ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΣΕΡΡΩΝ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

Δρ. Χαράλαμπος Π. Στρουθόπουλος
Καθηγητής

ΣΕΡΡΕΣ, ΙΟΥΝΙΟΣ 2008

ΠΕΡΙΕΧΟΜΕΝΑ

1. Κεφάλαιο 1 – Εισαγωγή και ορισμοί
    1.1 Είδη ψηφιακών εικόνων.
    1.2 Βασικές έννοιες
    1.3 Ιστόγραμμα
2. Κεφάλαιο 2 – Επεξεργασίες σε δυαδικές εικόνες
    2.1. Κέντρο βάρος αντικειμένου
    2.2. Κωδικοποίηση δυαδικής εικόνας κατά μήκος διαδρομής.
    2.3. Συνδεδεμένα στοιχεία.
    2.4. Κωδικοποίηση αλυσίδας
    2.5. Ο μετασχηματισμός του Hough.
3. Κεφάλαιο 3 – Επεξεργασίες σε εικόνες αποχρώσεων του γκρι
    3.1. Βελτίωση εικόνων
        3.1.1. Εξομάλυνση
        3.1.2. Εξισορρόπηση ιστογράμματος
    3.2. Κατωφλίωση
        3.2.1 Κατωφλίωση με βάση τη διασπορά
        3.2.2 Κατωφλίωση με βάση την εντροπία
        3.2.3 Πολυκατωφλίωση με χρήση νευρωνικού δικτύου
    3.3. Κωδικοποίηση εικόνων αποχρώσεων του γκρι
        3.3.1 Κωδικοποίηση Huffman
        3.3.2 Κωδικοποίηση 2Δ-διακριτού μετασχηματισμό συνημιτόνου
    3.4. Ανίχνευση ακμών

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή και ορισμοί

Η ψηφιακή επεξεργασία εικόνας (ΨΕΕ) αποτελεί έναν ευρύ επιστημονικό κλάδο που αναπτύχθηκε με την ραγδαία εξέλιξη των υπολογιστών. Ο όρος εικόνα χρησιμοποιείται ευρύτερα από την απλή απεικόνιση ενός σκηνικού και περιλαμβάνει την αποτύπωση κάθε είδους πληροφοριών. Τα υπερηχογραφήματα, οι μαγνητικές τομογραφίες, οι δορυφορικές φωτογραφίες κ.α. μπορούν να επεξεργαστούν ως ψηφιακές εικόνες.

Οι στόχοι της ΨΕΕ είναι οι εξής:

1) Η ψηφιοποίηση και κωδικοποίηση εικόνων με σκοπό την αποθήκευση, μετάδοση και εκτύπωσή τους.
2) Η βελτίωση και η αποκατάσταση των εικόνων με σκοπό την καλύτερη απεικόνισή τους.
3) Η ανάλυση και κατανόηση των εικόνων

Η ΨΕΕ συνεργάζεται με τους παρακάτω επιστημονικούς κλάδους:

1) Ψηφιακή Επεξεργασία Σημάτων (ΨΕΣ)
2) Ρομποτική όραση
3) Τεχνητή Νοημοσύνη
4) Αναγνώριση Προτύπων
5) Νευρωνικά Δίκτυα
6) Ασαφής Λογική
7) Κωδικοποίηση
8) Γραφικά Η/Υ

Η ψηφιακή εικόνα είναι ένα πεπερασμένο σύνολο περιοχών όπου κάθε περιοχή είναι χρωματισμένη με χρώμα που προέρχεται από ένα πεπερασμένο σύνολο χρωμάτων. Στις περισσότερες των περιπτώσεων, μια ψηφιακή εικόνα είναι ένα ορθογώνιο, διαιρεμένο με γραμμές και στήλες σε ορθογώνιες περιοχές που κάθε μία έχει συγκεκριμένο χρώμα. Μια τέτοια περιοχή ονομάζεται στοιχείο της εικόνας ή εικονοστοιχείο. Στην αγγλική λέγεται *pixel* ή *pel*, όρος ο οποίος προέρχεται από τη σύντμηση των λέξεων picture element. Από τα παραπάνω καταλαβαίνουμε πως μια ψηφιακή εικόνα είναι ένα διδιάστατο ψηφιακό σήμα.

Αν κάθε χρώμα κωδικοποιηθεί με έναν αριθμό τότε η ορθογώνια ψηφιακή εικόνα περιγράφεται από έναν πίνακα αριθμών $J \times K$, όπου J το πλήθος των γραμμών και K το πλήθος των στηλών της ψηφιακής εικόνας. Η τιμή $I(j,k)$ με $k=0,1,2....K-1$ και $j=0,1,2....J-1$ είναι ο κωδικός του χρώματος της ψηφιακής εικόνας.

Η μετατροπή μιας εικόνας σε ψηφιακή μορφή ουσιαστικά είναι η μετατροπή ενός δισδιάστατου αναλογικού σήματος σε ψηφιακό και απαιτεί τις διαδικασίες της δειγματοληψίας και του κβαντισμού.

1.1 Είδη ψηφιακών εικόνων

Υπάρχουν τρία είδη ψηφιακών εικόνων που χαρακτηρίζονται από το πλήθος των χρωμάτων που περιέχουν:

1) *Δυαδικές εικόνες* (*binary images*): Κάθε εικονοστοιχείο των εικόνων μπορεί να χρωματιστεί με ένα από δύο χρώματα.(συνήθως άσπρο ή μαύρο). Για κάθε εικονοστοιχείο απαιτείται ένα bit πληροφορίας, π.χ. με τιμή μηδέν (0) για το μαύρο και ένα (1) για λευκό. Οι εικόνες των εγγράφων που αποτελούνται μόνο_ από το χρώμα του χαρτιού και της μελάνης αναπαρίστανται σε δυαδική ψηφιακή μορφή.
2) *Εικόνες αποχρώσεων του γκρι* (*gray level images*): Κάθε εικονοστοιχείο των εικόνων μπορεί να χρωματιστεί με μία από τις αποχρώσεις του γκρι οι οποίες ξεκινούν από το μαύρο και καταλήγουν στο λευκό. Από αυτές τις αποχρώσεις συνήθως λαμβάνονται 256 αντιπροσωπευτικές που κωδικοποιούνται με τιμές 0,1,2….255. Η απόχρωση κάθε εικονοστοιχείου προφανώς απαιτεί πληροφορία ενός byte.
3) *Έγχρωμες εικόνες* (*color images*) στις οποίες κάθε εικονοστοιχείο χρωματίζεται με χρώματα που προέρχονται από την ανάμειξη των αποχρώσεων του κόκκινου, πράσινου και μπλε (*RGB*). Για κάθε ένα από τα τρία αυτά χρώματα λαμβάνονται 256 αποχρώσεις δηλαδή πληροφορία του ενός byte. Συνεπώς κάθε εικονοστοιχείο της έγχρωμης εικόνας, απαιτεί 3 bytes.

1.2 Βασικές έννοιες

Το σύνολο των χρωμάτων που μπορούν να χρησιμοποιηθούν για τον χρωματισμό των εικονοστοιχείων της εικόνας λέγεται χρωματική παλέτα. Εάν C είναι το πλήθος των χρωμάτων, τότε για την κωδικοποίησή τους απαιτούνται B bits και ισχύουν οι σχέσεις

$$C=2^B \Leftrightarrow B=log_2C$$

Το B ονομάζεται *βάθος bit* (*bit depth*) της ψηφιακής εικόνας. Εάν η εικόνα έχει K στήλες και J γραμμές τότε για την απεικόνισή της απαιτούνται $J\times K\times B$ bits. Ο Πιν.1.2.1 παρουσιάζει ενδεικτικές τιμές των παραπάνω μεγεθών.

Είδος εικόνας	J	K	B	Bits	bytes
Δυαδική	100	100	2	20000	2500
Αποχρώσεων του γκρι	100	100	8	80000	10000
Έγχρωμη RGB	100	100	24	240000	30000
Δυαδική	512	512	2	524288	65536
Αποχρώσεων του γκρι	512	512	8	2097152	262144
Έγχρωμη RGB	512	512	24	6291456	786432

Πίνακας 1.2.1.

Ευκρίνεια της εικόνας είναι το πλήθος των εικονοστοιχείων ανά μονάδα επιφάνειας και καθορίζει πόσο λεπτομερής είναι η ψηφιακή αναπαράσταση της εικόνας. Η ευκρίνεια E μιας εικόνας διαστάσεων $J\times K$ και εμβαδού A δίνεται από την σχέση

$$E=\frac{J\times K}{A} \qquad (1.2.1)$$

και μετριέται σε πλήθος εικονοστοιχείων ανά μονάδα επιφάνειας π.χ. pixels/mm^2 ή dpi (dots per inch : κουκίδες ανά ίντσα). Η ευκρίνεια εξαρτάται τόσο από το πλήθος των εικονοστοιχείων όσο και από τις φυσικές διαστάσεις της εικόνας.

Μια ομάδα γειτονικών εικονοστοιχείων λέγεται γειτονιά. Σε μια γειτονιά $S_{M\times N}$ με M γραμμές και N στήλες μιας εικόνας διαστάσεων $J\times K$ υπάρχει ένα κεντρικό εικονοστοιχείο (j_c,k_c) όταν M, N είναι περιττοί αριθμοί. Η θέση των εικονοστοιχείων της S αναφέρονται συχνά, σχετικά με την θέση του κεντρικού εικονοστοιχείου της. Η πιο συνήθης γειτονιά είναι τριών (3) γραμμών και τριών (3) στηλών και λέγεται γειτονιά 3×3. Στο Σχ.1.2.1 φαίνονται 3×3 και 5×5 γειτονιές.

Γειτονιά 3×3 Γειτονιά 5×5

Σχήμα 1.2.1: Γειτονιές 3×3 και 5×5 με τα κεντρικά τους εικονοστοιχεία.

Σε κάθε εικονοστοιχείο μιας γειτονιάς $S_{M\times N}$ μπορούμε να αντιστοιχίσουμε έναν συντελεστή w_{mn} , $m=0,...,M-1$, $n=0,...,N-1$. Ο πίνακας $W_{M\times N}$ λέγεται μάσκα. Οι συντελεστές της μάσκας και οι τιμές φωτεινότητας των εικονοστοιχείων μπορούν να εμπλακούν σε χρήσιμους υπολογισμούς για την επεξεργασία της εικόνας. Ο συνηθέστερος υπολογισμός δίνεται από την σχέση

$$A = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} w_{mn} \cdot I_{j-1+m,k-1+n}}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} w_{mn}}$$

όπου (j,k) το κεντρικό εικονοστοιχείο της γειτονιάς.

Στο παρακάτω σχήμα δίνονται μάσκες 3×3 και η εφαρμογή τους σε μία 3×3 περιοχή εικονοστοιχείων.

1	1	1
1	1	1
1	1	1

240	250	240
240	100	150
240	250	250

$$\frac{240+250+240+240+100+150+240+250++0}{1+1+1+1+1+1+1+1+1+1} \approx 191$$

1.41	1	1.41
1	8	1
1.41	1	1.41

240	250	240
240	100	150
240	250	250

$$\frac{1.41\cdot 240+250+1.41\cdot 240+240+8\cdot 100+150+1.41\cdot 240+250+1.41\cdot 250}{1.41+1+1.41+1+8+1+1.41+1+1.41} \approx 138$$

Σχήμα 1.2.2.

Αν (j_c,k_c) είναι το κεντρικό εικονοστοιχείο της γειτονιάς S της εικόνας $I_{J\times K}$ η τιμή A μπορεί να αποδοθεί ως τιμή φωτεινότητας του εικονοστοιχείου (j_c,k_c) μιας νέας εικόνας $I'_{J\times K}$. Αν αυτό εφαρμοσθεί για όλες τις γειτονιές της εικόνας I, τότε λέμε ότι η νέα εικόνα I' προέκυψε από το φιλτράρισμα της I με την μάσκα W.

1.3. Ιστόγραμμα

Αν η ψηφιακή εικόνα αποτελείται από αποχρώσεις του γκρι μπορούμε να δημιουργήσουμε την κατανομή του πλήθους των εικονοστοιχείων που έχουν την ίδια τιμή απόχρωσης για κάθε απόχρωση. Η κατανομή αυτή λέγεται ιστόγραμμα των αποχρώσεων της εικόνας και δίνεται μαθηματικά απο την σχέση

$$h(g) = \sum_{\forall I(j,k)=g} 1 \qquad (1.3.1)$$

όπου $g=0,1,2..G-1$, G το πλήθος των αποχρώσεων, $j=0,...,J-1$, $k=0,...,K-1$, J το πλήθος των γραμμών, K το πλήθος των στηλών της εικόνας, $I(j,k)$ η τιμή της φωτεινότητας στο σημείο (j,k) και $h(g)$ το πλήθος των εικονοστοιχείων με απόχρωση g. Το ιστόγραμμα δηλαδή μας πληροφορεί πόσα εικονοστοιχεία υπάρχουν με συγκεκριμένη τιμή απόχρωσης. Η συνάρτηση

$$H(g) = \frac{h(g)}{\sum_{g=0}^{G-1} h(g)} \qquad (1.3.2)$$

λέγεται κανονικοποιημένο ιστόγραμμα και δίνει την πιθανότητα ένα τυχαίο εικονοστοιχείο της εικόνας να έχει απόχρωση g.

ΚΕΦΑΛΑΙΟ 2

ΕΠΕΞΕΡΓΑΣΙΕΣ ΣΕ ΔΥΑΔΙΚΕΣ ΕΙΚΟΝΕΣ

2.1 Κέντρο βάρους αντικειμένου.

Μία ψηφιακή δυαδική εικόνα με J γραμμές και K στήλες παριστάνεται με έναν πίνακα $I_{J\times K}$ και κάθε στοιχείο I_{jk} , j=0,...,J-1, k=0,...,K-1 παίρνει τιμή μηδέν (0) ή ένα (1). Αν η εικόνα αναπαριστά ένα αντικείμενο τότε το πλήθος των εικονοστοιχείων του αντικειμένου που έχουν τιμή ΄1΄ δίνεται από την σχέση

$$N = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} I_{jk}$$

Το κέντρο βάρους του αντικειμένου βρίσκεται στην θέση $(\bar{j},\bar{k})$ της εικόνας σύμφωνα με τις σχέσεις

$$\bar{k} = \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{K-1} k \cdot I_{jk}}{N},$$

$$\bar{j} = \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{K-1} j \cdot I_{jk}}{N}$$

2.2 Κωδικοποίηση δυαδικής εικόνας κατά μήκος διαδρομής

Μία δυαδική εικόνα μπορεί να κωδικοποιηθεί με κατάλληλο αλγόριθμο ώστε να ελαττωθεί η απαιτούμενη ποσότητα πληροφορίας για την αποθήκευση ή την μετάδοσή της και να διευκολυνθεί η αναγνώριση του περιεχομένου της. Ένας τέτοιος αλγόριθμος είναι η *κωδικοποίηση κατά μήκος διαδρομής* (*RLE:Run Lenght Ecoding*). Σύμφωνα με αυτόν ονομάζουμε *συστοιχία* μία ομάδα από διαδοχικά εικονοστοιχεία με την ίδια τιμή (όλα ΄1΄ ή όλα ΄0΄) και το πλήθος των εικονοστοιχείων ως *μήκος της συστοιχίας*. Διατρέχουμε κάθε σειρά της εικόνας και γράφουμε την θέση του πρώτου εικονοστοιχείου και το μήκος κάθε

συστοιχίας από ΄1΄ ή εναλλακτικά την θέση του πρώτου και του τελευταίου εικονοστοιχείου κάθε συστοιχίας. Μία άλλη προσέγγιση είναι να γράψουμε το μήκος των διαδοχικών συστοιχειών από ΄1΄ και ΄0΄ δεχόμενοι ότι η πρώτη συστοιχία αποτελείται πάντα από ΄1΄ έστω και μηδενικού μήκους . Στο Σχ.2.2.1 που ακολουθεί φαίνεται η κωδικοποίηση των τριών πρώτων γραμμών μιας δυαδικής εικόνας με τους τρόπους που αναφέρθηκαν.

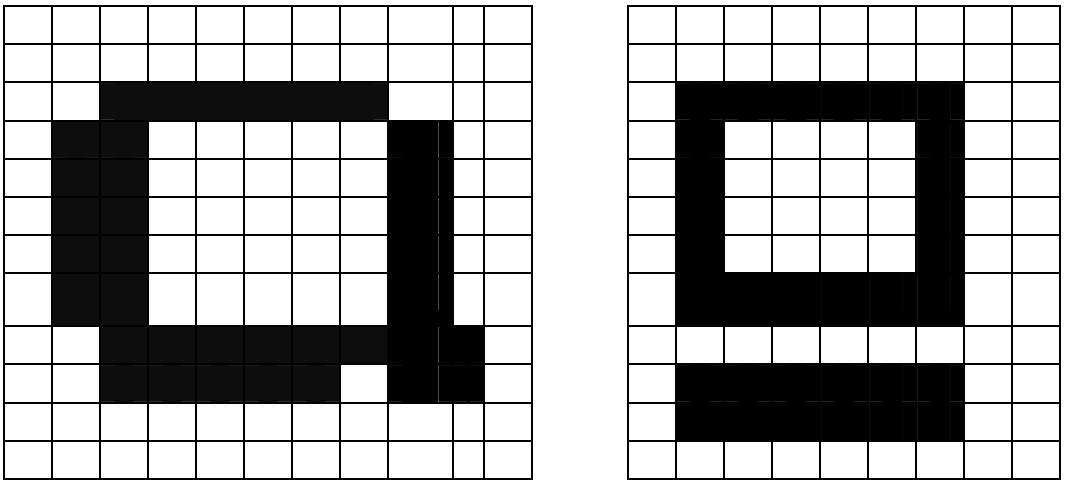
1	1	1	0	0	0	0	1	1	0	0
0	0	0	0	0	1	1	0	0	1	1
1	1	1	1	1	0	0	1	1	1	0

(0,3),(7,2)	(0,2),(8,9)	3,4,2,2
(5,2),(9,2)	(5,6),(9,10)	0,5,2,2,2
(0,5),(7,3)	(0,4),(7,9)	5,2,3,1

Σχήμα 2.2.1.

2.3 Συνδεδεμένα στοιχεία.

Θεωρούμε ένα σύνολο S εικονοστοιχείων μιας δυαδικής εικόνας με την ίδια τιμή (0 ή 1). Δύο εικονοστοιχεία ρ_1 και ρ_2 που ανήκουν στο S ονομάζονται *συνδεδεμένα* όταν υπάρχει διαδρομή από εικονοστοιχεία του S που οδηγεί από το ρ_1 στο ρ_2. Ένα σύνολο εικονοστοιχείων λέγεται *συνδεδεμένο συστατικό* (*connected component*) όταν όλα τα εικονοστοιχεία του είναι μεταξύ τους συνδεδμένα. Στο ακόλουθο Σχ.2.3.1 φαίνεται ένα συνδεδεμένο και ένα μη συνδεδεμένο συστατικό δυαδικής εικόνας.

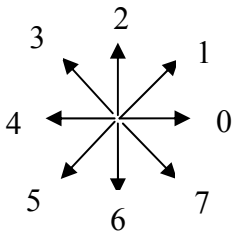


(α) Συνδεμένο συστατικό (β) Μη συνδεδεμένο συστατικό

Σχήμα 2.3.1

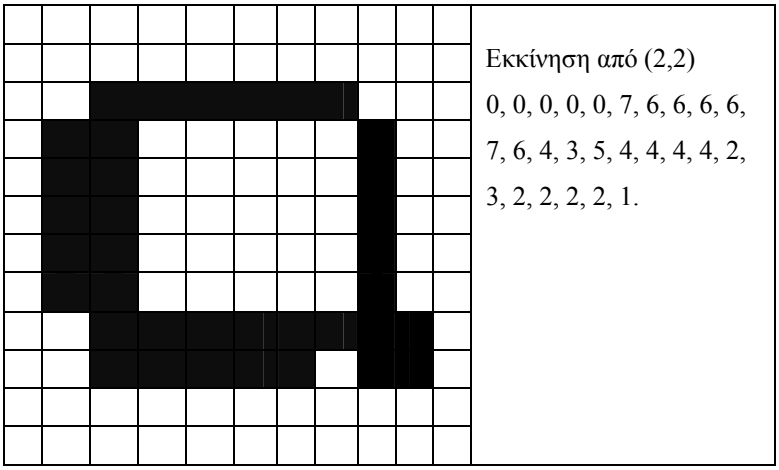
2.4 Κωδικοποίηση αλυσίδας.

Με την κωδικοποίηση αλυσίδας κωδικοποιούμε το περίγραμμα ενός αντικειμένου της εικόνας που είναι συνδεδεμένο. Για τον σκοπό αυτό καθορίζουμε και κωδικοποιούμε της διευθύνσεις που ξεκινούν από το κεντρικό εικονοστοιχείο και καταλήγουν στα γειτονικά του σε μια 3×3 γειτονιά. όπως φαίνεται στο Σχ.2.4.1



Σχήμα 2.4.1.

Ακολούθως ξεκινώντας από οποιοδήποτε εικονοστοιχείο του εξωτερικού περιγράμματος του αντικειμένου (συνήθως το πάνω αριστερό εικονοστοιχείο) διατρέχουμε το περίγραμμα γράφοντας τον κωδικό της σχετικής διεύθυνσης κάθε εικονοστοιχείου με το επόμενό του. Στο Σχ.2.4.2 δείχνεται η εφαρμογή της κωδικοποίησης αλυσίδας.



Σχήμα 2.4.2

2.5 Ο μετασχηματισμός του Hough

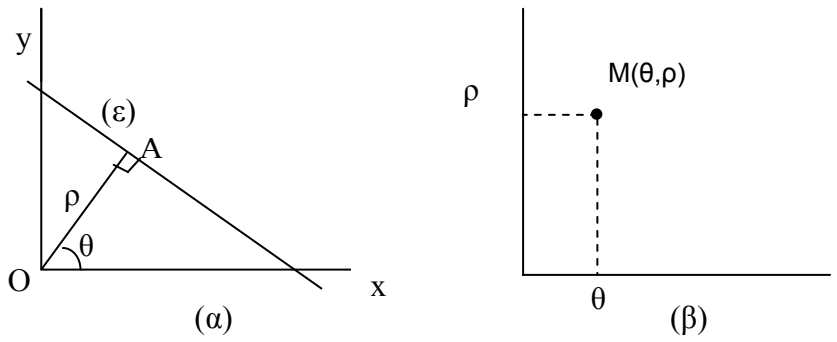
Συχνά οι ψηφιακές εικόνες περιέχουν γραμμές που ανήκουν σε σχήματα, τεχνικά σχέδια, γραφήματα, σειρές κειμένου ή άλλες αναπαραστάσεις. Σε πολλές εφαρμογές είναι επιθυμητή η εύρεση της θέσης και η αναγνώριση της μορφής των γραμμών (π.χ. ευθύγραμμα τμήματα, τόξα). Τέτοιες εφαρμογές είναι: η διανυσματική κωδικοποίηση τυπωμένων τεχνικών σχεδίων που ψηφιοποιήθηκαν από σαρωτές (scanners), η εύρεση περιοχών κειμένου σε έγγραφα, ο προσδιορισμός της υφής, βιομηχανικές εφαρμογές κατασκευών και ρομποτικής κ.α. Ο μετασχηματισμός του Hough (HT: Hough Transform) καταδεικνύει την ύπαρξη ευθειών σε μια εικόνα και αποτελεί την βάση πολλών τεχνικών για τον προσδιορισμό ευθυγράμμων τμημάτων, καμπυλών και μορφών που αναλύονται ανάλογα.

Ο HT προτάθηκε από τον Paul Hough το 1962 ως μέρος της κατασκευής μιας συσκευής ανίχνευσης της κίνησης σωματιδίων υψηλής ενέργειας και στόχευε στην αυτοματοποίηση και αντικατάσταση της οπτικής διαδικασίας που απαιτούσε πολλές ανθρωποώρες. Ο αρχικός αλγόριθμος εξελίχθηκε και η σημερινή διατύπωση του είναι η ακόλουθη.

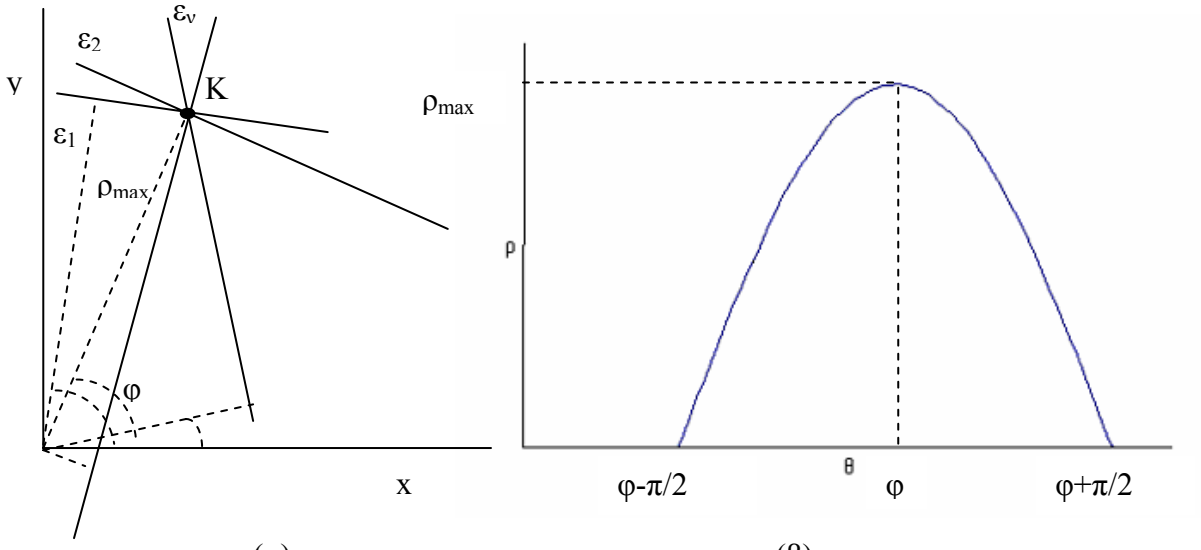
Κάθε ευθεία (ε) του καρτεσιανού επιπέδου Π_{xy} περιγράφεται από την πολική της εξίσωση

$$\rho = x \text{ συν}\theta + y \text{ ημ}\theta$$

όπου (x,y) σημείο της ευθείας και ρ, θ οι πολικές της παράμετροι (Σχ.2.5.1(α)).



Σχ. 2.5.1



(α) Σχήμα 2.5.2. (β)

Το ρ είναι το μήκος του ευθύγραμμου τμήματος ΟΑ, $OA \perp (\varepsilon)$ και θ η γωνία που σχηματίζει το ΟΑ με τον άξονα Οχ. Ισχύει ότι $\rho \geq 0$ και $-\pi \leq \theta < \pi$. Θεωρούμε το επίπεδο $\Pi_{\theta\rho}$ όπου στον κάθετο άξονα σημειώνουμε τις τιμές του ρ και στον οριζόντιο τις τιμές του θ. Οι πολικές παράμετροι (θ,ρ) της ευθείας (ε) καθορίζουν ένα σημείο Μ(θ,ρ) στο επίπεδο $\Pi_{\theta\rho}$ (Σχ.2.5.1.(β))

Με τον τρόπο αυτό η ευθεία (ε) του επιπέδου Π_{xy} αντιστοιχίζεται (μετασχηματίζεται) με την χρήση της πολικής της εξίσωσης σε ένα σημείο του επιπέδου $\Pi_{\theta\rho}$. Ισοδύναμα κάθε σημείο Μ(θ,ρ) του επιπέδου $\Pi_{\theta\rho}$ ορίζει μια ευθεία $\varepsilon \rightarrow \rho = x\cos\theta + y\eta\mu\theta$ στο επίπεδο Π_{xy}. Από ένα σημείο $K(x_\kappa,y_\kappa)$ του επιπέδου Π_{xy} διέρχονται άπειρες ευθείες (ε_ν), ν=1,2… που ικανοποιούν την σχέση

$$\rho_\nu = x_\kappa \, \sigma\upsilon\nu\theta_\nu + y_\kappa \, \eta\mu\theta_\nu$$

(Σχ.2.5.2(α)). Οι τιμές ρ_ν και θ_ν των μεταβλητών ρ και θ ανήκουν στην καμπύλη $\rho = x_\kappa \, \sigma\upsilon\nu\theta + y_\kappa \eta\mu\theta$ του επιπέδου $\Pi_{\theta\rho}$. Η καμπύλη είναι ημιτονοειδής (Σχ.2.5.2(β)), όπως δείχνεται ακολούθως. Έστω φ η γωνία χΟΚ. Ισχύει ότι

$$\varepsilon\varphi\phi = \frac{y_\kappa}{x_\kappa} \Leftrightarrow x_\kappa = y_\kappa \sigma\varphi\phi = y_\kappa \varepsilon\varphi(\frac{\pi}{2} - \phi) \Leftrightarrow$$

$$\rho = y_\kappa \frac{\eta\mu(\frac{\pi}{2} - \phi)}{\sigma\upsilon\nu(\frac{\pi}{2} - \phi)} \sigma\upsilon\nu\theta + y_\kappa \eta\mu\theta \Leftrightarrow$$

$$\rho = \frac{y_\kappa}{\eta\mu\phi} \eta\mu(\theta + \frac{\pi}{2} - \phi)$$

Αν x_κ και y_κ είναι θετικά (1° τεταρτημόριο) και επειδή ρ≥0 έπεται ότι

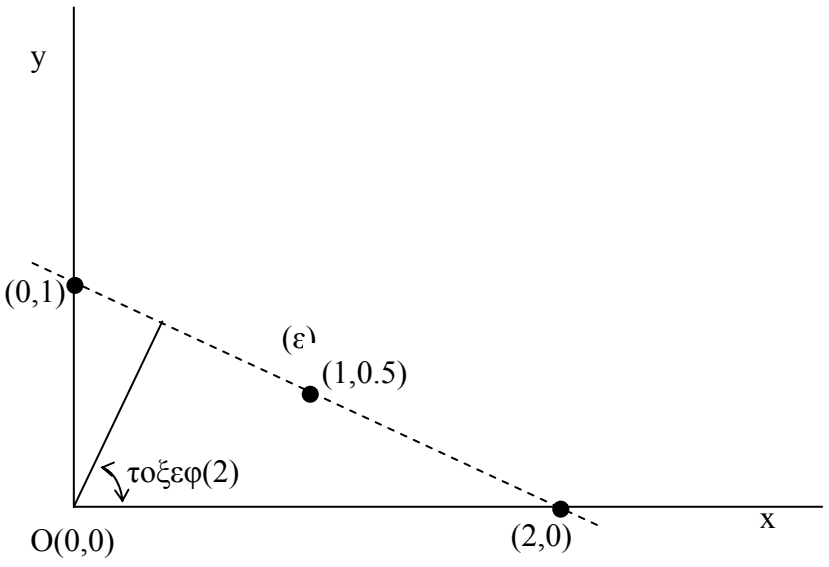
$$0 \le \phi \le \frac{\pi}{2} \qquad (\alpha)$$

$$\phi - \frac{\pi}{2} \le \theta \le \phi + \frac{\pi}{2}, \qquad (\beta)$$

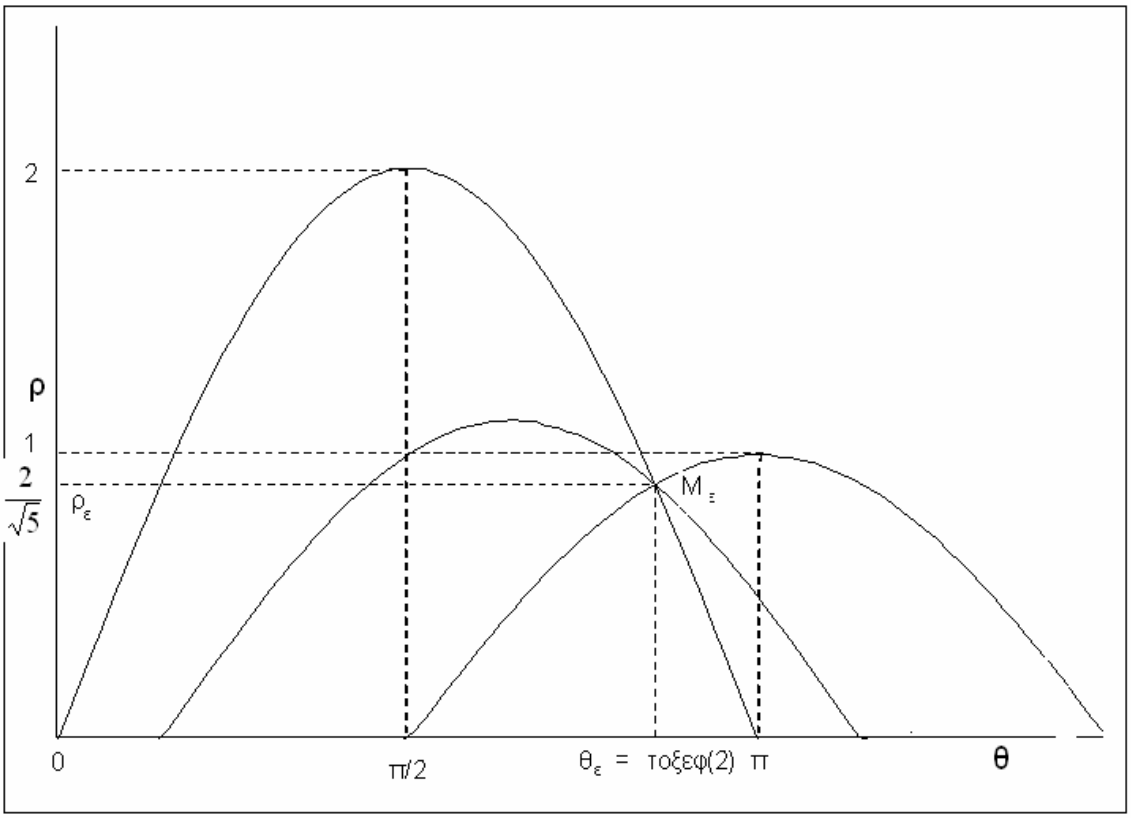
η μεγαλύτερη τιμή του ρ προκύπτει για θ=φ και είναι

$$\rho_{max} = \sqrt{x_\kappa^2 + y_\kappa^2} \qquad (\gamma)$$

Ο αναλυτικός υπολογισμός όλων των δυνατών τιμών των ρ και θ των ευθειών που ορίζουν ανά δύο τα σημεία της εικόνας και η εύρεση κοινών τιμών είναι επίπονος και πρακτικά ασύμφορος. Για αυτό ακολουθούμε το ακόλουθο σκεπτικό. Έστω τρία τουλάχιστον συνευθειακά σημεία στο επίπεδο Π_{xy} με τις αντίστοιχες καμπύλες των ευθειών που διέρχονται από αυτά (Σχ.2.5.3(α)). Οι καμπύλες αυτές τέμνονται σε σημείο $M_\varepsilon(\theta_\varepsilon,\rho_\varepsilon)$ με θ_ε και ρ_ε τις πολικές παραμέτρους της ευθείας που διέρχεται από αυτά (Σχ.2.5.3(β)). Άρα αν κατασκευάσουμε τις καμπύλες του επιπέδου $\Pi_{\rho\theta}$ για όλα τα σημεία του επιπέδου Π_{xy} τα σημεία τομής τους έχουν συντεταγμένες τις πολικές παραμέτρους των ευθειών που ορίζουν τα σημεία αυτά. Προφανώς, όσες καμπύλες διέρχονται από ένα σημείο τομής στο $\Pi_{\rho\theta}$, τόσα σημεία του Π_{xy} ανήκουν στην ευθεία που ορίζεται από το σημείο τομής.



Σχ.2.5.2(α).



Σχ.2.5.2(β).

Προγραμματιστικά και υπολογιστικά η μέθοδος μπορεί να εφαρμοσθεί με τεχνικές όπως η ακόλουθη. Ορίζεται ένας δισδιάστατος πίνακα $H_{M \times N}$ που κβαντίζει το επίπεδο $\Pi_{\theta\rho}$ σε M γραμμές και N στήλες με μηδενικές αρχικές τιμές. Η ψηφιακή εικόνα με πίνακα $I_{J \times K}$ τοποθετείται στο 1° τεταρτημόριο. Οι συντεταγμένες ενός εικονοστοιχείου P(k,j) με I(k,j)=1 θα είναι x_P=k, y_P=J-j. Επειδή η εικόνα βρίσκεται στο 1° τεταρτημόριο το διάστημα μεταβολής της γωνίας θ μπορεί να παίρνει τιμές από –π/2 έως π. Με βήμα

$$\Delta\theta = \frac{3\pi}{2N}$$

κάθε δείκτης ν=1,…,N αντιστοιχίζεται στο διάστημα [θ_ν, θν+Δθ), θ_ν= -π/2+(ν-1)Δθ.

Με βήμα

$$\Delta\rho = \frac{\sqrt{K^2 + J^2}}{M}$$

κάθε δείκτης μ=1,…,M αντιστοιχίζεται στο διάστημα [(μ-1)Δρ, μΔρ).

Για εκείνα τα ν που ισχύει

$$\phi - \frac{\pi}{2} \le \theta_\nu \le \phi + \frac{\pi}{2}\ ,\ \phi = \tau o\xi\varepsilon\phi(\frac{y_P}{x_P})$$

υπολογίζεται η τιμή $\rho = x_P\, \mathit{συν}\theta_\nu + y_P\, \mathit{ημ}\theta_\nu$ και ο δείκτης μ του διαστήματος που ανήκει το ρ. Η τιμή Η(μ,ν) αυξάνεται κατά ένα. Η διαδικασία επαναλαμβάνεται για όλα τα εικονοστοιχεία της εικόνας. Μετά το τέλος της διαδικασίας οι υψηλές τιμές του πίνακα Η προσδιορίζουν τις πολικές παραμέτρους ευθειών της εικόνας. Συγκεκριμένα αν Η(μ,ν) έχει υψηλή τιμή ισχύει ότι

$$\theta = \theta_\nu + \frac{\Delta\theta}{2} = -\frac{\pi}{2} + \nu\Delta\theta - \frac{\Delta\theta}{2} \ ,$$

$$\rho = \rho_\mu + \frac{\Delta\rho}{2} = \mu\Delta\rho - \frac{\Delta\rho}{2}$$

ΚΕΦΑΛΑΙΟ 3

ΕΠΕΞΕΡΓΑΣΙΕΣ ΣΕ ΕΙΚΟΝΕΣ ΑΠΟΧΡΩΣΕΩΝ ΤΟΥ ΓΚΡΙ

3.1 Βελτίωση εικόνων.

3.1.1 Εξομάλυνση

Με κατάλληλες τεχνικές είναι δυνατή η βελτίωση της οπτικής εμφάνισης μιας εικόνας ώστε να είναι καλύτερα αντιληπτή από τον άνθρωπο ή κατάλληλη για περαιτέρω επεξεργασία. Συχνά στις εικόνες εμφανίζονται μικροκουκίδες που οφείλονται σε τυχαίους παράγοντες (θόρυβος). Το *φίλτρο της μέσης τιμής* (*median filter*) είναι μια απλή τεχνική εξάλειψης και εξομάλυνσης του θορύβου αυτού από ψηφιακές εικόνες αποχρώσεων του γκρι. Σύμφωνα με την τεχνική αυτή από την αρχική εικόνα I παράγεται μία νέα εικόνα I' ιδίων διαστάσεων κάθε εικονοστοιχείο (j,k) της οποίας έχει φωτεινότητα $I'(j,k)$ που είναι η μέση τιμή των τιμών φωτεινότητας μιας γειτονιάς του εικονοστοιχείου (j,k) της εικόνας I. Αν για παράδειγμα κάθε περιοχή S είναι 3×3 τότε το $I'(j_c,k_c)$ δίνεται από την σχέση

$$I'_{jk} = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} I_{j-1+m,k-1+n}}{9} \qquad (3.1.1.1)$$

Με άλλα λόγια η εικόνα I' είναι το αποτέλεσμα του φιλτραρίσματος της αρχικής εικόνας I με την μάσκα $W = \frac{1}{9}\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}$.

Ειναι δυνατόν να χρησιμοποιηθούν και άλλες μάσκες για την εξομάλυνση μιας εικόνας όπως για παράδειγμα η $W = \frac{1}{16}\begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{vmatrix}$. Μάσκες 5×5 ή μεγαλύτερες εξομαλύνουν ακόμα περισσότερο την εικόνα.

Μια άλλη μέθοδος εξομάλυνσης είναι το *φίλτρο της ενδιάμεσης τιμής*. Σύμφωνα με την τεχνική αυτή οι τιμές των εικονοστοιχείων μιας γειτονιάς ταξινομούνται και η τιμή του εικοστοιχείου $I'(j,k)$ είναι η μεσαία από τις τιμές των εικονοστοιχείων της γειτονιάς του $I(j,k)$. Στο Σχ.3.1.1 που ακολουθεί φαίνονται οι τιμές που προκύπτουν από την εξομάλυνση ενός μέρους μιας εικόνας α) με το φίλτρο της μέσης τιμής και β) με το φίλτρο της ενδιάμεσης τιμής.

250	250	250	250	250	250
250	250	250	100	100	250
250	250	50	100	10	250
250	250	50	100	10	250
250	250	100	250	100	250
250	250	250	250	250	250
250	10	250	50	50	250
250	250	250	250	250	250

Τιμές φωτεινότητας της αρχικής εικόνας.

250	250	250	250	250	250
250	228	194	151	173	250
250	206	156	86	130	250
250	189	156	86	147	250
250	211	194	151	190	250
250	207	184	172	189	250
250	223	201	206	206	250
250	250	250	250	250	250

Οι τιμές φωτεινότητας μετά από εξομάλυνση με το φίλτρο της μέσης τιμής.

250	250	250	250	250	250
250	250	250	100	250	250
250	250	100	100	100	250
250	250	100	100	100	250
250	250	250	100	250	250
250	250	250	250	250	250
250	250	250	250	250	250
250	250	250	250	250	250

Οι τιμές φωτεινότητας μετά από εξομάλυνση με το φίλτρο της ενδιάμεσης τιμής.

Σχήμα 3.1.1.

Το φίλτρο της μέσης τιμής απομακρύνει τον θόρυβο χωρίς να επηρεάσει ιδιαίτερα τις ακμές της εικόνας. Το φίλτρο της ενδιάμεσης τιμής απομακρύνει τον θόρυβο και εξομαλύνει τις ακμές της εικόνας.

3.1.2. Εξισορρόπηση του ιστογράμματος

Η διάκριση γειτονικών περιοχών σε εικόνες αποχρώσεων του γκρι, είναι δύσκολη όταν η διαφορά τιμών των αποχρώσεων είναι μικρή. Η αύξηση των χρωματικών αντιθέσεων μιας εικόνας διευκολύνει την διάκριση των περιοχών αυτών. Μια τεχνική για την επίτευξη αυτού του σκοπού είναι η *εξισορρόπηση του ιστογράμματος* της εικόνας (*Histogram equalization)*. Σύμφωνα με αυτήν οι τιμές των αποχρώσεων των εικονοστοιχείων μεταβάλλονται έτσι ώστε να αυξηθεί η αντίθεση μεταξύ των περιοχών με διαδοχικές τιμές αποχρώσεων ανάλογα με το ποσοστό του πλήθους των εικονοστοιχείων τους. Συγκεκριμένα, αν G είναι το πλήθος όλων των αποχρώσεων της παλέτας του γκρι, $h(g)$ το κανονικοποιημένο ιστόγραμμα της αρχικής εικόνας I_{KXJ}, $g=0,1,\ldots,G-1$ και $P(g)$ η συνάρτηση αθροιστικής πιθανότητας του g, προκύπτει μία νέα εικόνα I' ίδιων διαστάσεων με τη I, κάθε εικονοστοιχείο (k,j) της οποίας έχει απόχρωση $I'(k,j) = [(G-1)P(I(k,j)]$. Με άλλα λόγια κάθε τιμής g μιας απόχρωση της αρχικής εικόνας αντικαθίσταται από μια τιμή $g' = [(G-1)P(g)]$. Υπενθυμίζεται ότι για την $P(g)$ ισχύουν οι σχέσεις

$$P(g) = \sum_{i=0}^{g} h(i) \text{ άρα και } P(g) = P(g\text{-}1) + h(g). \qquad (3.1.2.1)$$

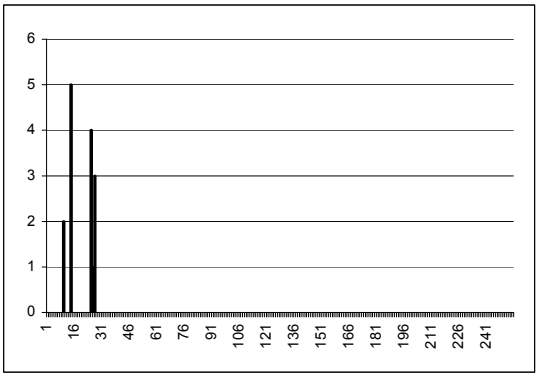
Παράδειγμα: Έστω ο ακόλουθος πίνακας τιμών των αποχρώσεων του γκρι μιας ψηφιοποιημένης εικόνας:

10	10	14	14	14
14	14	25	25	25
25	26	27	27	27

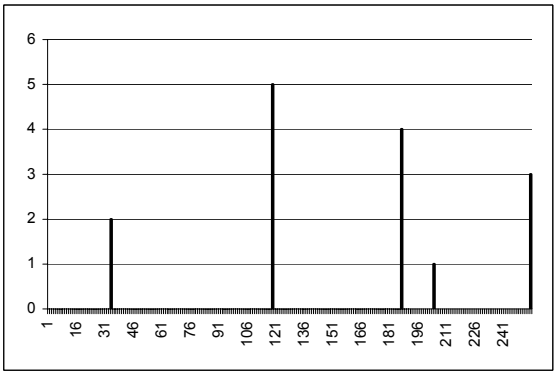
Οι αποχρώσεις g' προκύπτουν σύμφωνα με τους ακόλουθους υπολογισμούς:

G	$H(g)$	$h(g)$	$P(g)$	g'
10	2	2/15	2/15	255*2/15 = **34**
14	5	5/15	7/15	255*7/15 = **119**
25	4	4/15	11/15	255*11/15 = **187**
26	1	1/15	12/15	255*12/15 = **204**
27	3	3/15	15/15	255*15/15 = **255**

Στο Σχ.3.1.2.1 δείχνονται τα ιστογράμματα $h(g)$ και $h(g')$.



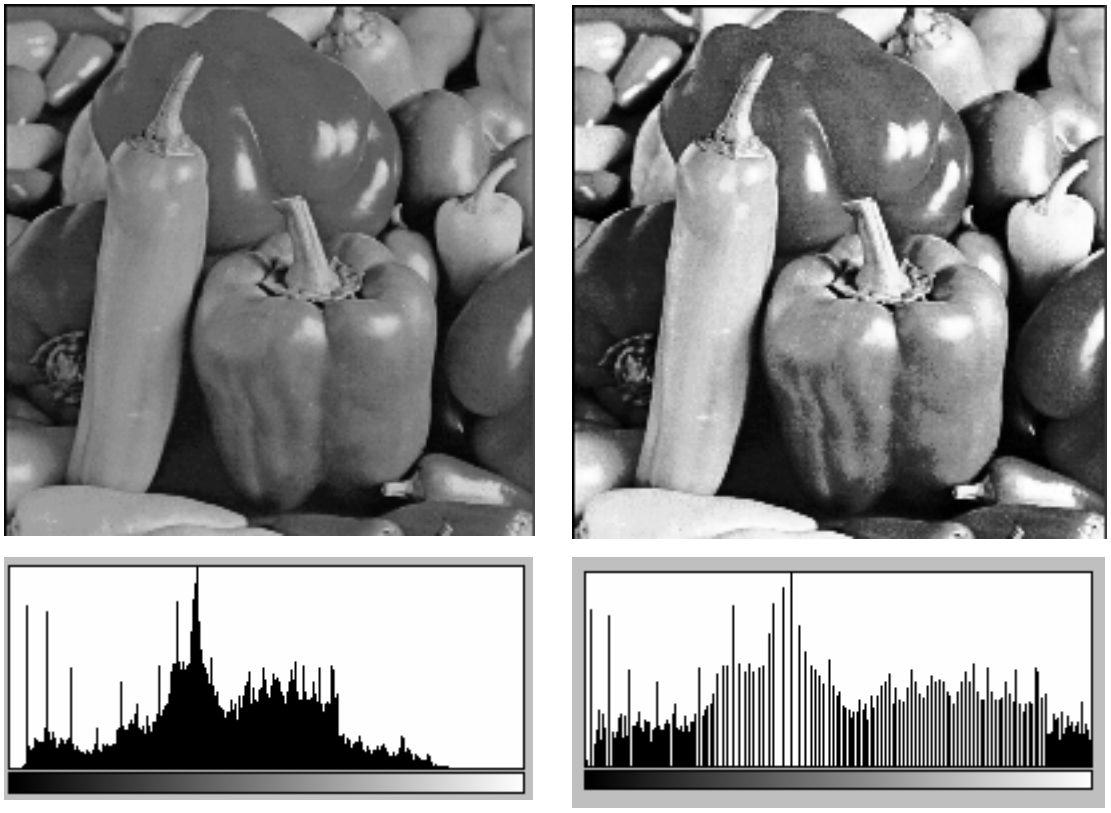
Το ιστόγραμμα $h(g)$



Το ιστόγραμμα $h(g')$

Σχήμα 3.1.2.1

Στο Σχ.3.1.2.2 φαίνεται η εφαρμογή της εξισορρόπησης του ιστογράμματος στην εικόνα με τις πιπεριές



Σχήμα 3.1.2.1.

3.2 Κατωφλίωση

Συχνά τα εικονοστοιχεία ενός αντικειμένου μιας εικόνας παίρνουν τιμές σε ένα μικρό διάστημα αποχρώσεων. Αυτό οδηγεί συνήθως στη δημιουργία ενός τοπικού μέγιστου στην περιοχή του ιστογράμματος της εικόνας. Η εύρεση τέτοιων τοπικών μεγίστων διευκολύνει τον εντοπισμό των αντικειμένων της εικόνας και την απόδοσή της με λιγότερες κύριες αποχρώσεις. Παρακάτω θα περιγράψουμε διάφορες τεχνικές για τον καθορισμό τιμών του πεδίου των αποχρώσεων μεταξύ των οποίων εμφανίζονται τοπικά μέγιστα του ιστογράμματος. Οι τιμές αυτές λέγονται κατώφλια.

Για την μετατροπή μιας εικόνας αποχρώσεων του γκρι σε δυαδική είναι επιθυμητός ο προσδιορισμός των 2 κύριων αποχρώσεών της. Με τον τρόπο αυτό οι αποχρώσεις της εικόνας χωρίζονται σε 2 κλάσεις C1, C2 που κάθε μια αποτελείται από τις αποχρώσεις που βρίσκονται πιο κοντά σε μια από τις 2 κύριες αποχρώσεις. Ο διαχωρισμός αυτός ισοδυναμεί με την εύρεση μιας τιμής κατωφλιού T για την οποία θα ισχύει

$$g \in \begin{cases} C_1 & \text{αν} \quad g < T \\ C_2 & \text{αν} \quad T \leq g \end{cases} \tag{3.2.1}$$

Ακολούθως από την αρχική εικόνα *I* προκύπτει η δυαδική εικόνα *I΄* κάθε εικονοστοιχείο της οποίας δίνεται από τη σχέση

$$I'_{jk} = \begin{cases} 0 & \text{αν} \quad I_{jk} \in C_1 \ (I_{jk} < T) \\ 1 & \text{αν} \quad I_{jk} \in C_2 \ (I_{jk} \geq T) \end{cases} \tag{3.2.2}$$

3.2.1 Κατωφλίωση με βάση τη διασπορά

Μια μέθοδος για την εύρεση της τιμής κατωφλίου T είναι η μέθοδος του Otsu. Σύμφωνα με αυτήν η τιμή *T* πρέπει να είναι τέτοια ώστε η συνολική εσωτερική διασπορά σ_w^2 εντός των κλάσεων να είναι ελάχιστη και η συνολική διασπορά σ_b^2 μεταξύ των κλάσεων μέγιστη. Σύμφωνα με την μέθοδο δημιουργούμε το κανονικοποιημένο ιστόγραμμα *h(g)* της εικόνας*)*. Η συνάρτηση *h(g)* δίνει την πιθανότητα να έχει ένα τυχαίο εικονοστοιχείο της εικόνας απόχρωση *g*.

Έστω T μια τιμή κατωφλιού που διαχωρίζει τις αποχρώσεις σε δύο κλάσεις C_1 ($g<T$) και C_2 ($T\leq g$). Η πιθανότητα ένα τυχαίο εικονοστοιχείο να έχει απόχρωση της κλάσεως C_1 είναι p_1 και δίνεται από τη σχέση

$$p_1 = \sum_{g=0}^{T-1} h(g) \tag{3.2.1.1}$$

όμοια για την κλάση C_2 είναι p_2 και δίνεται από τη σχέση

$$p_2 = \sum_{g=T}^{G-1} h(g) \tag{3.2.1.2}$$

Είναι προφανές ότι ισχύει

$$p_1+p_2=1. \tag{3.2.1.3}$$

Η μέση τιμή μ_1 της πρώτης κλάσης είναι

$$\mu_1 = \frac{\sum_{g=0}^{T-1} h(g)g}{p_1} \tag{3.2.1.4}$$

Η μέση τιμή μ_2 της δεύτερης κλάσης είναι και ισούται με

$$\mu_2 = \frac{\sum_{g=T}^{G-1} h(g)g}{p_2} \tag{3.2.1.5}$$

Η συνολική μέση τιμή μ δίνεται από τη σχέση

$$\mu = \sum_{g=0}^{G-1} h(g)g \tag{3.2.1.6}$$

Από τις τρείς προηγούμενες σχέσεις εύκολα αποδεικνύεται η σχέση

$$\mu=p_1\mu_1+p_2\mu_2 \tag{3.2.1.7}$$

Η διασπορά της πρώτης κλάσης δίνεται από τη σχέση

$$\sigma_1^2 = \frac{1}{P_1}\sum_{g=0}^{T-1}(g-\mu_1)^2 h(g) \tag{3.2.1.8}$$

και η διασπορά της δεύτερης κλάσης είναι

$$\sigma_2^2 = \frac{1}{P_2}\sum_{g=T}^{G-1}(g-\mu_2)^2 H(g) \tag{3.2.1.9}$$

Η συνολική διασπορά σ^2 είναι

$$\sigma^2 = \sum_{g=0}^{G-1}(g-\mu\ )^2 h(g) \tag{3.2.1.10}$$

Η συνολική εσωτερική διασπορά των κλάσεων ορίζεται από τη σχέση

$$\sigma_w^2 = p_1\sigma_1^2 + p_2\sigma_2^2 \qquad (3.2.1.11)$$

Η συνολική διασπορά μεταξύ των κλάσεων ορίζεται από τη σχέση

$$\sigma_b^2 = p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 = p_1 p_2(\mu_1 - \mu_2)^2 \qquad (3.2.1.12)$$

Αναζητούμε τώρα την κατάλληλη τιμή του Τ ώστε η σ_w^2 να είναι ελάχιστη και σ_b^2 μέγιστη. Προς τούτο θεωρούμε το πηλίκο

$$\lambda = \frac{\sigma_b^2}{\sigma_w^2} \qquad (3.2.1.13)$$

και ζητάμε την τιμή του Τ για την οποία το λ μεγιστοποιείται.

Για να μειωθούν οι υπολογισμοί θεωρούμε τις σχέσεις

$$\left.\begin{aligned} \frac{\lambda}{1+\lambda} &= \frac{\sigma_b^2}{\sigma_w^2 + \sigma_b^2} \\ \sigma^2 &= \sigma_w^2 + \sigma_b^2 \end{aligned}\right\} \Rightarrow \frac{\lambda}{1+\lambda} = \frac{\sigma_b^2}{\sigma^2} \qquad (3.2.1.14)$$

Η σχέση $\sigma^2 = \sigma_w^2 + \sigma_b^2$ αποδεικνύεται στο παράρτημα Α. Επειδή το σ^2 είναι ανεξάρτητο του T και η συνάρτηση λ/(1+λ) είναι αύξουσα συνεπάγεται πως για να είναι το λ μέγιστο αρκεί να είναι το σ_b^2 μέγιστο. Με τον τρόπο αυτό η τελικά η κατάλληλη τιμή του κατωφλίου T είναι εκείνη που μεγιστοποιεί την ποσότητα

$$\sigma_b^2 = \frac{p_1(\mu_1 - \mu)^2}{1 - p_1} \qquad (3.2.1.15)$$

ΠΑΡΑΔΕΙΓΜΑ 3.2.1.1.

Δίνεται ο πίνακας ψηφιακής εικόνας αποχρώσεων του γκρι. Ποια η τιμή κατωφλίωσης της με τη μέθοδο της διασποράς;

210	210	88	100	100
210	200	100	100	210
88	210	210	200	210
100	88	100	210	100
88	100	88	100	100

ΛΥΣΗ:

g	H(g)	h(g)
88	5	0.2
100	10	0.4
200	2	0.08
210	8	0.32

$$\mu = \frac{88 \cdot 5 + 100 \cdot 10 + 200 \cdot 2 + 210 \cdot 8}{25} = 140.8$$

T	μ_1	ρ_1	$\sigma_b^2 = \frac{\rho_1(\mu - \mu_1)^2}{1 - \rho_1}$
88	88	0,2	1936
100	(88×5+100×0,10)/0,6=96	0,24+0,26=0,6	13824
200	(88×5+100×0,10+200×2)/0,68=108,24	0,6+0,2=0,68	24894

Η μέγιστη τιμή του σ_b^2 προκύπτει για T=200 που αποτελεί σύμφωνα με τη μέθοδο η καλλίτερη τιμή για την κατωφλίωση

3.2.2. Κατωφλίωση με βάση την εντροπία

Μια διαφορετική προσέγγιση για την εύρεση κατάλληλης τιμής κατωφλίου T βασίζεται στην έννοια της εντροπίας. Σύμφωνα με την εκδοχή που παρουσίασε ο Kapur, η τιμή T χωρίζει το κανονικοποιημένο ιστόγραμμα h(g) σε δυο κλάσεις c_1 και c_2. Η πρώτη κλάση μπορεί να θεωρηθεί ως πηγή πληροφορίας με σύμβολα g = 0, …T-1 και αντίστοιχες πιθανότητες

$$\rho(g) = \frac{h(g)}{P_1} \qquad (3.2.2.1)$$

όπου

$$P_1 = \sum_{g=0}^{T-1} h(g) \qquad (3.2.2.2)$$

Η μέση πληροφορία (εντροπία) που προκύπτει της c_1 είναι

$$E_1 = -\sum_{g=0}^{T-1} \frac{h(g)}{P_1} \log(\frac{h(g)}{P_1}) = \log(P_1) - \frac{1}{P_1} \sum_{g=0}^{T=1} h(g) \log(h(g)) \qquad (3.2.2.3)$$

Όμοια η μέση πληροφορία που προκύπτει από την c_2 είναι

$$\begin{aligned} E_2 &= -\sum_{g=T}^{255} \frac{h(g)}{P_2} \log \frac{h(g)}{P_2} \ , P_2 = 1 - P_1 \\ &= \log(1 - P_1) - \frac{1}{1 - P_1} \sum_{1-P_1}^{255} h(g) \log(h(g)) \end{aligned} \qquad (3.2.2.4)$$

Ζητείται η πρώτη κλάση που θα αντιστοιχηθεί στην τιμή ‘0’ και η δεύτερη που θα αντιστοιχηθεί στο ‘1’ να προσφέρουν συνολικά την μεγαλύτερη δυνατή μέση πληροφορία. Δηλαδή η ζητούμενη τιμή του κατωφλίου Τα είναι εκείνη για την οποία $E_1(T)+E_2(T)$ γίνεται μέγιστο.

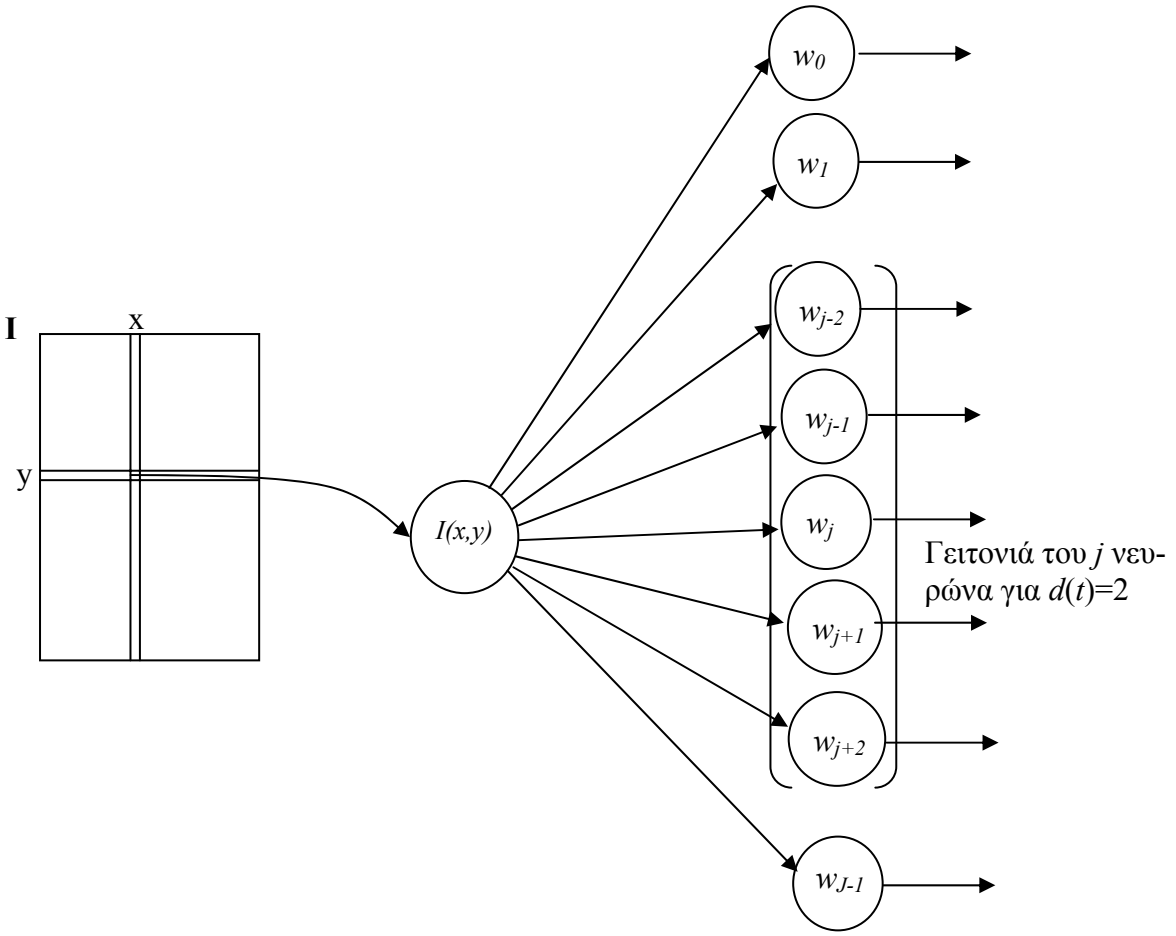
3.2.3. Πολυκατωφλίωση με χρήση νευρωνικού δικτύου

Εάν κάθε αντικείμενο της εικόνας αποτελείται από μία κύρια και κάποιες παραπλήσιες αποχρώσεις του γκρι , τότε δημιουργείται η συσσώρευση των τιμών του ιστογράμματος σε συγκεκριμένες περιοχές. Η επιλογή πολλαπλών κατωφλίων μπορεί να θεωρηθεί ως ένα πρόβλημα εύρεσης τιμών $T(k),\ k = 1,2,...,J-1$ με σκοπό το μετασχηματισμό της αρχικής εικόνας με L αποχρώσεις του γκρι σε μια νέα με J αποχρώσεις. Ειδικότερα για $T(1)<T(2)<...<T(J-1)$, τότε η νέα εικόνα ορίζεται ως

$$I(x,y)=\begin{cases} G(0)\ \alpha\nu\ f(x,y)\leq T(1) \\ G(1)\ \alpha\nu\ T(1)< f(x,y)\leq T(2) \\ . \\ . \\ G(J-1)\ \alpha\nu\ T(J-1)< f(x,y) \end{cases} \qquad (3.2.3.1)$$

όπου G μία συνάρτηση τιμών των επιπέδων του γκρι. Μία τεχνική πολλαπλών κατωφλίων που βασίζεται στο ιστόγραμμα πρέπει να πετυχαίνει έναν ικανοποιητικό διαχωρισμό των διαφόρων περιοχών του ιστογράμματος. Οι μέθοδοι της διασποράς και της εντροπίας είναι δυνατόν να εφαρμοσθούν και στην επίλυση του προβλήματος της πολυκατωφλίωσης, έχουν όμως μεγάλο υπολογιστικό κόστος.

Μία αποτελεσματική προσέγγιση είναι να θεωρήσουμε την εύρεση πολλαπλών κατωφλίων ως πρόβλημα εύρεσης κλάσεων (clustering problem). Ένα κατάλληλο νευρωνικό δίκτυο μπορεί να λύσει αποτελεσματικά προβλήματα εύρεσης κλάσεων. Είναι γνωστό ότι ο κύριος σκοπός ενός νευρωνικού δικτύου *Kohonen* για την δημιουργία ενός αυτό-οργανούμενου πίνακα απεικόνισης χαρακτηριστικών (SOFM: Self Organized Feature Map) είναι η αντιπροσώπευση ενός πολυπληθούς συνόλου διανυσμάτων εισόδου με ένα ολιγομελέστερο σύνολο “πρωτοτύπων ανυσμάτων” ώστε να επιτευχθεί μια καλή προσέγγιση του αρχικού χώρου εισόδου που να ικανοποιεί τα κύρια στατιστικά χαρακτηριστικά του. Θεωρούμε ένα δίκτυο με νευρώνες (κόμβους) όπως στο Σχ.3.2.3.1



Σχήμα 3.2.3.1

- Συγκροτείται το σύνολο εκπαίδευσης του ΑΠΑΧ από τις τιμές των αποχρώσεων των εικονοστοιχείων της εικόνας.
- Θεωρούμε τη μεταβλητή επανάληψης t που παίρνει ακέραιες τιμές από 0 μέχρι μία προκαθορισμένη τελική τιμή T (π.χ. T=100.000).
- Θεωρούμε τη μεταβλητή του ρυθμού εκμάθησης $\alpha(t)$.
- Θεωρούμε το μήκος $d(t)$ που ορίζει μια υποπεριοχή (γειτονιά) στην διάταξη των νευρώνων.
- Έστω w_j το βάρος της σύναψης μεταξύ του j νευρώνα του επιπέδου ανταγωνισμού και της εισόδου του.

Εκτελούνται τα παρακάτω βήματα:

Βήμα 1. Αρχικοποιούνται ($t=0$) τα βάρη των συνάψεων $w_j(0)$ με τυχαίες τιμές από 255.

Βήμα 2. Αρχικοποιείται η $\alpha(0)$ με μια μεγάλη τιμή, συνήθως μεταξύ 0.2 και 0.5.

Βήμα 3. Αρχικοποιείται η $d(0)$ με την τιμή J/2, που είναι ίση με το μισό του εύρους του κανάβου.

Βήμα 4. Επιλέγεται τιμή απόχρωσης *I(x,y)* ενός τυχαίου εικονοστοιχείου από το σύνολο εκπαίδευσης.

Βήμα 5. Υπολογίζεται η έξοδος $o_j(t)$ κάθε νευρώνα από τη σχέση

$$o_j(t) = \left\| I(x,y) - w_j(t) \right\|. \qquad (3.2.3.2)$$

Βήμα 6. Ο νευρώνας *c* ανακηρύσσεται νικητής εάν ικανοποιείται η συνθήκη

$$o_c(t) = min\{o_\xi(t)\}. \qquad (3.2.3.3)$$

Εάν οι έξοδοι δύο νευρώνων είναι ίσες, τότε κατά σύμβαση επιλέγεται αυτός με το μικρότερο δείκτη.

Βήμα 7. Τα βάρη w_j των συνάψεων ανανεώνονται σύμφωνα με τις παρακάτω σχέσεις

$$\Delta w_j = \begin{cases} \alpha(t) \cdot (I(x,y) - w_j(t)) \text{ αν } j \in N_c \\ 0 \text{ αν } j \notin N_c \end{cases} \qquad (3.2.3.4)$$

$$w_j(t+1) = w_j(t) + \Delta w_j(t) \qquad (3.2.3.5)$$

όπου N_c το σύνολο των δεικτών των νευρώνων που βρίσκονται μέσα στην γειτονιά του νικητή νευρώνα και πλευρά *d(t)*.

Βήμα 8. Αυξάνεται η μεταβλητή επανάληψης κατά ένα και αποδίδονται νέες τιμές στις μεταβλητές *α(t), d(t)* σύμφωνα με τις σχέσεις:

$$a(t) = a(0) \cdot (1 - \frac{t}{T}) \qquad (3.2.3.6)$$

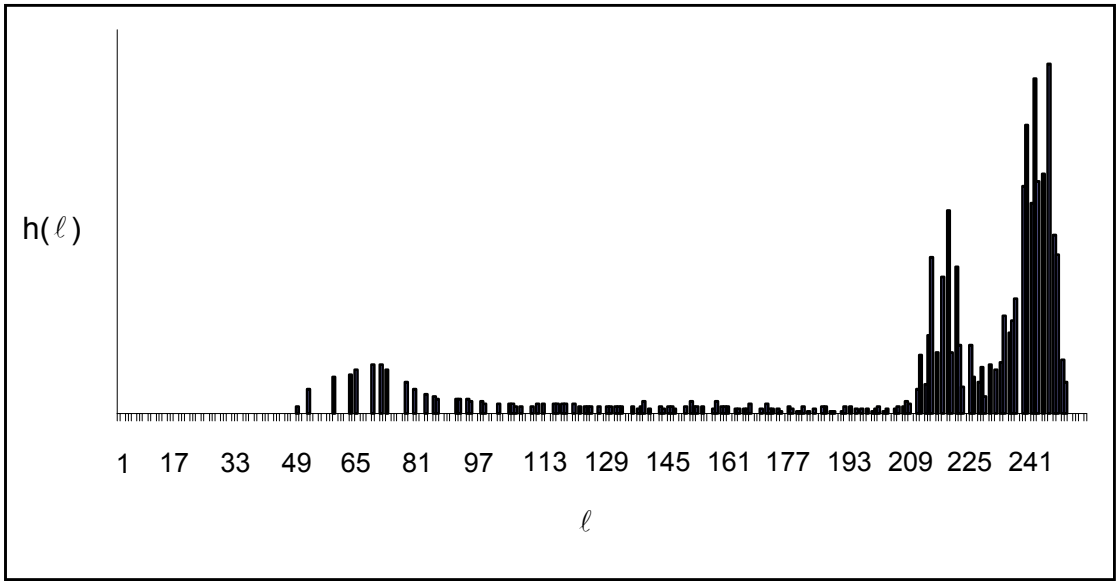
$$d(t) = d(0) \cdot (1 - \frac{t}{T}) \qquad (3.2.3.7)$$

Τα βήματα 4 έως 8 επαναλαμβάνονται έως ότου η μεταβλητή *t* πάρει τη μέγιστη τελική τιμή *T*. Είναι φανερό πως οι μεταβλητές *α(t)* και *d(t)* συγκλίνουν στο μηδέν καθώς η *t* τείνει στην τιμή *T*.

Μετά την εκπαίδευση κάθε άνυσμα εισόδου του ΑΠΑΧ αποδίδεται στον νικητή νευρώνα. Κάθε νευρώνας του επιπέδου εξόδου αντιπροσωπεύει μία ομάδα προτύπων (cluster). Πρότυπα με μεγάλη ομοιότητα αντιπροσωπεύονται από τον ίδιο νευρώνα. Ο χάρτης χαρακτηριστικών του Kohonen οργανώνει τους νευρώνες του επιπέδου ανταγωνισμού με τέτοιο τρόπο ώστε οι ομοιότητες μεταξύ των προτύπων να απεικονίζονται με σχέσεις γειτνίασης επάνω στον κάναβο του επιπέδου ανταγωνισμού

Στο Σχ. 3.2.3.1 φαίνεται η αρχική εικόνα στην οποία θα εφαρμοσθεί πολυκατωφλίωση και στο Σχ. 3.2.3.2.

Σχήμα 3.2.3.1.



Σχήμα 3.2.3.2. Το ιστόγραμμα της εικόνας του Σχ. 3.2.3.1.

Μετά την εκπαίδευση ενός ΝΔ *Kohonen* για την πολυκατωφλίωση των νευρώνων προκύπτουν οι τιμές των βαρών που αναγράφονται στον Πιν. 3.2.3.1 και οι τιμές των κατωφλίων του Πιν.3.2.3.2. Οι τιμές αυτές αποτελούν ουσιαστικά ένα μικρότερο πλήθος αποχρώσεων για την απόδοση της αρχικής εικόνας.

W_0	w_1	W_2	w_3	W_4	W_5	W_6	w_7
69	111	153	200	218	231	240	246

Πίνακας 3.2.3.1: Οι τιμές των συντελεστών του ΝΔ μετά την εκπαίδευση.

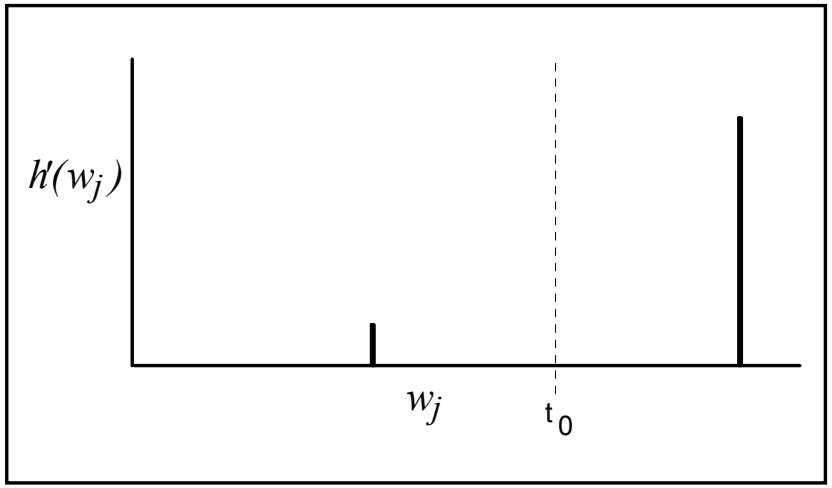
T_0	T_1	T_2	T_3	T_4	T_5	T_6
90	132	176	209	224	235	243

Πίνακας 3.2.3.2. Οι τιμές των κατωφλίων.

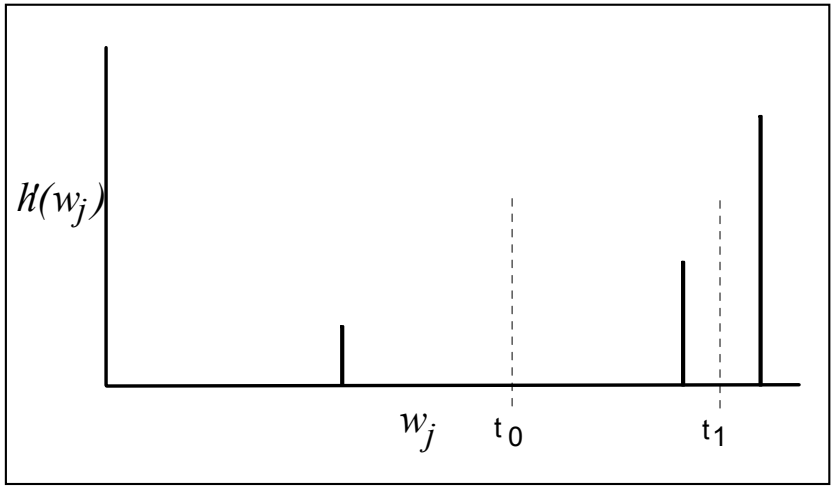
Στον Πιν. 3.2.3.3 φαίνονται οι τιμές των κατωφλίων για διαφορετικό πλήθος (*J*) νευρώνων του νευρωνικού δικτύου

J	T_0	T_1	T_2	T_3	T_4
2	164				
3	151	227			
4	106	179	229		
5	106	178	224	237	
6	97	143	191	225	239

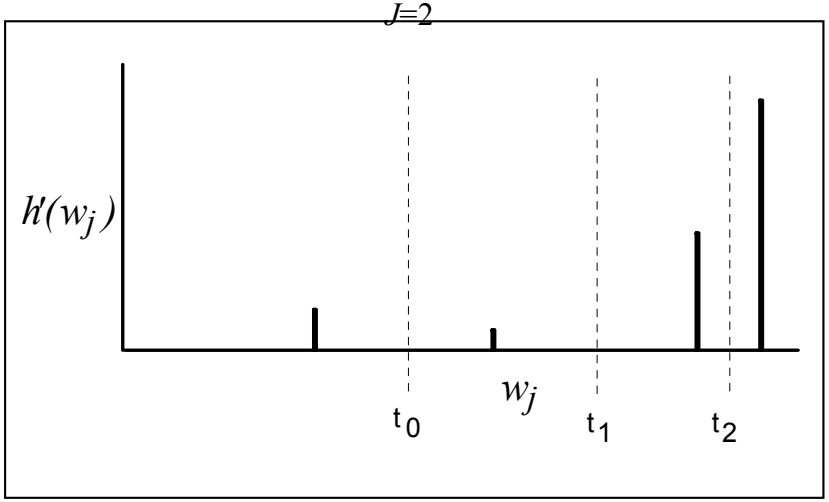
Πίνακας 3.2.3.3. Οι τιμές των κατωφλίων για $J=2,\ldots,6$.



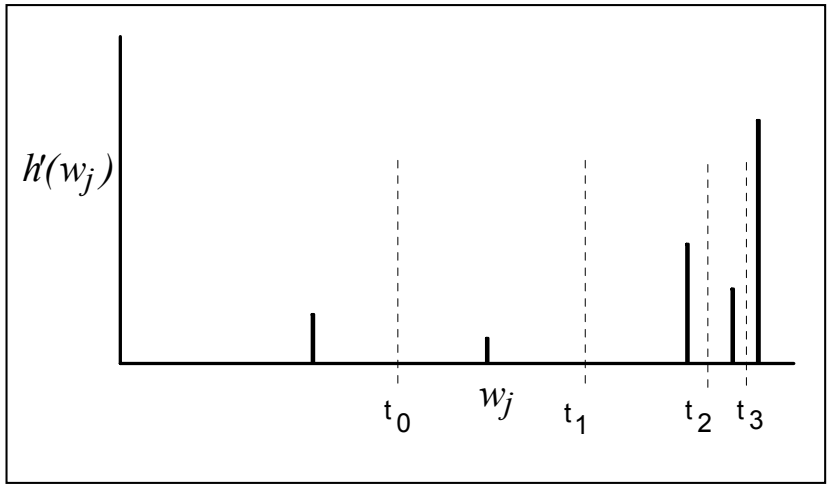
J=2



J=3



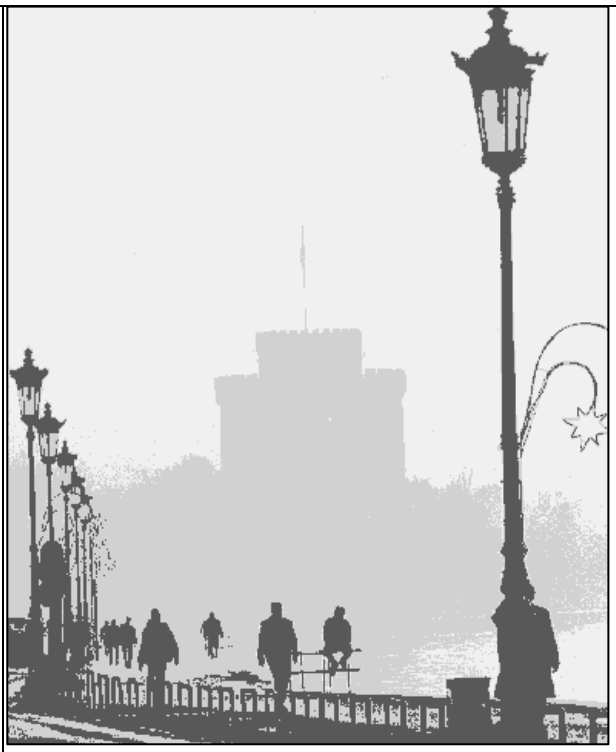
J=4



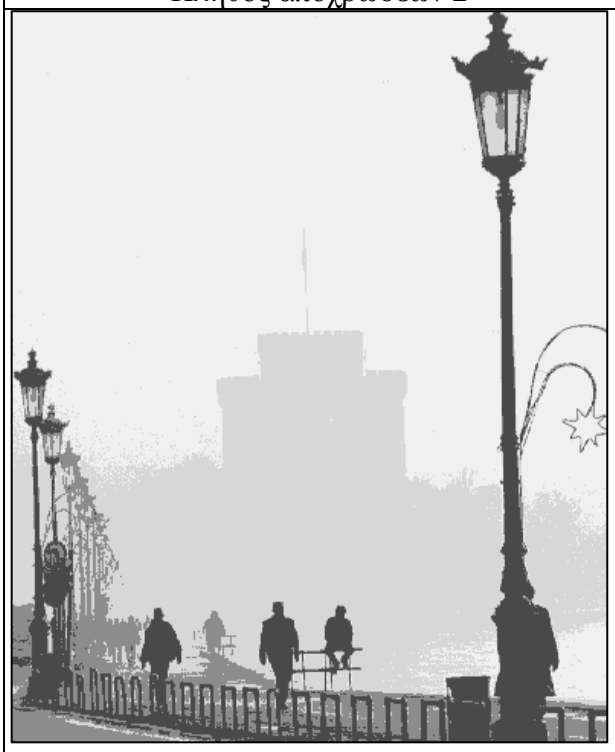
J=5

Σχήμα 3.2.3.3. Τα ιστογράμματα και οι αντίστοιχες τιμές κατωφλίων για διαφορετικές τιμές του *J*.

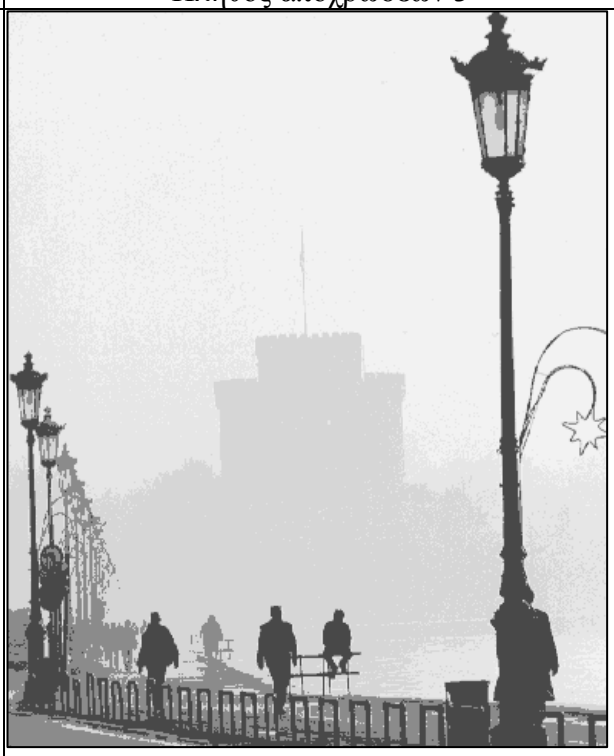
Πλήθος αποχρώσεων 2



Πλήθος αποχρώσεων 3



Πλήθος αποχρώσεων 3



Πλήθος αποχρώσεων 4

Σχήμα 3.2.3.4:Οι εικόνες που προέκυψαν από την πολυκατωφλίωση της εικόνας του Σχ.3.2.3.1 για διαφορετικό πλήθος κατωφλίων και αποχρώσεων.

Η βέλτιστη τιμή του J για ένα δεδομένο σύνολο τιμών $\{2,\ldots,J_{max}\}$ ελαχιστοποιεί το σφάλμα προσαρμογής μεταξύ των συναρτήσεων $H(\ell)$ και $H_J(\ell)$. Η συνάρτηση $H(\ell)$ υπολογίζεται από τη σχέση

$$H(\ell) = \frac{h(\ell)}{\sum_{\ell=0}^{L-1} h(\ell)} \qquad (3.2.3.8)$$

και είναι συνάρτηση πυκνότητας πιθανότητας επειδή $\sum_{\ell=0}^{L-1} H(\ell) = 1$.

Η πιθανότητα f_j εμφάνισης της κλάσης C_j, η μέση τιμή μ_j της κλάσης και η διασπορά της σ_j^2 δίνονται αντίστοιχα από τις σχέσεις

$$f_j = P_r(C_j) = \sum_{C_j} h(\ell) \qquad (3.2.3.9)$$

$$\mu_j = \frac{\sum_{C_j} \ell h(\ell)}{\sum_{C_j} h(\ell)} \qquad (3.2.3.10)$$

$$\sigma_j^2 = \frac{\sum_{C_j} (\ell - \mu_j)^2 h(\ell)}{\sum_{C_j} h(\ell)} \qquad (3.2.3.11)$$

Η συνάρτηση προσαρμογής $H_J(\ell)$ είναι το άθροισμα

$$H_J(\ell) = \sum_{j=0}^{J-1} f_j H_j(\ell) \qquad (3.2.3.12)$$

όπου

$$h_j(\ell) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(\ell - \mu_j)^2}{2\sigma_j^2} \right) \qquad (3.2.3.13)$$

Μπορεί να δειχθεί εύκολα ότι η $h_J(\ell)$ είναι επίσης συνάρτηση πυκνότητας πιθανότητας. Το σφάλμα προσαρμογής $|e_J|$ για ένα πλήθος κλάσεων J ορίζεται ως

$$|e_J| = \sum_{\ell=0}^{L-1} \left[h(\ell) - h_J(\ell) \right]^2 \qquad (3.2.3.14)$$

και υπολογίζεται για J=2,…,J_{max}. Η τιμή J_o είναι βέλτιστη όταν το αντίστοιχο σφάλμα προσαρμογής $\left|e_{J_o}\right|$ ικανοποιεί τη συνθήκη

$$\left|e_{J_o}\right| = minimum\{\left|e_J\right|\} \qquad (3.2.3.15)$$

Ο Πίν. 3.2.3.3 δείχνει τις τιμές του $\left|e_J\right|$ για J=2,…,6. Η μικρότερη από αυτές τις τιμές προκύπτει για $J=3$ και θεωρείται ότι είναι η βέλτιστη τιμή J_o. Όπως μπορεί να παρατηρηθεί από το ιστόγραμμα $h(\ell)$ (Σχ. 3.2.3.3), αυτή η τιμή είναι κατάλληλη επειδή αντιστοιχεί στο πλήθος των σημαντικών υψωμάτων του ιστογράμματος.

3.3 Κωδικοποίηση εικόνων αποχρώσεων του γκρι

Για την αναπαράσταση μια ψηφιακής εικόνας J γραμμών και K στηλών με βάθος χρώματος (bit) B bits απαιτούνται $J \times K \times B$ bits. Η κωδικοποίηση μιας συγκεκριμένης ψηφιακής εικόνας με τέτοιο τρόπο δεν είναι η πλέον κατάλληλη όταν ενδιαφέρει το πλήθος των bits που δαπανώνται. Αυτό συχνά συμβαίνει σε περιπτώσεις μετάδοσης ή αποθήκευσης της εικόνας. Με τον όρο *συμπίεση της εικόνας* αναφερόμαστε σε ένα πλήθος τεχνικών που στοχεύουν στην μείωση του πλήθους των bits που απαιτούνται για την κωδικοποίηση της εικόνας. Οι τεχνικές αυτές διακρίνονται σ αυτές που επιτρέπουν απώλεια πληροφορίας και αυτές που δεν επιτρέπουν. Στην πρώτη περίπτωση η πληροφορία που χάνεται δεν θα πρέπει να είναι ζωτικής σημασίας για τον παρατηρητή της εικόνας. Ακολούθως θα περιγραφούν αλγόριθμοι συμπίεσης ψηφιακών εικόνων αποχρώσεων του γκρι.

3.3.1 Κωδικοποίηση Huffman

Σύμφωνα με την τεχνική αυτή κωδικοποιούμε τις αποχρώσεις του γκρί μιας συγκεκριμένης εικόνας με κωδικές λέξεις μεταβλητού μήκους. Μια απόχρωση που εμφανίζεται συχνά (με μεγάλη πιθανότητα) στην εικόνα κωδικοποιείται με λέξη μήκους μικρότερου από άλλες αποχρώσεις που εμφανίζονται σπανιότερα. Το κανονικοποιημένο ιστόγραμμα $h(g)$ των αποχρώσεων της ψηφιακής εικόνας παρέχει τις τιμές της πιθανότητας εμφάνισης $p(g)$ των τιμών της μεταβλητής g των αποχρώσεων της εικόνας διότι $p(g)=h(g)$. Αν $l(g)$ είναι το μήκος της κωδικής λέξης της g απόχρωσης και G το πλήθος των αποχρώσεων, τότε το μέσο μήκος του κώδικα που απαιτείται για την εικόνα είναι

$$\bar{l} = \sum_{g=0}^{G-1} p(g)l(g) = \sum_{g=0}^{G-1} h(g)l(g) \qquad (3.3.1.1)$$

Ένας κατάλληλος ευκρινής, μονοσήμαντος και στιγμιαία αποκωδικοποιήσιμος κώδικας που ελαχιστοποιεί το μέσο μήκος είναι ο κώδικας *Huffman.* Αν η εικόνα κωδικοποιηθεί σύμφωνα με αυτόν η τιμή του σύμφωνα με την θεωρία της πληροφορίας θα ικανοποιεί τη σχέση

$$H(G) \le \bar{l} \le H(G) + 1 \qquad (3.3.1.2)$$

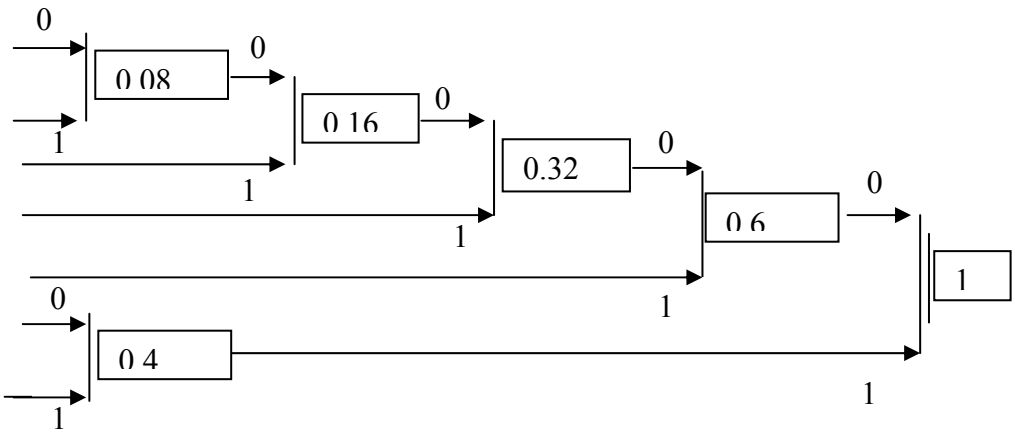
Όπου *H(G)* η εντροπία της εικόνας αν την θεωρήσουμε ως πηγή πληροφορίας που δίνεται από την σχέση

$$H(G) = -\sum_{g=0}^{G-1} p(g)\log_2 p(g) = -\sum_{g=0}^{G-1} h(g)\log_2 h(g) \qquad (3.3.1.3)$$

Ο εύρεση του κώδικα *Huffman* για μια συγκεκριμένη εικόνα γίνεται με την δημιουργία ενός δυαδικού δένδρου. Τα φύλλα του δένδρου είναι οι τιμές των αποχρώσεων με τις αντίστοιχες πιθανότητες εμφάνισης τους στην εικόνα. Τα φύλλα συνενώνονται σε δυαδικούς κόμβους. Σε κάθε κόμβο αντιστοιχίζεται το άθροισμα των πιθανοτήτων εμφάνισης των παιδιών του. Ο πρώτος κόμβος δημιουργείται από τη συνένωση των δύο αποχρώσεων (φύλλα) με τις μικρότερες πιθανότητες. Κάθε επόμενος κόμβος έχει δύο παιδιά που επιλέγονται από φύλλα ή διαθέσιμους κόμβους που δεν ενώθηκαν και έχουν τις μικρότερες πιθανότητες εμφάνισης.. Στο ακόλουθο παράδειγμα παρουσιάζεται εφαρμογή της κωδικοποίησης σε μια εικόνα αποχρώσεων του γκρι διαστάσεων 4 × 5.

100	100	100	100	100
200	10	20	150	100
200	50	50	150	100
200	200	200	150	150
80	80	80	80	80

G	*H(g)*	*h(g)*
10	1	0.04
20	1	0.04
50	2	0.08
80	4	0.16
100	7	0.28
150	5	0.2
200	5	0.2
	25	1



g	10	20	50	80	100	150	200
Κωδικός	00000	00001	0001	001	01	10	11

Για την εικόνα του παραδείγματος απαιτούνται 1×5+1×5+2×4+4×3+7×2+5×2+5×2=54 bits αντί των 25×3=75 bits που απαιτούνται για κωδικές λέξεις σταθερού μήκους τριών bits (3=[$\log_2$7]+1). Το μέσος μήκος κωδικολέξης $\bar{l} = 54/25 = 2.16$ και ο ποσοστό συμπίεσης είναι (75-54)/75%=28%

3.3.2 Κωδικοποίηση 2Δ-διακριτού μετασχηματισμό συνημιτόνου

Από όσα αναφέρθηκαν προηγουμένως η διαδικασία της κωδικοποίησης σχετιζόταν άμεσα με τις τιμές των εικονοστοιχείων της εικόνας. Στην ενότητα αυτή θα δούμε τον μετασχηματισμό του διδιάστατου ψηφιακού σήματος με την εύρεση των προβολών του (συντελεστών) σε μία ορθοκανονική βάση, πριν ακολουθήσει η διαδικασία κωδικοποίησης

Ο διδιάστατος μετασχηματισμός συνημιτόνου (2D-DCT: Discrete Cosine Transform) αποτελεί την βάση για συμπίεση μη δυαδικών εικόνων με αποδεκτή απώλεια πληροφορίας. Η τυποποίηση JPEG βασίζεται στις πρώτες εκδόσεις της στον 2D-DCT. Θα αρχίσουμε την παρουσίαση του μετασχηματισμού από την μονοδιάστατη εκδοχή του.

Οι σχέσεις

$$g_0(n) = \frac{1}{\sqrt{N}},\ g_k(n) = \sqrt{\frac{2}{N}}\cos(\frac{\pi}{2N}k(2n+1)) \qquad (3.3.2.1)$$

για k=1..N-1 και n=0,..,N-1 είναι ένας πυρήνας (kernel) δημιουργίας ορθοκανονικών διανυσμάτων για κάθε τιμή του N. Οι συντελεστές c_k για ένα διακριτό σήμα x(n) πεπερασμένου μήκους N (δηλαδή οι προβολές του **x** στα $\mathbf{g}_k$) προκύπτουν από τη σχέση:

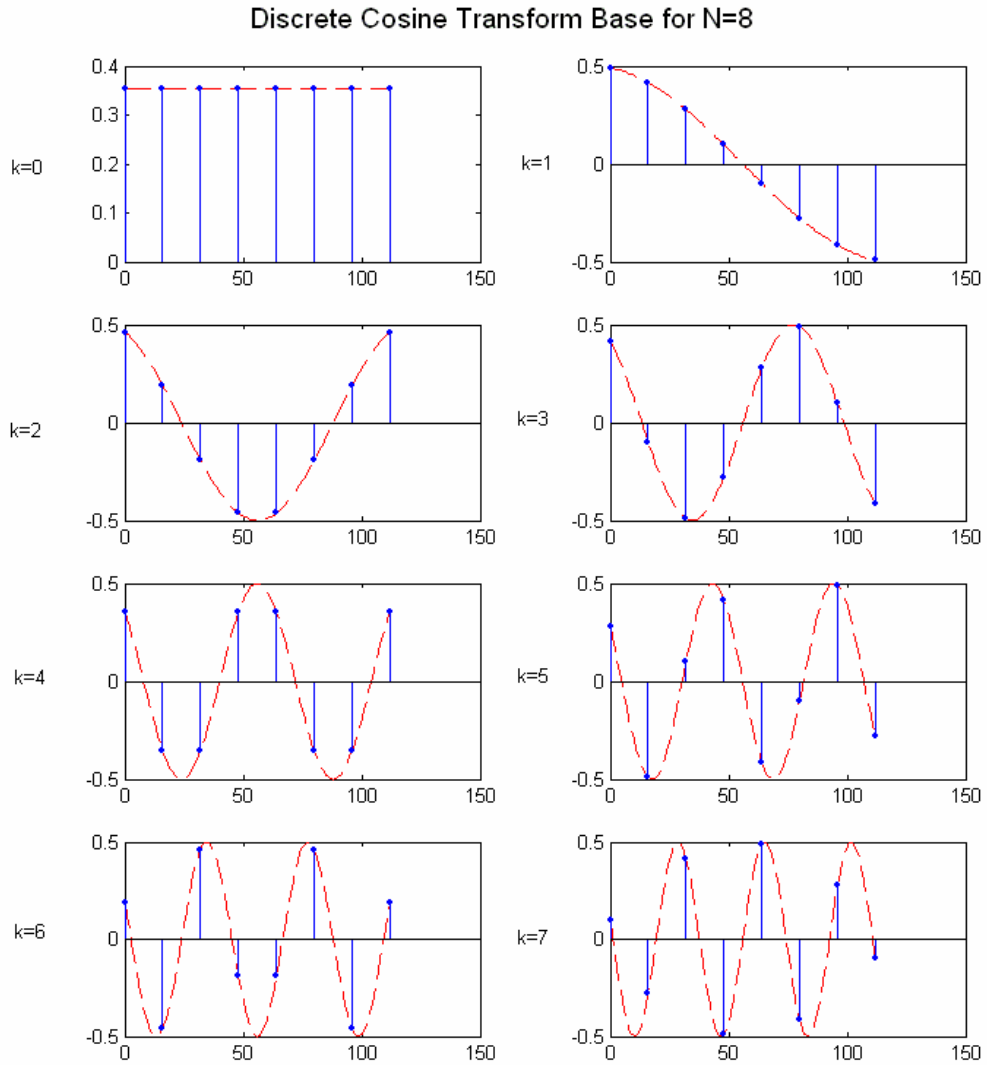
$$c_0 = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} x[n],\ c_k = \sqrt{\frac{2}{N}}\sum_{n=0}^{N-1}\cos\left(\frac{\pi}{2N}k\cdot(2n+1)\right)\cdot x[n] \qquad (3.3.2.1)$$

Η σχέση αυτή αποτελεί τον μονοδιάστατο διακριτό μετασχηματισμό συνημιτόνου (1D-DCT: Discrete Cosine Transform). Οι τιμές x(n) (δηλαδή το **x**) προκύπτουν από τη σχέση

$$x(n) = \frac{c_0}{\sqrt{N}} + \sqrt{\frac{2}{N}}\sum_{k=1}^{N-1} c_k\cdot\cos\left(\frac{\pi}{2N}k(2n+1)\right) \qquad (3.3.2.2)$$

Η σχέση αυτή αποτελεί τον αντίστροφο διακριτό μετασχηματισμό του συνημιτόνου.

Στο Σχ. 3.3.2.1 δείχνονται τα ανύσματα βάσης του DCT για N=8.



Σχήμα 3.3.2.1

Παράδειγμα 3.3.2.1

Για N=3 τα ανύσματα

$$\mathbf{g}_0=[\frac{1}{\sqrt{3}} \quad \frac{1}{\sqrt{3}} \quad \frac{1}{\sqrt{3}}],\ \mathbf{g}_1=[\frac{1}{\sqrt{2}} \quad 0 \quad -\frac{1}{\sqrt{2}}],\ \mathbf{g}_2=[\frac{1}{\sqrt{6}} \quad -\frac{\sqrt{2}}{\sqrt{3}} \quad \frac{1}{\sqrt{6}}]$$

προήλθαν από τις σχέσεις

$$g_0(n)=\frac{1}{\sqrt{N}},\ g_k(n)=\sqrt{\frac{2}{N}}\cos(\frac{\pi}{2N}k(2n+1))\ \text{για k=1..N-1 και n=0,..,N-1}$$

Και αποτελούν μία ορθοκανονική βάση και το σήμα $\mathbf{x}=[1,\ 3,\ 2]^T$ προβάλλεται σ' αυτά με προβολές που δίνουν οι συντελεστές $c_0=2\sqrt{3}$, $c_1=-\sqrt{2}/2$, $c_2=-\sqrt{6}/2$. Για n=0,1,2

$$x[n]=\ 2\sqrt{3}g_0[n]-\sqrt{2}/2g_1[n]-\sqrt{6}/2g_2[n]$$

Παράδειγμα 3.3.2.2

Για $\mathbf{x}=[1,\ 3,\ 2,\ -1]^T$ η τιμή N=4 και για n=0,1,2,3 τα ανύσματα βάσης είναι:

$$g_0[n]=\frac{1}{2} \qquad \text{Ή} \qquad \mathbf{g}_0=[0.50,\ 0.50,\ 0.50,\ 0.50]^T$$

$$g_1[n]=\sqrt{\frac{2}{N}}\cos(\frac{\pi}{2N}(2n+1)) \qquad \text{Ή} \qquad \mathbf{g}_1=[0.65,\ 0.27,\ -0.27,\ -0.65]^T$$

$$g_2(n)=\frac{1}{2}\cos(\frac{\pi}{2N}2\cdot(2n+1)) \qquad \text{Ή} \qquad \mathbf{g}_2=[0.50,\ -0.50,\ 0.50,\ -0.50]^T$$

$$g_3(n)=\frac{1}{2}\cos(\frac{\pi}{2N}3\cdot(2n+1)) \qquad \text{Ή} \qquad \mathbf{g}_3=[0.27,\ -0.65.\ 0.65,\ -0.27]^T$$

Οι συντελεστές c_k είναι τα στοιχεία του πίνακα

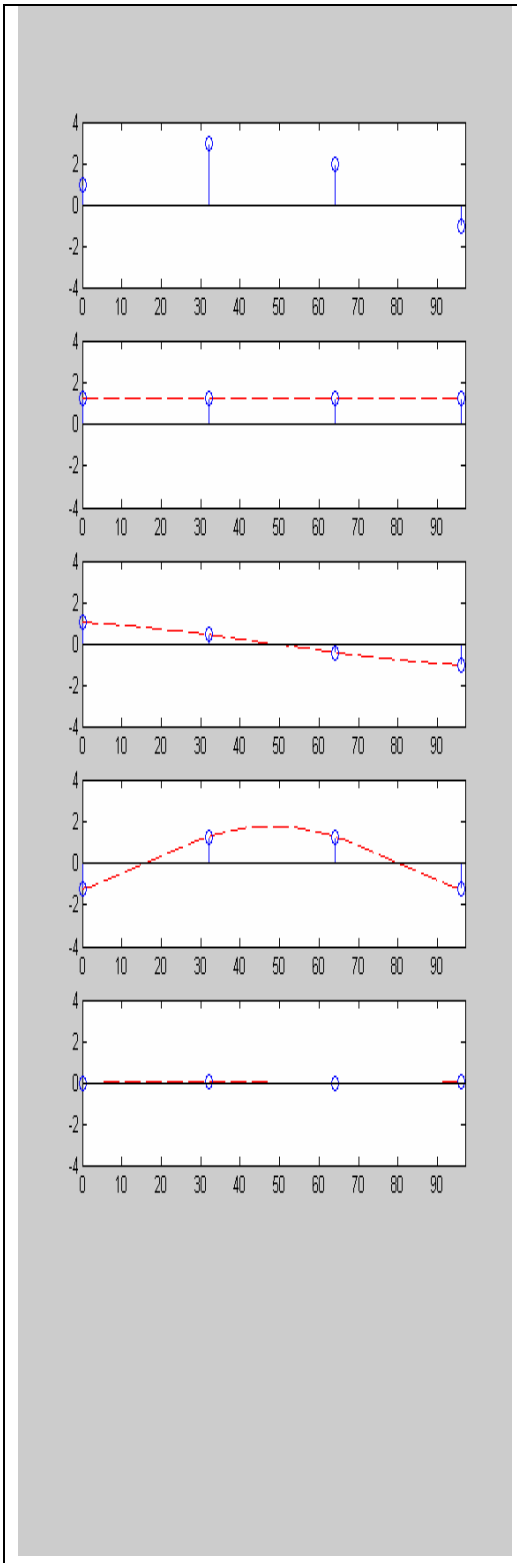
$$\mathbf{C}=\mathbf{G}^T\cdot\mathbf{x}=\begin{bmatrix}0.50 & 0.50 & 0.50 & 0.50\\ 0.65 & 0.27 & -0.27 & -0.65\\ 0.50 & -0.50 & 0.50 & -0.50\\ 0.27 & -0.65 & 0.65 & -0.27\end{bmatrix}\cdot\begin{Vmatrix}1\\2\\3\\-1\end{Vmatrix}=\begin{bmatrix}2.5\\1.58\\-2.5\\-0.11\end{bmatrix}$$

οι συντελεστές c_k προκύπτουν όμοια και από τις (3.3.2.1) Το σήμα ανακτάται από τις σχέσεις (3.3.2.2) ή από τον πίνακα

$$\mathbf{x} = \mathbf{G} \cdot \mathbf{C} = \begin{bmatrix} 0.50 & 0.65 & 0.50 & 0.27 \\ 0.50 & 0.27 & -0.50 & -0.65 \\ 0.50 & -0..27 & 0.50 & 0.65 \\ 0.50 & -0.65 & -0.50 & -0.27 \end{bmatrix} \cdot \begin{bmatrix} 2.5 \\ 1.58 \\ -2.5 \\ -0.11 \end{bmatrix} = \begin{Vmatrix} 1 \\ 2 \\ 3 \\ -1 \end{Vmatrix}$$

Στο Σχ. 3.3.2.2 δείχνονται τα ανύσματα βάσης του DCT για N=4, οι συντελεστές c_k του DCT του σήματος **x**=[1, 3, 2, -1] και η ανάλυση του στις τέσσερις συνιστώσες $c_k\mathbf{g}_k$. Κάθε τιμή του x[n] είναι $x[n]=2,5g_0[n]+ 1,58g_1[n]-2,5g_2[n]-0,11g_3[n]$.

.



x=[1, 3, 2, –1]

$\mathbf{g}_1[n]$

$2{,}5 \cdot g_1[n]$

$\mathbf{g}_2[n]$

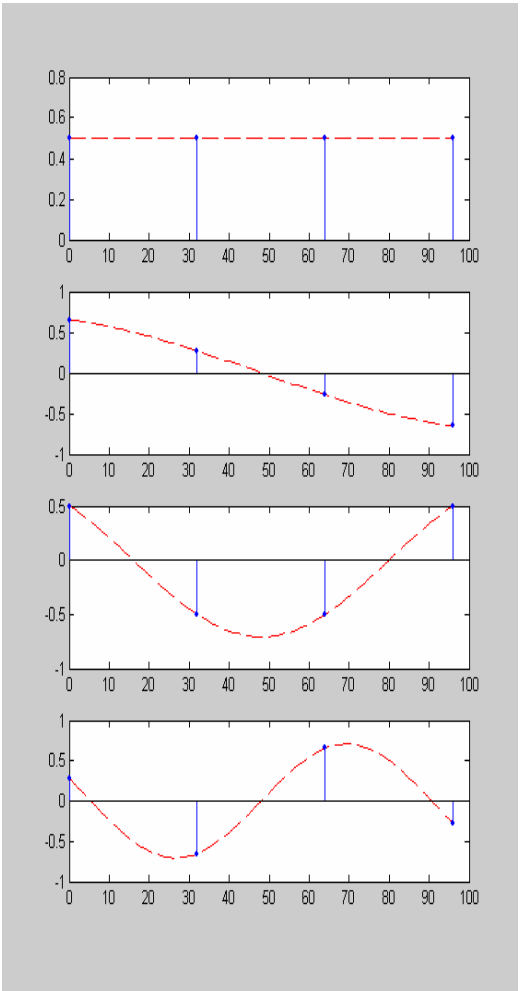
$1{,}58 \cdot g_2[n]$

$\mathbf{g}_2[n]$

$-2{,}5 \cdot g_2[n]$

$\mathbf{g}_3[n]$

$-0{,}11 \cdot g_3[n]$



Σχήμα 3.3.2.2

Για διακριτά σήματα δύο διαστάσεων με μήκη N_1, N_2 τα ανύσματα βάσης για $n_1=0,..,N_1-1$, $n_2=0,..,N_2-1$, $k_1=0,..,N_1-1$, $k_2=0,..,N_2-1$ είναι:

$$g_{k_1k_2}(n_1,n_2) = g_{k_1}(n_1)\cdot g_{k_2}(n_2)$$

ήτοι (3.3.2.3)

$$g_{00}(n_1,n_2) = \frac{1}{\sqrt{N_1N_2}},$$

$$g_{k_10}(n_1,n_2) = \frac{1}{\sqrt{N_2}}\sqrt{\frac{2}{N_1}}\cdot\cos(\frac{\pi k_1(2n_1+1)}{2N_1}),$$

$$g_{0k_2}(n_1,n_2) = \frac{1}{\sqrt{N_1}}\sqrt{\frac{2}{N_2}}\cdot\cos(\frac{\pi k_2(2n_2+1)}{2N_2}),$$

$$g_{k_1k_2}(n_1,n_2) = \sqrt{\frac{2}{N_1}}\cdot\cos(\frac{\pi k_1(2n_1+1)}{2N_1})\cdot\sqrt{\frac{2}{N_2}}\cdot\cos(\frac{\pi k_2(2n_2+1)}{2N_2})$$

Για ένα δισδιάστατο διακριτό σήμα $x(n_1,n_2)$ με μήκη N_1, N_2, οι προβολές του στα παραπάνω ορθοκανονικά ανύσματα βάσης είναι:

$$c_{00} = \frac{1}{\sqrt{N_1N_2}}\sum_{n_2=0}^{N_2-1}\sum_{n_1=0}^{N_1-1}x[n_1,n_2]$$

$$c_{0k_2} = \sqrt{\frac{2}{N_1N_2}}\sum_{n_2=0}^{N_2-1}\sum_{n_1=0}^{N_1-1}x[n_1,n_2]\cdot\cos(\frac{\pi k_2(2n_2+1)}{2N_2})$$

$$c_{k_10} = \sqrt{\frac{2}{N_1N_2}}\sum_{n_2=0}^{N_2-1}\sum_{n_1=0}^{N_1-1}x[n_1,n_2]\cdot\cos(\frac{\pi k_1(2n_1+1)}{2N_1}) \qquad (3.3.2.4)$$

$$c_{k_1k_2} = \frac{2}{\sqrt{N_1N_2}}\sum_{n_2=0}^{N_2-1}\sum_{n_1=0}^{N_1-1}x[n_1,n_2]\cdot\cos(\frac{\pi k_1(2n_1+1)}{2N_1})\cdot\cos(\frac{\pi k_2(2n_2+1)}{2N_2})$$

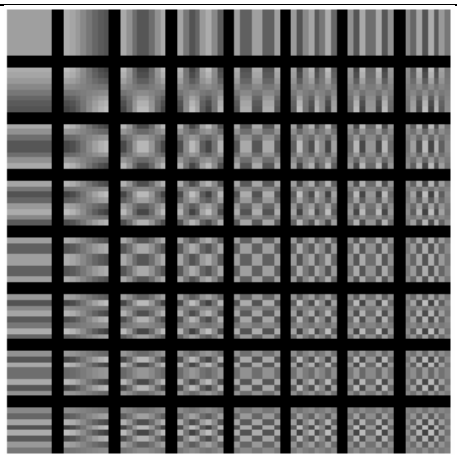
για $k_1,k_2 \neq 0$.

Οι παραπάνω σχέσεις αποτελούν τον δισδιάστατο (2Δ) διακριτό μετασχηματισμό συνημιτόνου (2D-DCT: Discrete Cosine Transform). Το σήμα $x[n_1,n_2]$ ανακτάται σύμφωνα με τη σχέση:

$$x(n_1,n_2) = \sum_{n_2=0}^{N_2-1}\sum_{n_1=0}^{N_1-1}c_{k_1k_2}\cdot g_{k_1k_2}(n_1,n_2) \qquad (3.3.2.5)$$

που είναι ο αντίστροφος δισδιάστατος (2Δ) διακριτός μετασχηματισμός συνημιτόνου.

Είναι προφανές από τα παραπάνω αν M×N είναι τα εικονοστοιχεία της εικόνας τότε M×N θα είναι και όλοι οι συντελεστές του 2D-DCT αυτής. Για παράδειγμα για μία εικόνα 8×8, ο μετασχηματισμός οδηγεί στον υπολογισμό 8×8 συντελεστών (προβολών) για κάθε μία από τις 8×8 διακριτές διδιάστατες ακολουθίες βάσης. Κάθε ακολουθία της βάσης έχει τιμές που δίνονται από τις (3.3.2.3) που αν παρασταθούν με τόνους του γκρι προκύπτουν οι εικόνες του Σχ.3.3.2.3.

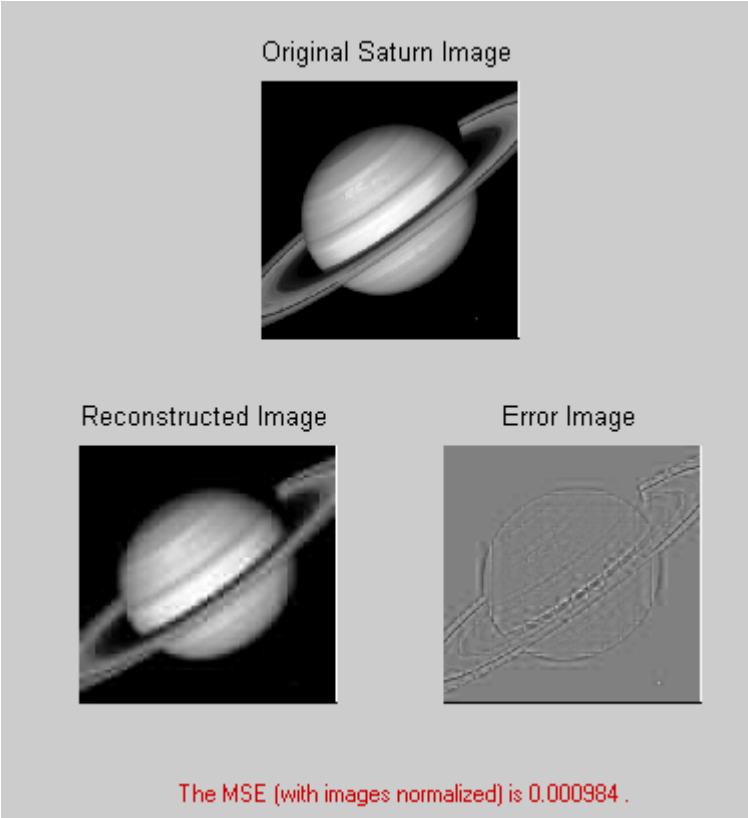


Σχήμα 3.3.2.3

Μετασχηματισμοί των οποίων η βάση ικανοποιεί την (3.3.2.1) λέγονται διαχωρίσιμοι και οι υπολογισμοί τους μπορούν να αναχθούν σε υπολογισμούς των μονοδιάστατων εκφράσεών τους. Η εφαρμογή του 2D-DCT στην συμπίεση εικόνων αποχρώσεων του γκρι σύμφωνα με την τυποποίηση JPEG (Join Photographic Experts Group) ακολουθεί τα ακόλουθα γενικά βήματα:

- Η εικόνα χωρίζεται σε μη επικαλυπτόμενα τμήματα 8×8 εικονοστοιχείων.
- Οι τιμές των εικονοστοιχείων μετατρέπονται από το διάστημα $[0,2^L-1]$ στο διάστημα $[-2^{L-1}, 2^{L-1}-1]$ με την αφαίρεση από κάθε τιμή του αριθμού 2^{L-1}, όπου L το βάθος χρώματος της εικόνας (bits/pel).
- Για κάθε τμήμα υπολογίζονται οι συντελεστές του 2D-DCT.
- Ακολούθως οι τιμές κβαντίζονται και τέλος
- Κωδικοποιούνται κατά Huffman.

Στο Σχ. 3.2.3.4 η αρχική εικόνα 256×256 pels χωρίζεται σε τμήματα των 8×8 pels και σε κάθε ένα από αυτά εφαρμόζεται ο 2D-DCT, ακολούθως με χρήση μόνο των 8 συντελεστών αναδημιουργείται η εικόνα με σφάλμα όπως φαίνεται στο σχήμα.


Original Saturn Image
Reconstructed Image
Error Image
The MSE (with images normalized) is 0.000984 .

Σχήμα 3.2.3.4

3.4 Προσδιορισμός ακμών.

Σε μια εικόνα αποχρώσεων του γκρι υπάρχουν περιοχές εικονοστοιχείων με απότομη αύξηση της φωτεινότητας. Οι περιοχές αυτές βρίσκονται στα όρια των τμημάτων της εικόνας που έχουν σημαντικά διαφορετικές αποχρώσεις. Η ανίχνευση των ορίων αυτών λέγεται *προσδιορισμός των ακμών της εικόνας* (*edge detection*). Η ανίχνευση ακμών είναι εξαιρετικά χρήσιμη εργασία στην ανάλυση των εικόνων διότι μέσω αυτής προσδιορίζονται τα περιγράμματα των αντικειμένων της εικόνας. Υπάρχει πληθώρα αλγορίθμων που αφορούν την επίλυση του προβλήματος, όμως όλοι βασίζονται στην έννοια της κλίσης της συνάρτησης φωτεινότητας I(k,j) στη θέση (k, j) ενός εικονοστοιχείου της εικόνας. Το αποτέλεσμα της όλης εργασίας είναι μια νέα δυαδική εικόνα, ιδίων διαστάσεων με την αρχική, όπου τα εικονοστοιχεία του αντικειμένου είναι οι ακμές της αρχικής εικόνας, Στο ακόλουθο σχήμα φαίνονται σε μία διάσταση, τύποι ακμών που διαφέρουν ως προς την κλίση τους.

Η κλίση μιας συνεχούς συνάρτησης $f(x,y):\Re^2 \to \Re$ είναι ένα διάνυσμα στο επίπεδο xy που ορίζεται σύμφωνα με την σχέση

$$\nabla f(x,y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Το μέτρο του διανύσματος κλίσης σε ένα σημείο (x,y) δείχνει πόσο απότομη είναι η ακμή και δίνεται από την σχέση

$$|\nabla f(x,y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

Η κλίση στο σημείο (x,y) είναι μέγιστη αν κινηθούμε κατά την διεύθυνση του xy επιπέδου που προσδιορίζει η γωνία φ σύμφωνα με την σχέση

$$\phi = \tau o \xi \varepsilon \phi \left(\frac{\partial f}{\partial y} \Big/ \frac{\partial f}{\partial x} \right)$$

Μια ψηφιακή εικόνα δεν είναι συνεχής συνάρτηση και ως εκ τούτου αντί των μερικών παραγώγων υπολογίζουμε τις διαφορές της απόχρωσης ενός εικονοστοιχείου με τα γειτονικά του, κατά την οριζόντια και κάθετη κατεύθυνση της εικόνας. Αν εφαρμοσθεί η προσέγγιση αυτή σε ένα εικονοστοιχείο (k, j) με το προηγούμενό του, στον πίνακα I μιας ψηφιακής εικόνας, τότε η κλίση σ'αυτό καθορίζεται από το διάνυσμα

$$\nabla_1 I(k,j) = (\nabla_{1K} I(k,j), \nabla_{1J} I(k,j))$$
$$\nabla_{1K} I(k,j) = I(k,j) - I(k-1,j)$$
$$\nabla_{1J} I(k,j) = I(k,j) - I(k,j-1)$$

Όμοια για ένα εικονοστοιχείο και τα εκατέρωθεν αυτού εικονοστοιχεία καθορίζεται η κλίση

$$\nabla_2 I(k,j) = (\nabla_{2K} I(k,j), \nabla_{2J} I(k,j))$$
$$\nabla_{2K} I(k,j) = I(k+1,j) - I(k-1,j)$$
$$\nabla_{2J} I(k,j) = I(k,j+1) - I(k,j-1)$$

Τα παραπάνω μπορούν να εκφρασθούν και με την χρήση μασκών όπως ακολούθως

$\nabla_{1K} =$

-1	1

$\nabla_{1J} =$

-1
1

$\nabla_{2K} =$

-1	0	1

$\nabla_{2J} =$

-1
0
1

Αν εφαρμόσουμε όλες τις παραπάνω μάσκες σε μία εικόνα 3Χ3, και αντιστοιχίσουμε το άθροισμα των αποτελεσμάτων στο κεντρικό εικονοστοιχείο, θα προκύψουν για κάθε κατεύθυνση αντίστοιχα οι τιμές

$$S_k = 1 \cdot I_{(k+1,j+1)} + 2 \cdot I_{(k+1,j)} + 1 \cdot I_{(k+1,j-1)} - 1 \cdot I_{(k-1,j+1)} - 2 \cdot I_{(k-1,j)} - 1 \cdot I_{(k-1,j-1)}$$
$$S_j = 1 \cdot I_{(k+1,j+1)} + 2 \cdot I_{(k+1,j)} + 1 \cdot I_{(k+1,j-1)} - 1 \cdot I_{(k+1,j-1)} - 2 \cdot I_{(k,j-1)} - 1 \cdot I_{(k-1,j-1)}$$

ή με την μορφή μάσκας

$S_{k=}$

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

$S_j=$

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

Η μάσκα αυτή λέγεται *ανιχνευτής ακμών του Sobel* (*Sobel edge detector*) και λειτουργεί ως διάνυσμα (S_k, S_j) με τις οικίες τιμές κατεύθυνσης και πλάτους. Πολλές μάσκες έχουν προταθεί από τους ερευνητές για την ανίχνευση ακμών βασισμένες σε διαφορετικές προσεγγίσεις για την εξυπηρέτηση διαφόρων ιδιαιτεροτήτων των εικόνων. Σε κάθε περίπτωση από την αρχική εικόνα παράγεται ένας πίνακας τιμών του μέτρου της κλίσης σε κάθε εικονοστοιχείο της εικόνας.

Ακολούθως μία τιμή κατωφλίου καθορίζει εκείνα τα εικονοστοιχεία που αντιστοιχούν σε ακμές και παράγουν την δυαδική εικόνα των ακμών.

Η επιλογή της κατάλληλης τιμής κατωφλίου είναι ένα επιπρόσθετο πρόβλημα. Η μεταβολή του φωτισμού, η μεταβολή της απόχρωσης του αντικειμένου ή του παρασκηνίου, απαιτούν την χρήση διαφορετικών τοπικών τιμών κατωφλίων.

Η εύρεση των περιοχών που το μέτρο της κλίσης (άνυσμα πρώτων μερικών παραγώγων) έχει τοπικά μεγάλες τιμές μπορεί βασισθεί στην εύρεση περιοχών όπου η δεύτερες μερικές παράγωγοι της συνάρτησης της εικόνας μηδενίζονται. Σε μία ψηφιακή εικόνα η δεύτερη παράγωγος κατά την οριζόντια διεύθυνση βασίζεται στην διπλή διαδοχική εφαρμογή του τελεστή ∇_{1K}. Συγκεκριμένα η $\frac{\partial^2 f(x,y)}{\partial x^2}$ αντικαθίσταται από την διαφορά

$\nabla_{1K}(k+1,j)- \nabla_{1K}(k,j)= I(k+1,j)-I(k,j)-(I(k,j)-I(k-1,j))= I(k+1,j)-2I(k,j)+I(k-1,j)$

Όμοια η $\frac{\partial^2 f(x,y)}{\partial y^2}$ αντικαθίσταται από την διαφορά

$\nabla_{1J}(k,j+1)- \nabla_{1J}(k,j)= I(k,j+1)-I(k,j)-(I(k,j)-I(k,j-1))= I(k,j+1)-2I(k,j)+I(k,j-1)$

Μπορούμε να εφαρμόσουμε τα παραπάνω σε μία μάσκα 3 X 3 με τιμές

$\nabla^2 =$

	1	
1	-4	1
	1	

Η μάσκα αυτή είναι μία έκφραση του τελεστή Laplace μιας συνάρτησης f(x,y)

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}$$

στο πίνακα I(k,j). Οι ακμές αποδίδονται στα εικονοστοιχεία που αντιστοιχούν σε μηδενικές τιμές του πίνακα που προέρχεται από την εφαρμογή του ∇^2 ή στα εικονοστοιχεία που έχουν διαφορετικό πρόσημο με τα γειτονικά τους.

Το πλεονέκτημα του τελεστή Laplace είναι ότι βρίσκει τα τοπικά μέγιστα της πρώτης παραγώγου και δημιουργεί λεπτές ακμές. Μειονεκτεί διότι είναι ευαίσθητος στο θόρυβο και στις μικρές αυξομειώσεις της έντασης της φωτεινότητας. Για το λόγο αυτό εφαρμόζεται σε συνδυασμό με τον ∇_1 ή ∇_2, τεχνικές τοπικής διασποράς και αφαίρεσης θορύβου.

Για το πρόβλημα ανίχνευσης των ακμών χρησιμοποιούνται σήμερα πιο πολύπλοκες τεχνικές που αντιμετωπίζουν προβλήματα όπως θορύβου, αλλοίωσης του φωτισμού κ.α. Από τις πλέον επιτυχημένες και διαδεδομένες είναι του *Canny* και θα την παρουσιάσουμε ακολούθως.

ΠΑΡΑΔΕΙΓΜΑ: Δίνεται ο ακόλουθος πίνακας ψηφιακής εικόνας. Να ανιχνευθούν οι ακμές του με την κλίση που καθορίζει ο $\nabla_2 I$.

I

80.0	80.0	100.0	100.0	80.0	100.0	100.0
80.0	100.0	100.0	80.0	80.0	80.0	100.0
80.0	100.0	220.0	220.0	200.0	220.0	80.0
80.0	200.0	220.0	200.0	220.0	200.0	80.0
80.0	80.0	200.0	220.0	220.0	100.0	100.0
80.0	100.0	100.0	80.0	100.0	80.0	100.0
80.0	80.0	80.0	80.0	100.0	80.0	100.0

$\nabla_{2K} I$

-	-	-	-	-	-	-
-	20.0	-20.0	-20.0	0.0	20.0	-
-	140.0	120.0	-20.0	0.0	-120.0	-
-	140.0	0.0	0.0	0.0	-140.0	-
-	120.0	140.0	20.0	-120.0	-120.0	-
-	20.0	-20.0	0.0	0.0	0.0	-
-	-	-	-	-	-	-

$\nabla_{2J} I$

-	-	-	-	-	-	-
-	20.0	120.0	120.0	120.0	120.0	-
-	100.0	120.0	120.0	140.0	120.0	-
-	-20.0	-20.0	0.0	20.0	-120.0	-
-	-100.0	-120.0	-120.0	-120.0	-120.0	-
-	0.0	-120.0	-140.0	-120.0	-20.0	-
-	-	-	-	-	-	-

$|\nabla_2 I|$

-	-	-	-	-	-	-
-	28.3	121.7	121.7	120.0	121.7	-
-	172.0	169.7	121.7	140.0	169.7	-
-	141.4	20.0	0.0	20.0	184.4	-
-	156.2	184.4	121.7	169.7	169.7	-
-	20.0	121.7	140.0	120.0	20.0	-
-	-	-	-	-	-	-

μ=	115.0	σ=	58.5
Κατώφλι	115.0		

0.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0
1.0	0.0	0.0	0.0	1.0
1.0	1.0	1.0	1.0	1.0
0.0	1.0	1.0	1.0	0.0

ΠΑΡΑΡΤΗΜΑ Α

$$\sigma_B^2 = \rho_1(\mu_1 - \rho_1\mu_1 - \rho_2\mu_2)^2 + \rho_2(\mu_2 - \rho_1\mu_1 - \rho_2\mu_2)^2 = \rho_1(\mu_1(1-\rho_1) - \rho_2\mu_2)^2 + \rho_2(\mu_2(1-\rho_2) - \rho_1\mu_1)^2 =$$
$$= \rho_1(\mu_1\rho_2 - \rho_2\mu_2)^2 + \rho_2(\mu_2\rho_1 - \rho_1\mu_1)^2 = \rho_1\rho_2{}^2(\mu_1 - \mu_2)^2 + \rho_2\rho_1{}^2(\mu_2 - \mu_1)^2 = \rho_1\rho_2(\mu_1 - \mu_2)^2(\rho_1 + \rho_2) =$$
$$= \rho_1\rho_2(\mu_1 - \mu_2)$$

$$\sigma^2 = \sum_{g=0}^{G-1}(g - \mu)^2 H(g) = \sum_{g=0}^{G-1}(g^2 + \mu^2 - 2g\mu)H(g) = \sum_{g=0}^{G-1} g^2 H(g) + \mu^2 \sum_{g=0}^{G-1} H(g) - 2\mu \sum_{g=0}^{G-1} H(g)g =$$
$$= \sum_{g=0}^{G-1} g^2 H(g) + \mu^2 - 2\mu\mu = \sum_{g=0}^{G-1} g^2 H(g) - \mu^2$$

$$\sigma_\omega{}^2 + \sigma_b{}^2 = \rho_1\sigma_1{}^2 + \rho_2\sigma_2{}^2 + \rho_1(\mu_1 - \mu)^2 + \rho_2(\mu_2 - \mu)^2 =$$
$$= \frac{\rho_1 \sum_{g=0}^{T-1}(g - \mu_1)^2 H(g)}{\rho_1} + \frac{\rho_2 \sum_{g=T}^{G-1}(g - \mu_2)^2 H(g)}{\rho_2} + \rho_1(\mu_1{}^2 + \mu^2 - 2\mu\mu_1) + \rho_2(\mu_2{}^2 + \mu^2 - 2\mu\mu_2) =$$
$$= \sum_{g=0}^{T-1}(g^2 + \mu_1{}^2 - 2g\mu_1)H(g) + \sum_{g=T}^{G-1}(g^2 + \mu_2{}^2 - 2g\mu_2)H(g) + \rho_1\mu_1{}^2 + \rho_1\mu^2 - 2\mu\mu_1\rho_1 + \rho_2\mu_2{}^2 + \rho_2\mu^2 + 2\mu\mu_2\rho_2 =$$
$$= \sum_{g=0}^{T-1} g^2 H(g) + \sum_{g=0}^{T-1} \mu_1{}^2 H(g) - \sum_{g=0}^{T-1} 2g\mu_1 H(g) + \sum_{g=T}^{G-1} g^2 H(g) + \sum_{g=T}^{G-1} \mu_2{}^2 H(g) - \sum_{g=T}^{G-1} 2g\mu_2 H(g) +$$
$$+ \rho_1\mu_1{}^2 + \rho_2\mu_2{}^2 + \mu^2(\rho_1 + \rho_2) - 2\mu(\mu_1\rho_1 + \mu_2\rho_2) =$$
$$= \sum_{g=0}^{G-1} g^2 H(g) + \mu_1{}^2 \sum_{g=0}^{T-1} H(g) - 2\mu_1 \sum_{g=0}^{T-1} gH(g) + \mu_2{}^2 \sum_{g=T}^{G-1} H(g) - 2\mu_2 \sum_{g=T}^{G-1} gH(g) + \rho_1\mu_1{}^2 + \rho_2\mu_2{}^2 + \mu^2 - 2\mu^2 =$$
$$= \sum_{g=0}^{G-1} g^2 H(g) + \mu_1{}^2\rho_1 - 2\mu_1{}^2\rho_1 + \mu_2^2\rho_2 - 2\mu_2{}^2\rho_2 + \rho_1\mu_1{}^2 + \rho_2\mu_2{}^2 - \mu^2 = \sum_{g=0}^{G-1} g^2 H(g) - \mu^2 = \sigma^2$$