

Μάθημα 5

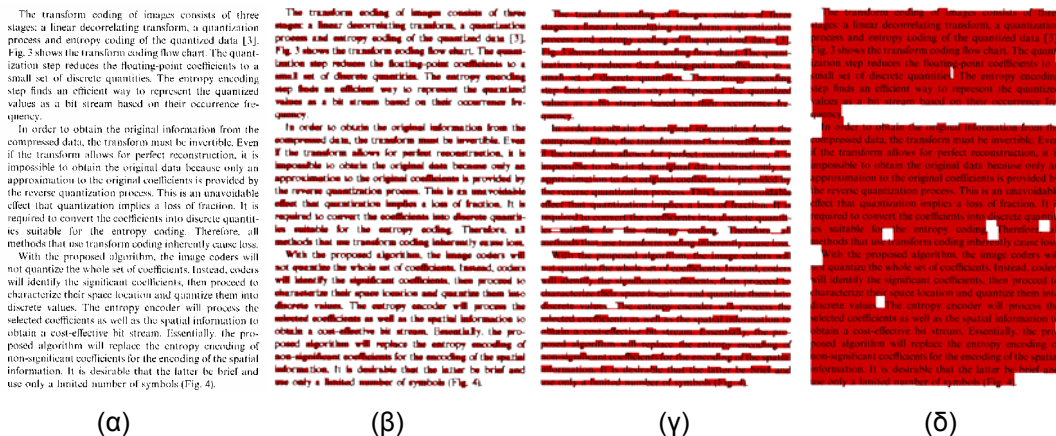
Κατάτμηση σελίδας εγγράφου

5.1 Το στάδιο της κατάτμησης της σελίδας του εγγράφου

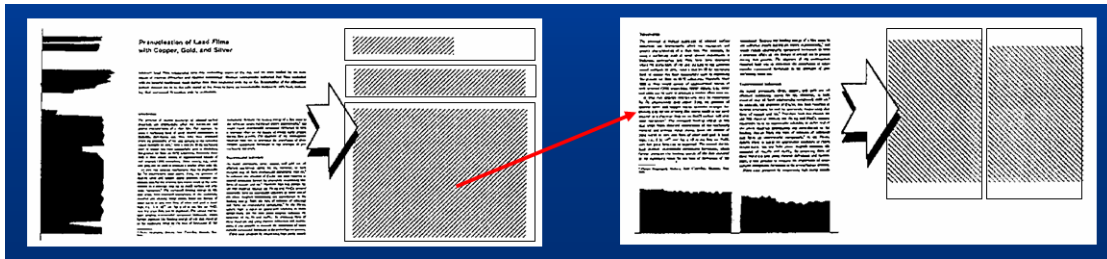
Το στάδιο της κατάτμησης της σελίδας των εγγράφων είναι από τα πιο σημαντικά στάδια στην επεξεργασία και κατανόηση των εγγράφων. Αφορά την διαδικασία κατάτμησης της εικόνας του εγγράφου σε ομογενείς περιοχές, δηλαδή σε περιοχές που περιέχουν ένα είδος πληροφορίας όπως κείμενο, εικόνες, γραμμές, γραφικά κ.λ.π. Επιπροσθέτως, περιλαμβάνει την εξαγωγή πληροφοριών σχετικά με την δομή του εγγράφου. Πιο συγκεκριμένα, τα επιμέρους βήματα της κατάτμησης της σελίδας του εγγράφου είναι τα ακόλουθα:

- Κατάτμηση της εικόνας σε επιμέρους ομογενή τμήματα.
- Κατηγοριοποίηση των τμημάτων ανάλογα με τον τύπο τους (κείμενο, εικόνα, γραφικό, γραμμή κ.λ.π.).
- Ονοματισμός των τμημάτων ανάλογα με τα ειδικά χαρακτηριστικά του εγγράφου (Τίτλος, υπότιτλος, λεζάντα, λογότυπο κ.λ.π.).
- Λογική ομαδοποίηση βάσει κανόνων που εξαρτώνται από την φύση του εγγράφου (π.χ. άρθρα).
- Ορισμός της σειράς ανάγνωσης (ροής κειμένου).

Οι προσεγγίσεις για την κατάτμηση της σελίδας του εγγράφου χωρίζονται στις «από κάτω προς τα πάνω» (Bottom-up) και στις «από πάνω προς τα κάτω» (Top-down). Σύμφωνα με τις «από κάτω προς τα πάνω» προσεγγίσεις, ξεκινώντας με τα συνδεδεμένα συστατικά της εικόνας, τα ενώνουμε για να σχηματίσουν χαρακτήρες, λέξεις, γραμμές κειμένου, παραγράφους κ.λ.π. Αντίθετα, σύμφωνα με τις από «πάνω προς τα κάτω» προσεγγίσεις, μια σελίδα χωρίζεται σε στήλες, μια στήλη σε παραγράφους, μια παράγραφος σε γραμμές κειμένου κ.λ.π. Ένα παράδειγμα «από κάτω προς τα πάνω» προσέγγισης είναι η σταδιακή εξομάλυνση της εικόνας ώστε να εντοπίσουμε διαδοχικά τις λέξεις, τις γραμμές και τέλος τα συνολικά τμήματα του κειμένου (σχήμα 5.1). Αντίστοιχα, ένα παράδειγμα «από πάνω προς τα κάτω» προσέγγισης είναι η χρήση των προβολών για τον εντοπισμό των οριζόντιων τμημάτων κειμένου, στην συνέχεια των στηλών, των παραγράφων κ.λ.π. (σχήμα 5.2).



Σχήμα 5.1. Σταδιακή εξομάλυνση της εικόνας (α) ώστε να εντοπίσουμε διαδοχικά τις λέξεις (β), τις γραμμές (γ) και τέλος τα συνολικά τμήματα του κειμένου (δ).



Σχήμα 5.2. Χρήση των προβολών για τον εντοπισμό των οριζόντιων τμημάτων κειμένου, στην συνέχεια των στηλών κ.λ.π.

Η ρύθμιση των παραμέτρων των αλγορίθμων κατάτμησης της εικόνας πρέπει να γίνεται έτσι ώστε η εφαρμογή των αλγορίθμων να είναι ανεξάρτητη από την ανάλυση της εικόνας και το μέγεθος των χαρακτήρων του εγγράφου. Για το λόγο αυτό υπολογίζουμε το επικρατέστερο ύψος του γράμματος του εγγράφου.

5.2 Υπολογισμός του ύψους του γράμματος

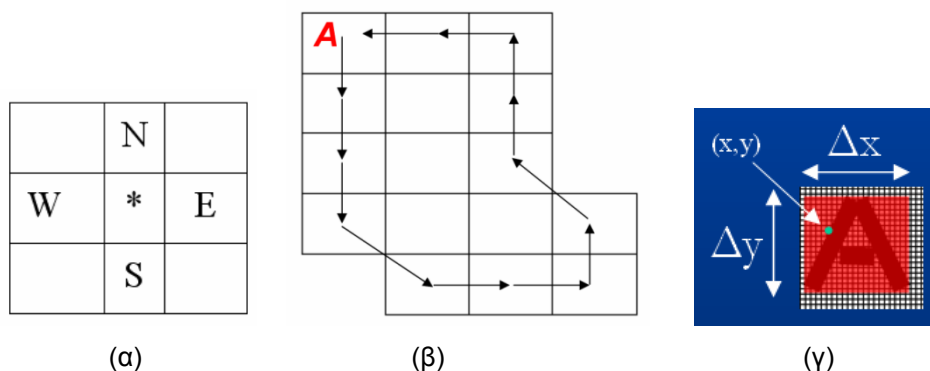
Για τον υπολογισμό του επικρατέστερου ύψους του γράμματος ενός εγγράφου, μπορούμε να χρησιμοποιήσουμε τον παρακάτω αλγόριθμο ο οποίος βασίζεται στην ακολούθηση του περιγράμματος τυχαίων συνδεδεμένων συστατικών της εικόνας:

- **BHMA 1:** Βρίσκω ένα τυχαίο σημείο της εικόνας (x,y) το οποίο έχει ένα τουλάχιστον σημείο υποβάθρου στην γειτονία των 4 pixels (σχήμα 5.3α).
- **BHMA 2:** Ξεκινώντας από το σημείο (x,y) ακολουθώ το περίγραμμα του συνδεδεμένου συστατικού στο οποίο ανήκει το σημείο (x,y) (σχήμα 5.3β).
- **BHMA 3:** Επαναλαμβάνω τα βήματα 1 & 2 και υπολογίζω το ιστογράμμο *Hist* με τα ύψη των ορθογώνιων παραλληλόγραμμων που περικλείουν τα συνδεδεμένα συστατικά στα οποία ανήκουν τα σημεία (x,y) μέχρι να έχω MAX_SAMPLES δείγματα:

$$Hist(\Delta y) = Hist(\Delta y) + 1 \quad (5.1)$$

όπου Δy είναι το ύψος του ορθογώνιου παραλληλόγραμμου που περικλείει το συνδεδεμένο συστατικό (σχήμα 5.3γ).

- **BHMA 4:** Βρίσκω την τιμή Δy που αντιστοιχεί στο μέγιστο του ιστογράμματος που δημιουργήθηκε.



Σχήμα 5.3. (α) Γειτονία των 4 Pixels. (β) Ακολούθηση του περιγράμματος ξεκινώντας από το σημείο $A(x,y)$. (γ) Το ορθογώνιο παραλληλόγραμμο $\Delta x \times \Delta y$ που περικλείει το συνδεδεμένο συστατικό.

Ο παραπάνω αλγόριθμος βελτιώνεται αν υπολογίζω το εξομαλισμένο ιστογράμμο των υψών προσθέτοντας κάθε φορά:

$$Hist(\Delta y - 1) = Hist(\Delta y - 1) + 1$$

$$Hist(\Delta y) = Hist(\Delta y) + 2 \tag{5.2}$$

$$Hist(\Delta y + 1) = Hist(\Delta y + 1) + 1$$

Αποτελέσματα της παραπάνω μεθοδολογίας για τον υπολογισμό του επικρατέστερου ύψους γράμματος δίνονται στο σχήμα 5.4.



Fig. 2. Tapered wavelet decomposition of the Lena image.

(α)

information. Next, the filters H and G are applied to the columns of both subimages. Final dimensions $n/2 \times n/2$ are obtained. The same procedure is applied iteratively to the subimage containing the most low-band information until the subimage's size reaches 1×1 . Therefore, the total dimensions of the image are required to be powers of two. In practice, it is not necessary to carry out all the possible decompositions until the size of 1×1 is reached. Usually, just a few levels are sufficient. Fig. 3 shows the "Lena" image decomposed in two levels. Each of the resulting subimages is known as a subband.

3. The transform coding process

The transform coding of images consists of three stages: a linear decomposing procedure, a quantization process and entropy coding of the quantized data [7]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy coding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the inverse quantization process. This is an unavoidable effect that quantization implies a loss of features. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image codes will not quantify the whole set of coefficients. Instead, codes will identify the significant coefficients, then proceed to characterize their exact locations and quantize their associated values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a non-redundant bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be used and use only a limited number of symbols (Fig. 4).

4. Significant coefficients identification

4.1. Applying the DWT

The DWT is applied to an image, and the number of wavelet decomposition levels depends on the dimensions of the image. In our experiments, most images are decomposed into four levels, yielding 16 subbands. Without loss of generality, Fig. 3 shows the numbering scheme for the decomposition of two levels. It is desirable that the relative subband be about $1/4 \times 1/4$. This is because DWT will have most of the energy content of the higher scales, meaning that all coefficients in each subband are more likely to be significant and therefore thresholding will have an effect in such scales.

Any type of wavelets can be used in the decomposition. However, for better results a set of wavelets is preferable. Wavelets such as the Daubechies are better choices than non-orthogonal wavelets such as the Haar. In general, smooth wavelets produce better reconstructions of the original image from a lower rate. Good sets of original coefficients. In the following sections, we adopt

In practice, it is not necessary possible decompositions until the Usuallv, just a few levels are suff

(β)

Τα μυστικά του scanner

Ψηφιοποιήστε σωστά κάθε εικόνα με τη χρήση του scanner Μέρος Α'

Πολλά φορές οι Υ.Γ. διαθέτουν scanner με τα οποία μπορούν να ψηφιοποιούν εικόνας εικόνας και γρήγορα. Ωστόσο, στη διάρκεια μιας σύντομης μέτρησης του scanner, η ποιότητα των εικόνων που παράγονται μπορεί να ποικίλλει σημαντικά από ένα μέτρο διαφάνειας από τον άλλο, όπως η εκτύπωση σε κίτρινο στο τυπωμένο ή σε κίτρινο εκτύπωση, υπάρχουν κίτρινα αποτυπώματα. Χρειάζεται, εκτός από ένα καλό scanner, μία σωστή μέθοδος επεξεργασίας, τους δε να πραγματοποιηθεί επεξεργασία εικόνας.

Το τελικό αποτέλεσμα

Το τελικό αποτέλεσμα είναι η εικόνα που έχει ψηφιοποιηθεί σωστά. Η εικόνα που έχει ψηφιοποιηθεί σωστά είναι η εικόνα που έχει ψηφιοποιηθεί σωστά. Η εικόνα που έχει ψηφιοποιηθεί σωστά είναι η εικόνα που έχει ψηφιοποιηθεί σωστά.



Bytes, bits και dpi... χρήση

Όσοι θέλουν να χρησιμοποιήσουν τον υπολογιστή τους για να ψηφιοποιήσουν εικόνας, πρέπει να γνωρίζουν ότι η ποιότητα των εικόνων που παράγονται εξαρτάται από τον αριθμό των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα. Ο αριθμός των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα είναι ο αριθμός των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα.

Η ποιότητα των εικόνων που παράγονται εξαρτάται από τον αριθμό των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα. Ο αριθμός των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα είναι ο αριθμός των bits που χρησιμοποιούνται για να ψηφιοποιηθεί η εικόνα.

Αριθμός γραμμών ανά ίντσο (DPI)

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο. Ο αριθμός γραμμών ανά ίντσο (DPI) είναι ο αριθμός των γραμμών που υπάρχουν σε ένα ίντσο.

(γ)

επεξεργασία εικόνας.

Το τελικό αποτέλεσμα της α

εικόνας επηρεάζουν ορισμένοι βασικοί

• Η ποιότητα του scanner. Αν δεν είναι ευαίσθητο το μηχανήματος, δε θα χρειαστεί και η φωτεινότητα της εικόνας της.

• Η ποιότητα της αρχικής εικόνας. Αν contrast ή είναι υπερφωτισμένη [εικόνες διορθωτικές κινήσεις, είτε κατά τη ρωσής είτε μετέπειτα, σε κάποιο πρόσας εικόνας.

Αναφερθούμε, όμως, σε μερικές τεχν

(δ)

Σχήμα 5.4. Υπολογισμός του επικρατέστερου ύψους γράμματος. (α),(γ) Εικόνες εγγράφου. (β),(δ) Εντοπισμένο επικρατέστερο ύψους γράμματος.

5.3 Εξομάλυνση εικόνας

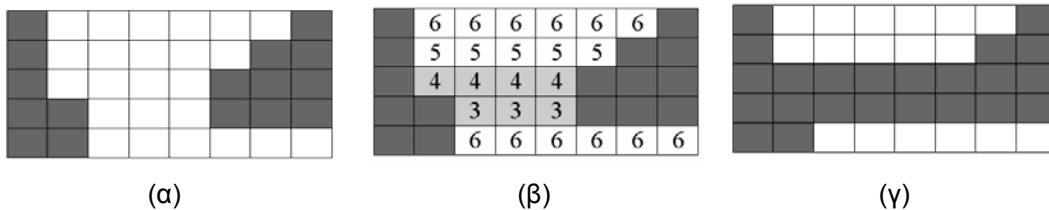
Οι τεχνικές εξομάλυνσης της εικόνας σκοπεύουν στο γέμισμα περιοχών του υποβάθρου ώστε να ενωθούν τμήματα της εικόνας που ανήκουν στην ίδια κατηγορία. Η πιο γνωστή τεχνική εξομάλυνσης είναι ο RLSA (Run Length Smoothing Algorithm – Αλγόριθμος εξομάλυνσης διαδοχικών rixels) (Wahl 1982). Για τον ορισμό του RLSA χρειάζονται να οριστούν δύο βασικοί παράμετροι: το μέγιστο μήκος διαδοχικών rixels υποβάθρου T_{max} και η διεύθυνση σάρωσης. Η εικόνα εξετάζεται ως προς την διεύθυνση σάρωσης που έχει οριστεί και τα διαδοχικά rixels υποβάθρου με μήκος μικρότερο από το μέγιστο μήκος διαδοχικών rixels υποβάθρου T_{max} μετατρέπονται σε σημεία εικόνας. Το όριο αυτό υπολογίζεται από στατιστικές πληροφορίες της εικόνας (Paramarkos 1996) και ρυθμίζεται ανάλογα με την εξομάλυνση που θέλουμε να πετύχουμε. Για παράδειγμα, εάν έχουμε επιλέξει την οριζόντια κατεύθυνση και $T_{max} = 4$, τότε η γραμμή:

11000111110000001001

γίνεται:

11111111110000001111

Ένα παράδειγμα εφαρμογής οριζόντιου RLSA σε ασπρόμαυρη εικόνα δίδεται στο σχήμα 5.5.



Σχήμα 5.5. Εφαρμογή οριζόντιου RLSA με $T_{max} = 5$ σε ασπρόμαυρη εικόνα. (α) Αρχική εικόνα. (β) Υπολογισμός του μήκους των διαδοχικών rixels υποβάθρου. (γ) Μετατροπή όλων των σημείων υποβάθρου με μήκος μικρότερο από T_{max} σε σημεία εικόνας.

Αναλυτικά όλες οι παράμετροι για το RLSA έχουν ως εξής:

- Διεύθυνση σάρωσης: Μπορεί να είναι οριζόντια, κάθετη ή διαγώνια. Ο ορισμός της διεύθυνσης είναι υποχρεωτικός για τον ορισμό του RLSA. Αν και έχει προταθεί και η χρήση της διαγώνιας διεύθυνσης (Waked 2001), συνήθως επιλέγουμε οριζόντια ή κάθετη διεύθυνση σάρωσης για την εξομάλυνση των εικόνων εγγράφων.
- Όρια μήκους: Το μέγιστο μήκος διαδοχικών rixels υποβάθρου T_{max} και το ελάχιστο μήκος διαδοχικών rixels υποβάθρου T_{min} . Η παράμετρος T_{max} είναι υποχρεωτική για τον ορισμό του RLSA ενώ η παράμετρος T_{min} μπορεί να παραλείπεται (τότε εννοείται $T_{min} = 0$). Αν έχουμε υπολογίσει ότι το μήκος των διαδοχικών σημείων υποβάθρου είναι L , τότε αυτά μετατρέπονται σε σημεία εικόνας εάν ισχύει $T_{min} < L < T_{max}$. Για παράδειγμα, εάν έχουμε επιλέξει την οριζόντια διεύθυνση και $T_{min} = 2$, $T_{max} = 4$, τότε η γραμμή:

11000111110000001001

γίνεται:

11111111110000001001

- Γεινίαση με το περιθώριο: Μπορούμε να επιλέξουμε εάν θα μετατρέψουμε σε σημεία εικόνας τα διαδοχικά σημεία υποβάθρου που πληρούν την συνθήκη για τα όρια του μήκους και ταυτόχρονα συνορεύουν με τα όρια της εικόνας. Για παράδειγμα, εάν έχουμε επιλέξει την οριζόντια διεύθυνση με $T_{max} = 4$ και χωρίς γεινίαση με το περιθώριο της εικόνας, η γραμμή:

00011100011100010

γίνεται:

00011111111111110

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

(α)

The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

(γ)

The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

(β)

The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

(δ)

Σχήμα 5.6. Εξομάλυνση εικόνας με χρήση RLSA. (α) Αρχική εικόνα. (β) Εντοπισμός blocks κειμένου (Οριζόντιο RLSA με $T_{\max} = 24$ και κάθετο RLSA με $T_{\max} = 24$). (γ) Εντοπισμός γραμμών (Οριζόντιο RLSA με $T_{\max} = 24$ και κάθετο RLSA με $T_{\max} = 6$). (δ) Εντοπισμός λέξεων (Οριζόντιο RLSA με $T_{\max} = 6$ και κάθετο RLSA με $T_{\max} = 6$).

Στο σχήμα 5.6 δίδεται ένα παράδειγμα διαδοχικής εφαρμογής οριζόντιου και κάθετου RLSA με σκοπό τον εντοπισμό λέξεων, γραμμών και blocks κειμένου. Στο σχήμα 5.7 χρησιμοποιείται η διαδοχική εφαρμογή οριζόντιου και κάθετου RLSA για τον εντοπισμό των

διαφόρων περιοχών της εικόνας. Για τις εικόνες των σχημάτων 5.6 και 5.7 το στατιστικά υπολογισμένο ύψος του γράμματος είναι 12.



Fig. 2. Two-level wavelet decomposition of the Lena image.

3. The transform coding process

The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent their quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

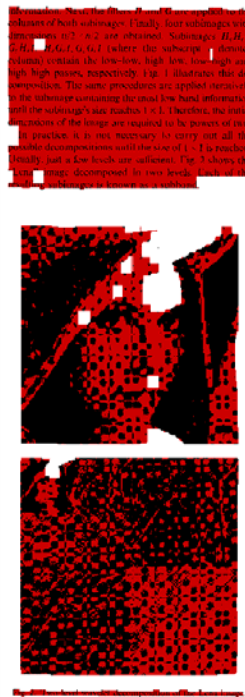
With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

4. Significant coefficients identification

4.1. Applying the DWT

The DWT is applied to an image, and the number of wavelet decomposition depends on the dimensions of the image. In our experiments, most images are decomposed into five levels, yielding 16 subbands. Without loss of generality, Fig. 5 shows the numbering scheme for the decomposition of two levels. It is desirable that the smallest subband be about 32×32 pixels. This is because DWT will have most of the energy compact at the higher scales, meaning that all coefficients in such subbands are most likely to be significant and therefore thresholding will have no effect in such scales.

Any type of wavelets can be used in the decomposition. However, for better results a smooth wavelet is preferable. Wavelets such as the B-spline are better choices than non-continuous wavelets such as the Haar. In general, smooth wavelets produce better reconstruction of the original image from a fewer (or modified) set of original coefficients. In the following sections, we adopt



The transform coding of images consists of three stages: a linear decorrelating transform, a quantization process and entropy coding of the quantized data [3]. Fig. 3 shows the transform coding flow chart. The quantization step reduces the floating-point coefficients to a small set of discrete quantities. The entropy encoding step finds an efficient way to represent the quantized values as a bit stream based on their occurrence frequency.

In order to obtain the original information from the compressed data, the transform must be invertible. Even if the transform allows for perfect reconstruction, it is impossible to obtain the original data because only an approximation to the original coefficients is provided by the reverse quantization process. This is an unavoidable effect that quantization implies a loss of fraction. It is required to convert the coefficients into discrete quantities suitable for the entropy coding. Therefore, all methods that use transform coding inherently cause loss.

With the proposed algorithm, the image coders will not quantize the whole set of coefficients. Instead, coders will identify the significant coefficients, then proceed to characterize their space location and quantize them into discrete values. The entropy encoder will process the selected coefficients as well as the spatial information to obtain a cost-effective bit stream. Essentially, the proposed algorithm will replace the entropy encoding of non-significant coefficients for the encoding of the spatial information. It is desirable that the latter be brief and use only a limited number of symbols (Fig. 4).

4. Significant coefficients identification

4.1. Applying the DWT

The DWT is applied to an image, and the number of wavelet decomposition depends on the dimensions of the image. In our experiments, most images are decomposed into five levels, yielding 16 subbands. Without loss of generality, Fig. 5 shows the numbering scheme for the decomposition of two levels. It is desirable that the smallest subband be about 32×32 pixels. This is because DWT will have most of the energy compact at the higher scales, meaning that all coefficients in such subbands are most likely to be significant and therefore thresholding will have no effect in such scales.

Any type of wavelets can be used in the decomposition. However, for better results a smooth wavelet is preferable. Wavelets such as the B-spline are better choices than non-continuous wavelets such as the Haar. In general, smooth wavelets produce better reconstruction of the original image from a fewer (or modified) set of original coefficients. In the following sections, we adopt

(α)

(β)

Τα μυστικά του scanner

Ψηφιοποιήστε σωστά κάθε εικόνα με τη χρήση του scanner Μέρος Α'

Πολλοί χρήστες ΗΥ διαθέτουν scanners με τα οποία μπορούν να ψηφιοποιούν εικόνες εύκολα και γρήγορα. Όμως, από τη στιγμή μιας εικόνας μέχρι την αναπαραγωγή της με επαγγελματική ποιότητα, δηλαδή τη μεταφορά της σε ένα μέσο διαφορετικό από τον οθόνι, όπως η εκτύπωση σε χαρτί στο τυπογραφείο ή σε έγγραφο εκπαιδευτή, υπάρχει κάποια απόσταση. Χρησιάζεται, εκτός από ένα καλό scanner, μια σωστή μεθοδολογία, ίσως δε και περσιτάωτο επεξεργασία εικόνας.

των Π. Αθανάσιου, Β. Γατάτου

Το τελικό αποτέλεσμα είναι η αναπαραγωγή μιας εικόνας επηρεάζοντας αρνητικά βλαπτικά παράγοντες, όπως:

- **Η ποιότητα του scanner.** Αν δεν είναι ικανοποιητική η ευαισθησία του μηχανήματος, δε θα "δείτε" σωστά τη κρούση και τη φωτεινότητα της εικόνας στο διάφορο σημείο της.
- **Η ποιότητα της αρχικής εικόνας.** Αν η εικόνα δεν έχει contrast ή είναι υπερφωτισμένη [εικ. 1], πρέπει να γίνει διαβαθμισμένη κλίμακα, είτε κατά τη διαδικασία της οργάνωσης είτε μετέπειτα, σε κάποιο πρόγραμμα επεξεργασίας εικόνας.



Βλέπεται σε κάποιο πρόγραμμα όπως το Photoshop. Το scanner [στην εικόνα 3] θέλει να μην αρχίσει τη ψηφιοποίηση της εικόνας μέχρι να είναι σωστά την αρχή βλαπτικά παράγοντες, κίνηση, πρόβλημα, μπλέκ, Red, Green, Blue, ε, εκ του RGB. Με τεχνική παράρτηση κίνησης που οι οθόνες χρησιμοποιούν για να αναπαράξουν μια εικόνα από pixels που παίρνουν χρωματικά τιμές στο χώρο από κρούση. Γι αυτό και όταν οργάνωση μια εικόνα, το αρχικό που αποθηκεύεται από το scanner είτε RGB [εκτός αν έχουμε ρυθμίσει το software να το μετατρέψει σε CMYK, δηλαδή σε τέσσερα κλάδια που αναπαράξουν στο χρώμα της τρέφω κρούσης επίπτωσης offset (C=κυανό, M=μαζέλι, Y=κίτρινο, K=μαύρο)]. Πάντα, όμως, ο scanner "δηλώνει" αρχικά μια εικόνα ως RGB. Μετατρέποντας τον RGB αρχικά σε CMYK επιβεβαιώνεται από pixel (1 byte ή 8 bits για κάθε κρούση, κίνηση, πρόβλημα, μπλέκ, ε, συνολικά 24bits). Οι εικόνες CMYK έχουν 4bytes πληροφορία για κάθε pixel λόγω των τεσσάρων κλάδων.

Bytes, bits και άλλα... χρήσιμα
 Αναπαράξουν πληροφορία (π.χ. μια εικόνα κίνηση ή ένα αρχείο που) μπορεί να περιγραφεί με μια δεδομένη γραμμή 0 και 1. Με βάση το **δεδομένο σύστημα** (σύστημα κρούσης) κλάση από το οποίο κλάση bit. Μια εικόνα δε από ψηφιακή κρούση RGB. Ο υπολογιστής μπορεί να "περιγράψει" οποιαδήποτε πληροφορία με 8 bits κίνηση, σε συνδυασμό των οποίων είναι $2^8=256$. Με άλλα λόγια, ο υπολογιστής είναι σε θέση να διαβαθμίσει συνολικά σε 256 διαβαθμίσεις τιμής. Σε αναπαράξουν επεξεργασία εικόνας οι αρχές αυτές θα μπορούσαν να αναπαράξουν σε 256 διαβαθμίσεις φωτεινότητας για κάθε pixel μιας εικόνας. Ένα byte (8bits) είναι η πιο κοινή μεθοδολογία της ψηφιακής ενός υπολογιστή. As παράρτηση ως παράρτηση μια ψηφιακή εικόνα που έχει πληροφορία μόνο για έναν του pixel, δηλαδή μονοχρωματική πληροφορία [εικ. 2]. Η πιο παλαιά εκδοχή είναι κλάση pixel της εικόνας να καταγράφεται 1 byte (8bits). Έτσι, όπως εξηγήσαμε προηγουμένως, κάθε pixel της εικόνας μπορεί να πάρει μια από 256 διαφορετικές τιμές τόνου διαβαθμίσεων. Στην τιμή 0 αντιστοιχεί το τελείως μαύρο (από 0 στα 255 είναι τόνος) και στην τιμή 255 το τελείως λευκό. Φανταστείτε μια κρούση 256 τόνων διαβαθμίσεων, η οποία ξεκινά από το τελείως μαύρο (την 0) και καταλήγει στο τελείως λευκό (την 255) [εικ. 2]. Οι διαβαθμίσεις αυτές αναπαράξουν με διαβαθμίσεις του ψηφ (σε μικρό αριθμό αντιστοιχούν σε ορισμένες διαβαθμίσεις και οι μεγάλοι σε φωτεινές). Στο 0, λοιπόν, θα αντιστοιχούσε η τιμή του "bit image" στην τιμή 0.

(γ)

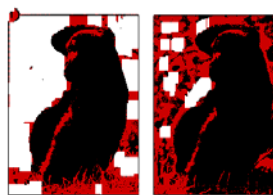
Τα μυστικά του scanner

Ψηφιοποιήστε σωστά κάθε εικόνα με τη χρήση του scanner Μέρος Α'

Πολλοί χρήστες ΗΥ διαθέτουν scanners με τα οποία μπορούν να ψηφιοποιούν εικόνες εύκολα και γρήγορα. Όμως, από τη στιγμή μιας εικόνας μέχρι την αναπαραγωγή της με επαγγελματική ποιότητα, δηλαδή τη μεταφορά της σε ένα μέσο διαφορετικό από τον οθόνι, όπως η εκτύπωση σε χαρτί στο τυπογραφείο ή σε έγγραφο εκπαιδευτή, υπάρχει κάποια απόσταση. Χρησιάζεται, εκτός από ένα καλό scanner, μια σωστή μεθοδολογία, ίσως δε και περσιτάωτο επεξεργασία εικόνας.

Το τελικό αποτέλεσμα είναι η αναπαραγωγή μιας εικόνας επηρεάζοντας αρνητικά βλαπτικά παράγοντες, όπως:

- **Η ποιότητα του scanner.** Αν δεν είναι ικανοποιητική η ευαισθησία του μηχανήματος, δε θα "δείτε" σωστά τη κρούση και τη φωτεινότητα της εικόνας στο διάφορο σημείο της.
- **Η ποιότητα της αρχικής εικόνας.** Αν η εικόνα δεν έχει contrast ή είναι υπερφωτισμένη [εικ. 1], πρέπει να γίνει διαβαθμισμένη κλίμακα, είτε κατά τη διαδικασία της οργάνωσης είτε μετέπειτα, σε κάποιο πρόγραμμα επεξεργασίας εικόνας.



Βλέπεται σε κάποιο πρόγραμμα όπως το Photoshop. Το scanner [στην εικόνα 3] θέλει να μην αρχίσει τη ψηφιοποίηση της εικόνας μέχρι να είναι σωστά την αρχή βλαπτικά παράγοντες, κίνηση, πρόβλημα, μπλέκ, Red, Green, Blue, ε, εκ του RGB. Με τεχνική παράρτηση κίνησης που οι οθόνες χρησιμοποιούν για να αναπαράξουν μια εικόνα από pixels που παίρνουν χρωματικά τιμές στο χώρο από κρούση. Γι αυτό και όταν οργάνωση μια εικόνα, το αρχικό που αποθηκεύεται από το scanner είτε RGB [εκτός αν έχουμε ρυθμίσει το software να το μετατρέψει σε CMYK, δηλαδή σε τέσσερα κλάδια που αναπαράξουν στο χρώμα της τρέφω κρούσης επίπτωσης offset (C=κυανό, M=μαζέλι, Y=κίτρινο, K=μαύρο)]. Πάντα, όμως, ο scanner "δηλώνει" αρχικά μια εικόνα ως RGB. Μετατρέποντας τον RGB αρχικά σε CMYK επιβεβαιώνεται από pixel (1 byte ή 8 bits για κάθε κρούση, κίνηση, πρόβλημα, μπλέκ, ε, συνολικά 24bits). Οι εικόνες CMYK έχουν 4bytes πληροφορία για κάθε pixel λόγω των τεσσάρων κλάδων.

Bytes, bits και άλλα... χρήσιμα
 Αναπαράξουν πληροφορία (π.χ. μια εικόνα κίνηση ή ένα αρχείο που) μπορεί να περιγραφεί με μια δεδομένη γραμμή 0 και 1. Με βάση το **δεδομένο σύστημα** (σύστημα κρούσης) κλάση από το οποίο κλάση bit. Μια εικόνα δε από ψηφιακή κρούση RGB. Ο υπολογιστής μπορεί να "περιγράψει" οποιαδήποτε πληροφορία με 8 bits κίνηση, σε συνδυασμό των οποίων είναι $2^8=256$. Με άλλα λόγια, ο υπολογιστής είναι σε θέση να διαβαθμίσει συνολικά σε 256 διαβαθμίσεις τιμής. Σε αναπαράξουν επεξεργασία εικόνας οι αρχές αυτές θα μπορούσαν να αναπαράξουν σε 256 διαβαθμίσεις φωτεινότητας για κάθε pixel μιας εικόνας. Ένα byte (8bits) είναι η πιο κοινή μεθοδολογία της ψηφιακής ενός υπολογιστή. As παράρτηση ως παράρτηση μια ψηφιακή εικόνα που έχει πληροφορία μόνο για έναν του pixel, δηλαδή μονοχρωματική πληροφορία [εικ. 2]. Η πιο παλαιά εκδοχή είναι κλάση pixel της εικόνας να καταγράφεται 1 byte (8bits). Έτσι, όπως εξηγήσαμε προηγουμένως, κάθε pixel της εικόνας μπορεί να πάρει μια από 256 διαφορετικές τιμές τόνου διαβαθμίσεων. Στην τιμή 0 αντιστοιχεί το τελείως μαύρο (από 0 στα 255 είναι τόνος) και στην τιμή 255 το τελείως λευκό. Φανταστείτε μια κρούση 256 τόνων διαβαθμίσεων, η οποία ξεκινά από το τελείως μαύρο (την 0) και καταλήγει στο τελείως λευκό (την 255) [εικ. 2]. Οι διαβαθμίσεις αυτές αναπαράξουν με διαβαθμίσεις του ψηφ (σε μικρό αριθμό αντιστοιχούν σε ορισμένες διαβαθμίσεις και οι μεγάλοι σε φωτεινές). Στο 0, λοιπόν, θα αντιστοιχούσε η τιμή του "bit image" στην τιμή 0.

(δ)

Σχήμα 5.7. Εξομάλυνση εικόνας για τον εντοπισμό των διαφόρων περιοχών με χρήση RLSA. (α,γ) Αρχική εικόνα. (β,δ) Εντοπισμός διαφόρων περιοχών (Οριζόντιο RLSA με $T_{max} = 24$ και κάθετο RLSA με $T_{max} = 24$).

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Για να οριοθετήσουμε τις περιοχές που έχουν εντοπιστεί με την χρήση της οριζόντιας και κάθετης εξομάλυνσης χρησιμοποιούμε την τεχνική του ονοματισμού των συνδεδεμένων συστατικών (connected component labelling) καθώς και της ακολούθησης του περιγράμματος των αντικειμένων (contour following).

Ονοματισμός των συνδεδεμένων συστατικών

Τα συνδεδεμένα συστατικά της ασπρόμαυρης εικόνας ορίζονται ως εξής: Δύο pixels της εικόνας ανήκουν στο ίδιο συνδεδεμένο συστατικό αν υπάρχουν διαδοχικά pixels της εικόνας τα οποία να ορίζουν κάποια διαδρομή ώστε να πάμε από το ένα pixel στο άλλο. Τα διαδοχικά pixels της διαδρομής μπορεί να έχουν μεταξύ τους γειτονία 4 ή 8 pixels (σχήμα 5.8). Σκοπός του ονοματισμού των συνδεδεμένων συστατικών (connected component labelling) είναι να ομαδοποιηθούν τα pixels που ανήκουν στο ίδιο συνδεδεμένο συστατικό. Η ασπρόμαυρη εικόνα μετατρέπεται έτσι ώστε η τιμή του κάθε pixel να αντιστοιχεί στην τιμή του συνδεδεμένου συστατικού στο οποίο ανήκει. Μια απλή υλοποίηση του ονοματισμού των συνδεδεμένων συστατικών (Shapiro 2001) περιλαμβάνει δύο σαρώσεις της εικόνας και έχει ως εξής (λαμβάνοντας υπόψη την γειτονία των 8 pixels):

- **ΒΗΜΑ 1:** Η εικόνα σαρώνεται από πάνω προς τα κάτω και από αριστερά προς τα δεξιά. Για κάθε pixel εικόνας p_i που συναντάμε εκτελούμε τα παρακάτω:
 - Εξετάζουμε τα NW, N, NE, W γειτονικά pixels (σχήμα 5.8β).
 - Αν και τα 4 αυτά pixels είναι pixels υποβάθρου, τότε δίνουμε στο p_i νέα τιμή συνδεδεμένου συστατικού.
 - Αν όσα από τα 4 αυτά pixels είναι pixel εικόνας, ανήκουν στο ίδιο συνδεδεμένο συστατικό τότε το p_i παίρνει την τιμή του συνδεδεμένου συστατικού.
 - Αν όσα από τα 4 αυτά pixels είναι pixel εικόνας, δεν ανήκουν στο ίδιο συνδεδεμένο συστατικό, τότε το p_i παίρνει την τιμή του συνδεδεμένου συστατικού που έχει ένα από αυτά τα pixels και ταυτόχρονα όλες οι τιμές των συνδεδεμένων συστατικών των pixels αυτών σημειώνονται ως ισοδύναμες.
- **ΒΗΜΑ 2:** Η εικόνα σαρώνεται και όλες οι τιμές των συνδεδεμένων συστατικών που έχουν σημειωθεί ως ισοδύναμες αντικαθίστανται από μία.

Στο σχήμα 5.9 δίδεται ένα παράδειγμα εφαρμογής του αλγορίθμου σε μία απλή εικόνα. Μετά την εφαρμογή του βήματος 1 (σχήμα 5.9β), έχουν σημειωθεί ως ισοδύναμα τα συνδεδεμένα συστατικά με τιμές 2 και 3. Μετά την εφαρμογή του βήματος 2 (σχήμα 5.9γ), οι τιμές 2 και 3 των συνδεδεμένων συστατικών αντικαθίστανται από την τιμή 2.

	N	
W	*	E
	S	

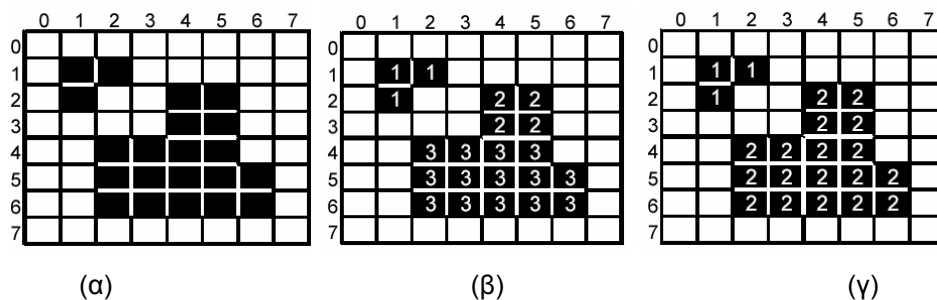
(α)

NW	N	NE
W	*	E
SW	S	SE

(β)

Σχήμα 5.8. (α) Γειτονία 4 pixels. (β) Γειτονία 8 pixels.

B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων



Σχήμα 5.9. Ονοματισμός των συνδεδεμένων συστατικών. (α) Αρχική εικόνα. (β) Αποτέλεσμα βήματος 1. (γ) Τελικό αποτέλεσμα μετά την εφαρμογή του βήματος 2.

Ακολουθήση του περιγράμματος

Για την οριοθέτηση μιας περιοχής, εκτός από τον ονοματισμό των συνδεδεμένων συστατικών, μπορούμε εναλλακτικά να ακολουθήσουμε το περίγραμμα των συνδεδεμένων συστατικών. Σύμφωνα με τον (Pavlidis 1981), βασιζόμαστε στην γειτονία των 8 pixels όπως αυτή ορίζεται στο σχήμα 5.10 και διατρέχουμε το περίγραμμα του αντικειμένου διαλέγοντας κάθε φορά το καταλληλότερο διαθέσιμο γειτονικό pixel.

3	2	1
4	P	0
5	6	7

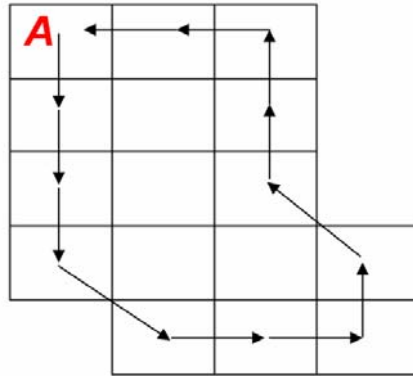
Σχήμα 5.10. Τιμές που δίδονται στα γειτονικά pixels για να οριστεί η σχετική τους θέση ως προς το pixel P.

Το αρχικό pixel A μπορεί να εντοπιστεί κάνοντας μία από πάνω προς τα κάτω σάρωση της εικόνας. Ο αλγόριθμος τερματίζει όταν το τρέχων pixel της ακολουθίας είναι το αρχικό pixel A. Αναλυτικά ο αλγόριθμος έχει ως εξής:

- BHMA 0: Επιλέγω το αρχικό pixel A το οποίο έχει ένα τουλάχιστον σημείο υποβάθρου στην γειτονία των 4 pixels.
- BHMA 1: Θέτω το τρέχων σημείο $C = A$, την διεύθυνση διαδρομής $S = 6$, την μεταβλητή $first = true$.
- BHMA 2: **While** $C \neq A$ ή $first = true$ **do** βήματα 3-10
Begin
- BHMA 3: Θέτω την παράμετρο $found = false$.
- BHMA 4: **While** $found = false$ **do** βήματα 5-9, το πολύ τρεις φορές
Begin
- BHMA 5: **If** το B, ο S-1 γείτονας του C, ανήκει στην εικόνα, **then**
Begin
- BHMA 6: Θέτω $C=B$, $S=S-2$, $found=true$
End.
- BHMA 7: **Else If** το B, ο S γείτονας του C, ανήκει στην εικόνα, **then**
Θέτω $C=B$, $found=true$
- BHMA 8: **Else If** το B, ο S+1 γείτονας του C, ανήκει στην εικόνα, **then**
Θέτω $C=B$, $found=true$
- BHMA 9: **Else** $S = S + 2$
End.
- BHMA 10: Θέτω $first = false$
End.

B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Στον παραπάνω αλγόριθμο θεωρούμε κάθε φορά το S ως το υπόλοιπο της διαίρεσης του S με το 8 . Το αποτέλεσμα του αλγορίθμου για εξωτερικά περιγράμματα είναι με μία ακολουθία σημείων αντι-ωρολογιακής φορά. Για την πλήρη περιγραφή των αντικειμένων πρέπει να βρεθούν και οι οπές τους οι οποίες περιγράφονται με μια ακολουθία σημείων ωρολογιακής φοράς. Στο σχήμα 5.11, δίδεται ένα παράδειγμα ακολούθησης του περιγράμματος ενός απλού αντικειμένου.



Σχήμα 5.11. Παράδειγμα ακολούθησης του περιγράμματος ενός απλού αντικειμένου.

Οριοθέτηση περιοχών

Για να οριοθετήσουμε τις περιοχές που έχουν βρεθεί με χρήση του ονοματισμού των συνδεδεμένων συστατικών ή της ακολούθησης του περιγράμματος, στις περισσότερες περιπτώσεις αρκεί να βρούμε το ορθογώνιο παραλληλόγραμμο που τις περικλείει (σχήμα 5.12).

(α)

(β)

Σχήμα 5.12. Οριοθέτηση περιοχών με χρήση ορθογώνιων παραλληλόγραμμων. (α) Η εικόνα του σχήματος 5.3α μετά από εξομάλυνση. (β) Περιοχές που ορίζονται από ορθογώνια παραλληλόγραμμο.

B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Υπάρχουν όμως και περιπτώσεις όπου οι εντοπισμένες περιοχές δεν μπορούν να οριστούν με ορθογώνια παραλληλόγραμμα. Αυτές οι περιπτώσεις αφορούν κυρίως μικτά έγγραφα (αποτελούνται από κείμενα και εικόνες) με περίπλοκη δομή ή τμήματα κειμένου σε μικρή απόσταση και με μικρή κλίση (Σχήμα 5.13). Σε αυτές τις περιπτώσεις, για την οριοθέτηση των εντοπισμένων περιοχών μπορούμε να χρησιμοποιήσουμε ισοθετικά πολύγωνα, τα οποία είναι πολύγωνα τα οποία έχουν μόνο οριζόντιες και κάθετες πλευρές. Τα ισοθετικά πολύγωνα χρησιμοποιούνται γιατί είναι εύκολα στον χειρισμό (εύκολα γίνονται αλλαγές από τον χρήστη), δεν έχουν μεγάλο αποθηκευτικό κόστος, ενώ μπορούν να αναπαρασταθούν με ευκολία στην οθόνη. Επιπλέον, τα ισοθετικά πολύγωνα ορίζουν περιοχές τις οποίες μπορούμε εύκολα να χειριστούμε σε άλλα στάδια επεξεργασίας της εικόνας (αναγνώριση, αυτόματη αποτίμηση του αποτελέσματος κ.λ.π.).

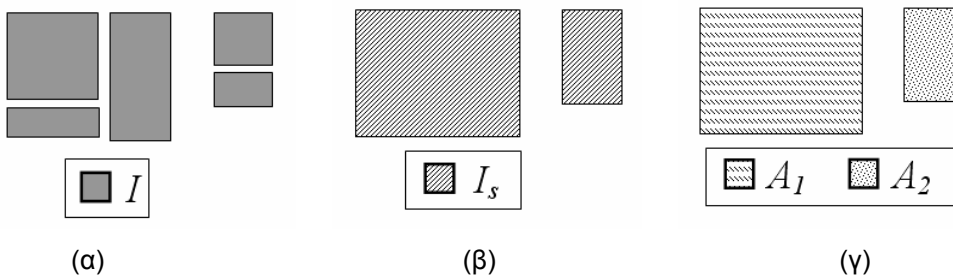


(α)

(β)

Σχήμα 5.13. Οριοθέτηση περιοχών με χρήση ισοθετικών πολυγώνων. (α) Μικτό έγγραφο με περίπλοκη δομή. (β) Έγγραφο με περιοχές κειμένου σε μικρή απόσταση και με μικρή κλίση.

Για τον ορισμό των βέλτιστων ισοθετικών πολυγώνων για την οριοθέτηση των εντοπισμένων περιοχών μπορούμε να ακολουθήσουμε την παρακάτω μεθοδολογία (Gatos 2000). Έστω μία δυαδική εικόνα I . Χρησιμοποιώντας την διαδικασία της εξομάλυνσης εξάγουμε την εικόνα I_s η οποία αποτελείται από τα συνδεδεμένα συστατικά A_i (σχήμα 5.14).



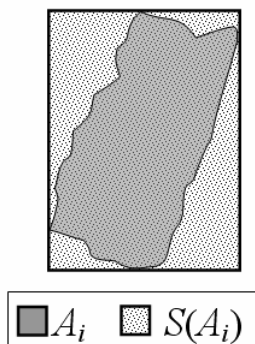
(α)

(β)

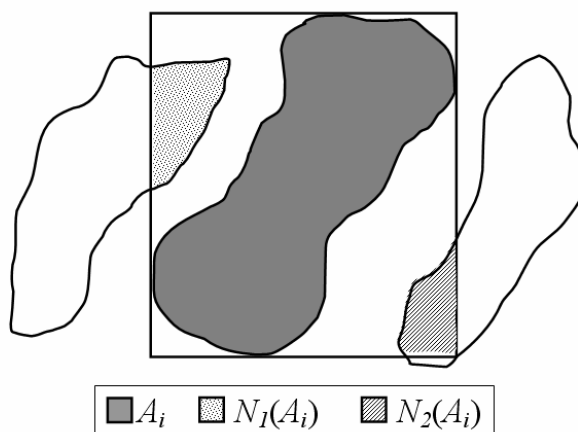
(γ)

Σχήμα 5.14. (α) Αρχική εικόνα I . (β) Εξομαλυσμένη εικόνα I_s . (γ) Συνδεδεμένα συστατικά A_1 και A_2 της εικόνας I_s .

Επίσης ορίζουμε ως $S(A_i)$ το παραλληλόγραμμο που περικλείει το συνδεδεμένο συστατικό A_i (σχήμα 5.15) και $N_j(A_i)$ όλες τα συνδεδεμένα συστατικά τα οποία ανήκουν στο $S(A_i)$ αλλά όχι στο A_i , όπου $j = 1 \dots F(A_i)$ και $F(A_i)$ είναι ο αριθμός των συνδεδεμένων συστατικών τα οποία ανήκουν στο $S(A_i)$ αλλά όχι στο A_i . Στο σχήμα 5.16 δίδεται ένα παράδειγμα του συνδεδεμένου συστατικού A_i και των αντίστοιχων $N_1(A_i)$, $N_2(A_i)$ περιοχών.



Σχήμα 5.15. Το συνδεδεμένο συστατικό A_i και το παραλληλόγραμμο που το περικλείει $S(A_i)$.



Σχήμα 5.16. Το συνδεδεμένο συστατικό A_i και οι αντίστοιχες περιοχές $N_1(A_i)$, $N_2(A_i)$ οι οποίες ανήκουν στο $S(A_i)$ και όχι στο A_i .

Το ισοθετικό πολύγωνο $P(A_i)$ το οποίο περιέχει μόνο το συνδεδεμένο συστατικό A_i και όχι άλλα τμήματα άλλων συνδεδεμένων συστατικών, δίδεται από τον αναδρομικό τύπο:

$$P(A_i) = S(A_i) - \bigcup_{j=1}^{F(A_i)} P(N_j(A_i)) \quad (5.3)$$

Η περιοχή $P(A_i)$ είναι ισοθετικό πολύγωνο γιατί προκύπτει από προσθαφαίρεση ορθογώνιων παραλληλογράμμων. Για τον υπολογισμό του $P(A_i)$, χρησιμοποιούμε τον παραπάνω αναδρομικό τύπο. Το τέλος της αναδρομής είναι όταν η περιοχή $S(A_i)$ περιλαμβάνει μόνο το A_i και όχι άλλα τμήματα συνδεδεμένων συστατικών. Τότε, ισχύει: $P(A_i)=S(A_i)$. Ένα παράδειγμα υπολογισμού του ισοθετικού πολυγώνου $P(A_i)$ δίδεται στο σχήμα 5.17. Στην περίπτωση αυτή ισχύει $F(A_i) = 2$, οπότε:

$$P(A_i) = S(A_i) - P(N_1(A_i)) - P(N_2(A_i)) \quad (5.4)$$

Εφόσον $F(N_1(A_i)) = 1$ και $F(N_2(A_i)) = 0$, τα ισοθετικά πολύγωνα $P(N_1(A_i))$, $P(N_2(A_i))$ υπολογίζονται ως εξής:

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

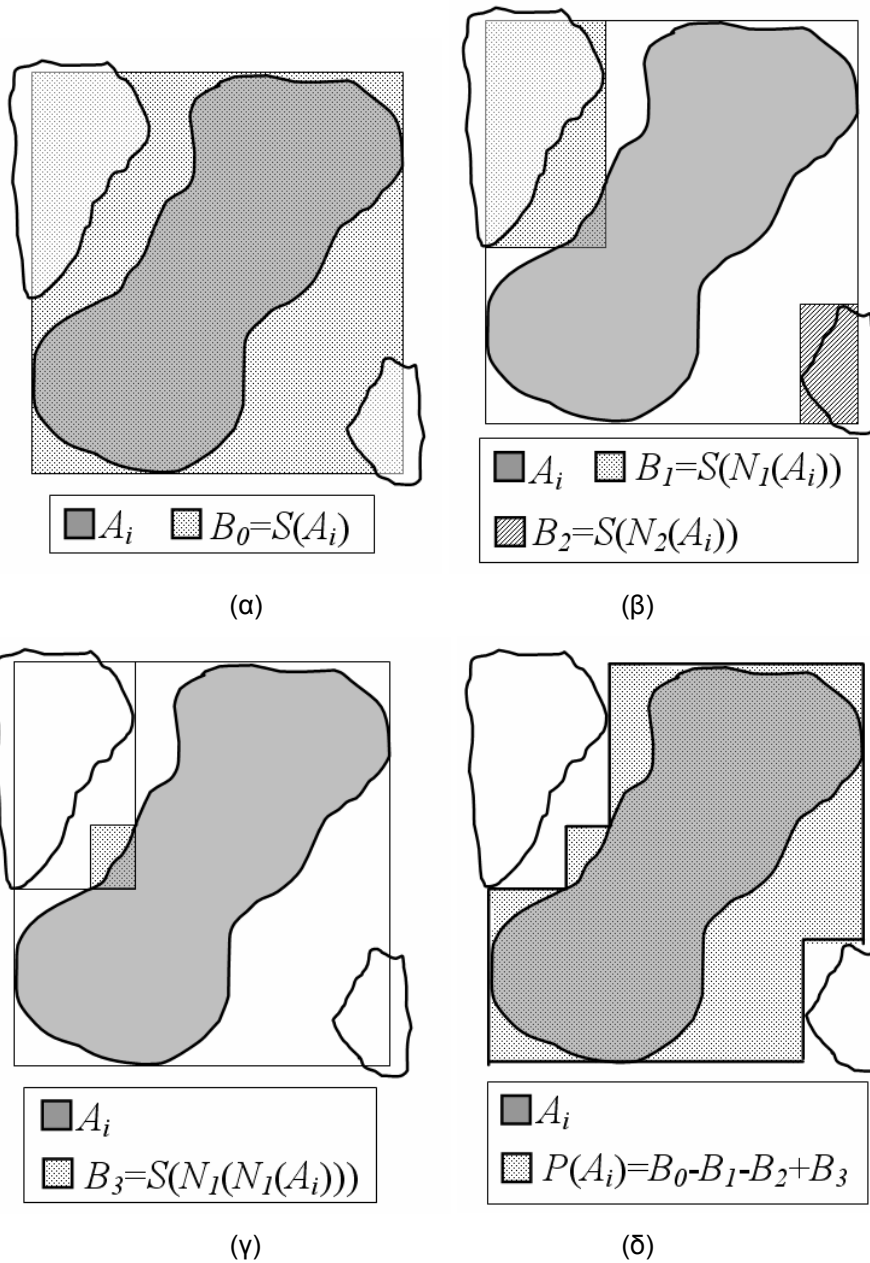
$$P(N_1(A_i)) = S(N_1(A_i)) - P(N_1(N_1(A_i)))$$

$$P(N_2(A_i)) = S(N_2(A_i)) \tag{5.5}$$

$$P(N_1(N_1(A_i))) = S(N_1(N_1(A_i)))$$

και το τελικό ισοθετικό πολύγωνο $P(A_i)$ γίνεται:

$$P(A_i) = S(A_i) - S(N_1(A_i)) + S(N_1(N_1(A_i))) - S(N_2(A_i)) \tag{5.6}$$



Σχήμα 5.17. Υπολογισμός του ισοθετικού πολυγώνου $P(A_i)$ το οποίο περιέχει το συνδεδεμένο συστατικό A_i .

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Αν B_v είναι όλα τα ορθογώνια παραλληλόγραμμα που πρέπει να υπολογιστούν προκειμένου να υπολογίσουμε το τελικό ισοθετικό πολύγωνο $P(A_i)$ και R_v οι φορές που θα χρειαστεί η $P()$ να καλέσει τον εαυτό της για να υπολογιστεί το B_v , τότε το $P(A_i)$ μπορεί να οριστεί ως εξής:

$$P(A_i) = \sum_{v=0}^N \left(4 \left\lfloor \frac{R_v}{2} \right\rfloor - 2R_v + 1\right) B_v \quad (5.7)$$

όπου $|x|$ είναι το ακέραιο μέρος του x . Η παράμετρος $\left(4 \left\lfloor \frac{R_v}{2} \right\rfloor - 2R_v + 1\right)$ είναι ίση με 1 αν το R_v είναι ζυγός και -1 αν είναι μονός. Για το παράδειγμα του σχήματος 5.17, έχουμε:

$$B_0 = S(A_i), R_0 = 0 \quad (5.8)$$

$$B_1 = S(N_1(A_i)), R_1 = 1 \quad (5.9)$$

$$B_2 = S(N_2(A_i)), R_2 = 1 \quad (5.10)$$

$$B_3 = S(N_1(N_1(A_i))), R_3 = 2 \quad (5.11)$$

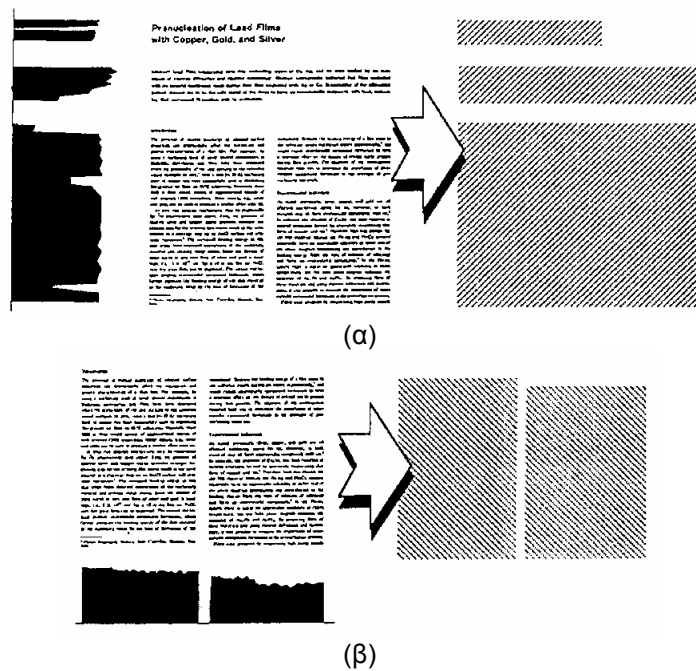
Το τελικό ισοθετικό πολύγωνο δίδεται από τον τύπο:

$$P(A_i) = B_0 - B_1 - B_2 + B_3 \quad (5.12)$$

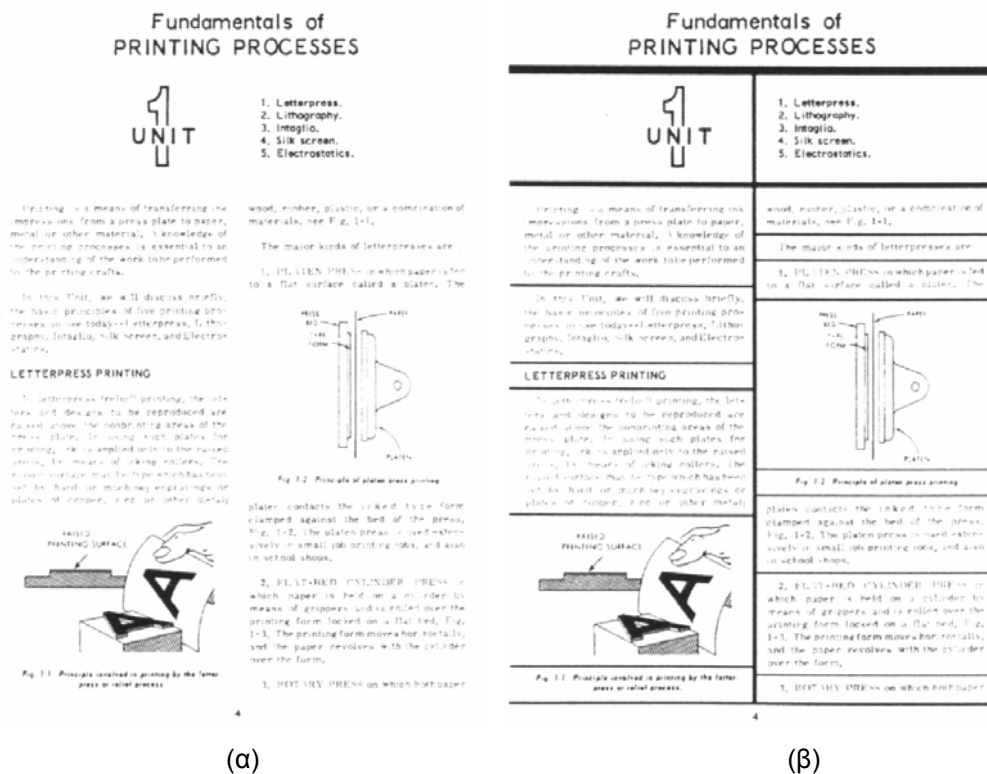
5.4 Χρήση προβολών

Οι οριζόντιες και κάθετες προβολές χρησιμοποιούνται για την τμηματοποίηση της εικόνας. Μία από τις πιο διαδεδομένες μεθόδους είναι η επαναληπτική εφαρμογή των οριζόντιων και κάθετων προβολών (Recursive X-Y cuts) (Nagy 1984). Σύμφωνα με αυτή την μέθοδο, υπολογίζονται οι προβολές (άθροισμα των pixels σε γραμμή ή στήλη της εικόνας του εγγράφου) σε οριζόντια και κάθετη διεύθυνση (σχήμα 5.18). Στη συνέχεια εντοπίζονται τα διαδοχικά μηδενικά των προβολών τα οποία θεωρούνται και όρια οριζόντιας ή κάθετης τμηματοποίησης όταν είναι πάνω από ένα ελάχιστο μήκος. Για κάθε τέτοιο όριο, η εικόνα χωρίζεται σε δύο υπο-εικόνες. Για κάθε υπο-εικόνα, η ίδια διαδικασία επαναλαμβάνεται στην άλλη διεύθυνση (αν το πρώτο χώνισμα είναι κάθετο, το επόμενο είναι οριζόντιο) και η όλη διαδικασία επαναλαμβάνεται μέχρι να μην μπορεί να χωριστεί η εικόνα άλλο σε οριζόντια ή κάθετη διεύθυνση. Αν η εικόνα περιέχει θόρυβο, τότε αντι να αναζητούμε μηδενικά στις προβολές, μπορούμε να αναζητούμε μικρό αριθμό pixels, όμως τότε πέφτει η ακρίβεια της μεθόδου. Το αποτέλεσμα της επαναληπτικής εφαρμογής των οριζόντιων και κάθετων προβολών μπορεί να αναπαρασταθεί με την μορφή δέντρων τα οποία ονομάζονται X-Y δένδρα. Στο σχήμα 5.19 δίδεται ένα παράδειγμα επαναληπτικής εφαρμογής των οριζόντιων και κάθετων προβολών για την τμηματοποίηση εικόνας εγγράφου. Στο παράδειγμα αυτό, οι γραμμές που ορίζουν τα όρια των διαφόρων περιοχών έχουν μεγαλύτερο πάχος όταν έχουν εντοπιστεί στις πρώτες επαναλήψεις της διαδικασίας. Στο σχήμα 5.20 δίδεται μία αναπαράσταση του αποτελέσματος της επαναληπτικής εφαρμογής των οριζόντιων και κάθετων προβολών με χρήση του X-Y δένδρου.

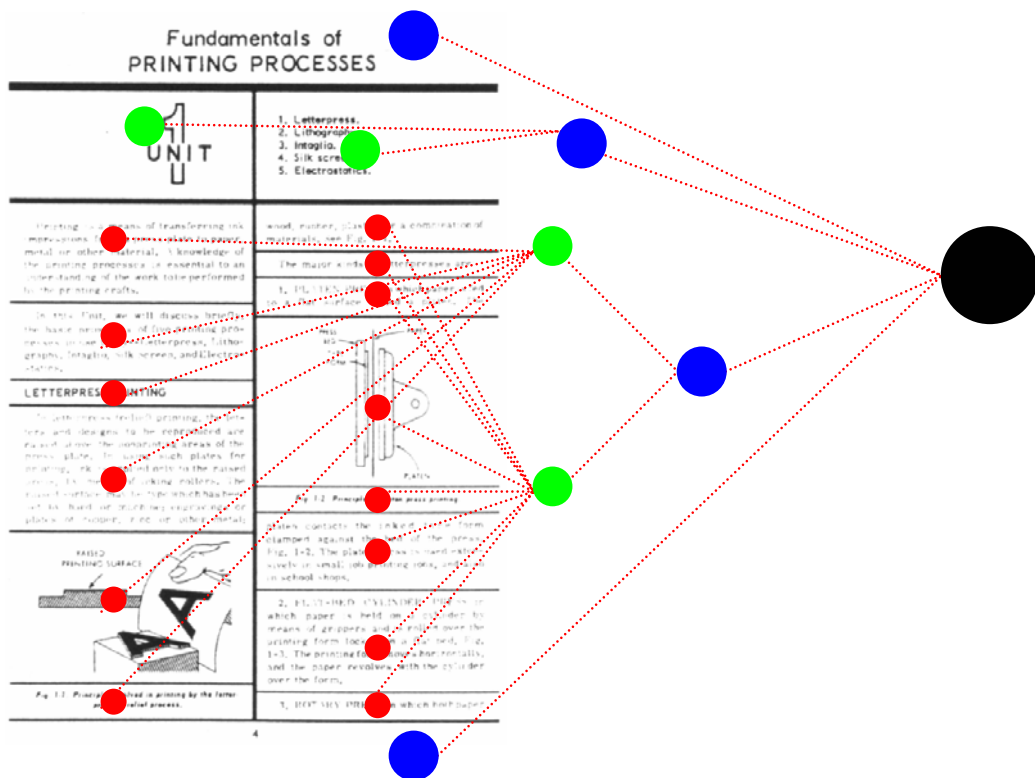
Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων



Σχήμα 5.18. Χρήση προβολών για την κατάτμηση της εικόνας. (α) Οριζόντιες προβολές. (β) Κάθετες προβολές.


(α)
(β)

Σχήμα 5.19. Επαναληπτική εφαρμογή των οριζόντιων και κάθετων προβολών για την τμηματοποίηση της εικόνας. (α) Αρχική εικόνα. (β) Αποτέλεσμα τμηματοποίησης.

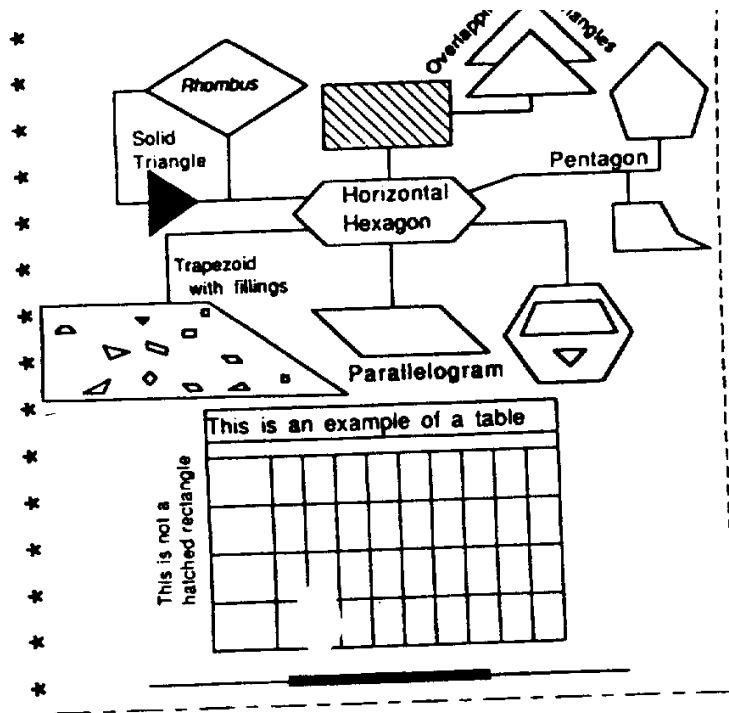


Σχήμα 5.20. Αναπαράσταση του αποτελέσματος της επαναληπτική εφαρμογή των οριζόντιων και κάθετων προβολών με χρήση του X-Y δένδρου.

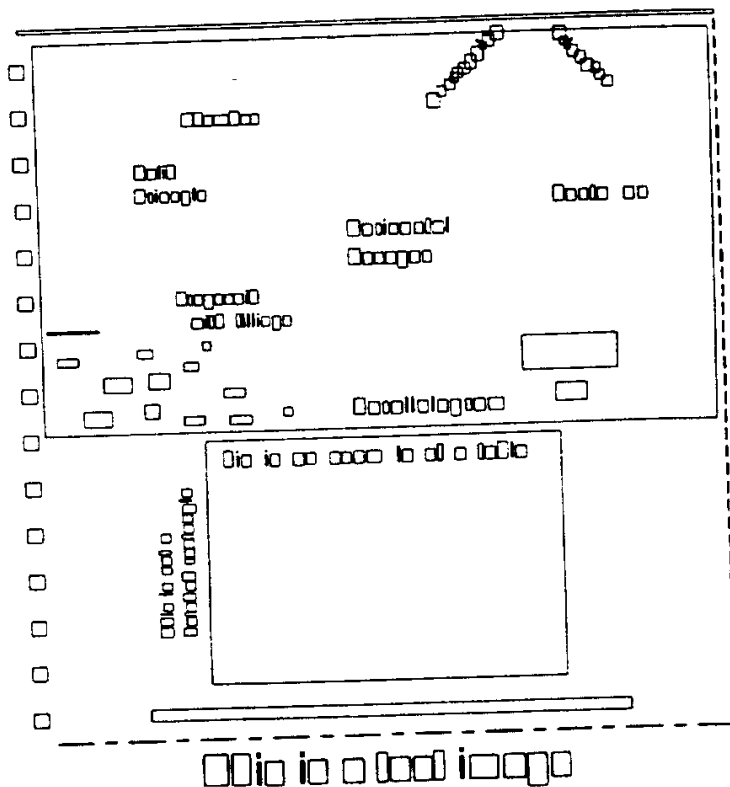
5.5 Χρήση μετασχηματισμού Hough

Ο μετασχηματισμός Hough χρησιμοποιείται για τον εντοπισμό των περιοχών κειμένου κυρίως σε σύνθετα έγγραφα όπου το κείμενο συνυπάρχει με γραφικά, πίνακες κ.λ.π. (Kasturi 1990). Η χρήση του μετασχηματισμού Hough είναι μία προσέγγιση «από κάτω προς τα πάνω» (bottom-up) γιατί βασίζεται στην ομαδοποίηση των συνδεδεμένων συστατικών της εικόνας ώστε να προκύψουν οι γραμμές του κειμένου. Συγκεκριμένα, αφού γίνει μία ανάλυση της εικόνας στα συνδεδεμένα συστατικά της, εντοπίζονται τα συστατικά που ανήκουν στο κείμενο (λόγω μεγέθους και γεωμετρίας) και στη συνέχεια εφαρμόζεται ο μετασχηματισμός Hough (Duda 1972) για τα κέντρα τους. Στο παράδειγμα του σχήματος 5.21 φαίνεται η ανάλυση ενός σύνθετου εγγράφου στα συνδεδεμένα συστατικά του. Τα συνδεδεμένα συστατικά που αντιστοιχούν στο κείμενο είναι μικρού μεγέθους και ορίζουν ευθείες που αντιστοιχούν στις γραμμές κειμένου του εγγράφου. Οι τιμές των (ρ, θ) που αντιστοιχούν στα τοπικά μέγιστα του πίνακα συσσώρευσης του μετασχηματισμού Hough δίνουν τις γραμμές κειμένου της εικόνας. Τα πλεονεκτήματα της μεθόδου αυτής είναι ότι οι γραμμές κειμένου μπορεί να βρίσκονται σε οποιαδήποτε γωνία και προσανατολισμό, εντός πινάκων ή σχημάτων κ.λ.π. Τα προβλήματα της μεθόδου είναι:

- η κατάλληλη επιλογή κβάντισης του πίνακα συσσώρευσης ως προς ρ και θ , ώστε να εντοπιστούν οι γραμμές του κειμένου.
- η μεγάλη καθυστέρηση που εισάγεται.



(α)



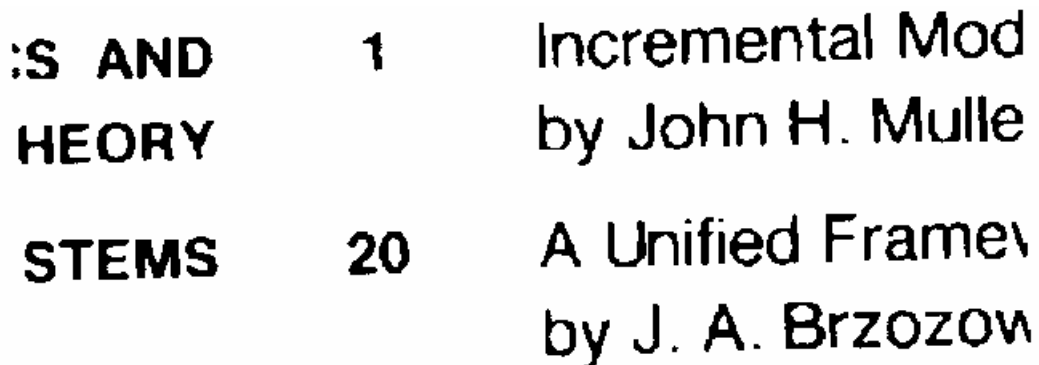
(β)

Σχήμα 5.21. Εντοπισμός συνδεδεμένων συστατικών σε σύνθετη εικόνα. (α) Αρχική εικόνα. (β) Τα ορθογώνια παραλληλόγραμμα που περικλείουν όλα τα συνδεδεμένα συστατικά.

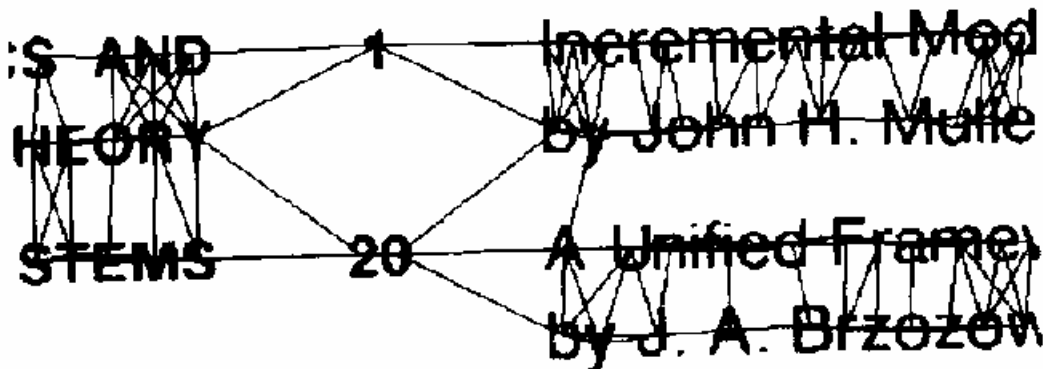
5.6 Φάσμα εγγράφου (Document Spectrum - DocStrum)

Το φάσμα του εγγράφου (Document Spectrum – DocStrum) είναι μια αναπαράσταση της σελίδας του εγγράφου η οποία περιγράφει τα καθολικά δομικά χαρακτηριστικά της σελίδας και μπορεί να χρησιμοποιηθεί για την κατάτμησή της (Gorman 1993). Η χρήση του φάσματος του εγγράφου είναι μία προσέγγιση «από κάτω προς τα πάνω» (bottom-up) που βασίζεται στην ομαδοποίηση των συνδεδεμένων συστατικών της εικόνας χρησιμοποιώντας του κ-κοντινότερους γείτονες (K-nearest neighbor – KNN) (Theodoridis 1997).

Οι κ-κοντινότεροι γείτονες για κάθε συνδεδεμένο συστατικό i είναι τα κ πλησιέστερα συνδεδεμένα συστατικά. Η απόσταση που χρησιμοποιείται είναι η ευκλείδεια απόσταση μεταξύ των κέντρων των συνδεδεμένων συστατικών. Στο σχήμα 5.22 δίδεται ένα παράδειγμα εύρεσης των κ-κοντινότερων γειτόνων σε μία εικόνα χρησιμοποιώντας $k=5$. Κάθε ζευγάρι $\{i,j\}$ συνδεδεμένων συστατικών που ανήκει στους κ-κοντινότερους γείτονες περιγράφεται από τις τιμές $\{d,\varphi\}$, όπου d είναι η απόσταση και φ η γωνία μεταξύ των δύο κέντρων των συνδεδεμένων συστατικών. Για παράδειγμα, δύο γειτονικοί χαρακτήρες της ίδια λέξης σχηματίζουν ένα ζευγάρι συνδεδεμένων συστατικών των οποίων η απόσταση d είναι σχετικά μικρή και η γωνία φ είναι κοντά στο 0. Χρησιμοποιώντας την τιμή $k=5$ (η τιμή που συνήθως χρησιμοποιείται) ένας χαρακτήρας συνήθως έχει δύο ή τρία ζευγάρια μέσα στην λέξη και σε γειτονικές λέξεις της ίδιας γραμμής και επίσης ζευγάρια με χαρακτήρες της πάνω και κάτω γραμμής. Τα ζευγάρια αυτά των συνδεδεμένων συστατικών που βρίσκονται σε διαφορετικές γραμμές έχουν σχετικά μεγάλη απόσταση d και γωνία περίπου στις 90 μοίρες. Το φάσμα του εγγράφου είναι η γραφική παράσταση όλων των τιμών $\{d,\varphi\}$ για όλα τα ζευγάρια συνδεδεμένων συστατικών που ανήκουν στους κ-κοντινότερους γείτονες της εικόνας (σχήμα 5.23). Είναι πολική παράσταση με κέντρο το (0,0), ακτινική απόσταση d και αντι-ωρολογιακή γωνία από την οριζόντια διεύθυνση φ . Για καλύτερη οπτικοποίηση, για κάθε σημείο εισάγουμε και το κατοπτρικό του ώστε η γωνία να είναι στο διάστημα $[0^\circ, 180^\circ]$. Ο προσανατολισμός του εγγράφου και η θέση των γραμμών κειμένου μπορούν να εξαχθούν άμεσα από τον εντοπισμό ομάδων σημείων στο γράφημα του φάσματος του εγγράφου.

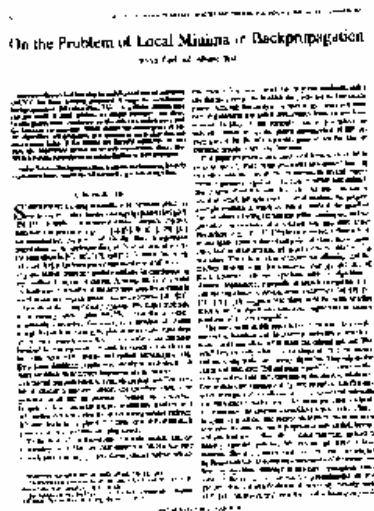


(α)

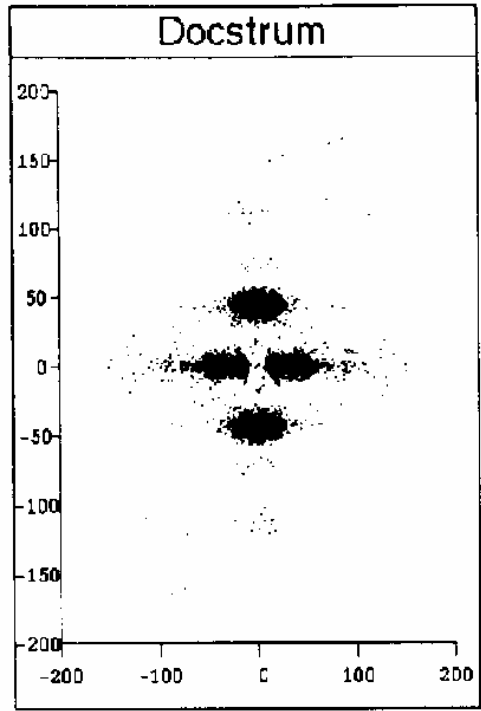


(β)

Σχήμα 5.22. Εύρεση των κ-κοντινότερων γειτόνων των συνδεδεμένων συστατικών εικόνας ($k=5$). (α) Αρχική εικόνα. (β) Αναπαράσταση των κ-κοντινότερων γειτόνων.



(α)



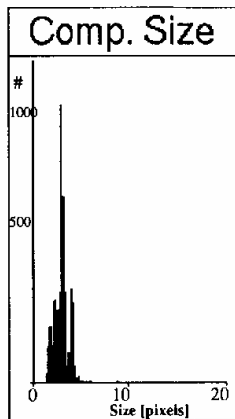
(β)

Σχήμα 5.23. Το φάσμα του εγγράφου. (α) Αρχική εικόνα. (β) Το γράφημα του φάσματος του εγγράφου.

Αναλυτικά τα βήματα που ακολουθούμε για την κατάτμηση της εικόνας με χρήση του φάσματος είναι τα ακόλουθα.

ΒΗΜΑ 1: Εντοπισμός των συνδεδεμένων συστατικών της εικόνας.

ΒΗΜΑ 2: Ομαδοποίηση συνδεδεμένων συστατικών ομοίου μεγέθους. Στις σελίδες που περιλαμβάνουν κείμενο με χαρακτηριστικές διαφορετικού μεγέθους, για παράδειγμα χαρακτηριστικές μεγάλου μεγέθους που ανήκουν σε τίτλους και χαρακτηριστικές κανονικού μεγέθους που ανήκουν στο κυρίως κείμενο, το φάσμα του εγγράφου πρέπει να υπολογιστεί για κάθε κατηγορία χαρακτηριστών ξεχωριστά. Ο λόγος είναι ότι η απόσταση και η γωνία μεταξύ των γειτονικών συνδεδεμένων συστατικών εξαρτάται από το μέγεθος των χαρακτηριστών. Η ανάλυση του φάσματος χαρακτηριστών με διαφορετικά μεγέθη μπορεί να καταλήξει σε εσφαλμένα αποτελέσματα. Για τον εντοπισμό των διαφόρων κατηγοριών που ανήκουν οι χαρακτηριστές, δημιουργούμε ένα ιστόγραμμα από τα ύψη των ορθογώνιων παραλληλογράμμων που περικλείουν τα συνδεδεμένα συστατικά (σχήμα 5.24).



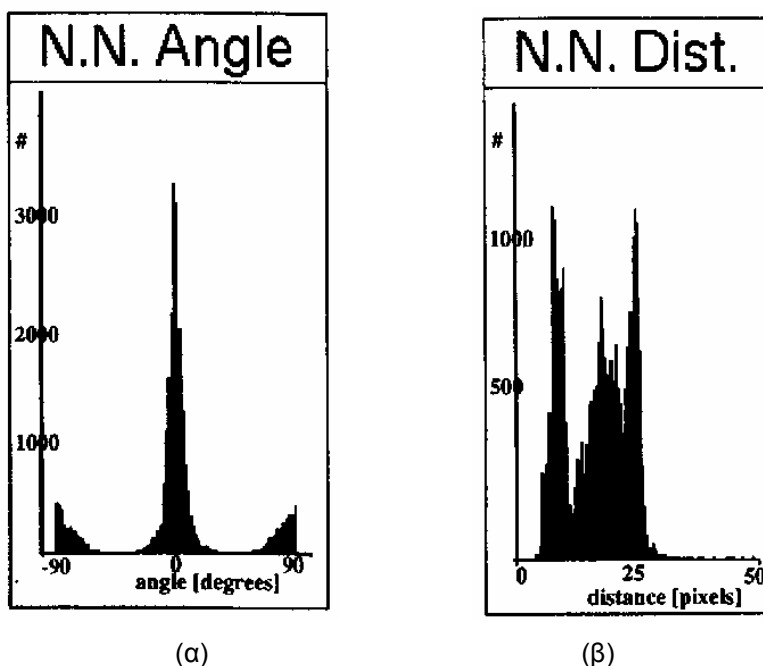
Σχήμα 5.24. Ιστόγραμμα με τα ύψη των ορθογώνιων παραλληλογράμμων που περικλείουν τα συνδεδεμένα συστατικά.

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Για ένα τυπικό έγγραφο, υπάρχει ένα μικρό μέγιστο που αντιστοιχεί σε θόρυβο, ένα μεγάλο μέγιστο σε μια πλατύτερη κατανομή που αντιστοιχεί στο κυρίως κείμενο, και συνήθως ένα μικρότερο μέγιστο που αντιστοιχεί στους τίτλους. Ανάλογα με τα εντοπισμένα τοπικά μέγιστα του ιστογράμματος των υψών ομαδοποιούνται και τα συνδεδεμένα συστατικά της εικόνας. Θα πρέπει να τονίσουμε ότι ο διαχωρισμός αυτός είναι αναγκαίος μόνο για μεγάλες διαφορές στο ύψος των συνδεδεμένων συστατικών. Για παράδειγμα, τα κεφαλαία και τα μικρά γράμματα καθώς και τα λίγο μικρότερα γράμματα κάποιας εξίσωσης ανήκουν στην ίδια κατηγορία.

ΒΗΜΑ 3: Δημιουργείται το φάσμα του εγγράφου για κάθε ομάδα συνδεδεμένων συστατικών. Για την επιτάχυνση της διαδικασίας υπολογισμού των κ-κοντινότερων γειτόνων όλα τα συνδεδεμένα συστατικά ταξινομούνται πρώτα ως προς την οριζόντια θέση τους. Από την ανάλυση του φάσματος ενός τυπικού εγγράφου μπορεί να εξαχθεί ένα αριθμός χαρακτηριστικών ο οποίος να συσχετιστεί με την πληροφορία της αρχικής εικόνας. Συνήθως, στο φάσμα του εγγράφου διακρίνονται τέσσερις ή έξι ομάδες σημείων. Στις μηδέν μοίρες έχουμε δύο βασικές κατανομές (για θετικό και αρνητικό x) και άλλες δύο πιθανές κατανομές που αντιστοιχούν σε μεγαλύτερη απόσταση. Οι δύο πρώτες αντιστοιχούν σε αποστάσεις μεταξύ γειτονικών χαρακτήρων που ανήκουν στην ίδια λέξη ενώ οι άλλες δύο σε αποστάσεις μεταξύ γειτονικών χαρακτήρων που ανήκουν σε διαφορετικές λέξεις της ίδιας γραμμής. Στον y -άξονα έχω δύο βασικές κατανομές που αντιστοιχούν στις αποστάσεις μεταξύ γειτονικών χαρακτήρων που ανήκουν σε διαφορετικές λέξεις διαφορετικής γραμμής.

ΒΗΜΑ 4: Εντοπισμός προσανατολισμού του εγγράφου. Δημιουργούμε ένα ιστόγραμμα με την κατανομή των γωνιών του φάσματος του εγγράφου (σχήμα 5.25α). Η μέγιστη κορυφή του ιστογράμματος αντιστοιχεί στην διεύθυνση προσανατολισμού του εγγράφου.



Σχήμα 5.25. (α) Ιστόγραμμα με την κατανομή των γωνιών φ του φάσματος του εγγράφου. (β) Ιστόγραμμα με την κατανομή των αποστάσεων d του φάσματος του εγγράφου.

ΒΗΜΑ 5: Εντοπισμός των αποστάσεων μεταξύ των χαρακτήρων. Δημιουργούμε δύο ιστογράμματα με την κατανομή των αποστάσεων που αντιστοιχούν στο φάσμα του εγγράφου (σχήμα 5.25β). Ένα για τα σημεία που αντιστοιχούν σε ένα εύρος γωνιών γύρω από την εντοπισμένη διεύθυνση προσανατολισμού του προηγούμενου βήματος και ένα σε ένα εύρος γωνιών γύρω από γωνία κάθετη στην εντοπισμένη διεύθυνση προσανατολισμού. Το πρώτο μας δίνει πληροφορίες για τις αποστάσεις των συνδεδεμένων συστατικών εντός της γραμμής κειμένου και το δεύτερο ανάμεσα σε διαφορετικές γραμμές κειμένου. Το μέγιστο που προκύπτει από το πρώτο ιστόγραμμα αντιστοιχεί στην απόσταση μεταξύ διαδοχικών χαρακτήρων της ίδιας λέξης, ενώ το μέγιστο που προκύπτει από το δεύτερο ιστόγραμμα αντιστοιχεί στην απόσταση μεταξύ δύο διαδοχικών γραμμών.

ΒΗΜΑ 6: Εντοπισμός των γραμμών κειμένου και της κλίσης τους. Ενώνουμε τα γειτονικά συνδεδεμένα συστατικά που έχουν μεταξύ τους γωνίες κοντά στην διεύθυνση προσανατολισμού του εγγράφου. Για κάθε γραμμή κειμένου, τα κέντρα των συνδεδεμένων συστατικών προσεγγίζονται από ευθεία της οποίας μπορεί να εντοπιστεί η κλίση. Στο παράδειγμα του σχήματος 5.26 φαίνονται οι ευθείες που έχουν προσεγγιστεί για κάθε γραμμή κειμένου.

S AND	1	Incremental Mod
HEORY		by John H. Mülle
STEMS	20	A Unified Framev
		by J. A. Brzozow

(α)

S AND	1	Incremental Mod
HEORY		by John H. Mülle
STEMS	20	A Unified Framev
		by J. A. Brzozow

(β)



(γ)

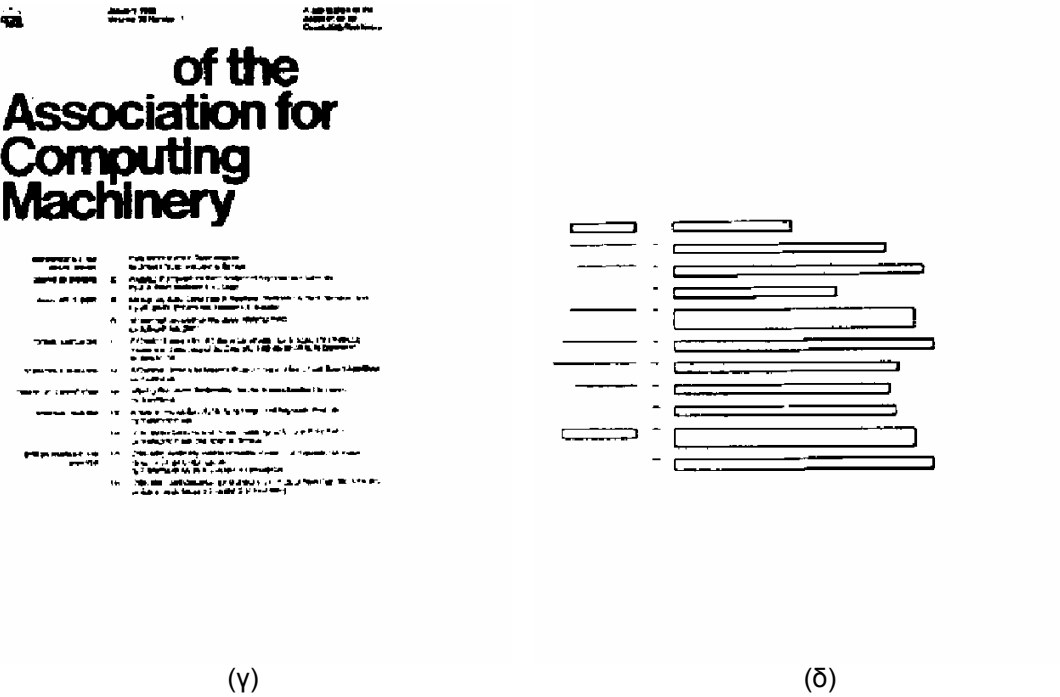
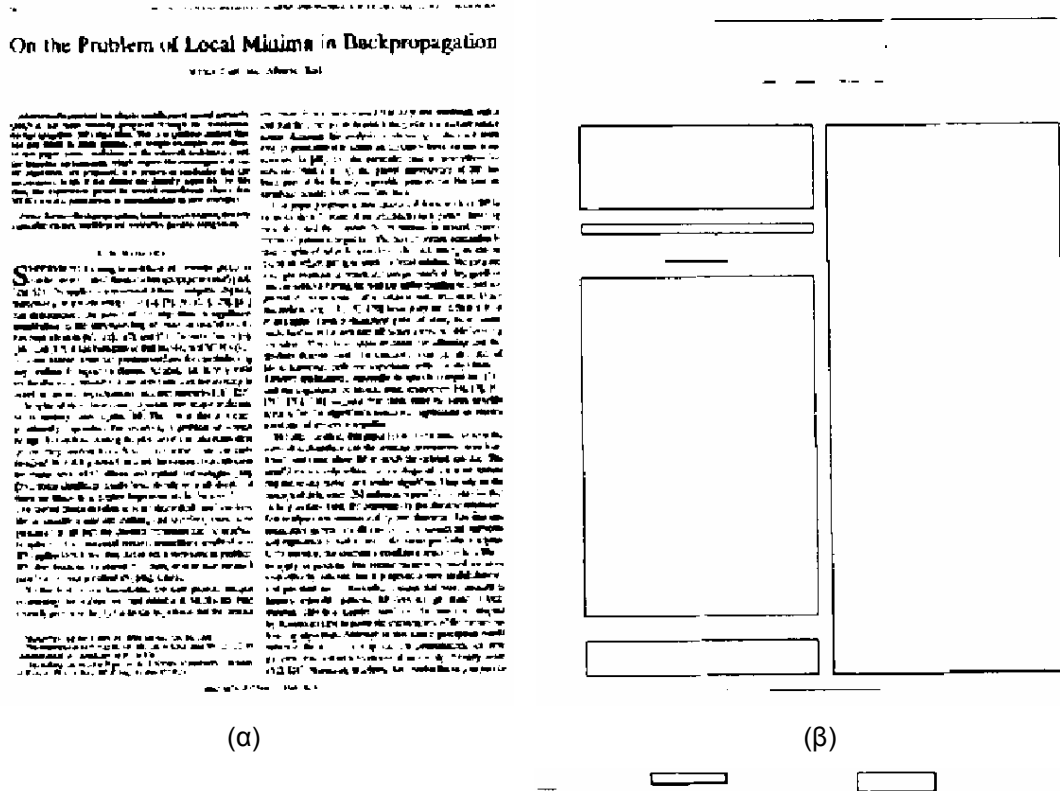
Σχήμα 5.26. Παράδειγμα εντοπισμού των γραμμών του κειμένου. (α) Αρχική εικόνα. (β),(γ) Οι εντοπισμένες ευθείες που αντιστοιχούν στις γραμμές του κειμένου.

ΒΗΜΑ 7: Εντοπισμός των blocks κειμένου. Οι εντοπισμένες γραμμές κειμένου ομαδοποιούνται βάσει χωρικών και γεωμετρικών χαρακτηριστικών. Συγκεκριμένα, δύο γραμμές κειμένου ανήκουν στο ίδιο block κειμένου εάν ισχύουν οι παρακάτω συνθήκες:

- Είναι σχεδόν παράλληλες
- Η κατακόρυφη απόσταση τους είναι μικρότερη από κάποιο όριο βάσει της εντοπισμένης απόστασης μεταξύ διαδοχικών γραμμών σε προηγούμενο στάδιο.
- Υπάρχει κατακόρυφη επικάλυψη ή είναι σε μικρή κατακόρυφη απόσταση.

Στο σχήμα 5.27 δίδονται δύο παραδείγματα εντοπισμού των blocks κειμένου με χρήση της ανάλυσης του φάσματος του εγγράφου.

Τα βασικά πλεονεκτήματα της μεθοδολογίας αυτής είναι: (α) Οι παράμετροι εφαρμογής (όρια για τις αποστάσεις ανάμεσα σε γράμματα και γραμμές κειμένου) δεν χρειάζεται να δοθούν από τον χρήστη αλλά εντοπίζονται από την μέθοδο. (β) Είναι ανεξάρτητη από τον προσανατολισμό και την κλίση του εγγράφου.



Σχήμα 5.27. Παραδείγματα εντοπισμού blocks κειμένου με την μέθοδο της ανάλυσης του φάσματος της εικόνας. (α),(γ) Αρχικές εικόνες. (β),(δ) Εντοπισμένα blocks κειμένου.

5.7 Ταξινόμηση περιοχών

Τα στάδια που έχουν περιγραφεί μέχρι στιγμής αφορούν στην τμηματοποίηση και οριοθέτηση των εντοπισμένων περιοχών. Η ταξινόμηση των περιοχών περιλαμβάνει την κατάταξη των εντοπισμένων περιοχών σε μία από τις παρακάτω βασικές κατηγορίες: Κείμενο, γραφικά, εικόνα, οριζόντια ή κάθετη γραμμή. Η διαδικασία αυτή είναι απαραίτητη προκειμένου να προχωρήσουμε στα στάδια της αναγνώρισης και κατανόησης των εγγράφων.

Οριζόντιες και κάθετες ευθείες

Μία εντοπισμένη περιοχή η οποία οριοθετείται από ένα ορθογώνιο παραλληλόγραμμο διαστάσεων $DX \times Dy$ μπορεί να είναι οριζόντια γραμμή αν ο λόγος πλάτους προς ύψος DX/DY είναι πάνω από κάποιο μεγάλο όριο. Αν η οριζόντια γραμμή έχει ασυνέχειες, τότε ο λόγος DX/DY μπορεί να μην είναι τόσο μεγάλος, όμως το ύψος DY θα είναι μικρότερο από το αναμενόμενο ελάχιστο ύψος των γραμμάτων (Yin 2001). Άρα για να ταξινομηθεί μία περιοχή ως οριζόντια ευθεία θα πρέπει:

$$DX/DY > r_1 \text{ OR } (DX/DY > r_2 \text{ AND } DY < h_1) \quad (5.13)$$

όπου r_1 και r_2 είναι το μεγάλο και το μεσαίο όριο για το λόγο πλάτους προς ύψος και h_1 είναι το ελάχιστο αναμενόμενο ύψος γράμματος. Οι παράμετροι αυτοί ορίζονται πειραματικά.

Αντίστοιχα, για να ταξινομηθεί μία περιοχή ως κάθετη ευθεία θα πρέπει:

$$DY/DX > r_3 \text{ OR } (DY/DX > r_4 \text{ AND } DX < h_2) \quad (5.14)$$

όπου r_3 και r_4 είναι το μεγάλο και το μεσαίο όριο για το λόγο ύψους προς πλάτος και h_2 είναι το ελάχιστο αναμενόμενο πλάτος γράμματος.

Η παραπάνω μεθοδολογία προϋποθέτει ότι κάθε οριζόντια ή κάθετη γραμμή είναι ανεξάρτητο συνδεδεμένο συστατικό της εικόνας. Αυτό όμως δεν συμβαίνει όταν έχουμε οριζόντιες και κάθετες γραμμές που εφάπτονται ή γραμμές που ορίζουν πίνακες κ.λ.π. Στην περίπτωση αυτή μπορούμε να ακολουθήσουμε την παρακάτω διαδικασία (Gatos 1999):

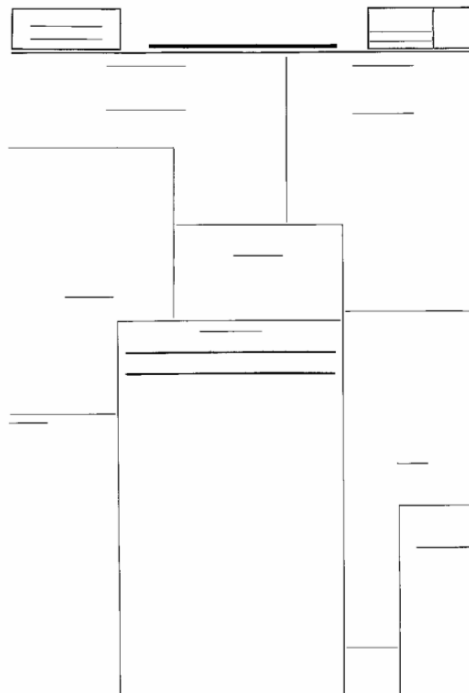
- **ΒΗΜΑ 1:** Δειγματοληψία της εικόνας διατηρώντας τα pixels εικόνας. Κάθε pixel της εικόνας προβάλλεται σε εικόνα χαμηλότερης διάστασης έτσι ώστε ακόμα και διακεκομμένες γραμμές μετασχηματίζονται σε συνεχόμενες.
- **ΒΗΜΑ 2:** Από την εικόνα που προκύπτει εξάγουμε δύο gray scale εικόνες ImV και ImH , επιστρέφοντας για κάθε pixel εικόνας το μήκος της κάθετης (για την ImV) ή οριζόντιας γραμμής (για την ImH) στην οποία ανήκει.
- **ΒΗΜΑ 3:** Οι γραμμές τις οποίες αναζητούμε χαρακτηρίζονται από ένα ελάχιστο μήκος και ένα μέγιστο πάχος. Από τις εικόνες του βήματος 2 κρατάμε τα σημεία τα οποία ανήκουν σε ευθείες των οποίων το μήκος και το πάχος είναι αντίστοιχα μεγαλύτερο και μικρότερο από δεδομένο όριο. Με αυτό τον τρόπο εξάγουμε τις δυαδικές εικόνες LH και FV .
- **ΒΗΜΑ 4:** Οι οριζόντιες και κάθετες γραμμές προκύπτουν ακολουθώντας το περίγραμμα των αντικειμένων των εικόνων LH και FV .

Η μέθοδος έχει ανοχή σε μικρή κλίση των ευθειών καθώς και σε διακεκομμένες ή γραμμές με ασυνέχειες. Παράδειγμα εφαρμογής της μεθόδου σε σελίδα εφημερίδας δίδεται στο σχήμα 5.28.

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων



(α)



(β)

Σχήμα 5.28. (α) Αρχική εικόνα. (β) Εντοπισμένες οριζόντιες και κάθετες γραμμές.

Περιοχές κειμένου/εικόνων/γραφικών

Για τον εντοπισμό των περιοχών κειμένου, εικόνων και γραφικών υπολογίζουμε για κάθε περιοχή τις παρακάτω παραμέτρους (Sauvola 1995):

α. Ποσοστό μαύρων pixels κανονικοποιημένο από 0 μέχρι 1 :

$$B = ABP/E \quad (5.15)$$

όπου ABP είναι το σύνολο των μαύρων pixels (ABP: Amount of Black Pixels) και E το σύνολο των pixels της περιοχής.

β. Διασυσχέτιση του σήματος (signal cross-correlation): Υπολογίζεται μία κανονικοποιημένη διασυσχέτιση της πληροφορίας μεταξύ των οριζόντιων γραμμών σε απόσταση y και $y+a$ από το πάνω όριο των περιοχών.

$$C(a, y) = 1 - \frac{2}{M} \sum_{k=0}^{M-1} (I(k, y) \text{ XOR } I(k, y + a)) \quad (5.16)$$

όπου $I(x, y)$ η ασπρόμαυρη εικόνα, M το πλάτος της περιοχής που εξετάζουμε, a η κατακόρυφη απόσταση μεταξύ των κατακόρυφων γραμμών των οποίων τις πληροφορίες συσχετίζουμε. Η διασυσχέτιση $C(a, y)$ παίρνει τιμές από -1 (οι δύο οριζόντιες γραμμές είναι εντελώς διαφορετικές) μέχρι 1 (οι δύο οριζόντιες γραμμές είναι ίδιες). Η συνολική διασυσχέτιση προκύπτει από την μέση τιμή των τιμών $C(a, y)$ για όλες τις οριζόντιες γραμμές της περιοχής.

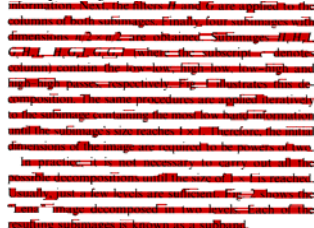
Στα σχήματα 5.29-5.30 δίδονται παραδείγματα υπολογισμού του ποσοστού των μαύρων pixels κανονικοποιημένο από 0 μέχρι 1 (B) καθώς της διασυσχέτισης του σήματος (C) για περιοχές με κείμενα, εικόνες και γραφικά, χωρίς εξομάλυνση, με οριζόντια εξομάλυνση και με οριζόντια και κάθετη εξομάλυνση. Παρατηρούμε ότι το B παίρνει μικρές τιμές (< 0.3) για τις

B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

περιοχές κειμένου και γραφικών. Επίσης το C παίρνει μικρές τιμές (< 0.4) μόνο για τις περιοχές κειμένου μετά από οριζόντια εξομάλυνση. Οι δύο αυτές συνθήκες μας βοηθούν να ξεχωρίζουμε τις περιοχές κειμένου, εικόνων και γραφικών στις περισσότερες περιπτώσεις.

information. Next, the filters H and G are applied to the columns of both subimages. Finally, four subimages with dimensions $n/2 \times n/2$ are obtained. Subimages $H, H, I, G, H, I, G, G, I$ (where the subscript c denotes column) contain the low-low, high-low, low-high and high-high passes, respectively. Fig. 1 illustrates this decomposition. The same procedures are applied iteratively to the subimage containing the most low band information until the subimage's size reaches 1×1 . Therefore, the initial dimensions of the image are required to be powers of two.

In practice, it is not necessary to carry out all the possible decompositions until the size of 1×1 is reached. Usually, just a few levels are sufficient. Fig. 2 shows the "Lena" image decomposed in two levels. Each of the resulting subimages is known as a subband.



information. Next, the filters H and G are applied to the columns of both subimages. Finally, four subimages with dimensions $n/2 \times n/2$ are obtained. Subimages $H, H, I, G, H, I, G, G, I$ (where the subscript c denotes column) contain the low-low, high-low, low-high and high-high passes, respectively. Fig. 1 illustrates this decomposition. The same procedures are applied iteratively to the subimage containing the most low band information until the subimage's size reaches 1×1 . Therefore, the initial dimensions of the image are required to be powers of two.

In practice, it is not necessary to carry out all the possible decompositions until the size of 1×1 is reached. Usually, just a few levels are sufficient. Fig. 2 shows the "Lena" image decomposed in two levels. Each of the resulting subimages is known as a subband.

B=0.12
C=0.69

B=0.40
C=0.33

B=0.93
C=0.95

Οποιοδήποτε πληροφορία [π.χ. ένα αρχείο εικόνας ή ένα αρχείο ήχου] μπορεί να περιγραφεί με μια διαδοχή των ψηφίων 0 και 1 - με βάση το **δυαδικό σύστημα** (binary system), καθώς από τα οποία καλείται **bit**. Μια σειρά δε από ψηφία καλείται **λέξη**. Ο υπολογιστής μπορεί να "περιγράψει" οποιαδήποτε πληροφορία με λέξεις μήκους 8bits έκαστα, οι συνδυασμοί των οποίων είναι $2^8=256$. Με άλλα λόγια, ο υπολογιστής είναι σε θέση να διακρίνει ανάμεσα σε 256 διαφορετικές τιμές. Σε ένα πρόγραμμα επεξεργασίας εικόνας οι τιμές αυτές θα μπορούσαν να αντιστοιχούν σε 256 διαβαθμίσεις φωτεινότητας για κάθε pixel μιας εικόνας. Ένα **byte** (8bits) είναι η πιο κοινή μονάδα μέτρησης της μνήμης ενός υπολογιστή. Αν πάρουμε ως παράδειγμα μία ψηφιακή εικόνα που έχει πληροφορία μόνο για τόνοους του γκρι, δηλαδή μονοχρωματική πληροφορία [εικ. 2]. Η πιο πιθανή εκδοχή είναι κάθε pixel της εικόνας να καταλαμβάνει 1byte (8bits). Έτσι, όπως εζηγήσαμε προηγουμένως, κάθε pixel της εικόνας μπορεί να πάρει μία από 256 διαφορετικές τιμές τόνικης διαβαθμίσης. Στην τιμή 0 αντιστοιχεί το τελείως μαύρο (ο πιο σκοτεινός τόνος) και στην τιμή 255 το τελείως λευκό. Φανταστείτε μία κλίμακα 256 τόνικων διαβαθμίσεων, η οποία ξεκινά από το τελείως μαύρο (τιμή 0) και καταλήγει στο τελείως λευκό (τιμή 255) [εικ. 2].

B=0.15
C=0.58

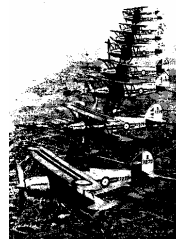
B=0.41
C=0.32

B=0.96
C=0.96

Σχήμα 5.29. Υπολογισμός ποσοστού μαύρων pixels κανονικοποιημένο από 0 μέχρι 1 (B) και διασυσχέτισης του σήματος (C) για περιοχές με κείμενο οι οποίες είναι: χωρίς εξομάλυνση, με οριζόντια εξομάλυνση καθώς και με οριζόντια και κάθετη εξομάλυνση.



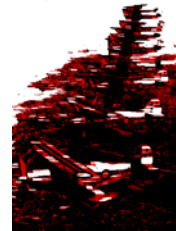
B=0.52
C=0.56



B=0.54
C=0.63



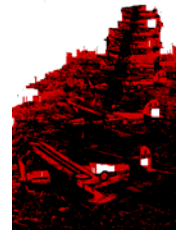
B=0.76
C=0.68



B=0.68
C=0.85



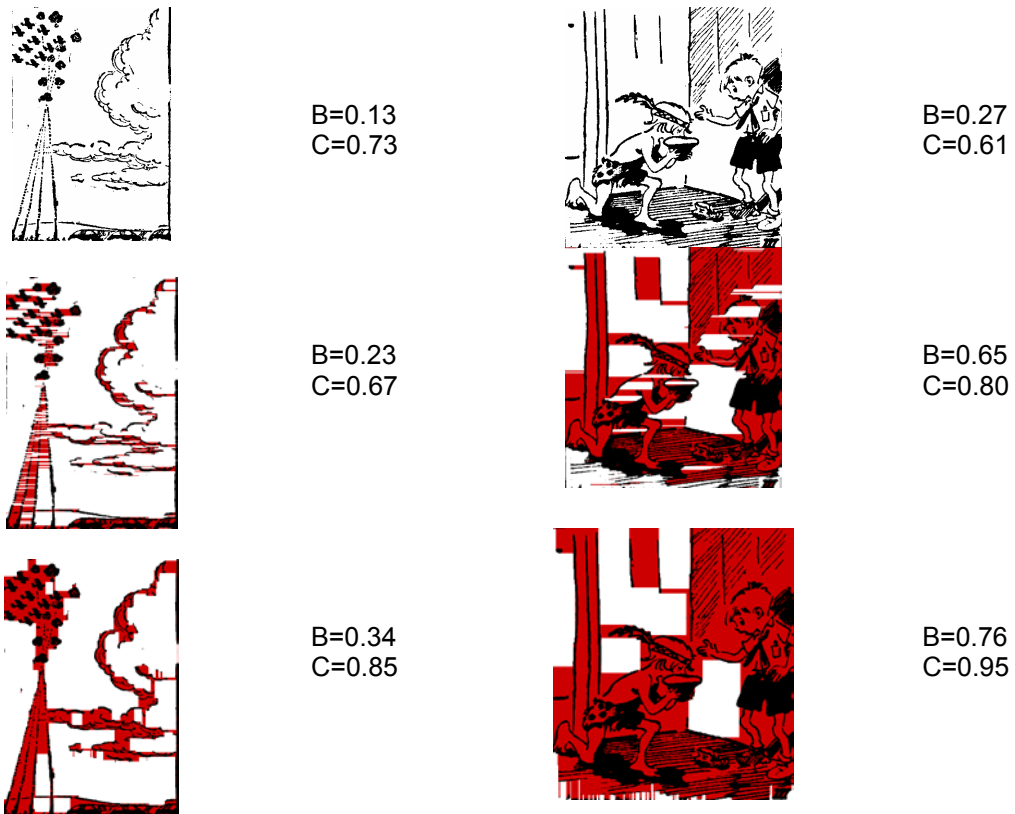
B=0.91
C=0.96



B=0.74
C=0.98

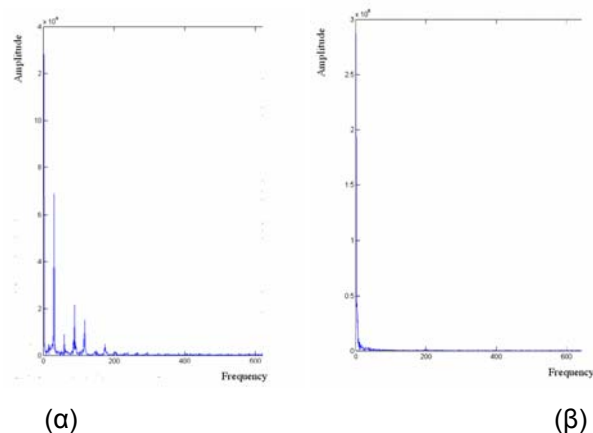
Σχήμα 5.30. Υπολογισμός ποσοστού μαύρων pixels κανονικοποιημένο από 0 μέχρι 1 (B) και διασυσχέτισης του σήματος (C) για περιοχές με εικόνες οι οποίες είναι: χωρίς εξομάλυνση, με οριζόντια εξομάλυνση καθώς και με οριζόντια και κάθετη εξομάλυνση.

B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων



Σχήμα 5.31. Υπολογισμός ποσοστού μαύρων pixels κανονικοποιημένο από 0 μέχρι 1 (B) και διασυσχέτισης του σήματος (C) για περιοχές με γραφικά οι οποίες είναι: χωρίς εξομάλυνση, με οριζόντια εξομάλυνση καθώς και με οριζόντια και κάθετη εξομάλυνση.

Μία διαφορετική προσέγγιση για τον εντοπισμό των περιοχών κειμένου βασίζεται στον υπολογισμό του Fast Fourier Transform (FFT) των οριζόντιων προβολών των μαύρων pixels των τμηματοποιημένων περιοχών (Gatos 1999). Οι οριζόντιες προβολές των περιοχών κειμένου παρουσιάζουν μία περιοδική δομή λόγω της ύπαρξης των ισαπεχόντων γραμμών κειμένου. Μετασχηματίζοντας αυτές τις προβολές στο πεδίο της συχνότητας χρησιμοποιώντας τον FFT μετασχηματισμό παρατηρούμε ότι επικρατούν συγκεκριμένες μη μηδενικές συχνότητες. Από την άλλη πλευρά, οι οριζόντιες προβολές των εικόνων ή των σχημάτων δεν παρουσιάζουν περιοδική δομή και κατά συνέπεια ο μετασχηματισμός τους στο πεδίο της συχνότητας δεν καταλήγει στην παρουσία συχνοτήτων που επικρατούν. Στο σχήμα 5.32 δίδεται ένα παράδειγμα του FFT για μία περιοχή κειμένου άλλα και για μια περιοχή εικόνας.



Σχήμα 5.32. Το FFT πεδίο για μια περιοχή κειμένου (α) και μια περιοχή εικόνας (β). Στο (α) η επικρατούσα συχνότητα ανιχνεύεται στα 30 Hz, ενώ στο (β) δεν υπάρχουν υψηλές συχνότητες.

Βιβλιογραφία

(Duda 1972) Duda, R.D, Hart, P.E: Use of the Hough transform to detect lines and curves in pictures. *Comm. ACM* 15 (1972) 11-15

(Gatos 1999) Gatos, B., Mantzaris, S. L., Chandrinou, K. V., Tsigris, A., Perantonis, S. J.: Integrated Algorithms for Newspaper Page Decomposition and Article Tracking. *Proc. of the Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, Bangalore, India (1999) 559-562

(Gatos 2000) Gatos, B., Mantzaris, S. L.: A novel recursive algorithm for area location using isothetic polygons. *Proc. of the 15th International Conference on Pattern Recognition (ICPR2000)*, Barcelona, Spain (2000) 496-499.

(Gorman 1993) O'Gorman, L.: The Document Spectrum for Page Layout Analysis. *IEEE Trans. On Patt. Anal. and Mach. Intell.*, Vol. 15, No. 11 (1993) 1162-1173

(Nagy 1984) Nagy, G., Seth, S.: Hierarchical representation of optically scanned documents. *Proc. 7th Int. Conf. Patt. Recogn.* (1984) 347-349

(Kasturi 1990) Kasturi, R., Bow, S. T., El-Masri, W., Shah, J., Gattiker, J. R., Mokate, U.: A System for Interpretation of Line Drawings. *IEEE Trans. On Pattern Analysis And Mach. Intell.*, Vol. 12, No. 10 (1990) 978-991

(Papamarkos 1996) Papamarkos, N. Tzortzakis, J., Gatos, B.: Determination of Run-Length smoothing values for document segmentation, *Proc. of the Third IEEE Int. Conf. On Electronics, Circuits and Systems, ICECS 96* (1996) 684-687

(Pavlidis 1981) Pavlidis, T.: *Algorithms for Graphics and Image Processing*. Springer-Verlag (1981)

(Sauvola 1995) Sauvola J., Pietikainen, M.: Page segmentation and classification using fast feature extraction and connectivity analysis. *Proc. of the Third International Conference on Document Analysis and Recognition (ICDAR'95)*, (1995) 1127-1131

(Shapiro 2001) Shapiro L., Stockman, G.: *Computer Vision*. Prentice Hall (2001)

(Theodoridis 1997) Theodoridis, S., Koutroumbas K.: *Pattern Recognition*, Academic Press (1997)

(Wahl 1982) Wahl, F.M., Wong, K.Y., Casey R.G.: Block Segmentation and Text Extraction in Mixed Text/Image Documents. *Computer Graphics and Image Processing*, 20 (1982) 375-390

(Waked 2001) Waked, B.; Ching Y Suen; Bergler, S.: Segmenting document images using diagonal white runs and vertical edges. *Sixth International Conference on Document Analysis and Recognition* (2001) 194 - 199

(Witten 1999) Witten, I. H., Moffat, A., Bell, T.C.: *Managing Gigabytes, Second Edition* (1999) Morgan Kaufmann Publishers

(Yin 2001) Yin P. Y.: Skew detection and block classification of printed documents. *Image and Vision Computing* 19 (2001) 567-579