

## Κεφάλαιο 2

# Δυαδική μετατροπή

### 2.1 Γενικά για την δυαδική μετατροπή

Η δυαδική μετατροπή των εικόνων (binarization - thresholding) είναι το πρώτο βήμα των περισσότερων συστημάτων ανάλυσης και επεξεργασίας εγγράφων και αναφέρεται στην μετατροπή των gray scale εικόνων σε ασπρόμαυρες. Η δυαδική μετατροπή αποτελεί βασικό στοιχείο των συστημάτων επεξεργασίας εγγράφων γιατί μία βέλτιστη δυαδική μετατροπή αποτελεί την βάση για σωστή κατάτμηση και αναγνώριση των εγγράφων. Οι βασικοί λόγοι που το στάδιο της δυαδικής μετατροπής είναι απαραίτητο είναι οι παρακάτω:

- Εξοικονόμηση αποθηκευτικών μέσων.
- Βελτίωση της εμφάνισης λόγω αφαίρεσης θορύβου.
- Διευκόλυνση και επιτάχυνση αλγορίθμων κατάτμησης και αναγνώρισης.

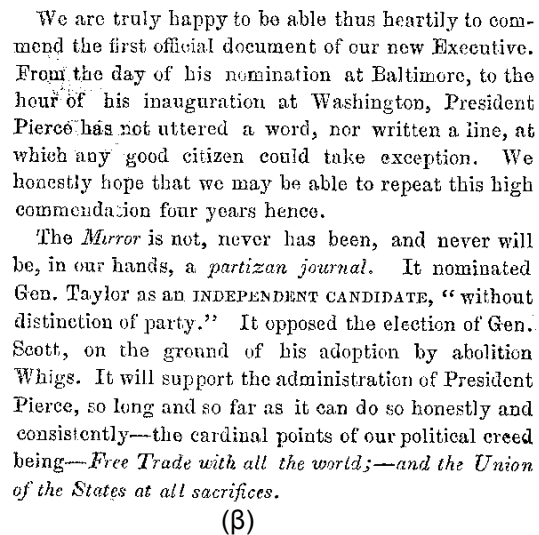
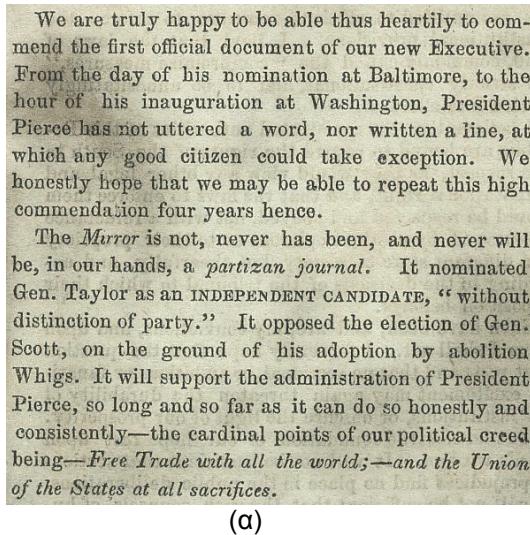
Στην βιβλιογραφία η δυαδική μετατροπή μπορεί και να χρησιμοποιείται και ως τεχνική εντοπισμού του κειμένου αφού συνήθως διακρίνει τις περιοχές του κειμένου από τις περιοχές του υποβάθρου της εικόνας.

Έστω μία εικόνα με 256 στάθμες του γκρι όπου οι τιμές κοντά στο 0 είναι οι σκούρες περιοχές που ανήκουν στο κείμενο ενώ οι τιμές κοντά στο 255 είναι οι πιο ανοιχτές περιοχές και ανήκουν στο υπόβαθρο της εικόνας. Ο πιο εύκολος και συνηθισμένος τρόπος μετατροπής μιας εικόνας πολλών σταθμών του γκρι σε ασπρόμαυρη είναι η επιλογή ενός κατωφλιού (threshold), σύμφωνα με το οποίο όλες οι γκρι στάθμες κάτω από αυτό μετατρέπονται σε 1 (κείμενο) ενώ εκείνες που είναι πάνω από το κατώφλι μετατρέπονται σε 0 (υπόβαθρο):

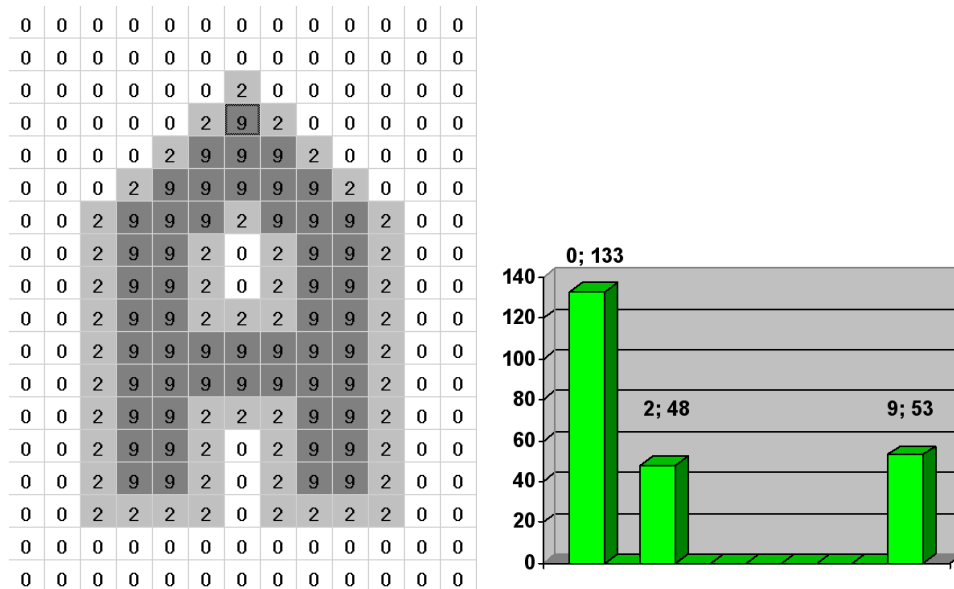
$$f(x_{i,j}) = \begin{cases} 1, & x_{i,j} \leq T \\ 0, & x_{i,j} > T \end{cases} \quad (2.1)$$

Οι μέθοδοι που χρησιμοποιούν μία τιμή κατωφλιού  $T$  για την μετατροπή της εικόνας σε ασπρόμαυρη σύμφωνα με τον τύπο (1) ονομάζονται μέθοδοι καθολικής κατωφλίωσης (global thresholding methods). Ένα παράδειγμα καθολικής κατωφλίωσης φαίνεται στο σχήμα 2.1.

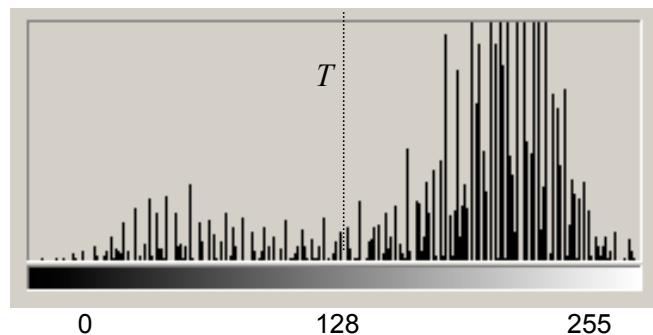
Το ιστόγραμμα  $H$  μιας gray scale εικόνας εκφράζει την κατανομή των αποχρώσεων του γκρι της εικόνας. Στο σχήμα 2.2 δίδεται ένα παράδειγμα ιστογράμματος μιας απλής gray scale εικόνας. Στο παράδειγμα αυτό εμφανίζονται μόνο 3 επίπεδα φωτεινότητας (0,2,9) τα οποία αντιστοιχούν σε 133,48 και 53 pixels αντίστοιχα. Σε περιπτώσεις καθαρών εγγράφων το ιστόγραμμα  $H$  εμφανίζει δύο βασικές κατανομές που αντιστοιχούν στις περιοχές του υποβάθρου της εικόνας και του κειμένου (σχήμα 2.3). Το κατώφλι  $T$  που χρησιμοποιείται στις μεθόδους καθολικής κατωφλίωσης θα πρέπει να διαχωρίζει ικανοποιητικά τις δύο βασικές κατανομές του ιστογράμματος. Η σωστή επιλογή του κατωφλιού είναι σημαντική τόσο για τον περιορισμό του θορύβου όσο και για την ποιότητα και το πάχος των χαρακτήρων της τελικής ασπρόμαυρης εικόνας. Στο σχήμα 2.4 φαίνονται οι διαφορετικές τελικές εικόνες που προκύπτουν για διαφορετικές τιμές του κατωφλιού  $T$ . Παρατηρούμε ότι για μικρές τιμές του  $T$  ( $T = 60, 80$ ) δεν έχουμε θόρυβο, όμως έχουμε ασυνέχειες στο σώμα των χαρακτήρων. Αντίθετα για μεγάλες τιμές του  $T$  ( $T = 140, 160$ ) έχουμε εισαγωγή θορύβου, αλλά η ποιότητα των χαρακτήρων είναι καλύτερη.



**Σχήμα 2.1.** Παραδείγματα δυαδικής μετατροπής με χρήση καθολικής κατωφλίωσης: (α) αρχική gray scale εικόνα, (β) ασπρόμαυρη εικόνα μετά την δυαδική μετατροπή.



**Σχήμα 2.2.** Το ιστόγραμμα της απλής gray scale εικόνας. (α) Η gray scale εικόνα. (β) Το ιστόγραμμά της όπου εμφανίζονται 3 επίπεδα φωτεινότητας.



**Σχήμα 2.3.** Το ιστόγραμμα της εικόνας του σχήματος 2.1α και η τιμή  $T$  του κατωφλιού που έχει επιλεγεί για την δυαδική μετατροπή της.

## Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Οι μέθοδοι καθολικής κατωφλίωσης μπορούν να εφαρμοστούν μόνο στις περιπτώσεις όπου υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο. Αυτό όμως δεν συμβαίνει στις περισσότερες περιπτώσεις των εγγράφων. Ιδίως στα παλιά έγγραφα έχουμε χαμηλή ποιότητα χαρακτήρων, σκιές, μη ομοιόμορφη φωτεινότητα υποβάθρου, έντονο θόρυβο κ.λ.π. (σχήμα 2.5). Στο ιστόγραμμα αυτών των εικόνων οι τιμές του γκρι που αντιστοιχούν στο κείμενο και στο υπόβαθρο επικαλύπτονται (σχήμα 2.5α). Για τους παραπάνω λόγους αναπτύχθηκαν οι μέθοδοι προσαρμοσμένης ή τοπικής κατωφλίωσης (local or regional or adaptive thresholding methods) στις οποίες χρησιμοποιούνται πολλές τιμές κατωφλίων ανάλογα με την τοπική πληροφορία της εικόνας. Έτσι επιτυγχάνεται η βέλτιστη δυαδική μετατροπή ακόμα και στις περιπτώσεις που δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο (σχήμα 2.6).

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(α)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(γ)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(ε)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(β)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(δ)

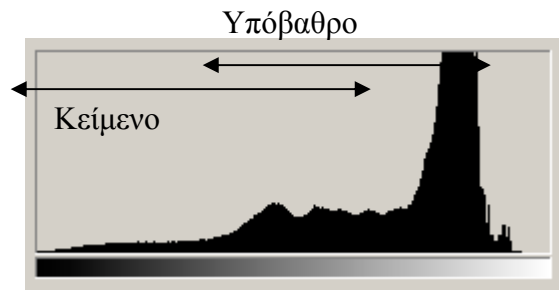
We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—Free Trade with all the world;—and the Union of the States at all sacrifices.

(ζ)

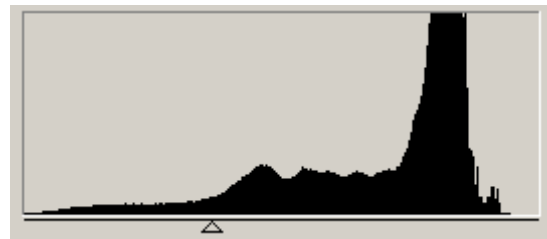
**Σχήμα 2.4.** Ασπρόμαυρες εικόνες που προκύπτουν από την εικόνα 2.1α για εφαρμογή καθολικής κατωφλίωσης επιλέγοντας διαφορετικές τιμές για το κατώφλι  $T$ . (α)  $T = 60$ , (β)  $T = 80$ , (γ)  $T = 100$ , (δ)  $T = 120$ , (ε)  $T = 140$ , (ζ)  $T = 160$ .

not so, or can not well do,  
fall into two classes, those  
entirely to Europe, and those  
to each of these branches  
with variety of subdivisions  
in relation to Europe—Europe  
is, misdeeds, and non  
misdeeds—The other spheres are  
law, and without wrong, regions  
as public roads and high  
roads, charities, pauperism,  
is of the deceased, and the  
government itself—  
years that if all men were just  
can be some, though not some.



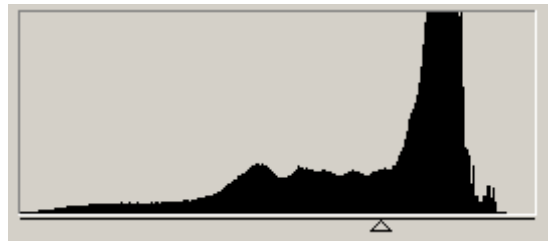
(α)

not so, or can not well do,  
fall into two classes, those  
entirely to Europe, and those  
to each of these branches  
with variety of subdivisions  
in relation to Europe—Europe  
is, misdeeds, and non  
misdeeds—The other spheres are  
law, and without wrong, regions  
as public roads and high  
roads, charities, pauperism,  
is of the deceased, and the  
government itself—  
years that if all men were just  
can be some, though not some.



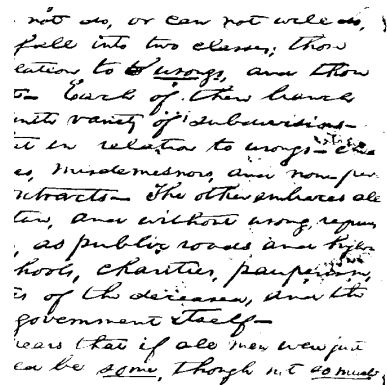
(β)

not so, or can not well do,  
fall into two classes, those  
entirely to Europe, and those  
to each of these branches  
with variety of subdivisions  
in relation to Europe—Europe  
is, misdeeds, and non  
misdeeds—The other spheres are  
law, and without wrong, regions  
as public roads and high  
roads, charities, pauperism,  
is of the deceased, and the  
government itself—  
years that if all men were just  
can be some, though not some.



(γ)

**Σχήμα 2.5.** Διαδική μετατροπή ιστορικού εγγράφου με χρήση καθολικής κατωφλίωσης. (α) Στο ιστόγραμμα οι τιμές του γκρι που αντιστοιχούν στο κείμενο και στο υπόβαθρο επικαλύπτονται. (β),(γ) Είναι αδύνατη η επιλογή ενός κατωφλιού για επιτυχή δυαδική μετατροπή.



not so, or can not well be, }  
 fall into two classes; those  
 either to ~~be~~ <sup>be</sup> ~~used~~, and then  
 the ~~best~~ of ~~the~~ <sup>best</sup> ~~ones~~  
 into variety of ~~subdivisions~~  
 it is relative to ~~usage~~ <sup>usage</sup> ~~the~~  
 e, ~~misadventures~~, and ~~non~~ <sup>non</sup> ~~for~~  
~~interaction~~ the ~~other~~ <sup>other</sup> ~~embodies~~ ~~de~~  
~~ten~~, and ~~without~~ ~~among~~ ~~regions~~  
 , as public roads and ~~high~~ <sup>high</sup> ~~ways~~  
 hosts, charities, ~~pauperism~~ <sup>the</sup>  
 is of the ~~disasters~~, and the  
 government itself  
 sees that if all men were just  
 can be ~~some~~, though not ~~some~~

**Σχήμα 2.6.** Διαδική μετατροπή του ιστορικού εγγράφου του σχήματος 2.4α με χρήση προσαρμοσμένης κατωφλίωσης.

## 2.2 Καθολική κατωφλίωση – Global Thresholding

Στη συνέχεια του κεφαλαίου παρατίθενται χαρακτηριστικά παραδείγματα μεθοδολογιών καθολικής κατωφλίωσης. Ζητούμενο όλων των μεθοδολογιών είναι ο υπολογισμός του καθολικού κατωφλίου ώστε να επιτυγχάνεται βέλτιστη διαδική μετατροπή της εικόνας.

### Χρήση σημείων ακμών – Weszka

Τα σημεία ακμών είναι σημεία ανάμεσα στο αντικείμενο και στο υπόβαθρο της εικόνας ή μεταξύ δύο αντικειμένων. Το ιστόγραμμα της εικόνας που προκύπτει μόνο από τα σημεία αυτά μας δίνει καλύτερη διαχωρισιμότητα κειμένου-υποβάθρου σε σχέση με το ιστόγραμμα της συνολικής εικόνας. Η ιδέα αυτή χρησιμοποιήθηκε από τον (Weszka 1974) ο οποίος πρότεινε μία μέθοδο καθολικής κατωφλίωσης βασισμένη στην Laplacian της αρχικής εικόνας, η οποία είναι μέθοδος προσδιορισμού ακμών. Πρώτα υπολογίζεται η Laplacian της εικόνας των σταθμών του γκρι. Αυτό μπορεί να γίνει με την εφαρμογή της παρακάτω μάσκας στην επιφάνεια της εικόνας:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Στη συνέχεια προσδιορίζεται το ιστόγραμμα της αρχικής εικόνας λαμβάνοντας υπόψη μόνο τα σημεία που έχουν υψηλές Laplacian τιμές. Στο σχήμα 2.6 φαίνεται η Laplacian αρχικής εικόνας καθώς και το ιστόγραμμα που προκύπτει από την συνολική εικόνα καθώς και από τα pixels που έχουν υψηλές Laplacian τιμές. Στην δεύτερη περίπτωση υπάρχει μεγαλύτερη διαχωρισιμότητα κειμένου-υποβάθρου.

Για τον εντοπισμό των δύο κορυφών του ιστογράμματος οι δύο τιμές της φωτεινότητας  $i_1$  και  $i_2$  μπορούν να εντοπιστούν ως εξής:

$$\begin{aligned} i_1: \text{Hist}(i) &= \max \text{ για } i = i_1 \\ i_2: \text{Hist}(i) * (i - i_1)^2 &= \max \text{ για } i = i_2 \end{aligned}$$

Με τον αλγόριθμο αυτό εξασφαλίζεται ότι η δεύτερη κορυφή του ιστογράμματος απέχει από την πρώτη (σχήμα 2.7).

### Επιλογή μετά από επανάληψη – Ridler

Η επιλογή μετά από επανάληψη (Ridler 1978) είναι μια μεθοδολογία η οποία ξεκινά από μία αρχική εκτίμηση για την τιμή του κατωφλίου και στη συνέχεια γίνεται διόρθωση αυτής της τιμής λαμβάνοντας υπόψη τις περιοχές κειμένου και υποβάθρου που προκύπτουν κάθε φορά. Η αρχική εκτίμηση του κατωφλίου είναι η μέση τιμή των gray scale τιμών. Αυτό το κατώφλι χρησιμοποιείται για να υπολογιστούν στατιστικές για τις περιοχές κειμένου και υποβάθρου που προκύπτουν. Υπολογίζεται η μέση τιμή των gray scale τιμών των pixels κάτω από το αρχικό κατώφλι και αποθηκεύεται στην μεταβλητή  $T_b$ . Αντίστοιχα, υπολογίζεται και η μέση

## Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

τιμή των gray scale τιμών των pixels πάνω από το αρχικό κατώφλι και αποθηκεύεται στην μεταβλητή  $T_o$ . Στη συνέχεια υπολογίζεται μία νέα προσέγγιση του κατωφλιού στην τιμή  $(T_o + T_o)/2$  και η διαδικασία επαναλαμβάνεται μέχρι να μην υπάρχει αλλαγή στο κατώφλι μεταξύ δύο διαδοχικών επαναλήψεων (σχήμα 2.8).

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image  $I$  belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

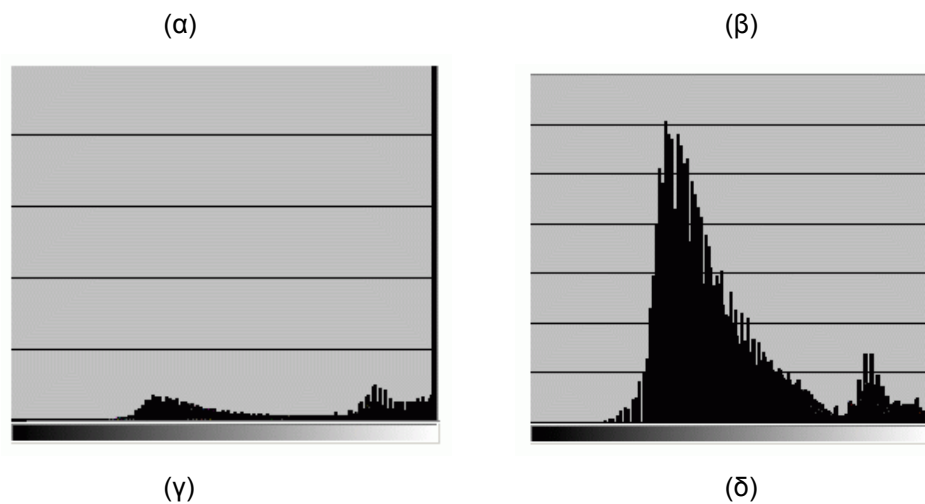
The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image  $I$  belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

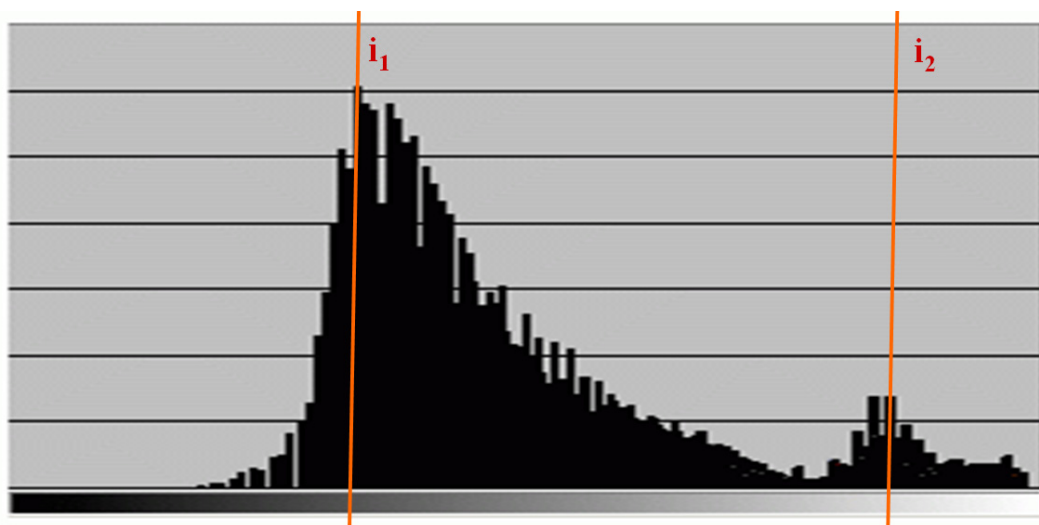
The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to



**Σχήμα 2.6.** Καθολική κατωφλίωση με χρήση των σημείων ακμής. (α) Αρχική gray scale εικόνα. (β) Laplacian της αρχικής εικόνας. (γ) Ιστογράμμο αρχικής εικόνας. (δ) Ιστογράμμο της αρχικής εικόνας λαμβάνοντας υπόψη μόνο τα σημεία που έχουν υψηλές Laplacian τιμές.



**Σχήμα 2.7.** Εύρεση των δύο κορυφών του ιστογράμματος.

## Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(α)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(β)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(γ)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(δ)

**Σχήμα 2.8.** Καθολική κατωφλίωση με χρήση της επιλογής μετά από επανάληψη για την εικόνα του σχήματος 2.6α. (α) 1<sup>η</sup> επανάληψη, κατώφλι = 238. (β) 2<sup>η</sup> επανάληψη, κατώφλι = 209. (γ) 3<sup>η</sup> επανάληψη, κατώφλι = 189. (δ) 4<sup>η</sup>, 5<sup>η</sup> επανάληψη, κατώφλι = 182.

### Επεξεργασία του ιστογράμματος – Otsu

Η μέθοδος του Otsu (Otsu 1979) βασίζεται στην επεξεργασία του ιστογράμματος της εικόνας και στον προσδιορισμό του κατωφλιού βάσει του κριτηρίου της μεγιστοποίησης της διαχωρισιμότητας μεταξύ των περιοχών κειμένου και υποβάθρου. Αρχικά, υπολογίζουμε το ολικό τετράγωνο της τυπικής απόκλισης (global variance) των επιπέδων του γκρι της εικόνας  $\sigma_t^2$ :

$$\sigma_t^2 = \sum_{i=0}^{255} (i - u_t)^2 p(i) \quad (2.2)$$

όπου  $p(i)$  η πιθανότητα εμφάνισης της  $i$  στάθμης της εικόνας και  $u_t$  το συνολικό μέσο επίπεδο γκρι της εικόνας:

$$p(i) = \frac{H(i)}{N}, \quad u_t = \sum_{i=0}^{255} i p(i) \quad (2.3)$$

και  $N$  το συνολικό πλήθος των pixels της εικόνας.

Για κάθε υποψήφιο κατώφλι  $t$  το ιστόγραμμα της εικόνας χωρίζεται σε δύο περιοχές - κλάσεις. Μπορούμε να υπολογίζουμε το τετράγωνο της τυπικής απόκλισης μεταξύ των δύο κλάσεων (between classes variance)  $\sigma_b^2$  το οποίο είναι το τετράγωνο της τυπικής απόκλισης των μέσων τιμών κάθε κλάσης από την συνολική μέση τιμή όλων των pixels και υπολογίζεται από τον παρακάτω τύπο:

## B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

$$\sigma_b^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \quad (2.4)$$

όπου τα  $\omega_0$ ,  $\omega_1$ ,  $\mu_0$  και  $\mu_1$  υπολογίζονται από τους τύπους:

$$\omega_0 = \sum_{i=0}^t p(i), \quad \omega_1 = 1 - \omega_0, \quad \mu_0 = \frac{\sum_{i=0}^t i p(i)}{\omega_0}, \quad \mu_1 = \frac{\sum_{i=t+1}^{255} i p(i)}{\omega_1} \quad (2.5)$$

Σύμφωνα με τον Otsu, η βέλτιστη τιμή για το κατώφλι  $t$  επιτυγχάνεται όταν μεγιστοποιείται ο λόγος του τετραγώνου της τυπικής απόκλισης μεταξύ των δύο κλάσεων (between classes variance) προς το ολικό τετράγωνο της τυπικής απόκλισης (global variance) των επιπέδων του γκρι της εικόνας:

$$n(t) = \frac{\sigma_b^2}{\sigma_t^2} \quad (2.6)$$

Η μέθοδος του Otsu έχει αξιολογηθεί σαν μία από τις καλύτερες τεχνικές εύρεσης καθολικού κατωφλιού. Στο σχήμα 2.9 φαίνεται το αποτέλεσμα εύρεσης κατωφλιού με την παραπάνω μεθόδο για την εικόνα του σχήματος 2.6α.

### Χρησιμοποιώντας την εντροπία – Pun, Karu

Η εντροπία είναι μία μετρική που σχετίζεται με το περιεχόμενο της πληροφορίας. Αν έχουμε  $n$  πιθανά σύμβολα (π.χ. γράμματα ή ψηφία) και το σύμβολο  $i$  εμφανίζεται με πιθανότητα  $p(x_i)$ , η εντροπία που σχετίζεται με τα σύμβολα  $X$  είναι:

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (2.7)$$

όπου η εντροπία μετρείται σε bits/σύμβολο.

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold  $T$ . What is being assumed here is that the pixels in an image  $I$  belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

**Σχήμα 2.9.** Διαδική μετατροπή της εικόνας του σχήματος 2.6α εφαρμόζοντας με την μέθοδο του Otsu.



## Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Η εικόνα μπορεί να θεωρηθεί σαν ένα σύνολο συμβόλων ή τιμών του γκρι. Θεωρώντας ένα κατώφλι  $t$ , η εντροπία που σχετίζεται με τα pixels κειμένου/φόντου μπορεί να δοθεί από τον τύπο (Pun 1981):

$$H_b = - \sum_{i=1}^t p_i \log(p_i) \quad (2.8)$$

όπου  $p_i$  είναι η πιθανότητα εμφάνισης της στάθμης  $i$ . Παρόμοια, η εντροπία των pixels του υποβάθρου της εικόνας είναι:

$$H_w = - \sum_{i=t+1}^{255} p_i \log(p_i) \quad (2.9)$$

Ο προτεινόμενος αλγόριθμος (Pun 1981) προσπαθεί να βρει την βέλτιστη τιμή κατωφλιού  $t$  η οποία μεγιστοποιεί την  $H = H_b + H_w$ .

Μία παραλλαγή του παραπάνω αλγορίθμου προτάθηκε από τον (Karur 1985) ο οποίος ορίζει τις πιθανότητες των δύο βασικών κατανομών ως εξής:

$$A: \frac{p_0}{P_t}, \frac{p_1}{P_t}, \dots, \frac{p_t}{P_t}, B: \frac{p_{t+1}}{1-P_t}, \frac{p_{t+2}}{1-P_t}, \dots, \frac{p_{255}}{1-P_t} \quad (2.10)$$

όπου  $P_t$  η αθροιστική πιθανότητα εμφάνισης μέχρι την στάθμη  $t$ :

$$P_t = \sum_{i=0}^t p_i \quad (2.11)$$

Οι εντροπίες των περιοχών κειμένου/φόντου και υποβάθρου υπολογίζονται ανάλογα:

$$H_b = - \sum_{i=1}^t \frac{p_i}{P_t} \log\left(\frac{p_i}{P_t}\right), H_w = - \sum_{i=t+1}^{255} \frac{p_i}{1-P_t} \log\left(\frac{p_i}{1-P_t}\right) \quad (2.12)$$

### Χρησιμοποιώντας την ασαφή λογική – Huang

Σύμφωνα με την ασαφή λογική, ένα στοιχείο  $x$  ανήκει σε ένα σύνολο  $S$  με συγκεκριμένη πιθανότητα  $u_x$ . Κατά την δυαδική μετατροπή της εικόνας προσπαθούμε να κατατάξουμε τα pixels ότι ανήκουν στο κείμενο/φόντο ή στο υπόβαθρο της εικόνας, άρα η χρήση της ασαφούς λογικής έχει άμεση εφαρμογή στο πρόβλημα. Σύμφωνα με την μέθοδο του (Huang 1995) σαν μετρική της ασάφειας (fuzziness) μπορεί να θεωρηθεί η απόσταση μεταξύ της αρχικής gray scale εικόνας με την παραγόμενη δυαδική εικόνα. Ελαχιστοποιώντας την ασάφεια, μπορούμε να εντοπίσουμε την βέλτιστη τιμή κατωφλιού.

Καταρχήν ορίζουμε την membership function η οποία είναι η πιθανότητα που σχετίζεται με το αν ένα pixel έχει ταξινομηθεί σαν κείμενο/φόντο ή υπόβαθρο. Έστω οι μέσες στάθμες του γκρι για τις περιοχές υποβάθρου είναι  $\mu_0$  και για του κειμένου  $\mu_1$ . Όσο μικρότερη η απόσταση μεταξύ ενός pixel  $x$  από την αντίστοιχη μέση τιμή της κατηγορίας του, τόσο μεγαλύτερη πρέπει να είναι η τιμή της membership function  $u_x$ . Σύμφωνα με τον (Huang 1995), προτείνεται η εξής συνάρτηση  $u_x$ :

$$u_x(x) = \begin{cases} \frac{1}{1 + |x - \mu_0|/C}, & \text{if } x \leq t \\ \frac{1}{1 + |x - \mu_1|/C}, & \text{if } x > t \end{cases} \quad (2.13)$$

για ένα δεδομένο κατώφλι  $t$ .  $C$  είναι η διαφορά μεταξύ των μεγίστων και των ελάχιστων σταθμών του γκρι. Ο βαθμός ενσωμάτωσης ενός pixel σε μία κατηγορία δίδεται από την συνάρτηση  $u_x$  και κυμαίνεται από 0,5 μέχρι 1.

Δεδομένης της membership function, θέλουμε να ορίσουμε μία μετρική για τον βαθμό της ασάφειας της δυαδικής μετατροπής που πραγματοποιείται χρησιμοποιώντας το κατώφλι  $t$ . Για παράδειγμα η μετρική αυτή θα πρέπει να είναι 0 αν η εικόνα είναι ήδη ασπρόμαυρη και θέτουμε  $t = 0$ . Αντίστοιχα, η μέγιστη τιμή της ασάφειας θα πρέπει να είναι 1. Ένας τρόπος μέτρησης της ασάφειας βασίζεται στην εντροπία του ασαφούς συνόλου και στην συνάρτηση του Shannon:

$$H_f(x) = -x \log(x) - (1-x) \log(1-x) \quad (2.14)$$

Η εντροπία του συνολικού ασαφούς συνόλου (εικόνα) δίδεται από τον τύπο:

$$E(t) = \frac{1}{MN} \sum_g H_f(u_x(g)) h(g) \quad (2.15)$$

για όλες τις στάθμες του γκρι  $g$ , όπου  $N$  και  $M$  είναι οι γραμμές και οι στήλες της εικόνας και  $h$  το gray scale ιστόγραμμα. Η  $E$  είναι συνάρτηση του  $t$  επειδή η  $u_x$  είναι. Η τιμή  $t$  που ελαχιστοποιεί την  $E(t)$ , ελαχιστοποιεί και την ασάφεια της δυαδικής μετατροπής.

## 2.2 Προσαρμοσμένη κατωφλίωση – Adaptive Thresholding

Στη συνέχεια του κεφαλαίου παρατίθενται χαρακτηριστικά παραδείγματα μεθόδων προσαρμοσμένης κατωφλίωσης. Σύμφωνα με αυτές αναζητούνται πολλές τιμές κατωφλίων ανάλογα με την τοπική πληροφορία της εικόνας και έτσι επιτυγχάνεται η βέλτιστη δυαδική μετατροπή ακόμα και στις περιπτώσεις που δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο.

### Διαίρεση της εικόνας – Chow

Σύμφωνα με την μέθοδο του (Chow 1972) η εικόνα διαιρείται σε επικαλυπτόμενες περιοχές. Για παράδειγμα για μία εικόνα 256x256 προτείνεται η διαίρεση σε 49 επικαλυπτόμενες περιοχές οι οποίες αποτελούνται από 64x64 pixels. Καταρχήν, υπολογίζεται το ιστόγραμμα για κάθε περιοχή και γίνεται έλεγχος ύπαρξης δύο βασικών κατανομών. Στην περίπτωση που υπάρχουν οι δύο κατανομές, θα υπάρχουν και δύο Gaussian καμπύλες που θα προσεγγίζουν τα δεδομένα χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων και κατά συνέπεια θα προσεγγίζεται και το βέλτιστο κατώφλι ανάμεσα στις δύο κατανομές. Στην περίπτωση που δεν υπάρχουν δύο κατανομές τιμών στο ιστόγραμμα των περιοχών, θα γίνεται προσεγγιστικός υπολογισμός (με παρεμβολή) του κατωφλίου χρησιμοποιώντας τα κατώφλια των γειτονικών περιοχών. Τέλος, σε κάθε pixel αντιστοιχίζεται μία τιμή κατωφλίου κάνοντας μία παρεμβολή με τις τιμές των γειτονικών pixels. Η μέθοδος αυτή είναι ο πρόδρομος των μεθόδων προσαρμοσμένης κατωφλίωσης και προτάθηκε κυρίως για την βελτίωση της ποιότητας των εικόνων από καρδιακές ακτινογραφίες.

### Χρήση ακμών - Yanowitz & Bruckstein

Σύμφωνα με την μέθοδο αυτή (Yanowitz 1989) πρώτα εντοπίζονται οι ακμές της εικόνας (σημεία που ανήκουν στα όρια των αντικειμένων). Οι αντίστοιχες στάθμες του γκρι για τα σημεία των ακμών θεωρούνται ενδεικτικές και για τα υπόλοιπα σημεία των αντικειμένων στο οποίο ανήκουν οι ακμές. Στη συνέχεια παράγεται μία επιφάνεια η οποία ταιριάζει στα επίπεδα φωτεινότητας των ακμών και δίνει προσεγγιστικές τιμές για τα υπόλοιπα σημεία των αντικειμένων που δεν ανήκουν στις ακμές της εικόνας. Τα σημεία που έχουν φωτεινότητα σημαντικά πάνω από την επιφάνεια αυτή ανήκουν στο υπόβαθρο, αλλιώς στο κείμενο/φόντο. Η μέθοδος αυτή έχει κυρίως εφαρμογή σε εικόνες όπου η φωτεινότητα του υποβάθρου μεταβάλλεται.

### Χρήση παραθύρου – Niblack, Sauvola

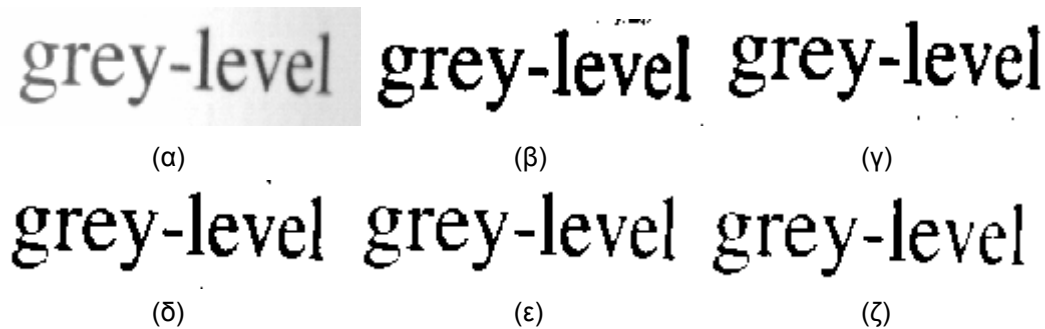
Ο αλγόριθμος του Niblack (Niblack 1986) υπολογίζει ένα κατώφλι για κάθε pixel κάνοντας χρήση ενός παραθύρου το οποίο μετακινείται σε όλη την εικόνα. Το κατώφλι  $T$  του κεντρικού pixel του παραθύρου υπολογίζεται χρησιμοποιώντας την μέση τιμή  $m$  και την τυπική απόκλιση  $s$  των τιμών των επιπέδων φωτεινότητας στο παράθυρο:

$$T = m + k s \quad (2.16)$$

όπου  $k$  είναι σταθερά ίση με -0.2. Η τιμή του  $k$  εξαρτάται από το ποσοστό του ορίου των χαρακτήρων που θεωρείται σαν μέρος των χαρακτήρων (σχήμα 2.10). Η μέθοδος αυτή μπορεί να ξεχωρίσει το αντικείμενο από το φόντο ακόμα και όταν οι φωτεινότητές τους δεν απέχουν αρκετά. Τα αποτελέσματα δεν είναι ιδιαίτερα ευαίσθητα στο μέγεθος του παραθύρου αρκεί το παράθυρο να περιλαμβάνει τουλάχιστον 1-2 χαρακτήρες. Όμως, τυχόν θόρυβος που υπάρχει στο υπόβαθρο της αρχικής εικόνας μπορεί να παραμείνει και στην τελική ασπρόμαυρη εικόνα. Ο αλγόριθμος του Sauvola (Sauvola 2000) δίνει λύση σε αυτό το πρόβλημα εισάγοντας την υπόθεση στα επίπεδα φωτεινότητας του κειμένου-φόντου και του υποβάθρου (τα pixels κειμένου έχουν στάθμες του γκρι κοντά στο 0 ενώ τα pixels του υποβάθρου έχουν στάθμες του γκρι κοντά στο 255). Η εξίσωση υπολογισμού του κατωφλιού γίνεται:

$$T = m + (1 - k (1 - s/R)) \quad (2.17)$$

όπου  $R$  είναι το δυναμικό εύρος της τυπικής απόκλισης ίσο με 128 και  $k$  σταθερά ίση με 0.5.



**Σχήμα 2.10.** Αποτελέσματα δυαδικής μετατροπής με τον αλγόριθμο του Niblack. (α) Αρχική εικόνα. (β) Αποτέλεσμα για  $k=-0.2$ . (γ) Αποτέλεσμα για  $k=-0.5$ . (δ) Αποτέλεσμα για  $k=-0.8$ . (ε) Αποτέλεσμα για  $k=-0.11$ . (ζ) Αποτέλεσμα για  $k=-0.14$ .

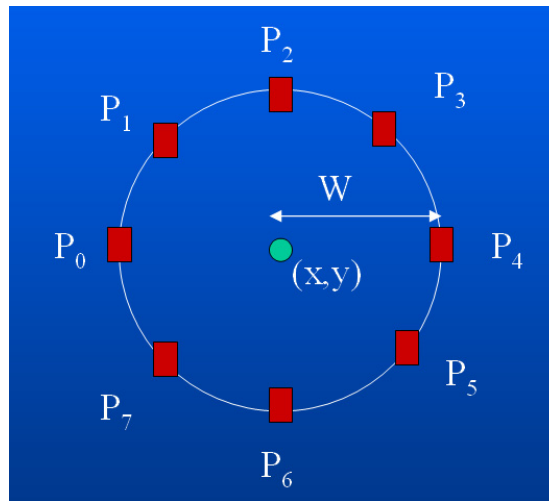
### Πλάτος των γραμμών των χαρακτήρων – Kamel & Zhao

Η μέθοδος (Kamel 1993) βασίζεται στην σύγκριση της gray scale τιμής του pixel  $(x,y)$  με τις μέσες τιμές των pixels των γειτονικών παραθύρων που βρίσκονται σε αποστάσεις όσο το αναμενόμενο πλάτος των γραμμών των χαρακτήρων  $W$  (σχήμα 2.11). Το pixel  $(x,y)$  ανήκει στο κείμενο μόνο αν παρουσιάζει τοπική διαφορά φωτεινότητας αλλά ανήκει και σε χαρακτήρες με γραμμές συγκεκριμένου πλάτους. Τα αντικείμενα που έχουν πάχος μεγαλύτερο από το πλάτος των γραμμών των χαρακτήρων ταξινομούνται σαν υπόβαθρο.

Αν  $\text{ave}(P_i)$  η μέση τιμή των pixels στο παράθυρο  $W \times W$  στο  $P_i$  και  $L(P_i) = \text{ave}(P_i) - f(x,y) > T$  τότε η δυαδική εικόνα  $b(x,y)$  προκύπτει ως εξής:

$$\begin{aligned}
 &\text{if } ((L(P_0) \cap L(P_1) \cap L(P_4) \cap L(P_5)) \\
 &\quad \cup (L(P_1) \cap L(P_2) \cap L(P_5) \cap L(P_6)) \\
 &\quad \cup (L(P_2) \cap L(P_3) \cap L(P_6) \cap L(P_7)) \\
 &\quad \cup (L(P_3) \cap L(P_4) \cap L(P_7) \cap L(P_0))) \\
 &\quad b(x, y) = 1; \\
 &\text{else} \\
 &\quad b(x, y) = 0;
 \end{aligned} \tag{2.18}$$

Η τεχνική αυτή μπορεί να εξάγει χαρακτήρες από θορυβώδες περιβάλλον όμως εξαρτάται από το πλάτος των γραμμών των χαρακτήρων  $W$  και από το όριο  $T$  το οποίο πρέπει να τεθεί από τον χρήστη.

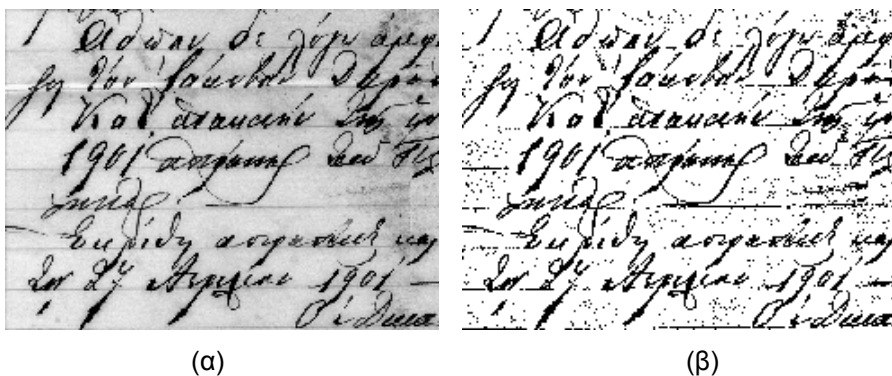


**Σχήμα 2.11.** Γειτονίες σε απόσταση  $W$  από το pixel  $(x,y)$  κατά την μέθοδο προσαρμοσμένης κατωφλίωσης που βασίζεται στο πλάτος  $W$  του γράμματος.

### Εντοπισμός υποβάθρου

Η μέθοδος προσαρμοσμένης κατωφλίωσης που βασίζεται στον εντοπισμό του υποβάθρου (Gatos 2004) αναπτύχθηκε για τον εντοπισμό του κειμένου σε χαμηλής ποιότητας ιστορικά έγγραφα. Αρχικά, χρησιμοποιείται η προσέγγιση του (Niblack 1986) ώστε να γίνει μία πρώτη προσέγγιση των pixels που ανήκουν στο κείμενο-φόντο. Συνήθως, τα πραγματικά pixels του κειμένου είναι υποσύνολο του αποτελέσματος του Niblack αφού ο Niblack εισάγει θόρυβο εντοπίζοντας ταυτόχρονα όλες τις περιοχές κειμένου. Στη συνέχεια, υπολογίζεται η επιφάνεια του υποβάθρου της εικόνας κάνοντας χρήση της παρεμβολής για τον εντοπισμό των φωτεινότητων των pixels που έχουν εντοπιστεί κατά το πρώτο στάδιο. Στο τελικό στάδιο, πραγματοποιείται η τελική κατωφλίωση συνδυάζοντας την εντοπισμένη επιφάνεια του υποβάθρου της εικόνας με την αρχική εικόνα. Οι περιοχές κειμένου εντοπίζονται αν η απόσταση της φωτεινότητας της αρχικής εικόνας από την αντίστοιχη φωτεινότητα της επιφάνειας του υποβάθρου ξεπερνάει κάποιο κατώφλι. Το κατώφλι αυτό προσαρμόζεται ανάλογα με τις τιμές της επιφάνειας του υποβάθρου ώστε να γίνεται εντοπισμός του κειμένου ακόμα και σε αρκετά σκοτεινές περιοχές. Αναλυτικά τα διάφορα στάδια έχουν ως εξής:

**Στάδιο 1.** Αρχική εκτίμηση του κειμένου: Επεξεργαζόμαστε την αρχική εικόνα  $I(x,y)$  ώστε να παράγουμε την δυαδική εικόνα  $N(x,y)$  σύμφωνα με την προσέγγιση του (Niblack 1986). Η εικόνα αυτή έχει 1 για τις εκτιμώμενες περιοχές κειμένου (σχήμα 2.12).



**Σχήμα 2.12.** Προσαρμοσμένη κατωφλίωση χρησιμοποιώντας την προσέγγιση του Niblack. (α) Αρχική εικόνα. (β) Εκτιμώμενες περιοχές κειμένου.

**Στάδιο 2.** Προσέγγιση της επιφάνειας του υποβάθρου: Στο στάδιο αυτό υπολογίζουμε προσεγγιστικά την επιφάνεια του υποβάθρου  $B(x,y)$  της εικόνας. Τα σημεία της αρχικής εικόνας  $I(x,y)$  ανήκουν στην επιφάνεια του υποβάθρου  $B(x,y)$  μόνο αν τα αντίστοιχα σημεία της εικόνας  $N(x,y)$  της αρχικής εκτίμησης του κειμένου έχουν μηδενικές τιμές. Τα υπόλοιπα σημεία της επιφάνειας  $B(x,y)$  υπολογίζονται με παρεμβολή από τα γειτονικά τους. Ο τύπος υπολογισμού της  $B(x,y)$  έχει ως εξής:

$$B(x,y) = \begin{cases} I(x,y) & \text{if } N(x,y) = 0 \\ \frac{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (I(i,j)(1-N(i,j)))}{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (1-N(i,j))} & \text{if } N(x,y) = 1 \end{cases} \quad (2.19)$$

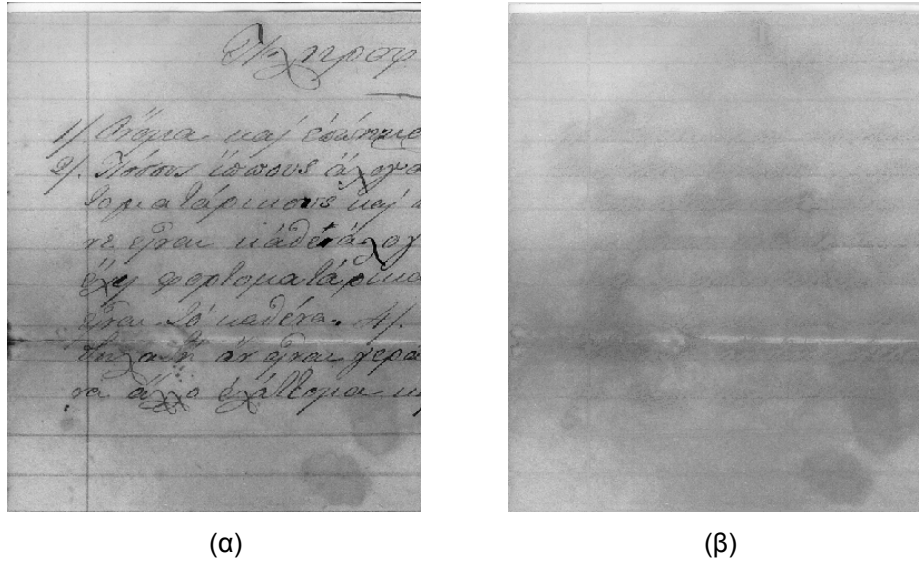
Το παράθυρο  $dx \times dy$  που χρησιμοποιείται για την παρεμβολή ορίζεται ώστε να καλύπτει τουλάχιστον δύο χαρακτήρες του εγγράφου. Ένα παράδειγμα του προσδιορισμού της επιφάνειας του υποβάθρου δίδεται στο σχήμα 2.13.

**Στάδιο 3.** Τελική κατωφλίωση: Στο στάδιο αυτό πραγματοποιούμε την τελική κατωφλίωση συνδυάζοντας την υπολογισμένη επιφάνεια  $B$  με την αρχική εικόνα  $I$ . Οι περιοχές κειμένου εντοπίζονται αν η απόσταση της αρχικής εικόνας από την επιφάνεια του υποβάθρου είναι μεγαλύτερη από ένα κατώφλι  $d$ . Προτείνεται ότι η τιμή του κατωφλίου  $d$  να αλλάζει ανάλογα με την αντίστοιχη τιμή της επιφάνειας  $B$  έτσι ώστε να διασώζεται η πληροφορία του κειμένου ακόμα και σε περιοχές με πολύ σκούρο υπόβαθρο. Για τον λόγο αυτό προτείνεται η τιμή του κατωφλίου  $d$  να παίρνει μικρότερες τιμές στις πιο σκούρες περιοχές. Η τελική δυαδική εικόνα  $T$  δίδεται από τον παρακάτω τύπο:

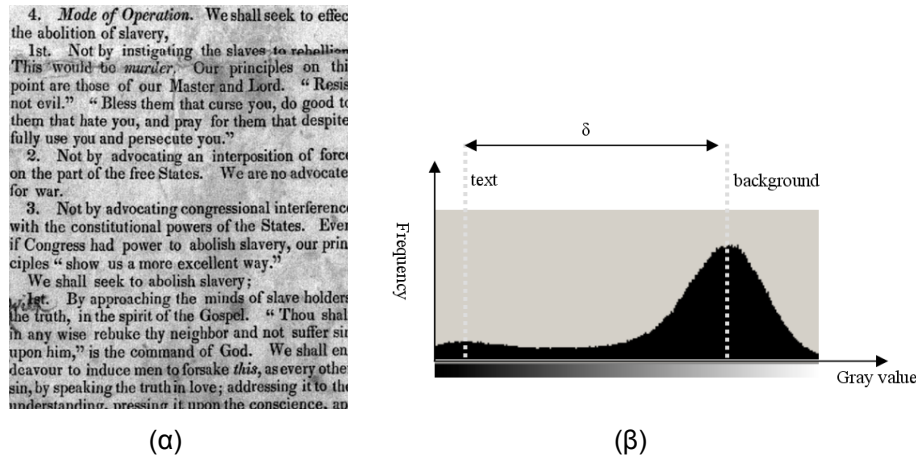
$$T(x,y) = \begin{cases} 1, & \text{if } B(x,y) - I(x,y) > d(B(x,y)) \\ 0, & \text{otherwise} \end{cases} \quad (2.20)$$

Ένα τυπικό ισόγραμμα εικόνας εγγράφου (σχήμα 2.14) έχει δύο κορυφές, μία για το κείμενο και μία για το υπόβαθρο. Η μέση απόσταση  $\delta$  ανάμεσα στις περιοχές κειμένου και φόντου μπορεί να υπολογιστεί από τον παρακάτω τύπο:

$$\delta = \frac{\sum_x \sum_y (B(x,y) - I(x,y))}{\sum_x \sum_y N(x,y)} \quad (2.21)$$



**Σχήμα 2.13.** Υπολογισμός της επιφάνειας του υποβάθρου. (α) Αρχική εικόνα. (β) Η υπολογισμένη επιφάνεια του υποβάθρου της εικόνας.



**Σχήμα 2.14.** Ιστόγραμμα εικόνας εγγράφου. (α) Αρχική εικόνα. (β) Ιστόγραμμα σταθμών του γκρι.

Για τις συνηθισμένες περιπτώσεις εικόνων εγγράφων με ομοιόμορφη φωτεινότητα υποβάθρου, το ελάχιστο κατώφλι  $d$  μεταξύ των pixels κειμένου και υποβάθρου μπορεί να οριστεί με επιτυχία σαν  $q \cdot \delta$  όπου  $q$  μεταβλητή κοντά στο 0.8 η οποία βοηθάει την διάσωση του συνολικού σώματος των χαρακτήρων ώστε να έχουμε βέλτιστα OCR αποτελέσματα (Seeger 2001). Στην περίπτωση των παλιών εγγράφων με χαμηλή ποιότητα και μη ομοιόμορφη κατανομή φωτεινότητας θέλουμε να έχουμε μικρότερη τιμή για το κατώφλι  $d$  για τις πιο σκούρες περιοχές υποβάθρου αφού είναι συνηθισμένη περίπτωση σε τέτοιες εικόνες να έχουμε κείμενο σε σκούρες περιοχές υποβάθρου όπου η διαφορά στην φωτεινότητα μεταξύ κειμένου και υποβάθρου είναι μικρή. Για να το πετύχουμε αυτό, πρώτα υπολογίζουμε την μέση τιμή  $b$  των σημείων της επιφάνειας του υποβάθρου  $B$  που αντιστοιχούν στις περιοχές κειμένου της εικόνας  $N$ :

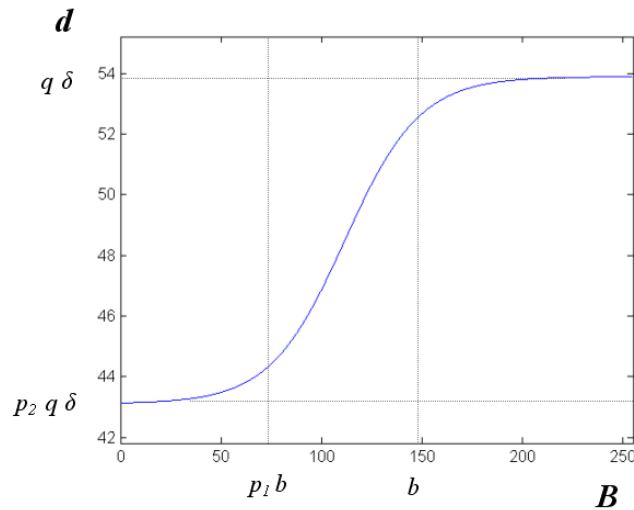
$$b = \frac{\sum_x \sum_y (B(x, y)(1 - N(x, y)))}{\sum_x \sum_y (1 - N(x, y))} \quad (2.22)$$

## Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Αυτό που θέλουμε να πετύχουμε είναι το κατώφλι να είναι περίπου ίσο με  $q \cdot \delta$  όταν η τιμή του υποβάθρου είναι μεγάλη (περίπου μεγαλύτερη από την μέση τιμή υποβάθρου  $b$ ) και περίπου ίσο με  $p_2 \cdot q \cdot \delta$  όταν η τιμή του υποβάθρου είναι μικρή (περίπου μικρότερη από μικρότερο από  $p_1 \cdot b$ ) με  $p_1, p_2 \in [0,1]$ . Για να εξομοιώσουμε την επιθυμητή συμπεριφορά του κατωφλίου, χρησιμοποιούμε την παρακάτω λογαριθμική σιγμοειδή συνάρτηση η οποία παρουσιάζει την επιθυμητή οριακή συμπεριφορά για μεγάλες και μικρές τιμές του υποβάθρου (σχήμα 2.15):

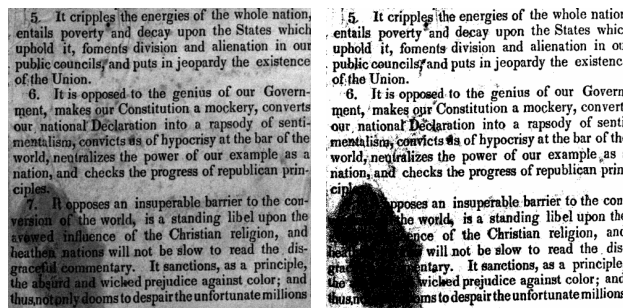
$$d(B(x, y)) = q \delta \left( \frac{(1 - p_2)}{1 + \exp\left(\frac{-4 B(x, y)}{b(1 - p_1)} + \frac{2(1 + p_1)}{(1 - p_1)}\right)} + p_2 \right) \quad (2.23)$$

Μετά από πειραματικά αποτελέσματα, για την περίπτωση χαμηλής ποιότητας ιστορικών εγγράφων, προτείνονται οι παρακάτω παράμετροι:  $q = 0.6$ ,  $p_1 = 0.5$ ,  $p_2 = 0.8$ .



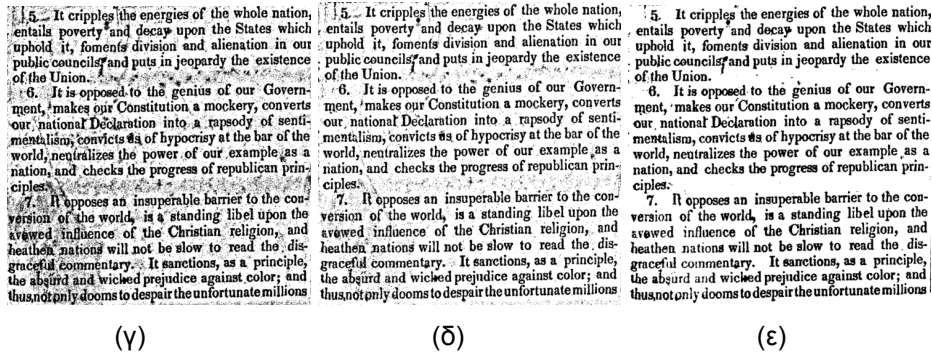
**Σχήμα 2.15.** Η συνάρτηση  $d(B(x, y))$ .

Στο σχήμα 2.16 φαίνονται τα συγκριτικά αποτελέσματα δυαδικής μετατροπής για εικόνα χαμηλής ποιότητας χρησιμοποιώντας μία μέθοδο καθολικής κατωφλίωσης (Otsu), τις προσαρμοστικές μεθόδους με χρήση παραθύρου (Niblack, Sauvola) και την μέθοδο που βασίζεται στον εντοπισμό του υποβάθρου.



(α)

(β)



**Σχήμα 2.16.** Αποτελέσματα δυαδικής μετατροπής. (α) Αρχική εικόνα. (β) Με χρήση της μεθόδου καθολικής κατωφλίωσης (Otsu). (γ) Με χρήση της μεθόδου του Niblack. (δ) Με χρήση της μεθόδου του Sauvola. (ε) Με χρήση της μεθόδου που βασίζεται στον εντοπισμό του υποβάθρου.

## Αποτίμηση Μεθοδολογιών Δυαδικής Μετατροπής

Η αποτίμηση των μεθοδολογιών δυαδικής μετατροπής συνήθως βασίζεται στην σύγκριση της ιδανικής ασπρόμαυρης εικόνας με το δυαδικό αποτέλεσμα. Η ιδανική ασπρόμαυρη εικόνα (Ground Truth - GT) παράγεται με μία ημιαυτόματη διαδικασία η οποία βασίζεται στην βοήθεια του χρήστη (Ntirogiannis 2008). Καταρχήν γίνεται εντοπισμός των True Positive (TP), False Positive (FP) και False Negative (FN) στοιχειοσημείων και στη συνέχεια γίνεται υπολογισμός της μετρικής της Ανάκτησης (Recall) και της Ακρίβειας (Precision).

Ένα στοιχειοσημείο καταγράφεται ως True Positive (TP) αν είναι ON και στην ιδανική ασπρόμαυρη εικόνα (GT) και στο αποτέλεσμα της δυαδικής μετατροπής.

Ένα στοιχειοσημείο καταγράφεται ως False Positive (FP) αν είναι ON μόνο στο αποτέλεσμα της δυαδικής μετατροπής.

Ένα στοιχειοσημείο καταγράφεται ως False Negative (FN) αν είναι ON μόνο στην ιδανική ασπρόμαυρη εικόνα (GT).

Αν  $C_{TP}$  είναι ο αριθμός των TP στοιχειοσημείων,  $C_{FP}$  ο αριθμός των FP στοιχειοσημείων και  $C_{FN}$  ο αριθμός των FN στοιχειοσημείων, η Ανάκτηση (RC) και η Ακρίβεια (PR) υπολογίζονται ως εξής:

$$RC = C_{TP} / (C_{FN} + C_{TP}) \quad (2.24)$$

$$PR = C_{TP} / (C_{FP} + C_{TP})$$

Μία συνολική μετρική είναι το F-Measure (FM) η οποία υπολογίζεται ως εξής:

$$FM = (2 \times RC \times PR / (RC + PR)) \times 100 \% \quad (2.25)$$



Παράδειγμα:

Ground Truth									

Binarization Result									

Evaluation Result									
		TP	TP	TP	TP				
	TP					TP			
	TP		FP	FP		TP	FP	TP	
FP	TP		FP	FP		TP	FP	TP	
	TP					TP	FP	TP	
		TP	TP			TP	TP		
			TP			TP			FN
FP	FP		TP	TP	TP			FN	FN
TP	TP				TP			FN	FN
	FN	FN			FP				FN

$C_{TP} = 27$ ,  $C_{FP} = 11$ ,  $C_{FN} = 8$ ,  $RC = .77$ ,  $PR = .71$ ,  $FM = 73.9\%$

## Βιβλιογραφία

- (Chow 1972)** Chow, C.K. and Kaneko, T.: Boundary detection of radiographic images by a threshold method. In Proceedings IFIP Congress - 71, Booklet - TA-7, (1972) 130--134
- (Gatos 2004)** Gatos, B., Pratikakis, I. and Perantonis, S. J.: Locating Text in Historical Collection Manuscripts. 3rd Hellenic Conference on Artificial Intelligence, Samos, Greece (2004)
- (Huang 1995)** Huang, L. K. and Wang, M. J.: Image Thresholding by Minimizing the Measures of Fuzziness. Pattern Recognition, Vol. 28 (1995) 41-51
- (Kamel 1993)** Kamel M, Zhao A.: Extraction of Binary Character/Graphics Images from Grayscale Document Images. Computer Vision, Graphics and Image Processing 55 (1993) 203-217
- (Kapur 1985)** Kapur, J. N. ,Sahoo, P. K. and Wong, A. K. C.: A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. Computer Vision, Graphics and Image Processing 29 (1985) 273-285
- (Niblack 1986)** Niblack, W.: An Introduction to Digital Image Processing. Englewood Cliffs, N. J., Prentice Hall (1986) 115-116
- (Otsu 1979)** Otsu, N.: A Threshold Selection Method from Gray-level Histograms. IEEE Trans. On Systems, Man and Cybernetics, Vol. 9, 1 (1979) 377-393
- (Pun 1981)** Pun, T.: Entropic Thresholding: A New Approach. CVGIP: Graphical Models and Image Processing 16, (1981) 210--239.
- (Ridler 1978)** Ridler, T. W. and Calvard, S.: Picture Thresholding Using an Iterative Selection Method. IEEE Trans. On Systems, Man and Cybernetics, Vol. SMC-8 (1978) 630-632
- (Sauvola 2000)** Sauvola, J. and Pietikainen: Adaptive document image binarization. Pattern Recognition 33 (2000) 225-236
- (Yanowitz 1989)** Yanowitz, S.D. and Bruckstein A.M.: A New method for Image Segmentation, Computer Vision, Graphics and Image Processing 46 (1989) 82-95
- (Weszka 1974)** Weszka, J.S., Nagel, R.N., Rosenfeld, A.: A threshold selection technique. IEEE Trans. on Computers, 23 (1974) 1322-1326
- (Ntirogiannis 2008)** K. Ntirogiannis, B. Gatos and I. Pratikakis, "An Objective Evaluation Methodology for Document Image Binarization Techniques", 8th International Workshop on Document Analysis Systems (DAS'08), pp. 217-224, Nara, Japan, September 2008.