Α: Οι τυπικές ασπρόμαυρες δοκιμαστικές εικόνες	221
2. Παράρτημα Β: Ο fractal αλγόριθμος με τη μορφή ψευδοκώδικα	224
3. Παράρτημα Γ: Η απόδειξη της σχέσης (4.15)	230

## ΕΙΣΑΓΩΓΗ

Η ανάπτυξη της τεχνολογίας που συντελέσθηκε τα τελευταία χρόνια σε τομείς όπως η ηλεκτρονική, η πληροφορική και οι τηλεπικοινωνίες επέδρασε καθοριστικά στο χώρο της ιατρικής παρέχοντας νέες και διαρκώς εξελισσόμενες διαγνωστικές και θεραπευτικές μεθόδους. Εξέχουσα θέση στη διαγνωστική, κατέχουν τα ιατρικά απεικονιστικά συστήματα τα οποία έχουν εξελιχθεί σημαντικά τα τελευταία χρόνια ενώ παράλληλα αναπτύσσονται και νέα τα οποία παρέχουν διαφορετική αλλά εξίσου χρήσιμη διαγνωστική πληροφορία.

Από τα παραπάνω γίνεται σαφές, ότι στη σύγχρονη ιατρική διαθέτουμε μια πληθώρα διαγνωστικών προσεγγίσεων προκειμένου τελικά να εξαχθεί το καλύτερο δυνατό συμπέρασμα για την κατάσταση του ασθενούς. Αυτή η πληθώρα διαγνωστικών εργαλείων, δημιούργησε αναπόφευκτα ένα πρόβλημα στην αποθήκευση του συνεχώς αυξανόμενου όγκου ιατρικών δεδομένων και στην διατήρηση ενός αξιόπιστου και εύχρηστου ιατρικού ιστορικού για κάθε ασθενή. Φυσικά, οι κλασικές μέθοδοι αποθήκευσης είναι εντελώς ακατάλληλες καθώς προκαλούν συσσώρευση έντυπου υλικού το οποίο χρειάζεται κατάλληλα διαμορφωμένους χώρους (*αρχείο*) καταλαμβάνοντας έτσι ζωτικό χώρο από την ιατρική μονάδα.

Τα ηλεκτρονικά μέσα αποθήκευσης (*οπτικά, μαγνητικά, κτλ*) αποτελούν, εν μέρει, μια λύση στο παραπάνω πρόβλημα. Ακόμη και αυτά όμως οδηγούνται σταδιακά στο έπακρο των δυνατοτήτων τους. Η ανάγκη αυτή οδήγησε στην ανάπτυξη μεθόδων συμπίεσης των ιατρικών δεδομένων και ειδικότερα των ιατρικών εικόνων. Η συμπίεση συνίσταται στη μετατροπή κάποιου αρχείου (*εικόνα στην προκειμένη περίπτωση*), με την εφαρμογή κατάλληλων αλγορίθμων, σε ένα άλλο που καταλαμβάνει λιγότερη μνήμη. Ο βαθμός συμπίεσης είναι ο μέγιστος δυνατός υπό την προϋπόθεση να μην προκαλεί απώλεια χρήσιμης διαγνωστικής πληροφορίας. Αυτό σημαίνει ότι ενώ ο χώρος που καταλαμβάνει στη μνήμη το αρχείο μειώνεται, τα απαραίτητα δεδομένα για την επιτυχή επαναφορά του στο αρχικό μέγεθος (*αποσυμπίεση*), διατηρούνται. Οι μέθοδοι συμπίεσης προσφέρουν πολλά πλεονεκτήματα ειδικά για τις ιατρικές εικόνες που αποτελούν μεν το πλέον εύχρηστο και εξαιρετικά πολύτιμο διαγνωστικό εργαλείο αλλά καταλαμβάνουν μεγάλο χώρο στη μνήμη.

Οι μέθοδοι συμπίεσης έχουν εξαιρετικά σπουδαία συνεισφορά στην ανάπτυξη της τηλεϊατρικής. Συγκεκριμένα, προσφέρεται η δυνατότητα ανάπτυξης μια κεντρικής βάσης δεδομένων στην οποία θα έχουν πρόσβαση (*on line*) όλες οι ιατρικές μονάδες της χώρας. Έτσι μπορεί να συγκροτηθεί ένα εθνικό δίκτυο ιατρικής κάλυψης από το οποίο μπορεί ο κάθε γιατρός να αντλήσει πληροφορίες για κάποιον ασθενή σε έγκαιρη κατάσταση. Αυτό βέβαια δεν θα ήταν αδύνατο χωρίς τη συμπίεση των εικόνων και των δεδομένων γενικότερα, αλλά σίγουρα πολύ δυσκολότερο καθώς η διακίνηση πληροφορίας αυτής της τάξης μεγέθους θα δημιουργούσε πρόβλημα στο δίκτυο. Επιπλέον, με την αποστολή των εικόνων ασύρματα, δίνεται η

δυνατότητα σύγκλησης ιατρικών συμβουλίων και ανταλλαγής απόψεων χωρίς την φυσική παρουσία όλων των γιατρών στον ίδιο χώρο με τον ασθενή. Είναι σαφές ότι η προοπτική αυτή είναι εξαιρετικά σημαντική καθώς έτσι επιταχύνεται σημαντικά ο χρόνος διάγνωσης και απόφασης για την εφαρμοζόμενη θεραπεία. Το δίκτυο ιατρικής κάλυψης που αναφέρθηκε παραπάνω μπορεί να επεκταθεί με διακρατικές συμφωνίες τόσο σε ευρωπαϊκή όσο και σε παγκόσμια κλίμακα, συντελώντας έτσι στην καλύτερη αντιμετώπιση των περιστατικών καθώς οι γιατροί θα μπορούν εύκολα και γρήγορα να ανταλλάσσουν πληροφορίες και απόψεις. Τέλος, οι μέθοδοι συμπίεσης σε συνδυασμό με την ρομποτική θα μπορούν σύντομα να χρησιμοποιηθούν στην ανάπτυξη της τηλε – χειρουργικής.

Από τα παραπάνω γίνεται αντιληπτή η σπουδαιότητα και η αναγκαιότητα των μεθόδων συμπίεσης των ψηφιακών εικόνων στην ανάπτυξη της σύγχρονης ιατρικής. Επιπλέον είναι σαφές, ότι η χρησιμότητα τους δεν περιορίζεται μόνο στην επίλυση του προβλήματος αποθήκευσης, ενώ οι μελλοντικές εφαρμογές τους αναμένεται να ξεπεράσουν τις προβλέψεις μας.

---

# ΚΕΦΑΛΑΙΟ 1

---

## Η επεξεργασία της εικόνας

Στο πρώτο αυτό κεφάλαιο γίνεται μία εισαγωγή στην έννοια της εικόνας. Επίσης παρουσιάζονται δύο μέθοδοι συμπίεσης, μέσα από την περιγραφή των οποίων δίνονται οι βασικές έννοιες της συμπίεσης της ψηφιακής εικόνας. Παράλληλα, εισαγόμαστε στον τρόπο με τον οποίο λειτουργεί και η μέθοδος συμπίεσης με βάση τη θεωρία των fractal, η οποία παρουσιάζει σημαντικές ομοιότητες με τις δύο αυτές τεχνικές και αναλύεται στο αμέσως επόμενο κεφάλαιο.

### 1.1 Η εικόνα ως κατανομή στο επίπεδο

Μία εικόνα ορίζεται ως μία κατανομή της έντασης και των χρωμάτων σε ένα ορθογώνιο πλαίσιο<sup>1</sup>. Για ένα τέτοιο πλαίσιο  $\Delta$ , διαστάσεων  $W*H$  μονάδων εμβαδού, η εικόνα είναι στοιχείο μίας κατανομής  $D$  με πεδίο ορισμού το  $\Delta$ . Σύμφωνα με το μοντέλο αυτό, η πληροφορία που λαμβάνει ο άνθρωπος από την όραση ενός αντικειμένου δεν είναι η ίδια με την εικόνα που αντανακλά το φως, καθώς μεταξύ του αντικειμένου και του εγκεφάλου παρεμβάλλεται το όργανο λήψης του σήματος η συνάρτηση μεταφοράς του οποίου, μεταβάλλει την τιμή της έντασης του ανακλώμενου φωτός. Για παράδειγμα, έχει διαπιστωθεί ότι το ανθρώπινο μάτι δεν μετρά σημειακά τη διακύμανση της τιμής της έντασης του φωτός αλλά κατά περιοχές, χρησιμοποιώντας ένα φίλτρο που μοιάζει με κατανομή Gauss σε ένα μικρό παράθυρο, είναι δηλαδή της μορφής:

$$\phi(x) = \begin{cases} e^{-\frac{1}{1-|x|^2}}, & |x| < 1 \\ 0, & \text{αλλου} \end{cases} \quad (1.1)$$

Στην πραγματικότητα, η εικόνα  $P$  ενός επίπεδου αντικειμένου, είναι μια συνάρτηση  $P: \Delta \rightarrow [0,1]$  και έχει την τιμή

$$P(x,y) = \iint_{\Delta} f(\xi,\eta)\omega(\xi-x,\eta-y)d(\xi,\eta) \quad (1.2)$$

όπου  $f$  η κατανομή της έντασης του φωτός που αντανακλά.

Το  $0$  στο πεδίο τιμών της  $P$  αντιστοιχεί στο πιο «σκοτεινό» μαύρο και το  $1$  στο πιο φωτεινό «άσπρο» που μπορεί να αντιληφθεί ο άνθρωπος. Η συνάρτηση  $\omega(-\varepsilon,\varepsilon)*(-\varepsilon,\varepsilon) \rightarrow [0,1]$  χαρακτηρίζει το σύστημα που λαμβάνει το σήμα και το μεταφέρει στο σύστημα επεξεργασίας της πληροφορίας (π.χ ο ανθρώπινος εγκέφαλος), αποτελεί δηλαδή το φίλτρο ή αλλιώς τη συνάρτηση δειγματοληψίας μεταξύ πομπού και δέκτη. Κατά συνέπεια, η σχέση (1.2) δεν περιγράφει τίποτε άλλο από τη συνέλιξη της κατανομής της πραγματικής εικόνας  $f$  με το φίλτρο  $\omega$ . Στη συνέχεια της εργασίας θεωρείται ότι δεν υπάρχει διαφορά μεταξύ της πραγματικής εικόνας και αυτής που αντιλαμβανόμαστε.



Σχ.1.1: Ανεξάρτητο της ανάλυσης σύστημα

Για το λόγο αυτό, όταν η εικόνα μοντελοποιείται ως συνάρτηση στο επίπεδο, μπορεί να περιγραφεί με ένα ειδικό σύστημα συντεταγμένων. Όπως δείχνει το παρακάτω σχήμα το πλαίσιο της βρίσκεται στο επίπεδο της σελίδας και η τιμή της συνάρτησης  $P$  σε κάθε σημείο  $(x,y)$  του πλαισίου  $\Delta$  αντιστοιχεί



στις διαφορετικές τιμές της έντασης του φωτός (*brightness*). Ένα τέτοιο σύστημα συντεταγμένων, θεωρείται ανεξάρτητο της ανάλυσης της εικόνας.

Αφού η εικόνα θεωρείται συνάρτηση στο επίπεδο, είναι εύκολο να ποσοτικοποιηθεί η διαφορά μεταξύ δύο εικόνων  $P$  και  $Q$ , το πλαίσιο των οποίων έχει τις ίδιες διαστάσεις. Συνήθως χρησιμοποιείται για το σκοπό αυτό η ρίζα του μέσου τετραγωνικού σφάλματος ως απόσταση μεταξύ δύο εικόνων δηλαδή:

$$d(P, Q) = \|P - Q\|_2 = \left( \int_{\Delta} |P(x, y) - Q(x, y)|^2 dx dy \right)^{\frac{1}{2}} \quad (1.3)$$

Ένα άλλο μέγεθος που χρησιμοποιείται συχνά για τη μέτρηση της διαφοράς μεταξύ 2 εικόνων είναι και ο μέγιστος σηματοθορυβικός λόγος (**Peak Signal to Noise Ratio, PSNR**) που ορίζεται ως ακολούθως:

$$PSNR = 20 \cdot \log_{10} \left( \frac{b}{\|P - Q\|_2} \right) \quad (1.4)$$

όπου το  $b$  είναι η μέγιστη πιθανή τιμή του σήματος, η οποία για το συνεχές μοντέλο που αναφέρεται εδώ είναι 1 και 255 για μία ψηφιακή εικόνα.

## 1.2 Η εικόνα ως σύνολο διακριτών εικονοστοιχείων

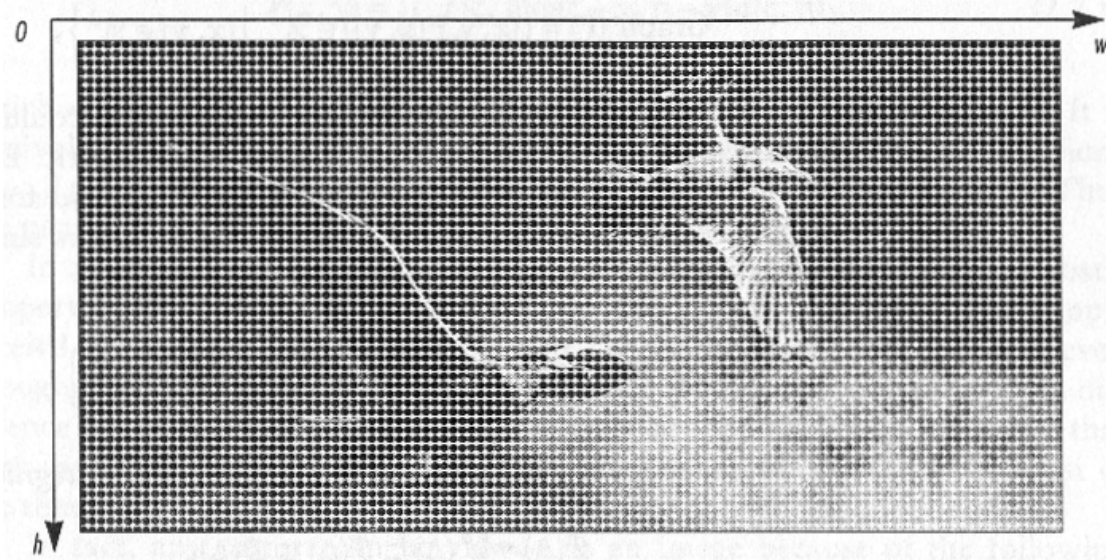
Στην απεικόνιση των αντικειμένων με χρήση υπολογιστών, οι εικόνες λαμβάνονται με τη χρήση των ψηφιακών μηχανών ή σαρωτών (scanner). Στη συνέχεια, αποθηκεύονται ως σύνολα με πεπερασμένο αριθμό στοιχείων (που το καθένα περιέχει τμήμα της εικόνας) ομοιόμορφα κατανεμημένων<sup>1</sup>. Στο σχήμα 1.2, η εικόνα  $P$  περιγράφεται από ένα δειγματοληπτικό πλέγμα από  $w$  στήλες και  $h$  γραμμές που είναι μια αναπαράσταση της εικόνας  $f$  με *τετράγωνες ψηφίδες* (*pixels*) σε ένα σύστημα συντεταγμένων που εξαρτάται από την *ανάλυση* (*resolution*) της εικόνας :

$$P : \{1, 2, \dots, w\} \times \{1, 2, \dots, h\} \rightarrow \{0, 1, \dots, 255\}^c \quad (1.5)$$

και σχηματίζεται από:

$$P = \{P[i, j]\}, 1 \leq i \leq w, 1 \leq j \leq h \quad (1.6)$$

$w * h$  ακεραίους όπου το μηδέν είναι το πιο σκοτεινό (μαύρο) και το 255 είναι το πιο φωτεινό (άσπρο) και το  $c$  είναι ο αριθμός των καναλιών που χρησιμοποιούνται για την απεικόνιση των χρωμάτων (συνήθως τρία, ένα για το κόκκινο, ένα για το πράσινο και ένα για το μπλε).



Σχ.1.2: Σύστημα που εξαρτάται από την ανάλυση

Όπως αναφέρθηκε στο προηγούμενο εδάφιο, η τιμή της έντασης της εικόνας για κάθε κανάλι στην τετραγωνική ψηφίδα που βρίσκεται στη θέση  $(i, j)$  της εικόνας  $P[i, j]$ , είναι στην πραγματικότητα η κατανομή της εικόνας  $f$  η οποία δειγματοληπτείται από μια δοκιμαστική συνάρτηση, που στην περίπτωση αυτή καλείται φίλτρο δειγματοληψίας  $\phi^c_{(i, j)}$  δηλαδή:

$$P^c[i, j] = \iint_{R^2} f(x, y) \phi^c_{(i, j)}(x, y) dx dy \quad (1.7)$$

Το σύνολο των δειγματοληπτικών αυτών τιμών στη θέση  $(i, j)$ :

$$P[i, j] = \{P^c[i, j]\} \quad (1.8)$$

καλείται *εικονοστοιχείο* ή *pixel*. Είναι προφανές ότι το σύνολο των δειγματοληπτικών φίλτρων  $\phi_{(i, j)} = \{\phi^c_{(i, j)}\}$  είναι απολύτως εξαρτώμενο από τη συσκευή λήψης της εικόνας. Στην εργασία αυτή εξετάζεται η περίπτωση της μονοχρωματικής εικόνας, δηλαδή  $C=1$ .

Είναι φυσικό, το μοντέλο που χρησιμοποιείται ως φίλτρο δειγματοληψίας να είναι ανεξάρτητο από τη θέση του *εικονοστοιχείου* καθώς εξαρτάται από τη συσκευή λήψης ή τα μάτια . Επομένως γράφοντας  $\phi = \phi_{(0,0)}$ ,

$$\phi_{(i,j)}(x,y) = \phi(x-i, y-j), \forall x,y,i,j \in \mathbb{R}^2 \quad (1.9)$$

Ως αποτέλεσμα, η σχέση (1.7) παίρνει ακριβώς τη μορφή συνέλιξης μεταξύ της συνάρτησης κατανομής  $f$  και της συνάρτησης δειγματοληψίας  $\phi$ .

Συχνά η συνάρτηση δειγματοληψίας  $\phi$  είναι συμμετρική, δηλαδή:

$$\phi(x,y) = \phi(-x,y) = \phi(x,-y) = \phi(-x,-y), \forall x,y \in \mathbb{R}^2 \quad (1.10)$$

Για την προσαρμογή των τιμών του συνελικτικού ολοκληρώματος στο πεδίο  $[0, 255]$ , μπορεί να χρησιμοποιηθεί ένα παράγοντας κανονικοποίησης

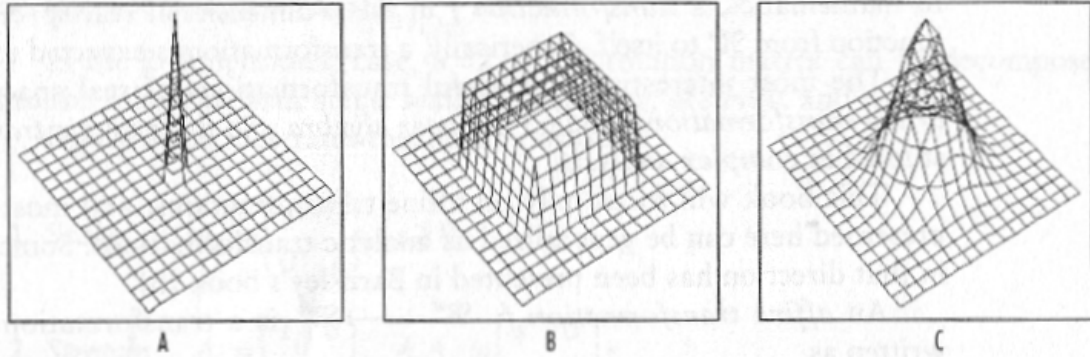
$$\iint_{\mathbb{R}^2} \phi(x,y) dx dy = 1 \quad \text{ή} \quad 255 \quad (1.11)$$

Συμπερασματικά, το δειγματοληπτικό φίλτρο χαρακτηρίζει πλήρως τη διακριτοποίηση των τιμών της κατανομής της φωτεινής έντασης μιας εικόνας από μία κάμερα ή ένα σαρωτή.

Στην ακραία περίπτωση όπου η  $\phi(x,y)$  είναι σταθερή, μπορεί, με τη βοήθεια της μέσης τιμής, να εκφρασθεί με τον ακόλουθο τρόπο:

$$\phi(x,y) = \begin{cases} \frac{1}{\varepsilon^2} & \forall -\frac{\varepsilon}{2} < x,y < \frac{\varepsilon}{2} \\ 0 & \text{αλλού} \end{cases} \quad (1.12)$$

Ένα τέτοιο φίλτρο καλείται φίλτρο τετραγωνικού μέσου πλάτους  $\varepsilon$ . Το όριο του φίλτρου αυτού καθώς το  $\varepsilon$  τείνει στο μηδέν είναι η συνάρτηση *Dirac*. Μπορούμε φυσικά να διαλέξουμε ένα πιο πολύπλοκο και ίσως πιο ακριβές φίλτρο όπως ένα φίλτρο τύπου Gauss  $\phi(x,y) = ae^{-\beta(x^2+y^2)}$ , για κάποιες σταθερές  $a$  και  $\beta$ . Όλα αυτά τα φίλτρα φαίνονται στο σχήμα 1.3.



Σχ.1.3: Ψηφιακά φίλτρα: A) Dirac, B) Μέσης τμής, C) Γκαουσιανό

Σε ένα σύστημα *εικονοστοιχείων* εξαρτώμενο από την ανάλυση της εικόνας, η απόσταση μεταξύ των δύο εικόνων  $P$  και  $Q$  έχει τη μορφή

$$d(P, Q) = \|P - Q\|_2 = \left( \sum_{j=1}^w \sum_{i=1}^w |P[i, j] - Q[i, j]|^2 \right)^{1/2} \quad (1.13)$$

(Ως συνάρτηση απόστασης έχει χρησιμοποιηθεί η Ευκλείδεια νόρμα).

Εανά ο μέγιστος σηματοθορυβικός λόγος (*Peak Signal to Noise Ratio*) που χρησιμοποιείται για να μετρήσει τη διαφορά των δύο εικόνων ορίζεται ως

$$PSNR = 20 \log_{10} \left( \frac{255}{\|P - Q\|_2} \right) \quad (1.14)$$

όπου  $b=255$  είναι η μέγιστη πιθανή τιμή του ψηφιακού σήματος .

Με βάση όσα προαναφέρθηκαν για την ψηφιακή εικόνα είναι φανερό ότι μια μονοχρωματική εικόνα διαστάσεων  $512*512$  εικονοστοιχείων, για την οποία έχουν δεσμευτεί 8 δυαδικά ψηφία για την απεικόνιση των διαβαθμίσεων του γκρι, καταλαμβάνει στη μνήμη χώρο ίσο με  $256K$ . Τέτοιου είδους εικόνες που μοιάζουν με αυτές που παίρνουμε από τα σύγχρονα ιατρικά όργανα, όπως αναφέρθηκε και στην Εισαγωγή, πρέπει να αποστέλλονται από τη βάση δεδομένων ενός νοσοκομείου στην αντίστοιχη ενός δεύτερου για την εξυπηρέτηση της τηλεϊατρικής. Η αποστολή των δεδομένων πρέπει να γίνεται σε όσο το δυνατό μικρότερο χρονικό διάστημα ώστε να επιταχύνεται η διάγνωση, γεγονός που δυσχεραίνεται από το μέγεθος των εικόνων στη μνήμη, το οποίο είναι ανάλογο του πλήθους ιατρικών εικόνων που πρέπει να αποσταλεί

στη μονάδα του χρόνου ώστε να έχουμε την καλύτερη δυνατή διάγνωση. Για το λόγο αυτό έχουν αναπτυχθεί διάφορες τεχνικές συμπίεσης των εικόνων, όπως και η συμπίεση με βάση τη θεωρία των *fractal*, η οποία αποτελεί και το αντικείμενο της εργασίας.

Υπάρχουν πολλές προοπτικές μέσα από τις οποίες μπορούμε να προσεγγίσουμε τη μέθοδο των *fractals*, όπως τα *Συστήματα Επαναληπτικών Συναρτήσεων (Iterated Function Systems)*, η *Μέθοδος Κωδικοποίησης με βάση τη Συνέλιξη (Convolution Coding)*, η *Διανυσματική Κβαντοποίηση (Vector Quantization)* κ.ά. Στην εργασία, θα δούμε τη *fractal* κωδικοποίηση από τη σκοπιά της Διανυσματικής Κβαντοποίησης όχι μόνο εξαιτίας των ομοιοτήτων μεταξύ τους, αλλά και επειδή η τελευταία αναφέρεται από τη φύση της στο διακριτό και κατά επέκταση ψηφιακό χώρο, γεγονός που την καθιστά απλούστερη και ιδανική για την κατανόηση της συμπίεσης με βάση τα *fractals*.

### 1.3 Διανυσματική Κβαντοποίηση

Το αντικείμενο της **ΔΚ (VQ)**<sup>2</sup> είναι η αναπαράσταση ενός συνόλου χαρακτηριστικών διανυσμάτων  $\mathbf{x} \in X$  (όπου  $X \subseteq \mathbb{R}^k$ ) από ένα σύνολο  $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_{N_c}\}$ , από  $N_c$  διανύσματα που ανήκουν στο  $\mathbb{R}^k$  και ονομάζονται *διανύσματα αναφοράς*. Το σύνολο  $Y$  ονομάζεται *κωδικό βιβλίο (codebook)* και τα στοιχεία του *κωδικές λέξεις* ή αλλιώς *κωδικά διανύσματα (code vectors)*. Τα διανύσματα του συνόλου  $X$  ονομάζονται *πρότυπα εισόδου (input pattern)* ή *διανύσματα εισόδου (input vectors)*. Άρα μια διανυσματική κβαντοποίηση μπορεί να παρασταθεί με τη μορφή συνάρτησης:  $q: X \rightarrow Y$ . Η γνώση της συνάρτησης  $q$  επιτρέπει τη διαίρεση του συνόλου  $X$  σε  $N_c$  υποσύνολα  $S_i$  (που καλούνται *κύτταρα*) και ορίζονται ως εξής:

$$S_i = \{\mathbf{x} \in X : q(\mathbf{x}) = \mathbf{y}_i\}, \quad i = 1, \dots, N_c \quad (1.15)$$

Το σύνολο  $S = \{S_1, S_2, \dots, S_{N_c}\}$  καλείται και *διαμέριση (partition)* του  $X$ .

Από εδώ και στη συνέχεια της ανάπτυξης της μεθόδου της Διανυσματικής Κβαντοποίησης θα θεωρείται ότι το σύνολο  $X$  των διανυσμάτων εισόδου είναι πεπερασμένου αριθμού στοιχείων, έστω  $N_p$  τον αριθμό.

### 1.3.1 Σφάλμα Κβαντοποίησης

Το **Σφάλμα Κβαντοποίησης** (*Quantization Error*), είναι η τιμή της παράστασης  $d(x, q(x))$  όπου  $d(\cdot)$  είναι ένας γενικός τελεστής απόστασης για διανύσματα. Το Μέσο ΣΚ (*Mean Quantization Error*) είναι ίσο με:

$$(MQE) \triangleq D(\{Y, S\}) = \frac{1}{N_p} \sum_{p=1}^{N_p} d(x_p, q(x_p)) = \frac{1}{N_p} \sum_{i=1}^{N_c} D_i \quad (1.16)$$

όπου  $D_i$  είναι το συνολικό σφάλμα για το  $i$ -οστό κύτταρο.

$$D_i = \sum_{n: x_n \in S_i} d(x_n, y_i) \quad (1.17)$$

Αρκετές συναρτήσεις μπορούν να χρησιμοποιηθούν και για την έκφραση του σφάλματος (Linde et al., 1980<sup>3</sup>) Αυτή που χρησιμοποιείται συνήθως είναι η τετραγωνική Ευκλείδεια απόσταση (**Τετραγωνικό Σφάλμα**, **ΤΣ**)

$$d(x, x') = \sum_{i=1}^k (x_i - x'_i)^2 \quad (1.18)$$

που χρησιμοποιείται και εδώ. Στην περίπτωση αυτή, το μέσο σφάλμα Κβαντοποίησης καλείται **Μέσο Τετραγωνικό Σφάλμα** (*Mean Square Error*). Μπορεί επίσης να χρησιμοποιηθεί και η τετραγωνική ρίζα του ΜΤΣ ως μέτρο σφάλματος και το **Κανονικοποιημένο ΜΤΣ** (*Normalized Mean Square Error*). Αυτό δεν είναι τίποτε άλλο από το ΜΤΣ διαιρεμένο από το ΜΤΣ που προκύπτει χρησιμοποιώντας ένα κωδικό βιβλίο που περιέχει μια μόνο κωδική λέξη  $c$ , τοποθετημένη στο «κέντρο μάζας» του συνόλου των δεδομένων εισόδου (Ο ορισμός του «κέντρου μάζας» δίνεται σε επόμενο εδάφιο). Δηλαδή το Κανονικοποιημένο ΜΤΣ δίνεται από την ακόλουθη σχέση:

$$NMSE = \frac{MSE}{\frac{1}{N_p} \sum_{p=1}^{N_p} d(x_p - c)} \quad (1.19)$$

### 1.3.2 Βέλτιστη συνάρτηση Κβαντοποίησης

Μία συνάρτηση κβαντοποίησης καλείται βέλτιστη αν για κάθε διαφορετική από αυτήν συνάρτηση με τον ίδιο αριθμό κωδικών λέξεων, το σφάλμα διακριτοποίησης είναι μεγαλύτερο από αυτό που αντιστοιχεί στην υπό εξέταση συνάρτηση. Με μαθηματικούς όρους, μπορεί να γραφεί ότι η συνάρτηση κβαντοποίησης  $q^*$  είναι βέλτιστη αν για κάθε  $q \neq q^*$ , έχουμε  $D(q^*) \leq D(q)$ .

Στη συνέχεια περιγράφονται οι δύο κύριες συνθήκες οι οποίες, από μαθηματική άποψη είναι αναγκαίες ώστε μία τέτοια συνάρτηση να θεωρηθεί βέλτιστη. Οι δυο αυτές συνθήκες καλούνται συνήθως *Συνθήκη του Πλησιέστερου Γείτονα*, (**ΣΠΓ**) (*Nearest Neighbor Condition (NNC)*) και *Συνθήκη Κέντρου Μάζας* (**ΣΚΜ**) (*Centroid Condition (CC)*).

### 1.3.3. Συνθήκη του Πλησιέστερου Γείτονα

Με δεδομένο ένα σταθερό κωδικό βιβλίο  $Y$ , η συνθήκη αυτή συνίσταται στην ανάθεση αντιστοίχισης του κάθε διανύσματος εισόδου με την πιο κοντινή σε αυτό κωδική λέξη. Διαιρείται λοιπόν το σύνολο των δεδομένων με τον ακόλουθο τρόπο:

$$S_i = \{x \in X : d(x, y_i) \leq d(x, y_j), j = 1, \dots, N_c, j \neq i\}, i = 1, \dots, N_c \quad (1.20)$$

Τα σύνολα  $S_i$  που μόλις ορίστηκαν, σχηματίζουν μια διαμέριση του συνόλου των δεδομένων. Αυτή ονομάζεται "*Voronoi Partition*" (Gersho & Gray, 1992<sup>4</sup>) και συμβολίζεται με  $P(y) = \{S_1, \dots, S_{N_c}\}$ . Αφού η  $P(y)$  είναι διαίρεση, δηλαδή  $S_1 \cap S_2 \cap \dots \cap S_{N_c} = \emptyset$ , ένα διάνυσμα εισόδου που έχει την ίδια απόσταση από δύο ή περισσότερες κωδικές λέξεις ανατίθεται (αντιστοιχίζεται) μονοσήμαντα σε μία από αυτές, δηλαδή περιλαμβάνεται σε ένα μόνο από τα σύνολα  $S_i$ .

Η ΣΠΓ (NNC) μας επιτρέπει το σχηματισμό μιας βέλτιστης διαίρεσης (Linde et al, 1980<sup>3</sup>) δηλαδή για κάθε διαίρεση  $S$  διαφορετική από την  $P(y)$ , ισχύει:

$$D(\{Y, S\}) \geq D(\{Y, P(Y)\}) \quad (1.21)$$

κάτι που είναι προφανές από τον τρόπο με τον οποίο σχηματίζεται το σύνολο  $P(Y)$ , του οποίου τα υποσύνολα περιέχουν τα διανύσματα εισόδου που προσεγγίζουν περισσότερο (μικρότερη απόσταση) την εκάστοτε κωδική λέξη.

### 1.3.4 Συνθήκη Κέντρου Μάζας

Με δεδομένη τη διαίρεση  $S$ , η ΣΚΜ ασχολείται με τη διαδικασία εύρεσης του βέλτιστου κωδικού βιβλίου, δηλαδή την εύρεση των  $y_1^*, y_2^*, \dots, y_{N_c}^*$ . Το κέντρο μάζας για ένα δεδομένο σύνολο  $A \subseteq \mathbb{R}^k$  είναι το διάνυσμα  $\underline{x}(A)$  για το οποίο ισχύει:

$$E\{d(x, \underline{x}(A)) \mid x \in A\} = \min_{u \in \mathbb{R}^k} E\{d(x, u) \mid x \in A\} \quad (1.22)$$

όπου  $E$  ένα σύνολο τα στοιχεία του οποίου είναι οι αποστάσεις των διανυσμάτων του  $A$  από το κέντρο μάζας. Αν ο αριθμός των στοιχείων του  $A$  είναι  $N_A$  και χρησιμοποιήσουμε την τετραγωνική Ευκλείδεια απόσταση τότε εύκολα αποδεικνύεται ότι:

$$\underline{x}(A) = \frac{1}{N_A} \sum_{x \in A} x \quad (1.23)$$

Με βάση τα παραπάνω, αν επιλεγθεί ως κωδικό βιβλίο  $\underline{X}(S)$  το σύνολο που οι κωδικές του λέξεις είναι τα κέντρα μάζας όλων των κυττάρων  $S_i$  του  $S$  δηλαδή αν

$$\underline{X}(S) = \{\underline{x}(S_i), i = 1, \dots, N_c\} \quad (1.24)$$

τότε το  $\underline{X}(S)$  είναι βέλτιστο καθώς με βάση τις (1.17), (1.22) και (1.24) για κάθε κωδικό βιβλίο  $Y \neq S$  ισχύει:



$$D\{Y, S\} \geq D\{(X(S), S)\} \quad (1.25)$$

### 1.3.5 Η κατασκευή του βέλτιστου κωδικού βιβλίου

Στη συνέχεια εξετάζεται ο τρόπος με τον οποίο από ένα δεδομένο σύνολο  $N_p$  διανυσμάτων εισόδου, που καλούνται και *εκπαιδευτικά διανύσματα* (*training vectors*) και ένα δεδομένο κωδικό βιβλίο  $Y_m = \{y^m_1, \dots, y^m_{N_c}\}$ , εξάγεται ένα βελτιωμένο κωδικό βιβλίο  $Y_{m+1}$ . Η διαδικασία αυτή είναι γνωστή και ως *Γενικευμένη Επαναληπτική Διαδικασία* του *Lloyd*<sup>3</sup>. Εντασσόμενος στην παραπάνω διαδικασία είναι και ο αλγόριθμος **LBG**, ο οποίος ονομάστηκε από τους ερευνητές που τον εμπνεύστηκαν (**Linde**, **Buzo** & **Grey**)<sup>3</sup>. Για την παρουσίασή του θα χρησιμοποιήσουμε τα εξής σύμβολα:

- $m$ : ο αριθμός των επαναλήψεων
- $Y_m$ : το  $m$ -οστό κωδικό βιβλίο
- $D_m$ : το Μέσο Σφάλμα Διακριτοποίησης που μας δίνει η  $m$ -οστή επανάληψη

Τα βήματα της διαδικασίας είναι τα ακόλουθα:

#### 1. Αρχικοποίηση. Οι ακόλουθες τιμές είναι σταθερές:

- $N_c$ : ο αριθμός των κωδικιών λέξεων
- $\epsilon \geq 0$ : η ακρίβεια της διαδικασίας βελτιστοποίησης
- $Y_0$ : το αρχικό κωδικό βιβλίο
- $X = (x_j : j = 1, \dots, N_p)$ : το σύνολο των διανυσμάτων εισόδου.

Επί πλέον γίνονται οι ακόλουθες αρχικοποιήσεις:

- $m = 0$
- $D_{-1} = +\infty$

#### 2. Υπολογισμός της διαίρεσης

Με δεδομένο το αρχικό κωδικό βιβλίο  $Y_m$  υπολογίζουμε τη διαίρεση  $P(Y_m)$  του  $X$  με βάση τη ΣΠΓ (NNC)

#### 3. Έλεγχος συνθήκης τερματισμού

Η παραμόρφωση του κβαντιστή  $D_m = D(\{Y_m, P(Y_m)\})$  υπολογίζεται με βάση τη σχέση (1.17). Αν

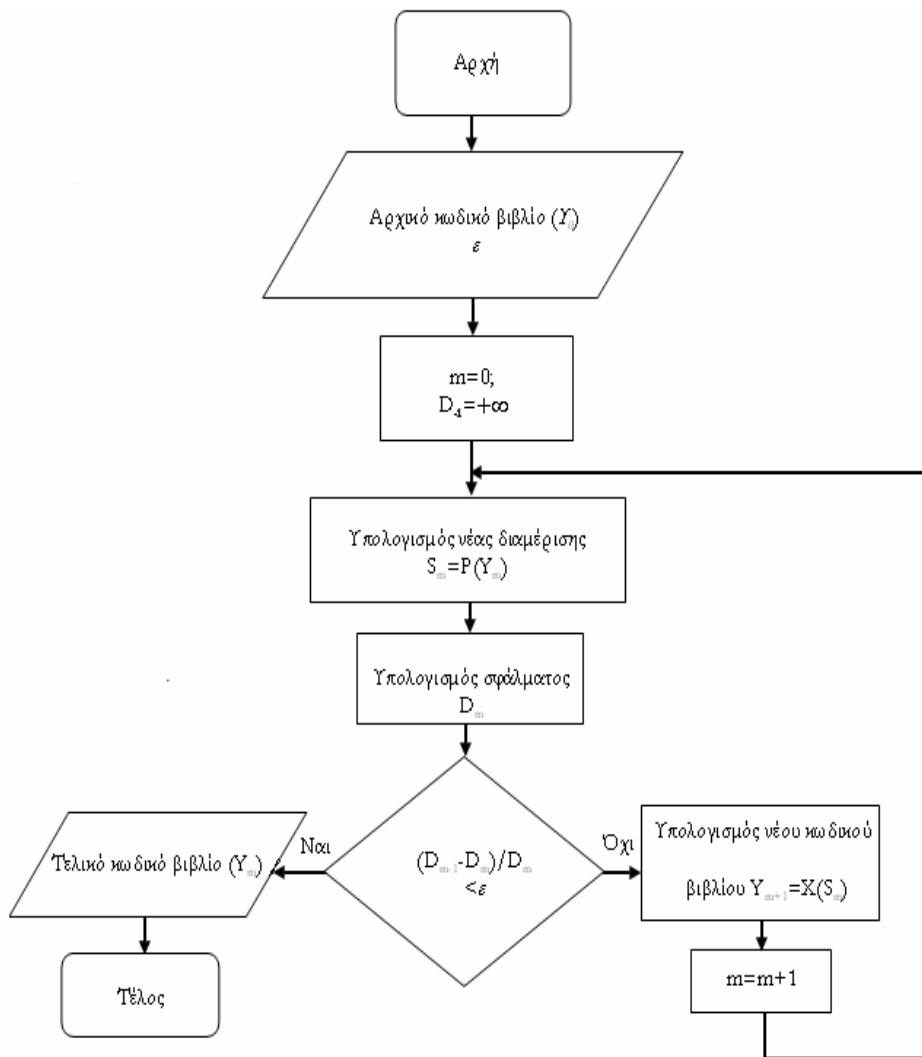
$$\frac{|D_{m-1} - D_m|}{D_m} < \varepsilon \quad (1.26)$$

τότε η διαδικασία τερματίζεται και το  $Y_m$  είναι το βέλτιστο κωδικό βιβλίο.

#### 4. Υπολογισμός του νέου κωδικού βιβλίου:

Με δεδομένη τη διαίρεση  $P(Y_m)$ , το νέο κωδικό βιβλίο υπολογίζεται με βάση τη συνθήκη του Κέντρου Μάζας. Δηλαδή έχουμε :

$$Y_{m+1} = X(P(Y_m)) \quad (1.27)$$



Σχ.1.4 : Η κατασκευή του βέλτιστου κωδικού βιβλίου με βάση τον αλγόριθμο LBG

Στη συνέχεια, η τιμή του μετρητή  $m$  αυξάνει κατά ένα και η διαδικασία συνεχίζεται από το βήμα 2. Αποδεικνύεται (Linde et al,1980),<sup>3</sup> ότι τα βήματα αυτά εξασφαλίζουν τη δημιουργία μιας φθίνουσας και συγκλίνουσας ακολουθίας σφαλμάτων  $D_m$ . Στο σχήμα 1.4 δίνονται τα βήματα του αλγορίθμου που μόλις περιγράφηκε.

### 1.3.6 Η συμπίεση σύμφωνα με τη Διανυσματική Κβαντοποίηση

Η παραπάνω μεθοδολογία είναι απλή στην εφαρμογή της για την κωδικοποίηση μονοχρωματικών εικόνων. Αν μία εικόνα διαιρεθεί σε μη επικαλυπτόμενες τετράγωνες περιοχές σταθερής διάστασης π.χ.  $4*4$  *εικονοστοιχείων*, τότε έχουμε δημιουργήσει διανύσματα διάστασης  $n=16$ , προσπελάζοντας για παράδειγμα τις τετράγωνες αυτές περιοχές κατά σειρές και αντιστοιχώντας την τιμή της έντασης κάθε εικονοστοιχείου που συναντούμε, στην εκάστοτε συνιστώσα του διανύσματος. Τέτοιου είδους διανύσματα μπορούμε να πάρουμε από διάφορες μονοχρωματικές εικόνες για το σχηματισμό του συνόλου των εκπαιδευτικών διανυσμάτων. Στη συνέχεια, επιλέγουμε το αρχικό κωδικό βιβλίο (για το σκοπό αυτό έχουν δημιουργηθεί διάφοροι αποτελεσματικοί αλγόριθμοι) και εφαρμόζουμε τη διαδικασία που μόλις περιγράψαμε μέχρι να ικανοποιηθεί το κριτήριο σύγκλισης (1.26) για την τιμή της ανοχής  $\epsilon$  που έχουμε επιλέξει. Με το βέλτιστο κωδικό βιβλίο  $Y^*$  που έχουμε σχηματίσει μπορούμε να κωδικοποιήσουμε μία δεδομένη μονοχρωματική εικόνα, η οποία είναι συνήθως διαφορετική από αυτές που χρησιμοποιούνται για να δώσουν τα εκπαιδευτικά διανύσματα (εκπαιδευτικές εικόνες). Η κωδικοποίηση γίνεται ως εξής: Διαιρούμε την προς κωδικοποίηση εικόνα σε τετράγωνες περιοχές ώστε η διάσταση των διανυσμάτων που προκύπτουν να είναι ίση με αυτή των αντίστοιχων του  $Y^*$ . Στη συνέχεια με βάση τη Συνθήκη του Πλησιέστερου Γείτονα αντιστοιχίζουμε τα διανύσματα, αποθηκεύοντας για κάθε διάνυσμα που διαβάζουμε και το δείκτη της θέσης του κωδικού διανύσματος με το οποίο έγινε η αντιστοίχιση. Δηλαδή αν στο  $i$ -οστό

διάνυσμα της εικόνας εισόδου  $x_i$  αντιστοιχίσουμε το  $y_k^*$ , αποθηκεύουμε το  $k$  ώστε κατά την αποκωδικοποίηση στη τετραγωνική περιοχή που βρίσκεται στη θέση  $i$  να τοποθετηθεί η τετραγωνική περιοχή που προκύπτει από τη μετατροπή του  $y_k^*$  από διάνυσμα σε πίνακα. Αρχίζοντας συνεπώς από μία αυθαίρετη εικόνα και έχοντας στη διάθεσή του τις διευθύνσεις των κωδικών λέξεων στο βιβλίο, ο αποκωδικοποιητής, ανακατασκευάζει τη συμπιεσμένη εικόνα εκτελώντας την παραπάνω διαδικασία για κάθε μη επικαλυπτόμενη τετραγωνική περιοχή διαστάσεων  $n^{1/2} * n^{1/2}$  ( $n=16$  στο παράδειγμά μας).

Ο βαθμός συμπίεσης, ο οποίος, ορίζεται ως ο λόγος του μεγέθους σε μνήμη της αρχικής εικόνας προς το μέγεθος σε μνήμη της κωδικοποιημένης, εξαρτάται τόσο από την επιλογή της διάστασης των τετραγωνικών περιοχών στην οποίες διαιρείται η εικόνα (όσο μεγαλύτερη η διάσταση τόσο λιγοστεύουν τα διανύσματα εισόδου), όσο και από το μέγεθος του κωδικού βιβλίου. Αν π.χ. τεμαχίσουμε μία εικόνα  $512 * 512$  εικονοστοιχείων, σε περιοχές  $4 * 4$  εικονοστοιχείων, παίρνουμε  $2^9 * 2^9 / 2^2 * 2^2 = 2^{14}$  διανύσματα εισόδου. Αν το πλήθος των κωδικών διανυσμάτων είναι  $512$ , για κάθε διάνυσμα εισόδου πρέπει να δεσμεύσουμε και  $9$  δυαδικά ψηφία (bits) για το δείκτη της θέσης του αντίστοιχου κωδικού διανύσματος και επομένως η συμπιεσμένη έκδοση της αρχικής εικόνας απαιτεί  $9 * 2^{14}$  δυαδικά ψηφία. Αν παριστάνουμε τη διαβάθμιση του γκρι με  $8$  bits/εικονοστοιχείο, το μέγεθος της αρχικής εικόνας είναι  $2^9 * 2^9 * 8$  bits, οπότε ο βαθμός συμπίεσης είναι ίσος με  $2^9 * 2^9 * 8 / 9 * 2^{14} \approx 14,2$ . Παρατηρούμε ότι αν διπλασιάσουμε τη διάσταση της τετραγωνικής περιοχής από  $4$  σε  $8$  εικονοστοιχεία (με την προϋπόθεση βέβαια ότι και τα κωδικά διανύσματα έχουν διάσταση  $n=8*8$ ) θα έχουμε  $2^{18} / 2^3 * 2^3 = 2^{12}$  διανύσματα εισόδου και ο βαθμός συμπίεσης θα τετραπλασιαστεί.

Κριτήριο της επίδοσης μιας μεθόδου συμπίεσης είναι και ο βαθμός παραμόρφωσης  $\lambda^1$ , που ορίζεται ως η θετική παράμετρος που δίνει την αναλογία μεταξύ της ποιότητας της αποκωδικοποιημένης εικόνας και του μεγέθους των συμπιεσμένων δεδομένων. Έστω δύο αποκωδικοποιήσεις της ίδιας εικόνας  $O, P$

και  $Q$ . Με δεδομένο ένα σταθερό βαθμό παραμόρφωσης, λέμε ότι η  $Q$  θεωρείται καλύτερη της  $P$  σε σχέση με το βαθμό παραμόρφωσης  $\lambda$ , αν και μόνο αν ικανοποιείται η σχέση:

$$\|O - Q\|_2 + \text{size}(Q) * \lambda < \|O - P\|_2 + \text{size}(P) * \lambda \quad (1.28)$$

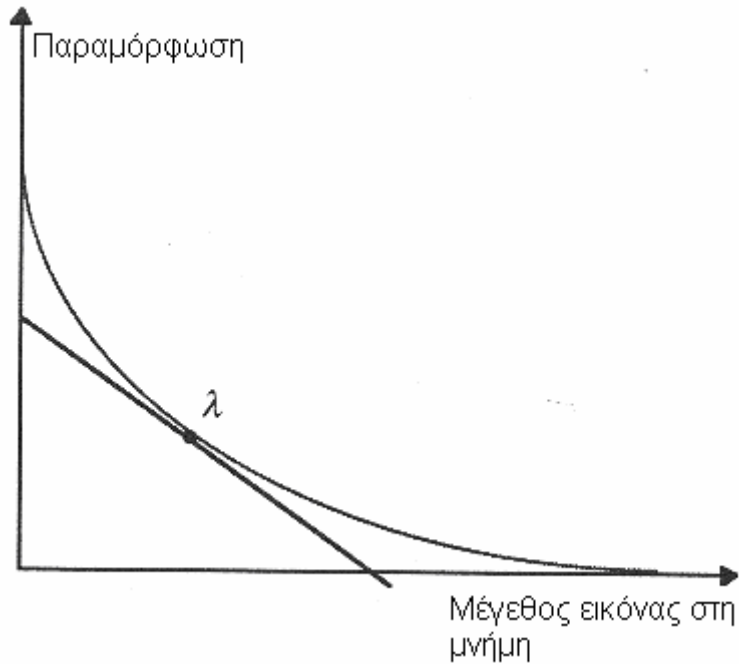
, όπου  $\text{size}(\bullet)$  το μέγεθος σε μνήμη της εικόνας. Η παραπάνω σχέση μπορεί να ξαναγραφεί ως εξής:

$$\lambda < \frac{\|O - P\|_2 - \|O - Q\|_2}{\text{size}(Q) - \text{size}(P)} \quad (1.29)$$

Η σχέση (1.29) καλείται *κριτήριο κόστους*. Σύμφωνα με αυτό, ο βαθμός παραμόρφωσης είναι η οριακή γραμμή μεταξύ της βελτίωσης του σφάλματος και της πρόσθετης απαίτησης σε μνήμη για την αποθήκευση των συμπιεσμένων δεδομένων. Το σχήμα 1.5 δείχνει ότι ο βαθμός παραμόρφωσης είναι η εφαπτομένη αυτής της καμπύλης.

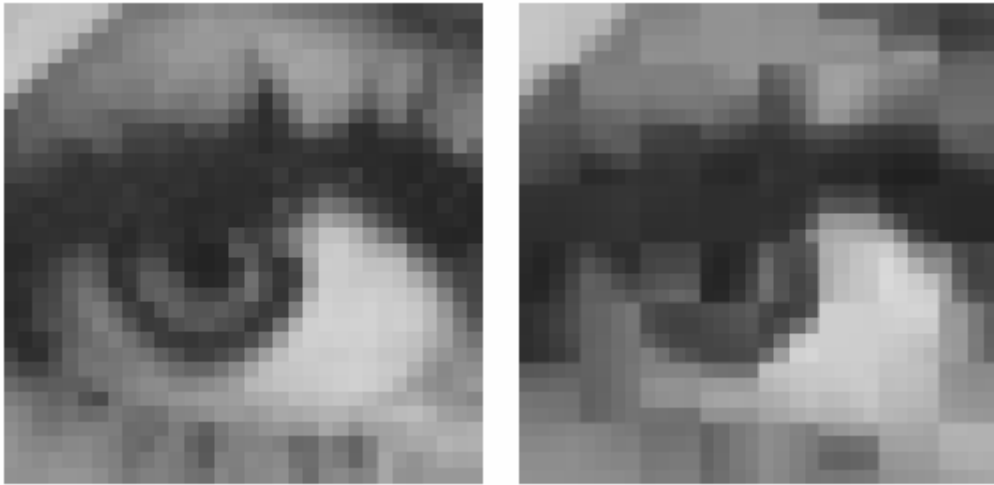
Ο βαθμός παραμόρφωσης συνεπώς, παίζει το ρόλο του ρυθμιστή της παραμόρφωσης της εικόνας και του μεγέθους της συμπίεσης. Στην ακραία περίπτωση κατά την οποία  $\lambda=0$ , επιλέγουμε συμπίεση που μας δίνει την καλύτερη δυνατή αναπαράσταση. Όταν  $\lambda \rightarrow \infty$ , επιλέγουμε συμπίεση με το μικρότερο δυνατό μέγεθος στη μνήμη, η οποία δίνει και φτωχή αναπαράσταση της αρχικής εικόνας.

Εκτός από το βαθμό συμπίεσης και το βαθμό παραμόρφωσης, ένα άλλο μέτρο της αποτελεσματικότητας μιας μεθόδου συμπίεσης είναι και το  $PSNR$ , όπου  $P$  στη σχέση ορισμού του (1.14) είναι η αρχική εικόνα και  $Q$  η αποκωδικοποίηση της συμπιεσμένης. Όσο μεγαλύτερος είναι αυτός ο λόγος τόσο καλύτερη είναι η ποιότητα της συμπίεσης.



Σχ.1.5: Ο βαθμός παραμόρφωσης στη γραφική παράσταση του μεγέθους στη μνήμη της συμπιεσμένης εικόνας και του σφάλματος ή παραμόρφωσης.

Μπορούμε, σύμφωνα με τις παραπάνω παρατηρήσεις, μεταβάλλοντας τον αριθμό των διανυσμάτων που περιέχει το κωδικό βιβλίο, τη διάσταση των τετραγωνικών περιοχών ή και τα δύο, να πάρουμε διαφορετικούς βαθμούς συμπίεσης, γεγονός το οποίο επηρεάζει και την τιμή του *PSNR*. Για να συγκρίνουμε λοιπόν δύο διαφορετικές τεχνικές συμπίεσης ή για να καταγράψουμε την επίδραση των διαφόρων παραμέτρων στην ίδια μεθοδολογία συμπίεσης, σχηματίζουμε γραφικές παραστάσεις του *PSNR* σε σχέση με το βαθμό συμπίεσης για κάθε τεχνική. Όσο υψηλότερη είναι η καμπύλη που αντιστοιχεί στην 1<sup>η</sup> τεχνική από αυτή που αντιστοιχεί στη 2<sup>η</sup>, τόσο καλύτερη είναι η 1<sup>η</sup> από τη 2<sup>η</sup>. Αντί για το βαθμό συμπίεσης, μπορούμε να χρησιμοποιήσουμε για το σχηματισμό των καμπυλών αυτών και το *bit rate*, το οποίο είναι απλώς ο λόγος του μεγέθους σε μνήμη της εικόνας προς τον αριθμό των εικονοστοιχείων.



Σχ.1.6 :Τυπική Διανυσματική Κβαντοποίηση ενός τμήματος της εικόνας της Lena (το αρχικό βρίσκεται στα αριστερά) με τη χρήση ενός κωδικού βιβλίου 512 διανυσμάτων διάστασης 4\*4. Το PSNR που δίνει η προσέγγιση αυτή είναι περίπου 31.2dB.

## 1.4 Διανυσματική Κβαντοποίηση με Αφαίρεση Μέσης

### Τιμής

Η Διανυσματική Κβαντοποίηση μπορεί να δώσει τις καλύτερες ίσως καμπύλες *PSNR-βαθμού συμπίεσης* από τις υπάρχουσες τεχνικές συμπίεσης, με το μειονέκτημα όμως ότι οι τετραγωνικές περιοχές στις οποίες πρέπει να τεμαχιστεί η εικόνα πρέπει να έχουν μεγάλες διαστάσεις. Μία τέτοια επιλογή, έχει ως απόρροια τις υψηλές απαιτήσεις σε μνήμη για την αποθήκευση ενός κωδικού βιβλίου με στοιχεία μεγάλων διαστάσεων και επιπλέον την αδυναμία σχεδίασης ενός τέτοιου κωδικού βιβλίου, καθώς εξαιτίας της τεράστιας πολυπλοκότητας, ο γενικευμένος αλγόριθμος *Lloyd* αποδεικνύεται εξαιρετικά χρονοβόρος. Έστω π.χ. ότι απαιτούμε σταθερό βαθμό συμπίεσης ίσο με  $\delta$  και χωρίζουμε την εικόνα σε περιοχές εμβαδού  $d$  εικονοστοιχείων. Έστω ακόμη ότι τα κωδικά διανύσματα, τα οποία έχουν την ίδια διάσταση  $d$ , είναι  $2^l$  τον

αριθμό. Εφ' όσον ο βαθμός συμπίεσης είναι  $\delta$  έχουμε  $\frac{512^2 * 8}{\frac{512^2}{d} l} = \delta \Rightarrow l = d$ .

Αυτό σημαίνει ότι για τετράγωνα διαστάσεων  $8*8=d=64$ , το πλήθος των κωδικών διανυσμάτων είναι  $2^{64}$  και η διάστασή τους  $64$ . Αν η συνιστώσα του

κάθε διανύσματος χρησιμοποιεί 8 bits για την απεικόνιση της διαβάθμισης του γκρι, τότε κάθε διάνυσμα απαιτεί  $64 \cdot 8$  bits για την αποθήκευσή του και κατά συνέπεια το κωδικό βιβλίο καταλαμβάνει χώρο ίσο με  $2^{64} \cdot 64$  bytes =  $2^{64} \cdot 2^6$  bytes =  $2^{70}$  bytes  $\approx 10^9$  Tb. Είναι φανερό ότι τέτοια μεγέθη υπερβαίνουν τις δυνατότητες ενός συμβατικού υπολογιστή.

Για το λόγο αυτό, έχουν ευρεθεί υποβέλτιστες (όσον αφορά την ποιότητα συμπίεσης) αλλά πρακτικά εφαρμόσιμες παραλλαγές της ΔΚ, όπως η *Mean-Removed Shape Gain VQ* (Διανυσματική Κβαντοποίηση με Αφαίρεση Μέσης Τιμής). Όπως δηλώνει το όνομα, ένα διάνυσμα  $R \in \mathfrak{R}^n$ , για να κωδικοποιηθεί γράφεται ως εξής:

$$\begin{aligned} R &= sD + o\mathbf{1}, \text{ με } \mathbf{1} = (1, 1, \dots, 1)^T \in \mathfrak{R}^n \\ s, o &\in \mathfrak{R} \text{ και } D = (d_1, d_2, \dots, d_n)^T \end{aligned} \quad (1.30)$$

Το διάνυσμα  $D$  καλείται διάνυσμα σχήματος (*shape-vector*) και έχει μηδενική μέση τιμή και μοναδιαία διασπορά, δηλαδή:

$$\sum_{i=1}^n d_i = 0, \quad \sum_{i=1}^n d_i^2 = 1 \quad (1.31)$$

Με δύο βαθμωτά κωδικά βιβλία  $S$  και  $O$  για τους συντελεστές  $s$  και  $o$  αντίστοιχα και ένα για τα διανύσματα σχήματος, το διάνυσμα εισόδου παίρνει τη μορφή:

$$R \approx s_i D_j + o_k \mathbf{1} \quad (1.32)$$

, όπου  $i, j, k$  είναι οι δείκτες της θέσης σε κάθε ένα από τα κωδικά βιβλία για τα  $s, D, o$  που δίνουν το ελάχιστο σφάλμα μεταξύ του  $R$  και της κωδικοποιημένης μορφής του  $sD + o\mathbf{1}$ .

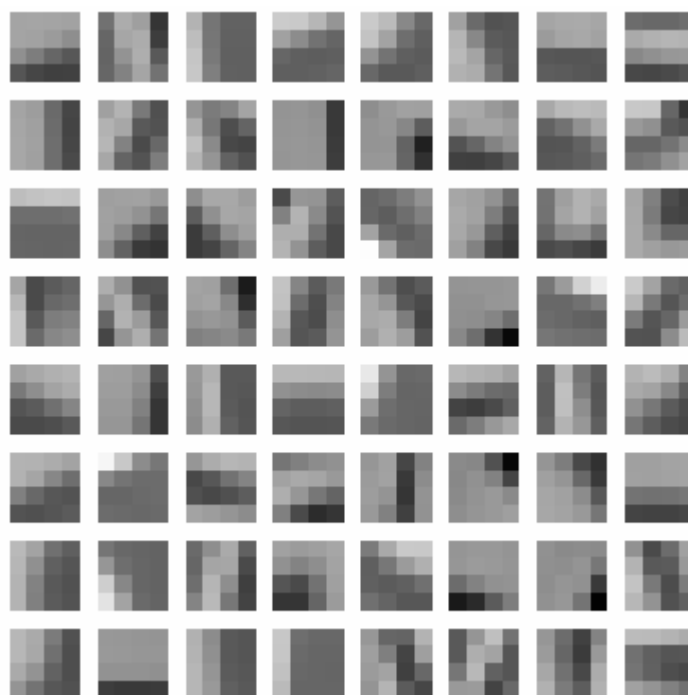
Με τη μέθοδο αυτή, έχουμε τη δυνατότητα από ένα και μόνο διάνυσμα  $D$ , να κατασκευάσουμε μία πληθώρα υποψήφιων διανυσμάτων για την



κωδικοποίηση του  $R$ , μεταβάλλοντας τις τιμές των  $s, a$ . Με αυτόν τον τρόπο μειώνουμε τις απαιτήσεις σε μνήμη αφού:

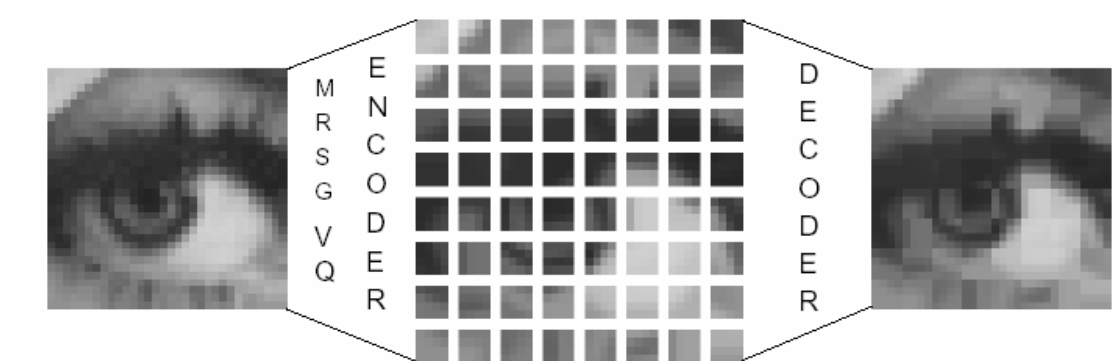
1. Το κωδικό βιβλίο δεν είναι ανάγκη να περιέχει τόσο μεγάλο αριθμό στοιχείων, όπως συμβαίνει στην περίπτωση της κλασσικής Δ.Κ.
2. Τα κωδικά βιβλία των  $s$  και  $a$  που χρησιμοποιούνται για την παραγωγή των εκπαιδευτικών διανυσμάτων, καταλαμβάνουν πολύ μικρό χώρο στη μνήμη καθώς το μέγεθος τους εξαρτάται από το πλήθος των βαθμωτών στοιχείων που περιέχουν και τα ψηφία που έχουμε δεσμεύσει για την απεικόνισή τους και όχι από τη διάσταση των στοιχείων, όπως θα συνέβαινε στην περίπτωση ενός συνόλου κωδικών διανυσμάτων.

Αν λοιπόν λάβουμε υπόψη και τα τρία κωδικά βιβλία ενσωματώνοντάς τα σε ένα κοινό σύνολο, θα μπορούσαμε να έχουμε ένα πολύ μεγάλο γενικευμένο κωδικό βιβλίο με μικρές απαιτήσεις αποθήκευσης σε σχέση με αυτό της κλασσικής μεθόδου. Για παράδειγμα, αν οι διαστάσεις των βιβλίων για τα  $s, a, D$  είναι 32, 128 και 4096 αντίστοιχα, μπορούμε να δημιουργήσουμε ένα σύνολο  $2^5 * 2^7 * 2^{12} = 2^{24}$  διανυσμάτων.



Σχ.1.7 : Οπτική αναπαράσταση ενός κωδικού βιβλίου της μεθόδου της Διανυσματικής Κβαντοποίησης με Αφαίρεση Μέσης Τιμής. Αποτελείται από 64 τετράγωνα –διανύσματα διάστασης 16 με μηδενική μέση τιμή

και μοναδιαία διασπορά. Για τις ανάγκες της απεικόνισης οι συνιστώσες του κάθε διανύσματος έχουν πολλαπλασιαστεί με 180 και στο γινόμενο έχει προστεθεί το 127.



Σχ.1.8: Οπτική αναπαράσταση της μεθόδου της Διανυσματικής Κβαντοποίησης με Αφαίρεση Μέσης Τιμής για το τμήμα της εικόνας της Lena του σχήματος 1.6 . Η εικόνα έχει κωδικοποιηθεί με βάση το κωδικό βιβλίο του προηγούμενου σχήματος. Τα κωδικά βιβλία για το κέρδος  $s$  και τη διόρθωση φωτεινότητας ο περιέχουν 32 και 128 διακριτές τιμές αντίστοιχα. Το PSNR της συμπίεσης αγγίζει τα 34.6 dB.

## Βιβλιογραφία

1. Ning Lu, *Fractal Imaging*, Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto.
2. Dietmar Saupe, Raouf Hamzaoui, Hannes Haternstein, *Fractal Image Compression- An Introductory Overview*, in: *Fractal Models for Image Synthesis, Compression and Analysis*, D. Saupe, J. Hart (eds.), ACM SIGGRAPH'96 Course Notes 27, New Orleans, Louisiana, Aug. 1996.
3. Linde, Y., Buzo, A., & Gray, R. (1980). *An Algorithm for Vector Quantizer Design*. IEEE Transactions on Communications, 28(1), 84-94.
4. Gersho, A., & Gray, R. (1992). *Vector Quantization and Signal Compression* Boston: Kluwer
5. Lloyd, S. (1957). *Least Squares Quantization in PCM's*. (Bell Telephone Labs Paper, Murray Hill)

---

## ΚΕΦΑΛΑΙΟ 2

---

### Η συμπίεση εικόνας με βάση τη θεωρία των fractal

Ο πυρήνας της μεθόδου συμπίεσης με fractal είναι παρόμοιος με αυτόν της μεθόδου της Διανυσματικής Κβαντοποίησης που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Η ειδοποιός διαφορά μεταξύ τους είναι, ότι ενώ στη Δ.Κ. το κωδικό βιβλίο είναι σταθερό και ανεξάρτητο της εικόνας εισόδου, το κωδικό βιβλίο στους fractal αλγορίθμους προσαρμόζεται ανάλογα με την προς συμπίεση εικόνα. Στην πραγματικότητα, τα διανύσματα του κωδικού βιβλίου των fractal αλγορίθμων προέρχονται από την ίδια την εικόνα εισόδου, την εικόνα δηλαδή που πρέπει να συμπιστεί. Αυτό φαίνεται να έρχεται σε αντίθεση με όσα έχουν προαναφερθεί, αφού ο αποκωδικοποιητής, ο οποίος είναι επιφορτισμένος με την ανακατασκευή της αρχικής εικόνας, δεν μπορεί να έχει πρόσβαση στην αρχική εικόνα. Όπως γνωρίζουμε από τη θεωρία της Διανυσματικής Κβαντοποίησης, στον αποκωδικοποιητή είναι διαθέσιμα το κωδικό βιβλίο και η διεύθυνση σε αυτό της κωδικής λέξης-τετραγωνικής περιοχής που πρέπει να τοποθετηθεί σε μια συγκεκριμένη θέση του πλαισίου της εικόνας για την ανακατασκευή της συμπιεσμένης. Όταν όμως το κωδικό βιβλίο προέρχεται από την προς συμπίεση εικόνα ο αποκωδικοποιητής δεν μπορεί να έχει πρόσβαση σε αυτό. Πώς είναι λοιπόν δυνατό, ενώ η εικόνα κωδικοποιείται ανά τετραγωνικές περιοχές, κάθε μία από τις οποίες γράφεται ως άθροισμα μίας άλλης ενδεχομένως περιοχής που ανήκει στην ίδια την εικόνα, πολλαπλασιασμένης με ένα παράγοντα  $s$  και μίας σταθερής περιοχής (ως προς τη διαβάθμιση του γκρι) ίδιων διαστάσεων, ο αποκωδικοποιητής να ανακατασκευάσει την αρχική εικόνα;

Στις σελίδες που ακολουθούν, δίνεται η απάντηση στο παραπάνω ερώτημα με την αναφορά στη μαθηματική βάση πάνω στην οποία αναπτύχθηκε η ιδέα της συμπίεσης μιας εικόνας με τη χρήση των fractal και μελετάται η βασική μέθοδος συμπίεσης που ενέπνευσε η θεωρία των τελευταίων: η τεχνική των *Partitioned Iterated Function Systems* (**PIFS**).

## 2.1 Ένα μονοδιάστατο παράδειγμα

Για να κατανοηθεί ο τρόπος λειτουργίας του fractal αλγόριθμου, παραθέτουμε το απλό παράδειγμα κωδικοποίησης ενός πραγματικού αριθμού, έστω του αριθμού  $\pi=3,1415\dots$ <sup>1</sup> Υποθέτουμε ότι τα κωδικά βιβλία για τους συντελεστές  $s$  και  $o$  είναι:

$$s \in \{0, 0.25, 0.5, 0.75\}, o \in \{0.0, 0.4, 0.8, 1.2, 1.6, 2.0\} \quad (2.1)$$

Το κωδικό βιβλίο των διανυσμάτων  $D$  αποτελείται από ένα μόνο στοιχείο, δηλαδή τον ίδιο τον αριθμό  $\pi$  (το κωδικό βιβλίο προέρχεται όπως αναφέρθηκε από την ίδια την εικόνα). Ο παρακάτω πίνακας δείχνει τους πιθανούς αριθμούς της μορφής  $s*\pi+o$  που παίρνουμε για όλες τις τιμές των  $s, o$  που ανήκουν στα σύνολα που εκλέξαμε.

Πίνακας 2.1: Απεικόνιση όλων των αριθμών (στρογγυλοποιημένων στο δεύτερο δεκαδικό ψηφίο) που δίνει η σχέση  $s*\pi+o$  με βάση τα κωδικά βιβλία της σχέσης (2.1).

Πολλαπλασιαστής $s$	Σταθερή Συνιστώσα $o$					
	0.00	0.40	0.80	1.20	1.60	2.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.79	1.19	1.59	1.99	2.39	2.79
0.50	1.57	1.97	2.37	2.77	3.17	3.57
0.75	2.36	2.76	3.16	3.56	3.96	4.36

Από τον πίνακα βλέπουμε ότι για  $s=0.75$  και  $o=0.8$  έχουμε την καλύτερη προσέγγιση του  $\pi$  δηλαδή  $s*\pi+o = 0.75\pi+0.8 = 3,1561\dots$  Επομένως, ο κωδικοποιητής μπορεί να δώσει στον αποκωδικοποιητή την ακόλουθη πληροφορία : Ο αριθμός που ζητούμε είναι περίπου  $0.75$  φορές ο εαυτός του  $+ 0.8$  . Το σφάλμα αυτής της προσέγγισης δεν έχει διευκρινιστεί και φυσικά υπάρχουν πολλοί αριθμοί που ικανοποιούν την παραπάνω περιγραφή. Μη γνωρίζοντας καμιά άλλη πληροφορία ο αποκωδικοποιητής θα μπορούσε να βρει οποιονδήποτε από αυτούς. Παρόλα αυτά ένας από αυτούς είναι ο μοναδικός αριθμός  $x$  που είναι ακριβώς ίσος με  $0.75$  φορές τον εαυτό του συν  $0.8$ , δηλαδή  $x = 0.75x + 0.8$ .

Η λύση της παραπάνω εξίσωσης είναι ο αριθμός 3.2, ο οποίος μπορεί να θεωρηθεί ως ο αποκωδικοποιημένος αριθμός. Επομένως ο κωδικοποιητής προσεγγίζει τον αριθμό εισόδου χρησιμοποιώντας τον ίδιο τον αριθμό και τα κωδικά βιβλία για το  $s$  και το  $\theta$ , ενώ ο αποκωδικοποιητής, ο οποίος δε μπορεί να γνωρίζει την προσέγγιση αυτή, δίνει ως αποτέλεσμα το μοναδικό αριθμό που χαρακτηρίζεται από την ιδιότητα εκείνη, σύμφωνα με την οποία, ο κωδικοποιητής με τις δεδομένες τιμές των  $s$  και  $\theta$  μπορούσε να κωδικοποιήσει το ζητούμενο αριθμό χωρίς σφάλμα.

Η παραπάνω εξίσωση  $x=0.75x+0.8$  είναι εύκολη στη λύση της. Όταν έχουμε όμως να αντιμετωπίσουμε εικόνες που περιέχουν χιλιάδες αριθμούς (η διαβάθμιση του γκρι για κάθε εικονοστοιχείο), το σύστημα που προκύπτει είναι τόσο περίπλοκο ώστε δε μπορεί να επιλυθεί απευθείας, παρά μόνο επαναληπτικά. Η δυνατότητα εύρεσης της λύσης με επανάληψη μπορεί να φανεί και στο παράδειγμα με τον αριθμό  $\pi$ . Με τον ορισμό ενός τελεστή  $T(\cdot)$  ως  $Tx = 0.75x+0.8$ , η κωδικοποίηση επιβάλλει την επαναληπτική επίλυση της εξίσωσης  $x=Tx$ . Αρχίζοντας από ένα αυθαίρετο αρχικό σημείο  $x_0$ , εφαρμόζουμε επαναληπτικά τον τελεστή  $T$  που δίνει  $x_1=Tx_0$ ,  $x_2=Tx_1$ ,  $x_3=Tx_2$ . Επιλέγοντας π.χ.,  $x_0=0$  παίρνουμε:  $x_1=0.8$ ,  $x_2=0.75*0.8+0.8=1.4$ ,  $x_3=0.75*1.4+0.8=1.5$ ,  $x_4=2.1875$  και συνεχίζοντας  $x_{10}=3.06\dots$ ,  $x_{20}=3.192\dots$ ,  $x_{30}=3.195$ . Παρατηρούμε ότι η λύση συγκλίνει στο σταθερό σημείο 3.2 το οποίο καλείται και *οριακό σημείο (attractor)* του τελεστή  $T$ . Αυτό δεν είναι τυχαίο, καθώς αφού  $|s|<1$  ο  $Tx=s*x+\theta$  δεν είναι τίποτε άλλο παρά ένας συστολικός μετασχηματισμός.

## 2.2 Ο Συστολικός Μετασχηματισμός-Οι μετασχηματισμοί τύπου affine

Δεδομένου ενός συνόλου  $X$ , μία συνάρτηση της μορφής  $d : X \times X \rightarrow [0, A)$  καλείται *μετρική ή συνάρτηση απόστασης* αν :

1. (ταυτότητα):  $d(x, y) = 0 \Leftrightarrow x = y, \forall x, y \in X$
2. (συμμετρία):  $d(x, y) = d(y, x), \forall x, y \in X$
3. (τριγωνική ανισότητα):  $d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in X$

Το ζευγάρι  $(X, d)$  καλείται *μετρικός χώρος*. Ένας μετασχηματισμός  $W$  είναι *συστολικός* (*contractive*) για δύο σημεία  $x$  και  $y$  που ανήκουν σε ένα μετρικό χώρο  $X$ , αν η απόσταση  $d(\cdot)$  έχει την ακόλουθη ιδιότητα :

$$d(W(x), W(y)) < s d(x, y), \quad 0 \leq s < 1 \quad (2.2)$$

Αυτό σημαίνει ότι τα σημεία  $x$  και  $y$  θα πλησιάζουν πάντα το ένα το άλλο κατά ένα παράγοντα  $s$  (*παράγοντας συστολής*). Επομένως, όταν ένας συστολικός μετασχηματισμός εφαρμόζεται επαναληπτικά σε δύο σημεία  $x$  και  $y$ , τα δύο αυτά σημεία θα συγκλίνουν σε ένα μοναδικό σημείο. Αυτό το σημείο παραμένει σταθερό και ανεξάρτητο από την εφαρμογή των περαιτέρω μετασχηματισμών.

Με βάση τα παραπάνω διατυπώνουμε το *Θεώρημα του Συστολικού Μετασχηματισμού Σταθερού Σημείου* :

**Θεώρημα 1:** *Αν ο  $X$  είναι ένας «πλήρης» μετρικός χώρος και ο μετασχηματισμός  $W: X \rightarrow X$  είναι συστολικός τότε ο  $W$  συγκλίνει σε ένα μοναδικό και σταθερό σημείο.*

**Απόδειξη.** Για να αποδείξουμε το θεώρημα αυτό θα χρησιμοποιήσουμε επαναληπτική μέθοδο. Έστω σημείο  $x \in X$ . Δημιουργούμε την ακολουθία σημείων:

$$W^0(x) = x, W^1(x) = W(x), W^i(x) = (\dots W(x) \dots) \quad (2.3)$$

Αφού όμως ο  $W$  είναι συστολικός μετασχηματισμός θα ισχύει:

$$d(W^{i+1}(x), W^{i+2}(x)) < s \cdot d(W^i(x), W^{i+1}(x)) \quad (2.4)$$

Δηλαδή σε κάθε βήμα, η απόσταση από το επόμενο σημείο της ακολουθίας που έχουμε δημιουργήσει, είναι μικρότερη της αποστάσεως από τα

προηγούμενα κατά ένα παράγοντα  $s < 1$ . Καθώς προχωρούμε με γεωμετρικά βήματα (δηλαδή ορίζεται η απόσταση από το  $W^i(x)$  στο  $W^{i+1}(x)$ ) και ο χώρος  $X$  που είναι μετρικός δεν έχει ασυνέχειες, μετά από άπειρη εφαρμογή του μετασχηματισμού  $W$  πάνω στο  $x$ , η ακολουθία των απείρων σημείων που δημιουργείται θα συγκλίνει (όπως δηλώνει και η σχέση της απόστασης) σε ένα σταθερό σημείο  $g$  που το συμβολίζουμε με  $g = \lim_{i \rightarrow \infty} W^i(x)$ . Πράγματι, κάθε φορά που εφαρμόζουμε το μετασχηματισμό  $W$  μετακινούμε γεωμετρικά το προηγούμενο σημείο  $W^i(x)$  στο επόμενο  $W^{i+1}(x)$ . Καθώς τα  $W^i(x)$  ανήκουν στο  $X, d(\cdot)$  η επόμενη μετακίνηση θα έχει μικρότερη απόσταση από την προηγούμενη και συνεπώς

$$\begin{aligned} d(W^{i+1}(x), W^{i+2}(x)) &< s \cdot d(W^i(x), W^{i+1}(x)) < s^2 \cdot d(W^{i-1}(x), W^i(x)) < \\ &< s^3 \cdot d(W^{i-2}(x), W^{i-1}(x)) < \dots < s^{i+1} \cdot d(W(x), x) \end{aligned} \quad (2.5)$$

Αν  $s=0$ , τότε  $W^i(x)=x=g, i=1, \dots$  που είναι και το μοναδικό οριακό σημείο.

Αν  $0 < s < 1$ , τότε καθώς το  $i \rightarrow \infty$ , το  $s^i \rightarrow 0$  και καθώς  $d(W(x), x) \geq 0$ , τότε

$$d(W^{i+1}(x), W^{i+2}(x)) < 0. \text{ Αφού όμως } d(\cdot) \geq 0 \xrightarrow{i \rightarrow \infty} W^{i+1}(x) = W^{i+2}(x) = g.$$

Έστω τώρα ότι για δύο διαφορετικά σημεία του χώρου  $X$ ,  $x_1$  και  $x_2$  ο ίδιος ο μετασχηματισμός  $W$  συγκλίνει σε δύο διαφορετικά σημεία  $g_1$  και  $g_2 \in X$ .

Αφού  $g_1 = \lim_{i \rightarrow \infty} W^i(x_1)$  και  $g_2 = \lim_{i \rightarrow \infty} W^i(x_2)$ ,  $W(g_1) = g_1$  και  $W(g_2) = g_2$ ,

καθώς η εφαρμογή του  $W$  στα δύο σημεία σύγκλισης δεν πρόκειται να τα μετακινήσει καθόλου. Αλλά τα  $g_1$  και  $g_2$  είναι σημεία του  $X, d(\cdot)$  που

σημαίνει ότι  $d(g_1, g_2) < s \cdot d(W(g_1), W(g_2))$ , δηλαδή  $d(g_1, g_2) < s \cdot d(g_1, g_2)$ .

Ο παράγοντας συστολής  $s$  όμως είναι αυστηρά μικρότερος της μονάδας και κατά συνέπεια η τελευταία ανισότητα δεν ισχύει, άρα  $g_1 = g_2 = g \square$ .

Ο συστολικός μετασχηματισμός ανήκει σε ένα ευρύτερο σύνολο μετασχηματισμών από το  $\mathfrak{R}^n$  στο  $\mathfrak{R}^n$  που καλούνται μετασχηματισμοί τύπου affine. Ένας μετασχηματισμός τύπου affine  $f: \mathfrak{R}^n \rightarrow \mathfrak{R}^n$  είναι ένας μετασχηματισμός που μπορεί να γραφεί ως εξής:



$$f(x) = Ax + b = \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix} \quad (2.6)$$

, όπου  $A = (a_{ij})_{i,j=1}^n$  είναι ένας  $n \times n$  πίνακας στο  $\mathfrak{R}^{n \times n}$ , που καλείται *πίνακας παραμόρφωσης (deformation matrix)* του  $f$  και το  $b = (b_j)_{j=1}^n$  είναι ένα διάνυσμα στο  $\mathfrak{R}^n$  το οποίο ονομάζεται διάνυσμα μεταφοράς (*translation vector*) του  $f$ . Δεδομένης μίας νόρμας  $\|\cdot\|$  στο χώρο  $\mathfrak{R}^n$ , η νόρμα ενός μετασχηματισμού τύπου affine  $f$ , ή ενός πίνακα παραμόρφωσης  $A$ , δίνεται από τη σχέση:

$$\|f\| = \|A\| = \sup_{x \in \mathfrak{R}^n} \frac{\|A(x)\|}{\|x\|} = \sup_{x \in \mathfrak{R}^n, \|x\|=1} \|A(x)\| \quad (2.7)$$

Ένας μετασχηματισμός τύπου affine  $f$ , ή ένας πίνακας παραμόρφωσης  $A$  λέγεται *συστολικός*, αν  $\|A\| < 1$ . Επιπλέον λέγεται ότι είναι συστολικός μέσα στον παράγοντα συστολής  $s$ , αν  $\|A\| < s$ ,  $0 \leq s < 1$ .

Οι πιο γνωστές νόρμες που χρησιμοποιούνται για τον πίνακα παραμόρφωσης είναι οι ακόλουθες:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|, \quad (2.8)$$

$$\|A\|_2 = \sqrt{\max_{1 \leq i \leq n} (\lambda_i)}, \quad \left| \lambda_i I - A^T A \right|_{i=1, \dots, n} = 0, \quad (2.9)$$

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|. \quad (2.10)$$

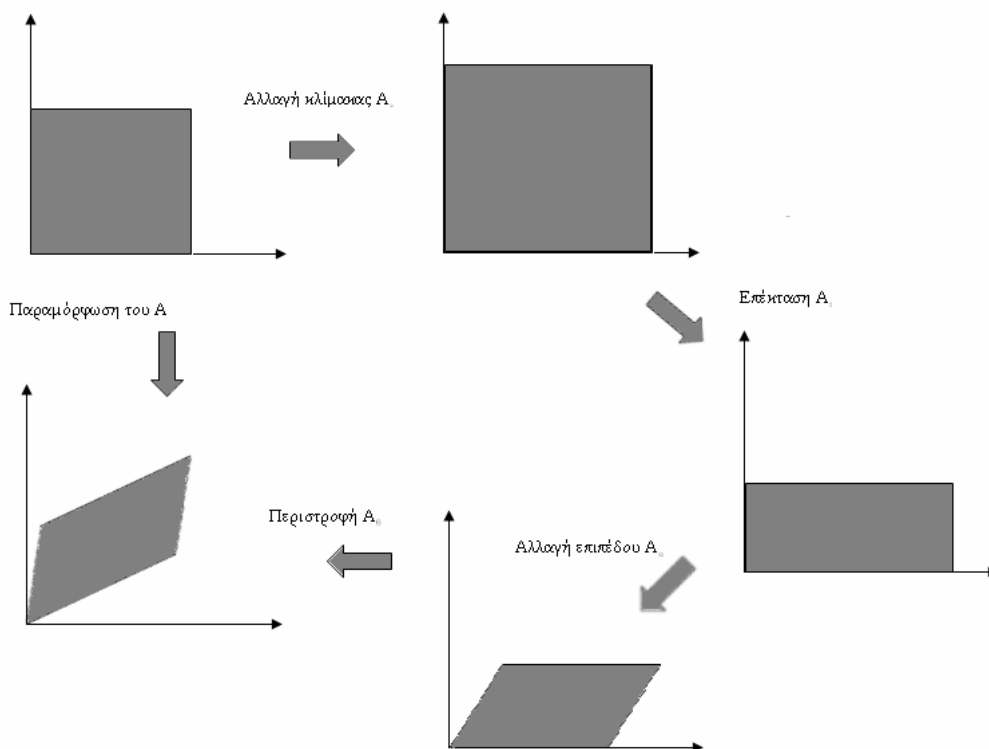
Στη δισδιάστατη περίπτωση ( $n=2$ ), ο πίνακας  $A$  μπορεί να αναλυθεί σε 4 βήματα: αλλαγή κλίμακας (*scaling*), επέκταση (*stretching*), αλλαγή επιπέδου (*skewing*) και περιστροφή (*rotating*). Το παρακάτω σχήμα δείχνει τα βήματα της ανάλυσης αυτής.

1. Αλλαγή Κλίμακας:  $A_s = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}, s \geq 0.$

2. Επέκταση:  $A_t = \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix}, A_t A_s = \begin{pmatrix} s & 0 \\ 0 & st \end{pmatrix}.$

3. Αλλαγή επιπέδου:  $A_u = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, A_u A_t A_s = \begin{pmatrix} s & stu \\ 0 & st \end{pmatrix}.$

4. Περιστροφή:  $A_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, 0 \leq \theta < 2\pi.$



Σχ.2.1: Ο πίνακας παραμόρφωσης

Για μία δεδομένη συνάρτηση απόστασης  $d(\cdot)$ , ένας μετασχηματισμός τύπου affine, ή ένας πίνακας παραμόρφωσης ονομάζεται *ισομετρικός*, αν διατηρεί αμετάβλητη τη συνάρτηση απόστασης, δηλαδή αν για οποιαδήποτε  $x, y \in \mathbb{R}^n$ ,  $d(x, y) = d(f(x), f(y)) = d(Ax, Ay)$ . Στην περίπτωση της χρησιμοποίησης της Ευκλείδειας απόστασης, που είναι και η συνηθέστερη, οι μόνες μετατροπές που μπορούν να χαρακτηριστούν ισομετρικοί μετασχηματισμοί είναι οι περιστροφές συνδυασμένες με την *αναστροφή* (*flip*)  $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$  και είναι οι ακόλουθοι:

$$\begin{array}{cccc} 1: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 2: \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} & 3: \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & 4: \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \\ 5: \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 6: \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & 7: \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 8: \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \end{array} \quad (2.11)$$

Παρατηρούμε ότι στις περιπτώσεις 1, 4, 6 και 7 οι πίνακες που προκύπτουν είναι ίδιοι με αυτούς που θα προέκυπταν με στροφή ενός αρχικού τετραγώνου γύρω από το κέντρο κατά γωνίες  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  και  $270^\circ$  αντίστοιχα. Οι 2, 3, 5 και 8 περιγράφουν την αναστροφή του αρχικού τετραγώνου ως προς τις ευθείες  $y=0$  (κεντρικός οριζόντιος άξονας τετραγώνου),  $x=0$  (κεντρικός κάθετος άξονας),  $x=y$  (διαγώνιος 1<sup>ου</sup> και 3<sup>ου</sup> τεταρτημορίου) και  $x=-y$  (διαγώνιος 2<sup>ου</sup> και 4<sup>ου</sup> τεταρτημορίου) αντίστοιχα.

## 2.3 Συστήματα Επαναληπτικών Συναρτήσεων

Εκτός των κανονικών συναρτήσεων απόστασης, όπως η Ευκλείδεια νόρμα, για σύνολα σημείων μπορούμε να χρησιμοποιήσουμε και τη συνάρτηση *Hausdorff*, ως συνάρτηση απόστασης. Δεδομένου ενός μετρικού χώρου  $(X, d)$ , ο χώρος Hausdorff  $H(X)$ , ορίζεται ως ο χώρος του οποίου σημεία είναι τα συμπαγή υποσύνολα του  $X$  εκτός του κενού συνόλου. Ένα σύνολο  $A$  καλείται *συμπαγές* όταν:

1. Είναι φραγμένο, δηλαδή υπάρχει σταθερά  $\delta$  τέτοια ώστε  $d(x,y) \leq \delta$  για οποιαδήποτε δύο σημεία  $x, y$  που ανήκουν στο  $A$ . Αυτή η πρόταση ισοδυναμεί με τη διατύπωση ότι το  $A$  έχει πεπερασμένη διάμετρο

$$\text{diam}(A) \triangleq \sup_{x,y \in A} d(x,y) < \infty$$

2. Είναι κλειστό, δηλαδή για μία οποιαδήποτε ακολουθία σημείων  $\{x_i\}_{i=1,2,\dots}$  που συγκλίνει στο  $\mathfrak{R}^n$  σε ένα οριακό σημείο  $x \in \mathfrak{R}^n$ , το  $x$  ανήκει και στο  $A$ .

Η συνάρτηση *Hausdorff* που χρησιμοποιείται για τον καθορισμό της απόστασης μεταξύ δύο συνόλων  $A$  και  $B$  που ανήκουν στο χώρο  $H(X)$ , ορίζεται από την ακόλουθη σχέση:

$$h(A, B) = \max \{d(A, B), d(B, A)\} \quad (2.12)$$

όπου

$$d(A, B) = \max \{d(x, B) : x \in A\} \quad (2.13)$$

και

$$d(x, B) = \min \{d(x, y) : y \in B\} \quad (2.14)$$

όπου  $d(\cdot)$  είναι μία τυπική συνάρτηση απόστασης, με την Ευκλείδεια νόρμα να χρησιμοποιείται πιο συχνά.

Έστω τώρα ο διανυσματικός χώρος  $\mathfrak{R}^n$  και μία συνάρτηση απόστασης  $d(\cdot)$  που έχει οριστεί σε αυτόν. Σύστημα Επαναληπτικών Συναρτήσεων (*Iterated Function System, IFS*) στο χώρο αυτό, ονομάζεται το πεπερασμένο σύνολο συστολικών (τύπου *affine*) μετασχηματισμών  $W = \{w_1, w_2, \dots, w_m\}$ . Ο παράγοντας συστολής  $s$  του συνόλου  $W$  ορίζεται ως η μέγιστη τιμή των παραγόντων συστολής κάθε ενός από τους συστολικούς μετασχηματισμούς,  $w_i$  ( $i=1, \dots, m$ ), που αποτελούν στοιχεία του  $W$ :

$$s = \max_{1 \leq i \leq m} \{\|w_1\|, \|w_2\|, \dots, \|w_m\|\} \quad (2.15)$$

Δεδομένου του συνόλου  $W = \{w_1, w_2, \dots, w_m\}$ , ορίζουμε το μετασχηματισμό που σχετίζεται με αυτό στο χώρο των συμπαγών υποσυνόλων  $H(\mathfrak{R}^n)$ , με τη σχέση:

$$W(B) = w_1(B) \cup w_2(B) \cup \dots \cup w_m(B), \forall B \in H(\mathbb{R}^n). \quad (2.16)$$

Έχοντας ορίσει το *IFS* και τις *Hausdorff* αποστάσεις, μπορούμε να διατυπώσουμε το παρακάτω Θεώρημα που είναι άμεση συνέπεια του Θεωρήματος του Συστολικού Μετασχηματισμού Σταθερού Σημείου :

**Θεώρημα 2-Θεώρημα Αποκωδικοποίησης IFS (IFS Decoding Theorem)<sup>2</sup>:**

Έστω  $W = \{w_1, w_2, \dots, w_m\}$  ένα σύστημα επαναληπτικών συναρτήσεων με παράγοντα συστολής  $s$ . Τότε ο μετασχηματισμός  $W: H(\mathbb{R}^n) \rightarrow H(\mathbb{R}^n)$  είναι συστολικός στο χώρο  $H(\mathbb{R}^n)$ , στον οποίο έχει οριστεί μία συνάρτηση απόστασης  $h$  τύπου *Hausdorff*, με τον ίδιο παράγοντα συστολής  $s$ . Αυτό σημαίνει ότι  $h(W(B), W(C)) \leq s \cdot h(B, C)$ , για δύο οποιαδήποτε συμπαγή σύνολα  $B, C$  του  $\mathbb{R}^n$  που ανήκουν στο  $H(\mathbb{R}^n)$ .

Κατά συνέπεια, ο  $W$  συγκλίνει σε ένα μοναδικό σταθερό σημείο  $A \in H(\mathbb{R}^n)$ , δηλαδή σε ένα συμπαγές σύνολο  $A$  στον  $\mathbb{R}^n$  που υπακούει στη σχέση:

$$A = W(A) = \bigcup_{k=1}^m w_k(A) \quad (2.17)$$

Το συμπαγές αυτό υποσύνολο  $A$  ονομάζεται *ελκυστής (attractor)*, ή *ντετερμινιστικό fractal (deterministic fractal)* του *IFS*  $W = \{w_1, w_2, \dots, w_m\}$ . Επιπλέον, καθώς ο  $W$  είναι συστολικός, για οποιαδήποτε άλλο συμπαγές υποσύνολο  $B \in H(\mathbb{R}^n)$ , η ακολουθία:

$$B, W(B), W^2(B) = W(W(B)), \dots, W^k(B) = W(W^{k-1}(B)), \dots \quad (2.18)$$

συγκλίνει στο όριο  $A \in H(\mathbb{R}^n)$ .

Είναι λοιπόν εμφανές ότι, αν πάρουμε μια αυθαίρετη εικόνα (σύνολο  $B$ ), η οποία στην περίπτωση μας δεν είναι τίποτε άλλο παρά ένας πίνακας που οι τιμές των στοιχείων του αντιστοιχούν στις διαβαθμίσεις του γκρι, και εφαρμόσουμε σε αυτήν ένα συστολικό μετασχηματισμό  $W = \{w_1, w_2, \dots, w_m\}$ , διαιρώντας π.χ. την εικόνα σε τετραγωνικές περιοχές και εφαρμόζοντας σε

κάθε μία από τις οποίες ένα συστολικό μετασχηματισμό  $w_i$  ( $i=1,\dots,m$ ), θα καταλήξουμε σε μία άλλη εικόνα (σύνολο  $A$ ), το όριο του  $W$ . Η μορφή της οριακής αυτής εικόνας εξαρτάται από τον ίδιο το μετασχηματισμό, καθώς διαφορετικοί μετασχηματισμοί συγκλίνουν σε διαφορετικό όριο.

Για ένα δεδομένο λοιπόν Σύστημα Επαναληπτικής εφαρμογής Συναρτήσεων μπορούμε, όπως διατείνεται το θεώρημα να καταλήξουμε σε μία συγκεκριμένη εικόνα. Είναι δυνατόν να επιλυθεί το αντίστροφο πρόβλημα, από ένα δεδομένο δηλαδή συμπαγές σύνολο να ευρεθεί το IFS που η εφαρμογή του σε ένα άλλο τυχαίο σύνολο να μας δίνει το αρχικό;

Μετά από επισταμένη έρευνα, ο *Michael Barnsley* κατόρθωσε να δώσει απάντηση σε ένα παραπλήσιο πρόβλημα το οποίο διατυπώνεται ως εξής: *Για ένα δεδομένο Επαναληπτικό Σύστημα Συναρτήσεων και ένα δεδομένο συμπαγές σύνολο, είναι δυνατό το σύνολο αυτό να είναι το οριακό σημείο του Επαναληπτικού Συστήματος που εξετάζουμε;* Το θεώρημα του *Barnsley*, το οποίο ονομάζεται *Θεώρημα Collage* ή *Θεώρημα Κωδικοποίησης με IFS (IFS Encoding Theorem)*, δίνει μία εκτίμηση μεταξύ της απόστασης ενός δεδομένου συμπαγούς συνόλου και του οριακού σημείου ενός δεδομένου IFS.

### **Θεώρημα 3-Θεώρημα Collage (Collage Theorem)<sup>3</sup>:**

Έστω  $W = \{w_1, w_2, \dots, w_m\}$  ένα IFS με παράγοντα συστολής  $0 \leq s < 1$ . Έστω ακόμη  $W: H(\mathbb{R}^n) \rightarrow H(\mathbb{R}^n)$ , ο μετασχηματισμός στο χώρο Hausdorff που σχετίζεται με το παραπάνω IFS και ας υποθέσουμε ότι το  $A$  είναι το αιτιοκρατικό fractal (attractor)

αυτού ( $A = W(A) = \bigcup_{k=1}^m w_k(A)$ ). Τότε

$$h(L, A) \leq \frac{h(L, W(L))}{1-s}, \forall L \in H(\mathbb{R}^n) \quad (2.19)$$

Το συμπέρασμα που εξάγεται από το παραπάνω θεώρημα είναι ότι για να είναι το  $L$  κοντά στο  $A$ , αρκεί το  $L$  και ο μετασχηματισμός του

$W(L) = \bigcup_{i=1}^m w_i(L)$  να είναι κοντά. Ορίζοντας ως σφάλμα κωδικοποίησης (*encoding error*) το μέγεθος:

$$\epsilon_c = h(L, W(L)) \quad (2.20)$$

και ως σφάλμα αποκωδικοποίησης (*decoding error*) το μέγεθος:

$$\epsilon_d = h(L, A) \quad (2.21)$$

από το θεώρημα προκύπτει ότι σφάλμα αποκωδικοποίησης έχει ένα άνω φράγμα που είναι το:

$$\epsilon_d = \frac{1}{1-s} \epsilon_c \quad (2.22)$$

Από την τελευταία σχέση προκύπτει ότι όσο μεγαλύτερος είναι ο παράγοντας συστολής τόσο μεγαλύτερο είναι και το ανώτατο όριο του σφάλματος αποκωδικοποίησης.

Το θεώρημα που μόλις διατυπώσαμε απετέλεσε το θεμέλιο της συμπίεσης με τη θεωρία των fractal. Οι περισσότερες μέθοδοι συμπίεσης σήμερα είναι *απωλεστικές*, δηλαδή οι αποκωδικοποιημένες εικόνες πρέπει να είναι οπτικά όμοιες, όχι όμως απαραίτητα πανομοιότυπες με τις αρχικές. Υπάρχει λοιπόν πάντοτε ένα είδος ανταλλαγής μεταξύ του βαθμού συμπίεσης και του σφάλματος ανακατασκευής της συμπιεσμένης εικόνας. Στη συμπίεση με βάση τη θεωρία των fractal, ο σκοπός είναι η εύρεση ενός *IFS* το όριο του οποίου θα βρίσκεται πολύ κοντά στην υπό συμπίεση εικόνα.

Με βάση το παραπάνω θεώρημα και τις ιδιότητες των Επαναληπτικών Συστημάτων Συναρτήσεων, η χρήση του *IFS* ως μεθόδου συμπίεσης απαιτεί όσο το δυνατό μικρότερο αριθμό μετασχηματισμών  $w_i \in W$ ,  $i=1, \dots, m$ , ώστε ο χώρος που καταλαμβάνουν στη μνήμη να είναι μικρός, αλλά ταυτόχρονα και το σφάλμα αποκωδικοποίησης να είναι αρκετά μικρό για την εξασφάλιση της υψηλής ποιότητας της συμπίεσης. Η λύση στο γενικό αυτό πρόβλημα αποδείχθηκε χρονοβόρα, αφού για μία δεδομένη εικόνα το πλήθος των

συνόλων  $W = \{w_1, w_2, \dots, w_m\}$  μετασχηματισμών τύπου affine που θα μπορούσαν να έχουν ως όριο μία εικόνα πολύ κοντά στη ζητούμενη, δεν ήταν καθορισμένο και όλες οι πιθανές περιπτώσεις έπρεπε να εξεταστούν.

## 2.4 Συστήματα Επαναληπτικών Συναρτήσεων με Διαίρεση της εικόνας

Όταν βλέπουμε μία εικόνα, η πρώτη μας ενέργεια είναι να αναγνωρίσουμε τα αντικείμενα που περιέχει. Η αναγνώριση των αντικειμένων είναι μία διαδικασία τμηματοποίησης της εικόνας και σύγκρισης των τμημάτων που προκύπτουν με παρόμοιες εικόνες στη μνήμη μας. Για το λόγο αυτό, λέμε ότι μία τέτοια διαδικασία βρίσκει ομοιότητες στο πεδίο του χρόνου. Εκτός όμως από τις ομοιότητες με εικόνες που έχει αποθηκεύσει στο παρελθόν, ο εγκέφαλος μπορεί να ανακαλύψει ομοιότητες και μέσα στην ίδια την εικόνα, πολύ ευκολότερα μάλιστα αν αυτή περιέχει συμμετρίες, ή αντικείμενα που είναι ίδια μεταξύ τους, (σχήματα 2.1, 2.2) μία διαδικασία που εκτελείται στο πεδίο του χώρου.



Σχ.2.2: Χωρική συμμετρία σε μία εικόνα





Σχ.2.3: Ομοιότητα μεταξύ διαφορετικών περιοχών της ίδιας εικόνας. Το είδος αυτό μπρόκολου που εικονίζεται στο σχήμα αποτελεί μία χαρακτηριστική φυσική fractal αναπαράσταση.

Οι ομοιότητες τμημάτων της ίδιας της εικόνας μεταξύ τους παίζουν πολύ σημαντικό ρόλο στην επεξεργασία της. Για παράδειγμα, αν σε μία μονοχρωματική εικόνα  $512 \times 512$  εικονοστοιχείων μειωθεί ο αριθμός των ψηφίων αναπαράστασης της διαβάθμισης του γκρι από 8 σε 7, ο χώρος στη μνήμη μειώνεται κατά  $512 \times 512 \text{ bits} = 32\text{Kb}$ . Αντιθέτως, αν η εικόνα περιέχει μία αντανάκλαση, τα δεδομένα μπορούν να μειωθούν στο μισό.

Αυτές οι ομοιότητες μέσα σε μία εικόνα που παρατηρούνται στις περισσότερες περιπτώσεις, ώθησαν τον *Michael Barnsley* και το μαθητή του *A.Jacquin*<sup>4</sup> στην εισαγωγή της θεωρία των *PIFS* για τη συμπίεση εικόνων. Η μέθοδος των *PIFS* είναι μία επέκταση των *IFS*, η οποία άρει κάποιες από τις δυσκολίες που ανέκυψαν από την προσπάθεια εφαρμογής της τελευταίας στη συμπίεση εικόνων.

Ο κωδικοποιητής λειτουργεί με παρόμοιο τρόπο με την *Διανυσματική Κβαντοποίηση με Αφαίρεση Μέσης Τιμής*. Εδώ το κωδικό βιβλίο των διανυσμάτων της μεθόδου δεν είναι, όπως προαναφέρθηκε, γνωστό εκ των

προτέρων. Αντίθετα, αποτελείται από τετραγωνικές περιοχές της ίδιας της εικόνας, πράγμα που σημαίνει ότι τα διανύσματα αυτά δεν έχουν κανονικοποιηθεί σε μηδενική μέση τιμή συνιστωσών και μοναδιαία διασπορά. Κάθε εικόνα λοιπόν έχει το δικό της κωδικό βιβλίο.

Στη συνέχεια θα δούμε πώς το προηγούμενο παράδειγμα με τον αριθμό  $\pi$ , επεκτείνεται στην πολυδιάστατη περίπτωση με την κωδικοποίηση μιας εικόνας με τη μέθοδο *PIFS*.

Υποθέτουμε ότι η εικόνα τεμαχίζεται σε τετραγωνικές περιοχές διαστάσεων  $4 \times 4$  *εικονοστοιχείων* που δεν καλύπτει η μία την άλλη. Οι περιοχές αυτές αποκαλούνται συνήθως *range περιοχές* ή *περιοχές τύπου R*. Όπως και στην περίπτωση της *Διανυσματικής Κβαντοποίησης με Αφαίρεση Μέσης Τιμής*, η περιοχή τύπου *R* πρέπει να προσεγγιστεί από τη σχέση  $R \approx sD + \mathbf{o1}$ , όπου το διάνυσμα  $D$  αντιστοιχεί με τη σειρά του σε μια τετραγωνική περιοχή  $4 \times 4$  *εικονοστοιχείων* που ανήκει, όπως έχει αναφερθεί στο κωδικό βιβλίο των *κωδικών διανυσμάτων*. Τα κωδικά διανύσματα  $D$  κατασκευάζονται ως εξής: Θεωρούμε μια οποιαδήποτε τετραγωνική περιοχή διαστάσεων  $8 \times 8$  *εικονοστοιχείων* μέσα στην εικόνα. Αυτές οι περιοχές, που δημιουργούνται συνήθως με την ολίσθηση ενός παραθύρου  $8 \times 8$  κατά  $1$  *εικονοστοιχείο* πρώτα οριζόντια και μετά κάθετα (ή το αντίθετο) μέχρι να καλυφθεί ολόκληρη η εικόνα, αποκαλούνται *περιοχές τύπου D* ή *domain περιοχές*. Διαιρούμε στη συνέχεια κάθε μία από τις περιοχές τύπου *D* σε τετραγωνικές περιοχές διαστάσεων  $2 \times 2$  *εικονοστοιχείων* και για κάθε τέτοια περιοχή, παίρνουμε το μέσο όρο των τιμών της έντασης των *εικονοστοιχείων* (*pixel averaging*) ώστε να μειωθεί η διάσταση των *domain* περιοχών στο μισό ( $4 \times 4$ ) και να σχηματιστούν οι τετραγωνικές περιοχές  $D$ . Με βάση τα παραπάνω συμπεραίνουμε ότι τα τετράγωνα  $D$  είναι περιοχές που προέρχονται από την *υποδειγματοληψία* (*sub sampling*) των περιοχών τύπου *D* (Στην εργασία δε γίνεται διάκριση μεταξύ των περιοχών τύπου *D* και των κωδικών διανυσμάτων  $D$  που έχουν προέλθει από την υποδειγματοληψία των πρώτων, εκτός αν η ανάπτυξη κάποιων εννοιών ή

θεωρίας το επιβάλλει). Το σύνολο των περιοχών τύπου  $D$  ονομάζεται *domain pool* και αν επιλέξουμε να το σχηματίσουμε με τον τρόπο που αναφέρουμε στο συγκεκριμένο παράδειγμα παίρνουμε ένα κωδικό βιβλίο με τεράστιο αριθμό στοιχείων, καθώς για εικόνα μεγέθους  $512 \times 512$  η διαδικασία αυτή παράγει  $(512-7)^2 = 255025$  κωδικά διανύσματα. Για να ελαττώσουμε το πλήθος των κωδικών λέξεων σε μία τιμή που να επιτρέπει την πραγματοποίηση υπολογισμών σε μικρό σχετικά χρονικό διάστημα, μπορούμε να θεωρήσουμε ότι το παράθυρο ολισθαίνει κατά  $l$  εικονοστοιχεία, με  $l \geq 1$ . Για  $l=8$  π.χ., παίρνουμε μία *domain pool* από 4096 περιοχές τύπου  $D$  που δεν καλύπτει η μία την άλλη καθώς έχουν διάσταση 8.

Έχοντας σχηματίσει το κωδικό βιβλίο, ο κωδικοποιητής πρέπει να λύσει το επόμενο πρόβλημα, την εύρεση της καλύτερης προσέγγισης  $R \approx sD + o\mathbf{1}$ . Στην κωδικοποίηση με τη μέθοδο *PIFS* οι συντελεστές  $s$  και  $o$  ονομάζονται *αντίθεση* και *φωτεινότητα* αντίστοιχα. Για την εύρεση των βέλτιστων  $s$ ,  $o$  και  $D$ , πρέπει να εξετάσουμε όλες τις κωδικές λέξεις  $D$ . Για κάθε μία από αυτές πρέπει να πρέπει να βρεθούν με τη σειρά τους οι βέλτιστες τιμές αντίθεσης και φωτεινότητας και στη συνέχεια να επιλεγεί εκείνο το διάνυσμα  $D$  που με το ζεύγος του βέλτιστων τιμών αντίθεσης και φωτεινότητας δίνει την καλύτερη προσέγγιση του  $R$ . Όπως και στο παράδειγμα με τον αριθμό  $\pi$ , μπορούμε να αναζητήσουμε για κάθε  $D$  τα βέλτιστα  $s$  και  $o$  μεταξύ τιμών που περιέχονται σε προϋπάρχοντα κωδικά βιβλία για το κάθε μέγεθος. Αν το κωδικό βιβλίο της αντίθεσης περιέχει  $\kappa$  στοιχεία και το αντίστοιχο της φωτεινότητας  $\nu$  στοιχεία τότε για κάθε  $D$  το άθροισμα  $sD + o\mathbf{1}$  πρέπει να υπολογιστεί  $\kappa \nu$  φορές. Κάτι τέτοιο όμως είναι υπολογιστικά δυνατό για πολύ μικρά βαθμωτά κωδικά βιβλία, αφού η όλη διαδικασία κοστίζει σε χρόνο. Για την εξοικονόμηση χρόνου, λύνουμε το πρόβλημα της προσέγγισης με αναλυτικές μεθόδους, βρίσκουμε δηλαδή το ελάχιστο της απόστασης:

$$E(D, R) = \min_{s, o} \|R - (sD + o\mathbf{1})\| \quad (2.23)$$

Χρησιμοποιώντας την Ευκλείδεια νόρμα ως συνάρτηση απόστασης και τη μέθοδο των ελαχίστων τετραγώνων, οι βέλτιστοι συντελεστές  $s^*$  και  $o^*$  δίνονται από τις σχέσεις:

$$s^* = \frac{\langle R, D \rangle - \frac{1}{k} \langle R, \mathbf{1} \rangle \langle D, \mathbf{1} \rangle}{\|D\|_2^2 - \frac{1}{k} \langle D, \mathbf{1} \rangle^2} \quad (\text{i}), \quad o^* = \frac{1}{k} (\langle R, \mathbf{1} \rangle - s^* \langle D, \mathbf{1} \rangle) \quad (\text{ii}) \quad (2.24)$$

, όπου το  $\langle \cdot, \cdot \rangle$  είναι το σύμβολο του εσωτερικού γινομένου, το  $\mathbf{1}$  συμβολίζει το μοναδιαίο διάνυσμα και  $k$  είναι η διάσταση του  $R$  και κατά συνέπεια όλων των διανυσμάτων, η οποία στο παράδειγμά που παρατίθεται είναι ίση με 16. Αν ο παρανομαστής στη σχέση (2.23i) είναι ίσος με 0, τότε  $s^* = 0$  και  $o^* = \frac{1}{k} \langle R, \mathbf{1} \rangle$ . Αυτή η διαδικασία δίνει 2 πραγματικούς αριθμούς  $s$  και  $o$ .

Επειδή η κωδικοποίηση επιβάλλει οι τιμές των  $s$  και  $o$  να είναι διακριτές (καθώς για την αναπαράστασή τους δεσμεύεται πεπερασμένος αριθμός ψηφίων), οι παραπάνω τιμές διακριτοποιούνται συνήθως με *ομοιόμορφη βαθμωτή διακριτοποίηση* που ισοδυναμεί με διαδικασία στρογγυλοποίησης (Το πρόβλημα της διακριτοποίησης και αποθήκευσης των παραμέτρων του μετασχηματισμού μελετάται εκτενώς σε επόμενο κεφάλαιο). Για να είναι συστολικός ο μετασχηματισμός πρέπει η αντίθεση, η οποία ισοδυναμεί με τον παράγοντα συστολής να είναι απολύτως μικρότερη της μονάδας. Παρ' όλα αυτά, είναι δυνατό να προκύψουν τιμές της αντίθεσης μεγαλύτερες απολύτως της μονάδας. Στην περίπτωση αυτή, και με την προϋπόθεση ότι για το σύνολο των τύπου  $R$  περιοχών ο αριθμός των τελευταίων που κωδικοποιούνται με αντίθεση μεγαλύτερη της μονάδας είναι πολύ μικρότερος σε σχέση με αυτές που κωδικοποιούνται με αντίθεση μικρότερη της μονάδας, επιτρέπουμε την κωδικοποίηση και με τιμές  $|s| > 1$ .

Τέλος, αφού η θεωρία των *PIFS* στηρίζεται στους μετασχηματισμούς τύπου affine, πρέπει να λάβουμε υπ' όψη ότι δεν είναι μόνο ο μετασχηματισμός

$sD + o\mathbf{1}$  τύπου affine, αλλά και οι ισομετρικοί. Όταν χρησιμοποιούμε την Ευκλείδεια νόρμα ως συνάρτηση απόστασης, πράγμα το οποίο συμβαίνει στην πλειονότητα των περιπτώσεων, κάθε υποψήφια περιοχή τύπου  $D$  υπόκειται στους  $\delta$  ισομετρικούς μετασχηματισμούς που έχουμε προαναφέρει πριν υποστεί υποδειγματοληψία και συγκριθεί στη συνέχεια (με την εφαρμογή του  $sD + o\mathbf{1}$ ) με την εκάστοτε περιοχή τύπου  $R$ . Αν λοιπόν ο αριθμός των περιοχών τύπου  $D$  που εξάγονται από την εικόνα (χωρίς την εφαρμογή ισομετρικών μετασχηματισμών) είναι  $n_D$ , τότε το πλήθος των στοιχείων του συνόλου των περιοχών τύπου  $D$ , συνυπολογίζοντας και τις ισομετρίες, είναι  $\delta * n_D$  με αποτέλεσμα για κάθε περιοχή τύπου  $R$ , η εύρεση των βέλτιστων  $s, o, D$  να απαιτεί  $\delta * n_D$  συγκρίσεις.

Συνοψίζοντας, ο βασικός αλγόριθμος της θεωρίας *PIFS* για την κωδικοποίηση εικόνας με χρήση τετραγωνικών περιοχών σταθερού μεγέθους, αποτελείται από τα παρακάτω βήματα:

1. *Διαμέριση της εικόνας.* Τεμαχίζουμε την εικόνα σε τετραγωνικές περιοχές σταθερών διαστάσεων, π.χ.  $4*4$ . Οι περιοχές που προκύπτουν καλούνται range περιοχές  $R_r$ .
2. *Domain pool και κωδικό βιβλίο.* Διασχίζοντας την εικόνα με ένα βήμα από  $l$  pixels οριζόντια και κάθετα δημιουργούμε μία λίστα από περιοχές τύπου  $D$  (domain pool) που έχουν το διπλάσιο μέγεθος από τις περιοχές τύπου  $R$  και τις οποίες μετασχηματίζουμε με βάση τους  $\delta$  ισομετρικούς μετασχηματισμούς δημιουργώντας μία διευρυμένη domain pool. Παίρνοντας το μέσο όρο της διαβάθμισης του γκρι τεσσάρων γειτονικών εικονοστοιχείων μέσα στις περιοχές τύπου  $D$  τις συρρικνώνουμε ώστε το μέγεθός τους να ταυτίζεται με αυτό των περιοχών τύπου  $R$ . Με τον τρόπο αυτό κατασκευάζουμε το κωδικό βιβλίο των τετραγωνικών περιοχών  $D_r$ .
3. *Η αναζήτηση.* Για κάθε περιοχή τύπου  $R$  βρίσκεται μία βέλτιστη προσέγγιση  $R \approx sD + o\mathbf{1}$  σύμφωνα με τα ακόλουθα βήματα:

- a) Για κάθε κωδική λέξη  $D_i$  υπολογίζουμε τη βέλτιστη αυτή προσέγγιση σε τρία βήματα:
- I. Εκτελούμε τη βελτιστοποίηση με βάση τη μέθοδο των ελαχίστων τετραγώνων χρησιμοποιώντας τους τύπους (2.24.i) και (2.24.ii) που δίνουν δύο πραγματικούς συντελεστές την αντίθεση  $s$  και τη φωτεινότητα  $o$ .
  - II. Διακριτοποιούμε τους συντελεστές αυτούς, χρησιμοποιώντας για παράδειγμα τη μέθοδο ομοιόμορφης διακριτοποίησης.
  - III. Χρησιμοποιώντας τους κβαντισμένους συντελεστές  $s$  και  $o$  υπολογίζουμε το σφάλμα  $E(R, D_i)$ .
- b) Ανάμεσα από όλα τα κωδικά διάνυσμα  $D_i$ , βρίσκουμε εκείνο το διάνυσμα  $D_k$  που ελαχιστοποιεί το σφάλμα  $E$ , δηλαδή  $E(R, D_k) = \min_i E(R, D_i)$ .
- c) Αποθηκεύουμε τον κώδικα για την περιοχή που εξετάζουμε, ο οποίος αποτελείται από τους δείκτες για τους διακριτοποιημένους συντελεστές  $s$  και  $o$  καθώς και το δείκτη  $k$  στη θέση της βέλτιστης κωδικής λέξης  $D_k$  στο κωδικό βιβλίο.

Ο ψευδοκώδικας του αλγόριθμου της κωδικοποίησης όπως και ο αντίστοιχος του αλγόριθμου της αποκωδικοποίησης, που περιγράφεται στη συνέχεια παρατίθενται στο Παράρτημα Β.

Η αποκωδικοποίηση γίνεται ως εξής:

Αρχίζοντας από μία αυθαίρετη εικόνα δημιουργούμε ξανά τα δύο σύνολα, των περιοχών τύπου  $R$  και  $D$ . Για κάθε περιοχή τύπου  $R$  βρίσκουμε τη θέση της αντίστοιχης περιοχής τύπου  $D$  στην αρχική εικόνα με βάση την τιμή του δείκτη την οποία έχουμε αποθηκεύσει κατά την κωδικοποίηση. Στη συνέχεια και αφού η τελευταία υποβληθεί σε υποδειγματοληψία, εφαρμόζουμε το μετασχηματισμό τύπου affine στο διάνυσμα  $D$  με τις τιμές των  $s$  και  $o$  που αντιστοιχούν στη διεύθυνση της περιοχής τύπου  $D$  από την οποία προήλθε. Η παραπάνω εργασία εκτελείται για όλες τις περιοχές τύπου  $R$  της αρχικής

εικόνας. Το αποτέλεσμα είναι μία νέα εικόνα, διαφορετική από την αρχική, καθώς τη θέση των περιοχών τύπου R της αρχικής έχουν καταλάβει οι μετασχηματισμένες περιοχές τύπου D. Στη συνέχεια υπολογίζουμε τη διαφορά της νέας εικόνας από την αρχική και τη συγκρίνουμε με ένα κατώφλι E το οποίο έχουμε επιλέξει τόσο μικρό, ώστε η ποιότητα της αποκωδικοποιημένης εικόνας να θεωρείται αποδεκτή. Αν η διαφορά των δύο εικόνων είναι μεγαλύτερη του E, τότε επαναλαμβάνουμε τη συστολική διαδικασία που περιγράψαμε για την τελευταία και τη συγκρίνουμε με την νέα εικόνα που προκύπτει, έως ότου η διαφορά των δύο να είναι μικρότερη του E ή ο αριθμός επαναλήψεων ξεπεράσει την τιμή πέρα από την οποία έχουμε προαποφασίσει ότι θα σταματά η όλη διαδικασία.

Όπως προαναφέρθηκε, το μέγεθος των περιοχών τύπου D επιλέγεται συνήθως διπλάσιο του μεγέθους των περιοχών τύπου R. Οι παράμετροι όμως της αντίθεσης και της φωτεινότητας που συνιστούν το συστολικό μετασχηματισμό για κάθε ζεύγος R-D περιοχών, δεν σχετίζονται με γεωμετρική μεταβολή (σμίκρυνση) της τύπου D (μεγαλύτερης) περιοχής, αλλά με τη μεταβολή των τιμών της έντασής των εικονοστοιχείων της. Επομένως μπορούμε να χρησιμοποιήσουμε περιοχές τύπου D οποιουδήποτε μεγέθους, για παράδειγμα μεγέθους ίδιου με αυτό των περιοχών τύπου R. Μολονότι κάτι τέτοιο είναι εφικτό οι πειραματικές μελέτες έχουν δείξει ότι όταν η γεωμετρική κλίμακα είναι πολύ μικρή η διασπορά του σφάλματος κατά την αποκωδικοποίηση είναι χειρότερη. Για το λόγο αυτό καθώς και για διευκόλυνση στους υπολογισμούς επιλέγουμε μείωση της διάστασης των περιοχών τύπου D στο μισό, γεγονός που δίνει τα καλύτερα αποτελέσματα «(από αισθητικής πλευράς)».

Ένα δεύτερο σχόλιο πάνω στο βασικό αλγόριθμο αφορά τους ισομετρικούς μετασχηματισμούς των τετραγωνικών περιοχών. Αυξάνοντας το σύνολο των περιοχών τύπου D κατά οκτώ φορές αυξάνουμε και τις πιθανότητες μίας καλύτερης προσέγγισης μεταξύ R-D περιοχών κατά την κωδικοποίηση με

αποτέλεσμα και την απόκτηση καλύτερων καμπυλών σηματοθορυβικού λόγου-βαθμού συμπίεσης. Ωστόσο, σύμφωνα με τις έρευνες του *Saupe*<sup>5</sup>, την ίδια ποιότητα και ακόμη υψηλότερη, μπορούμε να αποκτήσουμε με τη μείωση του βήματος  $l$  με το οποίο το τύπου D παράθυρο «σαρώνει» την εικόνα. Συνεπώς κατά το συγγραφέα, η πολυπλοκότητα που εισάγει ο υπολογισμός των ισομετρικών μετασχηματισμών μπορεί να αποφευχθεί.

Έστω  $d$  το μέγεθος σε εικονοστοιχεία της περιοχής τύπου R,  $CS$  το μέγεθος του κωδικού βιβλίου (ο αριθμός δηλαδή των διανυσμάτων που περιέχει) και  $\#s$  και  $\#o$  οι αριθμοί των ψηφίων που έχουν δεσμευθεί για την απεικόνιση των διακριτών τιμών της αντίθεσης και της φωτεινότητας. Έστω ακόμη  $r$  το πλήθος των περιοχών τύπου R και ότι συμπεριλαμβάνονται στην κωδικοποίηση και οι ισομετρικές. Το μέγεθος επομένως της αρχικής εικόνας στη μνήμη, για την οποία έχουν δεσμευθεί  $\delta$  ψηφία για την αναπαράσταση της διαβάθμισης του γκρι, είναι  $r \times d \times \delta$ . Όπως τονίστηκε προηγουμένως, έχει προιαθοριστεί ο αριθμός των ψηφίων για την αποθήκευση των βαθμωτών παραμέτρων του μετασχηματισμού και της διεύθυνσης του κάθε κωδικού διανύσματος στο κωδικό βιβλίο για κάθε περιοχή τύπου R. Για όλες λοιπόν τις τύπου R περιοχές, παίρνουμε το μέγεθος στη μνήμη της κωδικοποιημένης εικόνας που είναι ίσο με  $r \times (\#s + \#o + \lceil \log_2(8 \times CS) \rceil)$ , όπου το  $\lceil \cdot \rceil$  συμβολίζει τη στρογγυλοποίηση. Σύμφωνα με τον ορισμό του βαθμού συμπίεσης που δόθηκε στο προηγούμενο κεφάλαιο, ο βαθμός συμπίεσης της βασικής μεθόδου είναι:

$$\text{compression ratio} = \frac{8 \times d}{\#s + \#o + \lceil \log_2(8 \times CS) \rceil} \quad (2.25)$$

Ο πίνακας 2.2 δίνει κάποια αριθμητικά παραδείγματα της απόδοσης της απλής μεθόδου για διαφορετικό μέγεθος περιοχής τύπου R για την εικόνα της Lena που φαίνεται στο σχήμα 2.4 (Οι δοκιμαστικές εικόνες που χρησιμοποιούν οι περισσότεροι ερευνητές και που αποτελέσματα της επεξεργασία τους περιλαμβάνονται στην εργασία, βρίσκονται στο Παράρτημα A).

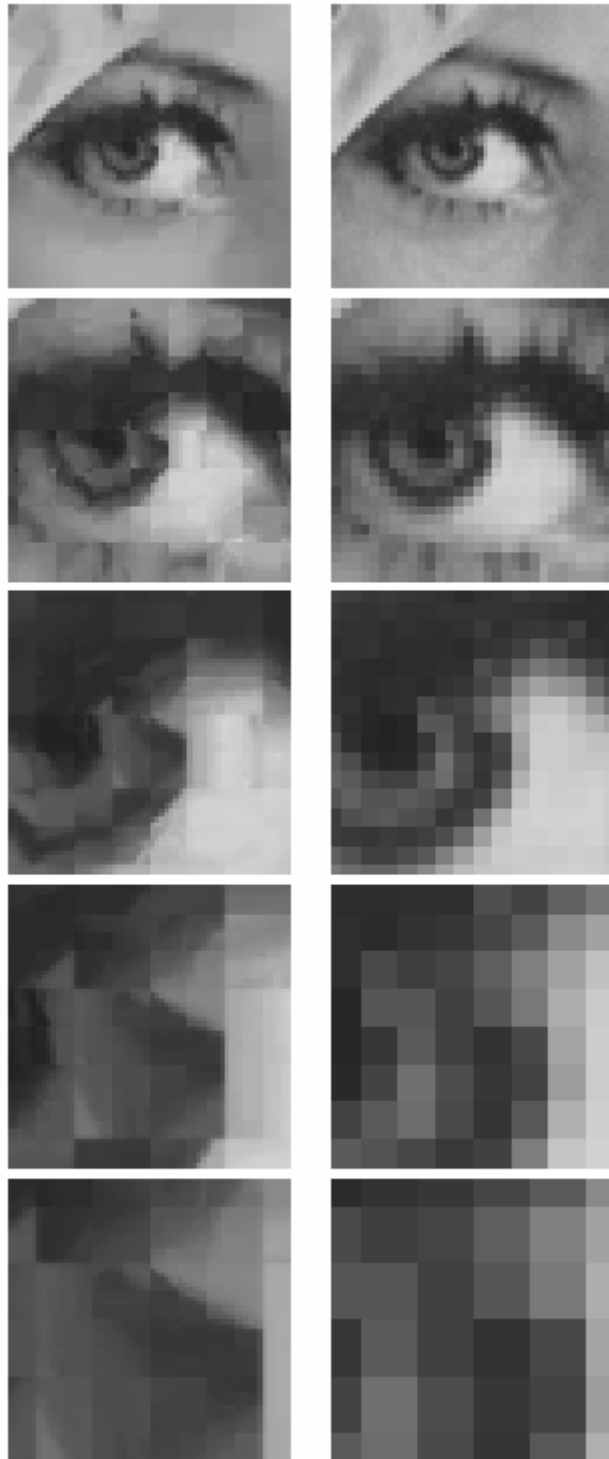


Πίνακας 2.2: Παραδείγματα επίδοσης του βασικού αλγορίθμου για την εικόνα της Lena

Μέγεθος περιοχής Τύπου R	PSNR (dB)		Χρόνος Κωδικοποίησης (sec)	Βαθμός Συμπίεσης
	Σφάλμα Προσέγγισης (collage)	Σφάλμα Μετ/σμού (attractor)		
4*4	36.96	36.66	147.48	4.4
8*8	31.15	31.27	69.93	17.7
16*16	27.02	26.89	59.61	70.5
32*32	23.55	23.32	54.76	281.0



Σχ.2.4:Lena



Σχ.2.5: Τα Fractal απέναντι στην παραδοσιακή μεγέθυνση. Από την αποκωδικοποίηση της εικόνας της Lena (PSNR=34.3dB, compression ratio=14.16), αποκωδικοποιούμε την εικόνα σε κλίμακες μεγέθυνσης 1, 2, 4, 8 και 12. Στην αριστερή στήλη εικονίζεται ένα τμήμα των αποτελεσμάτων. Η δεξιά παρουσιάζει την ίδια ακολουθία μεγεθύνσεων εφαρμοσμένη στην αρχική εικόνα.

Ένα ιδιαίτερο χαρακτηριστικό της συμπίεσης με fractal, είναι το γεγονός ότι οι εικόνες που παίρνουμε περιγράφονται ως το όριο ενός τελεστή που εφαρμόζεται επαναληπτικά σε μία εικόνα, χωρίς να γίνεται αναφορά στην

κλίμακα της εικόνας ή στο μέγεθός της ως προς τον αριθμό των *εικονοστοιχείων*. Κατά συνέπεια, ο fractal κώδικας μπορεί να αποκωδικοποιηθεί σε οποιαδήποτε ανάλυση παρουσιάζοντας λεπτομέρειες της εικόνας σε οποιαδήποτε κλίμακα. Η λεπτομέρεια όμως αυτή που παράγεται κατά την αποκωδικοποίηση σε μεγαλύτερη κλίμακα είναι τεχνητή και δεν αναπαριστά κάποια πραγματική λεπτομέρεια της αρχικής εικόνας (σχ. 2.5). Από την άλλη πλευρά, το χαρακτηριστικό αυτό της ανεξαρτησίας από την ανάλυση αποβαίνει ωφέλιμο για δύο λόγους:

1. Οι τεχνητές λεπτομέρειες της εικόνας, εξαιτίας της αυτο-αναφορικής δομής του κώδικα, βρίσκονται σε συνάφεια με τη συνολική εμφάνιση των αντικειμένων που περιέχει. Οι fractal εικόνες φαίνονται πιο «φυσικές» από εικόνες που προκύπτουν με μία απλή αντιγραφή κάποιων *εικονοστοιχείων* κατά την αποσυμπίεση ή με παρεμβολή.
2. Η μέθοδος μπορεί να χρησιμοποιηθεί και ως εργαλείο ενίσχυσης της λεπτομέρειας μιας εικόνας. Μια «φτωχή» σε ανάλυση εικόνα, μπορεί να κωδικοποιηθεί με τη χρήση των fractal και στη συνέχεια αποκωδικοποιώντας τη σε υψηλότερη ανάλυση να αποκτησουμε μία ενισχυμένη σε λεπτομέρεια έκδοσή της.

Η μέθοδος που παρουσιάστηκε στο κεφάλαιο αυτό αποτελεί την πιο απλή έκδοχή της συμπίεσης με βάση τη θεωρία των fractal. Υπάρχουν πολλά ζητήματα που πρέπει να μελετηθούν λεπτομερέστερα, ώστε να κάνουμε λόγο για μία μέθοδο ικανή να παράγει υψηλής ποιότητας συμπίεση, όπως είναι η διαμέριση της εικόνας, η αποδοτική αποθήκευση των fractal παραμέτρων, η μείωση του μεγέθους των υπολογισμών κ.ά. Τα παραπάνω ζητήματα είναι το αντικείμενο των κεφαλαίων που ακολουθούν.

## Βιβλιογραφία

1. Dietmar Saupe, Raouf Hamzaoui, Hannes Haternstein, *Fractal Image Compression, An Introductory Overview*, in: *Fractal Models for Image Synthesis, Compression and Analysis*, D. Saupe, J. Hart (eds.), ACM SIGGRAPH'96 Course Notes 27, New Orleans, Louisiana, Aug. 1996
2. Ning Lu, *Fractal Imaging*, Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto.
3. M. F. Barnsley, *Fractals Everywhere*, Academic Press, 1988, 1992.
4. A. E. Jacquin, *A Fractal Theory of Iterated Markov Operators with Applications to Digital Image Coding*, Ph.D. thesis, Georgia Tech, 1989.
5. Saupe, D., *The futility of square isometries in fractal image compression*, in: *Proc. ICIP-96 IEEE International Conference on Image Processing*, Lausanne, Sept. 1996.

---

## ΚΕΦΑΛΑΙΟ 3

---

### Η Διαμέριση της εικόνας

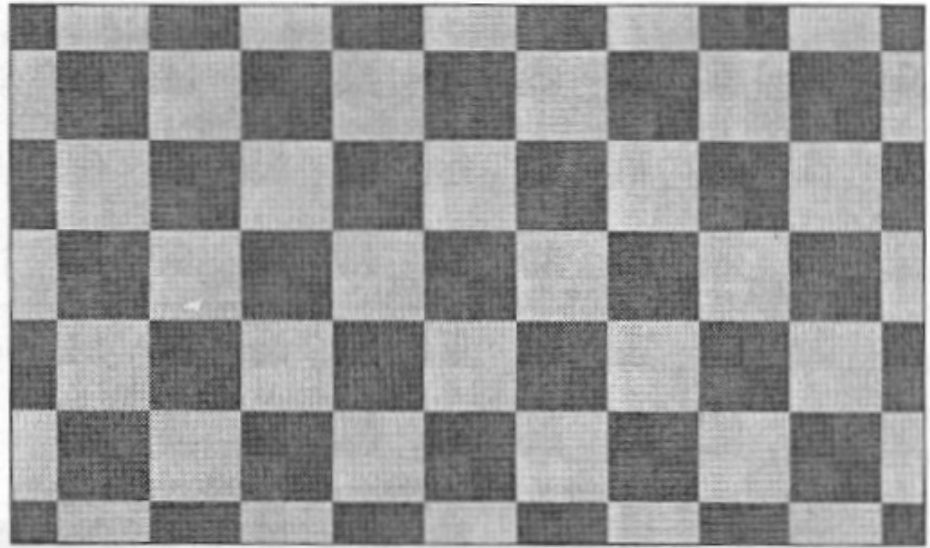
Υπάρχουν πολλές μέθοδοι διαμέρισης του πλαισίου της εικόνας. Τι χαρακτηρίζει όμως μία καλή διαίρεση για τη συμπίεση με τη θεωρία των fractal; Η απάντηση είναι ότι πρέπει να διαιρεί την εικόνα σε περιοχές που να μοιάζουν με άλλες περιοχές της ίδιας εικόνας. Όπως γνωρίζουμε ο κώδικας fractal αποτελείται από πληροφορίες για τη διαίρεση καθώς και τους συντελεστές του μετασχηματισμού τύπου affine. Πρέπει να είμαστε έτοιμοι να αποδεχθούμε ένα πρόσθετο υπολογιστικό κόστος που ενδεχομένως να επιφέρει η εφαρμογή τεχνικών ακανόνιστης διαίρεσης της εικόνας, αν αυτές οδηγούν σε καλύτερα αποτελέσματα όσον αφορά τις καμπύλες βαθμού συμπίεσης – PSNR. Η διαίρεση σε περιοχές σταθερού σχήματος και μεγέθους π.χ. τετράγωνα, τρίγωνα, ορθογώνια παραλληλόγραμμα, δεν απαιτεί πρόσθετη πληροφορία για τον τρόπο διαίρεσης, ο οποίος εξαιτίας της απλότητας του παραλείπεται με συνέπεια όλος ο κώδικας να συμπυκνώνεται στις παραμέτρους του μετασχηματισμού. Το μειονέκτημα όμως αυτών των τρόπων διαίρεσης είναι η αδυναμία προσαρμογής τους στο περιεχόμενο της εικόνας. Η αντίθετη προσέγγιση αντιπροσωπεύεται από τεχνικές διαίρεσης με βάση το περιεχόμενο της εικόνας. Η μερίδα του λέοντος της πληροφορίας πηγαίνει στο κώδικα της διαμέρισης. Όπως είναι αναμενόμενο, η βέλτιστη προσέγγιση βρίσκεται κάπου ανάμεσα στις δύο αυτές μορφές διαίρεσης. Πριν όμως τις εξετάσουμε λεπτομερώς, θα συζητήσουμε τα πλεονεκτήματα και μειονεκτήματα του χωρισμού της εικόνας σε περιοχές που επικαλύπτουν η μία την άλλη (*overlapped covering*) σε σχέση με το χωρισμό της σε μη επικαλυπτόμενες περιοχές. Στο τέλος του κεφαλαίου, θα παρουσιαστεί περιληπτικά και μία μέθοδος που στηρίζει ως επί το πλείστον τη διαμέριση και την όλη κωδικοποίηση στη μορφολογία της εικόνας.

### 3.1 Χωρισμός της εικόνας σε επικαλυπτόμενες περιοχές

Όπως είναι γνωστό, για την αναπαράσταση της εικόνας, η ένωση όλων των περιοχών τύπου  $R$ , επικαλυπτόμενων ή μη, πρέπει να την διατρέχει ολόκληρη. Οποιαδήποτε κάλυψη μπορεί να καταλήξει σε διαμέριση καθώς αν δύο τύπου  $R$  περιοχές επικαλύπτονται, εξετάζουμε για την τομή τους δύο κώδικες fractal που τις αντιπροσωπεύουν και επιλέγουμε τον ανώτερο. Η περιοχή της οποία ο κώδικας είναι ανώτερος προσαρτά εξολοκλήρου την τομή ενώ η άλλη ελαττώνεται, ώστε να αποφευχθεί η επικάλυψη. Με τον τρόπο αυτό αυξάνεται η πιθανότητα να βρεθεί ένας καλύτερος μετασχηματισμός για την περιοχή που έχει ελαττωθεί σε μέγεθος. Μολονότι δεν υπάρχει λόγος να αναπαριστούμε το ίδιο μέρος της εικόνας δύο φορές και ενώ οι περισσότεροι ερευνητές έχουν υιοθετήσει την τεχνική της διαμέρισης, όπως και στην εργασία αυτή, η επικάλυψη της εικόνας έχει με τη σειρά της ξεχωριστή ιδιότητα<sup>1</sup>.

Για να το διαπιστώσουμε, παραθέτουμε το ακόλουθο παράδειγμα:

Αρχίζουμε από μία τετραγωνική διαμέριση (*square partitioning*)  $8*8$  την οποία ενώνουμε με άλλη μία διαμέριση της ίδιας εικόνας  $8*8$  όπου τα τετράγωνα έχουν ολισθήσει κατά 4 εικονοστοιχεία προς τα κάτω και 4 προς τα δεξιά σε σχέση με την πρώτη. Προκύπτει μία κάλυψη επικαλυπτόμενων περιοχών τύπου  $R$  όπως φαίνεται στο σχήμα 3.1, στην οποία τα περισσότερα σημεία περιέχονται δύο φορές. Ποια από τις δύο θα επιλέξουμε;



Σχ.3.1: Χωρισμός της εικόνας σε τετράγωνα 8\*8 που επικαλύπτονται.

Η απάντηση είναι και τις δύο αφού εφαρμόσουμε μία σταθερή συνάρτηση επικάλυψης η οποία φαίνεται στο σχήμα 3.2 και μοιάζει με ένα χαμηλοδιαβατό φίλτρο. Κάθε τιμή της συνάρτησης που αντιστοιχεί σε μία τετραγωνική περιοχή, πολλαπλασιάζεται με την τιμή της έντασης των *εικονοστοιχείων* της. Το αποτέλεσμα της εφαρμογής της συνάρτησης στην ένωση των δύο, θα είναι το άθροισμα των αποτελεσμάτων από την εφαρμογή της στην επιμέρους διαμέριση. Το πλεονέκτημα της μεθόδου αυτής είναι η ανυπαρξία ανεπιθύμητων αντικειμένων, για παράδειγμα αναγνωρίσιμες ακμές ανάμεσα στις ακμές περιοχών τύπου  $R$  των δύο διαφορετικών διαμερίσεων, οι οποίες γίνονται ευκολότερα παρατηρήσιμες με την αύξηση του μεγέθους των περιοχών τύπου  $R$ .

0	1/8	1/4	1/2	1/2	1/4	1/8	0
1/8	1/4	1/2	3/4	3/4	1/2	1/4	1/8
1/4	1/2	3/4	7/8	7/8	3/4	1/2	1/4
1/2	3/4	7/8	1	1	7/8	3/4	1/2
1/2	3/4	7/8	1	1	7/8	3/4	1/2
1/4	1/2	3/4	7/8	7/8	3/4	1/2	1/4
1/8	1/4	1/2	3/4	3/4	1/2	1/4	1/8
0	1/8	1/4	1/2	1/2	1/4	1/8	0

Σχ.3.2: Συνάρτηση μάσκας για τετράγωνο 8\*8.

Χρησιμοποιώντας τη μέθοδο αυτή φαίνεται ότι αποθηκεύουμε διπλάσιο αριθμό παραμέτρων για την αναπαράσταση των περιοχών τύπου R. Από την άλλη πλευρά, δε χρειάζεται να αποθηκεύσουμε τη μετατόπιση της έντασης ή τη μέση τιμή για κάθε περιοχή τύπου R για τη δεύτερη διαμέριση, γεγονός που εξοικονομεί το 1/3 της πληροφορίας για κάθε fractal κώδικα. Αυτό οφείλεται στο γεγονός ότι η μέση τιμή ή μετατόπιση της φωτεινότητας (παράμετροι του συστολικού μετασχηματισμού για τις οποίες γίνεται εκτενής αναφορά στο επόμενο κεφάλαιο), μπορούν να υπολογιστούν κατά προσέγγιση από τις αντίστοιχες τιμές που έχουν αποθηκευθεί για την πρώτη διαμέριση, καθώς η δεύτερη δεν αποτελεί τίποτε άλλο παρά ολίσθηση της πρώτης. Κατά συνέπεια, αν οι fractal παράμετροι μιας διαμέρισης 8\*8 απαιτούν  $\beta$  bpp (*bits per pixel*), τότε στην περίπτωση της τμηματοποίησης με επικαλυπτόμενες περιοχές τύπου R, ο αριθμός των bpp είναι 5/3, μέγεθος συγκρίσιμο με αυτό της τετραγωνικής διαμέρισης 6\*6 που απαιτεί 64/36 bpp.





6\*6 τετραγωνική διαμέριση



Χωρισμός σε επικαλυπτόμενες τετραγωνικές περιοχές 8\*8

Σχ.3.3: Αναπαράσταση της Lena με διαφορετικό τρόπο χωρισμού της εικόνας

Το επόμενο βήμα είναι η σύγκριση της 6\*6 διαμέρισης με την επικαλυπτόμενη τμηματοποίηση 8\*8 που έχουμε εξετάσει σε αυτήν την ενότητα. Το σχήμα 3.3 δείχνει ότι η πρώτη παράγει μια εικόνα ακριβέστερη στη λεπτομέρεια από αυτή που μας δίνει η δεύτερη. Μια κοντινότερη ματιά όμως (σχ. 3.4), αποδεικνύει ότι η δεύτερη εικόνα είναι σαφώς πιο ευχάριστη χωρίς τα οπτικά εμπόδια (π.χ. ανεπιθύμητες ακμές) που εμφανίζει η πρώτη.



6\*6 τετραγωνική διαμέριση



Χωρισμός σε επικαλυπτόμενες τετραγωνικές περιοχές 8\*8

Σχ.3.4: Μία μεγέθυνση του σχήματος 3.3

Όπως αναφέρθηκε και προηγουμένως, η διαμέριση είναι το τελευταίο στάδιο της κάλυψης της εικόνας. Όταν γίνεται ακολουθώντας κάποιο γεωμετρικό

σχέδιο ανεξάρτητο της μορφής της εικόνας η κωδικοποίηση είναι απλούστερη και η αποκωδικοποίηση ευκολότερη. Όταν λαμβάνεται σοβαρά υπόψη το περιεχόμενο της κάθε εικόνας, η συμπίεση και αποσυμπίεση καθίστανται χρονοβόρες. Στις ενότητες που ακολουθούν μελετούμε δύο κατηγορίες διαίρεσης που συνδυάζουν σε ένα βαθμό τόσο την απλότητα όσο και την προσαρμοστικότητα: τις *Ιεραρχικές (Hierarchical)* διαιρέσεις και τις μεθόδους *διαίρεσης-και-συγχώνευσης (split-and-merge)*<sup>2</sup>.

## 3.2 Ιεραρχικές Μέθοδοι Διαμέρισης

Στην κατηγορία αυτή ανήκουν δύο από τις πιο γνωστές μεθόδους, το *τετραδικό δένδρο* και η *οριζόντια-κάθετη διαίρεση*.

### 3.2.1 Η διαμέριση με βάση το τετραδικό δένδρο

Ο προσαρμοστικός τρόπος διαίρεσης της εικόνας μπορεί να παρουσιάσει σημαντικά πλεονεκτήματα σε σχέση με τη διαίρεσή της σε περιοχές τύπου R σταθερού μεγέθους. Μπορεί π.χ. μέσα σε μία εικόνα να υπάρχουν περιοχές μεγέθους μεγαλύτερου από αυτό μίας τυπικής περιοχής τύπου R και οι οποίες να παρουσιάζουν τέτοια ομοιογένεια ώστε η κωδικοποίησή τους να δίνει μικρό collage σφάλμα, ενώ στην ίδια εικόνα να απαντώνται και περιοχές με μεγάλες αντιθέσεις με συνέπεια να απαιτείται τμηματοποίησή τους ώστε η κωδικοποίηση να διατηρεί την επιθυμητή ποιότητα. Μία πρώτη προσέγγιση (που είχε ήδη εισαχθεί από τον Jacquin) ήταν η διαίρεση της εικόνας σε τετραγωνικές περιοχές μεταβλητού μεγέθους παραδείγματος χάρη πλάτους 4, 8, ή και 16 εικονοστοιχείων. Αυτή η ιδέα οδήγησε στη γενικότερη μέθοδο *διαμέρισης του τετραδικού δένδρου (quadtree partition)*. Σε αντίθεση με τις κωδικοποιήσεις που χρησιμοποιούν τετράγωνα σταθερού μεγέθους, το αρχείο εξόδου στη μέθοδο του δένδρου πρέπει να περιέχει και τον τύπο του δένδρου που χρησιμοποιήθηκε κατά την κωδικοποίηση.

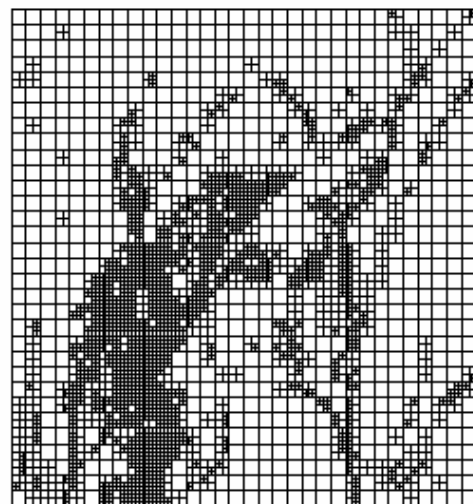
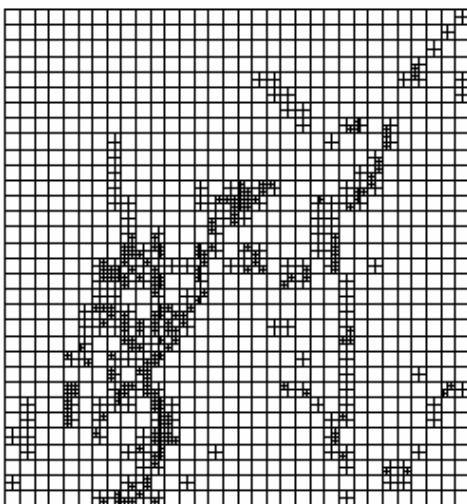
Η χρήση μεταβλητής διαίρεσης μας δίνει την δυνατότητα να σχεδιάσουμε ένα πρόγραμμα κωδικοποίησης που παρέχει μεταβλητές καμπύλες ποιότητας βαθμού συμπίεσης. Ο χρήστης μπορεί να καθορίζει κάθε φορά ποιος είναι ο στόχος του, είτε η ποιότητα της εικόνας, είτε ο βαθμός συμπίεσης. Ο κωδικοποιητής στη συνέχεια μπορεί να διαιρεί την εικόνα επαναληπτικά σε κατάλληλα κομμάτια έως ότου ένα από τα δύο γενικά αυτά κριτήρια να ικανοποιηθεί. Στη συνέχεια περιγράφουμε τον αλγόριθμο αυτό όταν το κριτήριο είναι η υψηλή πιστότητα της εικόνας κατά την αποκωδικοποίηση:

1. Καθορίζουμε μια ανοχή για το μέσο τετραγωνικό σφάλμα  $E(R,D)/\sqrt{\#pixels}$  στο  $R$  της αντιστοιχησης και ένα μέγιστο και ελάχιστο μέγεθος για την περιοχή τύπου  $R$ .
2. Αρχικοποιούμε μια στοίβα από περιοχές τύπου  $R$  τοποθετώντας σε αυτή τις περιοχές μεγίστου μεγέθους.
3. Όταν η στοίβα δεν είναι άδεια ακολουθούμε τα παρακάτω βήματα:
  - a) Παίρνουμε μία περιοχή τύπου  $R$  από τη στοίβα και αναζητούμε στο αντίστοιχο σύνολο των περιοχών τύπου  $D$ , εκείνη που δίνει το ελάχιστο σφάλμα μεταξύ  $R$  και  $sD+o1$ .
  - b) Αν το μέσο τετραγωνικό σφάλμα είναι μικρότερο της ανοχής, ή αν το μέγεθος της περιοχής τύπου  $R$  είναι ίσο με το ελάχιστο, τότε αποθηκεύουμε τον κώδικα για την περιοχή τύπου  $R$ , δηλαδή την αντίθεση, την φωτεινότητα και τη διεύθυνση της περιοχής τύπου  $D$ . Αν η αντίθεση είναι ίση με το 0 τότε δεν αποθηκεύουμε τα παραπάνω.
  - c) Διαφορετικά διαιρούμε την περιοχή  $R$  σε τριτατημόρια και τα τοποθετούμε στη στοίβα.

Χρησιμοποιώντας διαφορετικές ανοχές για την ποιότητα της προσέγγισης μπορούμε να αποκτήσουμε μια σειρά από κώδικες μεταβλητού βαθμού συμπίεσης και πιστότητας. Στο σχήμα 3.5 φαίνονται δύο τέτοιου είδους κωδικοποιήσεις με βάση το τετραδικό δένδρο που συνοδεύονται από δύο

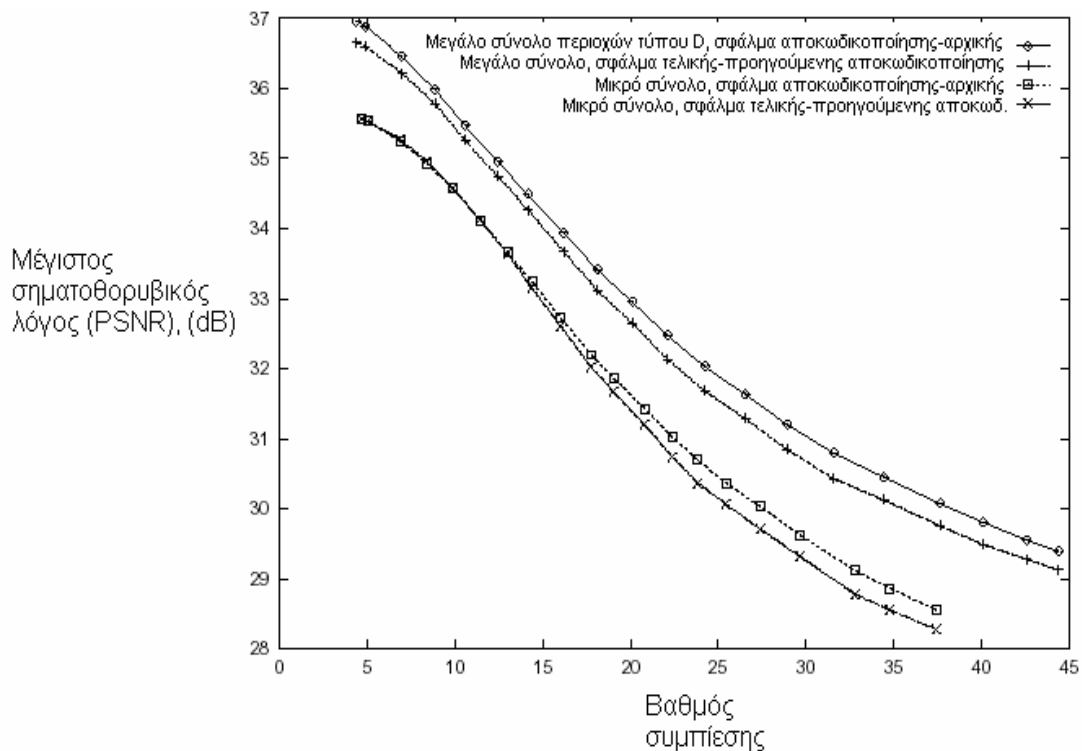
εικόνες του σφάλματος (μία για κάθε τρόπο). Στο σχήμα επίσης φαίνονται και οι περιοχές που έχουν χωριστεί σε τεταρτημόρια. Στα σχήματα 3.6 και 3.5 συνοψίζονται τα αποτελέσματα της εφαρμογής της μεθόδου για διαφορετικές τιμές ανοχής που φαίνονται στις καμπύλες βαθμού συμπίεσης-PSNR με βάση το χρόνο αποκωδικοποίησης.

Η αποκωδικοποίηση στην περίπτωση του τετραδικού δένδρου προχωρά με τον ίδιο τρόπο όπως και στους αλγορίθμους που χρησιμοποιούν σταθερές τύπου R περιοχές, δηλαδή με επαναληπτική εφαρμογή του μετασχηματισμού τύπου affine. Απαιτούνται μόνο 7 με 8 επαναλήψεις για τη σύγκλιση της μεθόδου. Στο σχήμα 3.8 μελετούμε την επίδραση στην εικόνα της αλλαγής της μέγιστης τιμής της αντίθεσης. Επιτρέποντας στην αντίθεση να παίρνει και τιμές απολύτως μεγαλύτερες της μονάδας, διακινδυνεύουμε την απώλεια της ιδιότητας της συστολής του τελεστή του μετασχηματισμού τύπου affine, αλλά όπως έχει προαναφερθεί αυτό δε σημαίνει ότι η σύγκλιση κατά την αποκωδικοποίηση είναι αδύνατο να επιτευχθεί. Ο λόγος για τον οποίο συμβαίνει αυτό είναι ότι ο τελεστής μπορεί να είναι συστολικός, αν για παράδειγμα σε μία επανάληψη έχουμε σύγκλιση, ενώ ο ίδιος ο μετασχηματισμός στο σύνολό του δε συγκλίνει. Υπάρχει δηλαδή η πιθανότητα κατά την τρέχουσα επανάληψη το σφάλμα αποκωδικοποίησης να είναι μικρότερο των προδιαγραφών, οπότε και μπορεί να τερματιστεί η διαδικασία, ενώ στις επόμενες επαναλήψεις το σφάλμα να αυξηθεί σε τιμές υψηλότερες από τα προδιαγεγραμμένο όριο.

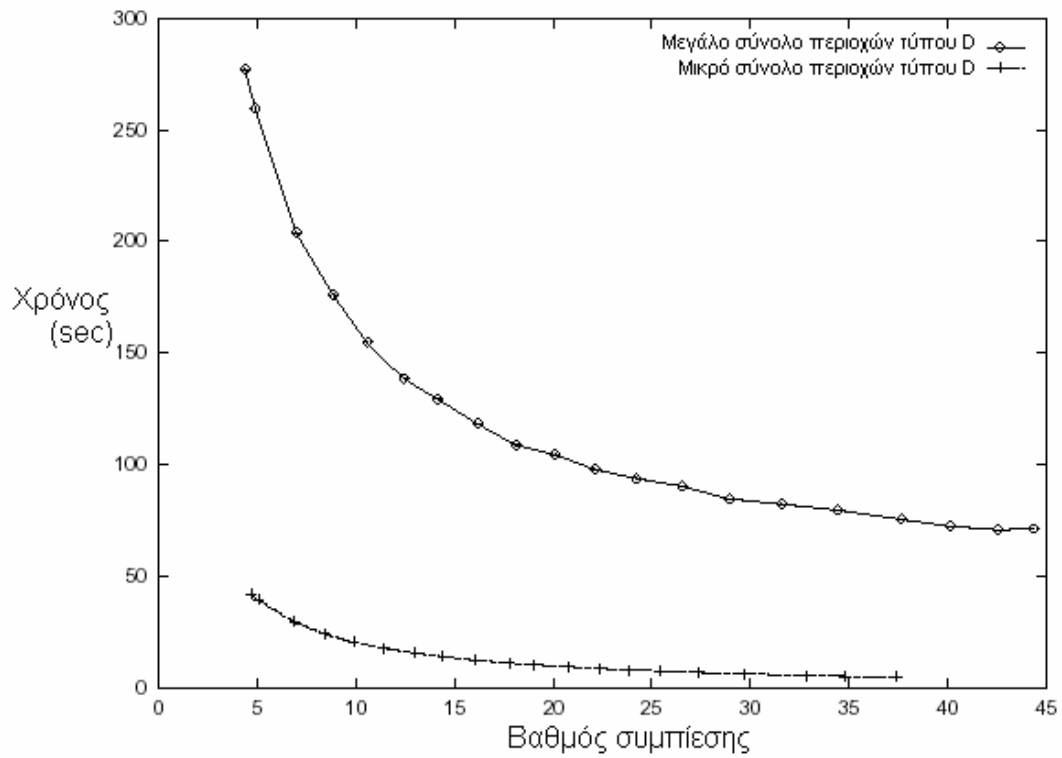


Σχ. 3.5: Δύο διαφορετικής ποιότητας (χαμηλή και μέτρια) κωδικοποιήσεις με τη χρήση του τετραεδρικού δένδρου μαζί με τις εικόνες του σφάλματος (όπου οι υψηλές τιμές σφάλματος έχουν πιο σκούρο χρώμα) και οι αντίστοιχες τους διαμερίσεις. Οι τιμές του PSNR είναι 28.3dB (αριστερά) και 32.0 (δεξιά). Οι αντίστοιχοι βαθμοί συμπίεσης είναι 37.5 και 17.8

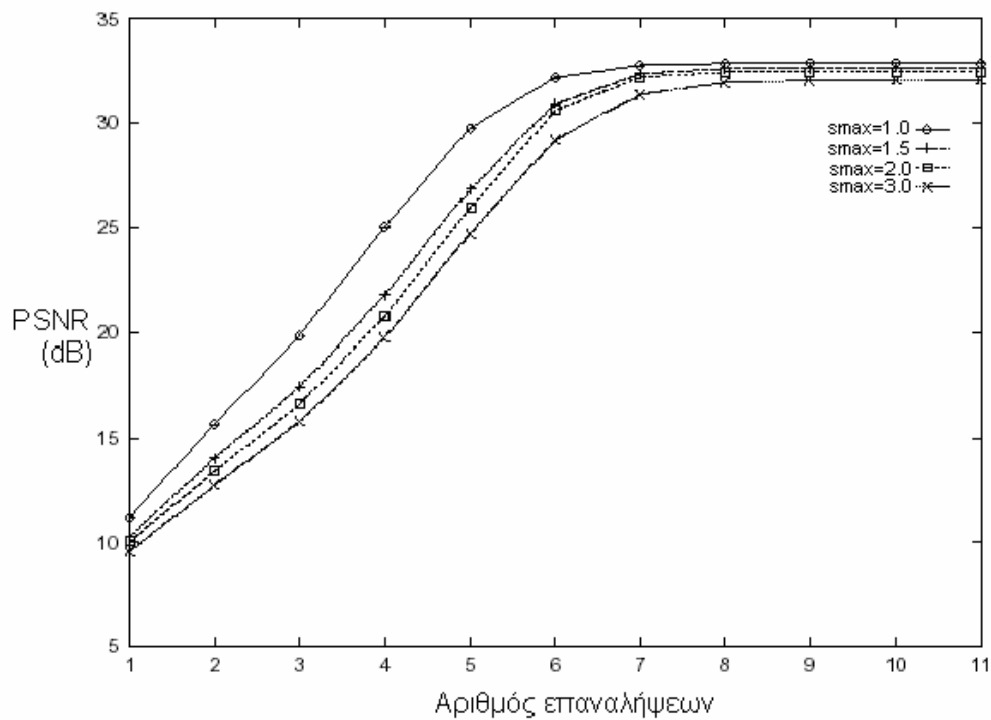
Η αποδοχή τιμών αντίθεσης μεγαλύτερων της μονάδος είναι δελεαστική, η διεύρυνση του πεδίου τιμών από την άλλη πλευρά, όταν δε συνοδεύεται από αντίστοιχη αύξηση του αριθμού των ψηφίων που δεσμεύονται για την αποθήκευσή της, μπορεί να επιδεινώσει στην πραγματικότητα το αποτέλεσμα της αποκωδικοποίησης όπως φαίνεται στο σχήμα. Ο λόγος είναι ότι η διακριτοποίηση των συντελεστών αντίθεσης είναι λιγότερο ακριβής.



Σχ. 3.6: Οι καμπύλες PSNR-βαθμού συμπίεσης για την κωδικοποίηση της εικόνας της Lena 512\*512 με τη χρήση του τετραεδρικού δένδρου. Τα αποτελέσματα αφορούν 2 διαφορετικά σύνολα domain περιοχών, ένα μικρό που προέρχεται από μη επικαλυπτόμενες μεταξύ τους περιοχές και ένα μεγαλύτερο το οποίο προήλθε από τη σάρωση της εικόνας σε περιοχές με βήμα τεσσάρων pixels και στις δύο διευθύνσεις. Η διαφορά στις καμπύλες μεταξύ του ορίου και της αποκωδικοποίησης οφείλεται στη διασπορά του σφάλματος κατά την αποκωδικοποίηση.



Σχ. 3.7: Χρόνοι εκτέλεσης για τις κωδικοποιήσεις του σχήματος 3.6



Σχ. 3.8: Ταχύτητα αποκωδικοποίησης για μεταβλητές μέγιστες τιμές αντίθεσης. Οι βαθμοί συμπίεσης για κάθε μία από αυτές τις κωδικοποιήσεις είναι περίπου ίδιοι, ανάμεσα στο 18 και το 19.

### 3.2.2 Οριζόντια-Κάθετη διαίρεση

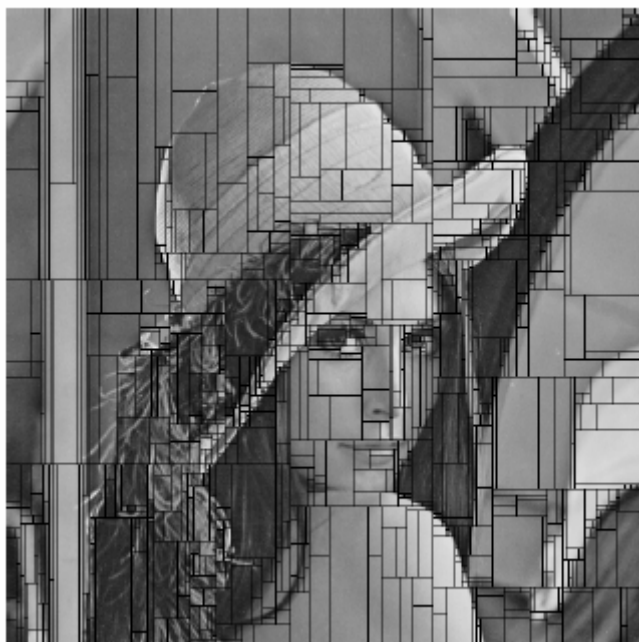
Η μέθοδος της Οριζόντιας-Κάθετης Διαμέρισης (*Horizontal-Vertical Partitioning*),<sup>3</sup> τεμαχίζει την εικόνα σε ορθογώνια παραλληλόγραμμα (σχήμα 3.9). Αν για μία δεδομένη ορθογώνια περιοχή τύπου R δεν καταστεί δυνατό να βρεθεί η κατάλληλη περιοχή τύπου D, η περιοχή χωρίζεται στα δύο με οριζόντια ή κάθετη τομή. Ο τεμαχισμός αυτός βασίζεται στην ομοιομορφία της αρχικής περιοχής. Για κάθε περιοχή  $R = (r_{ij})_{0 \leq i \leq N, 0 \leq j \leq M}$  οι διαφορές στην τιμή των αθροισμάτων των εντάσεων των εικονοστοιχείων των περιοχών που προέρχονται από κάθετη ή οριζόντια τομή υπολογίζονται αντίστοιχα από τους παρακάτω τύπους:

$$h_j = \frac{\min(j, M - j - 1)}{M - 1} \left( \sum_i r_{i,j} - \sum_i r_{i,j+1} \right) \quad (3.1)$$

$$v_i = \frac{\min(i, N - i - 1)}{N - 1} \left( \sum_j r_{i,j} - \sum_j r_{i+1,j} \right) \quad (3.2)$$

όπου το  $j$  φανερώνει την κάθετη τομή στην στήλη  $j$  του πίνακα  $R$  και το  $i$  την οριζόντια στην  $i$  γραμμή του πίνακα.





Σχ. 3.9: Μια Οριζόντια-Κάθετη διαμέριση αποτελούμενη από 1000 περιοχές τύπου R

Η μέγιστη τιμή των διαφορών αυτών, καθορίζει τον προσανατολισμό και το σημείο που θα γίνει η τομή. Στη συνέχεια, κατασκευάζεται ένα δένδρο απόφασης για τις περιοχές τύπου R που περιέχει την πληροφορία για κάθε μία από αυτές, αν δηλαδή χρειάστηκε να προχωρήσουμε σε τομή και ποια ή όχι.

Η διαδικασία που μόλις περιγράψαμε έχει αρνητική επίδραση στο χρόνο συμπίεσης, εξαιτίας των περιοχών τύπου R διαφορετικού σχήματος που προκύπτουν. Παρά το κόστος σε χρόνο και χώρο αποθήκευσης, η μέθοδος μας αποζημιώνει σε ποιότητα, καθώς τα αποτελέσματα της προσομοίωσης εμφανίζουν σημαντική βελτίωση των καμπυλών βαθμού συμπίεσης-PSNR σε σχέση με το τετραεδρικό δένδρο.

Ένα επόμενο βήμα στην προσαρμοστική διαδικασία είναι η χρήση πολυγώνων αντί για ορθογώνια παραλληλόγραμμα<sup>4</sup>. Η μέθοδος αυτή, βασισμένη στην εργασία των Wu και Yao είναι παρόμοια με την οριζόντια-κάθετη διαίρεση με την επιπρόσθετη ανάγκη για διαίρεση ως προς τις κατευθύνσεις των  $45^\circ$  και  $135^\circ$ .

### 3.3 Διαίρεση-και-Συγχώνευση

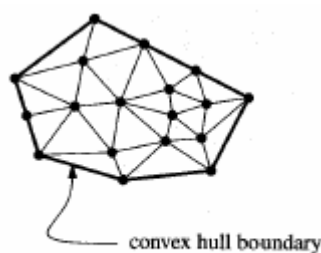
Ως αντιπροσωπευτικά παραδείγματα της κατηγορίας της Διαίρεσης –και-Συγχώνευσης (*Split-and-Merge*) παραθέτουμε τις ακόλουθες μεθόδους:

#### 3.3.1 Τριγωνοποίηση Delaunay

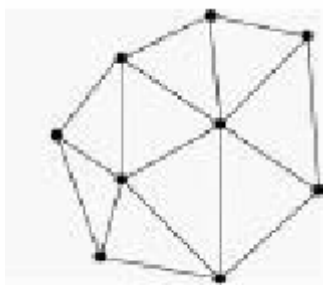
Ο Davoine et al.<sup>5</sup> συνηγορεί υπέρ της χρήσης της τριγωνοποίησης Delaunay ως μεθόδου διαίρεσης της εικόνας. Το πλεονέκτημα των τριγωνισμών είναι ο χωρίς περιορισμούς προσανατολισμός των ακμών.

Ορισμός: Καλούμε τριγωνοποίηση ενός πεπερασμένου συνόλου σημείων  $V$ , μια πλήρη επίπεδη υποδιαίρεση  $S$  που το σύνολο των κορυφών της είναι το  $V$ . Αυτό σημαίνει ότι δεν μπορούμε να προσθέσουμε καμία ακμή στην υποδιαίρεση αυτή για να ενώσουμε δύο κορυφές και να αποφύγουμε την τομή με τις προϋπάρχουσες.

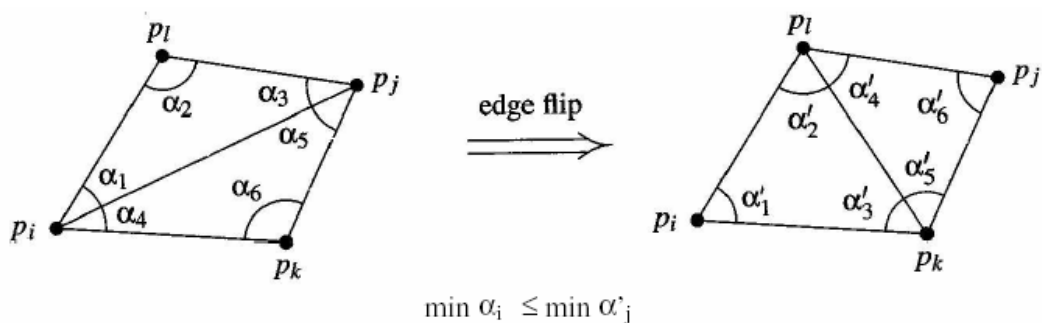
Είναι εύκολο να αποδειχθεί ότι αυτή η υποδιαίρεση αποτελείται μόνο από τρίγωνα: το κυρτό περίγραμμα (*convex hull boundary*) καθιστά όλες τις όψεις του  $S$  φραγμένες. Και σε όλες τις όψεις υπάρχουν πολύγωνα που έχουν τριγωνοποιηθεί, γιατί αν υπήρχε έστω και ένα που δεν ήταν τριγωνοποιημένο, θα μπορούσαμε να προσθέσουμε ακμές για την τριγωνοποίησή του, πράγμα που αντιβαίνει στον ορισμό.



Ορισμός: Η τριγωνοποίηση Delaunay είναι η τριγωνοποίηση  $S$  του συνόλου  $V$ , κάθε τρίγωνο της οποίας έχει περιγεγραμμένο κύκλο που δεν περιέχει κορυφή του  $V$  στο εσωτερικό του. Η τριγωνοποίηση Delaunay είναι μοναδική με την προϋπόθεση ότι τέσσερα ή περισσότερα σημεία του  $V$  δεν είναι ομοκυκλικά.



Η ιδιότητα της τριγωνοποίησης Delaunay είναι η μεγιστοποίηση της ελάχιστης εσωτερικής γωνίας. Αυτό σημαίνει ότι η κοινή ακμή δύο γειτονικών τέτοιων τριγώνων Delaunay που είναι και η διαγώνιος του κυρτού τετραπλεύρου που σχηματίζουν, μεγιστοποιεί την ελάχιστη από τις έξι γωνίες που προκύπτουν από την ένωση των δύο αυτών τριγώνων.



Η παραπάνω ιδιότητα επιβάλλει κάποια τάξη στη διαμέριση της εικόνας. Η μέθοδος λειτουργεί ως εξής: αρχίζοντας με ένα τρίγωνο Delaunay ενός συνόλου από ομοιόμορφα κατανεμημένα σημεία, η διαδικασία διαίρεσης βελτιώνεται με τη διαίρεση μη ομοιόμορφων τριγώνων (αυτών δηλαδή που η ένταση των *εικονοστοιχείων* που περιέχουν παρουσιάζει μεγάλη τυπική απόκλιση). Αυτό το βήμα συντελείται με την τοποθέτηση ενός επιπλέον σημείου στο βαρύκεντρο και με τον επαναυπολογισμό της τριγωνοποίησης για το νέο σύνολο σημείων. Η διαδικασία σταματά με τη βοήθεια ενός κριτηρίου ομοιομορφίας. Κατά τη διαδικασία συγχώνευσης, μία κορυφή  $p$  αγνοείται αν όλα τα τρίγωνα για τα οποία είναι κοινή, έχουν την ίδια ακριβώς μέση τιμή έντασης των *εικονοστοιχείων* που περιέχουν και στη συνέχεια εκτελούμε και πάλι την τριγωνοποίηση Delaunay για το νέο αυτό σύνολο σημείων. Στην εργασία τους, οι συγγραφείς επιτρέπουν τη συγχώνευση ακόμη και δύο τριγώνων, όταν

το τετράπλευρο που προκύπτει από την ένωσή τους είναι κυρτό και τα δύο τρίγωνα έχουν λίγο-πολύ την ίδια κατανομή διαβάθμισης του γκρι.

### 3.3.2 Ευρετική αναζήτηση

Ο fractal κώδικας που δημιουργείται με *Ευρετική Αναζήτηση* (*Heuristic Search*) βασισμένη σε περιοχές της εικόνας, προτάθηκε από τους Thomas και Deravi<sup>6</sup> και είναι μία προσέγγιση της μορφής διαίρεσης-και-συγχώνευσης. Αρχικά, η εικόνα χωρίζεται σε τετράγωνα μεγέθους  $4*4$  ή  $8*8$  για παράδειγμα. Στη συνέχεια τα γειτονικά τετράγωνα συγχωνεύονται διαδοχικά για το σχηματισμό περιοχών τύπου R ακανόνιστου σχήματος. Στο τέλος, καταλήγουμε στο σχηματισμό ενός συνόλου περιοχών τύπου R με μεγάλο μέγεθος και τυχαίο σχήμα οι οποίες και απαρτίζουν την εικόνα. Ο μικρός τους αριθμός μειώνει σε σημαντικό βαθμό τις απαιτήσεις σε χώρο αποθήκευσης, το μέγεθος και το σχήμα τους όμως, απαγορεύει τη συμβατική αναζήτηση ενός βέλτιστου domain κωδικού διανύσματος με απόρροια την ανάγκη εφαρμογής μίας ευρετικής στρατηγικής. Οι Thomas και Deravi δίνουν τρεις τέτοιες διαφορετικές μεθόδους που διαφοροποιούνται ως προς την πολυπλοκότητα. Σύμφωνα με την πιο απλή εκδοχή, ο αλγόριθμος ξεκινά από ένα μοναδικό αρχικό τετράγωνο τύπου R ή αλλιώς «σπόρο» (seed image) και αναζητείται για αυτό, το βέλτιστο τετράγωνο τύπου D. Στη συνέχεια εξετάζεται αν αυτός ο βέλτιστος μετασχηματισμός που έχει βρεθεί, μπορεί να εφαρμοσθεί και σε ένα γειτονικό τετράγωνο του αρχικού. Η επέκταση του προϋπάρχοντα μετασχηματισμού συνεχίζεται και σε γειτονικά τετράγωνα με αποτέλεσμα το πεδίο εφαρμογής του που είναι μία περιοχή της εικόνας να μεγαλώνει. Η επέκταση τερματίζεται όταν το σφάλμα προσέγγισης υπερβεί κάποιο προκαθορισμένο όριο. Η διαδικασία επαναλαμβάνεται με την επιλογή κάποιου άλλου «σπόρου» μέχρι να καλυφθεί ολόκληρη η εικόνα.

### 3.3.3 Ο Εξελικτικός Υπολογισμός

Μία άλλη μέθοδος προσαρμοστικής διαίρεσης, είναι η μέθοδος που χρησιμοποιεί τον εξελικτικό υπολογισμό<sup>7</sup>. Εδώ, για μία διαίρεση σταθερών σχημάτων σταθερού μεγέθους, τετραγώνων για παράδειγμα, βρίσκεται και πάλι ο fractal κώδικας όπως στη συμβατική κωδικοποίηση, αλλά για κάθε τύπου R περιοχή αποθηκεύονται σε μία λίστα τα  $d$  καλύτερα κωδικά διανύσματα για προσέγγιση, μαζί με τις τιμές αντίθεσης και φωτεινότητας που συνοδεύουν το καθένα. Ως αρχικό πληθυσμό της εξελικτικής διαδικασίας θεωρούμε το σύνολο αυτό περιοχών και μετασχηματισμών  $N$  φορές. Οι «απόγονοι» σχηματίζονται από την τυχαία συγχώνευση δύο γειτονικών τετραγώνων: ο κώδικας τροποποιείται έτσι ώστε να διατηρούνται μόνο οι λίστες μετασχηματισμών και κωδικών διανυσμάτων που έχουν αποθηκευθεί για τα δύο αυτά τετράγωνα. Στη συνέχεια επιλέγονται από αυτούς εκείνοι που δίνουν το μικρότερο σφάλμα.

Ο αλγόριθμος επομένως μπορεί να πάρει την ακόλουθη μορφή:

1. Διαμέριση της εικόνας σε τετραγωνικές περιοχές τύπου R σταθερού μεγέθους.
2. Για κάθε τύπου R περιοχή, εύρεση των βέλτιστων  $d$  τύπου D περιοχών και αποθήκευσή των διευθύνσεών τους, μαζί με τις τιμές της φωτεινότητας και αντίθεσης που τις συνοδεύουν.
3. Για όλες τις περιοχές τύπου R της εικόνας εκτελούμε: συγχώνευση δύο γειτονικών τύπου R περιοχών και αναζήτηση ενός νέου κώδικα για την προκύπτουσα περιοχή. Για το λόγο αυτό η διάσταση των υποψήφιων τύπου D περιοχών τροποποιείται ανάλογα. Όπως προαναφέρθηκε, οι υποψήφιες νέες τύπου D περιοχές αναζητώνται μέσα στο σύνολο των  $2d$  βέλτιστων κωδικών διανυσμάτων του προηγούμενου βήματος ( $d$  για κάθε τύπου R τετράγωνο) και επιλέγονται πάλι οι  $d$  καλύτερες με αποτέλεσμα στη νέα περιοχή τύπου R να έχουν αντιστοιχιστεί οι  $d$  νέες διευθύνσεις των τροποποιημένων τύπου D περιοχών μαζί με τους συντελεστές του affine μετασχηματισμού που τις συνοδεύουν.

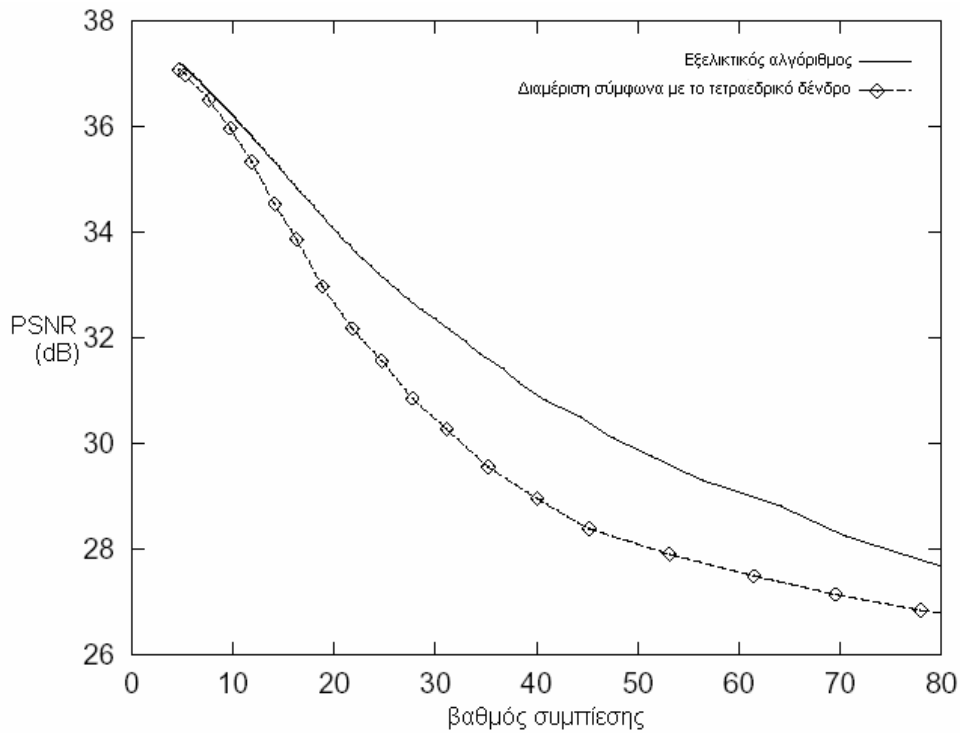
4. Αν το σφάλμα προσέγγισης υπερβεί κάποιο όριο ανοχής ή φθάσουμε στο επιθυμητό bit rate, ο αλγόριθμος σταματά, διαφορετικά επιστρέφουμε στο βήμα 3.

Είναι προφανές ότι με το ξεκίνημα της διαδικασίας η κωδικοποίηση έχει μεγάλο bit rate και μικρό σφάλμα προσέγγισης. Σύμφωνα όμως με το βήμα 3 σε κάθε νέα διαμέριση το σύνολο των τύπου R περιοχών μειώνεται κατά μία εξαιτίας της συγχώνευσης, με απόρροια τη μείωση του bit rate και την αύξηση του σφάλματος. Για το λόγο αυτό εισάγουμε στο βήμα 4 ένα κριτήριο τερματισμού που καθορίζεται από τις προδιαγραφές που πρέπει να πληρεί η συμπίεση.

Στο σχήμα 3.11 φαίνεται η σύγκριση της μεθόδου με αυτή του τετραεδρικού δένδρου ενώ στο σχήμα 3.10 εικονίζεται μία διαίρεση με τον εξελικτικό υπολογισμό στην οποία έχουμε τεμαχίσει την εικόνα σε 1000 περιοχές τύπου R.



Σχ. 3.10: Μία διαμέριση με βάση τον εξελικτικό αλγόριθμο αποτελούμενη από 1000 range περιοχές

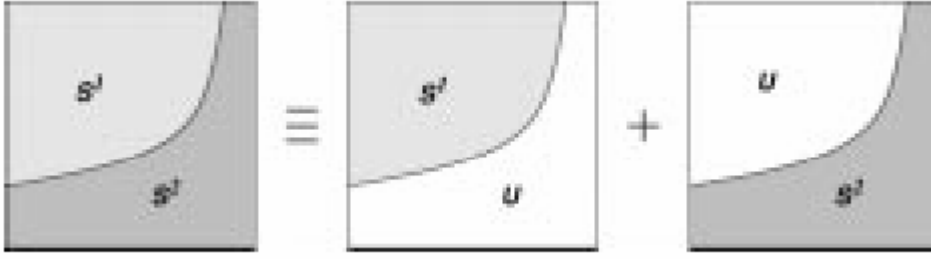


Σχ. 3.11: Οι καμπύλες του βαθμού παραμόρφωσης για τον εξελικτικό αλγόριθμο και το τετραεδρικό δένδρο για την εικόνα της Lena

### 3.4 Διαμέριση και κωδικοποίηση με βάση τα χαρακτηριστικά της εικόνας

Στην εργασία τους, οι *Belloulata* και *Konrad*<sup>8</sup> προτείνουν μία νέα προσέγγιση στη fractal συμπίεση με την a-priori τμηματοποίηση της εικόνας ώστε μεγάλες περιοχές της εικόνας να κωδικοποιούνται ανεξαρτήτως η μία της άλλης.

Η συνάρτηση τμηματοποίησης που χρησιμοποιούν τοποθετεί σε κάθε σημείο της εικόνας μια ετικέτα: οι ίδιες ετικέτες σχηματίζουν μια *περιοχή-τμήμα R* (*region*). Στην προσέγγισή τους χρησιμοποιούν τετράγωνα τύπου R και τύπου D περιοχές, όπως στην περίπτωση του συμβατικού αλγόριθμου, αλλά περιορίζουν το σφάλμα προσέγγισης E, ώστε να αφορά ένα υποσύνολο των θέσεων των εικονοστοιχείων στη τύπου R περιοχή που σχετίζεται με μια συγκεκριμένη περιοχή (*region*) της εικόνας που την περιέχει. Το υποσύνολο αυτό ονομάζεται *αυτόνομο τμήμα* ή *τεμάχιο*, (*segment*). Ένα παράδειγμα με περιοχή τύπου R 2 τεμαχίων φαίνεται στο σχήμα 3.12.



Σχ.3.12: Περιοχή δύο τεμαχίων  $S^1$  και  $S^2$ , που ανήκουν σε διαφορετικά τμήματα και η ανάλυσή της σε δύο περιοχές με καθορισμένη την ένταση των pixel είτε στην  $S^1$  είτε στην  $S^2$  και με απροσδιόριστη στην  $U$ .

Έστω  $S_{D_j}^n$  το  $n$ -οστό τεμάχιο στη τύπου  $D$  περιοχή  $D_j$  (μια περιοχή μπορεί να αποτελείται από περισσότερα των δύο τεμαχίων). Ομοίως έστω  $S_{R_i}^m$  το  $m$ -οστό τεμάχιο της τύπου  $R$  περιοχής  $R_i$ . Έστω  $\tilde{I}_{D_j}$  μια πρόβλεψη της συνάρτησης έντασης  $I_{D_j}$  της τύπου  $D$  περιοχής τέτοια ώστε:

$$\tilde{I}_{D_j}(x) = \begin{cases} I_{D_j}(x), & x \in S_{D_j}^n \\ u, & \text{αλλου} \end{cases} \quad (3.3)$$

όπου  $u$  είναι η τιμή της έντασης που προβλέπεται ότι έχει η υπόλοιπη περιοχή εκτός του τεμαχίου  $S_{D_j}^n$ . Ένα νέο λοιπόν σφάλμα προσέγγισης  $E$  ορίζεται ως εξής:

$$E(I_{R_i}, I_{D_j}, w_i^m) = \frac{1}{|S_{R_i}^m|} \sum_{x \in S_{R_i}^m} [I_{R_i}(x) - w_i^m(I_{D_j}, x)]^2 \quad (3.4)$$

όπου το  $w_i^m$  δηλώνει ένα μετασχηματισμό τύπου affine για το τεμάχιο  $S_{R_i}^m$ . Παρατηρούμε ότι σε σχέση με το συμβατικό αλγόριθμο, η παραπάνω παραμόρφωση αποτιμάται μόνο για τις θέσεις εκείνες των εικονοστοιχείων που βρίσκονται στο τεμάχιο  $S_{R_i}^m$  του  $R_i$ . Είναι φανερό ότι όταν τα σχήματα του  $S_{R_i}^m$  και του υπό συστολή  $S_{D_j}^n$  δεν ταιριάζουν, απαιτείται μια εκτίμηση για κάποια εικονοστοιχεία στην τύπου  $D$  περιοχή, γεγονός που εξηγεί και τη χρήση της προσθήκης  $u$ . Επίσης αξίζει να σημειωθεί ότι, όταν οι περιοχές τύπου  $R$  και τύπου  $D$  που εξετάζονται βρίσκονται εντός της ευρύτερης



περιοχής που κωδικοποιείται, τα  $S_{R_i}^m, S_{D_j}^n$  καλύπτουν όλο το εύρος της τύπου R και της τύπου D περιοχής αντίστοιχα και ο τύπος (3.4) γίνεται αυτομάτως ο τύπος του συμβατικού fractal αλγορίθμου.

Όπως και στην περίπτωση της συνηθισμένης κωδικοποίησης με fractal, για κάθε τεμάχιο  $S_{R_i}^m$  της τύπου R περιοχής  $R_i$  αναζητείται το τεμάχιο  $S_{D_j}^n$  της αντίστοιχης τύπου D περιοχής  $D_j$  και ένα μετασχηματισμός  $w_i^m$  που θα ελαχιστοποιεί το σφάλμα  $E$ . Ο δείκτης  $j$  της τύπου D περιοχής και ο ισομετρικός μετασχηματισμός βρίσκονται μετά από διεξοδική αναζήτηση, ενώ οι παράμετροι αντίθεσης και φωτεινότητας υπολογίζονται όπως και στην απλή περίπτωση, αλλά με τους υπολογισμούς να λαμβάνουν χώρα εντός των ορίων του υπό κωδικοποίηση τεμαχίου  $S_{R_i}^m$ . Η εικόνα αναπαράγεται κατά την αποκωδικοποίηση με επαναληπτική εφαρμογή του μετασχηματισμού  $w_i^m$  που περιορίζεται στο  $S_{R_i}^m$ . Η απόδειξη της σύγκλισης του αλγορίθμου αυτού έχει δοθεί από τον *Mansour*<sup>9</sup>.

### 3.4.1 Επιλογή του χώρου Αναζήτησης

Η υπολογιστική πολυπλοκότητα της κωδικοποίησης μιας εικόνας με την προτεινόμενη μέθοδο συνδέεται άμεσα με το μέγεθος του χώρου αναζήτησης μέσα στον οποίο η παραμόρφωση (3.4) ελαχιστοποιείται. Η πιο απαιτητική σε πράξεις περίπτωση είναι εκείνη στην οποία λαμβάνεται υπόψη κάθε τεμάχιο κάθε τύπου D περιοχής της εικόνας, δηλαδή το κωδικό βιβλίο των τύπου D περιοχών, είναι όλη η εικόνα. Αυτή η διεξοδική διαδικασία είναι θεωρητικά βέλτιστη, αλλά τρομερά επίπονη υπολογιστικά. Επιπλέον δεν επιτρέπει την ανεξάρτητη αποκωδικοποίηση των περιοχών .

Για να διασφαλισθεί η περιοχή ανά περιοχή κωδικοποίηση και αποκωδικοποίηση τόσο οι τύπου R όσο και οι τύπου D περιοχές πρέπει να είναι τοποθετημένες στο ίδιο τμήμα της εικόνας **R**. Επομένως, κατά την

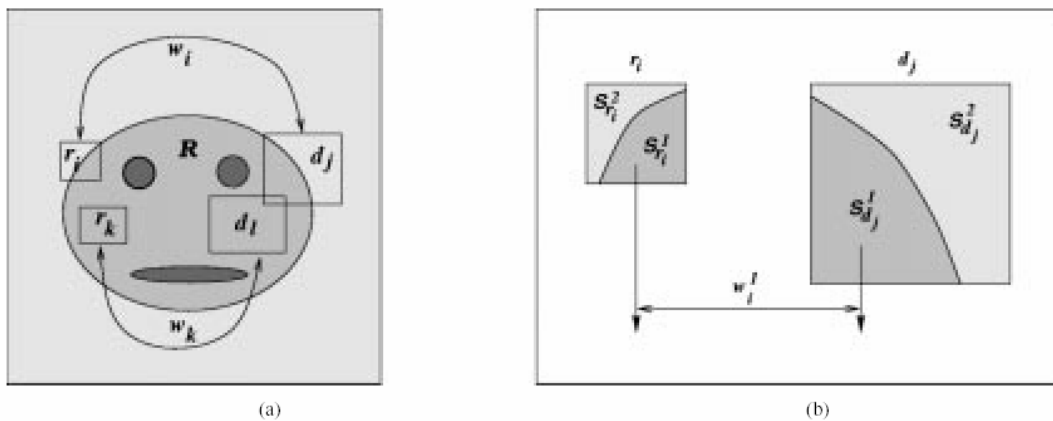
ελαχιστοποίηση της εξίσωσης (3.4), 4 περιπτώσεις μπορεί να προκύψουν που έχουν να κάνουν με τη θέση των περιοχών  $R_i$  και  $D_j$  σε σχέση με το τμήμα της εικόνας (region)  $\mathbf{R}$ .

1. Οι  $R_i$  και  $D_j$  είναι και οι δύο εσωτερικές περιοχές
2. Οι  $R_i$  και  $D_j$  είναι και οι δύο συνοριακές περιοχές
3. Η  $R_i$  είναι εσωτερική ενώ η  $D_j$  είναι συνοριακή περιοχή
4. Η  $R_i$  είναι συνοριακή ενώ η  $D_j$  είναι εσωτερική περιοχή

Ενώ στην πρώτη περίπτωση εκτελείται η συνηθισμένη αναζήτηση για όλες τις περιοχές, στη δεύτερη περίπτωση εκτελείται μόνο μερική προσέγγιση αφού τόσο οι τύπου D όσο και οι τύπου R περιοχές βρίσκονται στο σύνορο του τμήματος  $\mathbf{R}$ . Είναι φανερό ότι η πρόβλεψη  $\mu$  της έντασης είναι απαραίτητη, όταν το  $S_{D_j}^n$  που έχει υποστεί συστολή, δεν περιλαμβάνει ολοκληρωτικά το  $S_{R_i}^m$ . Αυτό σημαίνει ότι το υπό συστολή  $S_{D_j}^n$  δεν περιέχει τον ίδιο αριθμό εικονοστοιχείων με το  $S_{R_i}^m$ . Στην 3<sup>η</sup> περίπτωση, η εκτίμηση της έντασης της τύπου D περιοχής είναι πάντα απαραίτητη καθώς το  $R_i$  είναι εσωτερικό, ενώ το  $D_j$  συνοριακό. Αν και συνήθως αποτελεί μια καλύτερη λύση από το να χρησιμοποιούνται εντάσεις από μια γειτονική περιοχή (συνηθισμένη κωδικοποίηση με fractal), η πρόβλεψη της τιμής της έντασης είναι πιθανό να έχει ως αποτέλεσμα υποβέλτιστες εντάσεις (ανακρίβεια στην εκτίμηση της έντασης) σε σύγκριση με την περίπτωση που οι τύπου D περιοχές βρίσκονται εξ ολοκλήρου εντός του υπό κωδικοποίηση τμήματος  $\mathbf{R}$  της εικόνας. Στην 4<sup>η</sup> περίπτωση, δεν χρειάζεται εκτίμηση έντασης καθώς το  $R_i$  είναι συνοριακό. Αν και είναι ένα πραγματοποιήσιμο σενάριο, αυτή η περίπτωση απαιτεί μια εξέταση -που κοστίζει υπερβολικά σε πράξεις- όλων των εσωτερικών τύπου D περιοχών για κάθε συνοριακή τύπου R περιοχή. Τυπικά υπάρχουν πολύ περισσότερες εσωτερικές περιοχές από τις συνοριακές.

Στο πιο πολύπλοκο σενάριο, πραγματοποιείται μια 1+2+3+4 αναζήτηση (λαμβάνονται υπόψη όλες οι περιπτώσεις). Αναμένεται ότι η 3<sup>η</sup> και 4<sup>η</sup>

περίπτωση συνεισφέρουν πολύ λίγο στη διαμόρφωση του μέγιστου σηματοθορυβικού λόγου της συμπίεσης. Αποφεύγοντας την 3<sup>η</sup> περίπτωση, μπορεί να πραγματοποιηθεί μια 1+2+4 αναζήτηση με το επιπρόσθετο πλεονέκτημα της ελαττωμένης πολυπλοκότητας. Ομοίως αποφεύγοντας την 4<sup>η</sup> περίπτωση, η πολυπλοκότητα μπορεί να ελαττωθεί περαιτέρω με την πραγματοποίηση μιας 1+2+3 αναζήτησης αγνοώντας τις εσωτερικές τύπου D περιοχές για κάθε συνοριακή τύπου R περιοχή. Αγνοώντας και τις 2 τελευταίες περιπτώσεις, μια 1+2 αναζήτηση έχει ως αποτέλεσμα ακόμα λιγότερες πράξεις καθώς οι συνοριακές περιοχές προσεγγίζονται αποκλειστικά από συνοριακές, όπως συμβαίνει και με τις εσωτερικές περιοχές. Μία σχηματική αναπαράσταση μιας 1+2 αναζήτησης εικονίζεται στο σχήμα 3.13.



Σχ.3.13: Σχηματική απεικόνιση της μεθόδου των Belloulata και Konrad. (α) Οι εσωτερικές range περιοχές προσεγγίζονται από εσωτερικές domain περιοχές του ίδιο τμήματος και οι συνοριακές range περιοχές προσεγγίζονται μερικώς (3.4) από τις ίδιου τύπου domain. (β) Τεμάχια συνοριακών περιοχών επεξεργάζονται ανεξάρτητα: το τεμάχιο  $S_{r_i}^1$  προσεγγίζεται από την περιοχή  $d_j$  με το μετασχηματισμό τύπου affine  $w_i^1$ , ενώ το τεμάχιο  $S_{r_i}^2$  της ίδιας range περιοχής  $r_i$ , μπορεί να προσεγγιστεί από διαφορετική domain περιοχή με διαφορετικό μετασχηματισμό.

Για να εκτιμήσουν την απόδοση σε αντιδιαστολή με την πολυπλοκότητα του αλγορίθμου σε σχέση με τα διάφορα σενάρια αναζήτησης που προαναφέρθηκαν, οι συγγραφείς έχουν κωδικοποιήσει ανεξάρτητα το «προσκήνιο»(foreground) και το φόντο (background) τεσσάρων δοκιμαστικών εικόνων MPEG-4. Έχουν εξεταστεί και τα 4 σενάρια αναζήτησης: 1+2+3+4, 1+2+3, 1+2+4, 1+2, όλα με LPE πρόβλεψη (για την οποία θα αναφερθούμε στο επόμενο εδάφιο), αλλά παρουσιάζονται τα αποτελέσματα μόνο για την 1<sup>η</sup>

(πιο πολύπλοκη) και την 4<sup>η</sup> (πιο απλή) περίπτωση. Ο πίνακας 3.1 δείχνει το PSNR, την αναλογία ανά εικονοστοιχείο και το χρόνο επεξεργασίας για το «προσκήνιο» και το φόντο καθε μιας από τις 4 εικόνες. Παρατηρείται μια μεγάλη ελάττωση του υπολογιστικού χρόνου κατά 25-51% σε αντίθεση με μια μικρή πτώση της απόδοσης 0.04-0.06 dB. Η πολυπλοκότητα των άλλων δυο σεναρίων καθώς και η απόδοσή τους ήταν ανάμεσα στα δεδομένα του πρώτου και τελευταίου. Είναι φανερό από την παρατήρηση του πίνακα, ότι η 1+2 αναζήτηση είναι υπολογιστικά ελκυστική ενώ επιτυγχάνει και μια πολύ καλή απόδοση.

Πίνακας 3.1: Σύγκριση της επίδοσης της τεχνικής των Belloulata-Konrad για 2 σενάρια αναζήτησης (1+2+3+4 και 1+2) και 2 τρόπους πρόβλεψης της έντασης της περιοχής εκτός του εξεταζόμενου τεμαχίου (LPE και μηδενικής).

	1+2+3+4 αναζήτηση με πρόβλεψη LPE			1+2 αναζήτηση με πρόβλεψη LPE			1+2 αναζήτηση με Μηδενική πρόβλεψη		
	PSNR (dB)	bpp	Χρόνος Εκτέλεσης (sec)	PSNR (dB)	bpp	Χρόνος Εκτέλεσης (sec)	PSNR (dB)	bpp	Χρόνος Εκτέλεσης (sec)
προσκήνιο									
Κυκλάμινο	32.88	0.30	128	32.42	0.29	62	32.11	0.28	61
Ειδήσεις	30.47	0.25	108	30.10	0.24	56	30.00	0.24	55
Επιστάτης	39.11	0.23	96	39.00	0.23	61	38.86	0.23	61
Παιδιά	28.34	0.13	32	27.92	0.13	17	27.88	0.13	16
φόντο									
Κυκλάμινο	40.28	0.40	208	40.07	0.40	110	39.50	0.40	114
Ειδήσεις	33.66	0.40	316	33.38	0.40	205	33.22	0.40	204
Επιστάτης	29.40	0.40	277	29.36	0.39	208	29.35	0.38	206
Παιδιά	31.88	0.51	480	31.70	0.51	352	31.67	0.51	349

### 3.4.2 Επιλογή του τρόπου εκτίμησης της έντασης για τις τύπου D περιοχές

Όπως προαναφέρθηκε, επειδή κατά την 1+2 αναζήτηση συμβαίνει το μετασχηματισμένο τεμάχιο  $S_{D_j}^n$  της τύπου D περιοχής να μην περικλείει εξ ολοκλήρου το τεμάχιο  $S_{R_i}^m$ , οι συνοριακές τύπου D περιοχές χρειάζονται μια εκτίμηση για την τιμή της έντασής τους.

Πριν κλείσουμε την αναφορά μας στην εργασία των Konrad-Belloulata αναφέρουμε τις δύο περιπτώσεις εκτίμησης που χρησιμοποιούν και τις οποίες έχει υιοθετήσει και η μέθοδος MPEG-4. Την *εκτίμηση βαθυπερατού φίλτρου* (*Low-Pass Extrapolation, LPE*) και την εκτίμηση μηδενικής τιμής (Zero Mean), (πίνακας 3.1). Κατά την 1<sup>η</sup> εκτίμηση τα εικονοστοιχεία της τύπου D περιοχής εκτός του τεμαχίου  $S_{D_j}^n$  που μας ενδιαφέρει, έχουν τιμή έντασης ίση με το μέσο όρο αυτής στο τεμάχιο  $S_{D_j}^n$

$$m_{D_j} = \frac{1}{|S_{D_j}^n|} \sum_{x \in S_{D_j}^n} I_{D_j}(x) \quad (3.5)$$

και στη συνέχεια υποδειγματοληπτώνται (μία διαδικασία ισοδύναμη με τη χρήση βαθυπερατού φίλτρου) με τον τρόπο που ήδη έχουμε αναφέρει στο προηγούμενο κεφάλαιο. Η μηδενική πρόβλεψη προτάθηκε στη μέθοδο MPEG-4 για περιοχές μηδενικής μέσης τιμής και εξετάζεται σε μια πιθανή εναλλακτική της πρώτης μεθόδου, εξ αιτίας της χαμηλότερης υπολογιστικής της πολυπλοκότητας.

Στον πίνακα 3.1 συγκρίνονται οι δύο μέθοδοι ως προς το PSNR, το *bit rate* και το χρόνο επεξεργασίας για την 1+2 αναζήτηση. Παρατηρείται ότι ενώ με τη δεύτερη μέθοδο η πτώση στην ποιότητα (0.01-0.57 dB), ήταν αναμενόμενη, η ελάχιστη μείωση του χρόνου επεξεργασίας σε σχέση με την 1<sup>η</sup> μέθοδο οδηγεί στο συμπέρασμα ότι η εκτίμηση συνεισφέρει ελάχιστα στην υπολογιστική πολυπλοκότητα. Για το λόγο αυτό χρησιμοποιείται κυρίως

τεχνική LPE ως μέσο πρόβλεψης, καθώς η βελτίωση της ποιότητας που προσφέρει υπερβαίνει το κόστος της σε πολυπλοκότητα.

Στη συνέχεια της εργασίας τους, οι συγγραφείς συνδυάζουν την πρότασή τους με το Διακριτό Μετασχηματισμό Συνημιτόνου, ο οποίος όμως δεν αποτελεί κάποια ξεχωριστή τεχνική κωδικοποίησης με fractal, αλλά ένα εργαλείο επιτάχυνσης των υπολογισμών, το οποίο βρίσκει εφαρμογή σε πολλές μεθόδους συμπίεσης και εξετάζεται σε ξεχωριστό κεφάλαιο.

## Βιβλιογραφία

1. Ning Lu, *Fractal Imaging*, Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto.
2. Dietmar Saupe, Raouf Hamzaoui, Hannes Haternstein, *Fractal Image Compression, An Introductory Overview*.
3. Fisher., Y., Menlove, S., *Fractal encoding with HV partitions*, in: *Fractal Image Compression – Theory and Application*, Y. Fisher (ed.), Springer-Verlag , New York, 1994.
4. Reusens, E., *Partitioning complexity issue for iterated function systems based image coding*, in : *Proceedings of the VIIth European Signal Processing Conference EUSIPCO'94*, Edinburgh, Sept.1994.
5. Davoine, F., Chassery, J.-M., *Adaptive triangulation for attractor image coding*, in: *Proc. of 12<sup>th</sup> International Conference on Pattern Recognition (ICPR)*, Jerusalem, Oct.94.
6. Thomas, L., Deravi, F., *Region-based fractal image Compression using heuristic search*, *IEEE Transactions on Image Processing* (1995) 4,6 832-838
7. Saupe, D., Ruhl, M., *Evolutionary fractal image compression*, in: *Proc. ICIP-96 IEEE International Conference on Image Processing*, Lausanne, Sept. 1996
8. Belloulata and Konrad: *Fractal Image Compression with region-based functionality*, in: *IEEE Transactions on Image Processing*, Vol. 11, No. 4, April 2002.
9. A-R. Mansouri, *Some results in Fractal Coding*: IRNS-Telecommun., 1999

---

## ΚΕΦΑΛΑΙΟ 4

---

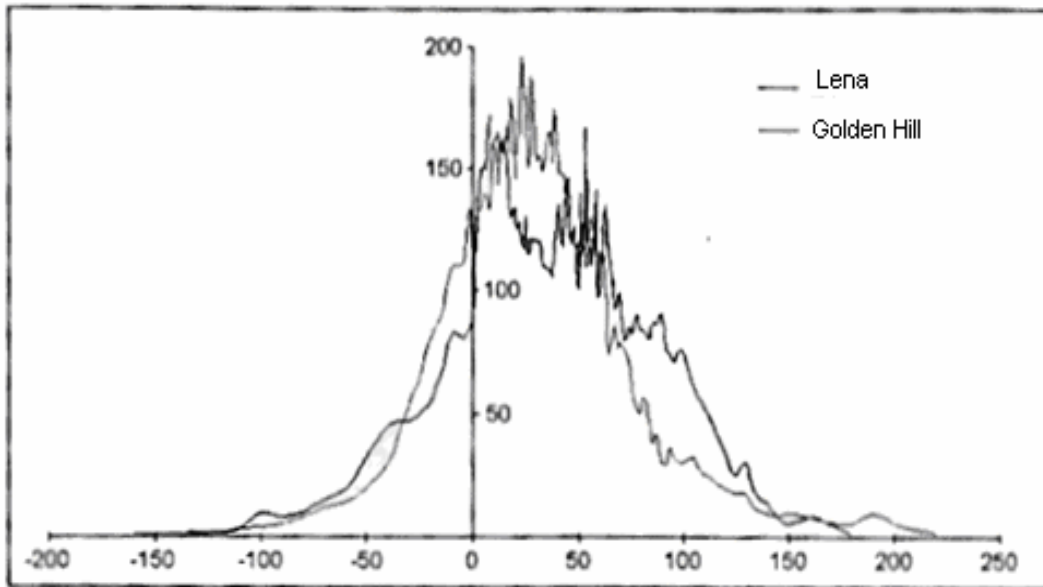
### Φωτεινότητα και Αντίθεση - Γενικότεροι μετασχηματισμοί έντασης

Οι δύο παράμετροι του συνηθέστερου μετασχηματισμού τύπου affine που χρησιμοποιεί το PIFS είναι η φωτεινότητα και η αντίθεση. Οι σχέσεις από τις οποίες προκύπτουν δίνουν εν γένει ένα αποτέλεσμα που ανήκει στο σύνολο των πραγματικών αριθμών. Για τις ανάγκες της συμπίεσης επιβάλλεται διακριτοποίηση των παραπάνω τιμών. Στο κεφάλαιο αυτό θα γίνει λόγος για τον τρόπο με τον οποίο μπορούμε να διακριτοποιήσουμε αποδοτικά τις δύο παραμέτρους. Το κεφάλαιο κλείνει με τη συζήτηση για τη γενικότερη μορφή που μπορεί να έχει ένας συστολικός μετασχηματισμός στις μεθόδους συμπίεσης με βάση τη θεωρία των fractal και την αναφορά σε μερικούς από τους μετασχηματισμούς που έχουν χρησιμοποιηθεί στη βιβλιογραφία.

#### 4.1 Η συνιστώσα της φωτεινότητας

Στις περισσότερες εφαρμογές συμπίεσης τα ψηφία που δεσμεύονται για τη φωτεινότητα  $o$  είναι τα περισσότερα ανάμεσα στα δεδομένα του fractal κώδικα<sup>1</sup>. Το σχήμα 4.1 δείχνει τη διασπορά των τιμών φωτεινότητας για ένα συμβατικό fractal αλγόριθμο για της εικόνας Lena και Golden Hill.

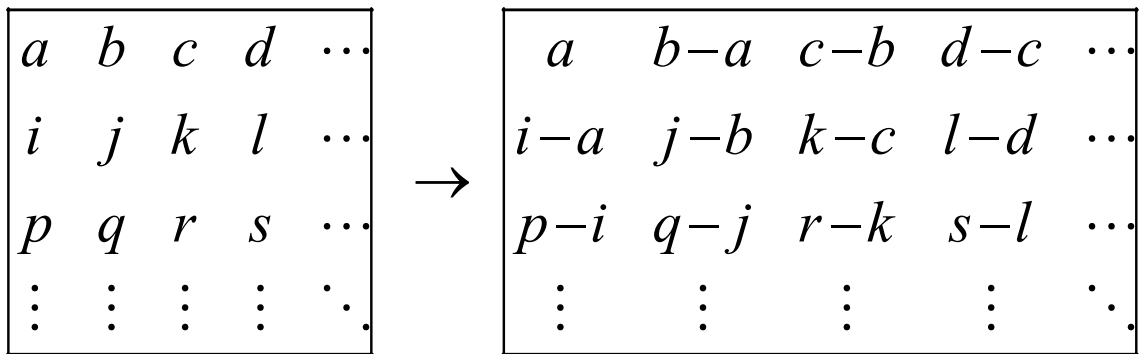




Σχ.4.1: Η διασπορά των τιμών της φωτεινότητας για τον απλό fractal αλγόριθμο.

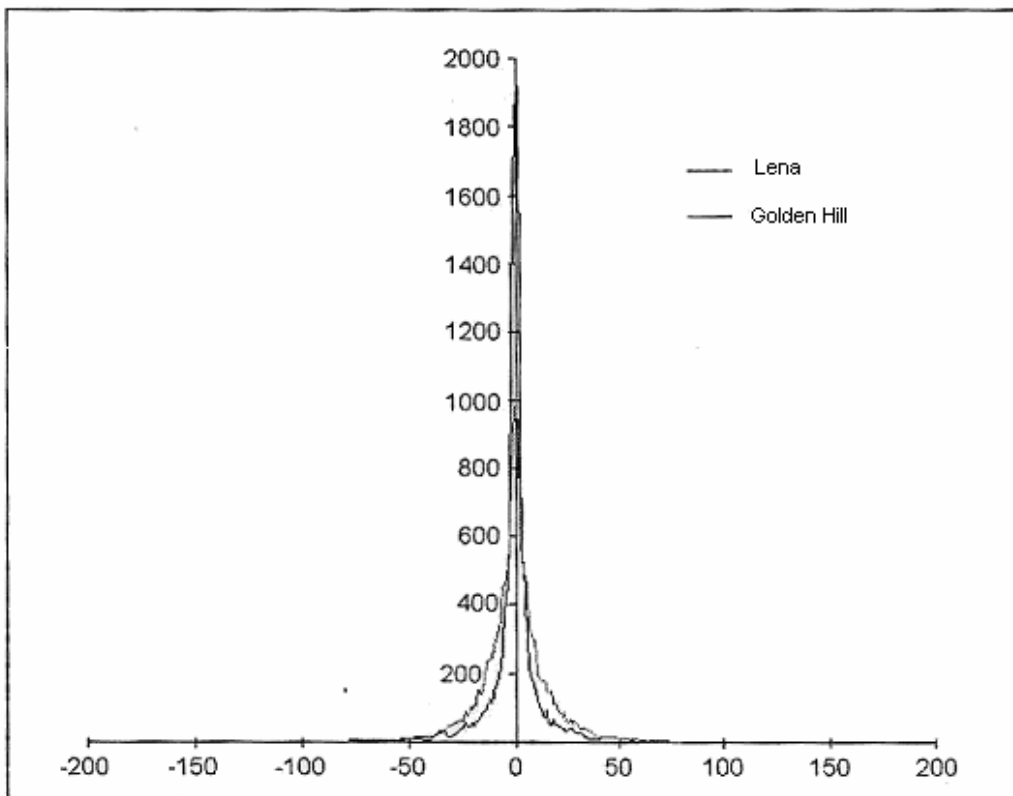
Το σύνολο των τιμών της φωτεινότητας μίας δεδομένης εικόνας μπορεί να οργανωθεί σε ένα πίνακα 2 διαστάσεων. Η διακριτοποίηση – συμπίεση γίνεται με τη χρήση ενός απλού αλγορίθμου αποσύζευξης (*decorrelating algorithm*) δύο διαστάσεων: της *Διαφορικής Πάλμοκωδικής Διαμόρφωσης (Differential Pulse Code Modulation)*.

Ο αλγόριθμος αυτός εικονίζεται στο σχήμα 4.2. Οι τιμές της φωτεινότητας αποθηκεύονται σε έναν πίνακα δύο διαστάσεων που προσπελάζεται κατά γραμμή. Κάθε θέση του πίνακα αυτού αντιστοιχεί σε μία περιοχή τύπου R. Ο πίνακας που δημιουργήθηκε με τον παραπάνω τρόπο, μετασχηματίζεται ως εξής: κάθε του στοιχείο γίνεται ίσο με τη διαφορά της τιμής που αντιστοιχεί στο στοιχείο αυτό μείον την τιμή του στοιχείου που βρίσκεται στην ίδια στήλη και στην αμέσως προηγούμενη γραμμή. Ο παραπάνω υπολογισμός δεν ισχύει για τα στοιχεία της πρώτης γραμμής τα οποία προκύπτουν ως εξής: το κάθε ένα από αυτά είναι ίσο με τη διαφορά του από το προηγούμενο εκτός από το πρώτο στοιχείο το οποίο και αποθηκεύεται.



Σχ.4.2: Ένας δισδιάστατος αλγόριθμος DPCM

Το σχήμα 4.3 δείχνει τη διασπορά των τιμών της φωτεινότητας όταν εφαρμόζουμε τον αλγόριθμο ΔΠΔ στον πίνακα των δεδομένων φωτεινότητας.



Σχ.4.3: Διασπορά τιμών της φωτεινότητας για τις εικόνες Lena και Golden Hill με τη χρήση του DPCM

Από το σχήμα συμπεραίνουμε ότι οι τιμές της φωτεινότητας παρουσιάζουν μικρότερη διασπορά με την χρήση του αλγορίθμου ΔΠΔ. Αυτό οφείλεται στην τοπική ομοιομορφία που παρουσιάζουν συνήθως οι εικόνες. Οι τιμές της

συνιστώσας της φωτεινότητας για γειτονικές τύπου R περιοχές (όπως αυτές που απέχουν για παράδειγμα μία γραμμή και βρίσκονται στην ίδια στήλη) είναι παραπλήσιες εξαιτίας της ομοιότητάς τους με αποτέλεσμα η διαφορά τους να οδηγεί σε μικρή τιμή φωτεινότητας.

Έχουν μελετηθεί στη διεθνή βιβλιογραφία διάφοροι τρόποι για την αποδοτική διακριτοποίηση των τιμών της παραμέτρου  $\sigma$  η οποία δεν είναι τίποτε άλλο από ένας προσθετικός όρος για τη διόρθωση της φωτεινότητας. Τι θα συμβεί αν αποθηκεύσουμε την μέση τιμή της φωτεινότητας της τύπου R περιοχής αντί του όρου  $\sigma$ ; Το επόμενο εδάφιο δίνει την απάντηση στο τελευταίο ερώτημα.

#### 4.1.1 Διακριτοποίηση μέσης τιμής

Όπως είναι προφανές η μέση τιμή της έντασης του γκρι μιας τύπου R περιοχής εξαρτάται από τον τρόπο με τον οποίο έχει διαμεριστεί η εικόνα και όχι από τη μορφή του μετασχηματισμού που εφαρμόζεται για την κωδικοποίησή της. Κατά συνέπεια, οι μέσες τιμές μπορούν να διακριτοποιηθούν ανεξάρτητα από την όποια διαδικασία συμπίεσης, σε αντίθεση με την φωτεινότητα  $\sigma$  η τιμή της οποίας είναι συνυφασμένη με τον απλό συστολικό μετασχηματισμό  $sD + \sigma \mathbf{1}$  που χρησιμοποιείται συνήθως.

Αν χρησιμοποιήσουμε ξανά τον αλγόριθμο ΔΠΔ δύο διαστάσεων για τον πίνακα των μέσων τιμών αυτήν τη φορά διαπιστώνουμε μικρότερη διασπορά τιμών από την περίπτωση της φωτεινότητας. Γενικά, οι πειραματικές μετρήσεις έχουν δείξει ότι οι μέσοι όροι καταλαμβάνουν κατά 15-20% μικρότερο χώρο απόθήκευσης σε σχέση με τις ισοδύναμες τιμές φωτεινότητας.

Ανατρέχοντας στους τύπους που δίνουν την αντίθεση και τη φωτεινότητα βρίσκουμε ότι αυτοί μπορούν να γραφούν και ως εξής<sup>2</sup>:

$$s^* = \frac{\langle R - \bar{r}\mathbf{1}, D - \bar{d}\mathbf{1} \rangle}{\|D - \bar{d}\mathbf{1}\|_2^2}, \sigma^* = \bar{r} - s^* \bar{d} \quad (4.1)$$

Κατά συνέπεια το σφάλμα μπορεί να απλοποιηθεί στην ακόλουθη μορφή:

$$\begin{aligned}
E(R, D) &= \|R - \bar{r}\mathbf{1}\|^2 - \frac{\langle R - \bar{r}\mathbf{1}, D - \bar{d}\mathbf{1} \rangle^2}{\|D - \bar{d}\mathbf{1}\|_2^2} = \\
&= \|R - \bar{r}\mathbf{1}\|^2 - s^2 \|D - \bar{d}\mathbf{1}\|^2
\end{aligned} \tag{4.2}$$

Το ελάχιστο σφάλμα ως προς όλες τις τύπου D περιοχές που ανήκουν στο σύνολο των περιοχών τύπου D (έστω  $\Omega$ ) γράφεται:

$$E(R) = \min_{D \in \Omega} E(R, D) = \|R - \bar{r}\mathbf{1}\|^2 (1 - c^2) \tag{4.3}$$

όπου

$$c = \max_{D \in \Omega} \frac{\langle R - \bar{r}\mathbf{1}, D - \bar{d}\mathbf{1} \rangle}{\|R - \bar{r}\mathbf{1}\| \|D - \bar{d}\mathbf{1}\|} \tag{4.4}$$

Επομένως το μέσο τετραγωνικό σφάλμα για κάθε τύπου R περιοχή γράφεται ως εξής:

$$MSE(R) = \frac{E(R)}{k^2} = var(R)(1 - c^2) \tag{4.5}$$

όπου  $var(R)$  είναι η διασπορά της τύπου R περιοχής.

Όταν η αντίθεση και η φωτεινότητα παίρνουν διακριτές τιμές  $\{s_i\}$  και  $\{o_j\}$  το σφάλμα έχει την ακόλουθη μορφή:

$$\begin{aligned}
\hat{E}(R) &= \|R - s_i D - o_j \mathbf{1}\|^2 = \\
&= \|R - \bar{r}\mathbf{1} - s_i(D - \bar{d}\mathbf{1}) + (\bar{r} - s_i \bar{d} - o_j)\mathbf{1}\|^2 = \\
&= \|R - \bar{r}\mathbf{1}\|^2 + s_i^2 \|D - \bar{d}\mathbf{1}\|^2 \\
&\quad + k^2 (\bar{r} - s_i \bar{d} - o_j)^2 - 2s_i \langle R - \bar{r}\mathbf{1}, D - \bar{d}\mathbf{1} \rangle
\end{aligned} \tag{4.6}$$

Αυτό το συνολικό σφάλμα είναι και το πραγματικό, καθώς δεν οφείλεται μόνο στην αδυναμία προσέγγισης μεταξύ τύπου R και τύπου D περιοχής αλλά και στη χρήση διακριτών τιμών για τις παραμέτρους του μετασχηματισμού τύπου affine.

Αντικαθιστώντας την αντίθεση και τη φωτεινότητα από τις σχέσεις τους και αφαιρώντας από το συνολικό σφάλμα το θεωρητικό προκύπτει ότι:

$$\widehat{E}(R) - E(R) = (s_i - s)^2 \|D - \bar{d}\mathbf{1}\|^2 + k^2 (\bar{r} - s_i \bar{d} - o_j)^2 \quad (4.7)$$

Από τη σχέση (4.1) που δίνει τη φωτεινότητα παρατηρούμε ότι αυτή δεν αντιπροσωπεύει τίποτε άλλο από τη διαφορά μεταξύ της μέσης τιμής της τύπου R περιοχής και της μέσης τιμής της τύπου D, πολλαπλασιασμένης με την αντίθεση. Κατά την αποκωδικοποίηση όμως, υπολογίζεται σε κάθε επανάληψη και για κάθε περιοχή τύπου R το άθροισμα:

$$R = s_i D + o_j \mathbf{1} \stackrel{(4.1)}{=} s_i (D - \bar{d}\mathbf{1}) + \bar{r}\mathbf{1} \quad (4.8)$$

Η παραπάνω σχέση φανερώνει ότι είναι δυνατό, αντί της φωτεινότητας να κωδικοποιηθεί η μέση τιμή της τύπου R περιοχής. (Η μέση τιμή της τύπου D περιοχής δεν χρειάζεται να αποθηκεύεται καθώς υπολογίζεται κατά την ανακατασκευή της εικόνας στη διαδικασία αποκωδικοποίησης). Όπως δείχνουν και τα πειράματα το δυναμικό εύρος της μέσης τιμής είναι κατα πολύ μικρότερο του αντιστοίχου της φωτεινότητας και επομένως η αποθήκευσή της σε διακριτές τιμές είναι πιο αποδοτική από πλευρά κόστους. Η όλη διεργασία ισοδυναμεί μαθηματικά με την αντικατάσταση  $\bar{r}_j = s_i \bar{d} + o_j$ . Επιπλέον, η χρήση της μέσης τιμής ως μία εκ των παραμέτρων του μετασχηματισμού τύπου affine αντί της αντίθεσης, επιτρέπει την αποδέσμευση της διαδικασίας βελτιστοποίησης από τις δύο affine παραμέτρους επιταχύνοντας περαιτέρω την σύγκριση, αφού δεν χρειάζεται να υπολογιστεί και να διακριτοποιηθεί η φωτεινότητα. Αντίθετα η σύγκριση γίνεται με βάση την (4.8), στην οποία η χρήση της μέσης τιμής της τύπου D περιοχής δεν αυξάνει τον υπολογιστικό φόρτο, καθώς τόσο αυτή όσο και η μέση τιμή της τύπου R περιοχής χρησιμοποιούνται για τον υπολογισμό της αντίθεσης, όπως μαρτυρά η σχέση (4.1). Η χρησιμοποίηση της μέσης τιμής έχει προταθεί και παλαιότερα από τους *Bani Eqbal*<sup>6</sup> και *Oien* και *Lepsoy*<sup>4</sup>.

Τέλος, πρέπει να αναφερθεί, ότι το μέσο τετραγωνικό συνολικό σφάλμα που προκύπτει από την εισαγωγή των διακριτών τιμών όπως και στην περίπτωση του θεωρητικού δίνεται από τη σχέση:

$$MSE'(R) = \frac{\widehat{E}(R)}{k^2} \quad (4.9)$$

Αφαιρώντας τη θεωρητική από την πραγματική τιμή του μέσου τετραγωνικού σφάλματος, εξάγεται η σχέση που εκφράζει το επιπρόσθετο σφάλμα που οφείλεται στη διακριτοποίηση της αντίθεσης και της μέσης τιμής της περιοχής τύπου R και καλείται σφάλμα διακριτοποίησης.

$$E_{\text{quant}} = MSE'(R) - MSE(R) = (s - s_i)^2 \text{var}(D) + (\bar{r} - \bar{r}_j)^2 \quad (4.10)$$

Το ερώτημα που προκύπτει είναι το ακόλουθο: Πώς η αποθήκευση της μέσης τιμής επηρεάζει την θεωρητική βάση στην οποία στηρίχτηκε η μέθοδος των PIFS συστημάτων; Την απάντηση δίνει το γενικευμένο θεώρημα Collage:

#### Θεώρημα 4-Γενικευμένο Θεώρημα Collage<sup>1</sup>:

Έστω ένα σύστημα PIFS που χρησιμοποιεί μέσους όρους αντί για τιμές φωτεινότητας

$$W = \{(D, R, \tau, s, \bar{r})\} \quad (4.11)$$

(το  $\mu$  συμβολίζει τη μέση τιμή και το  $\tau$  τον ισομετρικό μετασχηματισμό), ο παράγοντας συστολής  $p$  του οποίου είναι μικρότερος της μονάδας και μια δεδομένη προς συμπίεση εικόνα  $O$ . Το σύστημα θα είναι συστολικό σε σχέση με τον παράγοντα  $p = \max |s_i| < 1$ , αν

1. Οι μέσες τιμές του  $\widetilde{W}$  είναι στην πραγματικότητα οι μέσες τιμές της εικόνας  $O$  δηλαδή

$$\bar{r}_i = \frac{\iint_{(x,y) \in R_i} O(x,y) dx dy}{\iint_{(x,y) \in R_i} dx dy} \quad (4.12)$$

για κάθε  $i=1,2,\dots,m$  (όπου  $(x,y)$  οι συντεταγμένες των εικονοστοιχείων της εικόνας  $O$  που περιέχει η range  $R_i$ ) και

2.  $d(\widetilde{W}(O), O) < \varepsilon$ , για κάποιο  $\varepsilon$ , τότε για οποιαδήποτε εικόνα  $P$ ,

$$d(\tilde{W}^n(P), O) < \frac{\varepsilon}{1-p}, \quad (4.13)$$

για οποιοδήποτε αρκετά μεγάλο ακέραιο  $n$ .

### Απόδειξη:

Ένα συμβατικό PIFS  $W = \{(D, R, s, \tau, o)\}$ , μπορεί άμεσα να υπολογιστεί με βάση το πρωτότυπο  $\tilde{W}$ , καθώς η τιμή της φωτεινότητας για μίας τύπου R περιοχής R δίνεται από τη σχέση (4.1). Κατά την αποκωδικοποίηση όμως, ενώ το συμβατικό PIFS εφαρμόζει επαναληπτικά το μετασχηματισμό  $sD + o\mathbf{1}$  με τις σταθερές τιμές φωτεινότητας που έχουν υπολογιστεί κατά την κωδικοποίηση, το  $\tilde{W}$ , το οποίο δεν έχει στη διάθεσή του τις τιμές αυτές, εφαρμόζει τον ίδιο μετασχηματισμό χρησιμοποιώντας επαναληπτικά τη σχέση (4.1) με την ακόλουθη μορφή:

$$o_i = \frac{\iint_{(x,y) \in R} O(x,y) dx dy}{\iint_{(x,y) \in R} dx dy} - s \frac{\iint_{(x,y) \in D_{i-1}} \tilde{W}_{i-1}(P)(x,y) dx dy}{\iint_{(x,y) \in D_{i-1}} dx dy} \quad (4.14)$$

Ο δείκτης  $i$  αντιστοιχεί στην τρέχουσα επανάληψη και το  $D_{i-1}$  συμβολίζει την τύπου D περιοχή της αυθαίρετης εικόνας  $P$  μετά από εφαρμογή του  $\tilde{W}$  στην προηγούμενη επανάληψη. Το δεύτερο δηλαδή κλάσμα της (4.14) εκφράζει τη μέση τιμή της μετασχηματισμένης αυτής τύπου D περιοχής, ενώ το πρώτο τη μέση τιμή της περιοχής τύπου R της αρχικής εικόνας.

Μπορεί να αποδειχθεί (Παράρτημα Γ) ότι για κάθε εικόνα  $P$ , ισχύει η ακόλουθη σχέση:

$$d(\tilde{W}(P), O) \leq d(W(P), O), \quad (4.15)$$

που είναι το άμεσο αποτέλεσμα της 1<sup>ης</sup> συνθήκης. Στη συνέχεια, χρησιμοποιούμε επαναληπτικά τη σχέση αυτή. Με τον τρόπο αυτό έχουμε:

$$\begin{aligned}
d(\tilde{W}^n(P), O) &\leq d(W(\tilde{W}^{n-1}(P)), O) \\
&\leq d(W(\tilde{W}^{n-1}(P)), W(O)) + d(W(O), O) \\
&\leq p \cdot d(\tilde{W}^{n-1}(P), O) + \varepsilon \\
&\leq p \cdot (p \cdot d(\tilde{W}^{n-2}(P), O) + \varepsilon) + \varepsilon \\
&\vdots \\
&\leq p^n \cdot d(\tilde{W}^0(P), O) + p^{n-1} \cdot \varepsilon + \dots + p \cdot \varepsilon + \varepsilon \\
&\leq p^n \cdot d(P, O) + \frac{d(\tilde{W}(O), O)}{1-p}
\end{aligned} \tag{4.16}$$

Η 2<sup>η</sup> συνθήκη συμπληρώνει το θεώρημα  $\square$ .

Στην πράξη, όλα τα πειράματα έχουν δείξει ότι

$$d(\tilde{W}^n(P), O) \leq d(W^n(P), O), \tag{4.17}$$

για μεγάλο ακέραιο  $n$ .

Το συμπέρασμα της προηγούμενης μελέτης είναι ότι η χρήση μέσων τιμών βελτιώνει την απόδοση της συμπίεσης. Το γεγονός όμως αυτό επιφέρει κάποιες αναγκαίες αλλαγές στην αποκωδικοποίηση. Σε κάθε επανάληψη, αντί να προστίθεται η τιμή της φωτεινότητας  $o$  στην εκάστοτε περιοχή, αυτή πρέπει να υπολογιστεί πρώτα ως διαφορά μεταξύ της επιθυμητής τιμής της έντασης και της έντασης της περιοχής στην τρέχουσα επανάληψη και στη συνέχεια να εφαρμοστεί ο συστολικός μετασχηματισμός  $sD + o\mathbf{1}$ . Αν και η υπολογιστική πολυπλοκότητα αυξάνεται, το αντάλλαγμα είναι ο μικρότερος αριθμός επαναλήψεων.

Τα συμπεράσματα και οι υποθέσεις για την υπεροχή της συμπίεσης με βάση τη μέση τιμή ως προς αυτή με βάση τη φωτεινότητα έρχονται να επιβεβαιώσουν δύο πίνακες αποτελεσμάτων αποκωδικοποίησης για το συμβατικό fractal αλγόριθμο (πίνακες 4.1 και 4.2). Ο πρώτος αφορά αλγόριθμο που διακριτοποιεί τη φωτεινότητα ενώ ο δεύτερος τη μέση τιμή.

Πίνακας 4.1: Αποτελέσματα σύγκλισης για τον απλό fractal αλγόριθμο. Η στήλη με την ετικεταλίδα Αρχική εικόνα αντιστοιχεί στη διαφορά μεταξύ αποκωδικοποιημένης εικόνας και της αρχικής προς συμπίεση εικόνας, η 2<sup>η</sup> στήλη αντιστοιχεί στην απόσταση της αποκωδικοποιημένης εικόνας από το όριο του συστολικού μετασχηματισμού και η 3<sup>η</sup> στήλη στην απόσταση μεταξύ της αποκωδικοποιημένης εικόνας στην τρέχουσα επανάληψη και αυτής στην προηγούμενη.



Ευκλείδεια απόσταση	Lena			Golden Hill		
	Αριθμός Επαναλ.	Αρχική εικόνα	Όριο Μετ/σμού	Προηγούμενη	Αρχική εικόνα	Όριο Μετ/σμού
0	48.016402	47.803839		49.891805	49.595791	
1	33.164421	32.869468	50.160133	31.820888	31.390948	46.880547
2	17.771090	17.348753	31.199474	17.035684	16.403380	27.298880
3	7.801571	6.959135	14.851330	8.794332	7.672172	12.666736
4	4.604374	3.074657	5.086840	5.859413	4.059589	4.579958
5	3.792194	1.708593	1.685629	4.797178	2.340094	1.911426
6	3.528083	1.050378	.860085	4.398248	1.401360	1.071217
7	3.432070	0.735213	.613695	4.250532	.894098	.736818
8	3.395904	0.564582	.463661	4.197342	.652312	.553793
9	3.381499	0.436850	.357641	4.177849	.499110	.418407
10	3.375080	.338697	.275903	4.170598	.382284	.320890
11	3.372592	.263274	.213079	4.167727	.293353	.245123
12	3.371729	.207675	.161815	4.166111	.225981	.187052
13	3.371436	.164863	.126290	4.164938	.177777	.139508
14	3.371220	.133857	.096240	4.164147	.138837	.111036
15	3.371217	.109148	.077488	4.164061	.109532	.085314
16	3.371045	.089525	.062439	4.163969	.090394	.061856
17	3.370769	.071443	.053950	4.164169	.074552	.051118
18	3.370539	.054966	.045638	4.164149	.058659	.046013
19	3.370425	.042791	.034499	4.164252	.042702	.040217
20	3.370346	.031433	.029035	4.164183	.031128	.029232
21	3.370448	.015747	.027204	4.164124	.021036	.022944
22	3.370388	.005167	.014875	4.164116	.011220	.017794
23	3.370394	.000000	.005167	4.164091	.000000	.011220
24	3.370394	.000000	.000000	4.164091	.000000	.000000

Πίνακας 4.2: Αριθμός επαναλήψεων και σύγκλιση με χρήση των μέσων τιμών των εικόνων που δημιουργεί ο μετασχηματισμός τύπου affine (attractors) κατά την αποκωδικοποίηση.

Ευκλείδεια απόσταση	Lena			Golden Hill		
	Αριθμός Επαναλ.	Αρχική εικόνα	Όριο Μετ/σμού	Προηγούμενη	Αρχική εικόνα	Όριο Μετ/σμού

0	48.016402	47.803839		49.891805	49.595829	
1	9.313221	8.481671	47.236881	10.382104	9.210134	48.919721
2	4.433032	2.822330	8.275642	5.389032	3.298798	8.790541
3	3.421712	.632966	2.705847	4.215694	.716302	3.168183
4	3.375076	.287388	.607858	4.167173	.303169	.692886
5	3.371716	.196568	.279768	4.164997	.205654	.300229
6	3.370957	.147354	.189715	4.164862	.155883	.202476
7	3.371010	.112741	.146341	4.164570	.119827	.151943
8	3.370927	.090436	.114053	4.164164	.095484	.115118
9	3.371030	.071309	.091422	4.164281	.075240	.090310
10	3.370755	.053667	.070799	4.164249	.059786	.073418
11	3.370579	.045721	.059594	4.164221	.045184	.058561
12	3.370724	.037263	.050631	4.164218	.032973	.044495
13	3.370605	.027896	.043057	4.164164	.024316	.033716
14	3.370511	.022183	.033885	4.164135	.018113	.026851
15	3.370626	.012040	.025240	4.164101	.008286	.019137
16	3.370577	.005167	.013102	4.164073	.000000	.008286
17	3.370583	.000000	.005167			

Συγκρίνοντας τον πίνακα 4.2 με τον πίνακα 4.1, παρατηρούμε ότι η ποιότητα ενός αποκωδικοποιητή μέσω των τιμών στην 4<sup>η</sup> επανάληψη είναι η ίδια τόσο οπτικά όσο και αριθμητικά με αυτή ενός αποκωδικοποιητή διόρθωσης φωτεινότητας στην 10<sup>η</sup> επανάληψη.

Αν η μέση τιμή δεν υπολογίζεται σε κάθε επανάληψη, αλλά χρησιμοποιούνται πάντα οι μέσες τιμές των περιοχών της αρχικής εικόνας, τότε δε μπορεί να γίνει λόγος για σύγκλιση. Παρόλα αυτά το αποτέλεσμα είναι πάντα καλύτερο -κάπου ανάμεσα στο όριο του μετασχηματισμού τύπου affine και την αρχική εικόνα- όπως δείχνει ο πίνακας 4.3.

Πίνακας 4.3: Αριθμός επαναλήψεων και σύγκλιση με τη χρήση των μέσων τιμών έντασης των αρχικών εικόνων

Ευκλείδεια απόσταση	Lena		Golden Hill	
	Αρχική εικόνα	Προηγούμενη	Αρχική εικόνα	Προηγούμενη
Αριθμός Επαναλ.				

0	48.016402		49.891805	
1	11.496596	46.617727	12.037807	48.404100
2	6.815158	10.079661	7.849318	9.611075
3	3.749522	5.157243	4.728427	5.564765
4	3.396842	1.508483	4.209134	1.919773
5	3.364696	.682158	4.157121	.786823
6	3.361183	.533536	4.152344	.557941
7	3.362913	.486842	4.151988	.497280
8	3.362926	.461244	4.151451	.465315
9	3.361212	.441506	4.150711	.446070
10	3.361645	.430198	4.151849	.428095
11	3.361931	.419995	4.153813	.414541
12	3.361653	.414003	4.150401	.411605
...				
99	3.360668	.396114	4.152402	.409602
100	3.361808	.400400	4.152022	.409803

#### 4.1.2 Πίνακες Διακριτοποίησης

Οι μέσες τιμές εξαρτώνται από τη διαμέριση της εικόνας. Μία χρήσιμη τεχνική που χρησιμοποιείται στις απωλεστικές μεθόδους συμπίεσης, η οποία μειώνει το μέγεθος των δεδομένων που καταλαμβάνουν οι μέσες τιμές διατηρώντας παράλληλα την καλή οπτική ποιότητα της εικόνας, είναι η διακριτοποίηση της μέσης διαφοράς (διαφοράς μεταξύ μέσων τιμών). Οι έρευνες έχουν δείξει ότι το ανθρώπινο μάτι είναι πιο ευαίσθητο στις αλλαγές της φωτεινότητας σε λείες ομογενείς περιοχές, παρά σε περιοχές με μεγάλη διακύμανση στην υφή.

Επομένως οι μέσες διαφορές διακριτοποιούνται με τέτοιο τρόπο ώστε οι μικρές διαφορές να είναι πιο ακριβείς από τις μεγαλύτερες.

Οι πίνακες 4.4 και 4.5 αποτελούν δύο από τους πιο διαδεδομένους πίνακες διακριτοποίησης που χρησιμοποιούνται στις εφαρμογές. Καλούνται *πίνακας υψηλής* και *πίνακας χαμηλής διακριτοποίησης* αντίστοιχα (*fine quantization table* και *coarse quantization table*).

Πίνακας 4.4: Πίνακας υψηλής διακριτοποίησης. Οι στήλες με την έντονη γραμματοσειρά αντιστοιχούν

στις αντιπροσωπευτικές τιμές στις οποίες διακριτοποιούνται οι μέσες διαφορές της στήλης με κανονική γραμματοσειρά.

αντιπροσωπευτικό	τιμές	αντιπροσωπευτικό	τιμές
<b>0</b>	0		
<b>1</b>	1	<b>-1</b>	-1
<b>2</b>	2	<b>-2</b>	-2
<b>3</b>	3, 4	<b>-3</b>	-3, -4
<b>6</b>	5, 6, 7	<b>-6</b>	-5, -6, -7
<b>9</b>	8, 9, 10	<b>-9</b>	-8, -9, -10
<b>12</b>	11, 12, 13	<b>-12</b>	-11, -12, -13
<b>15</b>	14, 15, 16, 17	<b>-15</b>	-14, -15, -16, -17
<b>5k, k=4,..., 50</b>	5k-2,..., 5k+2	<b>-5k, k=4,..., 50</b>	-5k+2,..., -5k-2
<b>255</b>	253, 254, 255	<b>-255</b>	-253, -254, -255

Πίνακας 4.5: Πίνακας χαμηλής διακριτοποίησης

αντιπροσωπευτικό	τιμές	αντιπροσωπευτικό	τιμές
<b>0</b>	0		
<b>1</b>	1,2	<b>-1</b>	-1,-2
<b>4</b>	3, 4, 5	<b>-4</b>	-3, -4, -5
<b>7</b>	6, 7, 8	<b>-7</b>	-6, -7, -8
<b>10</b>	9, 10,11, 12	<b>-10</b>	-9, -10,-11, -12
<b>5k, k=3, 4, 5, 6</b>	5k-2,..., 5k+2	<b>-5k, k=3, 4, 5, 6</b>	-5k+2,...,-5k-2
<b>37+7k, k=0, 1, 2, 3, 4</b>	34+7k, 35+7k,...,40+7k	<b>-37-7k, k=0, 1, 2, 3, 4</b>	-34-7k, -35-7k,..., -40-7k
<b>74+9k, k=0, 1, 2, 3, 4</b>	70+9k, 71+9k,...,78+9k	<b>-74-9k, k=0, 1, 2, 3, 4</b>	-70-9k, -71-9k,..., -78-9k
<b>121+11k, k=0, 1, 2, 3, 4</b>	116+11k, 117+11k,...,126+11k	<b>-121-11k, k=0, 1, 2, 3, 4</b>	-116-11k, -117- 11k,...,-126-11k
<b>180+15k, k=0, 1, 2, 3</b>	173+15k, 174+15k, ...,187+15k	<b>-180-15k, k=0, 1, 2, 3</b>	-173-15k, -174- 15k, ..., -187-15k
<b>240</b>	233, 234,...,255	<b>-240</b>	-233, -234,...,-255

Ο πρώτος αντιστοιχεί 511 μέσες τιμές διαφορών σε 111 αντιπροσωπευτικούς αριθμούς, ενώ ο δεύτερος μειώνει τον αριθμό αυτών των αντιπροσωπευτικών τιμών στις 57. Οι πίνακες 4.6 και 4.7 που ακολουθούν αποδεικνύουν ότι η βελτίωση στα αριθμητικά αποτελέσματα είναι εμφανής. Από τις στατιστικές

που παρουσιάζουν καθώς και από οπτικά πειράματα συνάγεται ότι σε περιπτώσεις απωλεστικής συμπίεσης πρέπει να χρησιμοποιείται ένα είδος πίνακα διακριτοποίησης. Όταν το ζητούμενο είναι ο υψηλός βαθμός συμπίεσης, ο πίνακας χαμηλής διακριτοποίησης είναι ο πλέον κατάλληλος.

Πίνακας 4.6: Σύγκριση της απόδοσης της συμπίεσης για την εικόνα της Lena για ένα fractal αλγόριθμο που διαιρεί την εικόνα σε range περιοχές 4\*4 pixels. Το λ συμβολίζει το βαθμό παραμόρφωσης (distortion rate).

Χωρίς διακριτοποίηση			Υψηλή διακριτοποίηση			Χαμηλή διακριτοποίηση		
λ	PSNR	bpp	λ	PSNR	bpp	λ	PSNR	bpp
8	37.03	1.1983	5	37.21	1.1801	4	37.08	1.2137
15	36.52	1.0027	9	36.97	0.9916	8	36.86	0.9960
33	35.63	0.8063	18	36.51	0.8080	16	36.45	0.8044
112	33.85	0.6008	51	35.44	0.6029	42	35.55	0.6025

Πίνακας 4.7: Σύγκριση της απόδοσης της συμπίεσης του ίδιου αλγορίθμου για την εικόνα Golden Hill.

Χωρίς διακριτοποίηση			Υψηλή διακριτοποίηση			Χαμηλή διακριτοποίηση		
λ	PSNR	bpp	λ	PSNR	bpp	λ	PSNR	bpp
17	35.06	1.3596	8	35.32	1.3450	6	35.21	1.3514
27	34.51	1.2043	17	35.02	1.2031	15	34.97	1.2088
48	33.53	1.0022	33	34.37	1.0017	30	34.37	1.0053
86	32.23	0.8004	61	33.34	0.8024	57	33.41	0.8011

Ανακεφαλαιώνοντας, σε ένα απωλεστικό fractal σύστημα συμπίεσης, οι μέσες τιμές πρέπει να προτιμώνται από τη διόρθωση φωτεινότητας και πρέπει να συμπιέζονται υπό τη μορφή διακριτοποιημένων διαφορών.

## 4.2 Διακριτοποίηση της αντίθεσης

Όπως είναι γνωστό από το θεώρημα του *Barnsley*, το σφάλμα αποκωδικοποίησης είναι φραγμένο από το σφάλμα προσέγγισης (*collage error*) κατά ένα παράγοντα ίσο με  $1/(1-s)$ . Ως συνέπεια, για τιμές της αντίθεσης που είναι κοντά στη μονάδα πρέπει να ληφθεί υπόψη ότι το σφάλμα μπορεί να λάβει πολύ μεγάλες τιμές (*penalty values*). Πριν ασχοληθούμε με το πρόβλημα αυτό πρέπει να απαντήσουμε στο ερώτημα που αφορά τα δυαδικά ψηφία που δεσμεύουμε για την αποθήκευση των διακριτών τιμών της αντίθεσης. Ο πιο απλός τρόπος για τη μελέτη του προβλήματος είναι η αποκωδικοποίηση μίας

εικόνας με τη χρήση διαφορετικών μεταξύ τους τιμών αντίθεσης, με μικρή όμως απόκλιση από τιμή σε τιμή. Ο παρακάτω πίνακας περιέχει τα σφάλματα αποκωδικοποίησης μιας εικόνας που συμπίεστηκε χρησιμοποιώντας σταθερή τιμή για την αντίθεση και ίση με 0.75 και αποσυμπιέστηκε με τη χρήση μεταβλητών τιμών αντίθεσης σε ένα PIFS που χρησιμοποιεί μέσες τιμές έντασης του γκρι.

Πίνακας 4.8: Αποσυμπίεση εικόνας που έχει συμπιεστεί με  $s=3/4$  με μεταβλητές τιμές αντίθεσης.

s	Lena		Golden Hill	
	Rms σφάλμα	PSNR	Rms σφάλμα	PSNR
1/4	8.7213	29.32	9.4793	28.60
1/2	5.7074	33.00	6.5354	31.83
5/8	4.2454	35.57	5.0604	34.05
11/16	3.6726	36.83	4.4779	35.11
23/32	3.4738	37.31	4.2726	35.52
47/64	3.4071	37.48	4.2017	35.66
<b>3/4</b>	<b>3.3611</b>	<b>37.60</b>	<b>4.1519</b>	<b>35.77</b>
49/64	3.3546	37.62	4.1317	35.81
25/32	3.3694	37.58	4.1393	35.79
13/16	3.4867	37.28	4.2479	35.57
7/8	4.0761	35.93	4.8383	34.44
1	6.3184	32.12	7.3014	30.86
5/4	9.3508	28.71	10.9888	27.31

Τα δεδομένα του πίνακα καθιστούν φανερό ότι δεν υπάρχει λόγος να επιλέγονται τιμές αντίθεσης που απέχουν μεταξύ τους βήμα μικρότερο του 1/32. Αυτό που φαίνεται αρχικά περιεργό είναι ότι η βέλτιστη τιμή της αντίθεσης για την αποκωδικοποίηση δεν είναι η ίδια με αυτή που χρησιμοποιήθηκε κατά τη συμπίεση αλλά μία ελαφρώς μεγαλύτερη. (Ο πίνακας δείχνει ότι για την εικόνα της Lena που συμπίεστηκε με  $s=0.75$  η βέλτιστη τιμή αποκωδικοποίησης είναι η  $s=49/64$ . Το φαινόμενο δεν οφείλεται σε κάποιο σφάλμα στρογγυλοποίησης κατά την εφαρμογή της μεθόδου, αλλά στο γεγονός ότι το μοντέλο PIFS έχει την ιδιότητα να κάνει πιο ομοιόμορφη την εικόνα. Αυξάνοντας ελαφρώς την αντίθεση, τα επιπλέον σφάλματα από περιοχή σε περιοχή αλληλοαναιρούνται μερικώς, με αποτέλεσμα η ολική εικόνα να φαίνεται καλύτερη, γεγονός που φαίνεται να έρχεται σε αντίθεση με το θεώρημα Collage.

Την ιδιότητα όμως αυτή της αντίθεσης επιβεβαιώνει και το σχήμα 4.4: *μία μικρή τιμή αντίθεσης εξομαλύνει την εικόνα ενώ η μεγάλη τονίζει την υφή.*



Σχ.4.4: Η εικόνα της Lena συμπιεσμένη με  $s=3/4$  και αποσυμπιεσμένη με μεταβλητές τιμές αντίθεσης.



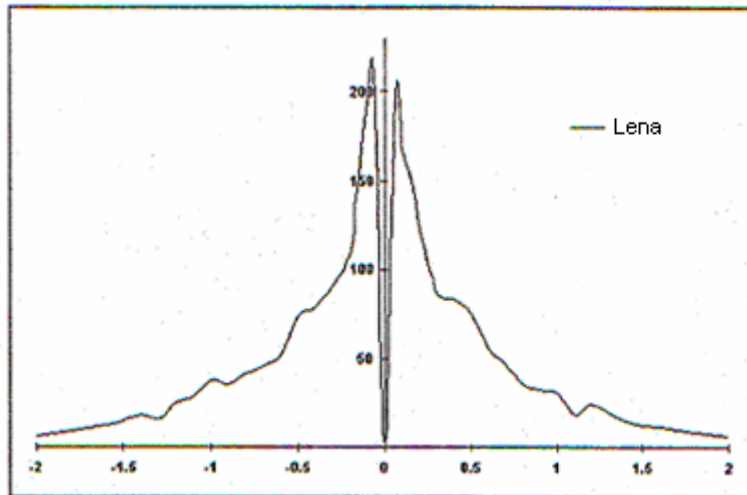
Σχ.4.4(συνέχεια): Η εικόνα της Lena συμπίεσμένη με  $s=3/4$  και αποσυμπιεσμένη με μεταβλητές τιμές αντίθεσης.

Παρατηρούμε ότι οι δύο τελευταίες τιμές της αντίθεσης στο παραπάνω παράδειγμα υπερβαίνουν την μονάδα, γεγονός που θέτει σε κίνδυνο τη σύγκλιση της διαδικασίας. Η μέση τιμή όμως της έντασης της τύπου R περιοχής διασφαλίζει το γεγονός ότι η αποκωδικοποιημένη εικόνα δε θα απέχει σημαντικά από την αρχική. Αυτό είναι και το κυριότερο πλεονέκτημα της χρήσης της μέσης τιμής αντί του όρου της φωτεινότητας. Στην πραγματικότητα, στην περίπτωση αυτή, το σφάλμα που οφείλεται σε μία μεγάλη τιμή αντίθεσης μπορεί να αγνοηθεί.

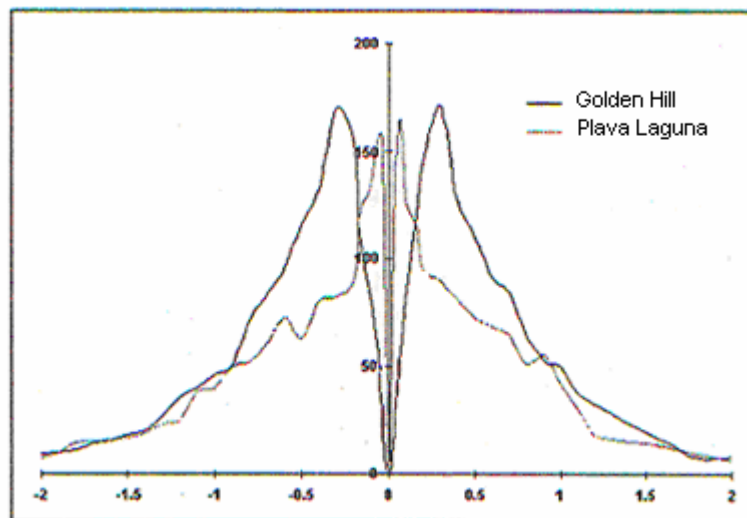


### 4.2.1 Διακριτές τιμές αντίθεσης

Έστω ότι συμπιέζονται δύο εικόνες, αυτή της Lena και της Plava Laguna, χρησιμοποιώντας τις βέλτιστες τιμές αντίθεσης για κάθε fractal κώδικα. Τα ακόλουθα ιστογράμματα εμφανίζουν την διασπορά των τιμών αντίθεσης και για τις δύο εικόνες. Παρατηρούμε ότι η αντίθεση παίρνει διάφορες τιμές.



Σχ.4.5: Ιστόγραμμα της διασποράς της αντίθεσης για την εικόνα της Lena



Σχ.4.6: Ιστόγραμμα της διασποράς της αντίθεσης για τις εικόνες Golden Hill και Plava Laguna.

Συγκρίνοντας μία εικόνα που έχει συμπιεστεί και αποσυμπιεστεί με μεταβλητές τιμές αντίθεσης με την ίδια εικόνα, όταν αυτή έχει υποστεί την ίδια επεξεργασία με μία μόνο σταθερή τιμή αντίθεσης, παρατηρούμε ότι δεν υπάρχει οπτική διαφορά όπως δείχνει το σχήμα 4.7. Στην πραγματικότητα, οι επιπρόσθετες τιμές παράγουν πιο ακριβή τοπική αντίθεση μέσα στην εικόνα χωρίς όμως να αλλάζουν την υφή. Καθώς το ανθρώπινο μάτι είναι ιδιαίτερα ευαίσθητο στα γεωμετρικά σχήματα και την υφή, συμπεραίνουμε ότι η δέσμευση ψηφίων για την αποθήκευση διακριτών τιμών αντίθεσης έχει σημαντική αξία. Κατά συνέπεια δεν υπάρχει λόγος για τη χρήση περισσότερων από μερικές σταθερές τιμές.

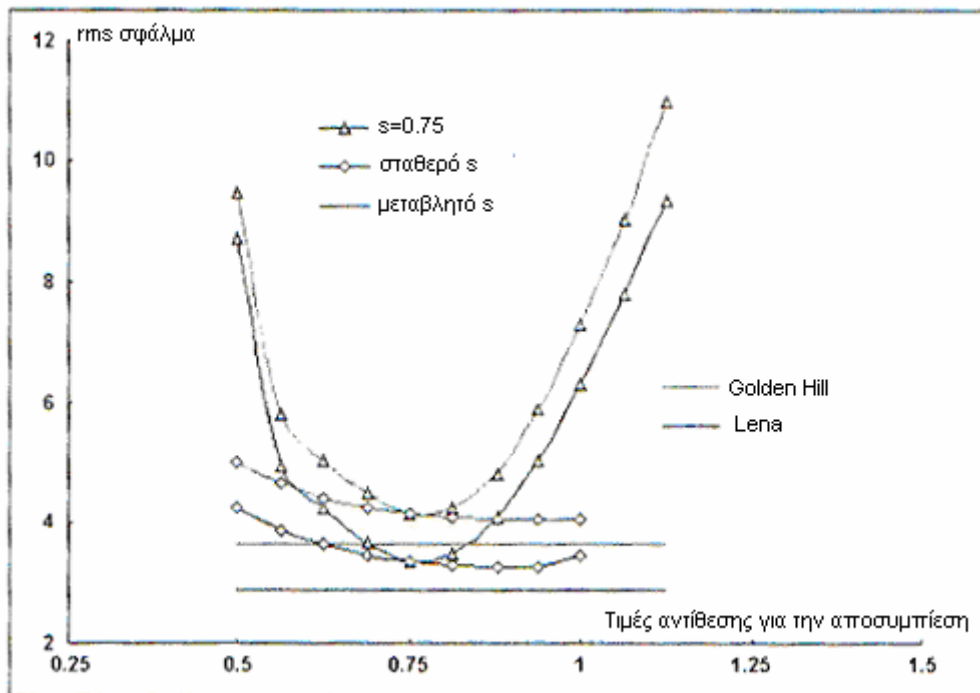


Μεταβλητό  $s$  (rms σφάλμα =2.91)



Σταθερό  $s=1$  (rms σφάλμα=3.28)

Σχ.4.7: Μεταβλητή τιμή αντίθεσης σε σχέση με σταθερή τιμή αντίθεσης.



Σχ.4.8: Η επιλογή της αντίθεσης.

Συνοπτικά όλες οι παραπάνω στατιστικές παρατηρήσεις περιλαμβάνονται στο σχήμα 4.8:

1. Οι ευθείες γραμμές αντιπροσωπεύουν τις βέλτιστες fractal αναπαραστάσεις. Η εικόνα έχει συμπιεστεί και αποσυμπίεστεί χρησιμοποιώντας μεταβλητές τιμές αντίθεσης.
2. Οι καμπύλες που μοιάζουν με παραβολές και οι οποίες ελαχιστοποιούνται στο  $s=3/4$ , είναι οι fractal αναπαραστάσεις που αντιστοιχούν σε συμπίεση με σταθερή τιμή αντίθεσης ίση με  $3/4$  και αποσυμπίεστεί με τη χρήση ενός μικρού συνόλου διαφορετικών σταθερών τιμών αντίθεσης.
3. Οι ακολουθίες από μικρούς ρόμβους αντιστοιχούν σε fractal αναπαραστάσεις που έχουν δημιουργηθεί με συμπίεση με σταθερές τιμές αντίθεσης και αποσυμπίεση με τις ίδιες αυτές τιμές.

Και για τα τρία ζεύγη παράλληλων καμπυλών, αυτή που βρίσκεται ψηλότερα αντιστοιχεί στην εικόνα Golden Hill ενώ η χαμηλότερη στη Lena.

Πίνακας 4.9: Διαφορά μεταξύ σφάλματος αποκωδικοποίησης και σφάλματος προσέγγισης.

Rms σφάλμα	Lena			Golden Hill		
s	Σφάλμα προσέγγισης	Σφάλμα αποκωδ.	Διαφορά	Σφάλμα προσέγγισης	Σφάλμα αποκωδ.	Διαφορά
1/2	3.7807	4.2684	0.4877	4.4683	5.0345	0.5662
9/16	3.5329	3.8880	0.3551	4.2471	4.6626	0.4155
5/8	3.3587	3.6348	0.2761	4.0892	4.4093	0.3201
11/16	3.2369	3.4624	0.2255	3.9739	4.2526	0.2787
3/4	3.1542	3.3611	0.2069	3.8897	4.1519	0.2622
13/16	3.0988	3.3000	0.2012	3.8328	4.0982	0.2654
7/8	3.0649	3.2698	0.2049	3.7954	4.0715	0.2761
15/16	3.0554	3.2685	0.2131	3.7764	4.0704	0.2940
1	3.0580	3.2790	0.2210	3.7627	4.0665	0.3038
όλες	2.6335	2.9082	0.2747	3.3103	3.6471	0.3368

Αν για λόγους εξοικονόμησης θέσεων στη μνήμη πρέπει να χρησιμοποιηθεί μία μόνο σταθερή τιμή αντίθεσης ποια πρέπει να επιλεχθεί; Υπάρχει και ένας πρόσθετος παράγοντας που δεν πρέπει να αγνοηθεί: η διαφορά ανάμεσα στο σφάλμα collage (που όπως είναι γνωστό είναι το σφάλμα προσέγγισης κατά τη συμπίεση) και το σφάλμα αποκωδικοποίησης (που όπως έχει προαναφερθεί αντιπροσωπεύει την πραγματική διαφορά μεταξύ αποσυμπιεσμένης και αρχικής εικόνας). Αυτή η διαφορά εικονίζεται στον πίνακα 4.9. Όσο μικρότερη είναι, τόσο καλύτερα συγκλίνει ο fractal αλγόριθμος. Στην πραγματικότητα, αυτός είναι και ο λόγος που επιλέχθηκε η τιμή  $s=3/4$  στην παραπάνω συζήτηση για την αντίθεση.

Ολοκληρώνοντας τη μελέτη της διακριτοποίησης των παραμέτρων του συστολικού μετασχηματισμού, θα αναφέρουμε και μία άλλη μέθοδο που διακριτοποιεί ταυτόχρονα την αντίθεση και τη μέση τιμή των τύπου R περιοχών της εικόνας, που προτάθηκε από τους Tong και P<sup>2</sup>.

### 4.3 Βέλτιστη δέσμευση ψηφίων για τη διακριτοποίηση των fractal παραμέτρων

Από τη σχέση (4.1) που μας δίνει τη φωτεινότητα, αναμένεται ότι θα υπάρχει μεγάλη συσχέτιση (αρνητική) μεταξύ φωτεινότητας και αντίθεσης, πράγμα που επιβεβαιώνουν και τα αποτελέσματα του πίνακα 4.10. Η συσχέτιση αυτή δυσκολεύει την προσπάθεια δέσμευσης ενός βέλτιστου αριθμού ψηφίων για τη φωτεινότητα και την αντίθεση με αποτέλεσμα πολλοί ερευνητές να καταφεύγουν στη μέθοδο της δοκιμής (και του σφάλματος).

Πίνακας 4.10: Συσχετίσεις μεταξύ αντίθεσης, φωτεινότητας και μέσης τιμής για διάφορες εικόνες.

Εικόνες	Συσχετίσεις μεταξύ αντίθεσης και φωτεινότητας	Συσχετίσεις μεταξύ αντίθεσης και μέσης τιμής περιοχών τύπου R
Lena	-0.9595	0.0008
Κυρία	-0.8725	0.0321
Καμεραμάν	-0.8093	0.0069
Μπαμπουίνος	-0.9721	-0.0029
Παλιό Σπίτι	-0.9696	0.0023
Σχέδιο	-0.9045	0.1091

Αν παρατηρήσουμε όμως τη συσχέτιση μεταξύ της μέσης τιμής και της αντίθεσης όπως αυτή εικονίζεται στην 3<sup>η</sup> στήλη του πίνακα 4.10, για διάφορες χαρακτηριστικές δοκιμαστικές εικόνες, διαπιστώνουμε ότι τα δύο αυτά μεγέθη είναι πρακτικά ασυσχέτιστα ή τουλάχιστον παρουσιάζουν συσχέτιση έως και 3 τάξεις μικρότερη αυτής μεταξύ αντίθεσης και φωτεινότητας. Λαμβάνοντας υπόψη την παραπάνω παρατήρηση καθώς και τη μορφή που έχει το συνολικό μέσο τετραγωνικό σφάλμα οι συγγραφείς πρότειναν την παρακάτω μέθοδο ταυτόχρονης διακριτοποίησης.

Για τις περισσότερες εικόνες, οι μέσες τιμές έντασης των τύπου R περιοχών έχουν μια ομοιόμορφη κατανομή και επομένως πρέπει και να έχουν μία ομοιόμορφη διακριτοποίηση (όπως συνήθως γίνεται για τη συνιστώσα της

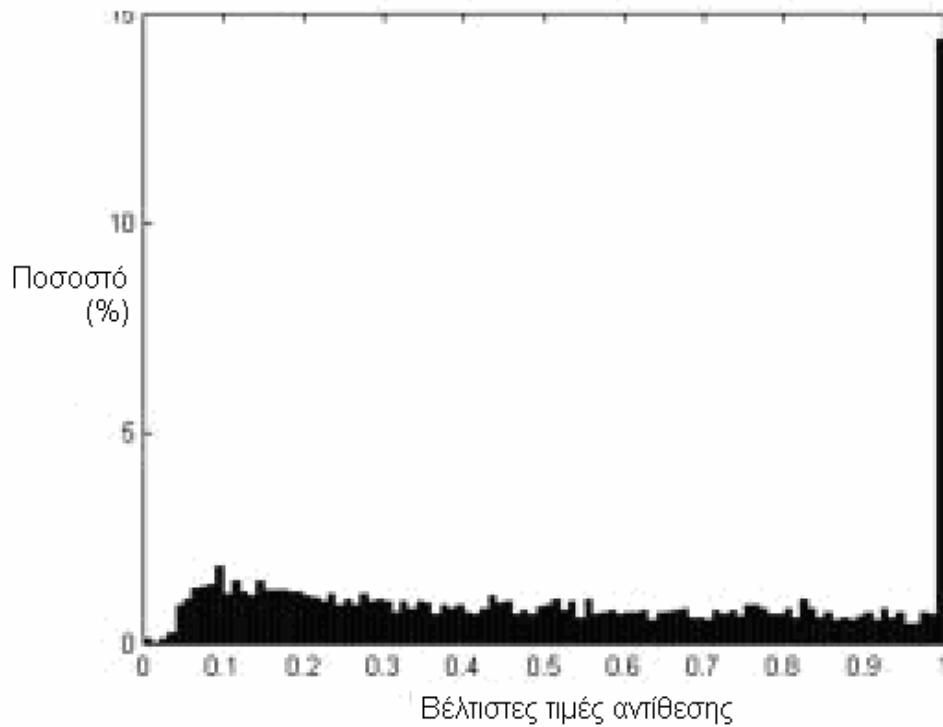
φωτεινότητας). Αν χρησιμοποιηθεί η συνήθης μέθοδος διακριτοποίησης Lloyd-Max<sup>5-6</sup> το σφάλμα διακριτοποίησης  $E_{\bar{r}}$  δίνεται από τον τύπο:

$$E_{\bar{r}} = \frac{d_{\bar{r}}^2}{12 \times 2^{2b_i}} \quad (4.18)$$

όπου  $d_{\bar{r}} = \max(\bar{r}) - \min(\bar{r})$ , είναι το δυναμικό εύρος των μέσων τιμών των range περιοχών και  $b_i$  ο αριθμός των δυαδικών ψηφίων που έχουν δεσμευθεί για τη διακριτοποίηση τους.

Η διακριτοποίηση της αντίθεσης είναι πιο πολύπλοκη. Από την (4.10) συμπεραίνουμε ότι η βέλτιστη διακριτοποίηση για την αντίθεση εξαρτάται από την τυπική απόκλιση των τιμών της έντασης της βέλτιστης περιοχής τύπου D. Αυτό εξάλλου επιβεβαιώνεται και από το γεγονός ότι στην περίπτωση ομοιόμορφης περιοχής τύπου D με πολύ μικρή δηλαδή διασπορά, η αντίθεση δεν παίζει σημαντικό ρόλο κατά την αποκωδικοποίηση:  $R = s_i(D - \bar{d}\mathbf{1}) + \bar{r}_j\mathbf{1}$  αλλά καθώς  $D \approx \bar{d}\mathbf{1}$ , τότε  $R \approx \bar{r}_j\mathbf{1}$  και επομένως η διακριτή της τιμή μπορεί να απέχει περισσότερο από τη θεωρητική. Μια τέτοια όμως προσαρμοστική μέθοδος διακριτοποίησης που θα εξαρτάται από την κατανομή των διασπορών των τύπου D περιοχών μιας εικόνας είναι εξαιρετικά επίπονη στην εφαρμογή της. Για το λόγο αυτό οι συγγραφείς όπως και οι περισσότεροι ερευνητές, προτείνουν τη δέσμευση σταθερού αριθμού ψηφίων για τη διακριτοποίηση της αντίθεσης.

Έχει αποδειχθεί μετά από πολλές πειραματικές μελέτες ότι οι μη αρνητικές τιμές της αντίθεσης κατανέμονται ομοιόμορφα εκτός μιας πολύ υψηλής κορυφής στο μέγιστο της τιμής της που οφείλεται σε ψαλιδισμό. Την αξιοπιστία της παραπάνω πρότασης έρχεται να επιβεβαιώσει και η κατανομή των τιμών της αντίθεσης για την εικόνα της Lena (σχ.4.9).



Σχ.4.9: Η κατανομή των μη αρνητικών βέλτιστων τιμών της αντίθεσης για την εικόνα της Lena.

Η συνήθης διακριτοποίηση Lloyd-Max για την αντίθεση είναι και αυτή μία ομοιόμορφη διακριτοποίηση δηλαδή η διαφορά μεταξύ δύο συνεχόμενων διακριτών τιμών είναι ίση με  $2^{-b_2}$  όπου  $b_2$  ο αριθμός των δυαδικών ψηφίων που δεσμεύθηκαν για την αναπαράσταση της αντίθεσης. Η κορυφή όμως στο  $s=1$  σημαίνει ότι για το τελευταίο διάστημα (το διάστημα δηλαδή μεταξύ των δύο τελευταίων διακριτών τιμών που συμβολίζεται με  $s \in [t_l, t_{l+1}], l = 2^{b_2}$ ), έχουμε  $s_l = E(s | s \in [t_l, t_{l+1}]) = 1$ . Σύμφωνα με την παραπάνω παρατήρηση, η διακριτοποίηση της αντίθεσης είναι μεν ομοιόμορφη αλλά με το περιορισμό  $s_l = 1$  για το τελευταίο στάδιο ανακατασκευής του συνόλου των τιμών της αντίθεσης. Για παράδειγμα, αν τα ψηφία που δεσμεύονται είναι δύο, τότε οι τέσσερις διακριτές τιμές πρέπει να είναι  $\{0.25, 0.5, 0.75, 1\}$ .

Στο τελευταίο μεταβατικό διάστημα το σφάλμα διακριτοποίησης είναι περίπου μηδεν εξαιτίας της κορυφής στο  $s=1$ . Κατά συνέπεια, το αντίστοιχο σφάλμα διακριτοποίησης δίνεται από τη σχέση:

$$E_s = \frac{(2^{b_2} - 1)}{12 \times 2^{3b_2}} \approx \frac{1}{12 \times 2^{2b_2}} \quad (4.19)$$

Η παραπάνω προσέγγιση θεωρείται ικανοποιητική όταν ο αριθμός των ψηφίων για την αντίθεση δεν είναι υπερβολικά μικρός (π.χ., μεγαλύτερος ή ίσος του 2).

Αντικαθιστώντας στην (4.10) τα αντίστοιχα επιμέρους σφάλματα διακριτοποίησης (4.18) και (4.19) και υποθέτοντας ότι το σφάλμα διακριτοποίησης για την αντίθεση είναι ανεξάρτητο της τυπικής απόκλισης της βέλτιστης τύπου D περιοχής, βρίσκεται ότι ο μέσος όρος του σφάλματος διακριτοποίησης είναι ο ακόλουθος:

$$\overline{E_{quant}} \approx \frac{1}{12 \times 2^{2b_2}} \overline{\text{var}(D)} + \frac{d_r^2}{12 \times 2^{2b_1}} \quad (4.20)$$

όπου  $\overline{\text{var}(D)}$  είναι η μέση τιμή της διασποράς της βέλτιστης περιοχής τύπου D.

Για την εύρεση του βέλτιστου συνδυασμού μεταξύ των  $b_1$ ,  $b_2$  για την ταυτόχρονη αναπαράσταση της αντίθεσης και της μέσης τιμής χρησιμοποιείται η μέθοδος των πολλαπλασιαστών Lagrange έχοντας υπόψη ότι  $b_1 + b_2 = b_{total}$ , όπου το  $b_{total}$  αντιστοιχεί στο συνολικό αριθμό ψηφίων που δεσμεύονται για τη διακριτοποίηση και των δύο παραμέτρων. Η εφαρμογή της μεθόδου δίνει:

$$\begin{aligned} \frac{-2 \ln 2}{12 \times 2^{2b_2}} \overline{\text{var}(D)} &= \frac{(-2 \ln 2) d_r^2}{12 \times 2^{2b_1}} \\ \Rightarrow \frac{1}{2^{2b_2}} \overline{\text{var}(D)} &= \frac{d_r^2}{2^{2b_1}} \\ \Rightarrow b_1 - b_2 &= \log_2 \left( \frac{d_r}{\sqrt{\overline{\text{var}(D)}}} \right) \end{aligned} \quad (4.21)$$

Στην πράξη, υπάρχουν δύο ακόμη ζητήματα τα οποία πρέπει να ξεκαθαριστούν πριν τη μελέτη του τρόπου εφαρμογής των παραπάνω αποτελεσμάτων στη συμπίεση. Πρώτον, για την απλοποίηση της αποκωδικοποίησης, το δυναμικό εύρος μπορεί να προσεγγιστεί από τον αριθμό 256 για εικόνες 8 δυαδικών ψηφίων αναπαράστασης της διαβάθμισης του γκρι.



Δεύτερον, χρειάζεται να καθοριστεί ο τρόπος με τον οποίο στρογγυλοποιούνται οι ακέραιοι ώστε να αντιστοιχούνται στο πλήθος των ψηφίων που επιλέξαμε. Υπενθυμίζοντας τη σχέση (4.19), που περιγράφει το σφάλμα διακριτοποίησης της αντίθεσης, μία πρόταση είναι η στρογγυλοποίηση του  $b_1$  στον πλησιέστερο ακέραιο (η οποία συμβολίζεται με  $\lceil \cdot \rceil$ ). Συνεχίζοντας επομένως, η αντικατάσταση από τη σχέση  $b_1 + b_2 = b_{total}$ , ενός από τους συντελεστές  $b_1, b_2$  σε συνάρτηση με τον άλλο, στη σχέση (4.21), έχει ως αποτέλεσμα την εύρεση του αριθμού των ψηφίων αναπαράστασης κάθε παραμέτρου:

$$b_1 = \left\lceil \frac{b_{total} + \log_2\left(\frac{d_r}{\sqrt{\text{var}(D)}}\right)}{2} \right\rceil \quad (4.22)$$

$$b_2 = b_{total} - b_1$$

Τα αποτελέσματα της εφαρμογής της παραπάνω τεχνικής στην εικόνα της Lena φαίνονται στον πίνακα 4.11. Όσες δεσμεύσεις ψηφίων έχουν γίνει με βάση τη σχέση (4.22) έχουν σημειωθεί με αστερίσκους. Η τελευταία στήλη δίνει το PSNR της ανακατασκευασμένης εικόνας. Τα αποτελέσματα συμφωνούν όπως φαίνεται και στον πίνακα με τις αρχικές υποθέσεις των συγγραφέων εκτός από την περίπτωση των δέκα ψηφίων όπου η ανάθεση δύο ψηφίων για την αντίθεση αποδεικνύεται πιο αποδοτική από αυτήν των τριών που προτείνει η μέθοδος που αναφέραμε. Η διαφορά όμως είναι αμελητέα.

Πίνακας 4.11: Οι δεσμεύσεις των ψηφίων για την 256\*256 εικόνα της Lena. Οι αστερίσκοι δείχνουν τις δεσμεύσεις που έγιναν σύμφωνα με τη σχέση (4.22).

Συνολικός αριθμός ψηφίων	Αριθμός ψηφίων για την αντίθεση	Αριθμός ψηφίων για τη μέση τιμή	PSNR (dB)
8	2*	6*	31.01
8	3	5	29.63
8	4	4	26.72
9	2*	7*	31.37
9	3	6	31.06
9	4	5	29.76
10	2	8	31.46
10	3*	7*	31.45
10	4	6	31.08
10	5	5	29.74

Επιπλέον πειραματικές δοκιμές σε διαφορετικές εικόνες ενισχύουν την πεποίθηση για την ορθότητα της παραπάνω μεθόδου, όπως φαίνεται στον πίνακα 4.12, στον οποίο οι αστερίσκοι συμβολίζουν και πάλι τις δεσμεύσεις που έγιναν με βάση την (4.22). Πρέπει ακόμη να σημειωθεί ότι επειδή στις περισσότερες εικόνες το τετράγωνο του δυναμικού εύρους των μέσων τιμών των τύπου R περιοχών είναι μεγαλύτερο της διασποράς των τύπου D περιοχών, η παραπάνω μέθοδος αναθέτει συνήθως λιγότερα ψηφία για την αντίθεση (οδηγώντας επομένως σε ταχύτερη αποκωδικοποίηση).

Πίνακας 4.12: Οι δεσμεύσεις των ψηφίων για διάφορες εικόνες. Οι αστερίσκοι δείχνουν και πάλι τις δεσμεύσεις που έγιναν σύμφωνα με τη σχέση (4.22).

Εικόνες	Συνολικός αριθμός ψηφίων	Αριθμός ψηφίων για την αντίθεση	Αριθμός ψηφίων για τη μέση τιμή	PSNR (dB)
Κυρία	8	2*	6*	36.32
	8	3	5	33.14
	8	4	4	28.17
Καμεραμάν	8	2*	6*	28.65
	8	3	5	28.07
	8	4	4	26.27
Γυναίκα	8	2*	6*	30.37
	8	3	5	29.16
	8	4	4	28.29
Σχέδιο	8	2*	6*	36.08
	8	3	5	33.03
	8	4	4	28.29

#### 4.4 Ο μετασχηματισμός των τετραγωνικών περιοχών

Όπως είναι γνωστό, η ουσία της συμπίεσης με βάση τη θεωρία των fractal είναι η εύρεση ενός συστολικού μετασχηματισμού  $T$  για την κωδικοποίηση μίας εικόνας  $f$ , το όριο  $g$  του οποίου ( $g=Tx$ ) θα αποτελεί και την πιθανή καλύτερη προσέγγιση της  $f$ . Στο θεώρημα του Barsnley που προαναφέρθηκε, επισημαίνεται ότι η ελαχιστοποίηση της απόστασης μεταξύ της  $f$  και του συστολικού μετασχηματισμού της  $Tf$ , ελαχιστοποιεί και την απόσταση μεταξύ του ορίου  $g$  του μετασχηματισμού και της εικόνας  $f$ . Όταν όμως επιλέγουμε το μετασχηματισμό  $T$  που θα χρησιμοποιήσουμε πρέπει να έχουμε υπ' όψη τους εξής περιορισμούς<sup>7</sup>:

- 1) Ο  $T$  δεν πρέπει να είναι γραμμικός καθώς στην περίπτωση αυτή το όριο του θα ήταν η μηδενική εικόνα (μία εικόνα με μηδενική ένταση όλων των εικονοστοιχείων της).

- 2) Ο  $T$  πρέπει να είναι δομικά και υπολογιστικά απλός, ώστε να παρέχει με απλό τρόπο βελτιστοποίηση του σφάλματος, γρήγορη αποκωδικοποίηση και απλή ανάλυση.
- 3) Το όριό του,  $g$ , πρέπει να είναι εύρωστο ως προς τη διακριτοποίηση των παραμέτρων, δηλαδή η μεταβολή των παραμέτρων από τη θεωρητική στην πλησιέστερη διακριτή τιμή να μην επιφέρει σημαντική αλλαγή στο όριο του μετασχηματισμού.

Στην αρχική του προσέγγιση ο *Jacquin* χρησιμοποίησε για την κωδικοποίηση μιας περιοχής τύπου  $\mathbf{R}$  ένα γραμμικό συνδυασμό ενός κωδικού διανύσματος και ενός διανύσματος σταθερής έντασης. Μπορεί εύκολα να αποδειχθεί ότι στην περίπτωση αυτή ο μετασχηματισμός είναι τύπου affine, δηλ.  $Tf = Af + b$  όπου ο  $A$  είναι ένας  $N \times N$  πίνακας και  $b \in \mathfrak{R}^n$  ( $N$  ο συνολικός αριθμός των εικονοστοιχείων της εικόνας). Στην παρούσα ενότητα περιγράφονται οι πιο γενικοί μετασχηματισμοί που προτείνει η βιβλιογραφία. Για λόγους σαφήνειας και απλότητας θα θεωρείται ότι αφού η εικόνα είναι ένωση όλων των τύπου  $\mathbf{R}$  περιοχών που την αποτελούν, οι επιδράσεις ενός οποιουδήποτε μετασχηματισμού  $T$  πάνω στην εικόνα θα είναι οι ίδιες με αυτές που θα έχει σε μία επιμέρους περιοχή. Για το λόγο αυτό, όλες οι έννοιες που αφορούν το μετασχηματισμό θα ορίζονται με αναφορά την τύπου  $\mathbf{R}$  περιοχή  $R$ .

Έστω λοιπόν τύπου  $\mathbf{R}$  περιοχή  $R = (r_1, \dots, r_n)^T$  και  $D_i = (d_1^i, \dots, d_n^i)^T, i = 1, \dots, N_D$  ένα κωδικό διάνυσμα. Μία πιο γενική δατύπωση του προβλήματος ελαχίστων τετραγώνων είναι η ακόλουθη:

$$\min_{(x_1, \dots, x_m) \in \mathfrak{R}^m} \sum_{k=1}^n \{r_k - t_k(x_1, \dots, x_m)\}^2 \quad (4.23)$$

Όπου

$$t_k(x_1, \dots, x_m) = t(d_k^1, \dots, d_k^{N_D}; x_1, \dots, x_m) \quad (4.24)$$

Επομένως στο μοντέλο που εξετάζουμε, το τύπου R διάνυσμα κωδικοποιείται ως:

$$\begin{pmatrix} t_1(x_1, \dots, x_m) \\ \vdots \\ t_n(x_1, \dots, x_m) \end{pmatrix} \text{ (σε μία απεικόνιση κατά συνιστώσες).}$$

Αν όλες οι συναρτήσεις  $t_k$  έχουν συνεχείς μερικές παραγώγους ως προς  $x_j$  τότε μία αναγκαία συνθήκη για να ισχύει η (4.23) για το  $x = (x_1, \dots, x_m)^T$  είναι:

$$\frac{\partial}{\partial x_j} \sum_{k=1}^n \{r_k - t_k(x_1, \dots, x_m)\}^2 = 0, j = 1, \dots, m \quad (4.25)$$

Ο μετασχηματισμός  $T$  καλείται τύπου affine αν υπάρχει  $n \times m$  πίνακας  $M$  τέτοιος ώστε:

$$\begin{pmatrix} t_1(x_1, \dots, x_m) \\ \vdots \\ t_n(x_1, \dots, x_m) \end{pmatrix} = Mx \quad (4.26)$$

Μια απλή περίπτωση τέτοιου μετασχηματισμού είναι η συμβατική μέθοδος κωδικοποίησης με fractal όπου:

$$M = \begin{pmatrix} d_1 & 1 \\ \vdots & \vdots \\ d_n & 1 \end{pmatrix} \quad (4.27)$$

Μία τροποποιημένη έκδοση της παραπάνω γενικής μεθόδου είναι η κωδικοποίηση που εισήχθη από τον Oien<sup>8</sup>. Αφαιρώντας το μέσο όρο  $\frac{d_1 + \dots + d_n}{n}$  από κάθε συντελεστή  $d_k$  στον παραπάνω πίνακα, μπορεί να αποδειχθεί, ότι για μια προσεγμένη επιλογή του συνόλου των τύπου D περιοχών, η αποκωδικοποίηση θα είναι γρήγορη.

Η χρησιμοποίηση αρκετών σταθερών περιοχών  $B_1, B_2, \dots, B_p$ , έχει προταθεί από πολλούς ερευνητές. Αντιστοιχεί στον πίνακα:

$$M = \begin{pmatrix} d_1 & b_1^1 & \cdots & b_p^1 \\ \vdots & \vdots & \vdots & \vdots \\ d_n & b_1^n & \cdots & b_p^n \end{pmatrix} \quad (4.28)$$

Άλλες προσπάθειες αφορούσαν τη χρήση πολλών κωδικών διανυσμάτων  $D_{i_1}, \dots, D_{i_r}$  για την ίδια περιοχή τύπου R, οπότε ο πίνακας  $M$  παίρνει τη μορφή που ακολουθεί:

$$M = \begin{pmatrix} d_1^{i_1} & \cdots & d_1^{i_r} & b_1^1 & \cdots & b_p^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_n^{i_1} & \cdots & d_n^{i_r} & b_1^n & \cdots & b_p^n \end{pmatrix} \quad (4.29)$$

Μία άλλη δυνατότητα είναι και η χρησιμοποίηση των τετραγώνων των συνιστωσών:

$$M = \begin{pmatrix} d_1^2 & d_1 & 1 \\ \vdots & \vdots & \vdots \\ d_n^2 & d_n & 1 \end{pmatrix} \quad (4.30)$$

Όλες αυτές οι μεταβλητές στοχεύουν στην απόκτηση μίας καλύτερης προσέγγισης για ένα δεδομένο διάνυσμα τύπου R. Δυστυχώς όμως, οι μέθοδοι που αναφέρθηκαν εδώ υποφέρουν από τον αυξημένο υπολογιστικό χρόνο. Επιπλέον, ο κώδικας για μία τύπου R περιοχή είναι πιο πολύπλοκος σε σχέση με αυτόν που χρησιμοποιεί ο συμβατικός αλγόριθμος. Για την περίπτωση π.χ. της χρήσης αρκετών κωδικών διανυσμάτων υπάρχει και η ανάγκη της διασφάλισης της ιδιότητας της συστολής του μετασχηματισμού. Παρ' όλα αυτά, αν ισχύει η συνθήκη γραμμικότητας (4.26), εξάγεται το ακόλουθο αποτέλεσμα:

### Θεώρημα 5:

Το πρόβλημα ελαχιστοποίησης (4.23) έχει τουλάχιστον μία λύση  $x_0$ . Επιπλέον, αν οι στήλες του πίνακα  $M$  είναι γραμμικά ανεξάρτητες τότε η  $x_0$  είναι μοναδική και δίνεται από τη σχέση:

$$x_0 = (M^T M)^{-1} M^T R \quad (4.31)$$

Η μη γραμμική περίπτωση δεν έχει μελετηθεί ακόμα επαρκώς. Ενδεικτικά αναφέρεται η προσπάθεια των Lin και Βενετσανόπουλου που χρησιμοποίησαν μια μέθοδο όπου  $m=4$  και ο μετασχηματισμός έχει τη μορφή:

$$t_k(x_1, \dots, x_4) = \pm \frac{1}{1 + e^{x_1 k_1 + x_2 k_2 + x_3}} + x_4 \quad (4.32)$$

Τα  $(k_1, k_2)$  αντιπροσωπεύουν τη δισδιάστατη αναπαράσταση του  $k$  (καθώς η τετραγωνική περιοχή μετατρέπεται σε διάνυσμα). Με ένα σταθερό *bit rate*, οι συγγραφείς αναφέρουν ότι αποκτούμε μία οπτικά καλύτερη και πιο γρήγορη αποκωδικοποίηση. Η δυσκολία όμως που αντιμετωπίζουν είναι η επίλυση του προβλήματος των ελαχίστων τετραγώνων.

Τέλος αναφέρουμε ότι ο μετασχηματισμός τύπου affine μπορεί υπό ορισμένες προϋποθέσεις να πάρει και τη μορφή πολωνύμου. Ο D. Vitulano<sup>9</sup> για παράδειγμα, χρησιμοποιεί πολώνυμα δεύτερης τάξης και ειδικότερα τη συνάρτηση:

$$p(z) = az^2 + c, a > 0 \quad (4.33)$$

ως εκφραστή της συστολικής διαδικασίας. Στη συνέχεια τίθενται τα όρια μέσα στα οποία πρέπει να κινούνται οι παράμετροι της συνάρτησης ώστε να μη παραβιάζει η συνάρτηση τη συνθήκη της συστολικότητας.

Έστω  $L$  ο αριθμός διαβαθμίσεων του γκρι, δηλαδή το πεδίο τιμών της έντασης είναι το  $[0, L]$ . Η παραβολή  $p(z)$  επιβάλλεται να συγκλίνει στο  $[0, L]$ , δηλαδή:

1. Το πεδίο τιμών του  $p([0, L])$  πρέπει να περιέχεται στο πεδίο ορισμού του  $[0, L]$ .

2. Το πεδίο ορισμού  $[0, L]$  πρέπει να περιέχεται στο διάστημα σύγκλισης του μετασχηματισμού.

Η 1<sup>η</sup> συνθήκη εξασφαλίζει ότι η ακολουθία που δημιουργεί η επαναληπτική εφαρμογή του μετασχηματισμού είναι καλώς ορισμένη για κάθε  $x_0 \in [0, L]$ . Αυτό επιβάλλει τις ακόλουθες συνθήκες:  $p(0) \in [0, L]$ ,  $p(L) \in [0, L]$  και αν  $\hat{x} \in [0, L]$ , τότε και  $\hat{y} \in [0, L]$ . Κατά συνέπεια οι συνθήκες για τις παραμέτρους της παραβολής είναι:

$$0 \leq c \leq L \quad (4.34)$$

και

$$0 < aL^2 + c \leq L \quad (4.35)$$

Ο μετασχηματισμός αυτός διαθέτει με τη σειρά του μόνο δύο παραμέτρους προς αποθήκευση, όπως και ο απλός και η υπολογιστική του πολυπλοκότητα είναι συγκρίσιμη με αυτή ενός συνηθισμένου μετασχηματισμού τύπου affine.



## Βιβλιογραφία

1. Ning Lu, *Fractal Imaging*, Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto.
2. Chong Sze Tong and Minghong Pi, *Fast Fractal Image Encoding Based on Adaptive Search*, in: IEEE Transactions on Image Processing, Vol.10, No.9, Sept. 2001.
3. B. Bani-Eqbal, *Enhancing the speed of fractal image compression*, Opt. Eng., vol. 34, no. 6, pp. 1705-1710, June 1995
4. G. E. Oien and S. Lepsoy, *A class of fractal image coders with fast decoder convergence*, in: *Fractal Image Compression: Theory and Applications* . Berlin, Germany: Springer-Verlag, 1995, pp. 153-174.
5. S.P. Lloyd: *Least square quantization in PCM*. IEEE Trans. Inf. Theory, vol. IT-28, pp. 129-137, March 1982.
6. J. Max: *Quantizing for minimum distortion*. IRE Trans. Inf. Theory, vol. IT-6, pp. 7-12, March 1960.
7. Oien, G. E, *Parameter Quantization in Fractal Image Coding*, in: *Proc. ICIP-94 IEEE International Conference on Image Processing*, Austin, Texas, Nov.1994
8. Oien, G. E,  *$L_2$  Optimal attractor Image Coding with fast decoder convergence*, PhD Thesis, *The Norwegian Institute Of Technology*, Trondheim, Norway, April 1993.
9. Vitulano, D., *Fractal image Coding schemes using non-linear grey scale functions*, Signal Processing 81 (2001) 1095-1099

---

## ΚΕΦΑΛΑΙΟ 5

---

### Τεχνικές μείωσης της πολυπλοκότητας

Η συμπίεση της εικόνας με τη μέθοδο των fractal ενώ επιτρέπει τη γρήγορη αποκωδικοποίηση, υποφέρει από τις μεγάλες απαιτήσεις σε χρόνο για την κωδικοποίηση. Το κομμάτι της κωδικοποίησης που είναι το περισσότερο χρονοβόρο, είναι η αναζήτηση της κατάλληλης τύπου  $D$  περιοχής για κάθε τύπου  $R$  περιοχή. Ο αριθμός των πιθανών αυτών τύπου  $D$  περιοχών που μπορούν θεωρητικά να είναι υποψήφιος για την προσέγγιση είναι απαγορευτικά μεγάλος. Για παράδειγμα, ο αριθμός των τετραγωνικών υποπεριοχών μιας εικόνας μεγέθους  $n*n$  *εικονοστοιχείων* είναι της τάξης του  $n^3$ . Επομένως πρέπει να τεθούν κάποιοι περιορισμοί ως προς τον καθορισμό των επιτρεπόμενων περιοχών τύπου  $D$  για σύγκριση. Μία απλή εφαρμογή του παραπάνω γενικού κανόνα είναι η θεώρηση ως τύπου  $D$  περιοχών για παράδειγμα των τετραγώνων εκείνων που θα είχαν περιορισμένο αριθμό μεγεθών και θέσεων. Αυτή η θεώρηση καθορίζει και το σύνολο των τύπου  $D$  περιοχών (domain pool). Για κάθε τύπου  $R$  περιοχή, πρέπει να εξεταστεί κάθε στοιχείο του συνόλου αυτού. Αν ο αριθμός των στοιχείων είναι  $N_D$ , τότε ο χρόνος που περνά για κάθε αναζήτηση είναι γραμμικός ως προς  $N_D$  ( $O(N_D)$ ). Για το λόγο αυτό έχουν αναπτυχθεί διάφορες μέθοδοι που ελαττώνουν την πολυπλοκότητα της κωδικοποίησης και οι οποίες θα εξεταστούν στο κεφάλαιο αυτό. Για τη διευκόλυνση της μελέτης, δίνονται αρχικά κάποιοι ορισμοί και επεξηγούνται έννοιες σχετικές με τις τεχνικές fractal συμπίεσης που ακολουθούν.

## 5.1 Μία μέθοδος που βασίζεται στις προβολές για τον υπολογισμό του σφάλματος ελαχίστων τετραγώνων

Για την ανάπτυξη της θεωρίας αυτής, γίνεται η υπόθεση ότι η εικόνα τεμαχίζεται σε μη επικαλυπτόμενα τετράγωνα μεγέθους  $N \times N$ . Αυτό δεν αποτελεί περιορισμό καθώς στη συνέχεια θα γίνει φανερό πώς οι αρχές που περιγράφονται παρακάτω επεκτείνονται σε πιο γενικές περιπτώσεις διαίρεσης.

Θεωρείται ότι κάθε τύπου  $R$  περιοχή  $R$  είναι διάνυσμα που ανήκει στο διανυσματικό χώρο  $R^n$  ( $R \in \mathfrak{R}^n$ ), όπου  $n=N \times N$ . Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο η μετατροπή μιας τετράγωνης υποπεριοχής εικόνας πλευράς  $N$  σε διάνυσμα μεγέθους  $N^2$  μπορεί να επιτευχθεί αν για παράδειγμα προσπελάσουμε την υποπεριοχή γραμμή προς γραμμή. Η εργασία με διανύσματα αντί για πίνακες δύο διαστάσεων απλοποιεί τους ορισμούς που θα αναφερθούν χωρίς βλάβη της γενικότητας.

Ανατρέχοντας σε προηγούμενα κεφάλαια, υπενθυμίζεται ότι το σύνολο των τύπου  $D$  περιοχών είναι μια συλλογή από τετράγωνα τυπικά μεγαλύτερα από τις τύπου  $R$ . Τα στοιχεία που περιέχει το σύνολο αυτό ομιταπλασιάζονται αν θεωρήσουμε και τους ισομετρικούς μετασχηματισμούς κάθε τετραγώνου. Τέλος, με υπολογισμό του μέσου όρου των γειτονικών εικονοστοιχείων, το μέγεθος αυτών των τετραγώνων ελαττώνεται σε μέγεθος ίσο προς αυτό των τύπου  $R$ . Ο τετραγωνικές περιοχές που προκύπτουν ονομάζονται τετράγωνα του κωδικού βιβλίου.

Κατά την κωδικοποίηση, πραγματοποιείται αναζήτηση, για κάθε τύπου  $R$  τετράγωνο, μέσα στο σύνολο των τετραγώνων του κωδικού βιβλίου. Το διάνυσμα που παριστάνει ένα τέτοιο τετράγωνο του συνόλου καλείται  $D$ . Εκτός από τα διανύσματα του κωδικού βιβλίου επιλέγεται και ένα μικρό σύνολο από  $p < n$  τετράγωνα τα οποία είναι σταθερά και ανεξάρτητα από την εικόνα. Αναπαριστώνται με τα διανύσματα  $B_1, B_2, \dots, B_p \in \mathfrak{R}^n$  και επιλέγονται ώστε να σχηματίζουν μία ορθοκανονική βάση σε ένα υποχώρο του  $R^n$  διάστασης  $p$ .

Αυτές οι τετραγωνικές περιοχές είναι γνωστές ως τετράγωνα σταθερής βάσης. Η αναζήτηση της καλύτερης δυνατής κωδικοποίησης ισοδυναμεί με την επίλυση του προβλήματος των ελαχίστων τετραγώνων:

$$E(D, R) = \min_{a, b_1, \dots, b_p \in \mathfrak{R}} \left\| R - (aD + \sum_{k=1}^p b_k B_k) \right\| = \min_{x \in \mathfrak{R}^{p+1}} \|R - Ax\| \quad (5.1)$$

όπου ο  $A$  έχει διαστάσεις  $n^*(p+1)$  και στήλες τα  $D, B_1, \dots, B_p$  και το  $x = (a, b_1, b_2, \dots, b_p) \in \mathfrak{R}^{p+1}$ , είναι το διάνυσμα των συντελεστών. Εδώ έχει αντικατασταθεί η αντίθεση  $s$  και η φωτεινότητα  $o$  που αντιστοιχούσαν στην περίπτωση της ύπαρξης ενός μόνο τετραγώνου βάσης (του μοναδιαίου διανύσματος) με τα  $a$  και  $b_i$ . Το παραπάνω πρόβλημα πρέπει να επιλυθεί για όλα τα κωδικά διανύσματα  $D$  και επιλέγεται αυτό που δίνει το μικρότερο σφάλμα

$$\left\| R - (aD + \sum_{k=1}^p b_k B_k) \right\|$$

με τον περιορισμό βέβαια ότι η αντίθεση είναι απολύτως

μικρότερη του 1, ώστε να είναι δυνατή η σύγκλιση κατά την αποκωδικοποίηση. Αυτή η συνθήκη για το  $a$  μπορεί να αρθεί αν χρησιμοποιήσουμε την ορθοκανονική αναπαράσταση του *Oien* (βιβλιογραφία 4<sup>ο</sup> κεφαλαίου). Ένα βασικό πόρισμα της γραμμικής άλγεβρας υπογραμμίζει ότι αν το κωδικό διάνυσμα  $D$  δεν ανήκει στη γραμμική επέκταση της σταθερής βάσης  $B_1, \dots, B_p$ , τότε το πρόβλημα ελαχιστοποίησης έχει μοναδική λύση τη

$$\bar{x} = (A^T A)^{-1} A^T R \quad (5.2)$$

όπου ο πίνακας  $A^+ = (A^T A)^{-1} A^T$  καλείται επίσης ψευδοανάστροφος του  $A$ . Επομένως η τύπου  $R$  περιοχή  $R$  προσεγγίζεται από την περιοχή  $AA^+R$  όπου το  $AA^+$  είναι ο πίνακας ορθοκανονικών προβολών πάνω στη τύπου  $R$  περιοχή. Έστω  $P$  ο τελεστής που προβάλλει το χώρο  $R^n$  στον υποχώρο  $B$  που σχηματίζεται από τα διανύσματα σταθερής βάσης που προαναφέρθηκαν. Θα ισχύει:

$$PR = \sum_{k=1}^p b_k B_k \quad (5.3)$$

Τότε η τύπου R περιοχή R θα έχει μια μοναδική ορθοκανονική γραφή (*decomposition*)  $R=OR+RP$ , όπου ο τελεστής  $O=I-P$  προβάλλει τον  $R^n$  στο ορθοκανονικό συμπλήρωμα του  $B, B^\perp$ . Για κάθε  $Z = (z_1, z_2, \dots, z_n) \in \mathfrak{R}^n \setminus B$ , ορίζεται ο τελεστής:

$$\phi(z) = \frac{OZ}{\|OZ\|} \quad (5.4)$$

Για ένα δεδομένο τύπου D κωδικό διάνυσμα  $D \notin B$  η ζητούμενη περιοχή  $AA^+R$  είναι:

$$AA^+R = \langle R, \phi(D) \rangle \phi(D) + \sum_{k=1}^p \langle R, B_k \rangle B_k \quad (5.5)$$

Για τον υπολογισμό του σφάλματος των ελαχίστων τετραγώνων χρησιμοποιείται η συνθήκη ορθοκανονικότητας των  $\phi(R), B_1, \dots, B_p$  με την τύπου R περιοχή R να εκφράζεται ως:

$$R = \langle R, \phi(R) \rangle \phi(R) + \sum_{k=1}^p \langle R, B_k \rangle B_k \quad (5.6)$$

Εισάγοντας το αποτέλεσμα για το R στο πρώτο μέλος της (5.5) και μετά από υπολογισμούς προκύπτει ότι:

$$\langle R, \phi(D) \rangle \phi(D) = \langle R, \phi(R) \rangle \langle \phi(D), \phi(R) \rangle \phi(D) \quad (5.7)$$

Επομένως η ζητούμενη τετραγωνική περιοχή μπορεί να ξαναγραφεί ως:

$$AA^+R = \langle R, \phi(R) \rangle \langle \phi(D), \phi(R) \rangle \phi(D) + \sum_{k=1}^p \langle R, B_k \rangle B_k \quad (5.8)$$

Το σφάλμα συνεπώς γράφεται:

$$E(D, R) = \|R - AA^+R\| = \sqrt{\langle R - AA^+R, R - AA^+R \rangle} \quad (5.9)$$

Μετά από υπολογισμούς βρίσκεται ότι:

$$E(D, R) = \langle R, \phi(R) \rangle \sqrt{1 - \langle \phi(D), \phi(R) \rangle^2} \quad (5.10)$$

Κατά συνέπεια, η ελαχιστοποίηση του σφάλματος  $E(D,R)$ , για όλα τα κωδικά διανύσματα  $D$ , μπορεί να επιτευχθεί με τη χρήση του κριτηρίου της γωνίας μεταξύ των  $\phi(R)$ ,  $\phi(D)$ . Το σφάλμα είναι ελάχιστο όταν το εσωτερικό γινόμενο  $\langle \phi(D), \phi(R) \rangle$  των δύο αυτών διανυσμάτων μοναδιαίου μέτρου είναι μέγιστο και καθώς

$$\langle \phi(D), \phi(R) \rangle^2 = \cos^2 \angle(\phi(D), \phi(R)) \quad (5.11)$$

μεγιστοποίηση του γινομένου συνεπάγεται άμεσα ελαχιστοποίηση της γωνίας  $\angle(\phi(D), \phi(R))$  ή ισοδύναμα  $\angle(OD, OR)$ .

## 5.2 Χαρακτηριστικά διανύσματα

Σε αυτήν την προσέγγιση με τη βοήθεια των χαρακτηριστικών διανυσμάτων η οποία προτάθηκε από τον *Saupe*<sup>1</sup>, χρησιμοποιείται ένα μικρό σύνολο από  $d$  «κλειδιά» (*keys*) (πραγματικοί αριθμοί) για κάθε τύπου  $D$  περιοχή τα οποία συνιστούν ένα χαρακτηριστικό διάνυσμα  $d$  διαστάσεων. Αυτά τα κλειδιά κατασκευάζονται προσεκτικά ώστε η αναζήτηση στο σύνολο των τύπου  $D$  περιοχών να μπορεί να περιοριστεί στους πιο κοντινούς γείτονες του σημείου αναζήτησης δηλαδή του χαρακτηριστικού διανύσματος της εκάστοτε τύπου  $R$  περιοχής. Επομένως η σειριακή αναζήτηση στο σύνολο των τύπου  $D$  περιοχών αντικαθίσταται από την αναζήτηση του πιο κοντινού γείτονα πολλαπλών διαστάσεων που μπορεί να γίνει σε λογαριθμικό χρόνο.

Θεωρούμε ένα σύνολο από  $N_D$  κωδικά διανύσματα  $D_1, \dots, D_{N_D} \in \mathfrak{R}^n$  και μία τύπου  $R$  περιοχή  $R \in \mathfrak{R}^n$ . Έστω  $E(D_i, R)$  το πιο μικρό πιθανό σφάλμα της προσέγγισης της τύπου  $R$  περιοχής από το κωδικό διάνυσμα  $D_i$  με τη χρήση ενός μετασχηματισμού τύπου affine, δηλαδή

$$E(D_i, R) = \min_{a, b_1, \dots, b_p \in \mathfrak{R}} \left\| R - (aD_i + \sum_{k=1}^p b_k B_k) \right\| \quad (5.12)$$

Το ακόλουθο θεώρημα θέτει τα θεμέλια της μεθόδου των χαρακτηριστικών διανυσμάτων:

**Θεώρημα 6:**

Έστω  $n \geq 2$  και  $X = \mathbb{R}^n \setminus B$ . Ορίζουμε τη συνάρτηση  $\Delta : X \times X \rightarrow [0, \sqrt{2}]$

με τη σχέση  $\Delta(D, R) = \min(\|\phi(R) + \phi(D)\|, \|\phi(R) - \phi(D)\|)$ . Για κάθε

$D_i, R \in X$ , το σφάλμα  $E(D_i, R)$  δίνεται από τη σχέση

$$E(D_i, R) = \langle R, \phi(R) \rangle g(\Delta(D_i, R)) \quad (5.13)$$

, όπου

$$g(\Delta) = \Delta \sqrt{1 - \frac{\Delta^2}{4}} \quad (5.14)$$

**Απόδειξη:** Η προσέγγιση μιας τύπου R περιοχής με τη μέθοδο των ελαχίστων τετραγώνων δίνεται από τη σχέση (5.8) στην ενότητα 5.1:

$$AA^+R = \langle R, \phi(D) \rangle \phi(D) + \sum_{k=1}^p \langle R, B_k \rangle B_k. \quad (5.15)$$

Εξαιτίας της ορθογωνιότητας μπορούμε να εκφράσουμε τη τύπου R περιοχή ως:

$$R = \langle R, \phi(R) \rangle \phi(R) + \sum_{k=1}^p \langle R, B_k \rangle B_k. \quad (5.16)$$

Χρησιμοποιώντας αυτή τη σχέση στην έκφραση (5.15) εξάγεται το αποτέλεσμα:

$$AA^+R = \langle R, \phi(R) \rangle \langle \phi(D), \phi(R) \rangle \phi(D) + \sum_{k=1}^p \langle R, B_k \rangle B_k. \quad (5.17)$$

, που δίνει την προσέγγιση του R. Το τετράγωνο της διαφοράς των (5.16) και (5.17) αποτελεί το σφάλμα των ελαχίστων τετραγώνων και υπολογίζεται ως:

$$E(D, R) = \langle R, \phi(R) \rangle \sqrt{1 - \langle \phi(D), \phi(R) \rangle^2} \quad (5.18)$$

Αφού

$$\|\phi(R) \pm \phi(D)\| = \sqrt{2(1 \pm \langle \phi(D), \phi(R) \rangle)} \quad (5.19)$$

έχουμε:

$$\Delta(D, R) = \sqrt{2(1 - |\langle \phi(D), \phi(R) \rangle|)} \quad (5.20)$$

Λύνοντας ως προς  $|\langle \phi(D), \phi(R) \rangle|$  και εισάγοντας το τετράγωνο του αποτελέσματος στη σχέση που δίνει το  $E(D, R)$  ολοκληρώνεται η απόδειξη  $\square$ .

Το θεώρημα δηλώνει ότι το σφάλμα ελαχίστων τετραγώνων  $E(D, R)$  είναι ανάλογο της απλής συνάρτησης  $g$  της Ευκλείδειας απόστασης  $\Delta$  μεταξύ των προβολών  $\phi(D_i)$  και  $\phi(R)$  ή  $-\phi(D_i)$  και  $\phi(R)$ . Αφού η  $g(\Delta)$  είναι γνησίως αύξουσα για  $0 \leq \Delta \leq \sqrt{2}$  συμπεραίνουμε ότι η ελαχιστοποίηση των σφαλμάτων  $E(D, R)$  για  $i = 1, \dots, N_D$  είναι ισοδύναμη με την ελαχιστοποίηση των αποστάσεων  $\Delta(D, R)$ . Επομένως, ο υπολογισμός και η ελαχιστοποίηση των  $N_D$  σφαλμάτων ελαχίστων τετραγώνων, μπορεί να αντικατασταθεί με την αναζήτηση του πλησιέστερου γείτονα του  $\phi(R) \in R^n$  στο σύνολο των  $2N_D$  διανυσμάτων  $\pm \phi(D_i) \in R^n$ . Το πρόβλημα της εύρεσης του πιο κοντινού γείτονα, στο οποίο έγινε αναφορά στο πρώτο κεφάλαιο, έχει μελετηθεί εκτενώς στην επιστήμη των υπολογιστών. Για παράδειγμα οι *Friedman, Bentley και Finkel*<sup>2</sup> παρουσιάζουν μία μέθοδο που χρησιμοποιεί  $k$ -διάστατα ( $k-d$ ) δένδρα και τρέχει σε λογαριθμικό χρόνο, μαζί με τον ψευδοκώδικά της. Μετά από ένα προπαρασκευαστικό βήμα που δημιουργεί το απαιτούμενο  $k-d$  δένδρο, το οποίο χρειάζεται  $O(N \log N)$  βήματα, η αναζήτηση του πλησιέστερου γείτονα για ένα συγκεκριμένο τύπου  $R$  χαρακτηριστικό διάνυσμα ολοκληρώνεται στο αναμενόμενο λογαριθμικό χρόνο,  $O(\log N)$ . Παρ' όλα αυτά, καθώς η διάσταση  $d$  αυξάνεται, η επίδοση μπορεί να μην είναι τόσο υψηλή. Μια μέθοδος πιο αποδοτική σε σχέση με την αύξηση της διάστασης που παρουσιάζεται από τον *Arpa et al.*<sup>3</sup>, δίνει την προσέγγιση ενός υποτιθέμενου πλησιέστερου γείτονα. Για μικρά σχετικά σύνολα τύπου  $D$  περιοχών, άλλες μέθοδοι που δε βασίζονται σε τέτοιου είδους δένδρα που χρησιμοποιούνται για τη διαίρεση του χώρου, ίσως και να αποδίδουν καλύτερα. Για παράδειγμα η τροποποιημένη αναζήτηση του



μέσου πιο κοντινού γείτονα (*Enhanced Nearest Neighbor Search*)<sup>4</sup> φαίνεται να είναι από τις καλύτερες. Εκτός από τα σημαντικά πλεονεκτήματα που απορρέουν από το παραπάνω θεώρημα η πρακτική εκμετάλλευση των οποίων θα συζητηθεί στη συνέχεια, πρέπει να τονιστεί ότι το αποτέλεσμα του Θεωρήματος 6 μπορεί να χρησιμοποιηθεί και για την αναγνώριση όλων εκείνων των κωδικών διανυσμάτων  $D_i$  που ικανοποιούν ένα συγκεκριμένο κριτήριο ανοχής, έστω  $E(R, D_i) \leq \delta$ . Με άλλα λόγια, λύνοντας την εξίσωση ως προς  $\Delta$  στην έκφραση του σφάλματος  $E(R, D)$ , παίρνουμε από το θεώρημα την αναγκαία και ικανή συνθήκη ώστε ένα κωδικό διάνυσμα να ικανοποιεί το κριτήριο ανοχής.

**Πόρισμα 1 (Μία αναγκαία και ικανή συνθήκη):**

Έστω  $\delta \geq 0$  και  $n \geq 2$ . Έστω ότι τα  $R$  και  $D$  ανήκουν στον  $\mathbf{R}^n \setminus B$  με  $\langle R, \phi(R) \rangle \geq \delta$ .

Τότε

$$E(D, R) = \min_{a, b_1, \dots, b_p \in \mathbf{R}} \left\| R - \left( aD + \sum_{k=1}^p b_k B_k \right) \right\| \leq \delta \quad (5.21)$$

αν και μόνο αν

$$\Delta(D, R) \leq \sqrt{2 - 2 \sqrt{1 - \frac{\delta^2}{\langle R, \phi(R) \rangle^2}}} \quad (5.22)$$

, όπου η συνάρτηση  $\Delta(D, R)$  έχει οριστεί από το Θεώρημα 6.

**Απόδειξη:** Από το Θεώρημα 6,  $E(D, R) = \langle R, \phi(R) \rangle g(\Delta(D, R))$  με

$g(\Delta) = \Delta \sqrt{1 - \frac{\Delta^2}{4}}$ . Επομένως, για  $0 \leq \Delta \leq \sqrt{2}$  έχουμε  $E(D, R) \leq \delta$  αν και

μόνο αν  $\frac{\Delta^4 - 4\Delta^2 + 4\delta^2}{\langle R, \phi(R) \rangle^2} \geq 0$ . Απαιτώντας η διακρίνουσα του παραπάνω

τριωνύμου ως προς  $\Delta^2$  να είναι αρνητική ώστε να ισχύει η τελευταία σχέση, εύκολα καταλήγουμε στο αποτέλεσμα του πορίσματος  $\square$ .

Όπως φάνηκε, η συνθήκη  $\langle R, \phi(R) \rangle \geq \delta$  δεν θέτει κανένα περιορισμό. Στην περίπτωση μάλιστα που  $\langle R, \phi(R) \rangle < \delta$  έχει ήδη εξασφαλιστεί ότι το σφάλμα θα είναι μικρότερο της ανοχής, καθώς:

$$E(D, R) = \langle R, \phi(R) \rangle \sqrt{1 - \langle \phi(D), \phi(R) \rangle^2} \leq \langle R, \phi(R) \rangle < \delta \quad (5.23)$$

για οποιοδήποτε κωδικό διάνυσμα  $D$ . Επομένως, η περιοχή  $R$  μπορεί να κωδικοποιηθεί χρησιμοποιώντας μόνο τα σταθερά διανύσματα  $B$ , δηλαδή με τη σχέση  $\sum_{k=1}^p b_k B_k$ .

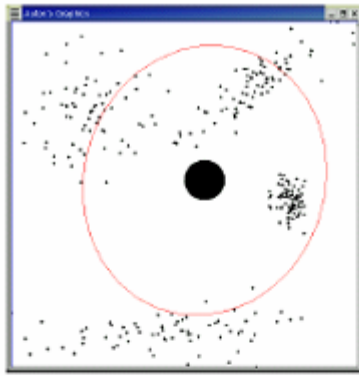
Στη συνέχεια, παρατίθενται κάποια σχόλια πάνω σε γενικεύσεις και εφαρμογές της θεωρίας που μόλις παρουσιάσαμε. Στην πραγματικότητα, υπάρχει ένα όριο ως προς την αποθήκευση των χαρακτηριστικών διανυσμάτων των τύπου  $R$  και τύπου  $D$  περιοχών. Για παράδειγμα, τα διανύσματα κλειδιά για τύπου  $R$  περιοχές μεγέθους  $8*8$  απαιτούν  $64$  πραγματικούς αριθμούς το κάθε ένα. Αν ένα σύνολο τύπου  $D$  περιοχών περιέχει μόνο  $32k$  τύπου  $D$  περιοχές, τότε έχουν ήδη καταληφθεί  $8M$  μνήμης, ενώ η διατήρηση ενός επιπέδου στην ποιότητα της συμπίεσης επιβάλλει συνήθως την εργασία με σύνολα  $100000$  τύπου  $D$  περιοχών και άνω. Για να αρθεί αυτή τη δυσκολία, ακολουθεί ο εξής συμβιβασμός: Υποδειγματοληπτώνται όλες οι τύπου  $D$  και τύπου  $R$  περιοχές σε μία προαποφασισμένη διάσταση π.χ.  $d=4*4=16$ . Επιπλέον κάθε ένα από τα  $d$  στοιχεία του χαρακτηριστικού διανύσματος διακριτοποιείται (π.χ.  $8$  δυαδικά ψηφία\στοιχείο είναι αρκετά). Οι παραπάνω ενέργειες επιτρέπουν την επεξεργασία ενός αυξημένου αριθμού χαρακτηριστικών διανυσμάτων τύπου  $D$  και τύπου  $R$  περιοχών, με τη διαφορά ότι η εφαρμογή του θεωρήματος δεν είναι ακριβής πλέον, αλλά προσεγγιστική. Αυτό μολαταύτα δεν αποτελεί σημαντικό μειονέκτημα όπως θα διαπιστώσουμε αμέσως τώρα, και όπως έχει δείχθει από διάφορα πειράματα.

Η παραπάνω προσεγγιστική διαδικασία, η οποία συμπεριλαμβάνει τον υπολογισμό του μέσου όρου των γειτονικών εικονοστοιχείων με σκοπό τη μείωση της διάστασης των τύπου R και τύπου D περιοχών ( που συχνά η τιμή της είναι 64 και άνω) σε μία τιμή που διευκολύνει τους υπολογισμούς (16 στην περίπτωση αυτή), μπορεί να βελτιωθεί ενσωματώνοντας με καλύτερο τρόπο τις πληροφορίες που περιέχει η εκάστοτε υποεικόνα στα  $d$  στοιχεία. Βασισμένος στην αναφορά του *Saupe*, ο *Barthel et al.* πρότεινε και εφάρμοσε ένα εναλλακτικό τρόπο μείωσης της διάστασης. Χρησιμοποιεί το δισδιάστατο Διακριτό Μετασχηματισμό Συνημιτόνου (**Discrete Cosine Transform, DCT**) των  $\pm\varphi(D)$ . Τα αποτελέσματα του Θεωρήματος 6 μεταφέρονται στο πεδίο της συχνότητας όπου οι ιδιότητες της απόστασης του μοναδιαίου μετασχηματισμού συνεχίζουν να ισχύουν και συνεπώς οι πιο κοντινοί μετασχηματισμένοι γείτονες θα δίνουν και τα πιο μικρά σφάλματα. Στην πράξη, υπολογίζεται αρχικά ο DCT για όλα τις τύπου D και τύπου R περιοχές. Στη συνέχεια, από τους συντελεστές που προκύπτουν αφαιρείται ο σταθερός (DC) όρος και κανονικοποιούνται οι  $d$  συντελεστές του κάθε διανύσματος. Περισσότερα για το Διακριτό Μετασχηματισμό Συνημιτόνου θα αναφέρουμε σε επόμενη ενότητα.

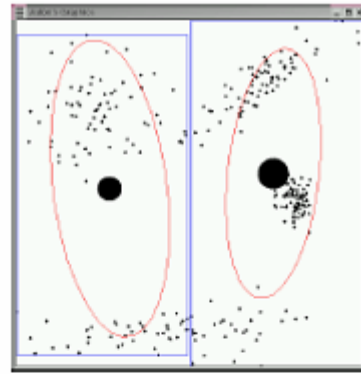
Εξαιτίας της μεθόδου μείωσης της διάστασης και διακριτοποίησης των συντελεστών  $a, b_1, \dots, b_p$  που εφαρμόζεται με σκοπό την παράλληλη μείωση του χώρου αποθήκευσης και τη διευκόλυνση των υπολογισμών, ενδέχεται ο πλησιέστερος γείτονας στο χώρο των χαρακτηριστικών διανυσμάτων να μην είναι το κωδικό διάνυσμα με το μικρότερο σφάλμα. Επιπλέον η τιμή της αντίθεσης  $a$  μπορεί να είναι πολύ μεγάλη για να γίνει αποδεκτή. Το πρόβλημα αυτό αντιμετωπίζεται με την αναζήτηση όχι μόνο του πλησιέστερου γείτονα αλλά και των 5 ή 10 πλησιέστερων γειτόνων (εργασία που μπορεί και πάλι να εκτελεστεί σε λογαριθμικό χρόνο αν διατηρηθεί μία σειρά προτεραιότητας). Από αυτό το σύνολο των πλησιέστερων γειτόνων, οι μη αποδεκτές τύπου D περιοχές απορρίπτονται και οι εναπομείνουσες συγκρίνονται με βάση τη μέθοδο

των ελαχίστων τετραγώνων. Αυτό το γεγονός εξηγεί και την αδυναμία της μεθόδου να δώσει ακριβή, παρά μόνο προσεγγιστικά αποτελέσματα. Μολονότι η τύπου  $D$  περιοχή που αντιστοιχίζεται σε μία συγκεκριμένη τύπου  $R$  ίσως να μην είναι η βέλτιστη, υπάρχουν συνήθως εναλλακτικές περιοχές από την ομάδα των πλησιέστερων γειτόνων που βρίσκονται «κοντά» στη βέλτιστη και μπορούν να αντικαταστήσουν την αρχική επιλογή που έχει γίνει.

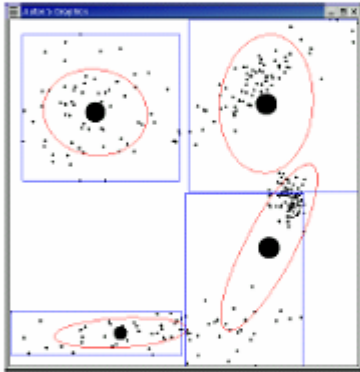
Εδώ πρέπει να κάνουμε και δύο σχόλια όσον αφορά τις απαιτήσεις σε μνήμη του  $k$ -διάστατου δένδρου, το οποίο δεν είναι παρά μία τροποποίηση ενός δυαδικού δένδρου αναζήτησης. Τα στοιχεία του έχουν διάσταση  $k$  και αποτελείται από  $d$  επίπεδα σε κάθε επίπεδο του οποίου λαμβάνεται μια διακλαδισμένη απόφαση με βάση κάποιο διάνυσμα «κλειδί» ώστε η σύγκριση των στοιχείων του χώρου με το τελευταίο να τα ταξινομεί στα υποδέντρα του συγκεκριμένου επιπέδου. Η δομή αυτή επομένως χρησιμοποιείται κατά κύριο λόγο ως μέσο διαμέρισης διανυσματικών χώρων, όπως φαίνεται στο σχήμα της επόμενης σελίδας.



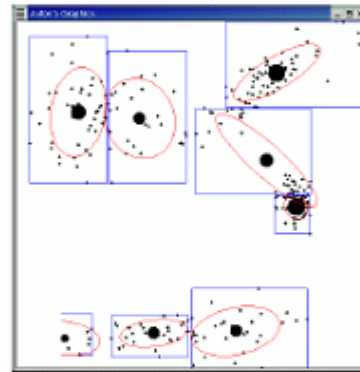
Ένα k-διάστατο δένδρο πρώτου επιπέδου



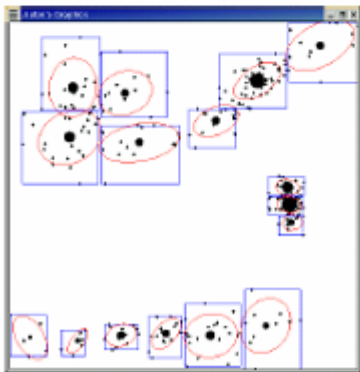
Ένα k-διάστατο δένδρο δεύτερου επιπέδου



Ένα k-διάστατο δένδρο τρίτου επιπέδου



Ένα k-διάστατο δένδρο τέταρτου επιπέδου



Ένα k-διάστατο δένδρο πέμπτου επιπέδου



Ένα k-διάστατο δένδρο έκτου επιπέδου

Σχ.5.1: Παράδειγμα ομαδοποίησης των σημείων του χώρου με ένα k-διάστατο δένδρο 6 επιπέδων.

Όταν χρησιμοποιείται το δένδρο αυτό για την ομαδοποίηση των χαρακτηριστικών διανυσμάτων πρέπει να λαμβάνεται πάντα υπόψη ότι:

1. Δεν είναι απαραίτητο να σχηματίζεται το δένδρο και για τα  $2N_D$  χαρακτηριστικά διανύσματα. Αρκεί να διατηρείται ένα μόνο πολυδιάστατο κλειδί ανά τύπου D περιοχή (διατηρώντας για παράδειγμα το κλειδί που έχει μια πρώτη μη αρνητική συνιστώσα,

πολλαπλασιάζοντας και με  $-1$  αν είναι απαραίτητο). Σε αυτήν την ταξινόμηση ένα  $k$ -διάστατο δένδρο έχει 2 συμμετρικά κύρια κλαδιά (χωρισμένα από ένα υπερεπίπεδο συντεταγμένων) και επομένως, αρκεί η αποθήκευση του ενός από τα δύο.

2. Υπάρχει κάποια ελευθερία σχετικά με την επιλογή του ισομετρικού μετασχηματισμού που αντιστοιχεί μία τύπου D σε μία τύπου R περιοχή. Αν συμπεριληφθούν και οι  $\delta$  ισομετρικοί μετασχηματισμοί τότε δημιουργούνται  $\delta$  είσοδοι για κάθε τύπου D περιοχή στο  $k$ -διάστατο δένδρο, με συνέπεια την αύξηση του μεγέθους του. Η λύση στο πρόβλημα που παρουσιάζεται, είναι η αγνόηση των ισομετρικών μετασχηματισμών των περιοχών τύπου D. Αντίθετα, συμπεριλαμβάνονται οι ισομετρικοί μετασχηματισμοί, εφαρμοζόμενοι όμως αυτή τη φορά στην προς κωδικοποίηση τύπου R περιοχή και αναζητείται για κάθε ένα από αυτούς ο πλησιέστερος γείτονας μέσα στο  $k$ -διάστατο δένδρο. Διατηρείται με τον τρόπο αυτό το μέγεθος το δένδρου σταθερό χωρίς να περιορίζονται οι επιλογές ως προς τον ισομετρικό μετασχηματισμό, πράγμα που ευνοεί μία καλύτερη προσέγγιση.

Ο χρόνος προεπεξεργασίας που απαιτείται για το σχηματισμό του δένδρου δεν αποτελεί παρά μικρό ποσοστό του κύριου χρόνου κωδικοποίησης κάτι το οποίο επιβεβαιώνουν και τα πειραματικά αποτελέσματα.

Κλείνοντας την ενότητα των χαρακτηριστικών διανυσμάτων, αναφέρουμε ότι ένας πρόδρομος αυτών είχε μελετηθεί από τους *Hurtgen* και *Stiller*<sup>6</sup>. Όπως και στην περίπτωση των Fisher, Jacobs και Boss που θα εξεταστεί σε επόμενη ενότητα, μια υποεικόνα διαιρείται σε τετρατημόρια και υπολογίζονται οι μέσες τους εντάσεις. Έπειτα κατασκευάζεται ένα διάνυσμα αποτελούμενο από τέσσερις συνιστώσες, με βάση τον ακόλουθο τρόπο: η  $i$ -οστή συνιστώσα είναι 1 αν η μέση τιμή του  $i$ -οστού τετρατημορίου είναι μεγαλύτερη της συνολικής μέσης τιμής της περιοχής και 0 σε διαφορετική περίπτωση. Επομένως με βάση

την ορολογία των συγγραφέων το διάνυσμα που κατασκευάζεται με τον παραπάνω τρόπο είναι το χαρακτηριστικό διάνυσμα που μελετήθηκε σε αυτήν την ενότητα που έχει προέλθει από υποδειγματοληψία σε διάσταση  $d=2*2=4$  και για κάθε συνιστώσα του οποίου έχει δεσμευθεί 1 ψηφίο. Το σύνολο λοιπόν των διαφορετικών χαρακτηριστικών διανυσμάτων είναι  $2^4=16$ . Εξαιτίας των αυστηρών περιορισμών κατά τη δημιουργία τους, η αναζήτηση με βάση τον πλησιέστερο γείτονα δεν είναι πρακτική, για αυτό και γίνεται ως εξής: Βρίσκεται το χαρακτηριστικό διάνυσμα της κάθε περιοχής τύπου R και στη συνέχεια αυτή συγκρίνεται, σύμφωνα με τη μέθοδο των ελαχίστων τετραγώνων, μόνο με τις περιοχές τύπου D που έχουν το ίδιο χαρακτηριστικό διάνυσμα με την υπό εξέταση τύπου R περιοχή.

### 5.3 Μέθοδοι ταξινόμησης

Οι μέθοδοι ταξινόμησης που παρουσιάζονται στη συνέχεια μελετούν μόνο την περίπτωση  $p=1$ , την περίπτωση δηλαδή κατά την οποία χρησιμοποιείται ένα μοναδικό σταθερό διάνυσμα βάσης ( $B=1/\sqrt{n}(1,\dots,1)$ ). Έχουμε όμως ήδη παρατηρήσει ότι η μέθοδος επεκτείνεται και στη γενική περίπτωση που χρησιμοποιούνται  $p>1$  διανύσματα βάσης, με την προϋπόθεση ότι έχουν γίνει κάποιου είδους τροποποιήσεις. Στην ουσία οι τροποποιήσεις αυτές ισοδυναμούν με την αντικατάσταση των τύπου D περιοχών από το σύνολο των  $2N_D$  χαρακτηριστικών διανυσμάτων  $\pm\varphi(D_j)$ .

#### 5.3.1 Η προσέγγιση του Jacquin

Στην αρχική του εργασία, ο Jacquin χρησιμοποίησε μία μέθοδο ταξινόμησης προερχόμενη από την εργασία των Rammaurthi και Gersho<sup>7</sup>. Οι περιοχές τύπου D ταξινομούνται με βάση τα γεωμετρικά χαρακτηριστικά τους όπως τα αντιλαμβάνεται ο άνθρωπος. Σύμφωνα με την άποψη αυτή, οι περιοχές μιας εικόνας διαιρούνται σε τρεις μεγάλες κατηγορίες: οι περιοχές με αποχρώσεις, οι περιοχές με αιμές και οι ενδιάμεσες περιοχές. Στις περιοχές

με αποχρώσεις οι μεταβολές της έντασης μέσα στην περιοχή είναι πολύ μικρές, ενώ στις αντίστοιχες με ακμές, παρουσιάζεται μία σημαντική μεταβολή στην ένταση π.χ. κατά μήκος του συνόρου ενός αντικειμένου που περιέχεται στην εικόνα. Η τάξη των περιοχών με ακμές διαιρείται σε δύο υποκατηγορίες: στις απλές περιοχές με ακμές και στις μικτές περιοχές με ακμές. Οι ενδιάμεσες περιοχές παρουσιάζουν μεγαλύτερες διακυμάνσεις στην ένταση από τις περιοχές με αποχρώσεις, χωρίς ωστόσο να παρατηρείται η ίδια σημαντική μεταβολή στην ένταση όπως συμβαίνει στην περίπτωση των περιοχών με ακμές. Επομένως, οι ενδιάμεσες περιοχές μπορούν να χαρακτηριστούν ως περιοχές που περιέχουν υφή. Αφού οι περιοχές τύπου R που περιέχουν αποχρώσεις μπορούν να προσεγγιστούν ικανοποιητικά από τη σταθερή περιοχή B που προαναφέρθηκε, πολλαπλασιασμένη με ένα κατάλληλο παράγοντα  $b$ , δεν είναι απαραίτητο να ξεκινά η αναζήτηση μιας κατάλληλης τύπου D περιοχής για αυτές (θεωρείται δηλαδή ότι η αντίθεση είναι 0). Με βάση λοιπόν τη μέθοδο αυτή, υπάρχουν μόνο δύο μεγάλες κατηγορίες περιοχών, μέσα σε μία από τις οποίες πρέπει να αναζητηθεί η κατάλληλη περιοχή τύπου D για κάθε τύπου R που δεν έχει απόχρωση.

### 5.3.2 Ταξινόμηση με βάση την ένταση και τη διακύμανσή της

Μία πιο λεπτομερής τεχνική ταξινόμησης προτάθηκε από τους *Boss Fisher* και *Jacobs*<sup>8</sup>. Πριν αναπτύξουμε τη διαδικασία, σημειώνουμε ότι θα χρησιμοποιήσουμε το σύμβολο  $\tau_k, k=1, \dots, 8$  για να συμβολίσουμε ένα από τους 8 ισομετρικούς μετασχηματισμούς και το  $\tau_B \in \{\tau_1, \dots, \tau_8\}$  για τον ισομετρικό μετασχηματισμό ενός τετραγώνου B. Η μέθοδος λειτουργεί με τον ακόλουθο τρόπο: Μία τετραγωνική περιοχή χωρίζεται σε τεταρτημόρια. Για κάθε τεταρτημόριο υπολογίζεται η μέση τιμή της έντασης των εικονοστοιχείων που περιέχει καθώς και η αντίστοιχη τυπική απόκλιση. Όπως φαίνεται στο σχήμα 5.2, για κάθε  $N * N$  περιοχή B υπάρχει μοναδική ισομετρία  $\tau_B \in \{\tau_1, \dots, \tau_8\}$  με την ακόλουθη ιδιότητα: Αν με  $B_i^*$  συμβολίσουμε τις 4 μέσες



τιμές ( $i=1,\dots,4$ ) της έντασης που αντιστοιχούν σε κάθε ένα από τα τεταρτημόρια της περιοχής, τότε αυτές μπορούν να ταξινομηθούν σε μια από τις κανονικές τάξεις (ομάδες). Αυτές οι ομάδες αναφέρονται στους τρόπους με τους οποίους μπορεί με βάση τη μέση τιμή  $B_i^*$  να καταταχθεί η κάθε περιοχή. Είναι εύκολο να παρατηρήσει κανείς, ότι τα τετράγωνα πάντοτε μπορούν να προσανατολιστούν, με βάση τους ισομετρικούς μετασχηματισμούς, ώστε οι μέσες εντάσεις των τεταρτημορίων να ταξινομούνται με έναν από τους τρεις παρακάτω τρόπους:

$$\text{Τάξη 1: } B_1^* \geq B_2^* \geq B_3^* \geq B_4^*$$

$$\text{Τάξη 2: } B_1^* \geq B_2^* \geq B_4^* \geq B_3^*$$

$$\text{Τάξη 3: } B_1^* \geq B_4^* \geq B_2^* \geq B_3^*$$

Από τη στιγμή που έχει τροποποιηθεί καταλλήλως ο προσανατολισμός της τετραγωνικής περιοχής, υπάρχουν  $4!=24$  διαφορετικές πιθανές διατάξεις των τυπικών αποκλίσεων οι οποίοι καθορίζουν και τις 24 υποτάξεις για κάθε μεγάλη κανονική τάξη. Κατά συνέπεια και για τις τρεις μεγάλες τάξεις υπάρχουν 72 υποομάδες. Αφού ταξινομηθεί η κάθε περιοχή  $D$  σε μία από τις 72 υποομάδες βρίσκεται σε ποια τάξη και ποια υποομάδα ανήκει η προς μελέτη τύπου R περιοχή. Είναι επόμενο, η προσέγγιση της τύπου R περιοχής  $aD+bB$  να ανήκει στην ίδια τάξη και ίσως και στην ίδια υποομάδα, πράγμα που σημαίνει ότι το ίδιο θα ισχύει και για το κωδικό διάνυσμα  $D$ . Τα παραπάνω ισχύουν στην περίπτωση που η αντίθεση είναι θετική. Αν ο παράγοντας αντίθεσης  $a$  της προσέγγισης  $aD+bB$  της τύπου R περιοχής  $R$ .

B	$\tau_B$	$\tau_B B$	Κανονική τάξη
	$\tau_3$		1
	$\tau_5$		2
	$\tau_5$		3
	$\tau_7$		1
	$\tau_2$		2
	$\tau_2$		3
	$\tau_3$		2
	$\tau_5$		1
	$\tau_4$		3
	$\tau_7$		2
	$\tau_2$		1
	$\tau_6$		3

B	$\tau_B$	$\tau_B B$	Κανονική τάξη
	$\tau_3$		3
	$\tau_4$		1
	$\tau_4$		2
	$\tau_7$		3
	$\tau_6$		2
	$\tau_6$		1
	$\tau_8$		3
	$\tau_8$		2
	$\tau_8$		1
	$\tau_1$		3
	$\tau_1$		2
	$\tau_1$		1

Σχ.5.2: Απεικόνιση των κανονικών τάξεων. Οι στήλες με το όνομα B αφορούν τις τετραγωνικές περιοχές με τις μέσες τιμές των εντάσεων των τεταρτημορίων. Υπάρχουν  $4!=24$  διατάξεις των μέσων τιμών των τεταρτημορίων. Σε κάθε περίπτωση, η επόμενη στήλη με το όνομα  $\tau_B$  εικονίζει τον ισομετρικό μετασχηματισμό που καθορίζει την κανονική τάξη της B η οποία παρατίθεται στην τελευταία στήλη.

είναι αρνητικός, τότε οι διατάξεις των τυπικών αποκλίσεων πρέπει να τροποποιηθούν αναλόγως, αφού η μεγαλύτερη απολύτως τυπική απόκλιση των τεταρτημορίων μίας περιοχής  $D$  είναι η μικρότερη για αρνητική αντίθεση. Επομένως, για μία δεδομένη τύπου R περιοχή πρέπει να γίνει αναζήτηση δύο τύπου D περιοχών από τις 72 υποομάδες. Αφού ταξινομηθούν κατά Fisher όλες οι τύπου D περιοχές υποθέτοντας ότι η αντίθεση είναι θετική, βρίσκεται η τύπου D περιοχή που την προσεγγίζει καλύτερα. Στη συνέχεια θεωρώντας ότι ο συντελεστής αντίθεσης είναι αρνητικός, η διάταξη των μέσων τιμών και τυπικών αποκλίσεων αντιστρέφεται με αποτέλεσμα την απόκτηση μίας νέας ταξινόμησης για την οποία βρίσκεται και πάλι η περιοχή τύπου D που δίνει το

μικρότερο σφάλμα. Τέλος, από τις δύο περιοχές που έχουν ξεχωρίσει, επιλέγεται εκείνη που δίνει το μικρότερο σφάλμα και αποθηκεύεται η διεύθυνσή της.

Οι δύο αυτές αναζητήσεις που προτείνει ο Fisher αναπαριστώνται με τα σύμβολα  $\tau_R^{-1}\tau_D$  και  $\tau_{-R}^{-1}\tau_D$  που ισοδυναμούν με την κωδικοποίηση του  $\tau_R R$  από το  $\tau_D D$  και του  $\tau_{-R} R$  από το  $\tau_D D$  αντιστοίχως. Στην πρώτη περίπτωση, σύμφωνα με τα προηγούμενα, ο συντελεστής της αντίθεσης αναμένεται να είναι θετικός ή μηδέν, ενώ στη δεύτερη αρνητικός ή μηδέν. Αυτή η τεχνική μειώνει τον αριθμό των τύπου D περιοχών με τους ισομετρικούς μετασχηματισμούς από  $8N_D$  σε  $2N_D$  ( όπου  $N_D$  ο αριθμός των αρχικών τύπου D περιοχών που δεν έχουν υποστεί ακόμα γεωμετρικό μετασχηματισμό) χωρίς σημαντική απώλεια στην ποιότητα. Επιπλέον, ο αριθμός των τύπου D περιοχών που εξετάζονται, μπορεί να περιοριστεί ακόμα περισσότερο αναζητώντας ένα μεταβλητό αριθμό τάξεων με βάση τους παρακάτω τρόπους:

1. *Αναζήτηση σε μια τάξη* : Η περιοχή τύπου R συγκρίνεται με:
  - a) Όλες τις περιοχές  $\tau_R^{-1}\tau_D D$  ώστε οι περιοχές με R και D να έχουν την ίδια κανονική τάξη και την ίδια υποτάξη.
  - b) Όλες τις περιοχές  $\tau_{-R}^{-1}\tau_D D$  ώστε οι -R ( είδωλο της R ως προς την αρχή των αξόνων, περίπτωση αρνητικής αντίθεσης) και οι D να έχουν την ίδια κανονική τάξη και την ίδια υποτάξη.
2. *Αναζήτηση σε 3 τάξεις* : Η τύπου R περιοχή R συγκρίνεται με:
  - a) όλες τις περιοχές  $\tau_R^{-1}\tau_D D$  ώστε οι R και D να έχουν την ίδια υποτάξη
  - b) όλες οι περιοχές  $\tau_{-R}^{-1}\tau_D D$  ώστε οι -R και D να έχουν την υποτάξη
3. *Αναζήτηση σε 24 υποτάξεις* : Η τύπου R περιοχή R συγκρίνεται :
  - a) όλες τις περιοχές  $\tau_R^{-1}\tau_D D$  ώστε οι R και D να έχουν την ίδια κανονική τάξη.
  - b) Όλες τις περιοχές  $\tau_{-R}^{-1}\tau_D D$  ώστε οι -R και D να έχουν την ίδια κανονική τάξη .

4. Αναζήτηση σε 72 τάξεις : Εδώ εξετάζονται όλες οι τύπου  $D$  περιοχές.

Επομένως οι  $R$  συγκρίνονται με όλες τις περιοχές  $\tau_R^{-1}\tau_D D$  και τις  $\tau_{-R}^{-1}\tau_D D$ .

Αν και αυτή η ταξινόμηση παρέχει ένα ποικίλο αριθμό τάξεων προς εξέταση, δεν υπάρχει η φυσική έννοια της γειτνίασης μεταξύ των τάξεων. Επομένως, αν δε βρεθεί μία ικανοποιητική προσέγγιση ανάμεσα στα στοιχεία της τάξης στα οποία έχει περιοριστεί η αναζήτηση, δεν είναι δυνατό να συνεχιστεί η αναζήτηση σε μία « γειτονική» τάξη. Αυτό το πρόβλημα γίνεται ιδιαίτερα έντονο όταν μια τάξη είναι κενή. Μία λύση του προβλήματος αυτού έχει προταθεί από τους *Caso, Obrador και Kim*<sup>9</sup> στην εργασία τους, όπου η μη ελαστική διάταξη των αποκλίσεων έχει αντικατασταθεί από ένα διάνυσμα που οι συνιστώσες του είναι οι αποκλίσεις. Αυτά τα διανύσματα των τυπικών αποκλίσεων διακριτοποιούνται με αποτέλεσμα τη δημιουργία ενός συνόλου ομάδων κάθε μία από τις οποίες γειτονεύει με πολλές άλλες στις οποίες μπορεί να επεκταθεί μια αναζήτηση. Μία άλλη λύση δίνεται σε επόμενη ενότητα.

### 5.3.3 Ταξινόμηση με βάση το αρχέτυπο

Πρόκειται για μία μέθοδο που καθορίζει εκ των προτέρων τις ομάδες, στηριζόμενη σε κάποιες εμπειρικές μελέτες που εκτελούνται πάνω σε μία συλλογή από εκπαιδευτικές εικόνες και παρουσιάστηκε από τους *Boss και Jacobs*<sup>10</sup>. Ένα πρωτότυπο για ένα σύνολο κωδικών διανυσμάτων δίνεται από εκείνο το ειδικό διάνυσμα το οποίο μπορεί να «καλύψει» όλα τα άλλα σύμφωνα με τη μέθοδο των ελαχίστων τετραγώνων. Για ένα σύνολο διανυσμάτων  $D_i$  είναι το διάνυσμα  $D_k$  για το οποίο ισχύει:

$$D_i = \arg \min_{D_k} \sum_{i \neq k} \min_{a,b} \|D_i - (aD_k + bB)\| \quad (5.24)$$

Αρχίζοντας από μία αυθαίρετη ταξινόμηση, όπως αυτή του Fisher για παράδειγμα, των υποπεριοχών διαφόρων εκπαιδευτικών εικόνων, υπολογίζεται το πρωτότυπο για κάθε ομάδα. Στη συνέχεια, τα διανύσματα-περιοχές

ανακατανέμονται με βάση το πρωτότυπο που τις προσεγγίζει καλύτερα. Οδηγούμαστε συνεπώς σε μία νέα ταξινόμηση και η επανάληψη της προσπάθειας εύρεσης ενός νέου πρωτοτύπου για κάθε μία από τις νέες ομάδες που έχουν προκύψει, επιφέρει νέα ανακατανομή των διανυσμάτων. Η επαναληπτική διαδικασία που μόλις περιγράφηκε, τερματίζεται όταν η επόμενη επανάληψη δεν έχει καμία επίδραση πάνω στην ήδη προϋπάρχουσα διάταξη των διανυσμάτων. Το τελικό σύνολο των πρωτοτύπων διανυσμάτων, κάθε ένα από τα οποία αντιπροσωπεύει και μία ομάδα, ενσωματώνεται στα δεδομένα της κωδικοποίησης. Για μία συγκεκριμένη προς συμπίεση εικόνα, ο κωδικοποιητής καθορίζει το σύνολο των περιοχών τύπου  $D$  και στη συνέχεια τις ταξινομεί με βάση τα πρωτότυπα διανύσματα. Με τον τρόπο αυτό, αναμένεται ότι η κάθε τύπου  $R$  περιοχή μπορεί να προσεγγιστεί ικανοποιητικά από μια τύπου  $D$ , η οποία ανήκει στην ομάδα του πρωτοτύπου που την προσεγγίζει καλύτερα. Κατά συνέπεια, για να επιτευχθεί μία αποκωδικοποίηση με συγκεκριμένη πιστότητα εικόνας, δε χρειάζεται να αναζητηθεί το κατάλληλο κωδικό διάνυσμα σε τόσες πολλές ομάδες όπως προηγουμένως (π.χ. ταξινόμηση Fisher), γεγονός που μειώνει το χρόνο κωδικοποίησης. Από την άλλη πλευρά, η διαδικασία ταξινόμησης είναι πιο περίπλοκη σε σχέση με τις συμβατικές μεθόδους. Ως συνέπεια, οι τελευταίες προτιμώνται για συμπίεση χαμηλής πιστότητας σε μικρό χρόνο, ενώ η υψηλή πιστότητα επιτυγχάνεται πιο γρήγορα με τη μέθοδο του πρωτοτύπου.

#### 5.3.4 Αμετάβλητες αποκλίσεις

Στην εργασία του, ο Novak<sup>11</sup> αναθέτει ένα τετρασδιάστατο διάνυσμα σε κάθε περιοχή. Οι συνιστώσες του διανύσματος αυτού είναι συγκεκριμένες αμετάβλητες αποκλίσεις που ορίζονται από τις διαβαθμίσεις του γκρι μέσα στην περιοχή. Μία χρήσιμη ιδιότητα των αποκλίσεων αυτών είναι, ότι η τιμή τους παραμένει αμετάβλητη ως προς τους ισομετρικούς μετασχηματισμούς της περιοχής, δηλαδή ένα χαρακτηριστικό διάνυσμα είναι αρκετό για κάθε

περιοχή τύπου D. Οι ισομετρίες της έχουν το ίδιο διάνυσμα απόκλισης. Παρ' όλα αυτά, οι αποκλίσεις δεν είναι οι ίδιες όσον αφορά το μετασχηματισμό τύπου affine της φωτεινότητας. Για να αντιμετωπιστεί το πρόβλημα αυτό, ο Novak πρότεινε μια διαδικασία κανονικοποίησης. Υπάρχουν όμως τρία προβλήματα με την προσέγγιση που επιχειρεί ο Novak:

1. Οι περιοχές αρνητικής έντασης δεν λαμβάνονται υπόψη με απόρροια την πιθανή μείωση της πιστότητας.
2. Οι τιμές των αμετάβλητων αποκλίσεων κυμαίνονται μεταξύ αρκετών δυνάμεων του 10 με αποτέλεσμα την ανάγκη για λογαριθμοποίηση.
3. Και πάνω από όλα, η μέθοδος είναι διαισθητική με την έννοια ότι δεν έχει θεμελιωθεί η θεωρία που να υποστηρίζει ότι η εγγύτητα στο χώρο αυτού του είδους των χαρακτηριστικών διανυσμάτων διασφαλίζει και καλή προσέγγιση ως προς τα ελάχιστα τετράγωνα.

## 5.4 Μέθοδοι ομαδοποίησης

Όπως παρατηρήθηκε και στην ενότητα 5.2, με τη βοήθεια των χαρακτηριστικών διανυσμάτων κατέστη δυνατή η μετατόπιση του προβλήματος των ελάχιστων τετραγώνων μεταξύ τύπου R και τύπου D διανυσμάτων στο πιο απλό υπολογιστικά πρόβλημα εύρεσης του πλησιέστερου γείτονα. Αναφέρθηκε προηγουμένως το παράδειγμα το *k*-διάστατου δένδρου, το οποίο αποτελεί έναν ενδεικτικό τρόπο με τον οποίο τα τύπου D χαρακτηριστικά διανύσματα μπορούν να ταξινομηθούν σε ομάδες και να αποφευχθεί ένας μεγάλος αριθμός συγκρίσεων με τη σύγκριση του τύπου R χαρακτηριστικού διανύσματος μόνο με το αντιπροσωπευτικό κάθε ομάδας. Στη συνέχεια, αφού βρεθεί ποια ομάδα προσεγγίζει καλύτερα το διάνυσμα που εξετάζεται, η διαδικασία προχωρεί με τη σύγκριση του τελευταίου με τα υπόλοιπα διανύσματα τη ομάδας ώστε να αποκτηθεί εκείνο που πιθανότατα θα δίνει και τη βέλτιστη προσέγγιση με την έννοια των ελαχίστων τετραγώνων. Στο εδάφιο αυτό θα εξετάσουμε μερικούς από τους πιο διαδεδομένους

τρόπους ομαδοποίησης και θα μελετήσουμε τους τρόπους με τους οποίους μπορούν να συνδυαστούν με τεχνικές που έχουμε αναφέρει ως τώρα. Πριν γίνει αναφορά σε μερικές από τις πιο διαδεδομένες μεθόδους ομαδοποίησης στη συμπίεση με fractal, τους παρατίθενται κάποιοι ορισμοί σχετικά με την ομαδοποίηση διανυσμάτων<sup>12</sup>.

**Ορισμός 1:** Έστω ένα σύνολο  $\Omega$  με πεπερασμένο αριθμό στοιχείων κάθε ένα από τα οποία είναι διάνυσμα στο χώρο  $\mathbb{R}^n$ . Καλούμε δομή ομάδας (cluster structure) ή ομαδοποίηση (clustering), του συνόλου  $\Omega$  την πεπερασμένη συλλογή  $\{\Omega_k\}_{0 \leq k < M}$  από  $M$  μη κενά υποσύνολα του  $\Omega$  και μία απεικόνιση

$$c: \mathbb{R}^N \rightarrow \{0, 1, \dots, M-1\} \quad (5.25)$$

ώστε  $\bigcup_{0 \leq k < M} \Omega_k = \Omega$  και  $c^{-1}(k) \neq \emptyset$  για κάθε  $0 \leq k < M$ . Η απεικόνιση  $c$  καλείται και χάρτης ταξινόμησης (classification map). Κάθε υποσύνολο  $\Omega_k$  καλείται ομάδα για  $k=0, 1, \dots, M-1$ .

**Ορισμός 2:** Η διάμετρος της κάθε ομάδας  $\Omega_k$ ,  $k=0, 1, \dots, M-1$  δίνεται από τη σχέση:

$$\delta(\Omega_k) = \sup_{d, r \in \Omega_k} d(d, r) \quad (5.26)$$

**Ορισμός 3:** Η ταχύτητα ομαδοποίησης καθορίζεται από τη σχέση:

$$\sigma(\Omega) = \frac{\max_{0 \leq k < M} \{|\Omega_k|\} - 1}{|\Omega|} \quad (5.27)$$

Η ομαδοποίηση καλείται ακαριαία αν  $\sigma(\Omega) = 0$ .

**Ορισμός 4:** Το σφάλμα ομαδοποίησης ορίζεται με βάση τον τύπο:

$$\varepsilon(\Omega) = \sup_{d \in \mathbb{R}^N} (d(\mathbf{d}, \Omega_{c(D)}) - d(\mathbf{d}, \Omega)) \quad (5.28)$$

Η ομαδοποίηση καλείται τέλεια αν  $\varepsilon(\Omega) = 0$ .

Στη συνέχεια, γίνεται λόγος για δύο αλγορίθμους ομαδοποίησης, τον αλγόριθμο της *Επέκτασης* (*Bubbling Algorithm*) και τον αλγόριθμο *Heckbert*, για τους οποίους θα θεωρηθεί ότι δεν υπάρχουν περιορισμοί σε χώρο και χρόνο. Τέλος, η ενότητα κλείνει με την αναφορά στην εργασία των *Saupe* και *Hamzaoui* πάνω στην ομαδοποίηση με βάση τα χαρακτηριστικά διανύσματα και στην σύνδεσή της με άλλες τεχνικές για την επίτευξη μιας καλύτερης fractal συμπίεσης.

#### 5.4.1 Ο Αλγόριθμος της Επέκτασης

Έστω ότι ο αριθμός των πιθανών διανυσμάτων διανυσμάτων στο  $R^n$  δεν είναι απαγορευτικός για την αποθήκευσή του σε υπολογιστή. Είναι δυνατή επομένως η δέσμευση μνήμης για την καταχώρηση μιας τιμής για κάθε διάνυσμα του  $R^n$ . Στην περίπτωση αυτή ο αλγόριθμος που εξετάζεται εδώ είναι ο τέλειος για μία ακαριαία και τέλεια ομαδοποίηση. Η μηδενική ταχύτητα  $\sigma(\Omega)=0$  σημαίνει ότι κάθε αρχική ομάδα περιέχει ένα ακριβώς διάνυσμα, το διάνυσμα αναφοράς. Επομένως αυτό που μένει είναι ο προσδιορισμός του χάρτη ταξινόμησης  $\epsilon$ .

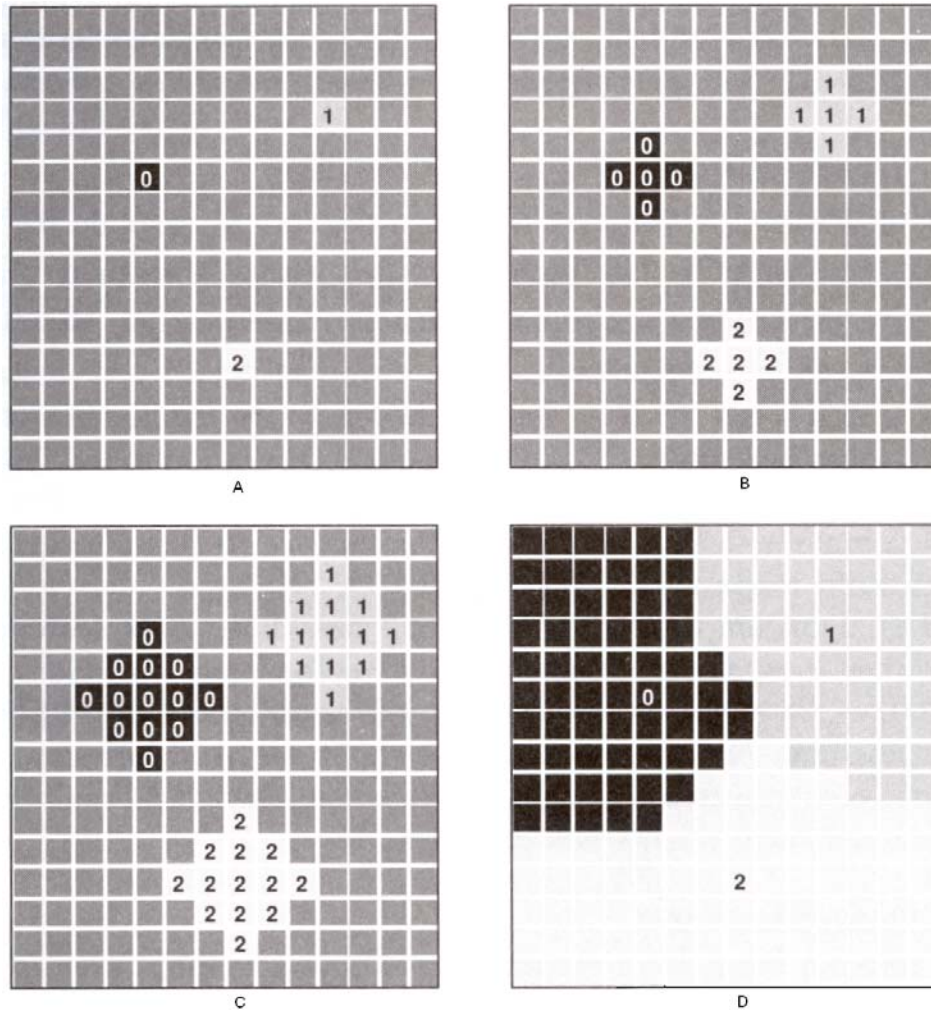
Ο *Αλγόριθμος της Επέκτασης* (*Bubbling Algorithm*) ξεκινά τοποθετώντας όλο το χώρο στην ομάδα με τον αύξοντα αριθμό  $M$ . Στη συνέχεια ο αύξων αριθμός τιμής  $k$  ανατίθεται στο μοναδικό διάνυσμα που εκπροσωπεί την αντίστοιχη ομάδα  $\Omega_k$  για  $k=0,1,\dots,M-1$ . Στη συνέχεια, όλες αυτές οι ομάδες με δείκτη μικρότερο του  $M$ , επεκτείνονται ώστε να συμπεριλαμβάνονται τα διανύσματα που ανήκουν στην ομάδα  $M$  με ένα βήμα τη φορά, μία διεύθυνση και κατά μία συνιστώσα ώσπου να καλυφθεί όλο ο χώρος και να μην υπάρχει άλλο διάνυσμα στον  $R^n$  που να ανήκει στην ομάδα  $M$ . Ακολουθεί η πιο λεπτομερής περιγραφή του αλγόριθμου σε βήματα.

Έστω ένα πεπερασμένο σύνολο διανυσμάτων στο  $R^n$  και μία προσωρινή μνήμη που έχει δεσμευτεί για την αποθήκευση των διανυσμάτων του χάρτη ταξινόμησης  $\epsilon$ . Υποτίθεται ότι το  $i$ -οστό μοναδιαίο διάνυσμα έχει την ακόλουθη μορφή:



$$u_i = (\overbrace{0,0,\dots,0}^{i-1 \text{ μηδενικά}}, 1, 0, \dots, 0) \quad (5.29)$$

1. Για όλα τα  $r \in \mathfrak{R}^N$  θέτουμε  $c(r) = M$ .
2. Για οποιοδήποτε  $o_k \in \Omega_k, k = 0, 1, \dots, M-1$  θέτουμε  $c(o_k) = k$ .
3. Έστω  $fin$  η μεταβλητή που δείχνει το τερματισμό του βρόχου. Θέτουμε  $fin=0$ . Για όλα τα  $r \in \mathfrak{R}^N$ , αν  $c(r) = M$ , θέτουμε  $fin=1$ , διαφορετικά, για όλα τα  $i = 0, 1, \dots, n-1$ , θέτουμε  $c(r + u_i) = c(r)$  αν  $c(r + u_i) = M$  και  $c(r - u_i) = c(r)$  αν  $c(r - u_i) = M$ .
4. Αν  $fin=1$  συνεχίζουμε από το βήμα 2. Διαφορετικά ο αλγόριθμος έχει ολοκληρωθεί.



Σχ.5.3: Ο αλγόριθμος Bubbling

Το σχήμα 5.3 παρουσιάζει μια διαισθητική εικόνα του αλγορίθμου. Σε αυτό το παράδειγμα τα διανύσματα αναφοράς είναι τρία και αντιπροσωπεύουν τις ομάδες 0, 1 και 2 (σχήμα 5.3.A). Το βήμα 3 του αλγορίθμου επαναλαμβάνεται ώσπου να σταματήσουν οι αλλαγές. Τα σχήματα 5.3.B και 5.3.C απεικονίζουν τις πρώτες δύο επαναλήψεις της διαδικασίας και το 5.3.D τον τελικό χάρτη ταξινόμησης.

Ο αλγόριθμος αυτός δεν είναι πρακτικός στις περισσότερες περιπτώσεις. Μπορεί να εφαρμοσθεί σε πολύ μικρές τετραγωνικές περιοχές-διανύσματα. Στην πραγματικότητα είναι χρήσιμος μόνο όταν η διάσταση δεν είναι πολύ μεγάλη, π.χ. 2 ή 3.

#### 5.4.2 Η ομαδοποίηση Heckbert

Όταν η μνήμη δεν επαρκεί για την αποθήκευση, ο χάρτης ταξινόμησης  $c$  πρέπει να περιγραφεί με ένα αλγοριθμικό τρόπο. Ο αλγόριθμος ομαδοποίησης του Heckbert (*Heckbert clustering*), χωρίζει το χώρο σε ομάδες χρησιμοποιώντας υπερεπίπεδα τα οποία είναι κάθετα σε κάποιο σύστημα συντεταγμένων. Η διαίρεση αυτή μπορεί να χρησιμοποιηθεί και ως μέρος του αλγορίθμου για την περιγραφή του χάρτη  $c$ .

Αρχικά επιλέγεται η μεγαλύτερη τρέχουσα ομάδα που στην περίπτωση της συμπίεσης με fractal είναι το σύνολο των domain διανυσμάτων. Στη συνέχεια, βρίσκεται το σύστημα συντεταγμένων ως προς το οποίο η ομάδα έχει τη μεγαλύτερη εξάπλωση και διαιρείται χρησιμοποιώντας ένα υπερεπίπεδο κάθετο στο σύστημα αυτό των συντεταγμένων και με τέτοιο τρόπο ώστε το πλήθος των στοιχείων των δύο νέων ομάδων που προκύπτουν να είναι όσο το δυνατό παραπλήσιο. Η διαδικασία αυτή επαναλαμβάνεται ώσπου να επιτευχθεί ένας προκαθορισμένος στόχος (συγκεκριμένη ταχύτητα, ή μέγεθος ομάδας, ή αριθμός ομάδων).

Σε γενικές γραμμές η μέθοδος αυτή δεν παρέχει την καλύτερη δυνατή ομαδοποίηση από μαθηματικής σκοπιάς (όπως ο προηγούμενος αλγόριθμος).

Παρ' όλα αυτά, η αποτελεσματικότητά της στην αντιμετώπιση τεράστιου αριθμού σημείων-διανυσμάτων στο χώρο την καθιστά πολύ χρήσιμη σε πολλές εφαρμογές.

### ***Αλγόριθμος Heckbert:***

Έστω ένα πεπερασμένο σύνολο διανυσμάτων  $\Omega$  στο χώρο  $R^n$  και ένας ακέραιος  $m$  που αντιστοιχεί στον επιθυμητό αριθμο ομάδων. Ο αλγόριθμος κατασκευάζει το χάρτη ταξινόμησης με μία σειρά από  $n-1$  διαιρέσεις. Κάθε μία χαρακτηρίζεται από τέσσερις παραμέτρους,  $(k, k', i, v)$ , όπου η  $k$  ομάδα χωρίζεται από το  $i$ -οστό υπερπίπεδο στο σημείο με τιμή  $v$  σε δύο ομάδες την  $k$  και  $k'$  για  $x_i < v$  και  $x_i \geq v$  αντίστοιχα. Εν συντομία ο αλγόριθμος διαιρεί από την  $k$  ομάδα στην  $k'$  στην  $i$ -οστή μεταβλητή, πάνω στην τιμή  $v$ . Το κριτήριο με βάση το οποίο καθορίζονται οι διαιρέσεις είναι το ακόλουθο:

1. Θέτουμε  $c_0^0 = \Omega$ .

Κατά την πρώτη διαίρεση,  $s = 1$ :

- 1-a. Επιλογή του αριθμού των ομάδων  $k_1 = 0$ .

- 1-b. Εύρεση της διεύθυνσης  $i_1$  με την ευρύτερη εξάπλωση, δηλ.

$$\max_{x, y \in c_0^0} \{x_{i_1} - y_{i_1}\} = \max_{0 \leq i \leq N} \left\{ \max_{x, y \in c_0^0} \{x_i - y_i\} \right\} \quad (5.30)$$

- 1-c. Βρίσκουμε την τιμή  $v_1$  για την τομή, έτσι ώστε:

$$\left| \#\{ \mathbf{x} \in c_0^0 \mid x_{i_1} < v_1 \} - \#\frac{c_0^0}{2} \right| = \inf_{v \in R} \left| \#\{ \mathbf{x} \in c_0^0 \mid x_{i_1} < v \} - \#\frac{c_0^0}{2} \right| \quad (5.31)$$

- 1-d. Διαμέριση της ομάδας σε δύο διαφορετικές:

$$c_0^1 = \{ \mathbf{x} \in c_0^0 \mid x_{i_1} < v_1 \} \text{ και } c_1^1 = \{ \mathbf{x} \in c_0^0 \mid x_{i_1} \geq v_1 \}. \quad (5.32)$$

Στις επόμενες διαιρέσεις,  $s = 2, 3, \dots, n-1$ :

s-a.  $s = 2, 3, \dots, n-1$ . Επιλέγουμε μία από τις μεγαλύτερες ομάδες  $k_s$ , δηλ.

$$\#c_{k_s}^{s-1} = \max_{0 \leq k < s} \#c_k^{s-1} \quad (5.33)$$

s-b. Εύρεση της διεύθυνσης  $i_s$  με μία από τις μεγαλύτερες εξαπλώσεις, δηλ.

$$\max_{\mathbf{x}, \mathbf{y} \in c_{k_s}^{s-1}} \{x_{i_s} - y_{i_s}\} = \max_{0 \leq i < N} \left\{ \max_{\mathbf{x}, \mathbf{y} \in c_{k_s}^{s-1}} \{x_i - y_i\} \right\} \quad (5.34)$$

s-c. Εύρεση της τμής που καθορίζει την τομή στη μέση  $v_s$ , έτσι ώστε

$$\left| \#\{\mathbf{x} \in c_{k_s}^{s-1} \mid x_{i_s} < v_s\} - \#\frac{c_{k_s}^{s-1}}{2} \right| = \inf_{v \in R} \left| \#\{\mathbf{x} \in c_{k_s}^{s-1} \mid x_{i_s} < v\} - \#\frac{c_{k_s}^{s-1}}{2} \right| \quad (5.35)$$

s-d. Διαίρεση της ομάδας  $c_{k_s}^{s-1}$  σε δύο:

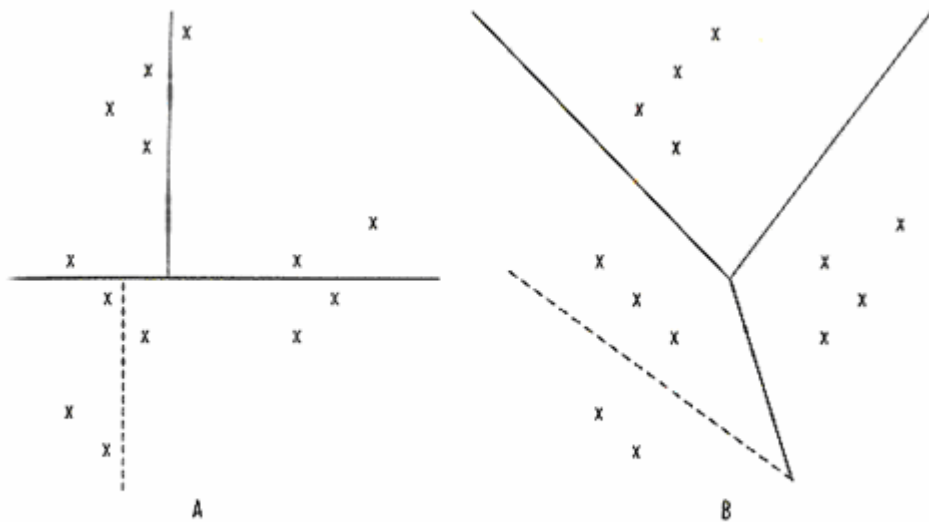
$$c_{k_s}^s = \{\mathbf{x} \in c_{k_s}^{s-1} \mid x_{i_s} < v_s\} \text{ και } c_s^s = \{\mathbf{x} \in c_{k_s}^{s-1} \mid x_{i_s} \geq v_s\}. \quad (5.36)$$

Και ταξινόμηση των υπολοίπων όπως προηγουμένως, δηλ.

$$c_k^s = c_k^{s-1}, \quad 0 \leq k < s \text{ και } k \neq k_s \quad (5.37)$$

Επομένως, δεδομένου ενός διανύσματος  $x$  του συνόλου  $\Omega$ , ο χάρτης ταξινόμησης ορίζεται από τα υπερεπίπεδα.

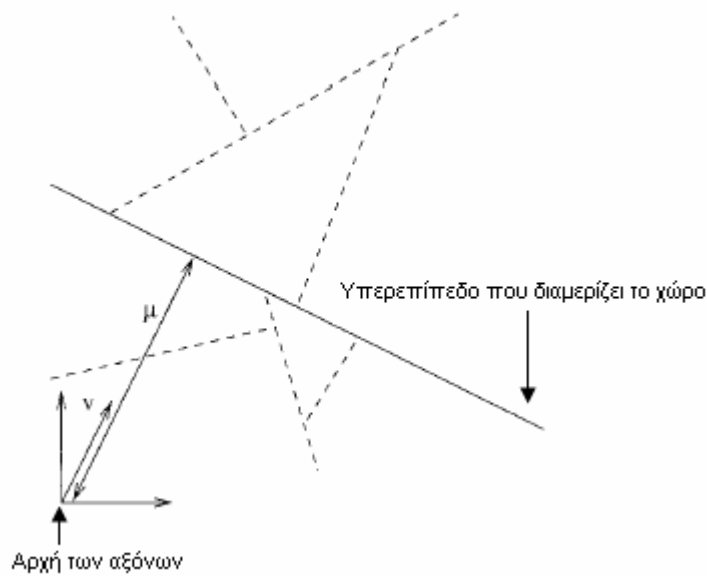
Το σχήμα 5.4.A δείχνει ένα απλό παράδειγμα από ένα σύνολο δεκατριών διανυσμάτων αναφοράς που έχουν χωριστεί σε τέσσερις ομάδες μετά τρεις διαιρέσεις. Το σχήμα 5.4.B δίνει την πιο πιθανή λύση με τρεις διαιρέσεις που βασίζεται στην κοινή λογική. Το σχήμα παρατίθεται για να δειχθεί ότι η λύση που προτείνει η παραπάνω μέθοδος μπορεί να απέχει πολύ από τη βέλτιστη αλλά εξακολουθεί να δίνει μια πρακτική απάντηση στο πρόβλημα της ομαδοποίησης.



Σχ.5.4: Ο αλγόριθμος Heckbert σε μία απλή περίπτωση: Α) Heckbert Β) Οπτική ομαδοποίηση

### 5.4.3 Ευρετική μέθοδος ομαδοποίησης

Εκτός από τον παραπάνω αλγόριθμο υπάρχουν και άλλες μέθοδοι που χρησιμοποιούν υπερεπίπεδα για τη διαμέριση του συνόλου των διανυσμάτων αναφοράς. Η διαφοροποίησή τους έγκειται στο κριτήριο της διαίρεσης. Όπως αναφέρθηκε και προηγουμένως, ο χώρος  $R^n$  χωρίζεται από υπερεπίπεδα διαστάσεως  $n-1$ . Σε κάθε βήμα το ζητούμενο υπερεπίπεδο μπορεί να αναπαρασταθεί από ένα κανονικοποιημένο διάνυσμα  $u \in R^n$  και μια βαθμωτή τιμή  $\mu \in R$ , ώστε το επίπεδο να είναι ορθογώνιο στο  $u$  και να περιέχει το σημείο  $\mu \cdot u$ . Το πρόβλημα επομένως της εύρεσης του κατάλληλου υπερεπιπέδου σε κάθε βήμα μετατοπίζεται στην εύρεση των αντιστοίχων  $\mu, u$ . Κάθε διάνυσμα αναφοράς θα τοποθετείται στη μία πλευρά του επιπέδου ή την άλλη ανάλογα με τον αν το εσωτερικό του γινόμενο με το ορθογώνιο στο επίπεδο διάνυσμα  $u$  είναι μεγαλύτερο ή μικρότερο του  $\mu$ . Το σχήμα 5.5 απεικονίζει την παραπάνω διαδικασία.



Σχ.5.5: Απεικόνιση της διαδικασίας διαμέρισης. Οι διακεκομμένες γραμμές αντιπροσωπεύουν τα υπερεπίπεδα που διαχωρίζουν το ένα μετά το άλλο το χώρο. Η σύγκριση ενός ζεύγους domain-range περιοχών πραγματοποιείται μόνο όταν τα αντίστοιχα χαρακτηριστικά τους διανύσματα ανήκουν στην ίδια ομάδα.

Η πολυπλοκότητα της διαδικασίας διαίρεσης είναι της τάξης  $O(n(N_R + N_D))$ , όπου  $N_R$  ο αριθμός των περιοχών τύπου R.

Ο *R.F Sproull* πρότεινε μία πολύ καλή λύση για την εύρεση του  $\mu$ , σύμφωνα με την οποία η διεύθυνση του επιπέδου δίνεται από το κύριο ιδιοδιάνυσμα του πίνακα συνδιακύμανσης της κατανομής των διανυσμάτων αναφοράς. Η λύση αυτή χαρακτηρίζεται ως βέλτιστη, καθώς ο προσανατολισμός του υπερεπιπέδου μεγιστοποιεί την παραμένουσα τυπική απόκλιση μεταξύ των διανυσμάτων αναφοράς δηλαδή, από όλους τους μονοδιάστατους άξονες που μπορούν να βρεθούν, είναι εκείνος που κάνει την μεγαλύτερη διάκριση μεταξύ των διανυσμάτων.

Ο *Jean Cardinal*<sup>3</sup> προτείνει μία ευρετική μέθοδο υπολογισμού του υπερεπιπέδου διαχωρισμού, καθώς ο υπολογισμός του βέλτιστου ιδιοδιανύσματος μπορεί να αποδειχθεί χρονοβόρος κυρίως για μεγάλες διαστάσεις. Ο αλγόριθμος είναι ο ακόλουθος:

1. Έστω  $\{\phi(R_i) \in \mathfrak{R}^n \mid i=1, \dots, N_R\}$  και  $\{\phi(D_i) \in \mathfrak{R}^n \mid i=1, \dots, N_D\}$  τα σύνολα των χαρακτηριστικών διανυσμάτων των range και domain περιοχών αντίστοιχα. Κανονικά θα έπρεπε να εξεταστεί και η περίπτωση αρνητικών χαρακτηριστικών διανυσμάτων. Στην περίπτωση αυτή το σφάλμα κωδικοποίησης θα ήταν πιθανότατα μικρότερο εξαιτίας της ύπαρξης περισσότερων υποψηφίων για σύγκριση. Ωστόσο η εργασία μόνο με τα θετικά διανύσματα, που ισοδυναμεί με τη μείωση του συνόλου των τύπου D περιοχών στο μισό, δίνει ταχύτερα αποτελέσματα χωρίς βλάβη της γενικότητας του αλγορίθμου.
2. Υπολογισμός του κέντρου βάρους της κατανομής των τύπου R χαρακτηριστικών διανυσμάτων

$$g = \frac{1}{N_R} \sum_{i=1}^{N_R} \phi(R_i) \quad (5.38)$$

3. Για κάθε  $i \in 1, \dots, N_R$ , υπολογίζεται η προβολή του  $\phi(R_i)$  στη μοναδιαία σφαίρα με κέντρο το  $g$ :

$$s_i = \frac{(\phi(R_i) - g)}{\|\phi(R_i) - g\|} \quad (5.39)$$

4. Το υπερεπίπεδο διαχωρισμού θα είναι ορθογώνιο στη διεύθυνση που δίνεται από τη σχέση

$$u = \frac{w}{\|w\|}, \quad w = \frac{1}{N_R} \sum_{i=1}^{N_R} s_i \quad (5.40)$$

και θα περιέχει το κέντρο βάρους  $g$ .

Η τιμή που αντιστοιχεί στο  $\mu$  είναι το εσωτερικό γινόμενο  $\langle g, u \rangle$ .

Ο στόχος της ευριστικής μεθόδου είναι η αναζήτηση μιας σημαντικής από πλευράς διαχωρισμού διεύθυνσης μέσα στο «νέφος» των σημείων. Υπολογίζεται το κέντρο βάρους των σημείων και στη συνέχεια βρίσκεται ο μέσος όρος των διευθύνσεων που «παρατηρούνται» από το σημείο αυτό. Πρέπει να σημειωθεί

ότι αν  $\|w\| = 0$ , πρέπει να επιλεγεί ένα αυθαίρετο διάνυσμα, περίπτωση όμως που δεν έχει ως τώρα αντιμετωπιστεί κατά την πειραματική διαδικασία.

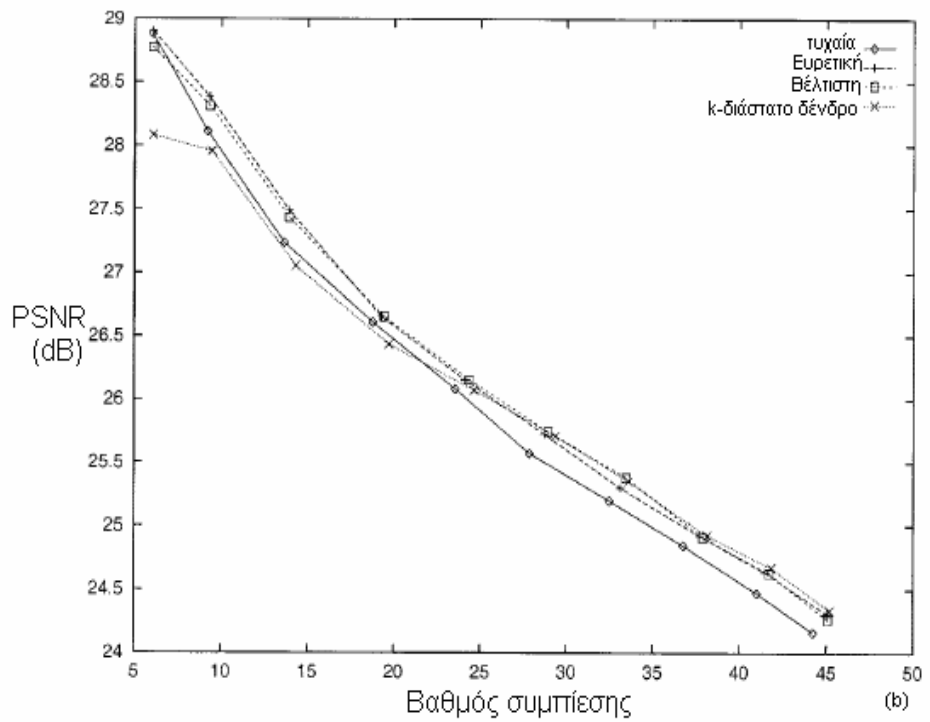
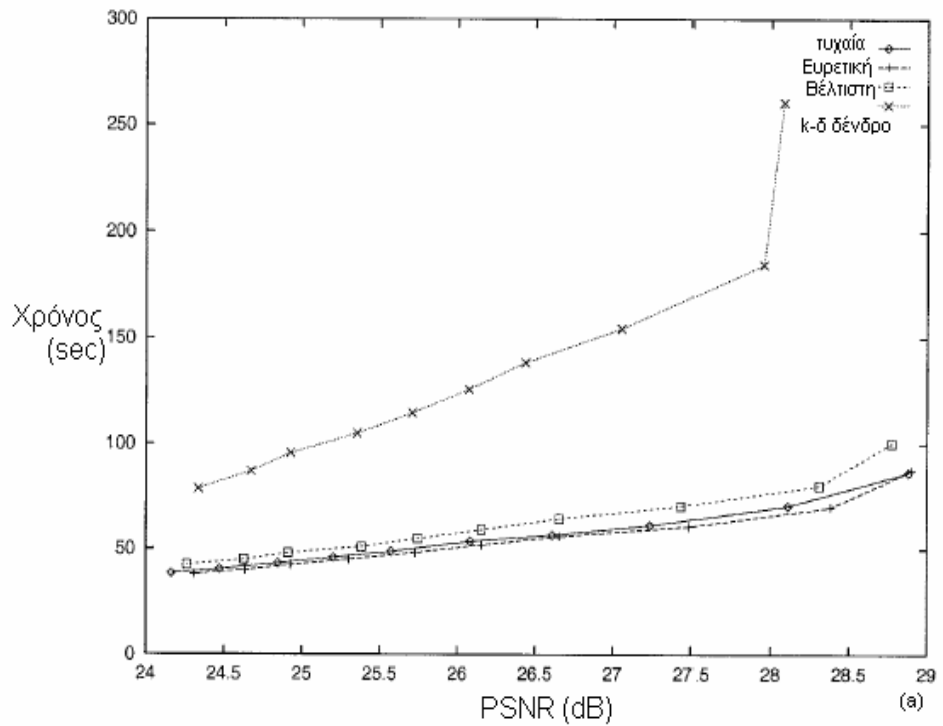
Όπως αναφέρθηκε προηγουμένως, η βέλτιστη μέθοδος  $u_{opt}$  τείνει να μεγιστοποιήσει τη μέση απόσταση μεταξύ ενός τύπου R χαρακτηριστικού διανύσματος και του συνόρου του τμήματος του χώρου-ομάδας που το περιέχει, μειώνοντας κατά συνέπεια την πιθανότητα ύπαρξης ενός γειτονικού διανύσματος σε γειτονική ομάδα που να προσφέρεται για επιτυχή σύγκριση. Η ευρετική μέθοδος έχει με τη σειρά της παρόμοια ιδιότητα.

Η βάση στην οποία στηρίζεται η ευρετική μέθοδος, είναι η υπόθεση ότι η κατανομή των χαρακτηριστικών διανυσμάτων τύπου R είναι ίδια με αυτή των τύπου D διανυσμάτων. Αυτή η υπόθεση δεν είναι αβάσιμη και μπορεί να θεωρηθεί ως πόρισμα της τοπικής ομοιότητας που παρατηρείται μέσα σε μία εικόνα και την οποία εκμεταλεύεται το μοντέλο *PIFS*. Ο αλγόριθμος λαμβάνει υπόψη την παραπάνω ιδιότητα και βρίσκει τα υπερεπίπεδα διαχωρισμού με βάση μόνο τα τύπου R διανύσματα τα οποία καθώς είναι γενικά λιγότερα στον αριθμό από τα τύπου D επιτρέπουν ταχύτερους υπολογισμούς.

Το επόμενο σχήμα συνοψίζει τα αποτελέσματα της σύγκρισης τεσσάρων μεθόδων ομαδοποίησης:

1. Της μεθόδου τυχαίας επιλογής του  $u$ .
2. Της μεθόδου της επιλογής του βέλτιστου διανύσματος  $u, u_{opt}$ .
3. Της ευριστικής μεθόδου που προτείνει ο *J.Cardinal*.
4. Της μεθόδου του  $k$ -διάστατου δένδρου.





Σχ.5.6: Ο μέσος όρος των αποτελεσμάτων της επεξεργασίας διαφόρων εικόνων. α) Χρόνος β) Ποιότητα

Παρατηρώντας τα αποτελέσματα, μπορούν να εξαχθούν τα ακόλουθα συμπεράσματα:

1. Οι τρεις πρώτες μέθοδοι είναι πάντα ταχύτερες του  $k$ -διάστατου δένδρου.  
Ο λόγος αύξησης της επιτάχυνσης κυμαίνεται μεταξύ 1.5 και 3.

2. Η βέλτιστη μέθοδος παράγει εικόνες υψηλότερης ποιότητας από τη βασισμένη στην τυχαία επιλογή μέθοδο.
3. Το πρόσθετο χρονικό διάστημα που απαιτείται για την ολοκλήρωση της εκτέλεσης της βέλτιστης μεθόδου προσεγγίζει το 20% σε σχέση με το χρόνο εκτέλεσης της ευρετικής και τυχαίας.
4. Κατά μέσο όρο, η ποιότητα που δίνει το  $k-d$  δένδρο για σταθερό χρόνο κωδικοποίησης, δεν υπερβαίνει την αντίστοιχη των άλλων τριών αλγορίθμων. Για μικρούς βαθμούς συμπίεσης μάλιστα, η ποιότητα του αλγορίθμου του δένδρου υπολείπεται των τριών.

#### 5.4.4 Ομαδοποίηση με βάση την απόσταση

Στον παρών εδάφιο, καθώς και σε αυτό που ακολουθεί, παρουσιάζονται οι προτάσεις των *Rauf Hamzaoui* και *Dietmar Saupe*<sup>14</sup> για αποδοτική και ταχεία ομαδοποίηση των χαρακτηριστικών διανυσμάτων, οι οποίες συνδυάζονται και με την ταξινόμηση του Fisher. Για το λόγο αυτό, υιοθετείται ο συμβολισμός της ενότητας 5.3.2. Επιπλέον για τις ανάγκες των δύο εδαφίων, θεωρείται ότι η κάθε τύπου R περιοχή R έχει διάσταση  $2^n * 2^n$ . Οι περιοχές τύπου D που έχουν το διπλάσιο μέγεθος των περιοχών τύπου R, έχουν προφανώς διαστάσεις  $2^{n+1} * 2^{n+1}$ . Οι περιοχές που προέρχονται από υποδειγματοληψία των  $N_D$  τύπου D περιοχών μετά από την εφαρμογή κάποιου ισομετρικού μετασχηματισμού και οι οποίες συγκρίνονται στην ουσία με τις τύπου R, συμβολίζονται ως:

$$C_j = S^{(n+1)}(\tau_{D_j} D_j), j = 1 \dots n_D \quad (5.41)$$

όπου  $S^{(n+1)}$ , ένας τελεστής που συμβολίζει την υποδειγματοληψία από τις διαστάσεις  $2^{n+1} * 2^{n+1}$  στις διαστάσεις  $2^n * 2^n$ .

Όπως αναφέρθηκε και στο 1<sup>ο</sup> κεφάλαιο, οι ομάδες ορίζονται ως κύτταρα *Voronoi* (πλησιέστερος γείτονας) στον Ευκλείδειο πολυδιάστατο χώρο. Κάθε σύνολο – κύτταρο  $\Omega_i \subseteq \mathcal{R}^n$ ,  $i = 1 \dots n_m$ , αντιπροσωπεύεται από ένα κέντρο  $\mathbf{m}_i \in \mathcal{R}^n$  και δίνεται από τη σχέση :

$$\Omega_i = \{\mathbf{x} \in \mathfrak{R}^n : \|\mathbf{x} - \mathbf{m}_i\| \leq \|\mathbf{x} - \mathbf{m}_k\|, k = 1 \dots n_m\} \quad (5.42)$$

Για να διασφαλιστεί η μη-επικάλυψη των κελιών ακολουθείται ο εξής κανόνας (Κεφάλαιο 1): Αν το όρισμα του  $\|\mathbf{x} - \mathbf{m}_i\|$  δεν είναι μοναδικό, δηλαδή αν υπάρχουν δύο διαφορετικά κέντρα  $\mathbf{m}_p, \mathbf{m}_j$ , που ικανοποιούν την παραπάνω σχέση, το  $\mathbf{x}$  αντιστοιχίζεται στο κύτταρο με το μικρότερο δείκτη. Επιπλέον, ο  $k$ -οστός πλησιέστερος γείτονας του  $\mathbf{x}$  στο σύνολο των κέντρων  $\|\mathbf{x} - \mathbf{m}_i\|$ , που σχηματίζουν ένα συγκεκριμένο σύνολο  $\Omega$ , συμβολίζεται με  $\Omega^k(x)$ .

Έστω ότι το κέντρο κάθε ομάδας είναι γνωστό. Η διαδικασία κωδικοποίησης μπορεί να επιταχυνθεί με τον τρόπο που περιγράφεται παρακάτω. Σε πρώτη φάση καθορίζεται η ομάδα στην οποία ανήκει κάθε ένα από τα χαρακτηριστικά διανύσματα  $\pm\phi(C_{j,p}), j = 1 \dots n_D, p = 1 \dots 8$  (το  $p$  συμβολίζει τον ισομετρικό μετασχηματισμό). Σε μία αναζήτηση μίας ομάδας συγκρίνεται η τύπου R περιοχή R μόνο με τις τύπου D περιοχές  $\tau_{D_j} D_j$  για τις οποίες τουλάχιστον ένα από τα χαρακτηριστικά διανύσματα  $\pm\phi(C_{j,p})$  είναι στην ομάδα με κέντρο  $\Omega^1(\varphi(R))$ . Αυτή η στρατηγική δικαιολογείται από τα σχόλια που ακολουθούν το *Θεώρημα 6* και την παρατήρηση ότι τα πλησιέστερα χαρακτηριστικά διανύσματα στο  $\varphi(R)$  είναι πιο πιθανό να βρίσκονται σε αυτήν την ομάδα. Η αναζήτηση μπορεί να επεκταθεί και σε περισσότερες ομάδες αν ληφθούν υπόψη και τα επόμενα αμέσως πλησιέστερα κέντρα. Όπως όμως έχει υπογραμμιστεί και προηγουμένως, μία τέτοια προσπάθεια, αν και δίνει καλύτερα αποτελέσματα κωδικοποίησης, κοστίζει σε χρόνο.

Είναι επόμενο, τα κέντρα των ομάδων να κατασκευάζονται με τρόπο τέτοιο, ώστε να προσεγγίζουν τη συνάρτηση πυκνότητας πιθανότητας της στοχαστικής πηγής που παράγει τα χαρακτηριστικά διανύσματα. Ο *Χάρτης Αυτό-Οργάνωσης (Self Organizing Map)* του *Kohonen (SOM)*<sup>15</sup> είναι ιδιαίτερα εύχρηστος για την εργασία αυτή καθώς παράγει σχετικά γρήγορα μία ομαδοποίηση υψηλής ποιότητας.

Εξαιτίας περιορισμών τόσο στο χώρο όσο και στο χρόνο αποθήκευσης, θα ήταν παράλογο η ομαδοποίηση και των  $8 \cdot 2N_D$  χαρακτηριστικών διανυσμάτων  $\pm\phi(C_{j,p})$ . Μία λύση στο πρόβλημα αυτό δίνεται από την ενσωμάτωση της μεθόδου του Fisher των κανονικών τάξεων στη διαδικασία ομαδοποίησης. Επομένως, όπως προτείνει ο Fisher χρησιμοποιούνται μόνο οι ισομετρίες  $\tau_R^{-1}\tau_{D_j}$  και  $\tau_{-R}^{-1}\tau_{D_j}$ . Στην πρώτη περίπτωση, οι καλύτερες υποψήφιες τύπου D περιοχές για την αντίστοιχη τύπου R περιοχή R καθορίζονται από την εύρεση των πιο κοντινών γειτόνων του  $\phi(\tau_R R)$  στο σύνολο των χαρακτηριστικών διανυσμάτων  $\pm\phi(C_j), j=1 \dots n_D$ .

Αλλά καθώς τα  $\tau_R R$  και  $\tau_{D_j} D_j$  έχουν και τα δύο κανονικό προσανατολισμό, ο παράγοντας της αντίθεσης είναι πιθανότατα μη αρνητικός. Επομένως, με βάση και την επομένη πρόταση, τα χαρακτηριστικά διανύσματα  $-\phi(C_j)$  μπορούν να αγνοηθούν.

### Πρόταση 1 :

Με βάση τα αποτελέσματα του Θεωρήματος 6 έχουμε  $\Delta(R, C_{j,p}) = \|\phi(R) - \phi(C_{j,p})\|$  αν και μόνο αν  $s \geq 0$  όπου ο  $s$  είναι ο παράγοντας αντίθεσης που υπολογίζεται από τη βελτιστοποίηση με βάση τη μέθοδο των ελαχίστων τετραγώνων.

**Απόδειξη :** Γνωρίζουμε ότι  $\Delta(R, C_{j,p}) = \|\phi(R) - \phi(C_{j,p})\|$  αν και μόνο αν  $\langle \phi(R), \phi(C_{j,p}) \rangle \geq 0$  ή ισοδύναμα  $\langle R - m_R \mathbf{1}, C_{j,p} - m_{C_{j,p}} \mathbf{1} \rangle \geq 0$  που είναι ισοδύναμο με  $s \geq 0$  με βάση τη σχέση που δίνει το  $s$  από την εφαρμογή της μεθόδου ελαχίστων τετραγώνων (σχέση (4.1))  $\square$ .

Στη δεύτερη περίπτωση (αναμονή αρνητικής τιμής αντίθεσης), με ένα παρόμοιο επιχείρημα βρίσκονται οι καλύτερες υποψήφιες περιοχές τύπου D, αναζητώντας τους πιο κοντινούς γείτονες του  $\phi(\tau_{-R}(-R))$  στο σύνολο των

χαρακτηριστικών διανυσμάτων  $\{\phi(C_j), j=1\dots n_D\}$ . Ο αλγόριθμος κωδικοποίησης επομένως έχει ως εξής :

Έστω  $q$  ένα βάθος αναζήτησης.

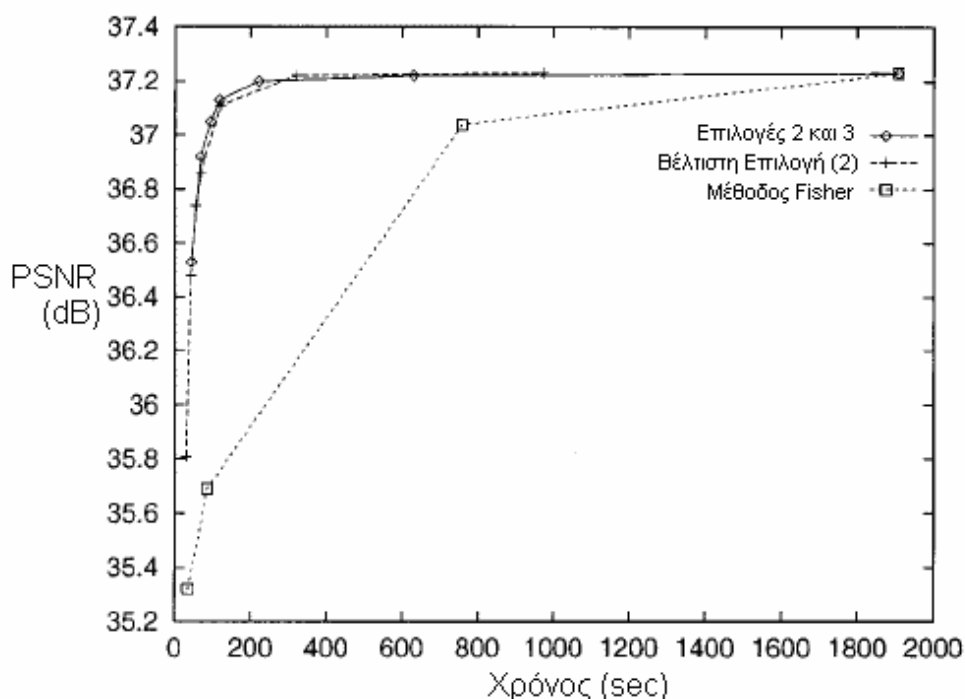
1. Κατασκευάζεται το σύνολο των κέντρων  $\{m_1, \dots, m_{n_m}\}$  χρησιμοποιώντας ως διανύσματα εκπαίδευσης τα  $\phi(C_j)$  και τον SOM αλγόριθμο.
2. Βρίσκεται η ομάδα του κάθε χαρακτηριστικού διανύσματος.

Για τη διαδικασία αυτή υπάρχουν 3 επιλογές :

1. Σύγκριση της τύπου R περιοχής  $R$  με κάθε περιοχή  $\tau_R^{-1}\tau_{D_j}D_j$  ώστε τα  $\phi(C_j)$  να ανήκουν στην ένωση των  $q$  ομάδων που τα κέντρα τους ανήκουν στους  $k$  πλησιέστερους γείτονες  $\Omega^k(\phi(\tau_R R))$  του  $\phi(\tau_R R)$ ,  $k=1$ .
2. Σύγκριση της τύπου R περιοχής  $R$  τόσο με τις τύπου D περιοχές που προκύπτουν από την 1<sup>η</sup> επιλογή όσο και με κάθε περιοχή  $\tau_{-R}^{-1}\tau_{D_j}D_j$  ώστε τα  $\phi(C_j)$  να ανήκουν στην ένωση των  $q$  ομάδων που τα κέντρα τους είναι  $\Omega^k(\phi(\tau_{-R}(-R)))$ ,  $k=1, \dots, q$ .
3. Καθορισμός το υποσύνολου των  $q$  πιο μικρών σφαλμάτων από το σύνολο  $\|\phi(\tau_R R) - m_k\|$  ,  $\|\phi(\tau_{-R}(-R)) - m_k\|$ . Για όλα τα  $k$  για τα οποία το  $\|\phi(\tau_R R) - m_k\|$  είναι το πιο μικρό ( ανήκει δηλαδή στο παραπάνω υποσύνολο) η  $R$  συγκρίνεται με όλες τις τύπου D περιοχές  $\tau_R^{-1}\tau_{D_j}D_j$  για τις οποίες τα διανύσματα  $\phi(C_j)$  ανήκουν στο  $\Omega^k$  . Ομοίως για όλα τα  $k$  για τα οποία το  $\|\phi(\tau_{-R}(-R)) - m_k\|$  ανήκει στο υποσύνολο η  $R$  συγκρίνεται με όλες τις τύπου D περιοχές  $\tau_{-R}^{-1}\tau_{D_j}D_j$  για τις οποίες τα διανύσματα  $\phi(-C_j)$  ανήκουν στο  $\Omega^k$ . Αυτή η τελευταία επιλογή εξετάζει τις μισές ομάδες από αυτές που εξετάζει η 2<sup>η</sup> . Επειδή εξετάζει το ελάχιστο σφάλμα και για τα  $\phi(\tau_R R)$  και  $\phi(\tau_{-R}(-R))$  είναι

πιο αργή από την 1<sup>η</sup> επιλογή αλλά δίνει μικρότερο σφάλμα προσέγγισης.

Το σχήμα 5.7 δείχνει το μέγιστο σηματοθορυβικό λόγο PSNR και το χρόνο κωδικοποίησης ως συνάρτηση του αριθμού των τάξεων (αντίστοιχα ομάδων) που εξετάζονται για την εικόνα της Lena 512\*512 εικονοστοιχείων. Οι καμπύλες αντιστοιχούν στις επιλογές 2 και 3 και στην ταξινόμηση με βάση τη μέση τιμή και την απόκλιση του Fisher. Για μια δίκαιη σύγκριση με τον αλγόριθμο του Fisher χρησιμοποιήθηκαν 72 διανύσματα ως κέντρα μάζας των ομάδων, που σχεδιάστηκαν με βάση τον αλγόριθμο SOM. Η τεχνική επιτάχυνσης της κωδικοποίησης που χρησιμοποίησαν οι συγγραφείς ξεπέρασε καθαρά αυτή του Fisher. Για παράδειγμα η 2<sup>η</sup> επιλογή είχε μία ακρίβεια των 37.05 dB στα 95sec, ενώ στην περίπτωση της ταξινόμησης του Fisher για ακρίβεια 37.04dB απαιτήθηκαν 758sec.



Σχ.5.7: Ο μέγιστος σηματοθορυβικός λόγος σε σχέση με τη διάρκεια κωδικοποίησης για την εικόνα της Lena. Οι τελείες προκύπτουν από το διαφορετικό αριθμό τάξεων (1, 3, 24, 72) ή ομάδων (1, 2, 3, 4, 8, 24, 72) που αναζητούνται.

Όταν η διαίρεση της εικόνας βασίζεται στην τεχνική του τετραδικού δένδρου, μπορούν να χρησιμοποιηθούν μέχρι και 4 διαφορετικά μεγέθη range

περιοχών (π.χ.  $4*4$ ,  $8*8$ ,  $16*16$ ,  $32*32$ ) με αποτέλεσμα τα χαρακτηριστικά διανύσματα να έχουν διάσταση έως και  $1024$ . Αυτό το γεγονός μπορεί να προκαλέσει πρόβλημα τόσο στον υπολογισμό όσο και στην αποθήκευση των αποτελεσμάτων, ειδικότερα αν η σχεδίαση των κέντρων μάζας καθώς και η ομαδοποίηση γίνουν με διανύσματα υψηλών διαστάσεων. Για να αντιμετωπιστεί η παραπάνω δυσκολία, ελαττώνεται πρώτα το μέγεθος των περιοχών σε  $4*4$  και στη συνέχεια υπολογίζονται τόσο τα χαρακτηριστικά διανύσματα όσο και τα κέντρα των ομάδων στο χώρο  $\mathfrak{R}^{16}$ . Με τον τρόπο αυτό προκύπτει μόνο ένα σύνολο από κέντρα στη χαμηλότερη διάσταση, τα οποία χρησιμοποιούνται στην ταξινόμηση των περιοχών όλων των διαστάσεων. Παρ' όλα αυτά, από τη στιγμή που θα έχουν επιλεγεί οι υποψήφιοι τύπου D περιοχές με τις ισομετρίες τους για σύγκριση με τη δεδομένη τύπου R περιοχή, το σφάλμα ελαχίστων τετραγώνων, υπολογίζεται στην πραγματική διάσταση των περιοχών. Ακριβέστερα, ως υποθεθεί ότι ένα σύνολο των κέντρων των ομάδων σχεδιάστηκε στον  $\mathfrak{R}^{16}$ . Έστω  $n_{D^n}$  ο αριθμός των τύπου D περιοχών  $D_j^n$  διαστάσεων  $2^{n+1} * 2^{n+1}$ . Πριν αρχίσει η επεξεργασία ακολουθούνται τα παρακάτω βήματα: Για όλα τα  $j = 1, \dots, n_{D^n}$ ,  $n = 2, 3, \dots, n_{\max}$ :

1. Υπολογίζονται τα διανύσματα διάστασης 16

$$C_j^n = S^{(3)} \dots S^{(n)} S^{(n+1)} (\tau_{D_j^n} D_j^n) \quad (5.43)$$

2. Ομαδοποιούνται τα χαρακτηριστικά διανύσματα  $\phi(C_j^n) \in \mathfrak{R}^{16}$ . Στη συνέχεια για βάθος αναζήτησης  $q$  και την  $1^{\eta}$  επιλογή (ο τρόπος εργασίας είναι παρόμοιος και για τις άλλες 2), η τύπου R περιοχή R μεγέθους  $2^n * 2^n$  κωδικοποιείται ως εξής:

- a) Υπολογισμός του διανύσματος, έστω

$$r^n = S^{(3)} \dots S^{(n-1)} S^{(n)} (\tau_{R^n} R^n), n \in 3 \dots n_{\max}.$$

- b) Υπολογισμός του αντίστοιχου χαρακτηριστικού  $\phi(r^n)$ .

c) Σύγκριση του  $R^n$  με κάθε περιοχή  $\tau_{R^n}^{-1} \tau_{D_j^n} D_j^n$  ώστε το  $\phi(C_j^n)$  να ανήκει στην ένωση των  $q$  ομάδων που τα κέντρα τους είναι  $\Omega^k(\phi(r^n)), k=1, \dots, q$ .

Τα χαμηλότερης διάστασης διανύσματα που θα αποτελέσουν τα κέντρα των ομάδων μπορούν να σχεδιαστούν με βάση τον SOM αλγόριθμο αν για παράδειγμα ως σύνολο των χαρακτηριστικών διανυσμάτων, θεωρηθεί το  $\{\phi(C_j^n); n=2, \dots, n_{\max}, j=1, \dots, n_{D^n}\}$ . Το σιεπτικό του παραπάνω αλγορίθμου στηρίζεται στην πεποίθηση ότι αν η χαμηλής διάστασης εκδοχή μίας τύπου D περιοχής δε μπορεί να αποτελέσει μία καλή προσέγγιση για τη δεδομένη τύπου R περιοχή χαμηλής διάστασης, τότε το ίδιο θα συμβαίνει και για τις περιοχές στην πραγματική τους διάσταση. Η πεποιήθηση αυτή ενισχύεται από το ακόλουθο θεώρημα που απέδειξε ο *Dekking*.

### Θεώρημα 7 :

Έστω  $R$  και  $D$  δύο περιοχές της εικόνας διαστάσεων  $2^{n+1} * 2^{n+1}$  και  $2^n * 2^n$  αντίστοιχα. Τότε:

$$\min_{s, o \in \mathbb{R}} \left( \frac{1}{2^n} \right) \|R - (sS^{(n+1)}D + o1)\| \geq \min_{s, o \in \mathbb{R}} \left( \frac{1}{2^{n-1}} \right) \|S^{(n)}R - (sS^{(n)}S^{(n+1)}D + o1)\| \quad (5.44)$$

Πίνακας 5.1: Το PSNR σε σχέση με τον αριθμό των ομάδων. Κωδικοποίηση με βάση την τεχνική ελάττωσης της διάστασης και τη  $2^n$  επιλογή για την εικόνα με τις πιερίες για ομοιόμορφη διαμέριση.

Αριθμός ομάδων	Μέγεθος περιοχής		
	8*8	16*16	32*32
1	31.04	26.46	22.31
2	31.31	26.61	22.48
3	31.38	26.73	22.54
4	31.42	26.77	22.57
8	31.46	26.82	22.62
72	31.48	26.83	22.61



Ο πίνακας 5.1 δείχνει ότι η αναζήτηση σε λίγες ομάδες με τη χρήση του αλγορίθμου ελάττωσης της διάστασης δίνει τιμές του PSNR που προσεγγίζουν εκείνες που αποκτώνται από την εξέταση και των 72 ομάδων. Στα πειράματα η απόσταση μεταξύ των τύπου D περιοχών ήταν *δεικνοστοιχεία*. Σύμφωνα με τους συγγραφείς, είναι εύλογο το γεγονός ότι αφού τα κέντρα των ομάδων υπολογίζονται με ένα αλγόριθμο Διανυσματικής Κβαντοποίησης, η ταξινόμηση που αναπτύχθηκε στην υποενότητα αυτή θα είναι το ίδιο επιτυχής ακόμη και αν ως διανύσματα εκπαίδευσης, χρησιμοποιηθούν διανύσματα που ανήκουν σε σύνολο άλλων εικόνων, που δεν περιλαμβάνει την προς εξέταση εικόνα. Αυτή η προσέγγιση εξαλείφει τον επιπρόσθετο χρόνο που απαιτεί η κατασκευή των κέντρων με βάση προσαρμοστικό έλεγχο, έλεγχο δηλαδή που είναι βασισμένος στη μορφή που έχει η προς συμπίεση εικόνα.

Πίνακας 5.2: Διάρκεια κωδικοποίησης και PSNR σε σχέση με τον αριθμό των ομάδων. Η στήλη με το όνομα «Adap» συμβολίζει κέντρα προσαρμοσμένα στην ελάχιστη εικόνα, ενώ η στήλη με το όνομα «Fix» κέντρα ανεξάρτητα της εικόνας.

Αριθμός ομάδων που αναζητώνται	Χρόνος (sec)		PSNR (dB)	
	Προσ.	Σταθ.	Προσ.	Σταθ.
1	43	32	35.76	35.89
2	67	59	36.23	36.29
3	89	81	36.37	36.41
4	113	105	36.42	36.45

Ο πίνακας 5.2 επιβεβαιώνει ότι η ταξινόμηση με συγκεκριμένα κέντρα είναι μία εναλλακτική της ταξινόμησης με προσαρμοστικά κέντρα. Εδώ οι συγγραφείς έχουν κωδικοποιήσει την  $512*512$  εικόνα του πλοίου με τη 2<sup>η</sup> επιλογή, με μοναδική διάσταση των τύπου R περιοχών την  $4*4$  *εικονοστοιχεία* και 4096 μη επικαλυπτόμενες τύπου D περιοχές  $8*8$  *εικονοστοιχείων*. Τα 72 σταθερά κέντρα σχεδιάστηκαν με βάση τον SOM αλγόριθμο από ένα σύνολο 9 εικόνων που δεν περιελάμβανε την εικόνα του πλοίου.

### 5.4.5 Ο Υβριδικός Αλγόριθμος των Saure-Hamzaui

Ο αλγόριθμος που ακολουθεί αποτελεί συνέχεια της προηγούμενης μεθόδου και συνδυάζει τόσο τη Διανυσματική Κβαντοποίηση (VQ), όσο και τις συμβατικές fractal τεχνικές. Οι συγγραφείς παρουσιάζουν δύο εκδοχές του αλγορίθμου, οι οποίες οφείλονται στον τρόπο με τον οποίο διαμερίζεται η εικόνα.

#### 5.4.5.1 Ομοιόμορφη διαίρεση

Στην αρχή γίνεται η υπόθεση ότι η εικόνα διαιρείται σε τετραγωνικές περιοχές σταθερού μεγέθους. Χρησιμοποιώντας τον αλγόριθμο SOM, σχεδιάζεται ένα μικρό σύνολο από κέντρα μάζας που είναι σταθερά και προέρχονται από αρκετές εκπαιδευτικές εικόνες, όπως προαναφέρθηκε. Μπορεί να αποδειχθεί ότι τα κέντρα αυτά έχουν μηδενική μέση τιμή. Τα διανύσματα αυτά μετατρέπονται στη συνέχεια με κανονικοποίηση σε μοναδιαία διανύσματα  $\mathbf{m}_1, \dots, \mathbf{m}_{n_m}$ . Επομένως έχουμε  $\phi(\mathbf{m}_p) = \mathbf{m}_p$  για όλα τα  $p = 1, \dots, n_m$ . Αφού ο σχεδιασμός των κέντρων είναι ανεξάρτητος της εικόνας που συμπίεζεται, τα κέντρα μάζας που κατασκευάστηκαν με τον παραπάνω τρόπο είναι διαθέσιμα και κατά την αποκωδικοποίηση.

Έστω  $R$ , μία υποψήφια για κωδικοποίηση τύπου  $R$  περιοχή διαστάσεων  $2^n * 2^n$ . Για λόγους απλότητας, θα αναλυθεί μόνο η αναζήτηση σε μία ομάδα με βάση την  $1^n$  επιλογή (μία από τις επιλογές αναζήτησης οι οποίες αναφέρθηκαν στην αμέσως προηγούμενη υποενότητα). Σε ένα πρώτο βήμα, καθορίζεται η ομάδα κάθε χαρακτηριστικού διανύσματος  $\varphi(C_j)$  που όπως προαναφέρθηκε, αντιστοιχεί στο διάνυσμα του συνόλου των τύπου  $D$  περιοχών  $S^{(n+1)}(\tau_{D_j} D_j), (j \in 1, \dots, n_D)$ . Στη συνέχεια, καθορίζεται το κέντρο της κάθε ομάδας  $m_c = \Omega^1 \{\varphi(\tau_R R)\}$ . Ακολουθεί η προσπάθεια προσέγγισης της τύπου  $R$  αυτής περιοχής από το παραπάνω κέντρο με τον υπολογισμό του σφάλματος ελαχίστων τετραγώνων:

$$E(R, m_c) = \min_{s, o \in \mathbb{R}^{*n}} \|\tau_R R - (s m_c + o I)\|_2^2 \quad (5.45)$$

Με βάση το Θεώρημα 6, το  $m_c$  είναι το κέντρο που μπορεί να προσεγγίσει καλύτερα το  $(\tau_R R)$ , ως προς το σφάλμα με τη μορφή ελαχίστων τετραγώνων. Οι τιμές  $s$  και  $o$  που προκύπτουν από τη λύση, διακριτοποιούνται δίνοντας μία αντίθεση  $\hat{s}(R)$ , μία φωτεινότητα  $\hat{o}(R)$  και ένα σφάλμα:

$$E(R, \hat{s}(R), \hat{o}(R), m_c, \tau_R^{-1}) = \|R - (\hat{s}(R) m_c + \hat{o}(R) I)\|_2^2 \quad (5.46)$$

Το επόμενο βήμα είναι η σύγκριση αυτού του rms σφάλματος που δίνει η παραπάνω προσέγγιση με την τιμή ενός δεδομένου κατωφλίου  $a$ . Αυτό μαθηματικά μεταφράζεται ως εξής:

$$\frac{1}{2^n} E(R, \hat{s}(R), \hat{o}(R), m_c, \tau_R^{-1}) \leq \alpha \quad (5.47)$$

Αν η ανισότητα ικανοποιείται, οι υπολοίπες τύπου D περιοχές δεν εξετάζονται και η περιοχή τύπου R κωδικοποιείται από το κέντρο, δηλαδή

$$R = \hat{s}(R) \tau_R^{-1} m_c + \hat{o}(R) I \quad (5.48)$$

Αν η ανισότητα δεν ικανοποιείται, τότε συγκρίνεται η προσέγγιση που δίνει το κέντρο με αυτές που δίνουν οι υπόλοιπες υποψήφιες τύπου D περιοχές που ανήκουν στο σύνολο που σχετίζεται με το κέντρο  $m_c$ . Αυτό γίνεται εξετάζοντας αν ισχύει η ανισότητα:

$$E(R, \hat{s}(R), \hat{o}(R), m_c, \tau_R^{-1})^{\frac{1}{2}} \leq (1 + \varepsilon) E(R, s(R), o(R), D(R), \tau_R^{-1} \tau_D(R))^{\frac{1}{2}} \quad (5.49)$$

Εδώ  $\varepsilon$  είναι ένα προαναφερθέν μικρό περιθώριο σφάλματος και το  $D(R)$  είναι η τύπου D περιοχή που δίνει το μικρότερο σφάλμα προσέγγισης ανάμεσα σε όλα τις τύπου D περιοχές που τα χαρακτηριστικά τους διανύσματα  $\varphi(C_j)$  ανήκουν στην ομάδα με κέντρο  $m_c$ . Όταν ικανοποιείται η τελευταία ανισότητα, τότε το συμπέρασμα που εξάγεται είναι ότι η προσέγγιση που δίνει η βέλτιστη τύπου D περιοχή δεν είναι τόσο ανώτερη αυτής που προέρχεται από το κέντρο και συνεπώς και σε αυτήν την περίπτωση είναι προτιμότερο να κωδικοποιείται η  $R$  με το  $m_c$ . Μία τέτοια κωδικοποίηση είναι συμφέρουσα για

δύο λόγους: Πρώτον, αφού τα κέντρα είναι συνήθως πολύ λιγότερα από τις τύπου  $D$  περιοχές, η πιθανή απώλεια σε ποιότητα αντισταθμίζεται από τη μείωση του bit rate. Δεύτερον, στην κωδικοποίηση με fractal, το σφάλμα αποκωδικοποίησης είναι στις περισσότερες των περιπτώσεων, όπως έχουν δείξει και τα πειράματα, μεγαλύτερο του σφάλματος προσέγγισης κατά την κωδικοποίηση. Αυτό δε συμβαίνει όταν οι τύπου  $R$  περιοχές ανακατασκευάζονται από τα διαθέσιμα κέντρα των ομάδων.

Αν καμία από τις δύο προηγούμενες ανισότητες δεν ικανοποιείται, τότε η τύπου  $R$  περιοχή  $R$  κωδικοποιείται από τη τύπου  $D$  περιοχή  $D(R)$  όπως στην περίπτωση της συμβατικής συμπίεσης με τη μέθοδο fractal. Με τον τρόπο αυτό, η κωδικοποίηση του  $R$  καθορίζεται από :

1. Ένα ψηφίο που έχει τιμή 0 ή 1, ανάλογα με το αν χρησιμοποιείται το κέντρο ή μια τύπου  $D$  περιοχή για την κωδικοποίηση.
2. Το δείκτη  $c$  του κέντρου της ομάδας ή της θέσης της τύπου  $D$  περιοχής  $D(R)$ .
3. Την αντίθεση και τη φωτεινότητα.

Αν η αντίθεση  $s$  είναι μηδέν, τότε δε χρειάζονται ούτε το ψηφίο που φανερώνει τη μέθοδο κωδικοποίησης ούτε η θέση και ούτε η ισομετρία. Η παραπάνω αναζήτηση μπορεί να επεκταθεί και σε γειτονικές ομάδες. Λαμβάνοντας υπόψη και το  $\tau_{-R}(-R)$  οι επιλογές 2 και 3 μπορούν επίσης να χρησιμοποιηθούν, δίνοντας αρνητικές τιμές για την αντίθεση.

Η αποκωδικοποίηση στον υβριδικό αυτό αλγόριθμο γίνεται επαναληπτικά, αρχίζοντας από μία αυθαίρετη εικόνα όπως και στους συμβατικούς αλγορίθμους. Εύκολα διαπιστώνεται, ότι αν οι τύπου  $R$  περιοχές που κωδικοποιούνται από περιοχές τύπου  $D$ , έχουν τιμές αντίθεσης μικρότερες απολύτως από τη μονάδα, η αποκωδικοποίηση συγκλίνει ακόμα και όταν οι τιμές της αντίθεσης που αντιστοιχούν στις τύπου  $R$  περιοχές που κωδικοποιούνται από τα κέντρα, είναι μεγαλύτερες απολύτως της μονάδας.

Τα πλεονεκτήματα του αλγορίθμου αυτού σε σχέση με τους αλγορίθμους που βασίζονται στην απόσταση είναι τα ακόλουθα :

1. Η κωδικοποίηση είναι ταχύτερη, αφού η αναζήτηση για μία κατάλληλη τύπου D περιοχή αρχίζει μόνο όταν το κέντρο δεν παρέχει μια ικανοποιητική προσέγγιση της περιοχής τύπου R.
2. Ο βαθμός συμπίεσης μπορεί να βελτιωθεί από την επιλογή της αναλογίας μεταξύ του αριθμού των ομάδων και τον αριθμό των τύπου D περιοχών. Αν ο αριθμός των περιοχών τύπου R είναι  $n_R$  και ο αριθμός των περιοχών που κωδικοποιούνται με βάση τη Διανυσματική Κβαντοποίηση είναι  $n_1$ , τότε ο υβριδικός αλγόριθμος χρησιμοποιεί λιγότερα ψηφία από το συμβατικό fractal αλγόριθμο αν

$$n_1 > \frac{n_R}{p-k}, p = \log_2 n_D, k = \log_2 n_m \quad (5.50)$$

Στην εξαγωγή της ανισότητας χρησιμοποιείται και ένα ψηφίο για κάθε τύπου R περιοχή για τον καθορισμό του τρόπου κωδικοποίησης (VQ ή fractal).

3. Η αποκωδικοποίηση είναι επίσης ταχύτερη. Πράγματι, δεν είναι απαραίτητος ο υπολογισμός των εντάσεων των *εικονοστοιχείων* των τύπου R περιοχών που κωδικοποιούνται από τα κέντρα σε κάθε επανάληψη, καθώς αυτές είναι αμετάβλητες μετά την 1<sup>η</sup> επανάληψη.

Στη συνέχεια συγκρίνεται η επίδοση του κωδικοποιητή που βασίζεται στην απόσταση με αυτή του υβριδικού αλγορίθμου. Με χρήση του SOM σχεδιάζεται ένα σύνολο από 256 σταθερά κέντρα που αντιστοιχούν στους κόμβους ενός πίνακα διαστάσεων 16\*16. Τα εκπαιδευτικά διανύσματα προέρχονται από 9 εικόνες μεγέθους 512\*512 και για την κωδικοποίηση χρησιμοποιήθηκε η 2<sup>η</sup> Επιλογή. Επιλέχθηκε αναζήτηση 4 ομάδων που μας δίνει μία αποδεκτή ισορροπία μεταξύ ταχύτητας και αξιοπιστίας. Ο πίνακας 5.3 δίνει τα αποτελέσματα κωδικοποίησης για την 512\*512 εικόνα της «βάριας». Συγκριτικά ο κωδικοποιητής fractal που βασίζεται στην απόσταση

χρειάστηκε 57sec, για την κωδικοποίηση είχε μια αναλογία συμπίεσης 4.74 : 1 και έδωσε PSNR 36.29 dB. Παρατηρείται ότι μια πλήρης αναζήτηση (στην περίπτωση αυτή 256 ομάδων) θα αύξανε το PSNR του αυτοτελούς fractal αλγορίθμου στα 36.62 dB. Από την άλλη πλευρά, η κωδικοποίηση των τύπου R περιοχών αποκλειστικά με τα κέντρα των ομάδων έδωσε PSNR 35.81dB. Ο πίνακας 5.4 συγκρίνει τη σύγκλιση της αποκωδικοποίησης των δύο αλγορίθμων όταν η αρχική εικόνα είναι μαύρη. Η σύγκλιση επήλθε στις 5 επαναλήψεις για τον υβριδικό αλγόριθμο και μετά από 9 για τον αυτούσιο Fractal αλγόριθμο.

Πίνακας 5.3: Η επίδοση του υβριδικού αλγορίθμου για διάφορες τιμές των  $a$  και  $\varepsilon$  για την εικόνα της «βάρακα». Στον πίνακα, το  $n_1$  συμβολίζει τον αριθμό των range περιοχών που κωδικοποιήθηκαν με βάση τη Διανυσματική Κβαντοποίηση.

$\varepsilon$	$a$	$n_1$	Βαθμός συμπίεσης	PSNR (dB)	Χρόνος (sec)
0.1	2	9075	5.74	36.72	49
	3	12662	6.31	36.54	39
0.15	2	9947	5.80	36.64	49
	3	13313	6.37	36.47	39
0.2	2	10786	5.85	36.53	49
	3	13937	6.43	36.38	39
0.25	2	11523	5.91	36.44	49
	3	14441	6.49	36.29	39

Τα πειράματα επιβεβαιώνουν ότι για μία ευρεία περιοχή κατωφλίων  $a$  και  $\varepsilon$ , ο υβριδικός αλγόριθμος ξεπερνά σε απόδοση τον αυτούσιο fractal αλγόριθμο. Ο ρυθμός συμπίεσης όπως και το PSNR είναι υψηλότερα και τόσο η κωδικοποίηση όσο και η αποκωδικοποίηση είναι ταχύτερες.

Πίνακας 5.4: Η σύγκλιση της αποκωδικοποίησης για την εικόνα της «βάρακας». Για τον υβριδικό αλγόριθμο  $a=3$  και  $\varepsilon=0.15$ . Οι στήλες δίνουν τον αριθμό των επαναλήψεων και το μέγιστο σηματοθορυβικό λόγο σε dB.

Αριθμός επαναλήψεων	Υβριδικός αλγόριθμος	Απλός fractal αλγόριθμος
k	PSNR (dB)	PSNR (dB)
1	22.46	11.56
3	35.48	21.48
5	36.47	35.20
9	-	36.29

#### 5.4.5.2 Διαίρεση με βάση το τετραδικό δένδρο

Στο εδάφιο αυτό, αναλύεται ο τρόπος με τον οποίο συνδέεται ο υβριδικός αλγόριθμος με τον αλγόριθμο του τετραδικού δένδρου του Fisher.

Πρώτα από όλα, σχεδιάζονται επιπρόσθετα σύνολα από κέντρα για διαφορετικά μεγέθη περιοχών τύπου R, ώστε τα κέντρα διαφορετικού μεγέθους να είναι ανεξάρτητα μεταξύ τους. Επομένως για μέγεθος περιοχής ίσο με  $2^n * 2^n$ ,  $n \in \{2, \dots, n_{\max}\}$  υπάρχουν  $n_m$  κέντρα (ο αριθμός των κέντρων δεν είναι απαραίτητα ο ίδιος για κάθε μέγεθος)  $m_1, m_2, \dots, m_{n_m}$ , κάθε ένα από τα οποία έχει μηδενική μέση τιμή και διασπορά 1 όπως και προηγουμένως. Για την επιταχύνση της ομαδοποίησης υποδειγματοληπτώνται όλες οι τύπου R περιοχές μεγάλης διάστασης ( $n > 2$ ) και οι αντίστοιχες περιοχές τύπου D στη χαμηλότερη διάσταση. Συνεπώς, τα χαρακτηριστικά διανύσματα των τύπου R και τύπου D περιοχών έχουν διάσταση  $2^2 * 2^2 = 16$ .

Ο πυρήνας του αλγορίθμου αυτού για μία αναζήτηση σε βάθος  $q$  ομάδων έχει ως εξής: Για απλότητα, αναφέρεται μόνο η 1<sup>η</sup> επιλογή, το ίδιο όμως ισχύει με κάποιες παραλλαγές και για τις 2 και 3. Έστω  $R^n$  μια τετραγωνική περιοχή μεγέθους  $2^n * 2^n$ ,  $n \in \{3, \dots, n_{\max}\}$ . Η διαδικασία αρχίζει θέτοντας  $n = n_{\max}$ . Για κάποιες δεδομένες τιμές κατωφλίου  $\alpha$  και  $t$ , αν:

$$\frac{1}{2^n} d(R^n, \hat{s}(R^n), \hat{o}(R^n), m_c^n, \tau_{R^n}^{-1}) \leq \alpha \quad (5.51)$$

τότε το  $R^n$  κωδικοποιείται από το κέντρο της ομάδας  $m_c^n$ . Εδώ το  $m_c^n$  είναι το κέντρο διάστασης  $2^n * 2^n$  που δίνει ελάχιστο rms σφάλμα ανάμεσα στα  $q$  κέντρα που τα χαρακτηριστικά τους διανύσματα  $\varphi(S^{(3)}S^{(4)}, \dots, S^{(n)}m_{p_i}^n)$ ,  $i=1, \dots, q$  είναι οι  $q$  πλησιέστεροι γείτονες του χαρακτηριστικού διανύσματος  $\varphi(S^{(3)}, \dots, S^{(n)}\tau_{R^n}R^n)$ . Αν η πιο πάνω συνθήκη δεν ικανοποιείται, τότε ορίζοντας το  $D(R^n)$  ως την καλύτερη τύπου D περιοχή αναμέσα στις περιοχές τύπου D που τα χαρακτηριστικά τους διανύσματα ανήκουν στις  $q$  επιλεγθείσες ομάδες, αν:

$$\frac{1}{2^n} E(R^n, s(R^n), o(R^n), D(R^n), \tau_{R^n}^{-1} \tau_D(R^n))^{\frac{1}{2}} > t \quad (5.52)$$

τότε το  $R^n$  δε γίνεται αποδεικτό και τεμαχίζεται σε τέσσερις μικρότερες τετραγωνικές περιοχές. Στην άλλη περίπτωση και αν

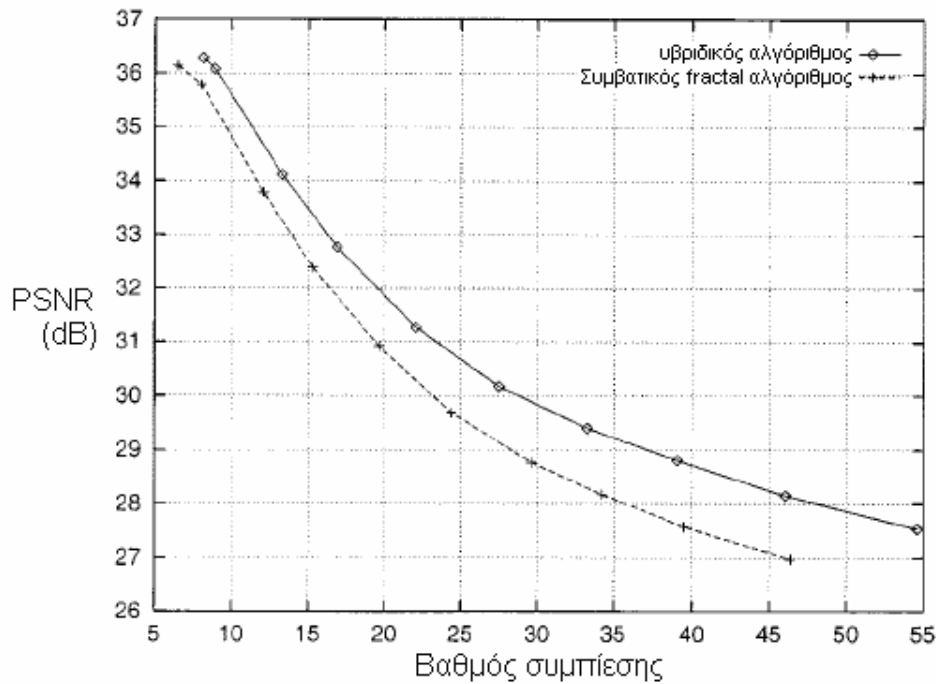
$$\begin{aligned} E(R^n, \hat{s}(R^n), \hat{o}(R^n), m_c, \tau_{R^n}^{-1})^{\frac{1}{2}} &\leq \\ &\leq (1 + \varepsilon) E(R^n, s(R^n), o(R^n), D(R^n), \tau_{R^n}^{-1} \tau_D(R^n))^{\frac{1}{2}} \end{aligned} \quad (5.53)$$

τότε η περιοχή  $R^n$  κωδικοποιείται με το κέντρο  $m_c^n$  της ομάδας διαφορετικά κωδικοποιείται με την τύπου D περιοχή  $D(R^n)$ . Όταν η διαδικασία φθάσει στη χαμηλότερη διάσταση, τότε η κωδικοποίηση προχωρά όπως αναφέρθηκε στην παράγραφο **5.4.5.1**.

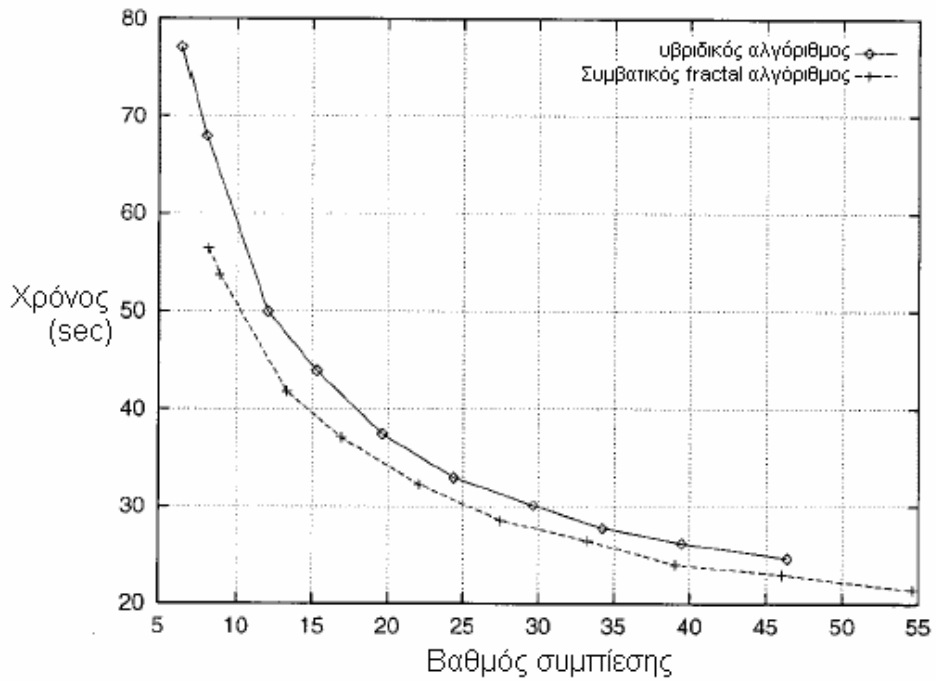
Το σχήμα 5.8 δείχνει την επίδραση του PSNR προς το βαθμό συμπίεσης του fractal αλγορίθμου που βασίζεται στην απόσταση και του υβριδικού αλγορίθμου μίας διαμέρισης τετραδικού δένδρου τριών επιπέδων  $n_{\max} = 4$ . Για το fractal αλγόριθμο χρησιμοποιήθηκε ένα μόνο σύνολο από 250 σταθερά κέντρα διάστασης 16. Για τον υβριδικό αλγόριθμο χρησιμοποιήθηκαν 3 σύνολα των 256 σταθερών κέντρων. Για κάθε μέγεθος τύπου R περιοχής, η περίοδος δειγματοληψίας για τη δημιουργία του συνόλου των περιοχών τύπου D, είναι 8 *εικονοστοιχεία*. Και για τους δύο αλγορίθμους που εξετάζονται, χρησιμοποιήθηκε η  $2^n$  επιλογή και μία αναζήτηση 4 ομάδων. Οι παράμετροι  $a$



και  $t$  ορίστηκαν ίσες με 3 και 0.15 αντίστοιχα. Ο βαθμός συμπίεσης διαφοροποιείται, δίνοντας στο rms κατώφλι  $t$  τιμές μεταξύ 22 και 1. Το σχήμα 5.9 δείχνει το χρόνο ως συνάρτηση του βαθμού συμπίεσης για την ίδια ακολουθία δοκιμών. Το σχήμα 5.10 συγκρίνει την ποιότητα των αποκωδικοποιημένων εικόνων για τον ίδιο βαθμό συμπίεσης.



Σχ. 5.8: Ο μέγιστος σηματοθορυβικός λόγος σε σχέση με το βαθμό συμπίεσης για την εικόνα της βάριας.



Σχ.5.9: Ο χρόνος σε σχέση με το βαθμό συμπίεσης για την εικόνα της βάρικας.



Σχ.5.10: A: Ο υβριδικός αλγόριθμος ( $PSNR=29.55dB$ , βαθμος συμπίεσης 32,06). B: Συμβατικός fractal αλγόριθμος ( $PSNR=28.48dB$ , βαθμος συμπίεσης 31,96)

## 5.5 Συναρτησιακές μέθοδοι

Οι *Belford, Dekking* και *Keane*<sup>16</sup>, πρότειναν τη χρησιμοποίηση ενός κριτηρίου που καθορίζει πότε ένα κωδικό διάνυσμα δεν μπορεί να αποτελέσει καλή προσέγγιση μιας τύπου  $R$  περιοχής. Η ιδέα τους είναι ότι δεν χρειάζεται να συγκρίνονται τύπου  $R$  με τύπου  $D$  περιοχές, αλλά ανεξάρτητα η μία κατηγορία από την άλλη με ένα συγκεκριμένο διάνυσμα μοναδιαίου μέτρου. Μόνο όταν οι συγκρίσεις αυτές φέρουν παραπλήσια αποτελέσματα μπορεί μια συγκεκριμένη περιοχή τύπου  $R$  να προσεγγιστεί από την αντίστοιχη τύπου  $D$ . Δίνεται λοιπόν με την ενέργεια αυτή η δυνατότητα μείωσης του υπολογιστικού κόστους με τον αποκλεισμό ενός μεγάλου αριθμού κωδικών διανυσμάτων. Εδώ δε θα δοθεί το αρχικό αποτέλεσμα της εργασίας των συγγραφέων, αλλά η γενίκευσή του για αυθαίρετα μοναδιαία διανύσματα καθώς επίσης και για την περίπτωση χρησιμοποίησης περισσότερων του ενός διανυσμάτων σταθερής βάσης.

### Θεώρημα 8:

Έστω  $\delta > 0$  και  $U$  ένα μοναδιαίο διάνυσμα στο χώρο  $R^n$ ,  $n \geq 2$ . Έστω ότι τα  $R$  και  $D$  ανήκουν στον  $R^n$  με  $\langle R, \phi(R) \rangle \geq \delta$ . Αν το

$$E(D, R) = \min_{a, b_1, \dots, b_p \in R} \left\| R - \left( aD + \sum_{k=1}^p b_k B_k \right) \right\| \leq \delta \quad (5.54)$$

τότε

$$\left| \langle \phi(R), U \rangle \right| - \left| \langle \phi(D), U \rangle \right| \leq \sqrt{2 - 2 \sqrt{1 - \frac{\delta^2}{\langle R, \phi(R) \rangle^2}}} \quad (5.55)$$

**Απόδειξη:** Χρησιμοποιώντας την ανισότητα των Cauchy-Schwarz:

$$\begin{aligned}
& \left( \left| \langle \phi(R), U \rangle \right| - \left| \langle \phi(D), U \rangle \right| \right)^2 \leq \left( \langle \phi(R), U \rangle - \langle \phi(D), U \rangle \right)^2 = \\
& = \left| \langle \phi(R) - \phi(D), U \rangle \right|^2 \leq \|\phi(R) - \phi(D)\|^2 \|U\|^2 = \\
& = \|\phi(R) - \phi(D)\|^2 = 2 - 2\langle \phi(R), \phi(D) \rangle
\end{aligned} \tag{5.56}$$

Αν το  $\rho$  συμβολίζει το  $\langle \phi(R), \phi(D) \rangle$ , τότε το τετράγωνο του σφάλματος είναι σύμφωνα με το Θεώρημα 6,  $E^2 = \langle R, \phi(R) \rangle^2 (1 - \rho^2)$ . Αφού όμως  $E < \delta$ , τότε και  $\langle R, \phi(R) \rangle^2 (1 - \rho^2) \leq \delta^2$ . Υποθέτοντας ότι  $\rho \geq 0$  (σε διαφορετική περίπτωση αντικαθιστούμε το D με το -D), λύνουμε την παραπάνω ανισότητα ως προς

$$\rho: \rho \geq \sqrt{1 - \frac{\delta^2}{\langle R, \phi(R) \rangle^2}}. \text{ Εισάγοντας την τελευταία ανισότητα στη σχέση (5.56)}$$

και παίρνοντας την τετραγωνική ρίζα κάθε μέλους οδηγούμαστε στην προς απόδειξη σχέση. Είναι εύκολο να παρατηρήσουμε ότι στην περίπτωση που  $\rho = 1$ , όταν δηλαδή υπάρχει ένα μόνο διάνυσμα σταθερής έντασης ( $B = (1, 1, \dots, 1) / \sqrt{n}$ ), ότι για οποιοδήποτε διάνυσμα  $Z$  που ανήκει στο χώρο  $R^n$  ισχύει:

$$\phi(Z) = \frac{1}{\sqrt{V(Z)}} (z_1 - m_z, \dots, z_n - m_z) \tag{5.57}$$

όπου το  $\mathbf{m}_z = (z_1 + \dots + z_n) / n$  είναι ο μέσος όρος της διαβάθμισης του γκρι και  $V(Z) = \sum_{k=1}^n (z_k - \mathbf{m}_z)^2$  η διασπορά. Αυτή είναι και η ειδική περίπτωση που εξετάζουν στην εργασία τους οι συγγραφείς, όπου η συνθήκη  $\langle R, \phi(R) \rangle \geq \delta$ , δίνεται ως  $V(R) \geq \delta^2 \square$ .

Σύμφωνα με το παραπάνω θεώρημα, ο αλγόριθμος για την κωδικοποίηση μιας τύπου R περιοχής έχει ως εξής:

#### Αλγόριθμος 4 (Ένας συναρτησιακός αλγόριθμος- a functional algorithm)

1. Επιλέγεται μία τιμή ανοχής  $\delta$  και ένα μοναδιαίο διάνυσμα  $U$ .
2. (Προεπεξεργασία) Για κάθε κωδικό διάνυσμα  $D$  υπολογίζεται το  $|\langle \phi(D), U \rangle|$ .

Για κάθε τύπου  $\mathbb{R}$  περιοχή  $R$ , εκτελούνται τα ακόλουθα βήματα:

3. Υπολογίζεται το  $|\langle \phi(R), R \rangle|$  και το άνω όριο της σχέσης (5.55).
4. Αν το  $\langle \phi(R), R \rangle < \delta$ , τότε δεν χρειάζεται να γίνει αναζήτηση αφού ένας μηδενικός συντελεστής αντίθεσης δίνει σφάλμα ελαχίστων τετραγώνων ίσο με  $E = \langle \phi(R), R \rangle \leq \delta$  για οποιοδήποτε κωδικό διάνυσμα  $D$ .
5. Αν το  $\langle \phi(R), R \rangle \geq \delta$ , τότε υπολογίζεται το  $|\langle \phi(R), U \rangle|$  και απορρίπτονται όλα εκείνα τα κωδικά διανύσματα  $D$  για τα οποία η ανισότητα (5.55) του Θεωρήματος 8 δεν ικανοποιείται.

Η απόδοση της παραπάνω μεθόδου είναι δυνατό να αυξηθεί περισσότερο χρησιμοποιώντας αρκετά μοναδιαία διανύσματα ως κριτήρια του θεωρήματος. Με τον τρόπο αυτό, για μία δεδομένη τύπου  $\mathbb{R}$  περιοχή, ο αριθμός των τύπου  $D$  περιοχών που αποκλείονται αναμένεται να είναι μεγαλύτερος.

Μία πιο αποδοτική εφαρμογή της παραπάνω μεθόδου δεν εξετάζει ολόκληρο το σύνολο των περιοχών τύπου  $D$  για να τις υποβάλλει στη σύγκριση που επιβάλλει το θεώρημα. Αντίθετα, για μία οποιαδήποτε συναρτησιακή μέθοδο όπως η προηγούμενη, υιοθετείται συνήθως ο παρακάτω βελτιωμένος γενικός αλγόριθμος:

#### Αλγόριθμος 5 (Γενική συναρτησιακή μέθοδος-General functional method)

Υποθέτουμε ότι η συνάρτηση  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  είναι τέτοια ώστε  $|F(R) - F(D)| \leq \epsilon_R$ , γεγονός που υποδηλώνει ότι η range περιοχή  $R$  μπορεί

να προσεγγιστεί επιτυχώς από τη domain  $D$ . Έστω το σύνολο των domain περιοχών  $\{D_1, \dots, D_{N_D}\}$ . Σε ένα προπαρασκευαστικό βήμα εκτελούμε:

1. Τον υπολογισμό του  $F(D)$  για κάθε  $D \in \{D_1, \dots, D_{N_D}\}$ .
2. Ταξινόμηση όλων των domain περιοχών σύμφωνα με την τιμή της συνάρτησης σε ένα πίνακα γραμμή και επωνομασία τους με τέτοιο τρόπο ώστε  $F(D_1) \leq F(D_2) \leq \dots \leq F(D_{N_D})$ .

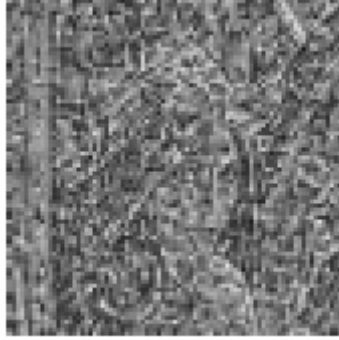
Για κάθε range περιοχή  $R$  εκτελούμε:

3. Τον υπολογισμό του  $F(R)$  και του ανωτάτου ορίου  $\in_R$  (π.χ. του δεξιού μέλους της (5.55) στο Θεώρημα 8)
4. Χρησιμοποιώντας μεθόδους όπως η δυαδική αναζήτηση για παράδειγμα, εύρεση των δεικτών  $k_0, k_1$  για τους οποίους ισχύει  $|F(R) - F(D_k)| \leq \in_R$  αν και μόνο αν  $k_0 \leq k \leq k_1$ .
5. Έλεγχος όλων των domain περιοχών  $D_k$  με  $k_0 \leq k \leq k_1$ .

Με τη διαδικασία αυτή παράγεται σε λογαριθμικό χρόνο  $O(\log N_D)$  μία λίστα υποψήφιων domain περιοχών, ενώ γνωρίζουμε ότι η πλήρης αναζήτηση στο σύνολο τους και η απόρριψη αυτών που δεν περνούν το συγκριτικό τεστ εκτελείται σε γραμμικό χρόνο.

### 5.5.1 Ταχεία Κωδικοποίηση Βασισμένη σε Προσαρμοστική Αναζήτηση

Στο πνεύμα της γενικής συναρτησιακής μεθόδου βρίσκεται και ο προσαρμοστικός αλγόριθμος για αποδοτική συμπίεση που προτείνουν η Tong και Pi (βιβλιογραφία 4<sup>ου</sup> κεφαλαίου).



Σχ.5.11: Η εικόνα της Lena με τις range περιοχές να έχουν αντικατασταθεί από τις domain που τις προσεγγίζουν καλύτερα.

Έστω το σχήμα 5.11 όπου κάθε τύπου R περιοχή της εικόνας της Lena έχει αντικατασταθεί από την τύπου D περιοχή (με συστολή διαστάσεων) που την προσεγγίζει καλύτερα χωρίς την εφαρμογή οπουδήποτε μετασχηματισμού. Αν και η εικόνα μοιάζει παραμορφωμένη, μπορούμε να διακρίνουμε μέσα της τη σιλουέτα της Lena. Αυτό το γεγονός υπενθυμίζει ότι η προσέγγιση μεταξύ δύο περιοχών εξαρτάται σε κάποιο βαθμό από την ομοιότητά τους. Εδώ οι συγγραφείς παρουσιάζουν ένα απλό και αποδοτικό μέτρο για την ομοιότητα δύο περιοχών συγκρίνοντας τις αντίστοιχες τυπικές αποκλίσεις τους. Η μαθηματική έκφραση του παραπάνω κριτηρίου θεμελιώνεται χρησιμοποιώντας τη γνωστή τριγωνική ανισότητα για το σφάλμα της σχέσης (4.6) για διακριτές τιμές αντίθεσης και φωτεινότητας  $\{s_i\}$  και  $\{o_j\}$ :

$$\begin{aligned}
 \sqrt{\widehat{E}(R)} &= \|R - s_i D - o_j \mathbf{1}\| = \|R - s_i D - (\bar{r}_j - s_i \bar{d} \mathbf{1})\| \\
 &= \|s_i (D - \bar{d} \mathbf{1}) - (R - \bar{r}_j \mathbf{1})\| \\
 &\geq |s_i| \|D - \bar{d} \mathbf{1}\| - \|R - \bar{r}_j \mathbf{1}\|
 \end{aligned} \tag{5.58}$$

(Υπενθυμίζεται ότι τα  $\bar{r}_j$  και  $\bar{d}$  συμβολίζουν τις μέσες τιμές των τύπου R και τύπου D περιοχών αντιστοίχως).

Η ποσότητα  $\|R - \bar{r}_j \mathbf{1}\| = \sqrt{\|R - \bar{r} \mathbf{1}\|^2 + k^2 (\bar{r} - \bar{r}_j)^2}$ , μπορεί να προσεγγιστεί από την ποσότητα  $\|R - \bar{r} \mathbf{1}\|$  καθώς το σφάλμα διακριτοποίησης της μέσης τιμής της περιοχής τύπου R είναι πολύ μικρό (διαφορετικά, το σφάλμα σύγκρισης

και ακολούθως η ανακατασκευή της εικόνας θα απείχε πολύ από την πραγματική). Επομένως διαιρώντας με τη διάσταση  $k$  του τύπου R διανύσματος, η συνθήκη γίνεται:

$$|std(R) - s_i std(D)| \leq RMS(R) = \sqrt{\frac{\widehat{E}(R)}{k^2}} \quad (5.59)$$

όπου το  $RMS(R)$  είναι η τετραγωνική ρίζα του σφάλματος σύγκρισης για κάθε περιοχή τύπου R και  $std()$  είναι τυπική απόκλιση των αντίστοιχων περιοχών.

Η παραπάνω ανισότητα παρέχει μία αναγκαία συνθήκη ώστε μια περιοχή τύπου D να αποτελεί πιθανή υποψήφια για προσέγγιση μίας τύπου R περιοχής. Επιλέγοντας ένα κατώφλι  $T_1$ , πέρα από το οποίο θεωρείται ότι το RMS σφάλμα είναι απαγορευτικό για καλή προσέγγιση, απορρίπτονται όλες εκείνες οι τύπου D περιοχές για τις οποίες ισχύει:

$$|std(R) - s_i std(D)| \geq T_1 \quad (5.60)$$

Ελαττώνοντας με τον τρόπο αυτό αρκετά τον υπολογιστικό φόρτο. Η συνθήκη αυτή εξαρτάται μόνο από την αντίθεση και την τυπική απόκλιση, γεγονός που την καθιστά αρκετά ικανοποιητική καθώς οι τυπικές αποκλίσεις όλων των περιοχών τύπου D υπολογίζονται και αποθηκεύονται μόνο μια φορά. Η ανισότητα γίνεται ακόμα πιο αποδοτική αν πάρει τη μορφή:

$$\left| \frac{std(R)}{s_i} - std(D) \right| \leq \frac{T_1}{s_i} \quad (5.61)$$

Μία σταθερή τιμή κατωφλίου  $T_1$ , όμως, έχει το έχει το εξής μειονέκτημα. Αν είναι πολύ υψηλή, η επιτάχυνση της διαδικασίας θα είναι μικρή, ενώ αν είναι πολύ χαμηλή τίθεται σε αμφισβήτηση η ποιότητα της προσέγγισης και κατά συνέπεια η πιστότητα της ανακατασκευασμένης εικόνας. Αυτό συμβαίνει γιατί η ανισότητα (5.59) εκφράζει μία αναγκαία, αλλά όχι ικανή συνθήκη.

Για την αύξηση λοιπόν της ταχύτητας, απαιτείται η εισαγωγή ενός προσαρμοστικού τρόπου καθορισμού της ανοχής ο οποίος να διατηρεί και την πιστότητα της ανακατασκευής.



Επιστρέφοντας στην εξίσωση (4.5), είναι εύκολο να παρατηρήσει κανείς ότι στη θεωρητική περίπτωση που οι παράμετροι του μετασχηματισμού είναι πραγματικοί αριθμοί, το ελάχιστο σφάλμα προσέγγισης εξαρτάται από δύο παράγοντες: την τυπική απόκλιση της τύπου R περιοχής και το συντελεστή συσχέτισης  $\sqrt{1-c^2}$  (το  $c$  ορίστηκε στην εξίσωση (4.4)), που εξαρτάται και ο ίδιος από την εικόνα. Αυτό σημαίνει ότι

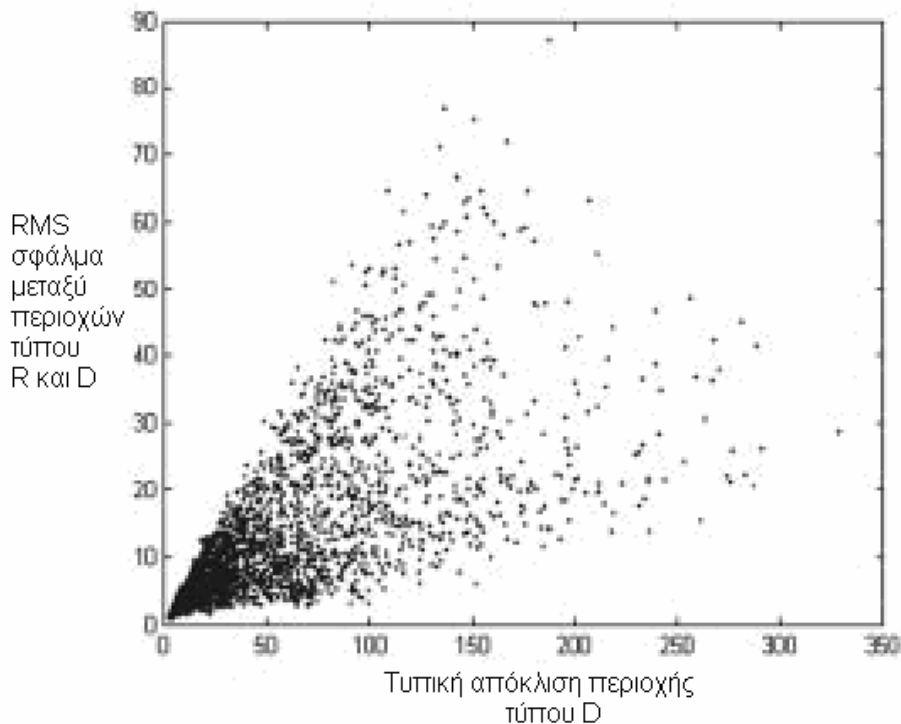
$$\sqrt{MSE(R)} \propto std(R) \quad (5.62)$$

Το παραπάνω συμπέρασμα συμφωνεί και με τη γενική παρατήρηση ότι περιοχές με μεγάλες τυπικές αποκλίσεις είναι πιο δύσκολο να προσεγγιστούν και επομένως το σφάλμα προσέγγισης για αυτές αναμένεται μεγαλύτερο. Το ερώτημα που προκύπτει είναι αν ο απλός αυτός κανόνας μπορεί να επεκταθεί στην κωδικοποίηση πλήρους αναζήτησης (όταν για όλες δηλαδή τις τύπου R περιοχές εξετάζονται όλες οι τύπου D περιοχές). Μία απάντηση στο ερώτημα που τέθηκε, δίνει το σχήμα 5.12 το οποίο απεικονίζει τη διασπορά του RMS σφάλματος σε συνάρτηση με την τυπική απόκλιση των range περιοχών για την εικόνα της Lena. Από το σχήμα φαίνεται καθαρά, ότι υπάρχει ένα ανώτερο όριο ως προς το σφάλμα προσέγγισης το οποίο μάλιστα είναι μία ευθεία γραμμή, για όλες τις τύπου R περιοχές ανεξαρτήτως τυπικής απόκλισης. Στην περίπτωση της Lena η ευθεία περιγράφεται από την εξίσωση:

$$RMS(R) \leq 0.6 * std(R) \quad (5.63)$$

και επομένως, η ανισότητα (5.59) γίνεται

$$|std(R) - s_i std(D)| \leq RMS(R) \leq 0.6 * std(R) \quad (5.64)$$



Σχ.5.12: Γραφική παράσταση της διασποράς του  $RMS(R)$  σε σχέση με την τυπική απόκλιση των range περιοχών για την εικόνα της Lena.

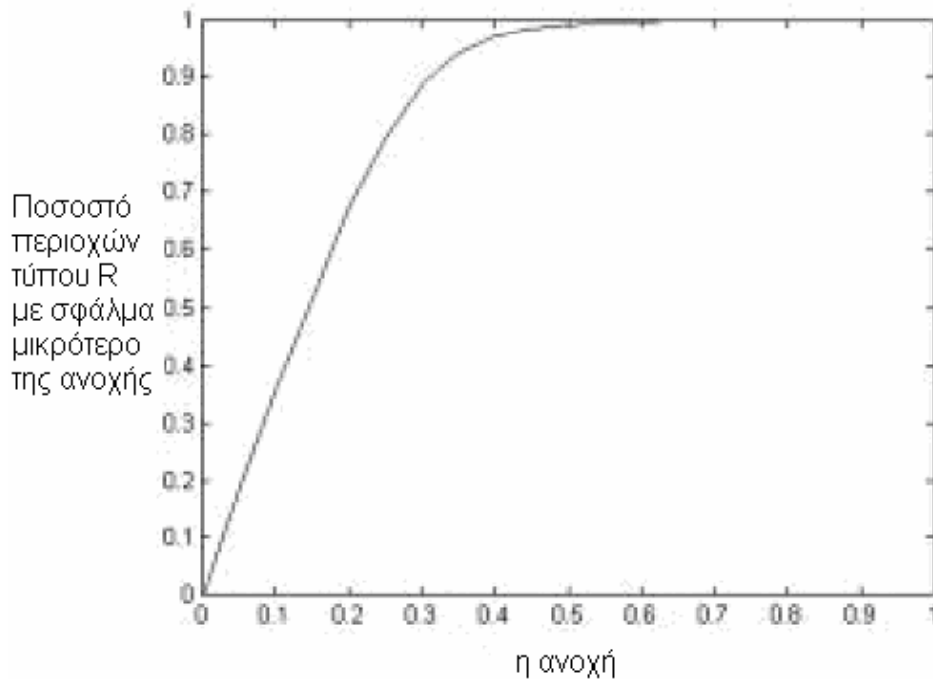
Παρόμοια γραμμικά ανώτατα όρια, έχει διαπιστωθεί μετά από πειραματικές δοκιμές, ότι υπάρχουν και για πολλές άλλες εικόνες, γεγονός που οδηγεί στην καθιέρωση του ακόλουθου προσαρμοστικού κριτηρίου αναζήτησης:

$$|std(R) - s_i std(D)| \leq T_1 = std(R)\theta \quad (5.65)$$

Για κάθε περιοχή τύπου R, ξεινά μία αναζήτηση που περιορίζεται όμως στο σύνολο εκείνων μόνο των domain περιοχών που ικανοποιούν το κριτήριο για μία συγκεκριμένη ανοχή  $\theta$ .

Για την εικόνα της Lena π.χ., η τιμή  $\theta=0.6$  εξασφαλίζει ότι όλες οι τύπου R περιοχές καθώς και οι τύπου D που τις προσεγγίζουν καλύτερα, ικανοποιούν τη συνθήκη. Αυτό σημαίνει ότι αν και πολλές περιοχές που δεν ικανοποιούν τη συνθήκη απορρίπτονται, η τελευταία εγγυάται ότι εκείνες που προσφέρονται για ικανοποιητική προσέγγιση γίνονται δεκτές. Με άλλα λόγια, επιτυγχάνεται σύγκλιση, δίχως απώλειες σε ποιότητα.

Προφανώς, αν κάποιος επιθυμεί να επιταχύνει περισσότερο την διαδικασία, έχει τη δυνατότητα μείωσης της τιμής της ανοχής, θυσιάζοντας όμως την πιστότητα της αποκωδικοποιημένης εικόνας, αφού μια τέτοια κίνηση θα απέκλειε μερικές από τις καλύτερες ως προς την προσέγγιση τύπου D περιοχές, που δε θα ικανοποιούσαν τη συνθήκη αναζήτησης. Στο σχήμα 5.13 έχει σχεδιαστεί το ποσοστό των περιοχών τύπου R που η σύγκριση τους με τις αντίστοιχες τύπου D ικανοποιεί το προσαρμοστικό κριτήριο, σε συνάρτηση με την ανοχή  $\theta$ . Το σχήμα αποκαλύπτει ότι η σχέση μεταξύ ποσοστού-ανοχής είναι γραμμική για μικρές τιμές της ανοχής, ενώ για μεγάλες τιμές της ανοχής (γύρω στο 0.5) το ποσοστό σταθεροποιείται στο 100%. Συνεπώς μία τιμή  $\theta=0.5$  θα συμπεριελάμβανε όλες τις καλύτερες υποψήφιες τύπου D περιοχές οπότε ένας αλγόριθμος κωδικοποίησης που θα χρησιμοποιούσε το κριτήριο, θα απολάμβανε το προνόμιο μιας ενισχυμένης ταχύτητας, με αμελητέα επίδραση στην ποιότητα συμπίεσης. Καθώς η καμπύλη αρχίζει να αποκλίνει από την ευθεία γραμμή για τιμή ανοχής μεγαλύτερη του 0.1, συμπεραίνεται ότι η πιο αποδοτική από πλευράς κόστους τιμή ανοχής είναι λίγο μεγαλύτερη του 0.1.



Σχ.5.13: Ποσοστό των ζευγών που δίνουν την καλύτερη προσέγγιση και ικανοποιούν την προσαρμοστική συνθήκη.

Έχοντας προτείνει ένα κριτήριο για τον περιορισμό της αναζήτησης, οι συγγραφείς προτείνουν και ένα κριτήριο τερματισμού της αναζήτησης. Υποθέτουμε ότι η ποιότητα της ανακατασκευασμένης εικόνας πρέπει να φθάνει ή και να υπερβαίνει κάποιο κατώφλι του PSNR ή ισοδύναμα το μέσο τετραγωνικό σφάλμα αποκωδικοποίησης πρέπει να έχει ένα άνω όριο. Αλλά αυτό, με βάση το *Θεώρημα Collage* σημαίνει ότι το RMS σφάλμα προσέγγισης υπόκειται και εκείνο με τη σειρά του σε περιορισμό σύμφωνα με τη σχέση (2.22).

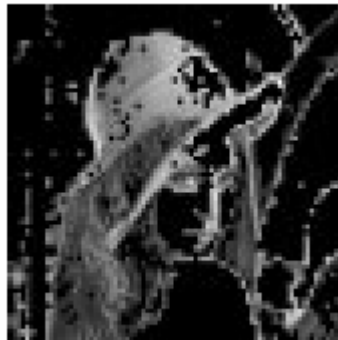
Επομένως, την αναζήτηση μπορεί να τερματιστεί από τη στιγμή που θα βρεθεί μία τέτοια προσέγγιση ώστε:

$$RMS(R) \leq T_0 \quad (5.66)$$

όπου το κατώφλι  $T_0$  εξαρτάται από την πιστότητα, τη μέγιστη τιμή του παράγοντα συστολής και την εικόνα. Αν δε βρεθεί τύπου D περιοχή ώστε να ικανοποιείται η παραπάνω ανισότητα, τότε η αναζήτηση συνεχίζεται κανονικά και επιλέγεται στο τέλος η καλύτερη περιοχή τύπου D. Θέτοντας  $T_0=0$ , στην ουσία ισοδυναμεί με την αναζήτηση σε ολόκληρο το σύνολο των τύπου D

περιοχών που είναι αποδεικτές. Για το λόγο αυτό, το κατώφλι  $T_0$  πρέπει να καθορίζεται από της ανάγκες της συμπίεσης, για παράδειγμα για πιστότητα 30dB το τελευταίο έχει τιμή περίπου ίση με 3.

Για την απεικόνιση της κατανομής του σφάλματος προσέγγισης για διαφορετικές τύπου R περιοχές, παρατίθεται το σχήμα 5.14, στο οποίο όλες οι τύπου R περιοχές για τις οποίες ισχύει  $RMS(R) \leq T_0 = 3$ , έχουν αντικατασταθεί από τετράγωνα μηδενικής έντασης. Οι μισές περιοχές της εικόνας που ικανοποιούν το κριτήριο συνιστούν τα ομοιόμορφα της μέρη και είναι προφανές ότι το κριτήριο στην περίπτωση αυτή είναι ωφέλιμο από πλευράς ταχύτητας κωδικοποίησης. Οι υπόλοιπες όμως περιοχές οι οποίες αντιπροσωπεύουν υφή ή περιέχουν αικμές, δεν επηρεάζονται από το κριτήριο τερματισμού, καθώς ο αλγόριθμος θα συνεχίσει την αναζήτηση της καλύτερης δυνατής τύπου D περιοχής.



Σχ.5.14: Η Lena, οι range με  $RMS(R) < 3$  έχουν αντικατασταθεί από μηδενικές περιοχές.

### 5.5.2 Το κριτήριο των Lai, Lam και Siu

Η αρχή του προτεινόμενου αλγορίθμου των παραπάνω συγγραφέων<sup>17</sup> είναι η παράκαμψη εκείνων των τύπου D περιοχών που ικανοποιούν μία συνθήκη απόρριψης, ώστε να μη λαμβάνονται υπόψη κατά τους υπολογισμούς για την προσέγγιση μίας δεδομένης τύπου R περιοχής. Όπως και στην εργασία των Tong και Pi (Κεφάλαιο 4), αντί της φωτεινότητας προτιμάται η συμπίεση με βάση τη μέση τιμή της τύπου R περιοχής, οπότε, για τον υπολογισμό του σφάλματος προσέγγισης μπορεί να χρησιμοποιηθεί η σχέση (4.2).

Από την (4.2), η συνάρτηση σφάλματος μπορεί να απλοποιηθεί περαιτέρω όπως φαίνεται παρακάτω:

$$\begin{aligned}
 E(R, D) &= \left[ \|R\|^2 - \frac{1}{k} \langle R, I \rangle^2 \right] - \frac{[\langle R, D \rangle - \frac{1}{k} \langle R, I \rangle \langle D, I \rangle]^2}{\left[ \|D\|^2 - \frac{1}{k} \langle D, I \rangle^2 \right]} = \\
 &= \left[ \|R\|^2 - \frac{1}{k} \langle R, I \rangle^2 \right] - \frac{[\langle R, D \rangle - \frac{1}{k} \langle R, I \rangle \langle D, I \rangle]^2}{\left[ \|D\|^2 - \frac{1}{k} \langle D, I \rangle^2 \right]^2} * \\
 &\left[ \|D\|^2 - \frac{1}{k} \langle D, I \rangle^2 \right]
 \end{aligned} \tag{5.67}$$

Αντικαθιστώντας από τη σχέση (2.24) την αντίθεση στον τελευταίο τύπο, προκύπτει η ακόλουθη έκφραση για το σφάλμα:

$$E(R, D) = A - s^2 B \tag{5.68}$$

όπου

$$A = \|R\|^2 - \frac{1}{k} \langle R, I \rangle^2 \tag{5.69}$$

και

$$B = \|D\|^2 - \frac{1}{k} \langle D, I \rangle^2 \tag{5.70}$$

Ο συντελεστής  $s$  είναι περιορισμένος στο διάστημα  $(-1, 1)$  για να εξασφαλισθεί η σύγκλιση κατά τη διαδικασία της αποκωδικοποίησης. Αν το  $A$  είναι μεγαλύτερο από το  $B$  το μέγιστο σφάλμα παρατηρείται όταν  $s=0$ , ενώ το ελάχιστο όταν  $s=1$ . Επομένως έχουμε:

Αν  $A-B \geq 0$ , τότε:

1. Το μέγιστο σφάλμα παρατηρείται όταν  $s=0$ ,  $E_{max} = A - s^2 B = A$
2. Το ελάχιστο σφάλμα παρατηρείται όταν  $s=\pm 1$ ,  $E_{min} = A - s^2 B = A - B$

Αυτό σημαίνει ότι για να βρεθεί η περιοχή τύπου  $D$  που να αποτελεί τη βέλτιστη προσέγγιση, η αναζήτηση γίνεται μόνο όταν το ελάχιστο σφάλμα  $E_{min}$

για την υπό εξέταση περιοχή, είναι μικρότερο από το τρέχον ελάχιστο σφάλμα, έστω  $d_{min}$ . Η συνθήκη απόρριψης που προτείνουν συνεπώς οι συγγραφείς είναι η ακόλουθη:

$$E_{min} = A - B \geq d_{min} \quad (5.71)$$

Βασισμένος στην (5.71), ο αλγόριθμος ταχείας αναζήτησης, ο οποίος μπορεί να απορρίπτει αποδοτικά τις ανόμοιες τύπου D περιοχές για μία δεδομένη τύπου R, διατυπώνεται ως εξής:

1. Επιλέγονται οι τύπου D περιοχές από αριστερά προς τα δεξιά και από πάνω προς τα κάτω (χωρίς αυτό να σημαίνει ότι ο αλγόριθμος επηρεάζεται αν διατρεξουμε διαφορετικά την εικόνα).
2. Στη συνέχεια υπολογίζονται τα σφάλματα προσέγγισης καθενός από τους οκτώ ισομετρικούς μετασχηματισμούς της 1<sup>ης</sup> τύπου D περιοχής  $D_1$ . Ο ισομετρικός μετασχηματισμός με το ελάχιστο σφάλμα θεωρείται ότι είναι η αρχική τύπου D περιοχή που ταιριάζει καλύτερα.
3. Η τρέχουσα ελάχιστη απόσταση καθορίζεται για αυτή την ελάχιστη παραμόρφωση έστω  $E(R, D_1)$  και η αναζήτηση προχωρά όπως η ανίχνευση σε φύλλο σχεδιάσεως. Για να καθοριστεί αν η επόμενη υποψήφια τύπου D περιοχή  $D_2$  βρίσκεται πλησιέστερα στην R από την καλύτερη μέχρι στιγμής προσέγγιση  $D_1$ , υπολογίζεται το ελάχιστο σφάλμα  $E_{min}(R, D_2)$  και συγκρίνεται με το  $d_{min}$ . Αν το  $E_{min}(R, D_2) \geq d_{min}$  τότε η  $D_2$  απορρίπτεται. Διαφορετικά, η πραγματική απόκλιση  $E(R, D_2)$  υπολογίζεται για κάθε έναν από τους οκτώ ισομετρικούς μετασχηματισμούς της  $D_2$  και συγκρίνεται με το  $d_{min}$ . Αν όλα τα σφάλματα προσέγγισης μεταξύ των μετασχηματισμών της  $D_2$  και της R είναι μεγαλύτερα από το τρέχον ελάχιστο σφάλμα  $d_{min}$ , τότε η  $D_2$  απορρίπτεται για τον ίδιο λόγο που αναφέρθηκε παραπάνω. Διαφορετικά το  $d_{min}$  αντικαθίσταται από το  $E(R, D_2)$  και η τρέχουσα τύπου D περιοχή που προσεγγίζει καλύτερα την R είναι η  $D_2$  με τον αντίστοιχο ισομετρικό της μετασχηματισμό.

4. Αυτή η διαδικασία επαναλαμβάνεται για όλες τις τύπου D περιοχές  $D_i$  στο σύνολο των τύπου D περιοχών για να βρεθεί εκείνη που προσεγγίζει καλύτερα τη δεδομένη τύπου R περιοχή που εισάγεται.

Με βάση αυτή τη συνθήκη τερματισμού οι υπολογισμοί για την αναζήτηση της βέλτιστης περιοχής τύπου D, μειώνονται σημαντικά.

Στην εφαρμογή του αλγορίθμου, οι συγγραφείς δεσμεύουν 5 ψηφία για την αναπαράσταση των διακριτών τιμών της αντίθεσης. Επομένως, οποιαδήποτε τιμή της αντίθεσης που εμπίπτει στο διάστημα  $(-0.03125, 0.03125)$ , γίνεται μηδενική μετά τη διακριτοποίηση. Με βάση την (5.68) και καθώς  $|s| < 1$ , θα έχουμε τη συνθήκη μηδενικής αντίθεσης μόνο όταν  $A < B$ :

$$E(R, D) = A - s^2 B \geq 0 \text{ ή } \sqrt{\frac{A}{B}} \geq |s| \quad (5.72)$$

Ο παράγοντας αντίθεσης  $s$  διακριτοποιείται στην τιμή 0 αν η απόλυτη τιμή του είναι μικρότερη του 0.03125, δηλαδή αν

$$0.03125 > \sqrt{\frac{A}{B}} \geq |s| \quad (5.73)$$

Όταν ο  $s$  διακριτοποιείται στο 0 το αντίστοιχο σφάλμα δίνεται από τη σχέση  $E(R, D) = A$ . Σε αυτήν την περίπτωση το σφάλμα σύγκρισης μεταξύ της τύπου R και της τύπου D περιοχής βρίσκεται χωρίς να γίνει κάποιος υπολογισμός και είναι η σταθερή τιμή  $A$ .

Στον πίνακα 5.5 εμφανίζονται τα αποτελέσματα της σύγκρισης του παραπάνω αλγορίθμου (συνθήκη απόρριψης και υπολογισμός σφάλματος με βάση τη πρόβλεψη μηδενικής αντίθεσης, Περίπτωση 1 και 2) με έναν αλγόριθμο πλήρους αναζήτησης.



Πίνακας 5.5: Σύγκρισης της επίδοσης ενός αλγορίθμου πλήρους αναζήτησης με τον προτεινόμενο αλγόριθμο (Περίπτωση 1& 2), για τρεις εικόνες που έχουν διαμεριστεί με βάση το τετραεδρικό δένδρο.

Δοκιμαστικές εικόνες		Αλγόριθμοι	
		Πλήρης αναζήτηση	Προτεινόμενος αλγόριθμος (Περίπτωση 1&2)
Lena	Χρόνος (sec)	5940	3242
	PSNR (dB)	34.91	34.91
	Βαθμός συμπίεσης	15.50	15.50
Βάρκα	Χρόνος (sec)	8463	3666
	PSNR (dB)	34.91	34.91
	Βαθμός συμπίεσης	9.71	9.71
Goldhill	Χρόνος (sec)	8976	5755
	PSNR (dB)	33.36	33.36
	Βαθμός συμπίεσης	9.11	9.11

Από τον πίνακα 5.5 συμπεραίνεται ότι η μέθοδος των Lai, Lam και Siu μειώνει σημαντικά το χρόνο υπολογισμού (γύρω στο 50%), διατηρώντας παράλληλα την ποιότητα συμπίεσης.

## 5.6 Μέθοδοι βασισμένες σε κατασκευή δένδρου

Εκτός από τη μέθοδο μείωσης της διάστασης και αυτή της ταξινόμησης με βάση την τυπική απόκλιση που προαναφέρθηκαν, οι Caso, Obrador και Kuo<sup>9</sup> πρότειναν και μια μέθοδο αναζήτησης που μοιάζει με την αναζήτηση σε δένδρο. Σε αυτή, το σύνολο των κωδικών διανυσμάτων οργανώνεται επαναληπτικά σε ένα δυαδικό δένδρο. Αρχικά, δύο «γονείς»-διανύσματα επιλέγονται από το σύνολο. Έπειτα, όλα τα διανύσματα ταξινομούνται σε ένα από τα δύο υποσύνολα που αντιπροσωπεύουν οι γονείς ανάλογα με το ποιος από τους δύο προσεγγίζει καλύτερα ως προς τα ελάχιστα τετράγωνα το υπό εξέταση διάνυσμα. Το αποτέλεσμα είναι η διαίρεση ολόκληρου του συνόλου

των τύπου  $D$  περιοχών σε δύο ομάδες-υποδένδρα τα οποία έχουν ρίζες τα δύο αρχικά διανύσματα. Επαναλαμβάνοντας την ίδια διαδικασία για κάθε ένα από τα δύο υποσύνολα δημιουργείται ένα δυαδικό δένδρο. Η διαδικασία σταματά όταν φθάσει σε ένα προκαθορισμένο ύψος του δένδρου. Δεδομένης μιας τύπου  $R$  περιοχής, συγκρίνονται με αυτή οι περιοχές που αποτελούν τους κόμβους του δυαδικού δένδρου έως ότου να συναντήσει κανείς ένα κόμβο, όλα τα διανύσματα που βρίσκονται εντός του οποίου θα έχουν ελεγχθεί. Αυτή η διαδικασία δεν εγγυάται την απόλυτα βέλτιστη προσέγγιση. Παρ' όλα αυτά, η καλύτερη ή τουλάχιστον η πιο αποδεκτή λύση μπορεί να βρεθεί με την επέκταση της αναζήτησης σε γειτονικούς κόμβους. Για το λόγο αυτό, χρησιμοποιείται ένα αριθμητικό τεστ βασισμένο στο κριτήριο της γωνίας που αναφέρθηκε σε προηγούμενη ενότητα. Η μέθοδος του δυαδικού δένδρου σχετίζεται με την αναζήτηση του πλησιέστερου γείτονα καθώς το κριτήριο ελαχίστων τετραγώνων (ελαχιστοποίηση του  $E(D,R)$ ) είναι ισοδύναμο με το κριτήριο της απόστασης (ελαχιστοποίηση του  $\Delta(\phi(R),\phi(D))$ ). Κατά συνέπεια, το δυαδικό δένδρο μπορεί να θεωρηθεί ως μια τυχαία έκδοση του  $k$ -διάστατου δένδρου που έχει μελετηθεί σε προηγούμενη ενότητα.

Ο *Van der Walle*<sup>18</sup> εργάστηκε πάνω στη μελέτη της συμπίεσης με fractal από τη σκοπιά του μετασχηματισμού *wavelet*, όπου όπως και στη μέθοδο των fractal τα τύπου  $R$  διανύσματα (που αντιστοιχούν στα υποδένδρα των *wavelet* συντελεστών), πρέπει να συγκριθούν με τα τύπου  $D$  διανύσματα (που αντιστοιχούν στους κόμβους του *wavelet* δένδρου) τα οποία μπορούν να πολλαπλασιαστούν με ένα αυθαίρετο παράγοντα. Για κάθε κόμβο κατασκευάζεται ένα χαρακτηριστικό διάνυσμα βασισμένο στις γωνίες μεταξύ των διανυσμάτων των συντελεστών και των αξόνων του χώρου *wavelet*. Αυτά τα διανύσματα ταξινομούνται στη συνέχεια σε μία πολυδιάστατη δομή δεδομένων μέσα στην οποία οργανώνεται η ταχεία αναζήτηση. Χρησιμοποιώντας τις ιδιότητες της απόστασης των χαρακτηριστικών διανυσμάτων  $\pm\phi(D)$ , η αναζήτηση αυτή έχει ως εξής: Καθορίζεται ένα σύνολο από σταθερά σημεία

(*anchor points*) στο χώρο των χαρακτηριστικών διανυσμάτων (π.χ., στις θέσεις των κύριων συνιστωσών-στοιχείων του συνόλου όλων των χαρακτηριστικών διανυσμάτων). Για κάθε κωδικό χαρακτηριστικό διάνυσμα καθώς και για κάθε τύπου R περιοχή υπολογίζονται οι αποστάσεις  $\perp$  από τα σταθερά σημεία. Αν ένα σημείο στο χώρο των χαρακτηριστικών διανυσμάτων προσεγγίζει ένα χαρακτηριστικό τύπου R διάνυσμα, πρέπει οι αποστάσεις του από τα σταθερά σημεία να είναι παρόμοιες με αυτές του τύπου R διανύσματος. Για τη διευκόλυνση της αναζήτησης τέτοιων χαρακτηριστικών τύπου D διανυσμάτων, τα παραπάνω μπορούν να οργανωθούν σε δομή δένδρου.

## 5.7 Προσεγγίσεις πολλαπλής ανάλυσης

Στην ενότητα αυτή, θα εξεταστούν δύο προσεγγίσεις πολλαπλής ανάλυσης που έχουν προταθεί από τους *Dekking*<sup>19</sup> και *Lin* και *Βενετσανόπουλο*<sup>20</sup> με σκοπό τη μείωση της πολυπλοκότητας της κωδικοποίησης. Οι μέθοδοι στηρίζονται στην ιδέα ότι για τη μείωση του κόστους αναζήτησης, μπορούμε να χρησιμοποιήσουμε την πυραμίδα των διαβαθμίσεων του γκρι που σχετίζεται με την προς συμπίεση εικόνα. Η αναζήτηση πραγματοποιείται αρχικά, σε μία χαμηλή ανάλυση της εικόνας. Αν δεν βρεθούν ικανοποιητικές προσεγγίσεις σε αυτό το επίπεδο ανάλυσης, το ίδιο θα συμβεί προφανώς και σε υψηλότερο. Η υπολογιστική οικονομία οφείλεται στο γεγονός ότι σε χαμηλή ανάλυση οι υπολογισμοί με τη μέθοδο των ελαχίστων τετραγώνων είναι λιγότεροι, αφού ο αριθμός των περιοχών της εικόνας είναι μικρότερος. Για μια πιο ακριβή περιγραφή της μεθόδου παρατίθενται οι ορισμοί κάποιων βασικών εννοιών: Η πυραμίδα των διαβαθμίσεων του γκρι μιας εικόνας  $f$ , η οποία θεωρείται δισδιάστατος πίνακας, ορίζεται ως η ακολουθία των εικόνων..  $f^{(0)}, \dots, f^{(r)}$ , όπου  $f^{(r)} = f$  και

$$f^{(k)}(i, j) = \frac{1}{4} \sum_{m,l=0}^1 f^{(k+1)}(2i+m, 2j+l) \quad (5.74)$$

για  $k=0, \dots, r-1$  και  $0 \leq i, j < 2^k$ . Ο αριθμός  $k$  υποδηλώνει το επίπεδο ανάλυσης της εικόνας και η σχέση που δόθηκε παραπάνω δείχνει ότι για να αποκτήσει κανείς μία εικόνα ανάλυσης  $k$ , πρέπει να ξειινήσει από την εικόνα ανάλυσης  $k+1$  αναθέτοντας σε κάθε *εικονοστοιχείο* της πρώτης, την ένταση που προκύπτει από το μέσο όρο των τεσσάρων γειτονικών *εικονοστοιχείων* της δεύτερης που προήλθαν από το τεμαχισμό σε τεταρτημόρια αυτού της πρώτης. Ομοίως, είναι δυνατό να προκύψουν τύπου R και τύπου D περιοχές της εικόνας στην ανάλυση  $k$ , από ήδη προϋπάρχουσες στην ανάλυση  $k+1$ . Το βασικό αποτέλεσμα της εργασίας του Dekking είναι το θεώρημα 7 που αναφέρθηκε σε προηγούμενη ενότητα και το οποίο υπενθυμίζεται εδώ:

**Θεώρημα 7:**

Έστω  $R^{(k)}$  και  $D^{(k)}$  οι *range* και *domain* περιοχή αντίστοιχα, μιας εικόνας σε ανάλυση  $k$ . Τότε  $E(D^{(k+1)}, R^{(k+1)}) \geq E(D^{(k)}, R^{(k)})$ .

Μολαταύτα, επειδή δεν υπάρχουν όλες οι τύπου D περιοχές στην ανάλυση  $k+1$  που να έχουν τις αντίστοιχες στην ανάλυση  $k$ , η εφαρμογή του θεωρήματος με την παραπάνω μορφή του λαμβάνει υπόψη εκείνες τις τύπου D περιοχές ανάλυσης  $k+1$  που έχουν τις γωνίες τους στη θέση  $(2i, 2j)$ . Για να παρακαμφθεί το πρόβλημα, κατασκευάζεται ένα δένδρο για την πυραμίδα, όπου κάθε τύπου D περιοχή ανάλυσης  $k+1$  έχει τέσσερα παιδιά ανάλυσης  $k$  στις θέσεις  $(2i, 2j)$ ,  $(2i+1, 2j)$ ,  $(2i, 2j+1)$ ,  $(2i+1, 2j+1)$ . Πρέπει ακόμα να επισημανθεί ότι δεν πρέπει να αποκλείεται από την αναζήτηση μία τύπου D περιοχή ανάλυσης  $k+1$  μόνο και μόνο επειδή ένα παιδί της, ανάλυσης  $k$ , έχει παράγοντα αντίθεσης μεγαλύτερο απολύτως της μονάδας. Στην πραγματικότητα, έχουν παρατηρηθεί περιπτώσεις όπου ενώ  $|s_k| > 1$ ,  $s_{k+1} = 0$ .

## 5.8 Ταχεία αναζήτηση μέσω ταχείας συνέλιξης

Οι περισσότερες τεχνικές με τις οποίες έχουμε ασχοληθεί μέχρι το σημείο αυτό, είναι απωλεστικές με την έννοια ότι θυσιάζουν ένα μέρος της πιστότητας της εικόνας στο βωμό της μεγαλύτερης ταχύτητας συμπίεσης. Αντίθετα, σε μία μη απωλεστική μέθοδο, το αντικείμενο αναζήτησης είναι το βέλτιστο κωδικό διάνυσμα-περιοχή, το οποίο και δίνει το ελάχιστο σφάλμα, παρά ένα υποβέλτιστο το σφάλμα το οποίο δίνει είναι αρκετά μικρό σε σχέση με τις προδιαγραφές που έχουν οριστεί, όχι όμως και το ελάχιστο. Η μέθοδος που αναπτύσσεται εδώ, εκμεταλλεύεται το γεγονός ότι τα κωδικά διανύσματα-περιοχές που απαρτίζουν την εικόνα επικαλύπτουν το ένα το άλλο. Η ταχεία συνέλιξη- που βασίζεται στο θεώρημα της συνέλιξης και εκτελείται στο πεδίο της συχνότητας- είναι η ιδανική για την εκμετάλλευση της σχέσης αυτής μεταξύ των κωδικών διανυσμάτων. Το ουσιώδες τμήμα του βασικού υπολογισμού της κλασικής μεθόδου συμπίεσης με fractal, είναι μια συγκεκριμένη μορφή συνέλιξης<sup>21</sup>. Για να γίνει αυτό κατανοητό, υπενθυμίζονται οι σχέσεις που δίνουν την αντίθεση και τη φωτεινότητα:

$$s^* = \frac{\langle R, D \rangle - \frac{1}{k} \langle R, \mathbf{1} \rangle \langle D, \mathbf{1} \rangle}{\|D\|_2^2 - \frac{1}{k} \langle D, \mathbf{1} \rangle^2} \quad (\text{i}), \quad o^* = \frac{1}{k} (\langle R, \mathbf{1} \rangle - s^* \langle D, \mathbf{1} \rangle) \quad (\text{ii}) \quad (5.75)$$

Για οποιαδήποτε  $(s, o)$  το σφάλμα  $E(D, R)$  μπορεί να θεωρηθεί συνάρτηση των  $\langle D, R \rangle$ ,  $\langle D, D \rangle$ ,  $\langle D, \mathbf{1} \rangle$ ,  $\langle R, R \rangle$  και  $\langle R, \mathbf{1} \rangle$ . Η αποτίμησή του απαιτεί 23 πράξεις κινητής υποδιαστολής. Τυπικά, οι υπολογισμοί μπορούν να οργανωθούν σε δύο φωλιασμένους βρόχους (*nested loops*):

- Ολική προεπεξεργασία: υπολογισμός των  $\langle D, D \rangle$ ,  $\langle D, \mathbf{1} \rangle$  για όλα τα κωδικά διανύσματα  $D$
- Τοπική προεπεξεργασία: υπολογισμός των  $\langle R, R \rangle$ ,  $\langle R, \mathbf{1} \rangle$ .

Για όλα τα κωδικά διανύσματα εκτελούμε:

- Υπολογισμό των  $\langle D, R \rangle$  και  $E(D, R)$

Ο υπολογισμός των εσωτερικών γινομένων  $\langle D, R \rangle$  αποτελεί το μεγαλύτερο ποσοστό του πλήθους των υπολογισμών της κωδικοποίησης. Τα κωδικά διανύσματα σχηματίζονται τυπικά από την υποδειγματοληψία της εικόνας στη μισή της ανάλυση. Οποιαδήποτε περιοχή της υποδειγματοληφθείσας εικόνας που έχει το ίδιο μέγεθος με μια τύπου  $R$  περιοχή μπορεί να θεωρηθεί κωδικό διάνυσμα για την τελευταία. Συνεπώς τα εσωτερικά γινόμενα  $\langle D, R \rangle$  δεν είναι τίποτε άλλο από την FIR απόκριση της υποδειγματοληφθείσας εικόνας σε σχέση με την τύπου  $R$  περιοχή. Με άλλα λόγια απαιτείται ο υπολογισμός της συνέλιξης ή καλύτερα της ετεροσυσχέτισης της τύπου  $R$  περιοχής  $R$ . Αυτή η διακριτή συνέλιξη σε δύο διαστάσεις μπορεί να εκτελεστεί αποτελεσματικά στο χώρο της συχνότητας όταν η τύπου  $R$  περιοχή δεν είναι πολύ μικρή (Θεώρημα Συνέλιξης). Η παραπάνω λοιπόν διαδικασία μεταθέτει τον υπολογισμό των γινομένων  $\langle D, R \rangle$  από την επεξεργασία, στην τοπική προεπεξεργασία όπου τα γινόμενα αυτά υπολογίζονται σε μία παρτίδα για όλα τα κωδικά διανύσματα  $D$ , με τη χρήση του ταχέως μετασχηματισμού Fourier. Κατά συνέπεια, η μέθοδος η οποία εξετάζεται εδώ, είναι φανερά μη απωλεστική.

Επιπλέον, και η ολική προεπεξεργασία, που απαιτεί σημαντικό χρόνο μέχρι να ολοκληρωθεί, μπορεί να επιταχυνθεί από την ίδια τεχνική της συνέλιξης. Τα γινόμενα  $\langle D, \mathbf{1} \rangle$  υπολογίζονται με συνέλιξη της υποδειγματοληφθείσας εικόνας με μία τύπου  $R$  περιοχή της οποίας όλα τα *εικονοστοιχεία* έχουν την ίδια ένταση. Το άθροισμα των τετραγώνων υπολογίζεται ομοίως με όλες τις εντάσεις των *εικονοστοιχείων* της υποδειγματοληφθείσας εικόνας να έχουν υψωθεί στο τετράγωνο πριν από τη συνέλιξη.

## 5.9 Συμπύεση χωρίς αναζήτηση

Οι μέθοδοι μείωσης της πολυπλοκότητας που είναι κάπως διαφορετικές σε σχέση με τις προηγούμενες βασίζονται στην ελάττωση του αριθμού των τύπου D περιοχών σε ένα πολύ μικρό σύνολο. Για παράδειγμα στην εργασία που ακολούθησε αυτήν των *Monro* και *Dudridge*<sup>22</sup>, για κάθε περιοχή, το κωδικό διάνυσμα που χρησιμοποιείται για την προσέγγισή της έχει προκαθοριστεί να προέρχεται από μία συγκεκριμένη περιοχή που περιέχει την τύπου R<sup>23</sup>. Παρόμοια βήματα ακολούθησαν και οι *Hurtgen* και *Stiller*<sup>6</sup> στην εργασία των οποίων η αναζήτηση για ένα κατάλληλο κωδικό διάνυσμα περιορίζεται στη γειτονιά της υπό εξέταση τύπου R περιοχής. Επιπροσθέτως, η αναζήτηση μπορεί να επεκταθεί σε μερικές διάσπαρτες περιοχές τύπου D που βρίσκονται μακριά από την τύπου R. Στην εργασία των *Saure*,<sup>24</sup> μελετήθηκε μία μη προσαρμοστική και παραμετρική εκδοχή μείωσης του συνόλου των τύπου D περιοχών με την επίτρεψη του αποκλεισμού ενός μεταβλητού αριθμού τους (μείωση που κυμαίνεται από 0 έως και 100%) και την παρατήρηση των επιπτώσεων στο χρόνο συμπίεσης και την πιστότητα της εικόνας. Τα πειραματικά αποτελέσματα έδειξαν ότι δε χρειάζεται το σύνολο των τύπου D περιοχών να περιέχει περιοχές με μικρή απόκλιση τιμών έντασης των *εικονοστοιχείων* που ανήκουν σε αυτές. Αποκλείοντας ένα κλάσμα  $1-a$ ,  $a \in [0,1)$ , από το σύνολο των τύπου D περιοχών που αποτελείται από στοιχεία με τη μικρότερη διασπορά προκύπτει ένα πιο λιτό και πιο παραγωγικό σύνολο. Τα πειράματα (πίνακες 5.6, 5.7) από την εφαρμογή της ιδέας αυτής σε συνδυασμό με το δένδρο του Fisher έδειξαν ότι:

1. Ο υπολογιστικός χρόνος μεταβάλλεται γραμμικά με το  $a$ .
2. Ακόμη και για πολύ μικρές τιμές του  $a$ , π.χ.  $a=0.15$  δεν παρατηρείται διαφορά στην ποιότητα της εικόνας. Αντίθετα, η πιστότητα βελτιώνεται ελαφρώς.
3. Για μεγαλύτερες τιμές του  $a$ , π.χ.,  $a=0.5$ , ο βαθμός συμπίεσης μειώνεται ελαφρώς περίπου 2%.

Πίνακας 5.6: Η επίδοση του αλγορίθμου του τετραεδρικού δένδρου για το λιτό σύνολο domain περιοχών που εξετάζουμε στο εδάφιο αυτό. Η παράμετρος  $\alpha$  αναφέρεται στο κλάσμα των domain περιοχών που εξακολουθούν να βρίσκονται στο σύνολο των domain περιοχών. Ο βαθμός συμπίεσης φαίνεται στην 4<sup>η</sup> στήλη. Όταν εφαρμόζουμε τον προτεινόμενο αλγόριθμο στις αρχικές domain περιοχές παίρνουμε τους βαθμούς συμπίεσης της 5<sup>ης</sup> στήλης με τη διαφορά στο μέγεθος στη μνήμη της εικόνας να μετράται σε bytes που φαίνονται στην τελευταία στήλη.

$\alpha$	Αποτελέσματα για την 512*512 εικόνα της Lena				
	Χρόνος Εκτέλεσης  (sec)	PSNR  (dB)	Συμπίεση		
			Βαθμός συμπίεσης	Νέος βαθμός συμπίεσης	Bytes που σώζονται
1.00	15.2	32.73	14.88	14.84	-39
0.90	14.0	32.71	14.86	14.84	-34
0.80	12.6	32.75	14.85	14.83	-21
0.70	11.3	32.76	14.83	14.82	-7
0.60	10.1	32.80	14.75	14.77	24
0.50	8.7	32.87	14.57	14.62	60
0.40	7.4	32.90	14.45	14.56	137
0.30	6.0	32.93	14.19	14.88	856
0.20	4.6	32.88	13.49	14.23	1009
0.15	3.9	32.78	13.10	13.89	1135
0.10	3.1	32.53	12.64	13.98	1982
0.08	2.8	32.40	12.36	13.69	2070
0.06	2.4	32.03	12.21	14.13	2921
0.04	2.1	31.80	11.86	13.79	3101
0.02	1.7	31.03	11.39	13.86	4103



Πίνακας 5.7 : Αποτελέσματα όπως και στον προηγούμενο πίνακα για περισσότερες δοκιμαστικές εικόνες

a	Χρόνος Εκτέλεσης (sec)	PSNR (dB)	Συμπίεση		
			Βαθμός συμπίεσης	Νέος βαθμός συμπίεσης	Bytes που σώζονται
Αποτελέσματα για την 512*512 εικόνα των Πιπεριών					
1.00	16.6	32.43	15.20	15.06	-161
0.50	10.0	32.49	14.93	14.95	17
0.20	5.1	32.55	14.00	14.73	921
0.10	3.3	32.45	13.22	14.56	1828
0.05	2.4	32.11	12.31	14.22	2859
Αποτελέσματα για την 512*512 εικόνα του Μπαμπούνου					
1.00	33.3	25.15	5.68	5.59	-727
0.50	17.2	25.13	5.61	5.75	1148
0.20	8.0	24.81	5.55	5.93	3028
0.10	4.7	24.37	5.52	6.16	4915
0.05	3.3	23.87	5.50	6.42	6766
Αποτελέσματα για την 512*512 εικόνα της Βάρκας					
1.00	25.6	32.03	10.11	10.18	185
0.50	14.6	32.07	10.06	10.19	335
0.20	7.3	31.83	9.92	10.51	1476
0.10	4.4	31.53	9.82	10.85	2531
0.05	2.8	30.95	9.79	11.29	3562
Αποτελέσματα για την 512*512 εικόνα του Αεριοθούμενου					
1.00	21.2	32.86	12.55	12.60	75
0.50	12.6	32.97	12.42	12.55	211
0.20	6.2	32.94	12.05	12.75	1186
0.10	3.8	32.63	11.98	13.23	2079
0.05	2.4	32.13	11.82	13.65	2967

## Βιβλιογραφία

1. Saupe, D., *Accelerating fractal image compression by multi-dimensional nearest neighbor search*, in: *Proceedings DCC'95 Data Compression Conference*, J. A. Storer and M. Cohn (eds.), IEEE Comp. Soc. Press, March 1995.
2. Friedman, J., H., Bentley, J. L., Finkel, R. A., *An algorithm for finding best matchew in Logarithmic expected time*, *ACM Trans. Math. Software* 3,3 (1977) 209-226.
3. Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., Wu, A., *An optimal algorithm for nearest neighbor searching*, *Proc. 5<sup>th</sup> Annual ACM-SLAM Symposium on Discrete Algorithms* (1994) 573-582.
4. Lee, C-H., Chen, L. H., *Fast closest codeword search algorithm for vector quantization*, *IEE Proc. Vis. Image Signal Process.* 141, 3 (1994) 143-148.
5. Barthel, K.U., Schuttemeyer, J., Voye, T., Noll, P., *A new image coding technique unifying fractal and transform coding*, in: *Proc. ICIP-94 IEEE International Conference on Image Processing*, Austin Texas, Nov.1994.
6. Hurtgen, B., Stiller, C., *Adaptive fractal coding of still pictures*, in: *EOS/SPIE Visual Communications and PACS for Medical Applications'93*, Berlin, 1993.
7. Ramammurthi, B., Gersho, A., *Classified vector quantization of images*, *IEEE Trans. Communications COM-34*, 1986.
8. Fisher, Y., *Fractal Image Compression*, *Fractals* 2,3 (1994) 325-334
9. Caso, G., Obrador, P., Kuo, C. C. J., *Fast methods for fractal image encoding*, in: *Proc. from IS&T/SPIE 1995 Symposium on Electronic Imaging: Science and Technology*, Vol. 2501, pp. 583-594, 1995.
10. Boss, R. D., Jacobs, E. W., *Archetype classification in an iterated transformation image compression algorithm*, in: *Fractal Image Compression-Theory and Application*, Fisher, Y., (ed.), Springer-Verlag, New York, 1994.
11. Novak, M., *Attractor coding of images*, in: *Proc. of the International Picture Coding Symposium PCS'93*, Lausanne, March 1993.
12. Ning Lu, *Fractal Imaging*, Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto.
13. Jean Cardinal, *fast fractal compression of grey scale images*, Correspondence, *IEEE Transactions on Image Proc.* Vol. 10, No. 1, Jan. 2001.
14. Saupe, D., Hamzaui, R., *Combining Fractal image compression and Vector Quantization*, *IEEE Transactions on Image Proc.*, Vol. 9, No. 2, Feb. 2000.
15. Σπύρος Τζαφέστας, *Υπολογιστική Νοημοσύνη, Τόμος Α'*, σελ. 175-176
16. Belford, T., Dekking, F. M., Keane, M. S., *Fractal image coding techniques and contraction operators*, *Nieuw Arch. Wisk.* (4) 10.3 (1992) 185-218.
17. Cheung-Ming Lai, Kin-Man Lam, Wan-Chi Siu, *A fast fractal image coding based on kick-out and zero contrast conditions*, *IEEE Transactions on Image Proc.*, Vol. 12, No. 11, Nov. 2003
18. van der Walle, A., *Merging Fractal image Compression and Wavelet Transform Methods*, in: *NATO ASI Conf. Fractal Image Encoding and Analysis*, Trondheim, July 1995, to appear in a special issue of *Fractals*.
19. Dekking, F. M., *Fractal Image coding: some mathematical remarks on its limits and its prospects*, Technical Report 95-95, Faculty of Technical Mathematics and Informatics, Delft University of Technology.
20. Lin, H., Venetsanopoulos, A. N., *A pyramid algorithm for fast fractal image compression*, *Proc. ICIP-95 IEEE International Conference on Image Proc.*, Washington D.C., 1995.
21. Saupe, D., *A new view of fractal image compression as convolution transform coding*, *IEEE Signal Processing Letters* 3,7 (1996).

22. Monro, D. M., Dudridge, F., *Fractal approximation of image blocks*, in: Proceedings of ICASSP-1992 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, pp. 485-488, 1992
23. Wooley, S. J., Monro, D. M., *Optimum parameters for hybrid fractal image coding*, in : *Proceedings of ICASSP-1995 IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, 1995.
24. Saupe, D., *Lean domain pools for fractal image compression*, in: *Proceedings from IS&T/SPIE 1996 Symposium on Electronic Imaging: Science & Technology - Still Image Compression II*, Vol. 2669, to appear Jan. 1996.

---

## ΚΕΦΑΛΑΙΟ 6

---

### Πιο Προχωρημένα Θέματα – Συμπεράσματα

Στο προηγούμενο κεφάλαιο, αναφέρθηκαν μερικές από τις πιο διαδεδομένες τεχνικές συμπίεσης στη βιβλιογραφία των fractal. Στο κεφάλαιο που ακολουθεί, γίνεται λόγος για μεθόδους που δεν εμπίπτουν σε κάποια από τις κατηγορίες που παρατέθηκαν στο προηγούμενο και πρώτα από όλα για το Διακριτό Μετασχηματισμό Συνημιτόνου, ο οποίος μεταφέροντας τους υπολογισμούς από το πεδίο του χώρου στο πεδίο της συχνότητας, αποτελεί ένα από τα κυριότερα εργαλεία επιτάχυνσης των υπολογισμών και μπορεί εύκολα να συνδυασθεί με τις περισσότερες από τις μεθόδους που εξετάστηκαν στο προηγούμενο κεφάλαιο. Στο τέλος του κεφαλαίου επιχειρείται ένας απολογισμός των πεπραγμένων στο χώρο της συμπίεσης με τη θεωρία των fractal.

#### 6.1 Ο Διακριτός Μετασχηματισμός Συνημιτόνου

Στην εργασία τους, οι *Truong, Jeng, Reed, Lee* και *Li*<sup>1</sup>, προτείνουν έναν ταχύ αλγόριθμο κωδικοποίησης με fractal βασισμένο στο Διακριτό Μετασχηματισμό Συνημιτόνου (*Discrete Cosine Transform, DCT*).

##### 6.1.1 Ο υπολογισμός του DCT των ισομετρικών

##### μετασχηματισμών

Για τις ανάγκες της ανάλυσης, υπενθυμίζονται οι ισομετρικοί μετασχηματισμοί μιας τετραγωνικής περιοχής:

$$\begin{array}{cccc} 1: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 2: \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} & 3: \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & 4: \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \\ 5: \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 6: \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & 7: \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 8: \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \end{array} \quad (6.1)$$

Όπως έχει προαναφερθεί, οι μισοί από αυτούς (1, 4, 6 και 7) αντιστοιχούν σε στροφή του αρχικού τετραγώνου κατά  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  και  $270^\circ$  αντίστοιχα, ενώ οι υπόλοιποι σε στροφή του ειδώλου του ως προς την ευθεία  $y=x$ , κατά τις ίδιες γωνίες.

Ο Διακριτός Μετασχηματισμός Συνημιτόνου των 8 αυτών προσανατολισμών χαρακτηρίζεται από μια αλλαγή συντεταγμένων. Για κάθε προσανατολισμό έχουν βρεθεί αναλυτικές φόρμουλες του ΔΜΣ που είναι βασισμένες στο ΔΜΣ του αρχικού τετραγώνου. Χρησιμοποιώντας τις, μπορεί κανείς να ομαδοποιήσει τα εκάστοτε δεδομένα και να εξαλείψει τους υπολογισμούς που επαναλαμβάνονται για διαφορετικούς προσανατολισμούς, μειώνοντας έτσι το υπολογιστικό κόστος.

Ο ΔΜΣ μιας εικόνας  $f(i,j)$  μεγέθους  $L \times L$  ( $i,j=0,\dots,L-1$ ) δίνεται από τον τύπο:

$$F(m,n) = \frac{2}{L} C_m C_n \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} f(i,j) \cos\left(\frac{(2i+1)m\pi}{2L}\right) \cos\left(\frac{(2j+1)n\pi}{2L}\right),$$

$$m,n = 0,1,\dots,L-1 \text{ και} \tag{6.2}$$

$$C_k = \begin{cases} 1/\sqrt{2}, & k=0 \\ 1, & k \neq 0. \end{cases}$$

Αρχικά, υπολογίζεται ο ΔΜΣ για τα είδωλα ως προς την ευθεία  $x=(L-1)/2$  και την ευθεία  $y=(L-1)/2$  (κέντρο του τετραγώνου). Οι νέες συντεταγμένες του ειδώλου ως προς την ευθεία  $x=(L-1)/2$  είναι οι  $u, v$  που δίνονται από τη σχέση:

$$\begin{pmatrix} u - \frac{L-1}{2} \\ v \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x - \frac{L-1}{2} \\ y \end{pmatrix} \tag{6.3}$$

δηλαδή:

$$u = -x + L - 1, \quad v = y \tag{6.4}$$

Έστω  $g_1(u,v)$  η νέα εικόνα που παίρνουμε από την  $f(m,n)$  και η οποία αποτελεί είδωλο της τελευταίας ως προς την ευθεία  $x=(L-1)/2$ . Η  $g_1(u,v)$

γράφεται:  $g_1(u,v) = g_1\{u(x,y),v(x,y)\} = f(x,y)$ . Από τις (6.2), (6.4) ο ΔΜΣ της  $g_1$  υπολογίζεται ως:

$$\begin{aligned}
 G_1(m,n) &= C_m C_n \sum_{u=0}^{L-1} \sum_{y=0}^{L-1} g_1(u,v) \cos \frac{(2u+1)m\pi}{2L} \cos \frac{(2v+1)n\pi}{2L} = \\
 &= C_m C_n \sum_{-x+L-1=0}^{L-1} \sum_{y=0}^{L-1} f(x,y) \cos \frac{[2(-x+L-1)+1]m\pi}{2L} \cos \frac{(2y+1)n\pi}{2L} = \\
 &= C_m C_n \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} f(x,y) \cos \left[ \frac{(2x+1)m\pi}{2L} + m\pi \right] \cos \frac{(2y+1)n\pi}{2L} = \quad (6.5) \\
 &= (-1)^m C_m C_n \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} f(x,y) \cos \frac{(2x+1)m\pi}{2L} \cos \frac{(2y+1)n\pi}{2L} = \\
 &(-1)^m F(m,n)
 \end{aligned}$$

Για την περίπτωση του ειδώλου  $g_2(u,v)$  της  $f(x,y)$  ως προς την ευθεία  $y=(L-1)/2$ , ο ΔΜΣ  $G_2(m,n)$  της  $g_2$ , δίνεται με παρόμοιο τρόπο από τον τύπο:

$$G_2(m,n) = (-1)^n F(m,n) \quad (6.6)$$

Στη συνέχεια, υπολογίζεται ο ΔΜΣ για την περίπτωση της στροφής γύρω από την ευθεία  $y=x$ . Οι νέες συντεταγμένες του ειδώλου  $(u, v)$  δίνονται από τη σχέση:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (6.7)$$

δηλαδή:

$$u = y, v = x \quad (6.8)$$

Έστω  $h(u,v)$  το είδωλο της εικόνας  $f(x,y)$  από τη στροφή ως προς την  $y=x$ . Η  $h(u,v)$  γράφεται  $h(u,v) = h\{u(x,y), v(x,y)\} = f(x,y)$ . Συνεπώς, με βάση τις (6.2) και (6.8) ο ΔΜΣ της  $h$  υπολογίζεται ως εξής:

$$\begin{aligned}
H(m,n) &= C_m C_n \sum_{u=0}^{L-1} \sum_{v=0}^{L-1} h(u,v) \cos \frac{(2u+1)m\pi}{2L} \cos \frac{(2v+1)n\pi}{2L} = \\
&= C_m C_n \sum_{y=0}^{L-1} \sum_{x=0}^{L-1} f(x,y) \cos \left[ \frac{(2y+1)m\pi}{2L} + m\pi \right] \cos \frac{(2x+1)n\pi}{2L} = \\
&= C_n C_m \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} f(x,y) \cos \left[ \frac{(2y+1)m\pi}{2L} + m\pi \right] \cos \frac{(2x+1)n\pi}{2L} = \\
&= F(n,m)
\end{aligned} \tag{6.9}$$

Συμβολίζοντας με  $f_k$ ,  $k = 1, \dots, 8$  την εικόνα που αντιστοιχεί σε κάθε ένα από τους 8 ισομετρικούς προσανατολισμούς της αρχικής και με  $F_k$  τους αντίστοιχους ΔΜΣ, αποδεικνύεται, με τη χρήση των σχέσεων (6.5), (6.6) και (6.9) ότι ισχύουν οι ακόλουθες σχέσεις:

$$\begin{aligned}
F_2(m,n) &= (-1)^m F_1(m,n) \\
F_3(m,n) &= (-1)^n F_1(m,n) \\
F_4(m,n) &= (-1)^{m+n} F_1(m,n)
\end{aligned} \tag{6.10}$$

και

$$\begin{aligned}
F_5(m,n) &= F_1(m,n) \\
F_6(m,n) &= F_2(m,n) = (-1)^m F_1(m,n) \\
F_7(m,n) &= F_3(m,n) = (-1)^n F_1(m,n) \\
F_8(m,n) &= F_4(m,n) = (-1)^{n+m} F_1(m,n)
\end{aligned} \tag{6.11}$$

### 6.1.2 Το σφάλμα ελαχίστων τετραγώνων στο πεδίο της συχνότητας

Όπως είναι γνωστό, το σφάλμα προσέγγισης δίνεται από τη σχέση (4.2) δηλαδή

$$E(R,D) = \|R - \bar{r}\mathbf{1}\|^2 - \frac{\langle R - \bar{r}\mathbf{1}, D - \bar{d}\mathbf{1} \rangle^2}{\|D - \bar{d}\mathbf{1}\|_2^2} \tag{6.12}$$

, όπου  $\bar{r}$ ,  $\bar{d}$  οι μέσες τιμές των τύπου R και τύπου D περιοχών αντίστοιχα. Αντικαθιστώντας:

$$R - \bar{r}\mathbf{1} = \widehat{\mathbf{g}} \quad (6.13)$$

και

$$D_k - \bar{d}_k\mathbf{1} = \widehat{\mathbf{f}}_k \quad (6.14)$$

όπου  $k=1, \dots, 8$  ο δείκτης που αντιστοιχεί στον ισομετρικό μετασχηματισμό που εφαρμόζεται σε κάθε τύπου D περιοχή. Σύμφωνα με τις παραπάνω αντικαταστάσεις, το σφάλμα παίρνει τη μορφή:

$$\langle \widehat{\mathbf{g}}, \widehat{\mathbf{g}} \rangle - \langle \widehat{\mathbf{f}}_k, \widehat{\mathbf{g}} \rangle^2 / \langle \widehat{\mathbf{f}}_k, \widehat{\mathbf{f}}_k \rangle \quad (6.15)$$

Είναι φανερό, ότι η ελαχιστοποίηση του Μέσου Τετραγωνικού Σφάλματος μεταξύ των  $\widehat{\mathbf{f}}_k$ ,  $\widehat{\mathbf{g}}$  είναι ισοδύναμη με τη μεγιστοποίηση του  $\langle \widehat{\mathbf{f}}_k, \widehat{\mathbf{g}} \rangle$ .

Χρησιμοποιώντας και στην περίπτωση αυτή το ΔΜΣ, οι περιττοί υπολογισμοί για την εύρεση του σφάλματος των  $\widehat{\mathbf{f}}_k$ ,  $\widehat{\mathbf{g}}$ , μπορούν να αποφευχθούν.

Έστω  $\widehat{F}_k$ ,  $\widehat{G}$  ο ΔΜΣ των  $\widehat{\mathbf{f}}_k$ ,  $\widehat{\mathbf{g}}$  αντιστοίχως για  $k = 1, \dots, 8$ . Για κάθε  $k \in [1 \dots 8]$ , υπολογίζεται το προς μεγιστοποίηση (με βάση το προηγούμενο επιχείρημα), εσωτερικό γινόμενο  $\langle \widehat{\mathbf{f}}_k, \widehat{\mathbf{g}} \rangle$ .

$$\delta_k = \langle \widehat{\mathbf{f}}_k, \widehat{\mathbf{g}} \rangle = \langle \widehat{F}_k, \widehat{G} \rangle = \sum_{i=0}^7 \sum_{j=0}^7 \widehat{F}_k(i, j) \widehat{G}(i, j) \quad (6.16)$$

Για  $k=1$ , από την (6.4) συνεπάγεται ότι :

$$\delta_1 = \sum_{i=0}^7 \sum_{j=0}^7 \widehat{F}_1(i, j) \widehat{G}(i, j) = \sum_{i=0}^7 \sum_{j=0}^7 a_{ij} \quad (6.17)$$

όπου  $a_{ij} = \widehat{F}_1(i, j) \widehat{G}(i, j)$ .

Από την (6.10) προκύπτει:



$$\begin{aligned}
\delta_1 &= \sum_{i=0}^7 \sum_{j=0}^7 a_{ij} & \delta_2 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^i a_{ij} \\
\delta_3 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^j a_{ij} & \delta_4 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^{i+j} a_{ij}
\end{aligned} \tag{6.18}$$

Για  $k=5$ , η (6.16) δίνει:

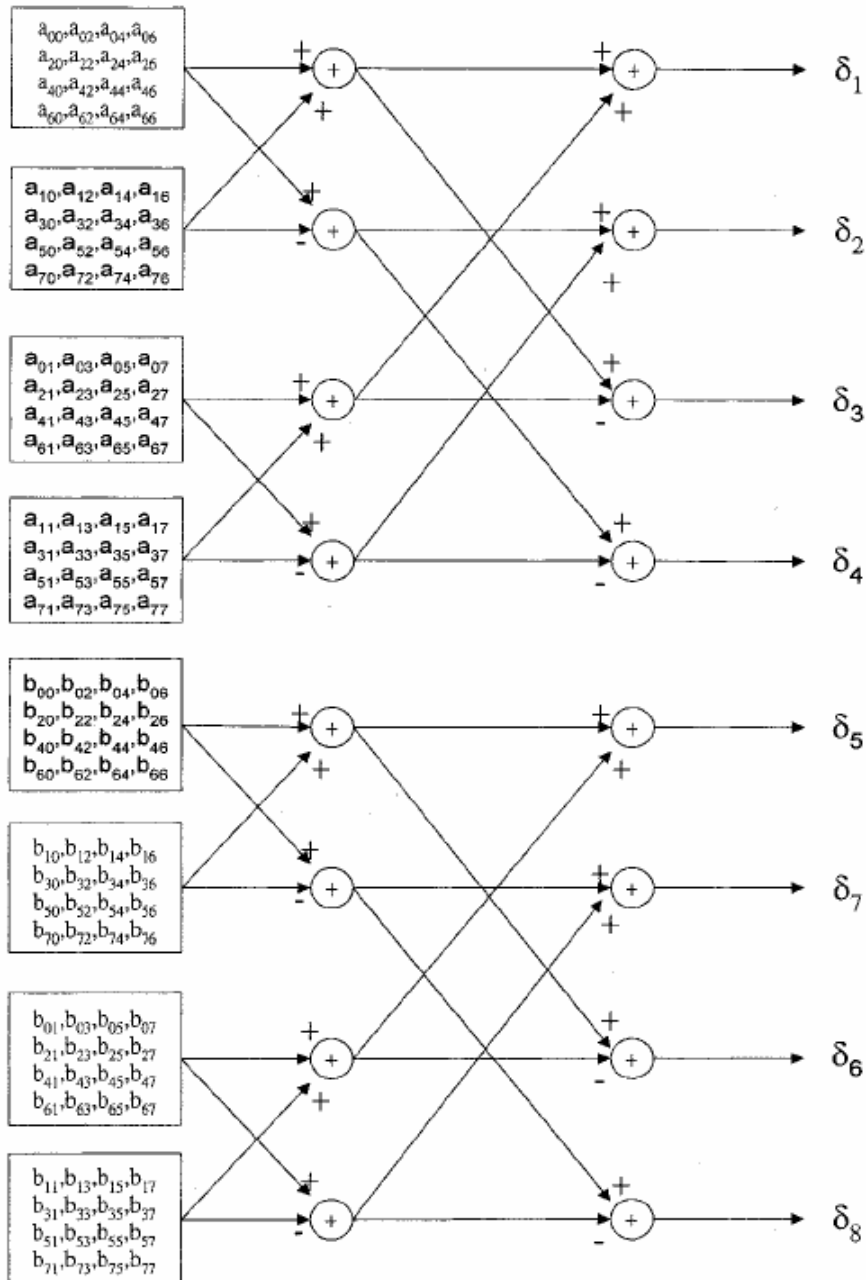
$$\begin{aligned}
\delta_5 &= \sum_{i=0}^7 \sum_{j=0}^7 \widehat{F}_5(i, j) \widehat{G}(i, j) = \sum_{i=0}^7 \sum_{j=0}^7 \widehat{F}_1(j, i) \widehat{G}(i, j) = \\
&= \sum_{i=0}^7 \sum_{j=0}^7 b_{ij}
\end{aligned} \tag{6.19}$$

όπου  $b_{ij} = \widehat{F}_1(j, i) \widehat{G}(i, j)$ .

Και τέλος, από την (6.11) εξάγονται οι παρακάτω σχέσεις:

$$\begin{aligned}
\delta_5 &= \sum_{i=0}^7 \sum_{j=0}^7 b_{ij} & \delta_6 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^j b_{ij} \\
\delta_7 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^i b_{ij} & \delta_8 &= \sum_{i=0}^7 \sum_{j=0}^7 (-1)^{i+j} b_{ij}
\end{aligned} \tag{6.20}$$

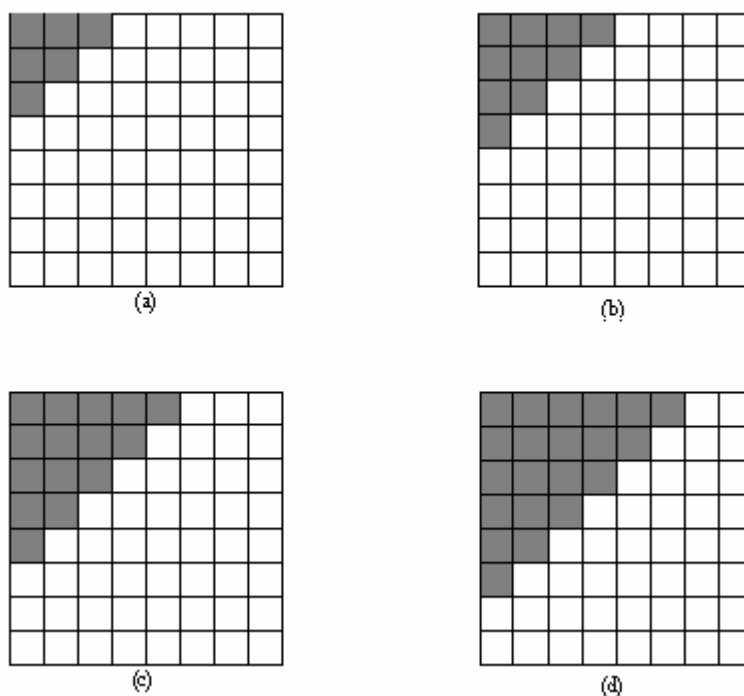
Από τις (6.18) και (6.20) παρατηρείται ότι οι περιττοί υπολογισμοί αφορούν τις ποσότητες  $a_{ij}$  και  $b_{ij}$ , κάτι το οποίο δεν είναι εύκολο να παρατηρηθεί στο πεδίο του χώρου. Με κατάλληλη διάταξη των συντελεστών  $a_{ij}$  και  $b_{ij}$  όπως δείχνει το σχήμα 6.1, όλοι οι περιττοί υπολογισμοί στο πεδίο του χώρου μπορούν να αποφευχθούν.



Σχ.6.1: Παράλληλη εφαρμογή των υπολογισμών του σφάλματος όλων των ισομετριών μιας τετραγωνικής περιοχής με τη χρήση όλων των συντελεστών του DCT. Οι μεταβλητές που βρίσκονται στο ίδιο τετράγωνο αθροίζονται μαζί. Παρατηρούμε ότι  $a_{00}=b_{00}=0$ .

Καθώς ο ΔΜΣ έχει μεγάλη πυκνότητα ενέργειας, σε πρακτικές καταστάσεις, μόνο ένα μέρος των συντελεστών του ΔΜΣ απαιτείται να υπολογιστούν για την εύρεση του βαθμωτού γινομένου. Συνεπώς, η υπολογιστική πολυπλοκότητα μπορεί να ελαττωθεί περισσότερο. Στο σχήμα 6.1 χρησιμοποιούνται και οι 64 συντελεστές του ΔΜΣ για τον υπολογισμό των βαθμωτών γινομένων των 8 μετασχηματισμών. Για τη βελτίωση της απόδοσης του αλγορίθμου

κωδικοποίησης, μπορούν να χρησιμοποιηθούν 4 διαφορετικές συναρτήσεις μάσκας. Στο τμήμα του τετραγώνου που καλύπτει η μάσκα τα *εικονοστοιχεία* διατηρούν την τιμή της έντασής τους, ενώ στο υπόλοιπο η ένταση είναι μηδενική. Όπως φαίνεται στο σχήμα 6.2, ο αριθμός των συντελεστών του ΔΜΣ που χρησιμοποιούνται για τον έλεγχο με τη μέθοδο των ελαχίστων τετραγώνων είναι 6, 10, 15 και 21 για κάθε ένα από τα 4 διαφορετικά φίλτρα αντίστοιχα. Παρατηρείται ότι ο συνεχής όρος των συντελεστών είναι 0, καθώς από κάθε τετραγωνική περιοχή έχει αφαιρεθεί η μέση τιμή των εντάσεων των *εικονοστοιχείων* της.

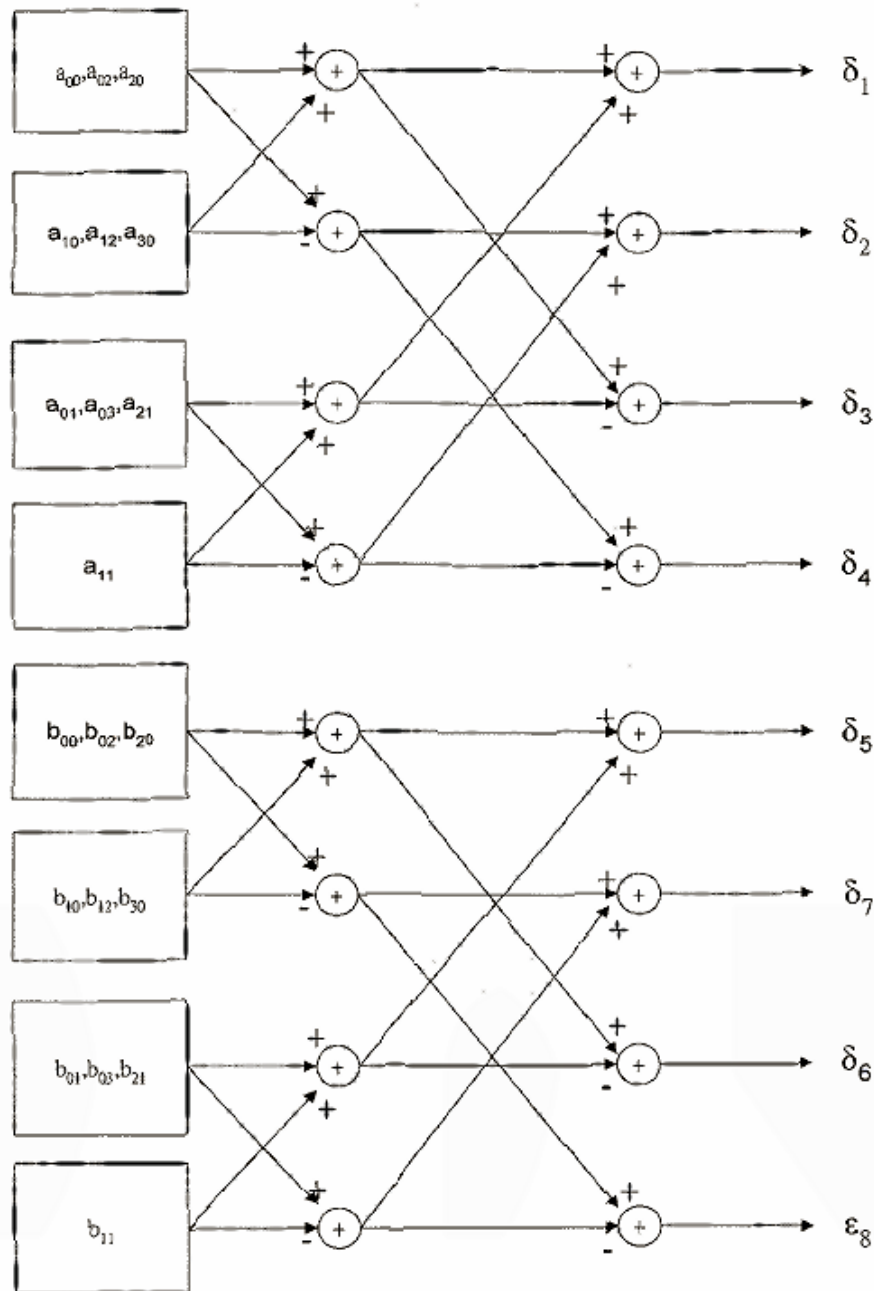


Σχ.6.2: Οι συναρτήσεις μάσκας που χρησιμοποιήθηκαν στα πειράματα που αφορούν τον παραπάνω αλγόριθμο. Η 1<sup>η</sup> περιέχει 6 συντελεστές, η 2<sup>η</sup> 10, η 3<sup>η</sup> 15 και η 4<sup>η</sup> 21.

Στο σχήμα 6.3 φαίνεται η μορφή που παίρνουν οι υπολογισμοί με τη χρήση συνάρτησης μάσκας, σε αντιδιαστολή με τους υπολογισμούς του σχήματος 6.1, ενώ στους πίνακες 6.1 και 6.2 συγκρίνεται ο αλγόριθμος των συγγραφέων με το συμβατικό fractal αλγόριθμο.

Πίνακας 6.1: Λίστα της διάρκειας κωδικοποίησης σε λεπτά και ο μέγιστος σηματοθορυβικός λόγος για την εικόνα της Lena, με τη χρήση τεσσάρων διαφορετικών συναρτήσεων μάσκας για τους DCT συντελεστές.

Lena	Χρόνος (sec)	PSNR (dB)
Βασικός fractal αλγόριθμος	22.42	28.90
Διακριτός Μετ/σμός Συνημιτόνου	6.46	28.93
Συνάρτηση μάσκας 4	3.80	28.18
Συνάρτηση μάσκας 3	3.05	27.96
Συνάρτηση μάσκας 2	2.60	25.43
Συνάρτηση μάσκας 1	2.15	21.90



Σχ.6.3: Παράλληλη εφαρμογή των υπολογισμών του σφάλματος όλων των ισομετριών μιας τετραγωνικής περιοχής με τη χρήση της  $2^{75}$  συνάρτησης μάσκας, δηλαδή μόνο οι 10 από τους 64 συντελεστές του DCT χρησιμοποιούνται για τους υπολογισμούς.

Πίνακας 6.2: Σύγκριση του χρόνου κωδικοποίησης μεταξύ του συμβατικού αλγόριθμου και του προτεινόμενου που χρησιμοποιεί στη μία περίπτωση όλους τους συντελεστές και στην άλλη την 4<sup>η</sup> συνάρτηση μάσκας.

	Συμβατική fractal συμπίεση		Διακριτός Μετ/σμός Συνημιτόνου		Συνάρτηση μάσκας 4	
	Χρόνος	PSNR	Χρόνος	PSNR	Χρόνος	PSNR
Lena	22.42	28.90	6.46	28.93	3.80	28.18
Μπαμπούinos	22.42	20.15	6.46	20.16	3.80	19.44
Αεριοθούμενο	22.42	25.50	6.46	24.54	3.80	24.39
Πιπεριές	22.42	29.86	6.46	29.86	3.80	29.27

## 6.2 Απαλλαγμένη από επαναλήψεις αποκωδικοποίηση, βασισμένη στον αποδοτικό σχεδιασμό του συνόλου των τύπου D περιοχών

Στην ενότητα αυτή, εξετάζεται διεξοδικά ο αλγόριθμος που προτείνουν οι Chang και Kuo<sup>2</sup>, σύμφωνα με τον οποίο ο κώδικας συμπίεσης που δημιουργείται, δεν απαιτεί την επαναληπτική εφαρμογή κάποιου μετασχηματισμού για την αποκωδικοποίησή του.

### 6.2.1 Ο τρόπος απαλλαγής από τις επαναλήψεις

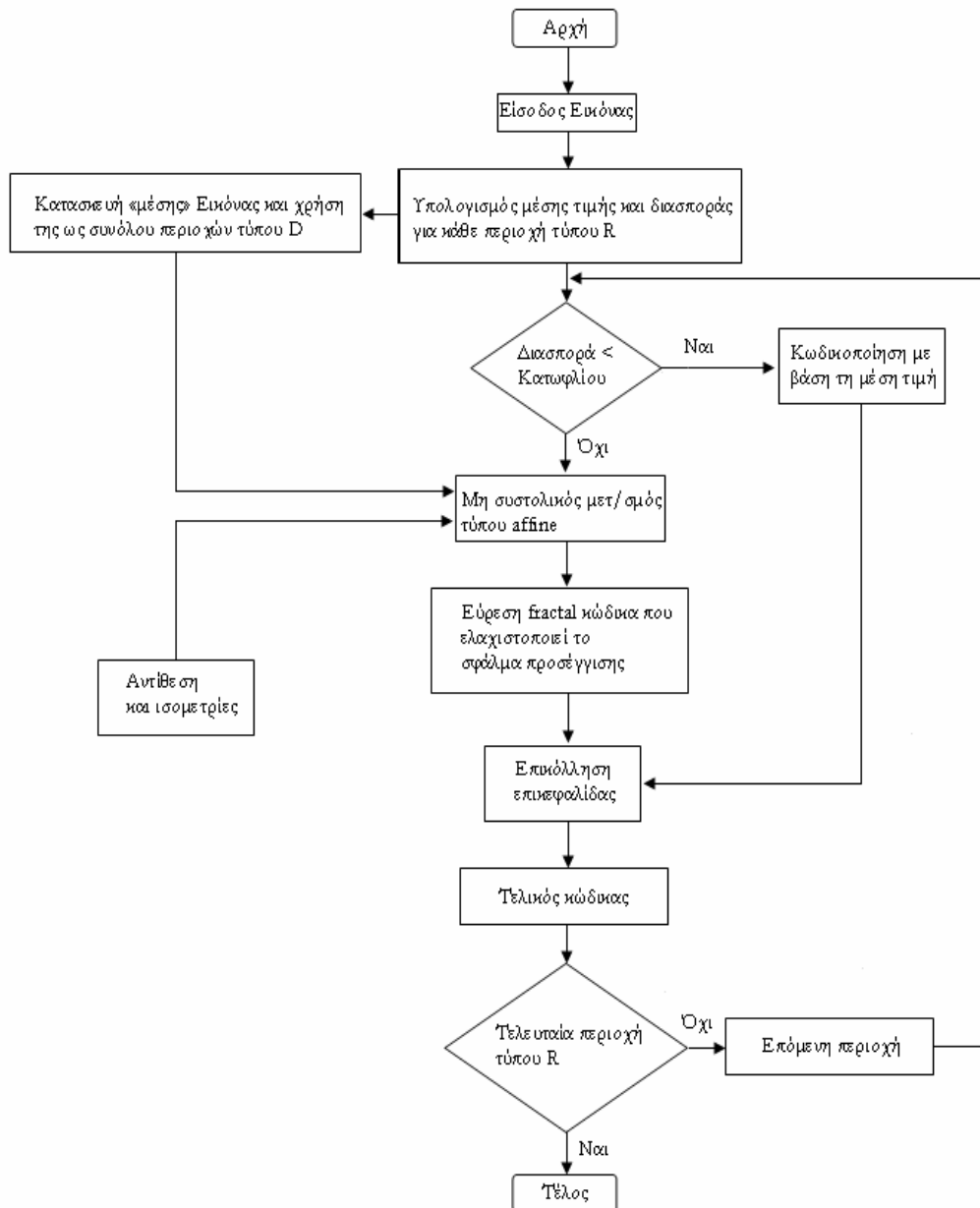
Για την ύπαρξη των ίδιων περιοχών τύπου D τόσο στην κωδικοποίηση όσο και κατά την αποκωδικοποίηση, χωρίς να να είναι απαραίτητο τα δεδομένα να μεταδίδονται off-line, οι συγγραφείς προτείνουν ένα αλγόριθμο που δε στηρίζεται σε επαναλήψεις και στον οποίο οι πληροφορίες για τις τύπου D περιοχές ενυπάρχουν στους κώδικες που χρησιμοποιεί. Επομένως, μέθοδοι όπως ο LBG αλγόριθμος (Κεφάλαιο 1) και ο μέσος όρος της έντασης των τετραγωνικών περιοχών (*block averaging*) μπορούν να εφαρμοστούν για τη μείωση των ομοιοτήτων μεταξύ των περιοχών τύπου D. Η προτεινόμενη κωδικοποίηση και αποκωδικοποίηση περιγράφονται παρακάτω.

### 6.2.1.1. Κωδικοποίηση

Το βασικό διάγραμμα ροής του προτεινόμενου αλγόριθμου φαίνεται στο σχήμα 6.4. Η προς επεξεργασία εικόνα διαστάσεων  $M \times M$  διαιρείται σε περιοχές τύπου  $R$  διάστασης  $B \times B$ . Πρώτα από όλα, υπολογίζεται διαδοχικά η μέση τιμή και η διασπορά της έντασης κάθε τύπου  $R$  περιοχής. Στη συνέχεια, όλες οι μέσες τιμές που έχουν βρεθεί, απαρτίζουν τη λεγόμενη «μέση» εικόνα διαστάσεων  $(M/B) \times (M/B)$ , της οποίας η τιμή της έντασης του κάθε *εικονοστοιχείου* αντιστοιχεί στη μέση τιμή μιας περιοχής τύπου  $R$ . Αν η διασπορά της τύπου  $R$  περιοχής:

$$\text{Var}(R) = \frac{1}{B^2} \sum_{0 \leq i, j \leq B} (r_{i,j} - \bar{r})^2 \quad (6.21)$$

(όπου το  $r_{i,j}$  δείχνει τη θέση  $(i,j)$  του *εικονοστοιχείου* της τύπου  $R$  περιοχής και  $\bar{r}$  η μέση τιμή της έντασης της τύπου  $R$  περιοχής), είναι μικρότερη από μια τιμή κατωφλίου  $E_{th}$ , τότε η περιοχή τύπου  $R$  κωδικοποιείται με την αποθήκευση του μέσου όρου. Διαφορετικά η παραπάνω περιοχή θα κωδικοποιηθεί με τη χρήση συστολικού μετασχηματισμού τύπου affine. Στην περίπτωση αυτή, η διάσταση της "μέσης" εικόνας πρέπει να είναι πολύ μεγαλύτερη από αυτή της τύπου  $D$  περιοχής, δηλαδή  $(M/B) \times (M/B) \gg B \times B$ , διαφορετικά η αντιστοίχιση μεταξύ τύπου  $R$  και της κατάλληλης τύπου  $D$  περιοχής καθίσταται δύσκολη καθώς το σύνολο των τύπου  $D$  περιοχών που εξάγεται από τη μέση εικόνα περιέχει λίγα στοιχεία. Σύμφωνα με τα παραπάνω, το μέγεθος της τύπου  $D$  περιοχής είναι το ίδιο με αυτό της τύπου  $R$  και επομένως δεν απαιτείται «συστολή» (μείωση της διάστασης), διαδικασία που ακολουθούν οι περισσότεροι συμβατικοί fractal αλγόριθμοι. Η διαδικασία συνεπώς, εγκαινιάζει μία νέα μορφή κωδικοποίησης με ένα νέο συστολικό μετασχηματισμό τύπου affine μεταξύ των τύπου  $R$  και των τύπου  $D$  περιοχών που δημιουργούνται από τη μέση εικόνα.



Σχ.6.4: Διάγραμμα ροής της κωδικοποίησης του προτεινόμενου αλγορίθμου.

Οι παράμετροι που χρησιμοποιούνται στο νέο αυτό μετασχηματισμό καθορίζονται ως εξής: Η φωτεινότητα αντικαθίσταται από τη μέση τιμή η οποία κωδικοποιείται με 6 ψηφία. Η αντίθεση είναι συνήθως μικρότερη της μονάδας για την αποφυγή της απόκλισης που οφείλεται στις επαναλήψεις των συμβατικών fractal αλγορίθμων. Παρ' όλα αυτά, είναι δυνατή και η χρήση τιμών αντίθεσης μεγαλύτερων της μονάδας καθώς στον υπό εξέταση αλγόριθμο δεν υπάρχουν επαναλήψεις. Άρα η αντίθεση κινείται σε ένα ευρύτερο πεδίο

τιμών. Στη συγκεκριμένη εργασία, η τιμή της αντίθεσης καθορίζεται από τη δοκιμή όλων των τιμών που ανήκουν στο σύνολο  $\{n/4, n=1,2,3,\dots,8\}$  για την εύρεση εκείνης που ελαχιστοποιεί την παραμόρφωση. Χρειάζονται συνεπώς 3 ψηφία για την αποθήκευσή της. Από την άλλη πλευρά, οι 8 ισομετρικοί μετασχηματισμοί της κάθε περιοχής τύπου R, είναι οι ίδιοι με αυτούς που χρησιμοποιούνται στους συμβατικούς αλγορίθμους και κωδικοποιούνται με τρία ψηφία. Ο νέος συστολικός μετασχηματισμός μπορεί να εκφραστεί από τον τύπο:

$$\hat{R} = \tau \{s \cdot D + \bar{r} - s \cdot \bar{d}\} = \tau \{s \cdot (D - \bar{d}) + \bar{r}\} \quad (6.22)$$

όπου  $\hat{R}$  είναι η κωδικοποιημένη τύπου R περιοχή και  $\bar{d}$  η μέση τιμή της έντασης της τύπου D περιοχής ( $\tau$ : ο ισομετρικός μετασχηματισμός). Παρατηρούμε ότι σε σχέση με τον τύπο που δίνει την κωδικοποιημένη τύπου R περιοχή στους τυπικούς fractal αλγόριθμους:

$$\hat{R} = \tau \{s \cdot S \circ D + o\} \quad (6.23)$$

η διαδικασία «συστολής» στο χώρο που εκφράζει ο τελεστής S, έχει αποφευχθεί και ότι ο όρος  $\bar{r} - s \cdot \bar{d}$  είναι ισοδύναμος με τη μεταβολή της φωτεινότητας  $o$ , όπως αναφέρθηκε στο 4<sup>ο</sup> Κεφάλαιο. Αφού εξεταστούν όλοι οι δυνατοί συνδυασμοί των παραμέτρων που περιέχει η παραπάνω εξίσωση, ο κώδικας καθορίζεται από το ελάχιστο δυνατό σφάλμα μεταξύ της αρχικής και της τελικής περιοχής R και  $\hat{R}$  αντίστοιχα. Το σφάλμα προσέγγισης μπορεί να γραφεί, όπως είναι γνωστό, με βάση το μέσο τετραγωνικό σφάλμα ως εξής:

$$\text{MSE}(R, \hat{R}) = \frac{1}{B^2} \sum_{0 \leq i, j \leq B} (r_{i,j} - \hat{r}_{i,j})^2 \quad (6.25)$$

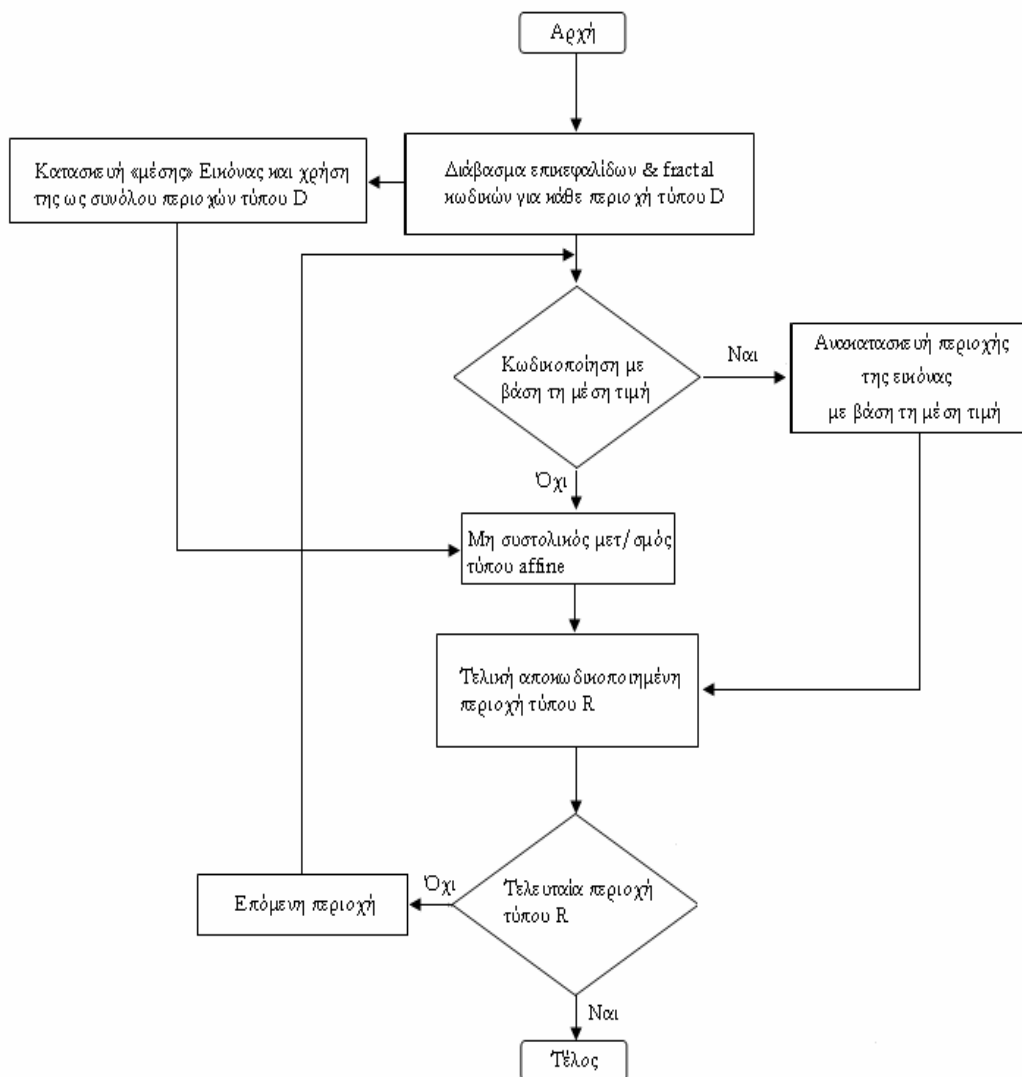
όπου το  $\hat{r}_{i,j}$  δείχνει την  $(i,j)$  θέση του *εικονοστοιχείου* στην κωδικοποιημένη τύπου R περιοχή. Τέλος, προστίθεται μια επικεφαλίδα για κάθε περιοχή τύπου R για να επισημάνει τον τρόπο κωδικοποίησής της (είτε αποθήκευση του μέσου όρου, είτε αποθήκευση της θέσης της τύπου D περιοχής και των παραμέτρων μετασχηματισμού). Με τον τρόπο αυτό, κάθε κωδικοποιημένη



τύπου  $R$  περιοχή, μπορεί να ανακατασκευαστεί σωστά, ανάλογα με την επικεφαλίδα.

### 6.2.1.2. Αποκωδικοποίηση

Το σχήμα 6.5 δείχνει το διάγραμμα ροής του αλγορίθμου αποκωδικοποίησης. Αρχικά, λαμβάνεται η κωδικοποιημένη πληροφορία και καθορίζεται αν η εικάστοτε περιοχή τύπου  $R$  έχει κωδικοποιηθεί με βάση τη μέση τιμή ή όχι, διαβάζοντας την επικεφαλίδα. Η «μέση» εικόνα κατασκευάζεται και κατά την αποκωδικοποίηση με βάση τις μέσες τιμές των περιοχών τύπου  $R$  που έχουν αποθηκευθεί είτε αυτούσιες είτε ως μέρος του affine μετασχηματισμού κατά την κωδικοποίηση. Είναι προφανές, ότι η μέση εικόνα που χρησιμοποιείται κατά την αποκωδικοποίηση είναι ταυτόσημη αυτής που κατασκευάστηκε κατά την κωδικοποίηση καθώς και οι δύο δημιουργήθηκαν από τις ίδιες μέσες τιμές. Συνεπώς οι περιοχές τύπου  $D$  που εξάγονται και από τις δύο εικόνες είναι επίσης ίδιες. Αν μια τύπου  $R$  περιοχή έχει κωδικοποιηθεί με βάση τη μέση τιμή έντασής της, τότε το κάθε *εικονοστοιχείο* της ανακατασκευασμένης έκδοσής της κατά την αποκωδικοποίηση, έχει τιμή έντασης ίση με τη μέση τιμή. Διαφορετικά, εκτελείται ο μετασχηματισμός που αναφέρθηκε για την ανακατασκευή της περιοχής τύπου  $R$ . Η αποκωδικοποίηση τελειώνει με την ανακατασκευή και της τελευταίας τύπου  $R$  περιοχής.



Σχ.6.5: Διάγραμμα ροής της αποκωδικοποίησης του προτεινόμενου αλγορίθμου.

Σε αυτόν τον αλγόριθμο, δεν απαιτούνται επαναλήψεις και για το λόγο αυτό δεν απασχολούμαστε με προβλήματα απόκλισης. Στους συμβατικούς αλγόριθμους, η διαδικασία αποκωδικοποίησης απαιτεί τη δέσμευση δύο θέσεων στη μνήμη, η 1<sup>η</sup> για την αρχική αυθαίρετη εικόνα και η 2<sup>η</sup> για αυτή που προκύπτει από την εφαρμογή του συστολικού μετασχηματισμού. Καθώς εφαρμόζεται περιοδικά ο μετασχηματισμός οι δύο θέσεις στη μνήμη ανανεώνονται με την 1<sup>η</sup> να περιέχει την εικόνα που δημιουργήθηκε με την προτελευταία επανάληψη και τη 2<sup>η</sup>, την εικόνα που δημιουργήθηκε με την τελευταία. Στον αλγόριθμο όμως που εξετάζεται εδώ, το μόνο που χρειάζεται

για την αποκωδικοποίηση είναι η «μέση» εικόνα που ανακατασκευάζεται από τις μέσες τιμές της έντασης των περιοχών τύπου R που έχουν ήδη αποθηκευθεί. Άρα η απαιτούμενη μνήμη είναι πολύ μικρότερη στην περίπτωση αυτή, σε σχέση με το χώρο που δεσμεύεται για την αποκωδικοποίηση στους συμβατικούς αλγορίθμους. Από την άλλη πλευρά, η ανυπαρξία επαναληπτικής διαδικασίας, συνεπάγεται τη δυνατότητα της παράλληλης αποκωδικοποίησης των τύπου R περιοχών. Η αρχιτεκτονική συνεπώς πολυπλοκότητα του αποκωδικοποιητή είναι εμφανώς χαμηλότερη από αυτή ενός συμβατικού fractal αποκωδικοποιητή ο οποίος απαιτεί επαναλήψεις. Για αυτό το λόγο και ο εν λόγω αποκωδικοποιητής είναι κατάλληλος για εφαρμογές υψηλής ταχύτητας.

### **6.2.2. Αποδοτικός σχεδιασμός του χώρου των τύπου D περιοχών**

Για να βρεθεί ένας αποδοτικός τρόπος δημιουργίας του χώρου των τύπου D περιοχών, χρησιμοποιείται κατ' αρχήν ο αλγόριθμος LBG και στη συνέχεια μια νέα μέθοδος που ονομάζεται μέσος όρος των περιοχών (*block averaging*).

#### **6.2.2.1. Σχεδιασμός με βάση τον αλγόριθμο LBG**

Οι αρχές του παραπάνω αλγορίθμου αναπτύχθηκαν στο 1<sup>ο</sup> Κεφάλαιο και επομένως δε θα ασχοληθούμε περαιτέρω με αυτόν.

Πρέπει να τονιστεί απλώς, ότι με την εφαρμογή του ομαδοποιούνται τα τύπου D διανύσματα, μειώνοντας με τον τρόπο αυτό τον αριθμό συγκρίσεων, καθώς σύμφωνα με όσα έχουν αναφερθεί σε προηγούμενο κεφάλαιο το κάθε τύπου R χαρακτηριστικό διάνυσμα συγκρίνεται με το αντιπροσωπευτικό τύπου D διάνυσμα κάθε ομάδας και στη συνέχεια λαμβάνουν χώρα περαιτέρω συγκρίσεις με διανύσματα της ομάδας που το προσεγγίζει βέλτιστα. Προφανώς λοιπόν, το σύνολο των τύπου D περιοχών (domain pool) που κατασκευάζει ο αλγόριθμος LBG είναι πιο αποδοτικό.

Μολαταύτα, η διαδικασία εύρεσης του βέλτιστου κωδικού βιβλίου απαιτεί έναν αριθμό επαναλήψεων ώστε το σφάλμα κβαντοποίησης να ελαχιστοποιηθεί. Όσο μεγαλύτερο είναι το μέγεθος του κωδικού βιβλίου τόσο μεγαλύτερος είναι και ο αριθμός των απαιτούμενων επαναλήψεων. Επειδή ο LBG εφαρμόζεται τόσο κατά την κωδικοποίηση όσο και κατά την αποκωδικοποίηση, η επαναληπτική διαδικασία για το μετασχηματισμό του συνόλου των τύπου D περιοχών εκτελείται δυο φορές. Αυτό το πλήθος υπολογισμών είναι ένα από τα κύρια μειονεκτήματα που παρουσιάζει η εφαρμογή του αλγορίθμου αυτού. Για να επιλυθεί το πρόβλημα αυτό παρουσιάζεται στη συνέχεια η μέθοδος του μέσου όρου των τετραγώνων (*block averaging*).

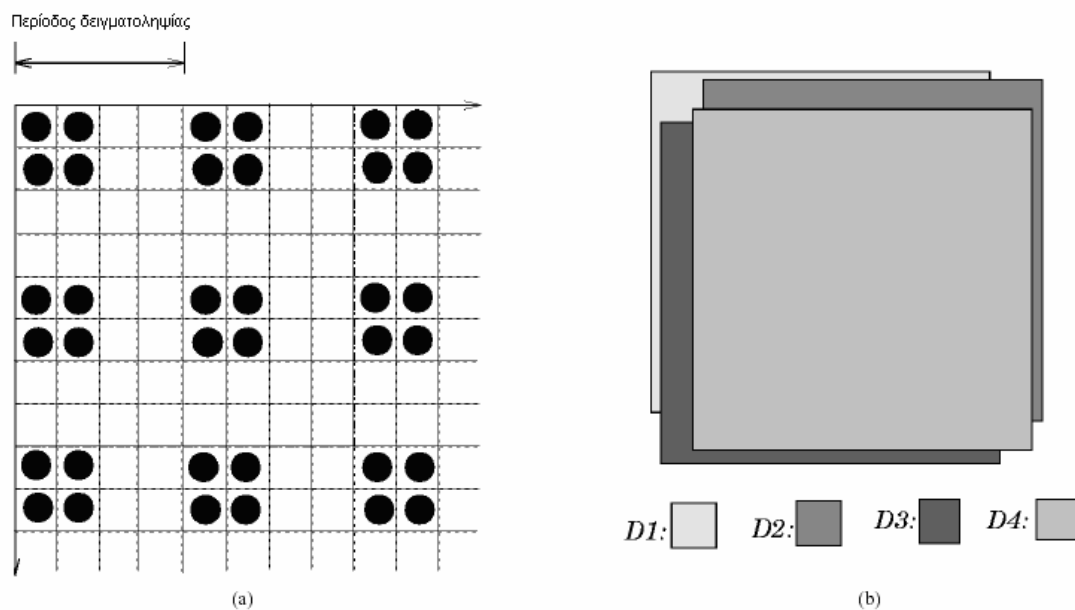
#### 6.2.2.2 Μέθοδος του μέσου όρου των τετραγωνικών περιοχών

Στη μέση εικόνα τα χαρακτηριστικά διανύσματα που επιλέγονται από τα γειτονικά τετράγωνα έχουν μεγάλες ομοιότητες μεταξύ τους γιατί ένα σημαντικό μέρος τους επικαλύπτεται. Μπορούν να θεωρηθούν ως μέλη της ίδιας ομάδας για τον αλγόριθμο LBG. Η μέθοδος του *block-averaging* που προτείνεται εδώ, βασίζεται σε αυτήν ακριβώς την έννοια της ομοιότητας που χρησιμοποιεί και ο LBG. Υπολογίζεται το κέντρο μάζας τεσσάρων τετραγώνων της μέσης εικόνας που είναι γειτονικά και επικαλύπτουν το ένα το άλλο μερικώς. Από τα τέσσερα αυτά τετράγωνα λαμβάνεται ένα νέο τύπου D τετράγωνο. Για να μειωθούν ακόμη περισσότερο οι ομοιότητες μεταξύ των παραγόμενων τύπου D περιοχών, μπορεί, αν αυτό είναι επιθυμητό, να υπολογιστεί το κέντρο μάζας περισσότερων από 4 τύπου D περιοχών. Είναι όμως αναμενόμενο, η συσχέτιση μεταξύ τεσσάρων γειτονικών τύπου D περιοχών να είναι υψηλότερη αυτής μεταξύ περισσότερων τύπου D περιοχών. Επομένως, οι συγγραφείς χρησιμοποιούν μόνο τέσσερα γειτονικά τετράγωνα στη μέθοδό τους. Το σχήμα 6.6(a) δείχνει κάποια σύνολα από 4 γειτονικά τετράγωνα στη μέση εικόνα με περίοδο δειγματοληψίας 4 *εικονοστοιχεία*. Σε αυτό το σχήμα, κάθε μαύρο σημείο αντιστοιχεί στην πάνω αριστερή γωνία ενός

τετραγώνου και χρησιμοποιείται για την αναπαράσταση ενός τετραγώνου διαστάσεων  $B \times B$ . Οι σχετικές θέσεις μεταξύ των τεσσάρων τύπου D περιοχών φαίνονται στο σχήμα 6.6(b). Η ένταση του *εικονοστοιχείου*  $d_{ij}$  του μέσου όρου  $D$  που προκύπτει από τις τέσσερες γειτονικές τύπου D περιοχές  $D_1, D_2, D_3, D_4$ , μπορεί να υπολογιστεί από τον τύπο:

$$d_{ij} = \frac{1}{4}(d1_{i,j} + d2_{i,j} + d3_{i,j} + d4_{i,j}), \quad 0 \leq i, j \leq B \quad (6.26)$$

, όπου τα  $d1_{i,j}-d4_{i,j}$  αναπαριστούν τις εντάσεις των *εικονοστοιχείων* στη θέση  $i,j$  για κάθε ένα από τα  $D_1-D_4$ . Επομένως το *εικονοστοιχείο* που υπολογίζεται με τον τρόπο αυτό, σχετίζεται με την πληροφορία που λαμβάνεται από τα τέσσερα γειτονικά *εικονοστοιχεία* των αρχικών περιοχών. Αντικαθιστώντας τις τελευταίες με το μέσο όρο, μειώνονται και οι περιττές συγκρίσεις μεταξύ μιας τύπου R περιοχής και των τεσσάρων αυτών παρόμοιων τύπου D περιοχών. Λαμβάνοντας τους μέσους όρους κάθε τετράδας τύπου D περιοχών κατασιευάζεται ένα σύνολο πολύ πιο αποδοτικό από αυτό που δημιουργείται απ' ευθείας από τη μέση εικόνα.



Σχ.6.6: (α) Μερικά σύνολα από τετράγωνα της εικόνας που χρησιμοποιούνται για την παραγωγή domain περιοχών με τη μέθοδο block averaging. (β) Οι σχετικές θέσεις τεσσάρων γειτονικών και μερικώς επικλυπτόμενων τετραγώνων της εικόνας D1-D4.

Οι τύπου D περιοχές επιλέγονται ομοιόμορφα από τις τύπου D περιοχές που έχουν προκύψει από το block averaging με μια περίοδο δειγματοληψίας  $T'$  στη μέση εικόνα. Έστω ότι ο αριθμός των στοιχείων του συνόλου των τύπου D περιοχών είναι  $N_D$ , τότε η περίοδος δειγματοληψίας τόσο στον οριζόντιο όσο και στον κάθετο άξονα είναι:

$$T' = \left[ \frac{M/B - B}{\sqrt{N_D} - 1} \right], \quad T' \geq 1 \quad (6.27)$$

Αντί του αλγορίθμου LBG, εδώ χρησιμοποιείται η αντιστοίχιση τεσσάρων περιοχών σε μία, το μέσο όρο, για την εξαγωγή των τύπου D περιοχών. Συνεπώς η απαιτούμενη υπολογιστική πολυπλοκότητα είναι πολύ μικρότερη από αυτή του αλγορίθμου LBG διατηρώντας με τον τρόπο αυτό την υψηλή ταχύτητα του αλγορίθμου χωρίς επαναλήψεις που έχουν προτείνει οι συγγραφείς.

Το σχήμα 6.8 είναι ενδεικτικό της αυξημένης επίδοσης της τεχνικής των Chang και Kuo σε σχέση με τους συμβατικούς fractal αλγόριθμους.



(a)



(b)

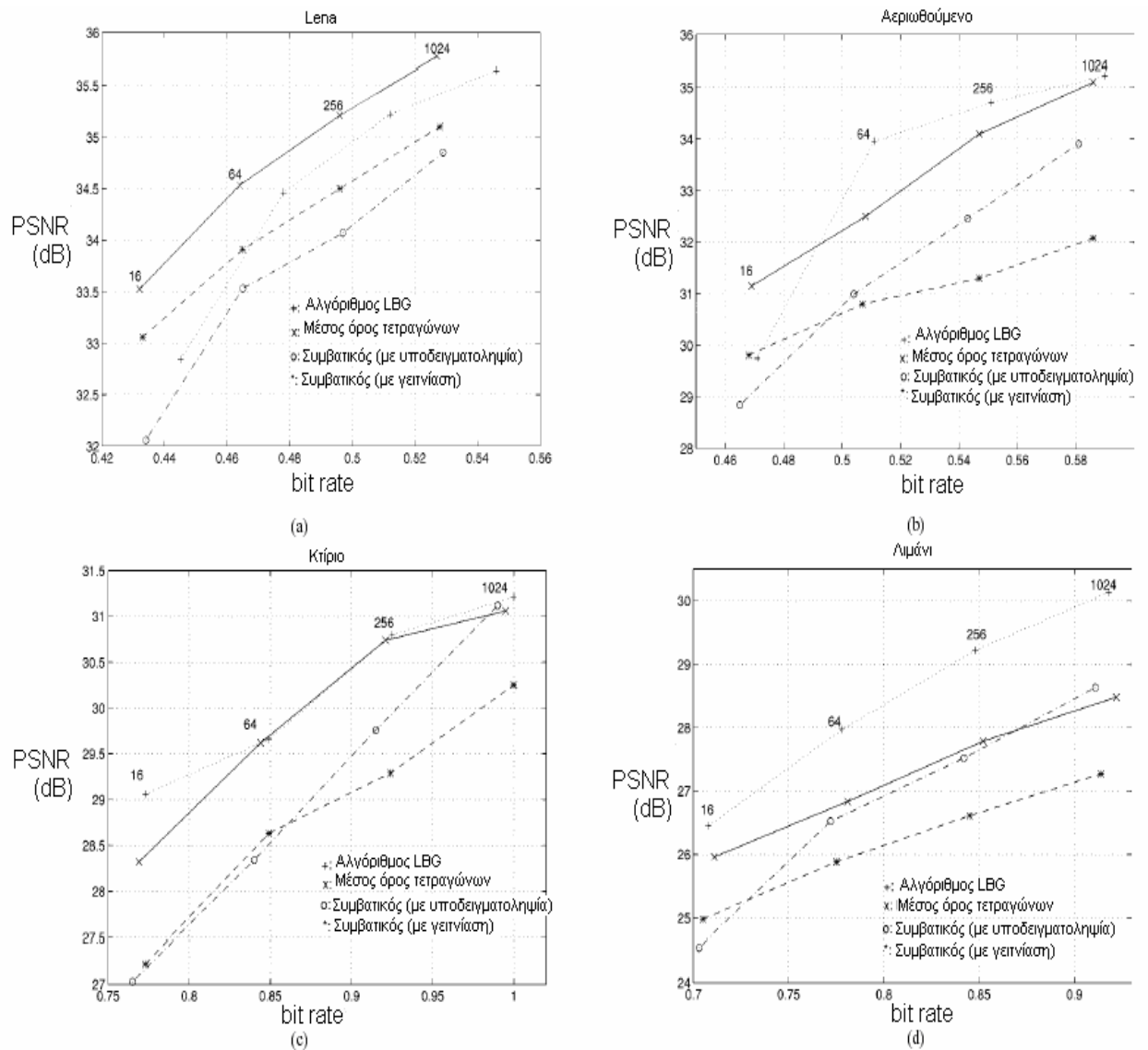


(c)



(d)

Σχ.6.7: Οι πειραματικές εικόνες που χρησιμοποιήσαν οι Chang, Κuo. (a) Lena, (b) αεριωθούμενο, (c) πύργο, (d) λιμάνι.



Σχ.6.8: Τα αποτελέσματα της κωδικοποίησης σύμφωνα με την απαλλαγμένη από επαναλήψεις τεχνική για τις εικόνες του προηγούμενου σχήματος.

### 6.3 Συμπεράσματα

Επιστρέφοντας στον πίνακα 2.2 του δεύτερου κεφαλαίου, παρατηρούμε ότι και το απλό αυτό παράδειγμα είναι ενδεικτικό του υψηλού βαθμού συμπίεσης που μπορούμε να επιτύχουμε με τη μέθοδο των fractal σε σχέση με άλλες τεχνικές συμπίεσης τις εικόνες, θυσιάζοντας φυσικά ένα μέρος της ποιότητας.



Πίνακας 6.3: Ο Σηματοθορυβικός λόγος , βαθμός και χρόνος συμπίεσης για την εικόνα της Lena. (Πίνακας 2.2)

Μέγεθος περιοχής Τύπου R	PSNR (dB)		Χρόνος Κωδικοποίησης (sec)	Βαθμός Συμπίεσης
	Σφάλμα Προσέγγισης (collage)	Σφάλμα Μετ/σμού (attractor)		
4*4	36.96	36.66	147.48	4.4
8*8	31.15	31.27	69.93	17.7
16*16	27.02	26.89	59.61	70.5
32*32	23.55	23.32	54.76	281.0

Το βαθμό που απέχει σε πιστότητα η μία αποκωδικοποίηση από την άλλη για διαφορετικά μεγέθη των τύπου R περιοχών μπορούμε να τον δούμε και από διαφορετική σκοπιά, αν από τη σχέση (1.4) που δίνει το μέγιστο σηματοθορυβικό λόγο, υπολογίσουμε το λόγο των αποστάσεων  $d_1, d_2$  δύο διαφορετικών αποκωδικοποιημένων εικόνων  $f_1, f_2$ , που προέρχονται από τη συμπίεση της ίδιας αρχικής εικόνας  $f$ , με βάση τη σχέση:

$$\frac{d_1}{d_2} = 10^{\frac{PSNR_2 - PSNR_1}{20}} \quad (6.28)$$

Για παράδειγμα, μπορούμε να υπολογίσουμε τους λόγους των αποστάσεων  $d_2, d_3, d_4$  των τριών άλλων αποκωδικοποιήσεων 8\*8, 16\*16, 32\*32 του πίνακα, προς την απόσταση  $d_1$  της πρώτης. Χρησιμοποιώντας τις τιμές της αριστερής στήλης για το PSNR βρισκουμε ότι το σφάλμα της 2<sup>ης</sup> είναι 1.95 φορές μεγαλύτερο του αντίστοιχου της πρώτης, το σφάλμα της 3<sup>ης</sup> 3.14 φορές μεγαλύτερο, ενώ της 4<sup>ης</sup> 4.6. Αν αντικαταστήσουμε εξάλλου την τιμή του PSNR για τη διαμέριση 4\*4 στη σχέση (1.4), βρισκουμε ότι η Ευκλείδεια απόσταση μεταξύ της αρχικής και της αποκωδικοποιημένης  $d_1$ , είναι περίπου 3.5, πολύ μικρότερη της  $d_4$  που ξεπερνά το 16.

Αν ο στόχος μας λοιπόν είναι η ποιότητα και όχι τόσο ο βαθμός συμπίεσης, θα προτιμήσουμε την πρώτη διαμέριση. Παρατηρούμε όμως από τον πίνακα, ότι ο χρόνος κωδικοποίησης είναι αρκετά μεγάλος σε σχέση μάλιστα και με άλλες τεχνικές, όπως η JPEG, η οποία για χαμηλούς βαθμούς συμπίεσης όπως

αυτός (περίπου 4), ο οποίος είναι μικρότερος της δεκάδας, είναι γενικά προτιμότερη των fractal. Ας υποθέσουμε ότι για μία υψηλής ποιότητας συμπύεση διαμερίζουμε μία  $512*512$  εικόνα σε τύπου R περιοχές  $4*4$  με αποτέλεσμα τη δημιουργία ενός συνόλου που αριθμεί  $128*128=16384$  στοιχεία. Αφού το ζητούμενο είναι η καλύτερη δυνατή προσέγγιση, υποθέτουμε ότι το  $8*8$  τύπου D παράθυρο «σαρώνει» την εικόνα με βήμα μόλις ενός *εικονοστοιχείου* στην οριζόντια και κάθετη διεύθυνση, για τη δημιουργία του συνόλου των τύπου D περιοχών που περιέχει, συνυπολογισμένων και των ισομετριών  $8*(512-8+1)*(512-8+1)=8*505*505=2040200$  στοιχεία. Ο απλός αλγόριθμος πλήρους αναζήτησης πρέπει να εκτελέσει τη μέθοδο των ελαχίστων τετραγώνων  $16384*2040200 = 33426636800$  φορές. Κάτι τέτοιο ακόμη και με την αύξηση της ταχύτητας των προσωπικών υπολογιστών την τελευταία δεκαετία, είναι αρκετά χρονοβόρο.

Η ανεξαρτησία όμως της fractal συμπύεσης από την ανάλυση της εικόνας, που τα πλεονεκτήματά της αναφέρθηκαν στο 2<sup>ο</sup> κεφάλαιο, είναι μία επιθυμητή ιδιότητα για τη μετάδοση εικόνων μέσω του Διαδικτύου, καθώς διαφορετικοί δέκτες ενδέχεται να έχουν και διαφορετικές αναλύσεις. Οι εικόνες κωδικοποιούνται μία φορά αλλά μπορούν να αποκωδικοποιηθούν περισσότερες, σε διαφορετικές αναλύσεις. Μία άλλη επιθυμητή ιδιότητά τους, είναι υψηλή επίδοσή τους στην αναπαράσταση φυσικών εικόνων που περιέχουν μεγάλη λεπτομέρεια, κάτι που δε μας ξενίζει γιατί η πηγή έμπνευσης των PIFS ήταν οι ομοιότητες που παρουσιάζουν μεταξύ τους οι περιοχές τις ίδιες φυσικής εικόνας. Μπορούμε επίσης να προσθέσουμε, ότι ο τρόπος με τον οποίο οι διάφορες εικόνες «φορτώνονται» στους χρήστες του Διαδικτύου είναι παρόμοιος με τη διαδικασία της fractal αποκωδικοποίησης.

Για το λόγο αυτό, οι υποστηρικτές της μεθόδου των PIFS, αφιέρωσαν το χρόνο τους στην προσπάθεια εύρεσης τρόπων μείωσης της διάρκειας συμπύεσης. Στην εργασία αναφέρθηκαν διάφορες στρατηγικές που αντλούσαν

την έμπνευσή τους από τη Γεωμετρία, για την αποδοτικότερη διαμέριση π.χ. της εικόνας, την Άλγεβρα και την Επιστήμη Υπολογιστών για μία διαφορετική αντιμετώπιση του προβλήματος ελάχιστων τετραγώνων, την Επεξεργασία Σήματος με την εφαρμογή μετασχηματισμών, κ.ά. Μελετήθηκαν επιπλέον, υβριδικές τεχνικές που συνδυάζαν τα αποτελέσματα και τη θεωρία άλλων μεθόδων όπως η Διανυσματική Κβαντοποίηση και Ο Μετασχηματισμός Wavelet, με τη θεωρία των fractal.

Τα πειραματικά αποτελέσματα που περιέχονται στην εργασία, διαφέρουν από συγγραφέα σε συγγραφέα ως προς το χρόνο συμπίεσης όχι μόνο εξαιτίας των διαφορετικών μεθόδων που ο καθένας χρησιμοποιεί, αλλά κυρίως επειδή οι δυνατότητες των υπολογιστικών συστημάτων που χρησιμοποιήθηκαν για την εξαγωγή τους ήταν διαφορετικές. Όλα όμως αποδεικνύουν, ότι οι παραπάνω μέθοδοι που εξετάστηκαν στην παρούσα εργασία κατόρθωσαν να μειώσουν σε ένα βαθμό το κόστος και το χρόνο υπολογισμού, διατηρώντας παράλληλα την ποιότητα σε υψηλά για τη συμπίεση με fractal, επίπεδα. Δεν έχουν καταφέρει ωστόσο να ανυψώσουν τα PIFS στο επίπεδο άλλων ευρέως γνωστών μεθόδων συμπίεσης, όπως για παράδειγμα η JPEG, αφού το πρόβλημα του χρόνου συμπίεσης εξακολουθεί να παραμένει ως ένα από τα σημαντικότερα εμπόδια εξάπλωσής τους.

## Βιβλιογραφία

1. Trieu-Kien Truong, Jyh-Horng Jeng, Irving S. Reed, P. C. Lee, Alan Q. Li, *A Fast Encoding Algorithm for Fractal Image Compression using The DCT Inner Product*, *IEEE Transactions on Image Processing* Vol. 9 No. 4 April 2000.
2. Hsuan , T. Chang, Chung J. Kuo, *Iteration – Free Fractal Image Coding Based on Efficient Domain Pool Design*, *IEEE Transactions on Image Processing* Vol. 9, No. 3, March 2000.

---

## ΠΑΡΑΡΤΗΜΑΤΑ

---

### Παράρτημα Α: Οι τυπικές ασπρόμαυρες δοκιμαστικές εικόνες

Στο παράρτημα αυτό παρουσιάζουμε τις συνηθέστερες gray-scale εικόνες που χρησιμοποιούνται από τους ερευνητές για την αξιολόγηση της επίδοσης των διαφόρων τεχνικών συμπίεσης και οι οποίες αναφέρονται στην εργασία αυτή.

Οι πιο σημαντικές από αυτές είναι η *Lena*, η εικόνα *Golden Hill*, η *Plava Laguna* και η εικόνα του *Μπαμπούνου (Mandrill)*.

Η πιο διαδεδομένη είναι η εικόνα της *Lena* (σχήμα Α.1), που ήταν και μία από τις πρώτες εικόνες που υπέστησαν ψηφιακή επεξεργασία. Αν και θεωρείται εύκολη στην επεξεργασία υπάρχουν μέρη της εικόνας που δυσκολεύουν πολλές τεχνικές συμπίεσης. Το πιο σημαντικό τμήμα της είναι τα μάτια, συμπεριλαμβανομένων της κόρης του άκρου του οφθαλμού και της βλεφαρίδας, που περιβάλλονται από τις ομοιόμορφες περιοχές του υπόλοιπου προσώπου. Η επιστήμονες επιλέγουν την εικόνα αυτή γιατί είναι εύκολη αλλά όχι και ασήμαντη.



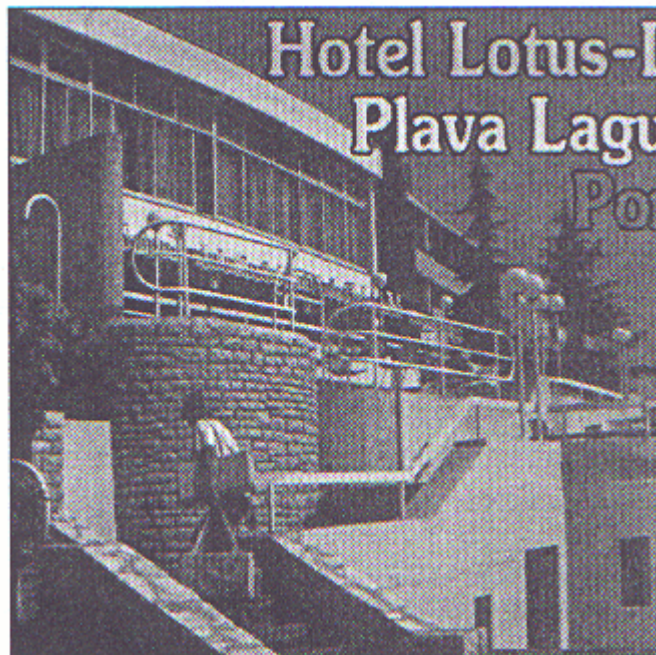
Σχ.Α.1:Lena



Σχ.Α.2: Golden Hill

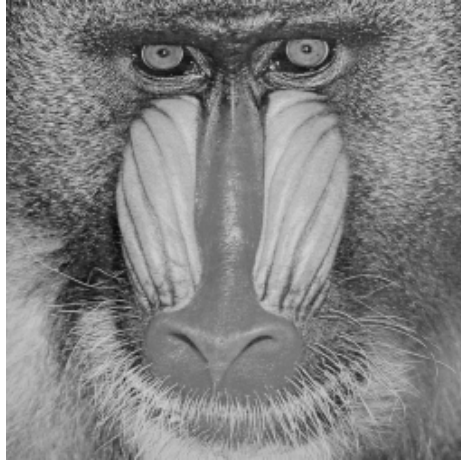
Η εικόνα *Golden Hill* (σχήμα Α.2) είναι μια πολύπλοκη φωτογραφία καθώς περιέχει πολλά μικρά αλλά αναγνωρίσιμα αντικείμενα, γεγονός που καθιστά δύσκολη την επιτυχημένη τη συμπίεσή της.

Το ξενοδοχείο *Lotus-Lila* στην *Plava Laguna* (σχήμα Α.3), είναι μία εικόνα που περιέχει πολλά αντικείμενα που κατασκευάστηκαν από τον άνθρωπο. Αποτελεί μία καλή εικόνα για την αντιμετώπιση περίπλοκων γεωμετρικών σχημάτων κατά την επεξεργασία της.



Σχ.Α.3: Plava Laguna

Ο Μπαμπουΐνος (Mandrill, σχήμα Α.4), είναι μία εικόνα με ποικίλες διαβαθμίσεις του γκρι. Η εικόνα αυτή έχει αποδειχθεί ιδανική για την αξιολόγηση της οπτικής ποιότητας μίας τεχνικής συμπίεσης εξαιτίας της συνύπαρξης μέσα σε αυτήν στοιχείων με μεγάλες αντιθέσεις.



Σχ.Α.4: Mandrill

Εκτός από τις παραπάνω πειραματικές εικόνες υπάρχουν και άλλες ευρέως διαδεδομένες ανάμεσα στους ερευνητές όπως για παράδειγμα οι *Βάρκες (Boats)*, οι *Πιπεριές (Peppers)* κ.ά.



Σχ.Α.5: Boats



Σχ.Α.6: Peppers

## Παράρτημα Β: Ο fractal αλγόριθμος με τη μορφή ψευδοκώδικα

Στο παράρτημα που ακολουθεί, παρουσιάζεται ένας τρόπος υλοποίησης του απλού fractal αλγορίθμου τόσο για την κωδικοποίηση, όσο και για την αποκωδικοποίηση. Οι δύο ψευδοκώδικες χρησιμοποιούν ψευδο-εντολές άμεσα επηρεασμένες από το περιβάλλον εργασίας του MATLAB, καθώς το τελευταίο αποτελεί ένα εύχρηστο εργαλείο για την επεξεργασία εικόνας. Τόσο το πρόγραμμα κωδικοποίησης, όσο και το πρόγραμμα αποκωδικοποίησης έχουν εμπνευστεί από το MATLAB. Για το λόγο αυτό και τα σχόλια που συνοδεύουν κάθε βήμα εκτέλεσης των δύο προγραμμάτων αναφέρονται σε συναρτήσεις και εντολές που διαθέτει το συγκεκριμένο περιβάλλον.



## B.1 Η κωδικοποίηση

1. Επιλογή της μίας διάστασης  $r$  της περιοχής τύπου  $R$ .
2. Επιλογή της περιόδου δειγματοληψίας έστω  $sample$  για το σχηματισμό του συνόλου των  $domain$  περιοχών.
3. Επιλογή των ψηφίων έστω  $b_1$  και  $b_2$  που δεσμεύονται για την αναπαράσταση των διακριτών τιμών της αντίθεσης και φωτεινότητας. Οι τιμές αυτές μπορούν να αποθηκευθούν σε δύο πίνακες γραμμή  $Sd$  και  $Od$  κάθε ένας από τους οποίους θα περιέχει  $2^{b_1}$  και  $2^{b_2}$  στοιχεία αντιστοίχως. Για παράδειγμα, αν η αντίθεση διακριτοποιείται με 5 ψηφία, τότε μπορεί να λάβει 32 διαφορετικές τιμές, η μικρότερη από αυτές είναι το  $-15/16$  οπότε μπορούμε, δημιουργώντας έναν επαναληπτικό βρόχο να αποθηκεύσουμε στον  $Sd$  τις τιμές αυτές (π.χ., για  $i=0:31 \quad Sd[i]=-15/16+i*1/32$ ).
4. Διάβασμα της εικόνας προς κωδικοποίηση  $I1$  (διαδικασία που εκτελείται ευκολότερα σε περιβάλλον Matlab, με την εντολή  $Imread()$ , σε σχέση με άλλες γλώσσες προγραμματισμού όπως η C για παράδειγμα).
5. Διαμέριση της εικόνας σε τετράγωνα  $r*r$ , μετατροπή τους σε διανύσματα και αποθήκευσή τους σε πίνακα  $R$  (πίνακας διανυσμάτων). Στο Matlab αυτό γίνεται με την εντολή  $im2col$ .
6. Χωρισμός της εικόνας σε επικαλυπτόμενα τετράγωνα  $2r*2r$  με περίοδο δειγματοληψίας 1 pixel (οριζόντια και κάθετα), μετατροπή τους σε διανύσματα και αποθήκευσή τους σε πίνακα έστω  $D1$ .
7. Εύρεση της μέγιστης διάστασης του  $D1$ ,  $dmax$  (π.χ., με την εντολή  $size()$  του Matlab).
8. Αρχικοποίηση ενός μετρητή  $k=1$  που σημειώνει την πρώτη  $domain$  περιοχή του συνόλου των  $domains$  που θα σχηματίσουμε.
9. Όσο το  $k < dmax$ , τότε:

- a) μετατροπή του διανύσματος  $D1[k]$  σε τετράγωνο (μία τέτοια εντολή στο Matlab είναι η `vec2mat()`) και εύρεση των ισομετρικών μετασχηματισμών  $T_p$ ,  $p=1,\dots,8$  και αποθήκευσή τους σε πίνακα  $T$ . Αν  $T1$  το αρχικό τετράγωνο τότε οι 2, 3, 4, 5, 8 μπορούν να βρεθούν με την ακόλουθη υπορουτίνα:

```

for i=1:2*r
    for j=1:2*r
        m=2*r+1-j;T2(i,j)=T1(i,m);
        n=2*r+1-i;T3(i,j)=T1(n,j);
        T4(i,j)=T1(n,m);
        T5(i,j)=T1(m,n);
        T8(i,j)=T1(j,i);
    end
end

```

Οι ισομετρικοί μετασχηματισμοί 6 και 7 οι οποίοι θα μπορούσαν με τη σειρά τους να βρεθούν με παρόμοιο τρόπο, στο Matlab βρίσκονται με μία απλή εντολή την `imrotate()` ( $T6=imrotate(T1,90)$ ;  $T7=imrotate(T1,270)$ ), καθώς είναι στροφές του αρχικού. Βέβαια, μία πιο εύκολη λύση για τον υπολογισμό και των υπολοίπων, θα ήταν η εύρεση του ειδώλου του  $T1$  και στροφή του, χωρίς την εμπλοκή σε πρόσθετους υπολογισμούς.

- b) Χωρισμός του πίνακα  $T$  σε πίνακες  $2 \times 2$  και υπολογισμός του μέσου όρου των στοιχείων του καθενός (subsampling). Το Matlab διαθέτει εντολές για την παραπάνω εργασία, όπως η `blkproc()`.
- c) Μετατροπή του πίνακα  $X$  που προκύπτει, σε πίνακα διανυσμάτων, έστω  $D_m$  (ο οποίος προφανώς περιέχει 8 διανύσματα διάστασης  $r \times r$ ), και αποθήκευσή του ως στοιχείο ενός νέου πίνακα γραμμής, έστω  $d$ . Για να το πετύχουμε αυτό,

μετατρέπουμε τον  $D_m$  σε διάνυσμα έστω  $y$ , το οποίο μπορεί να αποθηκευθεί ως στοιχείο του  $d$  και εισάγουμε ένα νεό μετρητή, έστω  $count$ , ο οποίος εξαρτάται από τον αρχικό  $k$  με βάση τις εξής σχέσεις:  $l=(k-1)/sample$ ;  $count=l+1$ ; Ο  $count$  δηλαδή είναι ένας μετρητής που μετρά οκτάδες και με την αύξησή του μία νέα οκτάδα (το ελάχιστο διάνυσμα  $y$ ) εισάγεται στην  $count$  στήλη του πίνακα  $d$ , καθώς  $k=k+sample$ ; και ο βρόχος κλείνει.

10. Σχηματισμός της domain pool, έστω  $D$  με τη μετατροπή του πίνακα  $d$  που περιέχει 8άδες διανυσμάτων σε πίνακα που περιέχει απλά διανύσματα δηλαδή πίνακα διαστάσεων  $r^2*(8*count)$ . Στο Matlab η αλλαγή διαστάσεων του πίνακα μπορεί να γίνει με την εντολή `reshape()`.
11. Εύρεση της μέγιστης διάστασης του  $R$ , έστω  $rmax$ .
12. Κατασκευή του μοναδιαίου διανύσματος  $I$  (π.χ., for  $i=1: r^2 \rightarrow I[i]=1$ );).
13. Επαναληπτική εφαρμογή των υπολογισμών για την προσέγγιση δηλ.

for  $i=1:rmax$

for  $j=1:8*count$

- a) Υπολογισμός των  $s(i,j)$ ,  $o(i,j)$  από τη (2.24). Σύγκρισή τους με τα στοιχεία των πινάκων  $S_d$ ,  $O_d$  και εύρεση των αντίστοιχων διακριτών τιμών  $sd(i,j)$  και  $od(i,j)$ . Εφαρμογή του μετασχηματισμού στο διάνυσμα  $D[j]$  και αποθήκευση του νέου διανύσματος στη στήλη  $j$  πίνακα γραμμής έστω  $R_d$ .
- b) Αποθήκευση της απόστασης μεταξύ των διανυσμάτων  $R[i]$ ,  $R_d[j]$  στην  $j$  στήλη ενός πίνακα γραμμής  $Error$ .
- c) Κλείνουμε τον εσωτερικό βρόχο και ανοίγουμε 2°. for  $p=1:(8*count) \rightarrow$

- d) Εύρεση της θέσης του μικρότερου στοιχείου του `Error` και αποθήκευσή του `p` στην `i` στήλη ενός πίνακα γραμμής π.χ., με την εντολή του Matlab `if Error(:,p)==min(Error) → Adress[i]=p` (η domain περιοχή αντιστοιχίζεται στην `i range`).
- e) Αποθήκευση τέλος των `sd(i,p)=S[i]`, `od(i,p)=O[i]` σε δύο πίνακες γραμμή για την αντίθεση και τη φωτεινότητα.

#### 14. Τερματισμός του αλγορίθμου.

## B.2 Η αποκωδικοποίηση

Στη συνέχεια παραθέτουμε σε ψευδοκώδικα τον αλγόριθμο της αποκωδικοποίησης: (Υποθέτουμε ότι από το πρόγραμμα κωδικοποίησης που παρατέθηκε μόλις προηγουμένως, διατηρούμε την ίδια περίοδο δειγματοληψίας `sample`, τη διάσταση `r` της περιοχής τύπου `R` και φυσικά τους πίνακες `Adress`, `S` και `O`, οι οποίοι αποτελούν και τη συμπιεσμένη πληροφορία):

1. Επιλογή του αριθμού των επαναλήψεων, έστω `IN` και της μέγιστης ανεκτής παραμόρφωσης (σφάλμα), έστω `E`.
2. Αρχικοποίηση μετρητή `icount=0`. Όσο `Error>E` ή `i<IN` τότε:
  - a) Διάβασμα της αρχικής εικόνας `Io`.
  - b) Χωρισμός της `Io` σε τετράγωνο `r*r` και αποθήκευση των `range` διανυσμάτων σε πίνακα `R` όπως και στο προηγούμενο ψευδοκώδικα.
  - c) Με τον ίδιο τρόπο που αναφέραμε κατά τη διαδικασία κωδικοποίησης, κατασκευή του συνόλου των περιοχών τύπου `D` μετά από υποδειγματοληψία (συνυπολογίζοντας και τις ισομετρίες) και αποθήκευση όλων των domain διανυσμάτων σε πίνακα `D`.
  - d) Εύρεση των μέγιστων διαστάσεων του `R`, `rmax`.
  - e) Ανακατασκευή της εικόνας: Για `i=1:rmax` εκτελούμε:

- I. Εύρεση διεύθυνσης της αντίστοιχης domain περιοχής, για παράδειγμα θέτουμε  $ad=Address[i]$ .
- II. Εφαρμογή του συστολικού μετασχηματισμού.  
 $Rrec[i]=S[i]*D[ad] + O[i]$ .
- III. Κλείσιμο βρόχου. Ο βρόχος αυτός σχηματίζει ένα πίνακα  $Rrec$  που περιέχει τα ανακατασκευασμένα range διανύσματα, τον οποίο μπορούμε να μετατρέψουμε σε πίνακα με τις αρχικές διαστάσεις της εικόνας έστω  $I1$  αν για παράδειγμα χρησιμοποιήσουμε την εντολή `reshape` ή `col2im` του Matlab.
  - f) Μετατροπή του πίνακα  $Rrec$  σε πίνακα διαστάσεων εικόνας  $I1$ .
  - g)  $Error=distance(Io, I1)$ , όπου ως συνάρτηση απόστασης `distance` χρησιμοποιούμε συνήθως την Ευκλείδεια.
  - h) Αντικατάσταση του  $Io$  από το  $I1$  δηλ.  $Io=I1$  ώστε ο νέος κύκλος υπολογισμών να εφαρμοσθεί στη νέα εικόνα που προέκυψε, τη  $I1$ .
  - i) Αύξηση τέλους του μετρητή επαναλήψεων,  $icount=i+1$ .
3. Στο τέλος το πρόγραμμα επιστρέφει την εικόνα  $Io$ , π.χ., με την εντολή στο Matlab `imshow(Io)`.

Πρέπει να τονίσουμε εδώ, ότι οι δύο ψευδοκώδικες που περιέχονται στην εργασία είναι ενδεικτικοί και σε καμία περίπτωση δεν αποτελούν εγχειρίδιο για τη συμπίεση και αποσυμπίεση της εικόνας με τη μέθοδο των fractal. Άλλωστε επιδέχονται πολλές βελτιώσεις, μία από τις οποίες είναι και η δέσμευση τριών ψηφίων για την αναπαράσταση των ισομετρικών μετασχηματισμών. Αυτή η ενέργεια δεν επηρεάζει το πρόγραμμα συμπίεσης που διατυπώσαμε εδώ, καθώς το τελευταίο έχει ενσωματώσει τις ισομετρίες στο σύνολο των περιοχών τύπου  $D$  που αριθμεί οκταπλάσια στοιχεία, ενώ θα μπορούσε να περιέχει αυτούσιες (χωρίς δηλαδή τους μετασχηματισμούς) τις περιοχές και να υπολογίζει τον ισομετρικό μετασχηματισμό εντός του βρόχου, καλώντας 8 υπορουτίνες (μία για τον καθένα) και αποθηκεύοντας για κάθε περιοχή τύπου  $R$  το δείκτη της υπορουτίνας που πρέπει να καλέσει κατά την αποκωδικοποίηση. Ο χρόνος

συμπίεσης δεν μεταβάλλεται από μία τέτοια αλλαγή, η μείωση όμως του χρόνου εκτέλεσης της αποκωδικοποίησης όπως την έχουμε διατυπώσει εδώ θα ήταν δραστική, γιατί σε κάθε επανάληψη θα εργαζόμασταν με ένα σύνολο, έστω  $N_D$  τον αριθμό περιοχών τύπου  $D$ , ενώ εδώ σχηματίζουμε κάθε φορά ένα οκταπλάσιο σύνολο.

## Παράρτημα Γ: Η απόδειξη της σχέσης (4.15)

Στο Παράρτημα αυτό αποδεικνύεται η σχέση (4.15)  $d(\tilde{W}(P), O) \leq d(W(P), O)$  στην επαναληπτική εφαρμογή της οποίας, στηρίχθηκε η απόδειξη του Γενικευμένου Θεωρήματος Collage για ένα PIFS  $\tilde{W}$  που χρησιμοποιεί μέσες τιμές. Η σχέση θα αποδειχθεί για μία τύπου  $R$  περιοχή  $R$  της αρχικής προς συμπίεση εικόνας  $O$ , οπότε η απόδειξη μπορεί να επεκταθεί και για όλη την εικόνα  $O$ .

Όπως αναφέρθηκε στο Κεφάλαιο 4, το  $W = \{(D, R, s, \tau, o)\}$  χρησιμοποιεί σταθερές τιμές φωτεινότητας σε κάθε επαναληπτική εφαρμογή του συστολικού μετασχηματισμού στην αυθαίρετη εικόνα  $P$ , οι οποίες δίνονται από τη σχέση:

$$o_{constant} = \bar{r} - s\bar{d}_0 \quad (\Gamma.1)$$

όπου το  $\bar{r}$  αντιστοιχεί στη μέση τιμή της περιοχής της αρχικής εικόνας και το  $\bar{d}_0$  στη μέση τιμή της υποδειγματολειφθείσης τύπου  $D$  περιοχής της αρχικής εικόνας που δίνει τη βέλτιστη προσέγγιση. Αντίθετα, το  $\tilde{W} = \{(D, R, \tau, s, \bar{r})\}$  υπολογίζει με βάση τη μέση τιμή, μια μεταβλητή τιμή φωτεινότητας σε κάθε επανάληψη. Η φωτεινότητα δηλαδή στην περίπτωση αυτή δίνεται από την ακόλουθη σχέση:

$$o_{variable} = \bar{r} - s\bar{d} \quad (\Gamma.2)$$

όπου το  $\bar{d}$  αντιστοιχεί στη μέση τιμή της τύπου  $D$  περιοχής της αυθαίρετης εικόνας  $P$  στην οποία εφαρμόζεται ο συστολικός μετασχηματισμός. Με βάση τα παραπάνω, οι Ευκλείδειες αποστάσεις της συγκεκριμένης τύπου  $R$  περιοχής

από τις μετασχηματισμένες τύπου D περιοχές  $W(D)$  και  $\widetilde{W}(D)$  δίνονται από τις σχέσεις:

$$\begin{aligned} E_1 &= \|R - (sD + o_{\text{constan}t} \mathbf{1})\| = \|R - [sD + (\bar{r} - s\bar{d}_0)\mathbf{1}]\| = \\ &= \|R - (sD + (\bar{r}\mathbf{1} - s\bar{d}_0\mathbf{1}))\| = \|(R - \bar{r}\mathbf{1}) - s(D - \bar{d}_0\mathbf{1})\| \end{aligned} \quad (\Gamma.3)$$

και

$$\begin{aligned} E_2 &= \|R - (sD + o_{\text{variable}} \mathbf{1})\| = \|R - [sD + (\bar{r} - s\bar{d})\mathbf{1}]\| = \\ &= \|R - (sD + (\bar{r}\mathbf{1} - s\bar{d}\mathbf{1}))\| = \|(R - \bar{r}\mathbf{1}) - s(D - \bar{d}\mathbf{1})\| \end{aligned} \quad (\Gamma.4)$$

Για να ισχύει η (4.15) αρκεί να αποδειχθεί ότι:

$$E_1 \geq E_2 \Leftrightarrow E_1^2 \geq E_2^2 \Leftrightarrow E_1^2 - E_2^2 \geq 0 \quad (\Gamma.5)$$

Για το λόγο αυτό, υπολογίζουμε αναλυτικά το τετράγωνο των δύο αυτών αποστάσεων:

$$\begin{aligned} E_1^2 &= \|(R - \bar{r}\mathbf{1}) - s(D - \bar{d}_0\mathbf{1})\|^2 = \sum_{i=1}^k [(r_i - \bar{r}) - s(d_i - \bar{d}_0)]^2 = \\ &= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \sum_{i=1}^k (r_i - \bar{r})(d_i - \bar{d}_0) + s^2 \sum_{i=1}^k (d_i - \bar{d}_0)^2 = \\ &= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \left( \sum_{i=1}^k r_i d_i - \sum_{i=1}^k r_i \bar{d}_0 - \sum_{i=1}^k \bar{r} d_i + \sum_{i=1}^k \bar{r} \bar{d}_0 \right) + \\ &+ s^2 \left( \sum_{i=1}^k d_i^2 - 2 \sum_{i=1}^k d_i \bar{d}_0 + \sum_{i=1}^k \bar{d}_0^2 \right) = \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + \\ &+ 2skr\bar{d}_0 + 2sk\bar{r}\bar{d} - 2skr\bar{d}_0 + s^2 \|D\|^2 - 2s^2 k\bar{d}\bar{d}_0 + ks^2 \bar{d}_0^2 = \\ &= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + 2skr\bar{d} + s^2 \|D\|^2 - 2s^2 k\bar{d}\bar{d}_0 + ks^2 \bar{d}_0^2 \end{aligned} \quad (\Gamma.6)$$

και

$$\begin{aligned}
E_2^2 &= \left\| (R - \bar{r}\mathbf{1}) - s(D - \bar{d}\mathbf{1}) \right\|^2 = \sum_{i=1}^k \left[ (r_i - \bar{r}) - s(d_i - \bar{d}) \right]^2 = \\
&= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \sum_{i=1}^k (r_i - \bar{r})(d_i - \bar{d}) + s^2 \sum_{i=1}^k (d_i - \bar{d})^2 = \\
&= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \left( \sum_{i=1}^k r_i d_i - \sum_{i=1}^k r_i \bar{d} - \sum_{i=1}^k \bar{r} d_i + \sum_{i=1}^k \bar{r} \bar{d} \right) + \\
&+ s^2 \left( \sum_{i=1}^k d_i^2 - 2 \sum_{i=1}^k d_i \bar{d} + \sum_{i=1}^k \bar{d}^2 \right) = \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + \\
&+ 2sk\bar{r}\bar{d} + 2sk\bar{r}\bar{d} - 2sk\bar{r}\bar{d} + s^2 \|D\|^2 - 2s^2 k\bar{d}\bar{d} + ks^2 \bar{d}^2 = \\
&= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + 2sk\bar{r}\bar{d} + s^2 \|D\|^2 - ks^2 \bar{d}^2
\end{aligned} \tag{Γ.7}$$

Στη συνέχεια, υπολογίζουμε τη διαφορά των (Γ.6)-(Γ.7):

$$\begin{aligned}
E_1^2 - E_2^2 &= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + 2sk\bar{r}\bar{d} + s^2 \|D\|^2 - \\
&- 2s^2 k\bar{d}\bar{d}_0 + ks^2 \bar{d}_0^2 - \\
&- \left( \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + 2sk\bar{r}\bar{d} + s^2 \|D\|^2 - ks^2 \bar{d}^2 \right) = \\
&= \sum_{i=1}^k (r_i - \bar{r})^2 - 2s \langle R, D \rangle + 2sk\bar{r}\bar{d} + s^2 \|D\|^2 - 2s^2 k\bar{d}\bar{d}_0 + \\
&+ ks^2 \bar{d}_0^2 - \sum_{i=1}^k (r_i - \bar{r})^2 + 2s \langle R, D \rangle - 2sk\bar{r}\bar{d} - s^2 \|D\|^2 + ks^2 \bar{d}^2 = \\
&= ks^2 \bar{d}^2 - 2s^2 k\bar{d}\bar{d}_0 + ks^2 \bar{d}_0^2 = ks^2 \left( \bar{d}^2 - 2\bar{d}\bar{d}_0 + \bar{d}_0^2 \right) = \\
&= ks^2 \left( \bar{d} - \bar{d}_0 \right)^2 \geq 0
\end{aligned} \tag{Γ.8}$$

Όπως φαίνεται από την τελευταία εξίσωση,  $E_1 \geq E_2$ , που σημαίνει ότι  $d(R, W(D)) \geq d(R, \tilde{W}(D))$  και κατά συνέπεια,  $d(\tilde{W}(P), O) \leq d(W(P), O) \square$ .