

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ  
ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ  
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ**

## **ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ**

### **Ενότητα 3.**

Χαράλαμπος Π. Στρουθόπουλος  
Καθηγητής



Ιούλιος 2015

## ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 3 – Επεξεργασίες σε εικόνες αποχρώσεων του γκρι

3.1. Βελτίωση εικόνων και αποκατάσταση 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 LZW

3.3.3 Κωδικοποίηση 2Δ-διακριτού μετασχηματισμό συνημιτόνου

3.4. Ανίχνευση ακμών

3.4.1 Ανίχνευση ακμών με την χρήση πρώτων παραγώγων

3.4.2 Ανίχνευση ακμών με την χρήση του τελεστή Laplace

3.5 Μεταβολή μεγέθους – Παρεμβολή τιμών

## ΚΕΦΑΛΑΙΟ 3

### ΕΠΕΞΕΡΓΑΣΙΕΣ ΣΕ ΕΙΚΟΝΕΣ ΑΠΟΧΡΩΣΕΩΝ ΤΟΥ ΓΚΡΙ

#### 3.1.1 Εξομάλωση

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 εξομάλωσης 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  $3 \times 3$  τότε το  $I'(j_c, k_c)$  δίνεται από την σχέση

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

Με άλλα λόγια η εικόνα  $I'$  είναι το αποτέλεσμα του φιλτραρίσματος της αρχικής εικόνας

$$I \text{ με την μάσκα } W = \frac{1}{9} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}.$$

Είναι 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 5x5 ή μεγαλύτερες εξομαλύνουν ακόμα

περισσότερο την εικόνα.

Η μάσκα  $W_G$ ,  $N \times N$  με συντελεστές που δίνονται από τις σχέσεις

$$w_G(j, k) = \frac{G(j, k)}{\sum_{j=-\frac{(N-1)}{2}}^{\frac{(N-1)}{2}} \sum_{k=-\frac{(N-1)}{2}}^{\frac{(N-1)}{2}} G(j, k)}$$

με

$$G(j, k) = \frac{1}{2\pi\sigma^2} e^{-\frac{j^2+k^2}{2\sigma^2}}$$

λέγεται μάσκα Gauss ή φίλτρο Gauss (Gaussian filter) και χρησιμοποιείται ευρέως για την εξομάλυνση της εικόνας. Οι τιμές της για  $N=3$  και  $\sigma=1$ , είναι

$$\begin{matrix} 0.0751 & 0.1238 & 0.0751 \\ 0.1238 & 0.2042 & 0.1238 \\ 0.0751 & 0.1238 & 0.0751 \end{matrix}$$

Οι μάσκες που περιγράψαμε αποτελούν γραμμικά συστήματα που λειτουργούν ως κατωδιαβατά φίλτρα. Ως εκ τούτου επηρεάζουν τις τιμές των περιοχών της εικόνας που εμφανίζονται απότομες μεταβολές τους (ακμές).

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.

Με την εξομάλυνση γενικά αμβλύνονται οι ακμές της εικόνας. 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  $I_{kxj}$ ,  $g=0,1,\dots,G-1$  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 απόχρωσης της αρχικής εικόνας αντικαθίσταται από μια τιμή  $g' = [(G-1)P(g)]$ . Υπενθυμίζεται ότι για την  $P(g)$  ισχύουν οι σχέσεις

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

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

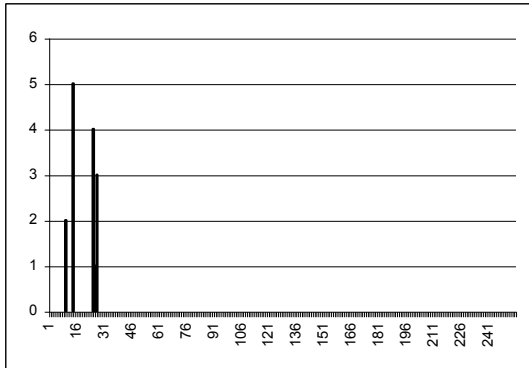
10	10	14	14	14
14	14	25	25	25
25	26	27	27	27

Οι αποχρώσεις  $g'$  προκύπτουν σύμφωνα με τους ακόλουθους υπολογισμούς:

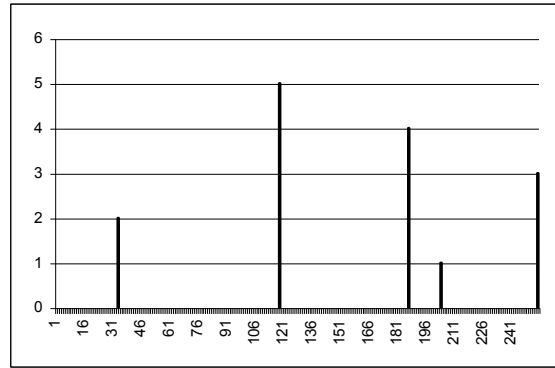
$g$	$H(g)$	$h(g)$	$P(g)$	$g'$
<b>10</b>	2	2/15	2/15	$255 \cdot 2/15 = \mathbf{34}$
<b>14</b>	5	5/15	7/15	$255 \cdot 7/15 = \mathbf{119}$
<b>25</b>	4	4/15	11/15	$255 \cdot 11/15 = \mathbf{187}$

<b>26</b>	1	1/15	12/15	255*12/15 = <b>204</b>
<b>27</b>	3	3/15	15/15	255*15/15 = <b>255</b>

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



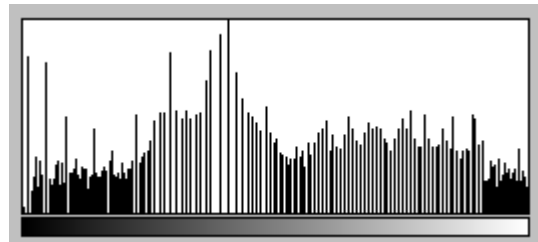
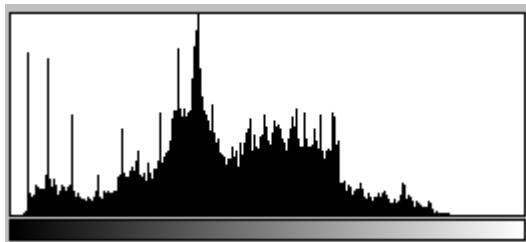
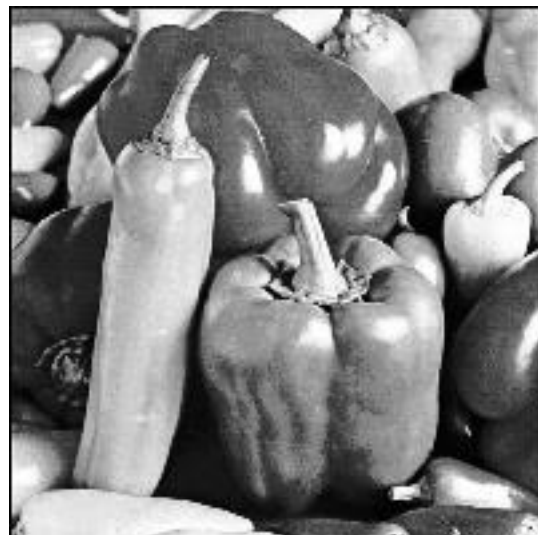
Το ιστόγραμμα  $h(g)$



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  $C_1$ ,  $C_2$  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

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

Ακολουθώντας από την αρχική εικόνα  $I$  προκύπτει η δυαδική εικόνα  $I'$  κάθε εικονοστοιχείο της οποίας δίνεται από τη σχέση

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

#### 3.2.1 Κατωφλίωση με βάση τη διασπορά

Μια μέθοδος για την εύρεση της τιμής κατωφλιού  $T$  είναι η μέθοδος του Otsu. Σύμφωνα με αυτήν η τιμή  $T$  πρέπει να είναι τέτοια ώστε η συνολική εσωτερική διασπορά  $\sigma_w^2$  εντός των κλάσεων να είναι ελάχιστη και η συνολική διασπορά  $\sigma_b^2$  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

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

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

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

Είναι προφανές ότι ισχύει

$$p_1 + p_2 = 1. \quad (3.2.1.3)$$

Η μέση τιμή  $\mu_1$  της πρώτης κλάσης είναι

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

Η μέση τιμή  $\mu_2$  της δεύτερης κλάσης είναι και ισούται με

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

Η συνολική μέση τιμή  $\mu$  δίνεται από τη σχέση

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

Από τις τρεις προηγούμενες σχέσεις εύκολα αποδεικνύεται η σχέση

$$\mu = p_1\mu_1 + p_2\mu_2 \quad (3.2.1.7)$$

Η διασπορά της πρώτης κλάσης δίνεται από τη σχέση

$$\sigma_1^2 = \frac{1}{p_1} \sum_{g=0}^{T-1} (g - \mu_1)^2 h(g) \quad (3.2.1.8)$$

και η διασπορά της δεύτερης κλάσης είναι

$$\sigma_2^2 = \frac{1}{p_2} \sum_{g=T}^{G-1} (g - \mu_2)^2 H(g) \quad (3.2.1.9)$$

Η συνολική διασπορά  $\sigma^2$  είναι

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

Η συνολική εσωτερική διασπορά των κλάσεων ορίζεται από τη σχέση

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

Η συνολική διασπορά μεταξύ των κλάσεων ορίζεται από τη σχέση

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

Αναζητούμε τώρα την κατάλληλη τιμή του  $T$  ώστε η  $\sigma_w^2$  να είναι ελάχιστη και  $\sigma_b^2$  μέγιστη.

Προς τούτο θεωρούμε το πηλίκο

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

και ζητάμε την τιμή του  $T$  για την οποία το  $\lambda$  μεγιστοποιείται.

Για να μειωθούν οι υπολογισμοί θεωρούμε τις σχέσεις

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

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

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

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

Η μέγιστη τιμή του  $\sigma_b^2$  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 Karur, 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  $g = 0, \dots, T-1$  και αντίστοιχες πιθανότητες

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

όπου

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

Η μέση πληροφορία (εντροπία) που προκύπτει της  $c_1$  είναι

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

Όμοια η μέση πληροφορία που προκύπτει από την  $c_2$  είναι

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

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  $T_a$  είναι εκείνη για την οποία  $E_1(T) + E_2(T)$  γίνεται μέγιστο.

### 3.2.3. Πολυκατωφλίωση με χρήση νευρωνικού δικτύου

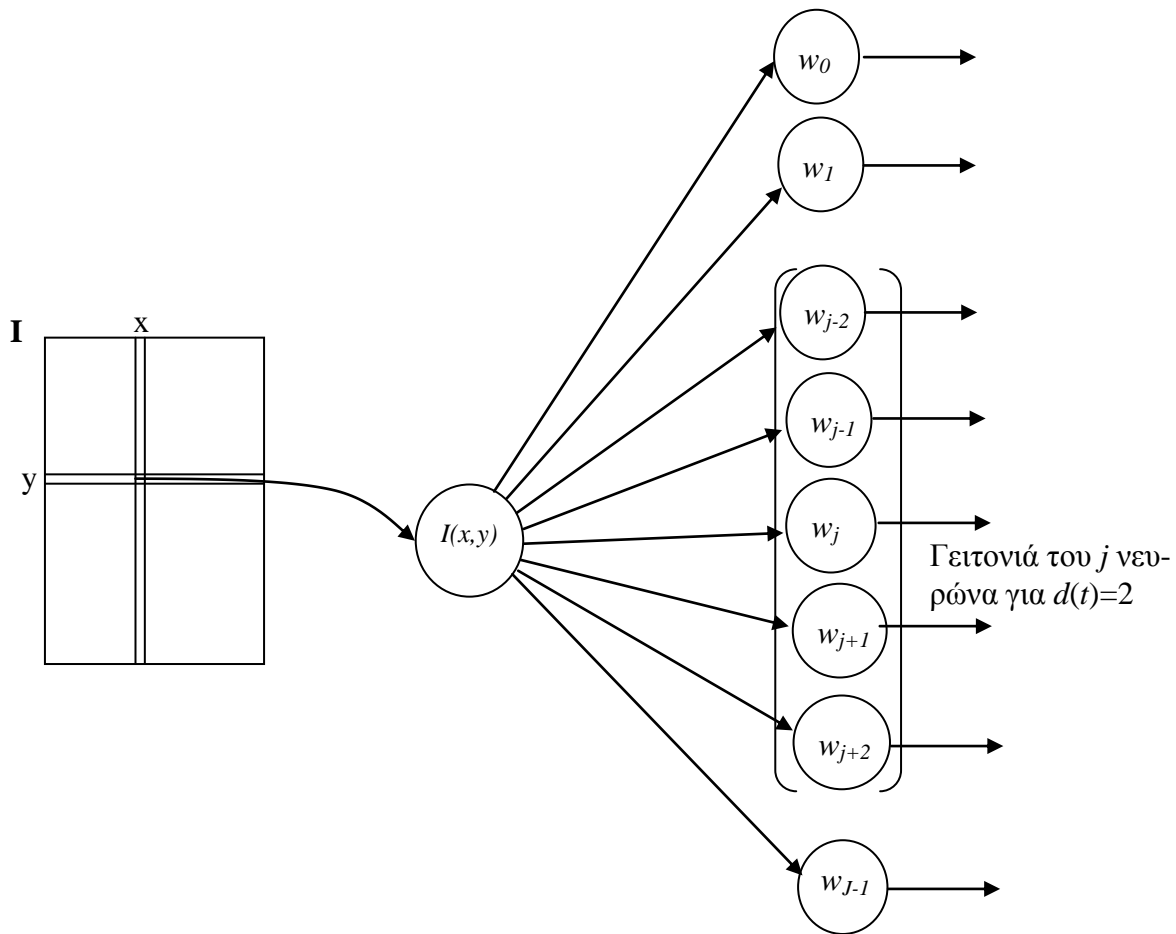
Εάν κάθε αντικείμενο της εικόνας αποτελείται από μία κύρια και κάποιες παραπλήσιες αποχρώσεις του γκρι, τότε δημιουργείται η συσσώρευση των τιμών του ιστογράμματος σε συγκεκριμένες περιοχές. Η επιλογή πολλαπλών κατωφλίων μπορεί να θεωρηθεί ως ένα πρόβλημα εύρεσης τιμών  $T(k)$ ,  $k = 1, 2, \dots, J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

$$I(x,y) = \begin{cases} G(0) \text{ αν } f(x,y) \leq T(1) \\ G(1) \text{ αν } T(1) < f(x,y) \leq T(2) \\ \vdots \\ G(J-1) \text{ αν } T(J-1) < f(x,y) \end{cases}
 \tag{3.2.3.1}$$

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

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

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

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

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

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

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

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  $\alpha(t)$ ,  $d(t)$  σύμφωνα με τις σχέσεις:

$$\alpha(t) = \alpha(0) \cdot \left(1 - \frac{t}{T}\right) \quad (3.2.3.6)$$

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

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

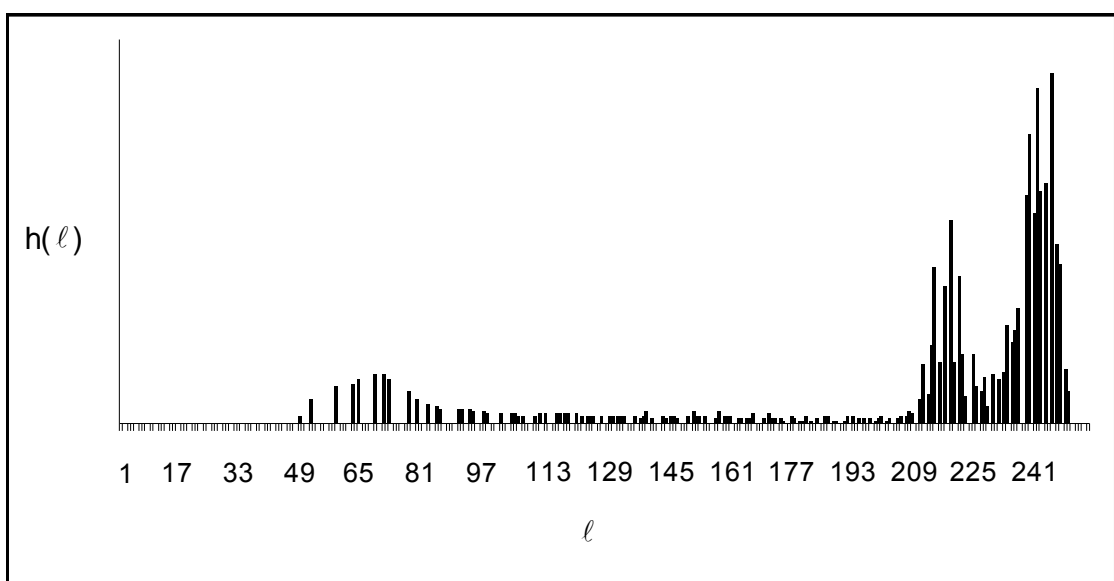
Μετά την εκπαίδευση κάθε άνυσμα εισόδου του ΑΠΑΧ αποδίδεται στον νικητή νευρώνα. Κάθε νευρώνας του επιπέδου εξόδου αντιπροσωπεύει μία ομάδα προτύπων (cluster). Πρότυπα με μεγάλη ομοιότητα αντιπροσωπεύονται από τον ίδιο νευρώνα. Ο χάρτης χαρακτηριστικών του Kohonen οργανώνει τους νευρώνες του επιπέδου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

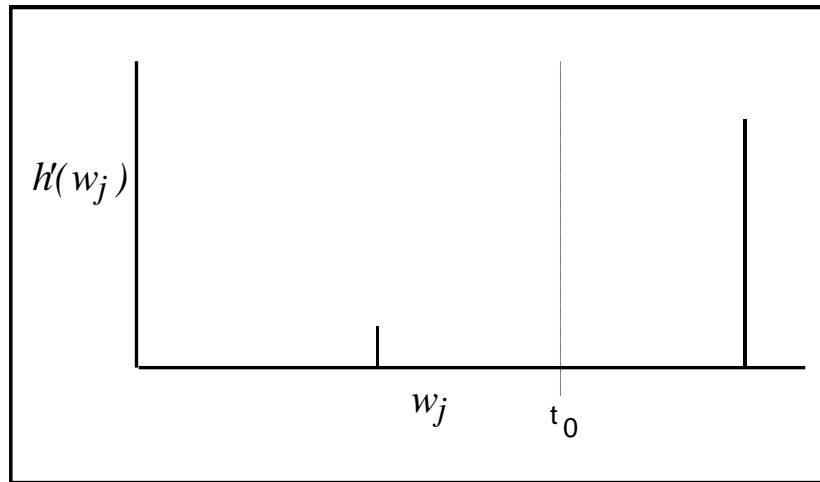
$T_0$	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$
90	132	176	209	224	235	243

Πίνακας 3.2.3.2. Οι τιμές των κατωφλίων.

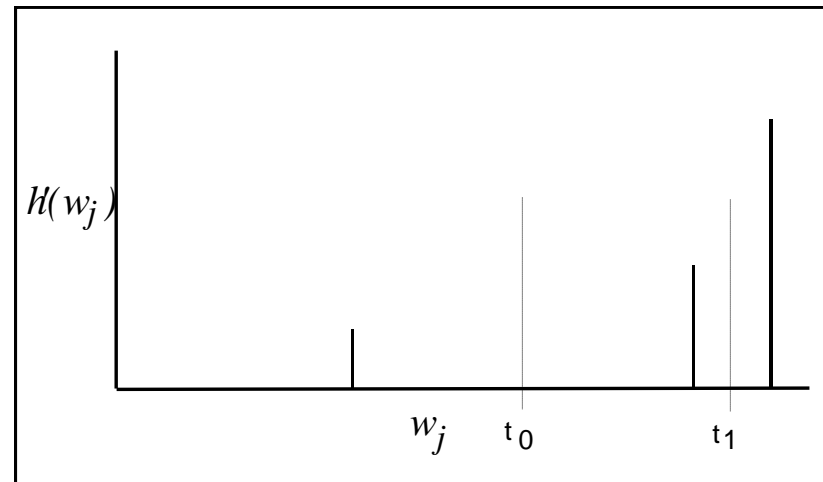
Στον Πιν. 3.2.3.3 φαίνονται οι τιμές των κατωφλίων για διαφορετικό πλήθος ( $J$ ) νευρώνων του νευρωνικού δικτύου

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

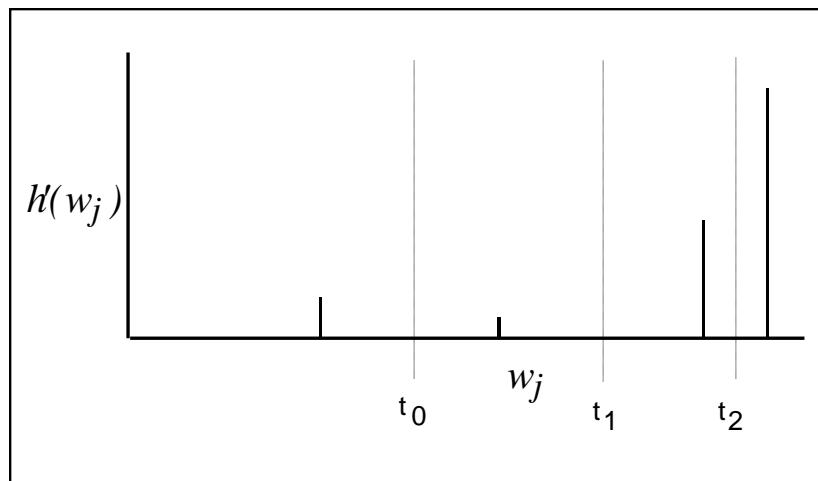
Πίνακας 3.2.3.3. Οι τιμές των κατωφλίων για  $J=2, \dots, 6$ .



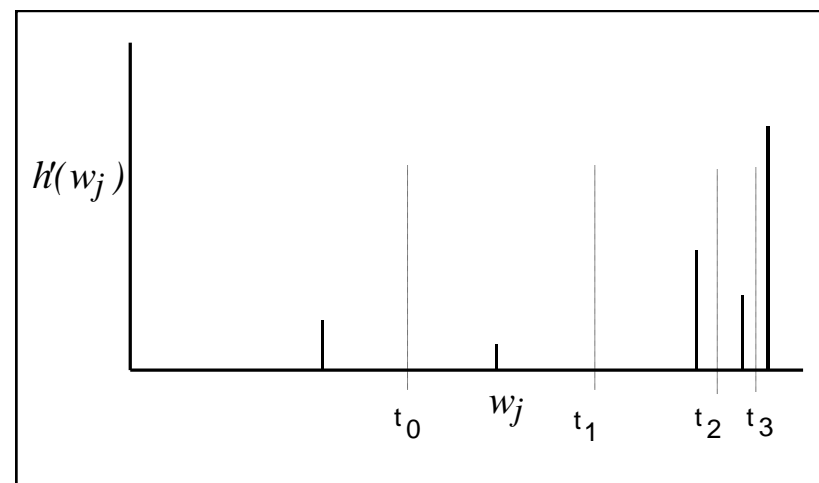
$J=2$



$J=3$



$J=4$

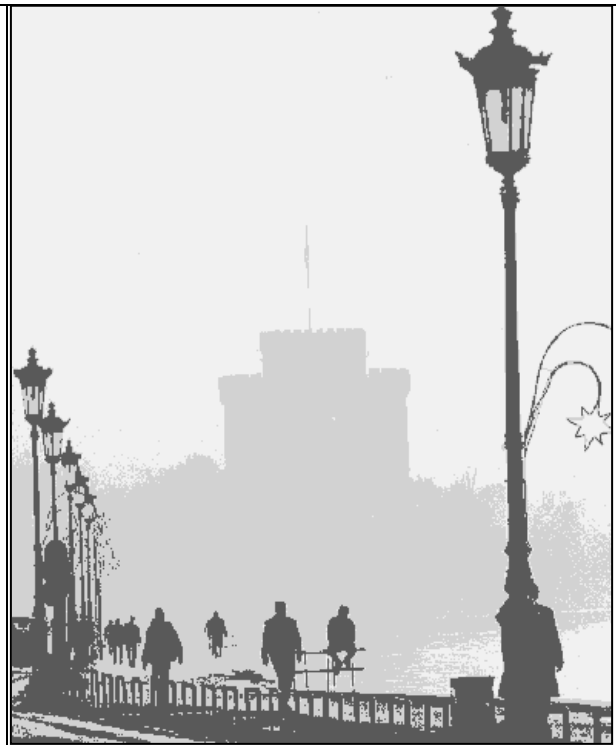


$J=5$

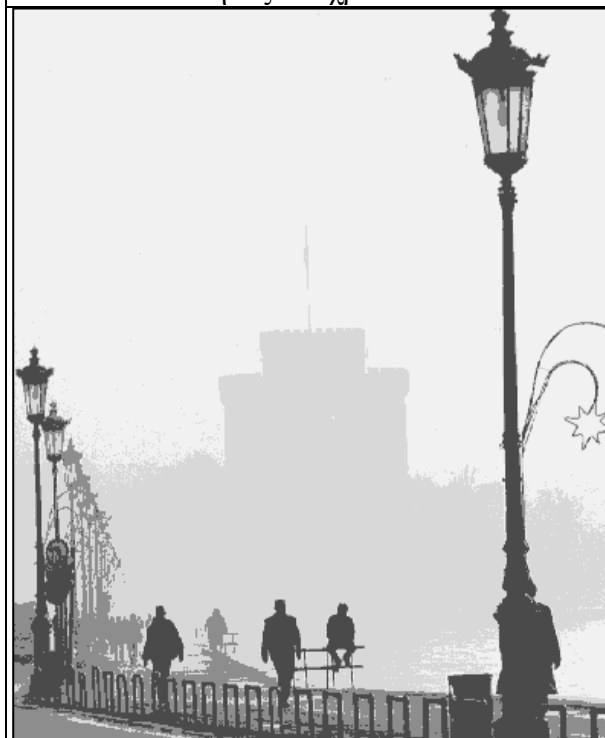
Σχήμα 3.2.3.3. Τα ιστογράμματα και οι αντίστοιχες τιμές κατοφλίων για διαφορετικές τιμές του  $J$ .



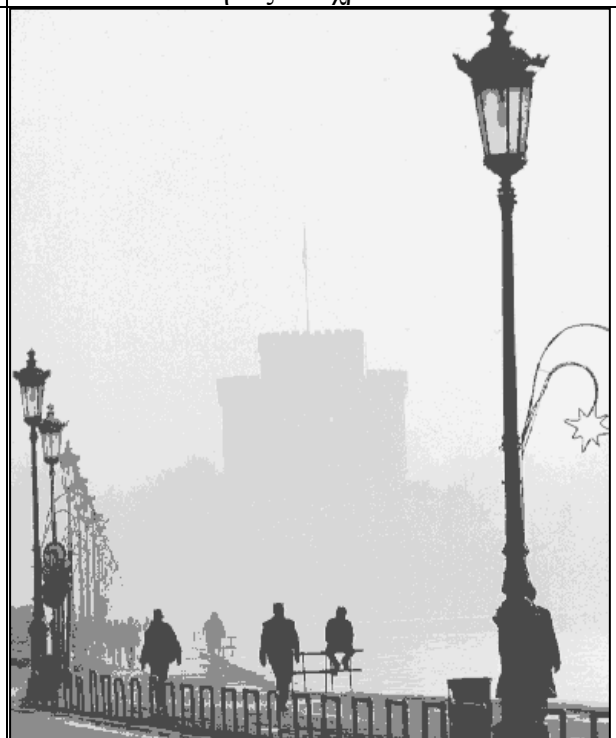
Πλήθος αποχρώσεων 2



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  $\{2, \dots, J_{max}\}$  ελαχιστοποιεί το σφάλμα προσαρμογής μεταξύ των συναρτήσεων  $H(\ell)$  και  $H_J(\ell)$ . Η συνάρτηση  $H(\ell)$  υπολογίζεται από τη σχέση

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

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  $\mu_j$  της κλάσης και η διασπορά της  $\sigma_j^2$  δίνονται αντίστοιχα από τις σχέσεις

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

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

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

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

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

όπου

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

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

$$|e_J| = \sum_{\ell=0}^{L-1} [h(\ell) - H_J(\ell)]^2 \quad (3.2.3.14)$$

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

$$|e_{J_o}| = \text{minimum}\{|e_J|\} \quad (3.2.3.15)$$

Ο Πίν. 3.2.3.3 δείχνει τις τιμές του  $|e_J|$  για  $J=2, \dots, 6$ . 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 σε 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 γκρι 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

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

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  $\bar{l}$  σύμφωνα με την θεωρία της πληροφορίας θα ικανοποιεί τη σχέση

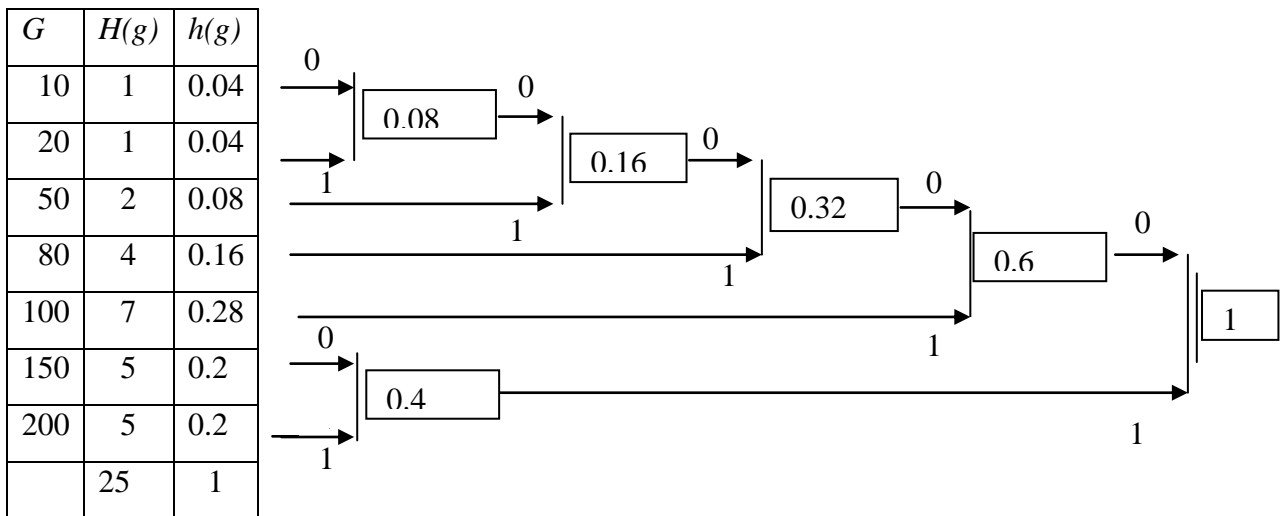
$$H(G) \leq \bar{l} \leq H(G)+1 \quad (3.3.1.2)$$

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

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

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  $4 \times 5$ .

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



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

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

### 3.3.2 Κωδικοποίηση LZW

Με την τεχνική LZW (Lebel – Ziv – Welch) επιδιώκεται η κωδικοποίηση ακολουθιών από διαδοχικές τιμές εικονοστοιχείων. Προς τούτο για συγκεκριμένη εικόνα δημιουργείται ένα ευρετήριο των ακολουθιών και των κωδικών τους. Για παράδειγμα αν η ακολουθία τιμών 120-30-30 ( $3 \times 8 = 24$  bits) κωδικοποιηθεί με τον κωδικό 410 (9 bits) και επαναλαμβάνεται 30 φορές στην εικόνα υπάρχει ένα όφελος  $30 \times 15 = 450$  bits. Συνήθως στο ευρετήριο έχουν τοποθετηθεί ήδη οι 256 τιμές των αποχρώσεων του γκρι στις θέσεις-κωδικούς από 0-255. Η εικόνα προσπελαίνεται από πάνω προς τα κάτω και από αριστερά προς τα δεξιά. Εάν μία ακολουθία τιμών διαδοχικών εικονοστοιχείων υπάρχει στο ευρετήριο αυτή αυξάνεται με την τιμή του επόμενου εικονοστοιχείου και η νέα ακολουθία αναζητάται στο ευρετήριο. Αν μία ακολουθία με  $N$  τιμές δεν βρεθεί στο ευρετήριο, τότε αυτή καταχωρείται α' αυτό και στην έξοδο του κωδικοποιητή τίθεται ο κωδικός της ακολουθίας των πρώτων  $N-1$  τιμών. Στο παρακάτω παράδειγμα φαίνεται η παραπάνω διαδικασία για μια μικρή εικόνα  $4 \times 4$ .



10	10	30	30
10	10	30	30
10	10	30	30
10	10	30	30

Ακολουθία τιμών				Τιμή επόμενου εικ/χειού	Έξοδος κωδικο- ποιητή	Ευρετήριο				
						Κωδικός	Ακολουθία			
10				10	10	256:	10	10		
10				30	10	257:	10	30		
30				30	30	258:	30	30		
30				10	30	259:	30	10		
10				10						
10	10			30	256	260:	10	10	30	
30				30						
30	30			10	258	261:	30	30	10	
10				10						
10	10			30						
10	10	30		30	260	262:	10	10	30	30
30				10						
30	10			10	259	263:	30	10	10	
10				30						
10	30			30	257	264:	10	30	30	
30					30					

Είναι σημαντικό να αναφέρουμε ότι κατά την αποκωδικοποίηση των τιμών του συμπιεσμένου αρχείου δεν είναι απαραίτητη η γνώση του ευρετηρίου και ως εκ τούτου δεν είναι αναγκαία η επισύναψή του. Το ευρετήριο δημιουργείται στον αποκωδικοποιητή από τις τιμές που αποκωδικοποιούνται με τον ίδιο τρόπο που δημιουργήθηκε στην φάση της κωδικοποίησης. Ακολούθως παρουσιάζεται η διαδικασία αποκωδικοποίησης των τιμών που δημιουργήθηκαν κατά την φάση κωδικοποίησης του προηγούμενου παραδείγματος.

Έισοδος αποκωδικο- ποιητή	Έξοδος κωδικο-ποιητή			Ευρετήριο			
				Κωδικός		Ακολουθία	
10	10						
10	10						
30	30						
30	30						
256				256:	10	10	
				257:	10	30	
				258:	30	30	
	10	10					
258	30	30					
260				259:	30	10	
				260:	10	10	30
	10	10	30				
259	30	10					
257	10	30					
30	30						

### 3.3.3 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

$$g_0(n) = \frac{1}{\sqrt{N}}, \quad g_k(n) = \sqrt{\frac{2}{N}} \cos\left(\frac{\pi}{2N} k(2n+1)\right) \quad (3.3.3.1)$$

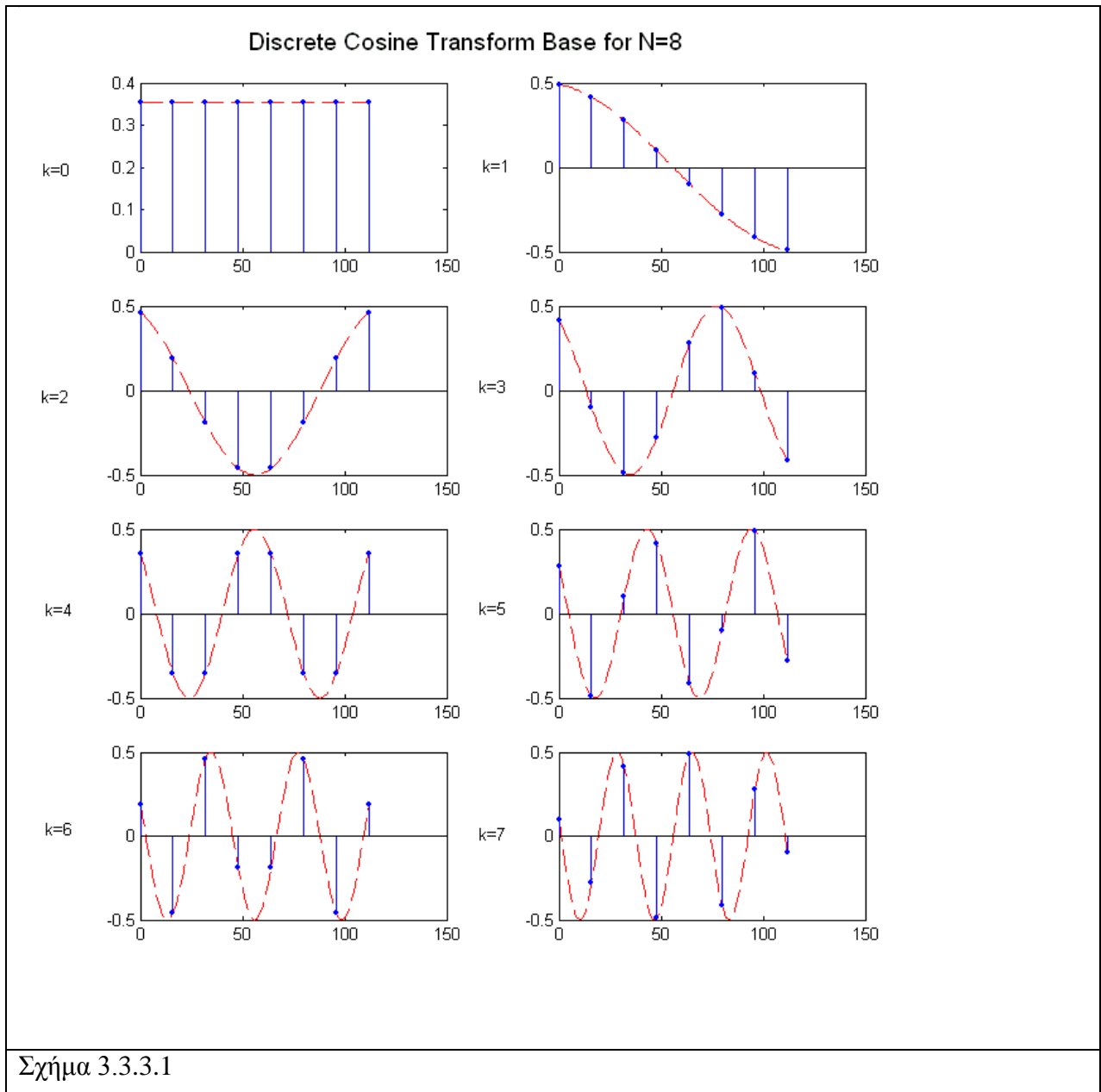
για  $k=1..N-1$  και  $n=0,..,N-1$  είναι ένας πυρήνας (kernel) δημιουργίας ορθοκανονικών διανυσμάτων για κάθε τιμή του  $N$ . Οι συντελεστές  $c_k$  για ένα διακριτό σήμα  $x(n)$  πεπερασμένου μήκους  $N$  (δηλαδή οι προβολές του  $\mathbf{x}$  στα  $\mathbf{g}_k$ ) προκύπτουν από τη σχέση:

$$c_0 = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n], \quad c_k = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} \cos\left(\frac{\pi}{2N} k \cdot (2n+1)\right) \cdot x[n] \quad (3.3.3.1)$$

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

$$x(n) = \frac{c_0}{\sqrt{N}} + \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} c_k \cdot \cos\left(\frac{\pi}{2N} k(2n+1)\right) \quad (3.3.3.2)$$

Η σχέση αυτή αποτελεί τον αντίστροφο διακριτό μετασχηματισμό του συνημιτόνου. Στο Σχ. 3.3.3.1 δείχνονται τα ανύσματα βάσης του DCT για  $N=8$ .



Παράδειγμα 3.3.3.1

Για N=3 τα ανύσματα

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

προήλθαν από τις σχέσεις

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

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 3.3.3.2

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

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

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

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

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

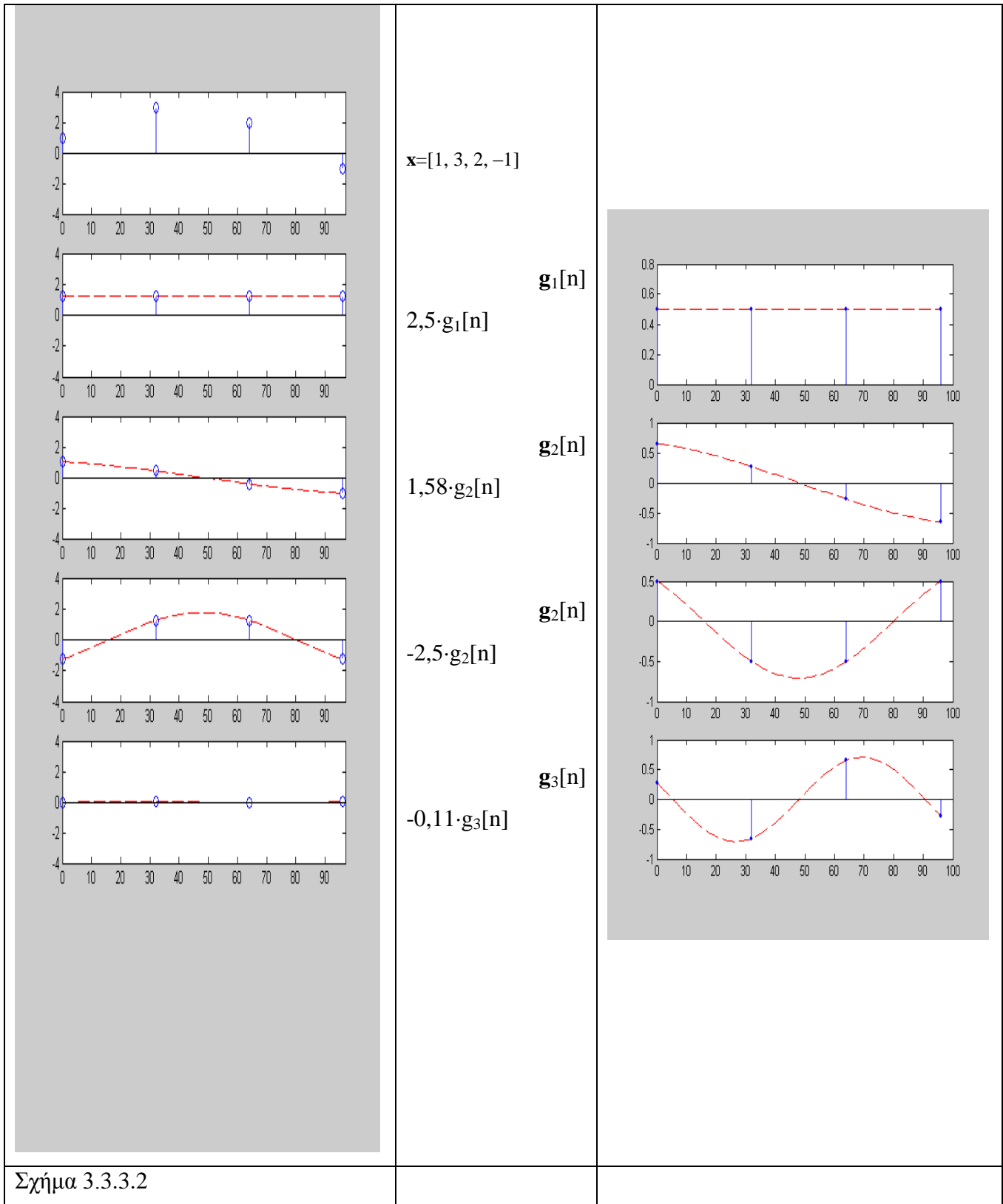
Οι συντελεστές  $c_k$  είναι τα στοιχεία του πίνακα

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

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

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

Στο Σχ. 3.3.3.2 δείχνονται τα ανύσματα βάσης του DCT για  $N=4$ , οι συντελεστές  $c_k$  του DCT του σήματος  $\mathbf{x}=[1, 3, 2, -1]$  και η ανάλυση του στις τέσσερις συνιστώσες  $c_k \mathbf{g}_k$ . Κάθε τιμή του  $x[n]$  είναι  $x[n]=2,5g_0[n]+ 1,58g_1[n]-2,5g_2[n]-0,11g_3[n]$ .



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

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

ήτοι (3.3.3.3)

$$g_{00}(n_1, n_2) = \frac{1}{\sqrt{N_1 N_2}},$$

$$g_{k_1 0}(n_1, n_2) = \frac{1}{\sqrt{N_2}} \sqrt{\frac{2}{N_1}} \cdot \cos\left(\frac{\pi k_1 (2n_1 + 1)}{2N_1}\right),$$

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

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

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:

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

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

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

(3.3.3.4)

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

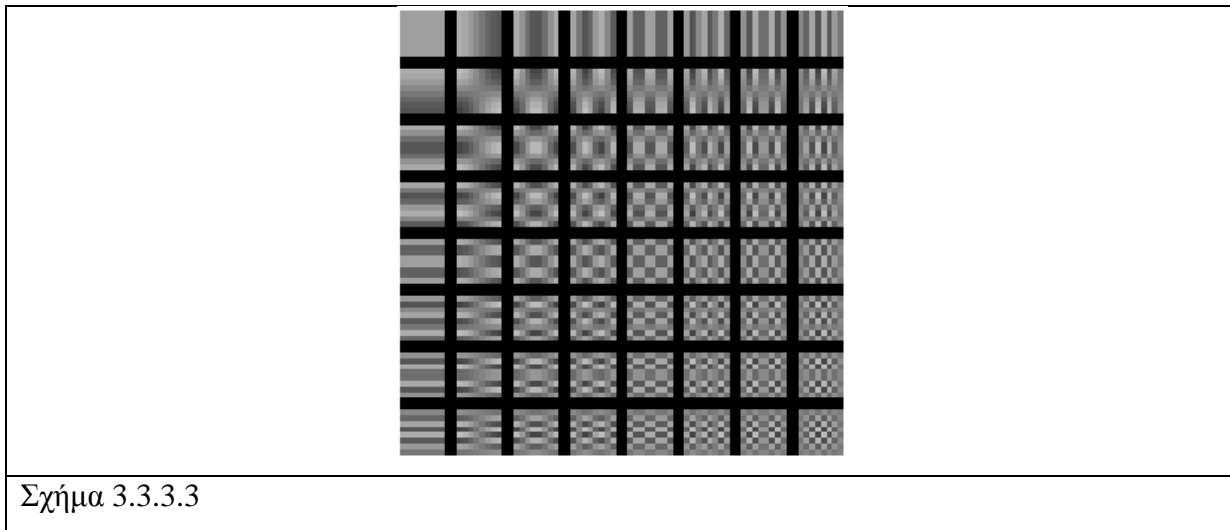
για  $k_1, k_2 \neq 0$ .

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:

$$x(n_1, n_2) = \sum_{n_2=0}^{N_2-1} \sum_{n_1=0}^{N_1-1} c_{k_1 k_2} \cdot g_{k_1 k_2}(n_1, n_2) \quad (3.3.3.5)$$

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

Είναι προφανές από τα παραπάνω αν  $M \times N$  είναι τα εικονοστοιχεία της εικόνας τότε  $M \times N$  θα είναι και όλοι οι συντελεστές του 2D-DCT αυτής. Για παράδειγμα για μία εικόνα  $8 \times 8$ , ο μετασχηματισμός οδηγεί στον υπολογισμό  $8 \times 8$  συντελεστών (προβολών) για κάθε μία από τις  $8 \times 8$  διακριτές διδιάστατες ακολουθίες βάσης. Κάθε ακολουθία της βάσης έχει τιμές που δίνονται από τις (3.3.3.3) που αν παρασταθούν με τόνους του γκρι προκύπτουν οι εικόνες του Σχ.3.3.3.3.

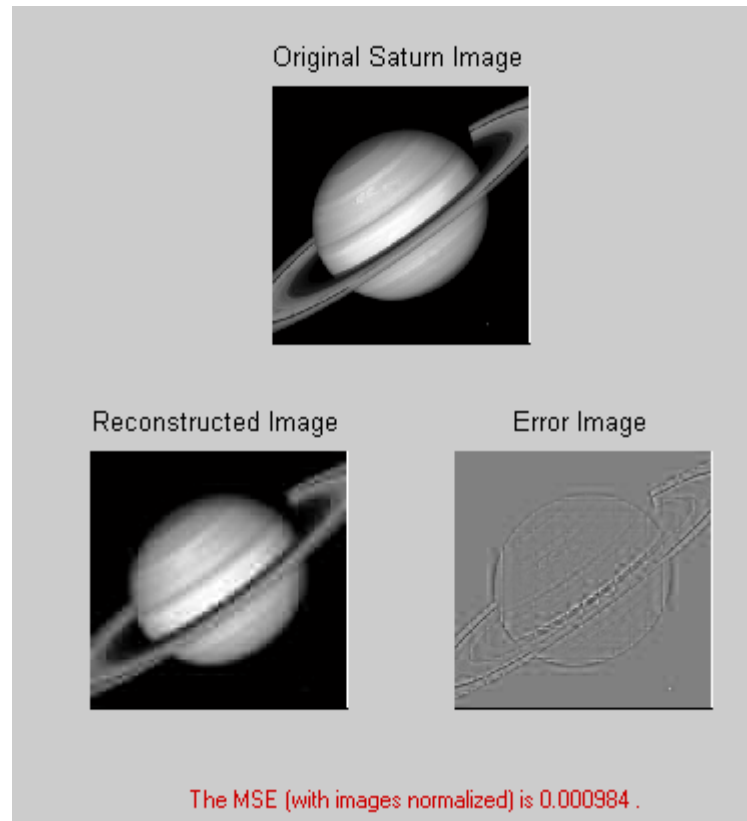


Μετασχηματισμοί των οποίων η βάση ικανοποιεί την (3.3.3.1) λέγονται διαχωρίσιμοι και οι υπολογισμοί τους μπορούν να αναχθούν σε υπολογισμούς των μονοδιάστατων εκφράσεών τους. Η εφαρμογή του 2D-DCT στην συμπίεση εικόνων αποχρώσεων του γκρι σύμφωνα με την τυποποίηση JPEG (Join Photographic Experts Group) ακολουθεί τα ακόλουθα γενικά βήματα:

- Η εικόνα χωρίζεται σε μη επικαλυπτόμενα τμήματα  $8 \times 8$  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.
- Ακολουθώς οι τιμές κβαντίζονται και τελικά
- Κωδικοποιούνται κατά Huffman.

Στο Σχ. 3.3.3.4 η αρχική εικόνα  $256 \times 256$  pels χωρίζεται σε τμήματα των  $8 \times 8$  pels και σε κάθε ένα από αυτά εφαρμόζεται ο 2D-DCT, ακολουθώς με χρήση μόνο των 8 συντελεστών αναδημιουργείται η εικόνα με σφάλμα όπως φαίνεται στο σχήμα.

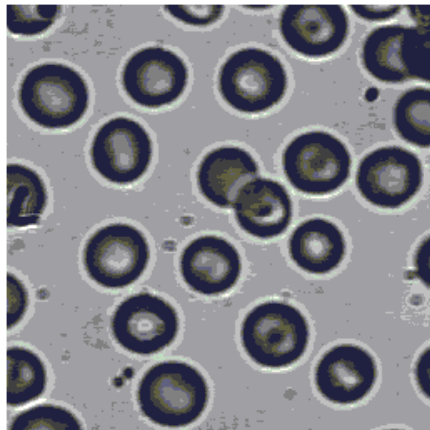




Σχήμα 3.3.3.4

### 3.4 Ανίχνευση ακμών.

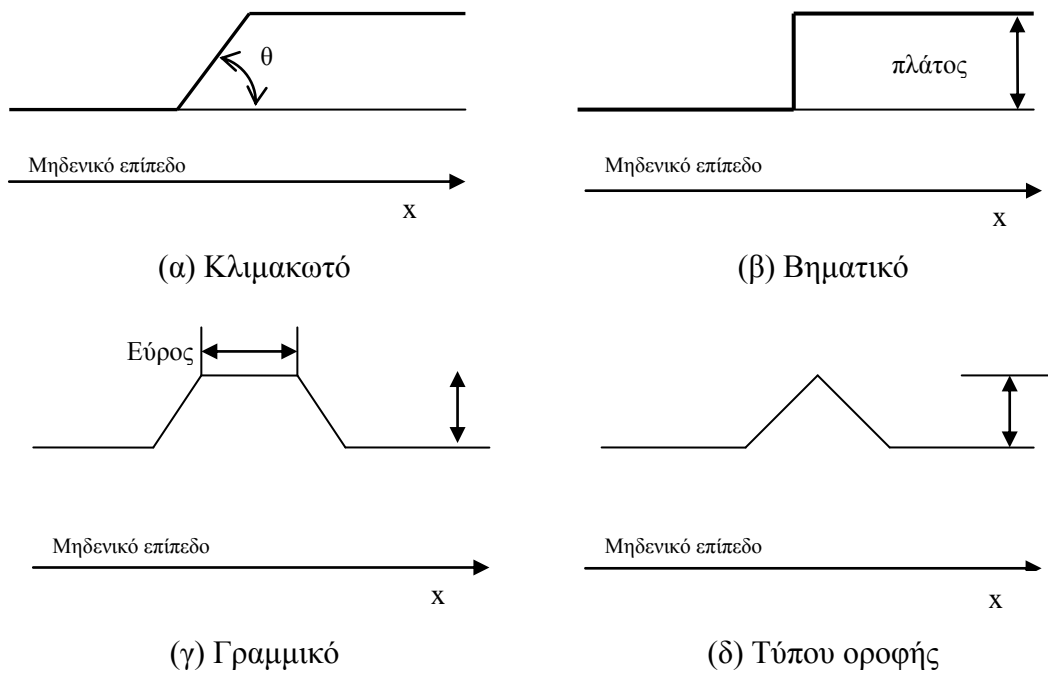
Σε μια εικόνα αποχρώσεων του γκρι υπάρχουν περιοχές εικονοστοιχείων με απότομη αύξηση της φωτεινότητας. Οι περιοχές αυτές βρίσκονται στα όρια των τμημάτων της εικόνας που έχουν σημαντικά διαφορετικές αποχρώσεις. Η ανίχνευση των ορίων αυτών λέγεται *προσδιορισμός των ακμών της εικόνας (edge detection)*. Με άλλα λόγια ακμή είναι μια καμπύλη επί της εικόνας της οποίας οι δυο πλευρές παρουσιάζουν σημαντικές διαφορές της φωτεινότητας ή ενδεχομένως και άλλων χαρακτηριστικών τους όπως η πυκνότητα και η υφή. Η ανίχνευση ακμών είναι μια από τις διαδικασίες ανάλυσης εικόνας που χρησιμοποιούνται ευρύτατα και οι αλγόριθμοι που υπάρχουν στην βιβλιογραφία για την ανίχνευση και βελτίωση τους είναι ίσως οι περισσότεροι από αυτούς άλλων διαδικασιών ανάλυσης εικόνας. Μια ακμή είναι το όριο ενός εικονιζόμενου αντικειμένου και του φόντου ή το όριο μεταξύ επικαλυπτόμενων αντικειμένων (Σχ. 3.4.1).



Σχήμα 3.4.1

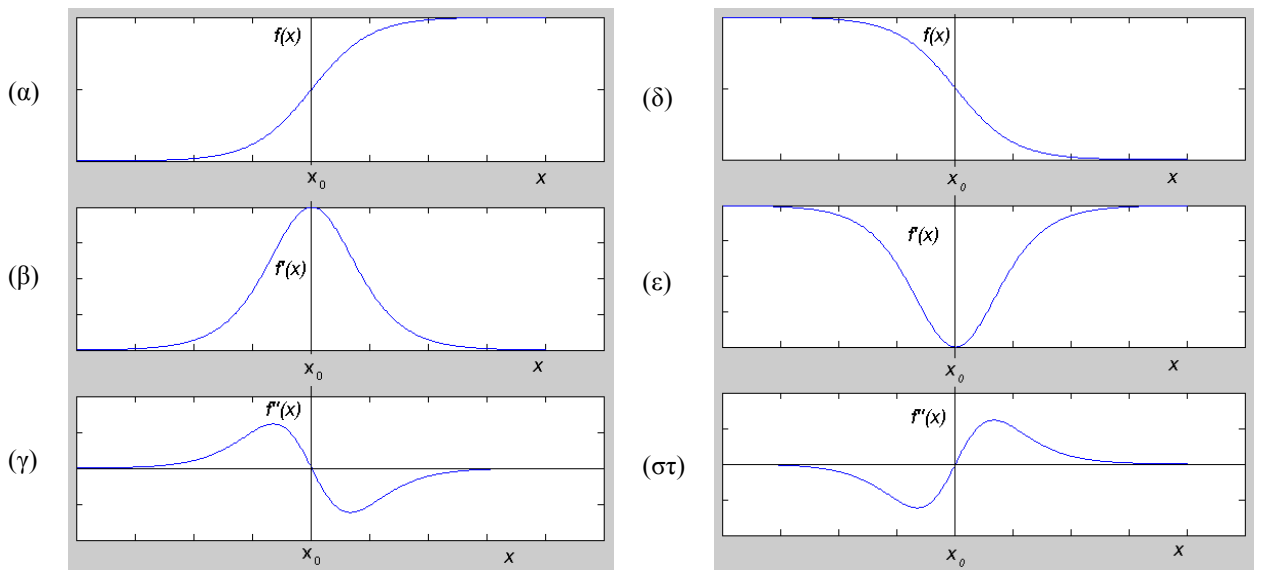
Αυτό σημαίνει ότι αν βρεθούν με ακρίβεια οι ακμές σε μια εικόνα τότε θα εντοπισθούν τα αντικείμενα της εικόνας και είναι δυνατόν να προσδιορισθούν χαρακτηριστικά όπως το εμβαδόν, η περίμετρος ή το σχήμα τους. Η ανίχνευση ακμών είναι εξαιρετικά χρήσιμη εργασία στην ανάλυση των εικόνων διότι μέσω αυτής προσδιορίζονται τα περιγράμματα των αντικειμένων της εικόνας και αποτελεί ένα αρχικό στάδιο τεχνικών τμηματοποίησης της εικόνας. Για τους λόγους αυτούς η εύρεση ακμών είναι βασικό εργαλείο για τον προσδιορισμό και την ταξινόμηση των αντικειμένων που είναι βασικοί στόχοι της μηχανικής όρασης.

Γενικό μαθηματικό μοντέλο ορισμού και ανίχνευσης των ακμών δεν υπάρχει αλλά γίνεται ανάλογα με το πεδίο εφαρμογής, επιλογή από ένα πλήθος μοντέλων και τεχνικών για την ανίχνευση τους. Στο Σχ.3.4.2 φαίνεται η μεταβολή της φωτεινότητας κατά δεδομένη διεύθυνση, ακμών διαφορετικού τύπου ως προς το ύψος, την κλίση και το πλάτος τους.



Σχήμα 3.4.2: Μονοδιάστατα συνεχή μοντέλα ακμών.

Από μαθηματική άποψη αν  $f(x,y)$  η συνάρτηση της φωτεινότητας της εικόνας ( $x, y$  συνεχείς μεταβλητές - συντεταγμένες σημείου του επιπέδου της εικόνας), η μεταβολή της ως προς μία κατεύθυνση εκφράζεται από την παράγωγο της κατά κατεύθυνση. Ως εκ τούτου η εύρεση των ακμών μπορεί να βασισθεί στην χρήση των παραγώγων της εικόνας. Στο Σχ. 3.4.3 φαίνονται οι γραφικές παραστάσεις της φωτεινότητας, της πρώτης και της δεύτερης παραγώγου της ως προς την  $x$  κατεύθυνση.



Σχήμα 3.4.3

Αν το σημείο  $x_0$  είναι σημείο ακμής, η φωτεινότητα παρουσιάζει μέγιστη μεταβολή στο  $x_0$  και συνεπώς η πρώτη παράγωγος έχει ακρότατη τιμή ενώ η δεύτερη παράγωγος μηδενίζεται. Ακολούθως θα δούμε την χρήση των παραγώγων ως βάση για την ανίχνευση των ακμών.

### 3.4.1 Ανίχνευση ακμών με την χρήση των πρώτων παραγώγων

Αν  $f(x,y)$  δισδιάστατη συνάρτηση των συνεχών ανεξάρτητων μεταβλητών  $x, y$ ,  $f(x,y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ , οι μερικοί της παράγωγοι  $\partial f/\partial x, \partial f/\partial y$  ορίζουν τον πίνακα

$$Df = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

Αν θεωρήσουμε ορθογώνιους άξονες  $x', y'$  με αντίστοιχα μοναδιαία διανύσματα  $\vec{u}_x$  και  $\vec{u}_y$ , το διάνυσμα κλίσης (*gradient*) της  $f$  συμβολίζεται  $\nabla f$  και δίνεται από τη σχέση

$$\nabla f = \frac{\partial f}{\partial x} \vec{u}_x + \frac{\partial f}{\partial y} \vec{u}_y$$

Αν κινηθούμε κατά την διεύθυνση που ορίζει η γωνία  $\phi$  που δίνεται από σχέση

$$\phi = \text{τοξεφ} \left( \frac{\partial f / \partial x}{\partial f / \partial x} \right)$$

η μεταβολή της  $f(x,y)$  είναι μέγιστη και το μέτρο της μεταβολής δίνεται από τη σχέση

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

Μια ψηφιακή εικόνα δεν είναι συνάρτηση συνεχών μεταβλητών και ως εκ τούτου αντί των μερικών παραγώγων υπολογίζουμε τις μερικές διαφορές της απόχρωσης ενός εικονοστοιχείου και των γειτονικών του κατά την οριζόντια και κάθετη κατεύθυνση της εικόνας.

Αν  $\mathbf{I}$  ο πίνακας μιας ψηφιακής εικόνας, οι μερικές μεταβολές της φωτεινότητας στο σημείο  $(k, j)$  μπορούν να ορισθούν εναλλακτικά από τα παρακάτω ζεύγη σχέσεων:

$$D_k(k, j) = I(k, j) - I(k-1, j)$$

$$D_j(k, j) = I(k, j) - I(k, j-1)$$

ή

$$D_k(k, j) = I(k+1, j) - I(k-1, j)$$

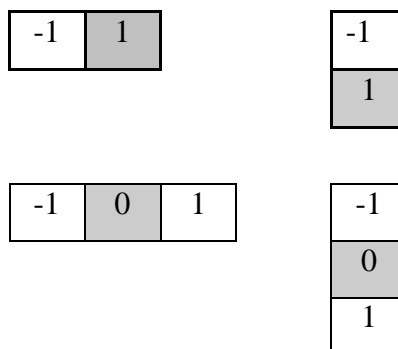
$$D_j(k, j) = I(k, j+1) - I(k, j-1)$$

Αν  $\mathbf{D}(k, j) = [D_k(k, j), D_j(k, j)]$ , η διεύθυνση και το πλάτος της κλίσης μπορούν να ορισθούν σύμφωνα με τα προηγούμενα από τις σχέσεις

$$\phi(k, j) = \arctan\left(\frac{D_j(k, j)}{D_k(k, j)}\right)$$

$$|\mathbf{D}(k, j)| = \sqrt{D_k(k, j)^2 + D_j(k, j)^2}$$

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



Η χρήση των μερικών διαφορών μπορεί να θεωρηθεί η είσοδος των τιμών των γραμμών ή των στηλών της εικόνας σε ένα σύστημα γραμμικό, ανεξάρτητο από μετατόπιση (LTI) που ορίζεται από τη σχέση

$$y(n) = x(n) - x(n-1) \text{ ή } y(n) = x(n+1) - x(n-1)$$

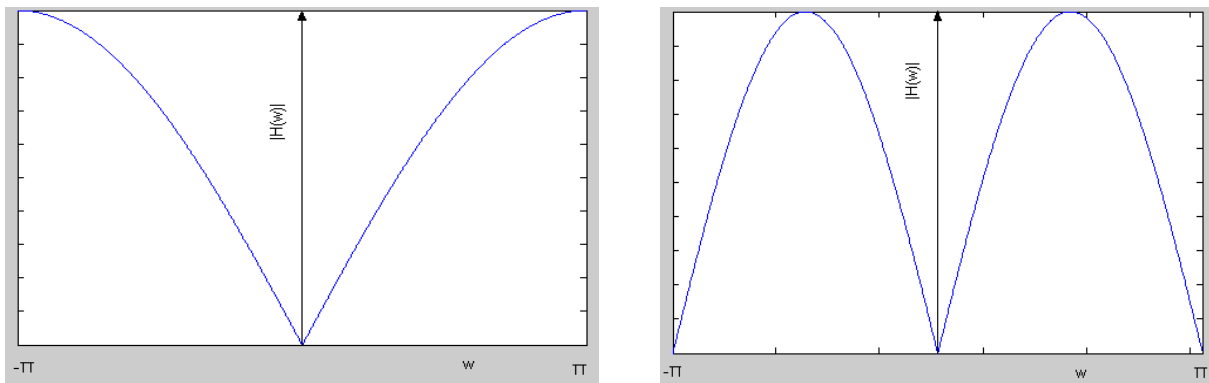
Η κρουστική απόκριση θα είναι τότε

$$h(n) = \delta(n) - \delta(n-1) \text{ ή } h(n) = \delta(n+1) - \delta(n-1)$$

ή με μορφή πίνακα

$$\mathbf{h} = [1, -1] \text{ ή } \mathbf{h} = [1, 0, -1]$$

Η μερική διαφορά  $y(n)$  στο σημείο  $n$  που είναι σημείο μιας γραμμής ή στήλης της εικόνας θα δίνεται ο από την συνέλιξη  $x(n) * h(n)$  δηλαδή την εφαρμογή των αντίστοιχων μασκών. Τα συστήματα αυτά λειτουργούν ως φίλτρα στο πεδίο της συχνότητας. Στο σχήμα 3.4.4 δείχνονται τα διαγράμματα πλάτους των αποκρίσεων συχνότητας των παραπάνω συστημάτων. Πρόκειται για ένα ανωδιαβατό και ένα ζωνοπερατό φίλτρο αντίστοιχα.



Σχήμα 3.4.4

Ο εντοπισμός των σημείων των ακμών μπορεί να βασισθεί στην εύρεση υψηλών τιμών του πίνακα με στοιχεία τις τιμές  $|\mathbf{D}(k, j)|$ . Υψηλές είναι οι τιμές που είναι μεγαλύτερες από μία καθορισμένη τιμή κατωφλίου  $T$ . Συνεπώς στη θέση  $(k, j)$  υπάρχει σημείο ακμής εάν  $|\mathbf{D}(k, j)| > T$ . Μία πρώτη προσέγγιση για την επιλογή μιας καθολικής τιμής κατωφλίου μπορεί να βασισθεί στον μέσο όρο των τιμών  $|\mathbf{D}(k, j)|$ . Η επιλογή της κατάλληλης τιμής κατωφλίου είναι κρίσιμη. Συχνά η μεταβολή του φωτισμού, η μεταβολή της απόχρωσης του αντικειμένου ή του παρασκηνίου, απαιτούν την χρήση διαφορετικών τοπικών τιμών κατωφλίων. Αν εφαρμόσουμε όλες τις παραπάνω μάσκες σε μία εικόνα  $3 \times 3$ , 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

$$S_k=1/4 \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_j=1/4 \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

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. Οι πλέον σημαντικοί είναι των

*Prewitt*

$$P_k=1/3 \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad P_j=1/3 \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

*Roberts*

$$R_k= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad R_j= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

*Frei - Chen*

$$F_k= \frac{1}{2+\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix} \quad F_j= \frac{1}{2+\sqrt{2}} \begin{bmatrix} -1 & -\sqrt{2} & -1 \\ 0 & 0 & 0 \\ 1 & \sqrt{2} & 1 \end{bmatrix}$$

Σε κάθε περίπτωση από την αρχική εικόνα παράγεται ένας πίνακας τιμών του μέτρου ή 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.

ΠΑΡΑΔΕΙΓΜΑ: Δίνεται ο ακόλουθος πίνακας ψηφιακής εικόνας.

$$I = \begin{pmatrix} 10 & 100 & 100 & 20 & 30 \\ 10 & 100 & 110 & 20 & 30 \\ 10 & 100 & 140 & 140 & 30 \\ 10 & 10 & 110 & 20 & 30 \\ 10 & 10 & 100 & 20 & 30 \end{pmatrix}$$

Να ανιχνευθούν οι ακμές του με την κλίση που καθορίζουν οι μάσκες

a)

$$h_1 = \begin{bmatrix} -1 & 1 \end{bmatrix} \quad h_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

b) του Sobel

$$S_{k=} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_{j=} = \begin{bmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & 1 \end{bmatrix}$$

a)

$D_k(k, j)$  με εφαρμογή της  $h_1$

$D_j(k, j)$  με εφαρμογή της  $h_2$

$$\begin{pmatrix} 10 & 90 & 0 & -80 & 10 \\ 10 & 90 & 10 & -90 & 10 \\ 10 & 90 & 40 & 0 & -110 \\ 10 & 0 & 100 & -90 & 10 \\ 10 & 0 & 90 & -80 & 10 \end{pmatrix} \quad \begin{pmatrix} 10 & 100 & 100 & 20 & 30 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 30 & 120 & 0 \\ 0 & -90 & -30 & -120 & 0 \\ 0 & 0 & -10 & 0 & 0 \end{pmatrix}$$

$|\mathbf{D}(k, j)|$

Οι ακμές για  $|\mathbf{D}(k, j)| \geq 90$



14.14	134.54	100	82.46	31.62	0	1	1	0	0
10	90	14.14	90	10	0	1	0	1	0
10	90	50	120	110	0	1	0	1	1
10	90	104.4	150	10	0	1	1	1	0
10	0	90.55	80	10	0	0	1	0	0

b) Για τη μάσκα του Sobel

$D_k(k, j)$  με εφαρμογή της  $S_k$

$D_j(k, j)$  με εφαρμογή της  $S_j$

1/4	300	280	-240	-220	-60	, 1/4	120	320	340	180	80
	400	420	-200	-340	-200		0	40	200	280	120
	310	460	10	-380	-320		-90	-180	-90	0	0
	130	420	70	-340	-200		-90	-220	-290	-280	-120
	30	280	30	-220	-60		-30	-140	-250	-180	-80

$|\mathbf{D}(k, j)|$

Οι ακμές για

$|\mathbf{D}(k, j)| \geq 75$

1/4	323.11	425.21	416.17	284.25	100.00	1	1	1	0	0
	400.00	421.90	282.84	440.45	233.24	1	1	0	1	0
	322.80	493.96	90.55	380.00	320.00	1	1	0	1	1
	158.11	474.13	298.33	440.45	233.24	0	1	0	1	0
	42.43	313.05	251.79	284.25	100.00	0	1	0	0	0

### 3.4.2 Ανίχνευση ακμών με την χρήση του τελεστή Laplace

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  $\nabla_{1K}$ . 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

$$\nabla^2 = \begin{bmatrix} & & \\ & 4 & \\ & & \end{bmatrix}$$

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 βρίσκονται στα σημεία με τιμή μηδέν και εκατέρωθεν των οποίων υπάρχουν ετερόσημες τιμές. Επειδή μια ψηφιακή εικόνα αποτελείται από δείγματα τιμών, η μια μηδενική τιμή της  $\nabla^2 I(j,k)$  πέφτει σπάνια σε συγκεκριμένο εικονοστοιχείο. Εκείνο που συμβαίνει κυρίως είναι σε δύο γειτονικές, οριζόντια ή κάθετα θέσεις, να εμφανίζονται ετερόσημες τιμές. Ένας απλός τρόπος να παράγουμε την εικόνα των ακμών είναι θεωρούμε σημείο ακμής κάθε στοιχείο  $\nabla^2 I(j,k)$  όταν

- $\nabla^2 I(j,k) \cdot \nabla^2 I(j,k+1) < 0$  ή
- $\nabla^2 I(j,k) \cdot \nabla^2 I(j+1,k) < 0$  ή
- $\nabla^2 I(j,k) = 0$  και  $\nabla^2 I(j,k-1) \cdot \nabla^2 I(j,k+1) < 0$  ή  $\nabla^2 I(j-1,k) \cdot \nabla^2 I(j+1,k) < 0$

Η ύπαρξη ακμών δεν εξετάζεται στα εξωτερικά στοιχεία της εικόνας

Για παράδειγμα με την εφαρμογή των ανωτέρω οι ακμές της εικόνας  $I$  είναι τα σημεία με την τιμή 1 στο πίνακα  $A$ .

$$I = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 10 & 10 & 1 & 1 \\ 1 & 10 & 10 & 10 & 1 \\ 1 & 1 & 10 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{vmatrix} \quad \nabla^2 I = \begin{vmatrix} -2 & 8 & 8 & -1 & -2 \\ 8 & -18 & -18 & 18 & -1 \\ 8 & -18 & 0 & -27 & 8 \\ -1 & 18 & -27 & 18 & -1 \\ -2 & -1 & 8 & -1 & -2 \end{vmatrix}$$

$$A = \begin{vmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Η χρήση του τελεστή Laplace πλεονεκτεί, διότι εντοπίζονται τα τοπικά μέγιστα της πρώτης παραγώγου και προκύπτουν λεπτές ακμές. Μειονεκτεί όμως, διότι παρουσιάζει ευαισθησία στο θόρυβο και στις μικρές αυξομειώσεις της έντασης της φωτεινότητας. Για το λόγο αυτό εφαρμόζεται σε συνδυασμό με τον  $\nabla_1$ , τεχνικές τοπικής διασποράς και τεχνικές αφαίρεσης θορύβου.

Για την αφαίρεση του θορύβου χρησιμοποιείται ευρύτατα η μάσκα εξομάλυνσης του Gauss. Στην εικόνα που προκύπτει από την εξομάλυνση εφαρμόζεται η μάσκα Laplace. Επειδή οι δύο μάσκες είναι γραμμικά συστήματα ανεξάρτητα από την μετατόπιση με κρουστικές αποκρίσεις  $h_G$  και  $h_L$  αντίστοιχα η τελική έξοδος θα προκύπτει από την διαδοχική εφαρμογή του δισδιάστατου συνελκτικού αθροίσματος σύμφωνα με την σχέση

$$I' = h_L ** (h_G ** I) = (h_L ** h_G) ** I$$

Την μάσκα που αντιστοιχεί σε ένα σύστημα με κρουστική απόκριση  $h_L ** h_G$  ονομάζουμε μάσκα LoG και αποτελεί την διακριτή έκδοση εφαρμογής του τελεστή Laplace σε δισδιάστατη Gaussian συνάρτηση. Ακολούθως φαίνονται οι τιμές μιας μάσκας Log 5 X 5

$$W_{LoG} = \begin{vmatrix} 0 & 0.0751 & 0.1238 & 0.0751 & 0 \\ 0.0751 & -0.0528 & -0.1410 & -0.0528 & 0.0751 \\ 0.1238 & -0.1410 & -0.3214 & -0.1410 & 0.1238 \\ 0.0751 & -0.0528 & -0.1410 & -0.0528 & 0.0751 \\ 0 & 0.0751 & 0.1238 & 0.0751 & 0 \end{vmatrix}$$

Στην γενική περίπτωση η εφαρμογή του τελεστή Laplace στην δισδιάστατη Gaussian συνάρτηση

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

ικανοποιεί την σχέση

$$\nabla^2 G(x, y) = -\frac{1}{\pi\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}} \left(1 - \frac{x^2 + y^2}{2\sigma^2}\right)$$

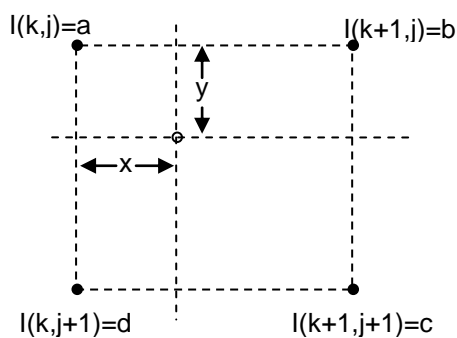
Με βάση όσα αναπτύχθηκαν παραπάνω έχουν προταθεί τεχνικές ανίχνευσης των ακμών που αντιμετωπίζουν προβλήματα όπως του θορύβου, της μεταβολής του φωτισμού, της συνέχειας των ακμών κ.α. Από τις πλέον επιτυχημένες και διαδεδομένες είναι ο ανιχνευτής ακμών του *Canny* (*Canny edge detector*).

### 3.5 Μεταβολή μεγέθους – Παρεμβολή τιμών

Η σμίκρυνση, η μεγέθυνση, η περιστροφή και άλλοι μετασχηματισμοί της εικόνας απαιτούν τον υπολογισμό της φωτεινότητας σε σημεία άλλα από αυτά που ορίζει ο αρχικός κάναβος των συντεταγμένων  $(k,j)$ . Οι συνηθέστερες μέθοδοι που χρησιμοποιούνται για το σκοπό αυτό είναι:

- του κοντινότερου γείτονα (nearest neighbor)
- της δι-γραμμικής παρεμβολής (bi-linear interpolation)
- της δι-κυβικής παρεμβολής (bi-cubic interpolation)

Σύμφωνα με την μέθοδο του κοντινότερου γείτονα η τιμή στην θέση  $(x,y)$  είναι ίση με την τιμή του κοντινότερου προς αυτήν εικονοστοιχείου. Για παράδειγμα στο σχήμα 3.5.1. η τιμή της φωτεινότητας στη θέση  $(x,y)$  είναι  $I(x,y) = I(k,j)$ .



Σχήμα 3.5.1

Στην μέθοδο της δι-γραμμικής παρεμβολής υπολογίζεται η εξίσωση της ευθείας που διέρχεται από τα σημεία  $I(k,j)$  και  $I(k+1,j)$ . Η εξίσωση της οριζόντιας αυτής ευθείας είναι:

$$I(x,j) = a + x \cdot (b - a)$$

Όμοια η εξίσωση της ευθείας που διέρχεται από τα σημεία  $I(k,j+1)$  και  $I(k+1,j+1)$  θα είναι

$$I(x,j+1) = d + x \cdot (c - d)$$

Η εξίσωση της ευθείας που διέρχεται από τα σημεία  $I(x,j)$  και  $I(x,j+1)$  είναι:

$$I(x,y) = I(x,j) + y \cdot [I(x,j+1) - I(x,j)] = a + (b - a) \cdot x + (d - a) \cdot y + x \cdot y \cdot (c + a - d - b)$$

Το αποτέλεσμα δεν αλλάζει αν η παρεμβολή ξεκινήσει από τις κάθετες ευθείες.