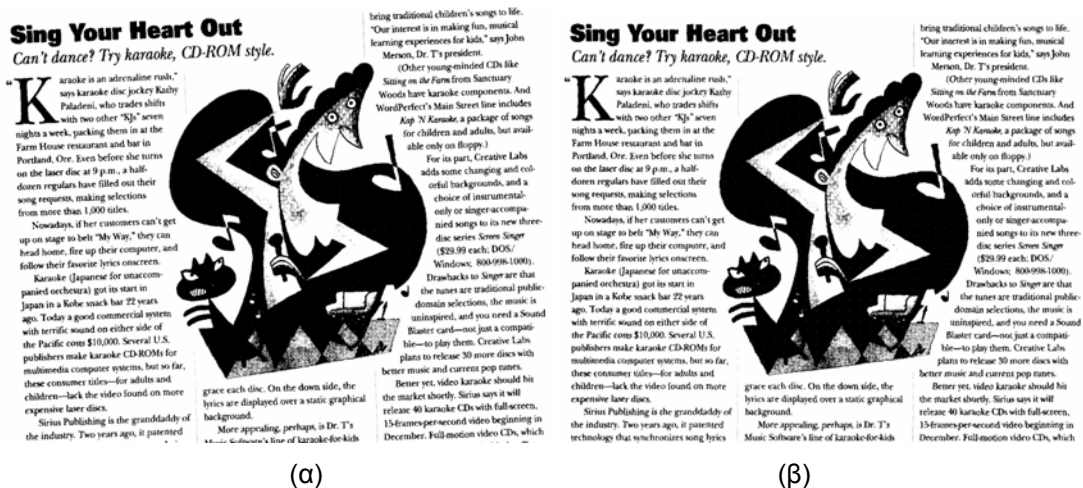


Μάθημα 4

Διόρθωση στροφής

4.1 Το πρόβλημα της στροφής των εγγράφων

Ένα από τα βασικά βήματα της προεπεξεργασίας του εγγράφου είναι ο εντοπισμός και η διόρθωση της στροφής του (σχήμα 4.1). Η στροφή αυτή συνήθως προέρχεται από μη ευθυγραμμισμένη τοποθέτηση του εγγράφου στον ψηφιακό μετατροπέα. Το πρόβλημα αυτό είναι πολύ συνηθισμένο κυρίως κατά την σάρωση φαξ ή συρραμμένων εντύπων ή βιβλίων καθώς και κατά την μαζική σάρωση πολλών εγγράφων με χρήση αυτόματου τροφοδότη. Η διόρθωση της στροφής του εγγράφου είναι ένα απαραίτητο στάδιο πριν προχωρήσουμε στα επόμενα στάδια κατάμησης και αναγνώρισης της εικόνας. Υπάρχουν τεχνικές κατάμησης και αναγνώρισης που δεν προϋποθέτουν την διόρθωση της στροφής του κειμένου, όμως προϋποθέτουν μεγάλο υπολογιστικό κόστος χωρίς να δίνουν σίγουρο αποτέλεσμα. Η στροφή που αναζητούμε στα έγγραφα δεν είναι μεγάλη και μπορούμε να πούμε ότι δεν ξεπερνά τις 5 μοίρες. Οι προτεινόμενες μέθοδοι για την εξαγωγή της στροφής των εγγράφων βασίζονται κυρίως στον μετασχηματισμό Hough, στην ανάλυση των προβολών της εικόνας σε διάφορες διευθύνσεις καθώς και στον πίνακα διασυσχέτισης των πληροφοριών της εικόνας που υπάρχουν σε σύνολο ισαπέχοντων κάθετων ευθειών. Αν έχουμε έγγραφα οριζόντιου και κάθετου προσανατολισμού (portrait – landscape) θα πρέπει πρώτα να εντοπίσουμε τον προσανατολισμό του εγγράφου.



Σχήμα 4.1. Παράδειγμα διόρθωσης στροφής εγγράφου. (α) Αρχική ασπρόμαυρη εικόνα. (β) Τελική εικόνα μετά την διόρθωση της στροφής.

4.2 Εύρεση προσανατολισμού του εγγράφου

Υπάρχουν δύο δυνατοί προσανατολισμοί των εγγράφων, ο οριζόντιος (portrait) και ο κάθετος (landscape). Ο εντοπισμός του προσανατολισμού είναι απαραίτητος πριν προχωρήσουμε στην αναζήτηση και διόρθωση της στροφής των εγγράφων. Μία απλή και αποτελεσματική μέθοδος βασίζεται στην οριζόντια και κάθετη εξομάλυνση της εικόνας (Yin 2001). Οι τεχνικές εξομάλυνσης της εικόνας σκοπεύουν στο γέμισμα περιοχών του υποβάθρου ώστε να

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

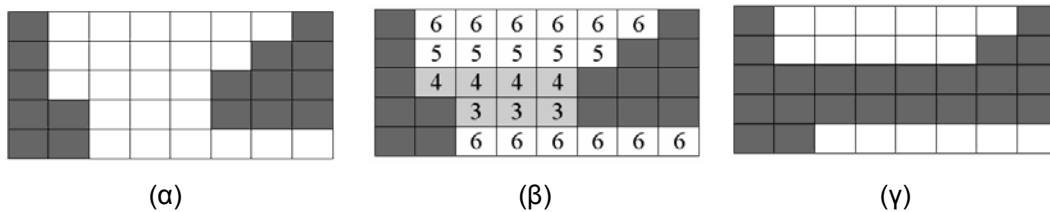
ενωθούν τμήματα της εικόνας που ανήκουν στην ίδια κατηγορία. Η πιο γνωστή τεχνική εξομάλυνσης είναι ο RLSA (Run Length Smoothing Algorithm – Αλγόριθμος εξομάλυνσης διαδοχικών pixels) (Wahl 1982). Για τον ορισμό του RLSA χρειάζονται να οριστούν δύο βασικοί παράμετροι: το μέγιστο μήκος διαδοχικών pixels υποβάθρου T_{max} και η διεύθυνση σάρωσης. Η εικόνα εξετάζεται ως προς την διεύθυνση σάρωσης που έχει οριστεί και τα διαδοχικά pixels υποβάθρου με μήκος μικρότερο από το μέγιστο μήκος διαδοχικών pixels υποβάθρου T_{max} μετατρέπονται σε σημεία εικόνας. 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 4.2.

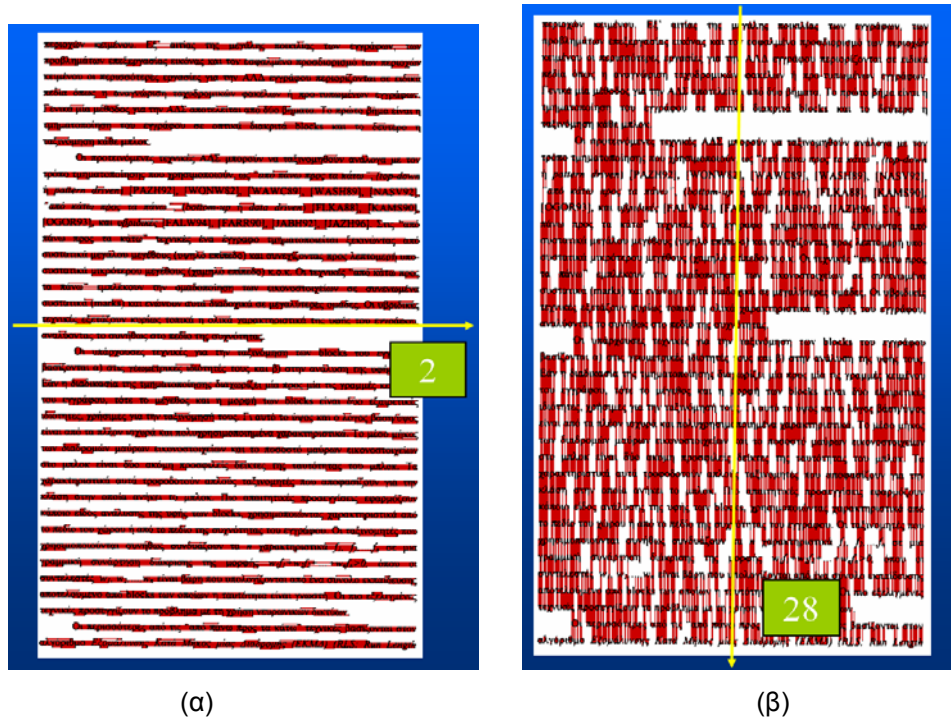


Σχήμα 4.2. Εφαρμογή οριζόντιου RLSA με $T_{max} = 5$ σε ασπρόμαυρη εικόνα. (α) Αρχική εικόνα. (β) Υπολογισμός του μήκους των διαδοχικών pixels υποβάθρου. (γ) Μετατροπή όλων των σημείων υποβάθρου με μήκος μικρότερο από T_{max} σε σημεία εικόνας.

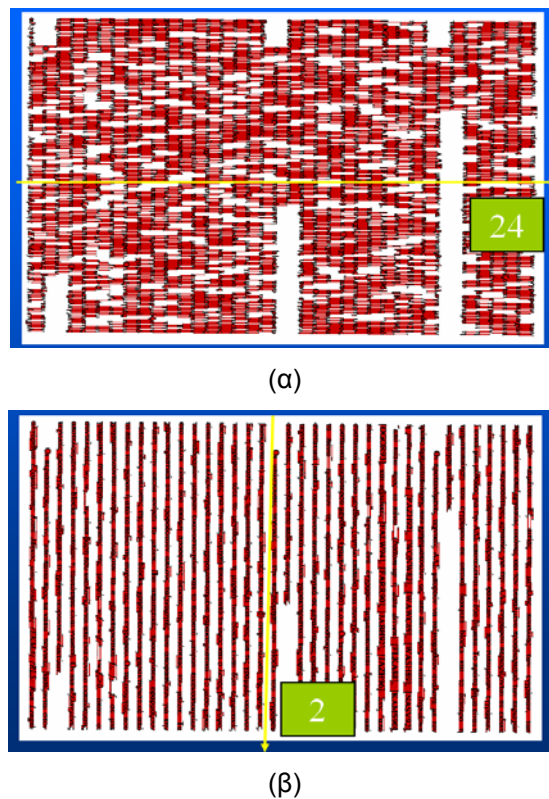
Για τον εντοπισμό του προσανατολισμού του εγγράφου ακολουθούμε τα παρακάτω βήματα:

- **Βήμα 1:** Εφαρμόζουμε οριζόντιο RLSA στην εικόνα και υπολογίζουμε το NoH που είναι όλες οι οριζόντιες μεταβάσεις από 0 σε 1 στην τελική εικόνα.
- **Βήμα 2:** Εφαρμόζουμε κάθετο RLSA στην εικόνα και υπολογίζουμε το NoV που είναι όλες οι κάθετες μεταβάσεις από 0 σε 1 στην τελική εικόνα.
- **Βήμα 3:** Αν $NoH < NoV$, τότε το έγγραφο έχει οριζόντιο προσανατολισμό (portrait), αλλιώς έχει κάθετο προσανατολισμό (landscape).

Στα παραδείγματα, στα σχήματα 4.3 και 4.4 δίδονται παραδείγματα εφαρμογής της μεθόδου σε οριζόντιο και σε κάθετο έγγραφο. Στο σχήμα 4.3α όπου το έγγραφο είναι οριζόντιο, εφαρμόζουμε οριζόντια εξομάλυνση. Σε μια τυχαία γραμμή της εικόνας οι μεταβάσεις από 0 σε 1 είναι 2. Στο σχήμα 4.3β εφαρμόζουμε κάθετη εξομάλυνση. Σε μια τυχαία στήλη της εικόνας οι μεταβάσεις από 0 σε 1 είναι 28. Αν υπολογίσουμε το NoH , το οποίο είναι το συνολικό άθροισμα των οριζόντιων μεταβάσεων από 0 σε 1 για την οριζόντια εξομάλυνση και το NoV , το οποίο είναι το συνολικό άθροισμα των κάθετων μεταβάσεων από 0 σε 1 για την κάθετη εξομάλυνση, ισχύει $NoH < NoV$. Αντίστοιχα, στο σχήμα 4.4α όπου το έγγραφο είναι κάθετο, εφαρμόζουμε οριζόντια εξομάλυνση. Σε μια τυχαία γραμμή της εικόνας οι μεταβάσεις από 0 σε 1 είναι 24. Στο σχήμα 4.4β εφαρμόζουμε κάθετη εξομάλυνση. Σε μια τυχαία στήλη της εικόνας οι μεταβάσεις από 0 σε 1 είναι 2. Αν υπολογίσουμε το NoH , το οποίο είναι το συνολικό άθροισμα των οριζόντιων μεταβάσεων από 0 σε 1 για την οριζόντια εξομάλυνση και το NoV , το οποίο είναι το συνολικό άθροισμα των κάθετων μεταβάσεων από 0 σε 1 για την κάθετη εξομάλυνση, ισχύει $NoH > NoV$.



Σχήμα 4.3. Εφαρμογή της μεθόδου εύρεσης του προσανατολισμού σε οριζόντιο (portrait) έγγραφο (α) Εφαρμογή οριζόντιου RLSA και υπολογισμός των οριζόντιων μεταβάσεων από 0 σε 1 για μια τυχαία γραμμή της εικόνας. (β) Εφαρμογή κάθετου RLSA και υπολογισμός των κάθετων μεταβάσεων από 0 σε 1 για μια τυχαία στήλη της εικόνας.



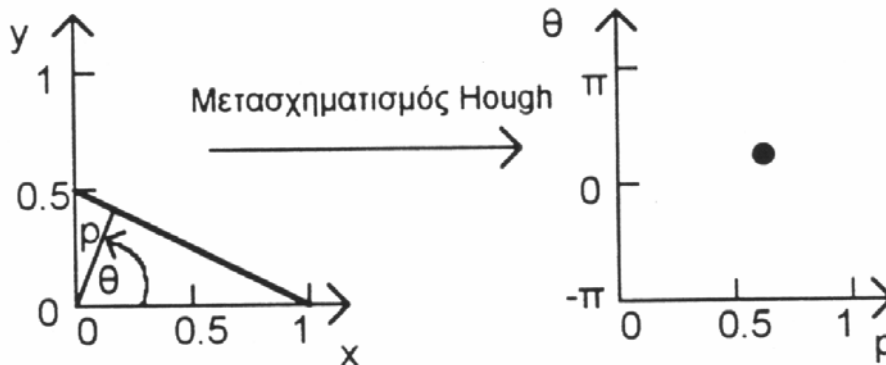
Σχήμα 4.4. Εφαρμογή της μεθόδου εύρεσης του προσανατολισμού σε κάθετο (landscape) έγγραφο (α) Εφαρμογή οριζόντιου RLSA και υπολογισμός των οριζόντιων μεταβάσεων από 0 σε 1 για μια τυχαία γραμμή της εικόνας. (β) Εφαρμογή κάθετου RLSA και υπολογισμός των κάθετων μεταβάσεων από 0 σε 1 για μια τυχαία στήλη της εικόνας.

4.3 Μετασχηματισμός Hough

Ο μετασχηματισμός Hough αντιστοιχεί τα σημεία μιας ευθείας του επιπέδου x-y σε ένα σημείο (ρ, θ) του επιπέδου των πολικών συντεταγμένων (Duda 1972). Μια ευθεία του επιπέδου x-y περιγράφεται από τη σχέση:

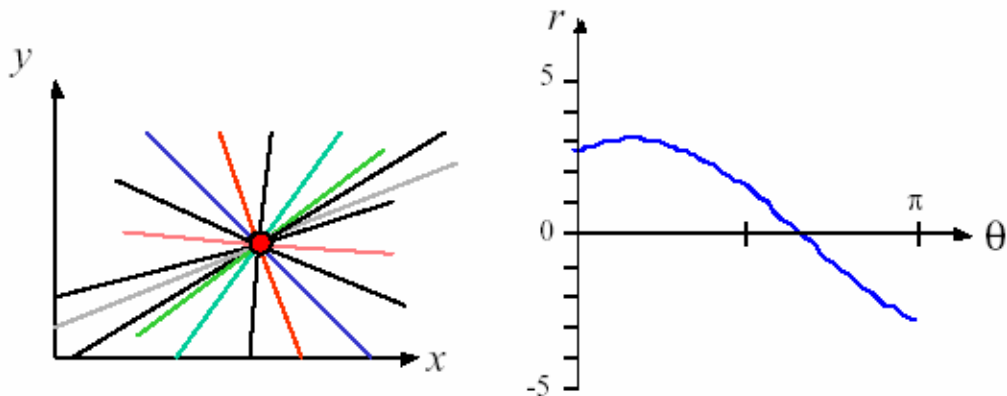
$$\rho = x \cos\theta + y \sin\theta \quad (4.1)$$

όπου ρ η κάθετη απόσταση της αρχής των αξόνων από την ευθεία και θ η γωνία που σχηματίζει η κάθετη στην ευθεία με τον άξονα των x. Όπως φαίνεται και στο σχήμα 4.5, ο μετασχηματισμός Hough μιας ευθείας του επιπέδου x-y είναι το σημείο (ρ, θ) του επιπέδου των πολικών συντεταγμένων.

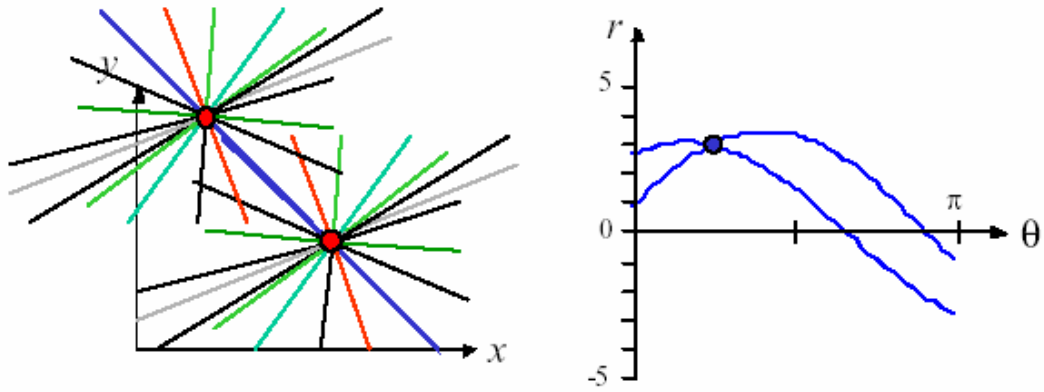


Σχήμα 4.5. Μετασχηματισμός Hough μιας ευθείας του επιπέδου x-y στο σημείο (ρ, θ) του επιπέδου των πολικών συντεταγμένων.

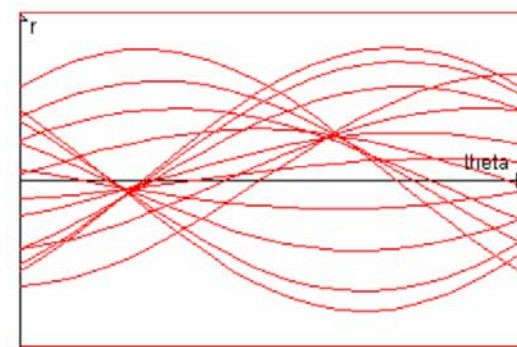
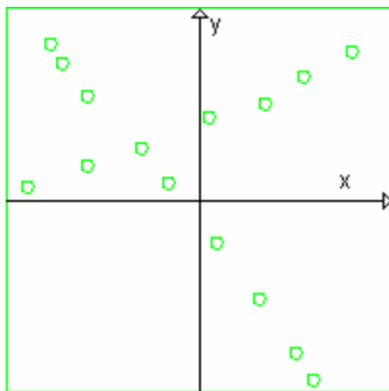
Για όλα τα σημεία της εικόνας υπολογίζεται ο μετασχηματισμός Hough όλων των ευθειών που περνούν από αυτά. Στο σχήμα 4.6 φαίνονται κάποια παραδείγματα της εφαρμογής του μετασχηματισμού Hough για εικόνες που αποτελούνται από ένα, δύο ή πιο πολλά σημεία. Στην πρώτη περίπτωση που έχουμε ένα μόνο σημείο, το πεδίο του Hough αντιστοιχεί σε μία ημιτονοειδής καμπύλη. Στην δεύτερη περίπτωση όπου έχουμε δύο σημεία, το πεδίο του Hough αντιστοιχεί σε δύο ημιτονοειδής καμπύλες οι οποίες τέμνονται στο σημείο (ρ, θ) που αντιστοιχεί στην ευθεία που ενώνει τα δύο σημεία. Στην τρίτη περίπτωση όπου έχουμε πολλά σημεία, το πεδίο του Hough παρουσιάζει την μορφή πολλών ημιτονοειδών καμπυλών οι οποίες τέμνονται σε δύο σημεία που αντιστοιχούν και στις επικρατέστερες ευθείες που ορίζουν τα σημεία της εικόνας.



(α)



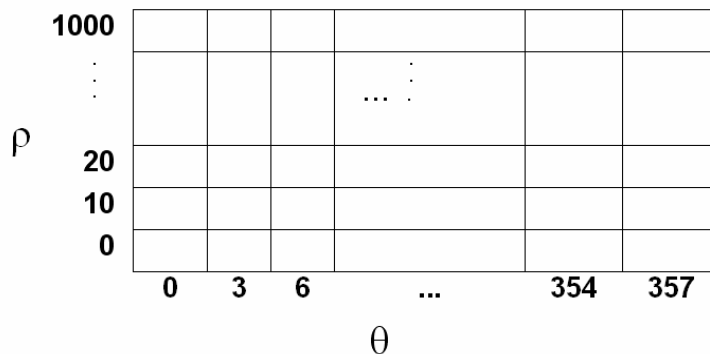
(β)



(γ)

Σχήμα 4.6. Μετασχηματισμός Hough μιας εικόνας που αποτελείται από ένα (α), από δύο (β) και από πολλά σημεία (γ).

Όλα τα εξαγόμενα ζευγάρια τιμών (ρ, θ) αποθηκεύονται σε ένα πίνακα συσσώρευσης ο οποίος δημιουργείται με κατάλληλο βήμα κβάντισης των ρ, θ . Στο σχήμα 4.7 φαίνεται ένα παράδειγμα πίνακα συσσώρευσης όπου η γωνία λαμβάνεται με βήμα 3 μοίρες και το ρ με βήμα 10. Οι τιμές των (ρ, θ) που αντιστοιχούν στα μέγιστα του πίνακα συσσώρευσης δίνουν τις επικρατέστερες ευθείες της εικόνας. Επειδή η κλίση των ευθειών του κειμένου αντιστοιχεί και στη συνολική κλίση του εγγράφου μπορούμε να δεχθούμε ότι το μέγιστο των αθροισμάτων του πίνακα συσσώρευσης ως προς θ αντιστοιχεί στην κλίση του εγγράφου.



Σχήμα 4.7. Παράδειγμα πίνακα συσσώρευσης.

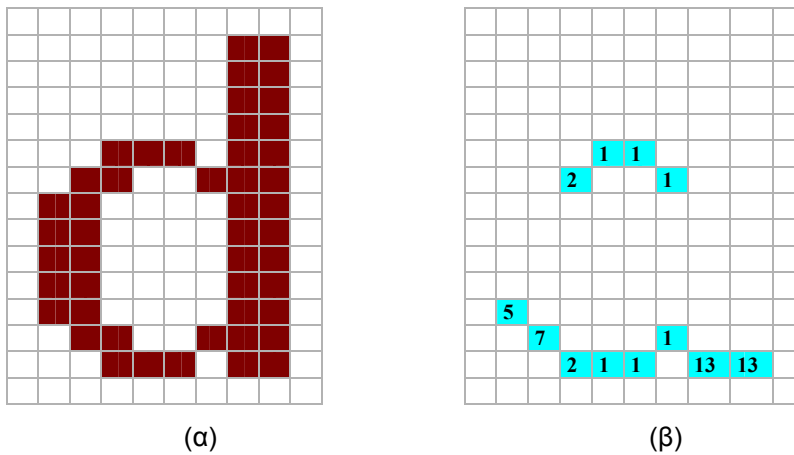
Εκτός από τον προσδιορισμό της κλίσης μιας εικόνας κειμένου, ο μετασχηματισμός Hough χρησιμοποιείται για τον εντοπισμό των ευθειών οι οποίες υπάρχουν ανάμεσα σε μεγάλο όγκο άλλων αντικειμένων. Με κατάλληλη προσαρμογή ο μετασχηματισμός Hough μπορεί να χρησιμοποιηθεί για τον εντοπισμό κύκλων, ελλείψεων κ.λ.π. Στην γενικευμένη του μορφή ο μετασχηματισμός Hough μπορεί να χρησιμοποιηθεί για τον εντοπισμό απλών σχημάτων. Τα

Β. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

προβλήματα που πρέπει να αντιμετωπιστούν προκειμένου να εφαρμοστεί ο μετασχηματισμός Hough είναι:

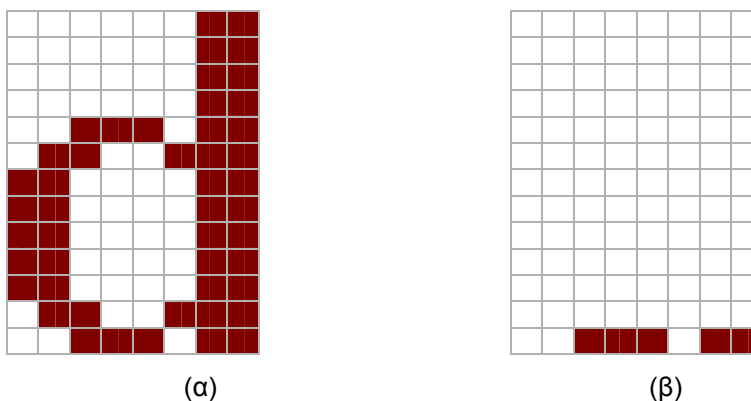
- η κατάλληλη επιλογή κβάντισης του πίνακα συσσώρευσης ως προς ρ και θ , ανάλογα με την ανάλυση της εικόνας και τον σκοπό χρήσης του μετασχηματισμού και
- η μεγάλη καθυστέρηση που εισάγεται όταν έχουμε εικόνες οι οποίες αποτελούνται από πολλά σημεία ή έχουν θόρυβο.

Για την μείωση του υπολογιστικού κόστους προσδιορισμού του μετασχηματισμού Hough κατά την ανίχνευση της στροφής της εικόνας προτείνεται η μείωση του αριθμού των πληροφοριών που συμμετέχουν στον μετασχηματισμό υπολογίζοντας τα οριζόντια ή κάθετα διαδοχικά pixels της εικόνας (Hinds 1990). Η εικόνα μετασχηματίζεται υπολογίζοντας τα διαδοχικά σημεία της εικόνας σε διεύθυνση κάθετη στις γραμμές κειμένου και τοποθετώντας τον αριθμό του μήκους των διαδοχικών σημείων στο τελευταίο σημείο προς τα κάτω. Έτσι δημιουργείται μια νέα εικόνα που ονομάζεται «burst image» (σχήμα 4.8). Εφαρμόζοντας τον μετασχηματισμό Hough στην νέα εικόνα («burst image») μειώνεται ο αριθμός των στοιχείων που μετασχηματίζονται και έτσι επιτυγχάνεται η διαδικασία υπολογισμού του πίνακα συσσώρευσης.



Σχήμα 4.8. Παράδειγμα «burst image». (α) Η αρχική εικόνα. (β) Η «burst image» που προκύπτει μετά τον υπολογισμό των διαδοχικών κάθετων σημείων.

Στην ίδια προσπάθεια μείωσης των πληροφοριών που συμμετέχουν στο μετασχηματισμό Hough κατά την ανίχνευση της στροφής της εικόνας εντάσσεται και η προτεινόμενη μεθοδολογία από τον (Thoma 1994). Σύμφωνα με αυτή, επιλέγουμε μόνο ένα παράθυρο της εικόνας και χρησιμοποιούμε για τον μετασχηματισμό Hough μόνο τα κάτω οριακά pixels των μεμονωμένων αντικειμένων (συνδεδεμένων συστατικών) της εικόνας (σχήμα 4.9). Στο παράδειγμα του σχήματος 4.10 φαίνεται ένα επιλεγμένο παράθυρο της εικόνας καθώς και τα σημεία που τελικά συμμετέχουν στο μετασχηματισμό Hough.



Σχήμα 4.9. Παράδειγμα κάτω οριακών pixels. (α) Η εικόνα ενός μεμονωμένου χαρακτήρα. (β) Τα κάτω οριακά pixel του.

numerous video and sound improvements since the first edition in 1989, says Pamela Sogge, marketing product manager for *Compton's Interactive Encyclopedia*. Released in September (with a Mac version following this month), the Windows-based version 3.0 is "more fun," says Sogge. "When you push an icon, there's a little animation. Our first encyclopedias just had [static] buttons." In addition to 53,000 articles, you'll get 8,000 pictures, 3-D animations, and more than 50 videos.

(α)

numerous video and sound improvements since the first edition in 1989, says Pamela Sogge, marketing product manager for *Compton's Interactive Encyclopedia*. Released in September (with a Mac version following this month), the Windows-based version 3.0 is "more fun," says Sogge. "When you push an icon, there's a little animation. Our first encyclopedias just had [static] buttons." In addition to 53,000 articles, you'll get 8,000 pictures, 3-D animations, and more than 50 videos.

(β)

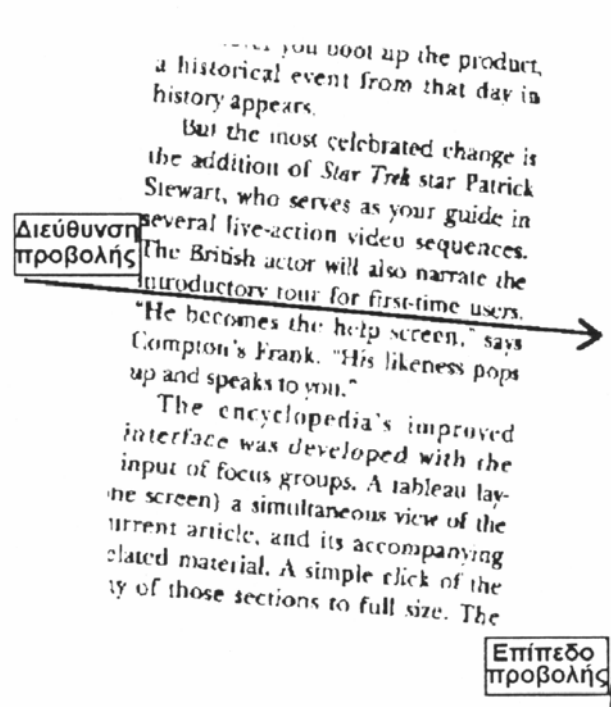
numerous video and sound improvements since the first edition in 1989, says Pamela Sogge, marketing product manager for *Compton's Interactive Encyclopedia*. Released in September (with a Mac version following this month), the Windows-based version 3.0 is "more fun," says Sogge. "When you push an icon, there's a little animation. Our first encyclopedias just had [static] buttons." In addition to 53,000 articles, you'll get 8,000 pictures, 3-D animations, and more than 50 videos.

(γ)

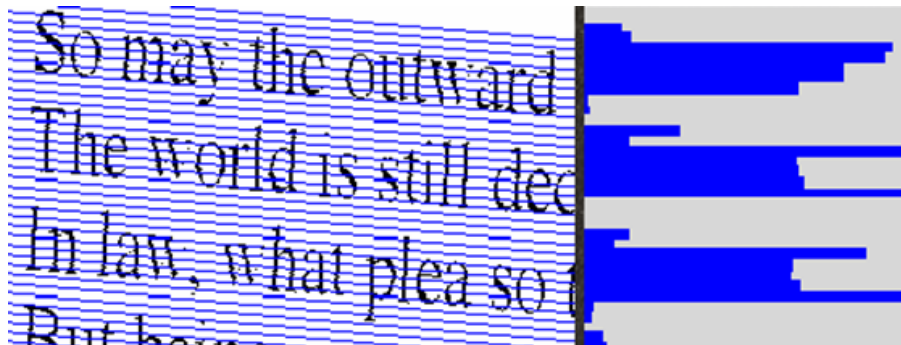
Σχήμα 4.10. Επιτάχυνση του μετασχηματισμού Hough. (α) Επιλέγουμε ένα παράθυρο της εικόνας. (β) Βρίσκουμε τα μεμονωμένα αντικείμενα. (γ) Επιλέγουμε μόνο τα κάτω οριακά pixels για κάθε μεμονωμένο αντικείμενο.

4.4 Ανάλυση των προβολών

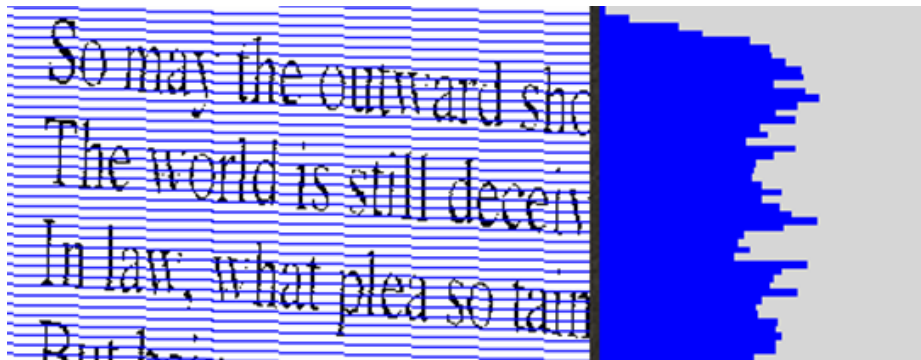
Οι τεχνικές εντοπισμού της στροφής του κειμένου που βασίζονται σε προβολές του εγγράφου, αναλύουν τις προβολές του εγγράφου ή τμημάτων αυτού σε συγκεκριμένες διευθύνσεις (Baird 1987) (σχήμα 4.11). Αν η διεύθυνση προβολής ταυτίζεται με την διεύθυνση της κλίσης του εγγράφου, τότε το ιστόγραμμα της προβολής της εικόνας παρουσιάζει τοπικά μέγιστα τα οποία αντιστοιχούν στις γραμμές του κειμένου (σχήμα 4.12α). Σε αντίθετη περίπτωση το ιστόγραμμα της προβολής της εικόνας δεν παρουσιάζει περιοδικότητα (σχήμα 4.12β).



Σχήμα 4.11. Προβολή εικόνας σε επίπεδο.



(α)



(β)

Σχήμα 4.12. Το ιστόγραμμα της προβολής της εικόνας όταν η διεύθυνση προβολή ταυτίζεται με την κλίση του εγγράφου (α) ή είναι τυχαία (β).

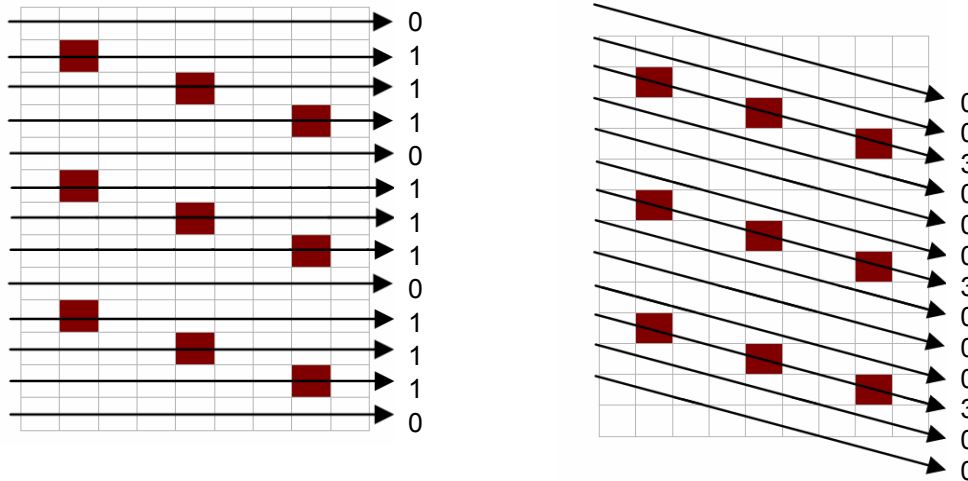
B. Γάτος, Ψηφιακή Επεξεργασία και Αναγνώριση Εγγράφων

Για κάθε γωνία θ της διεύθυνσης προβολής βρίσκουμε την τιμή της συνάρτησης ενέργειας $A(\theta)$:

$$A(\theta) = \sum_{i=1}^m c_i^2(\theta) \quad (4.2)$$

όπου m είναι ο αριθμός κβάντισης του επιπέδου προβολής (κάθετο στην διεύθυνση προβολής) και $c_i(\theta)$ ο αριθμός των σημείων που προβάλλονται στο i -στο σημείο του επιπέδου προβολής όταν έχουμε διεύθυνση προβολής θ . Στο σχήμα 4.13 φαίνεται ο αριθμός των σημείων που προβάλλονται όταν η διεύθυνση προβολής δεν συμπίπτει με την κλίση του εγγράφου και όταν συμπίπτει. Στην πρώτη περίπτωση $A(\theta)=9$ ενώ στην δεύτερη $A(\theta)=27$.

Υπολογίζουμε την συνάρτηση ενέργειας για κάθε γωνία θ , συνήθως από -5 μέχρι 5 μοίρες, χρησιμοποιώντας βήμα $d\theta$. Η γωνία θ για την οποία μεγιστοποιείται η συνάρτηση ενέργειας μας δίνει τη γωνία κλίσης του εγγράφου. Η παραπάνω διαδικασία επιταχύνεται αν αντί να προβάλλουμε ολόκληρη την εικόνα σε διάφορες διευθύνσεις, προβάλλουμε τα κεντρικά σημεία των παραλληλογράμμων που οριοθετούν τα μεμονωμένα αντικείμενα (συνδεδεμένα συστατικά) της εικόνας.



(α)

(β)

Σχήμα 4.13. Ο αριθμός των σημείων που προβάλλονται όταν η διεύθυνση προβολής δεν συμπίπτει με την κλίση του εγγράφου (α) και όταν συμπίπτει (β).

4.5 Πίνακας διασυσχέτισης

Σύμφωνα με αυτή την προσέγγιση (Gatos 1997), επεξεργαζόμαστε την πληροφορία που υπάρχει σε ένα σύνολο ισαπέχοντων κάθετων ευθειών. Δεχόμεθα αρχικά ότι κάθε έγγραφο περιέχει κυρίως οριζόντιες γραμμές κειμένου. Κατόπιν, μελετούμε μόνο τα pixels που βρίσκονται στις τομές των κάθετων ευθειών με τις γραμμές κειμένου (σχήμα 4.14). Με τη χρησιμοποίηση μόνο αυτών των pixels κατασκευάζουμε έναν πίνακα διασυσχέτισης. Ο πίνακας αυτός περιέχει τις αναγκαίες πληροφορίες για την ανίχνευση της κλίσης των εγγράφων. Δεν απαιτούμε να συσχετίσουμε κάθε pixel μιας κάθετης γραμμής με όλα τα pixels των άλλων γραμμών, αλλά μόνο με τα pixels που βρίσκονται σε συγκεκριμένες περιοχές που καθορίζονται από το αναμενόμενο εύρος στροφής (συνήθως -5 μέχρι 5 μοίρες). Με τον τρόπο αυτό, μειώνεται σημαντικά το υπολογιστικό κόστος χωρίς να μειώνεται η τελική ακρίβεια. Τελικά, διαμορφώνουμε μία κάθετη προβολή του πίνακα διασυσχέτισης, το μέγιστο της οποίας μας δίνει τη γωνία στροφής του εγγράφου.

statistical assumptions under which it operates, while both methods suffer from the problem that the number of segments must be known a priori. The method reported here combines a quad-tree smoothing operation with statistical classification performed at the highest level of the quad-tree, followed by a downward directed boundary estimation based on the segments obtained at the top level of the quad-tree. This segmentation is statistical, but is based on a

Σχήμα 4.14. Τομές των κάθετων ευθειών με τις γραμμές κειμένου.

Καταρχήν, υποθέτουμε ότι χρησιμοποιούμε μόνο δύο κάθετες γραμμές d_1 και d_2 σε απόσταση D_1 και D_2 από το αριστερό όριο της εικόνας (σχήμα 4.15). Θέλουμε να ορίσουμε έναν πίνακα ο οποίος θα αποθηκεύει όλες τις σχετικές θέσεις των pixels της κάθετης γραμμής d_1 έναντι της κάθετης γραμμής d_2 . Παρατηρούμε ότι λόγω της στροφής της εικόνας, μια γραμμή κειμένου τέμνει τις δύο κάθετες ευθείες με κατακόρυφη απόσταση λ . Υποθέτοντας ότι η εικόνα μπορεί να στραφεί μέχρι $\pm 5^\circ$, δηλαδή $\theta_{\max} = 5^\circ$, η κατακόρυφη απόσταση λ πρέπει να ικανοποιεί τον περιορισμό:

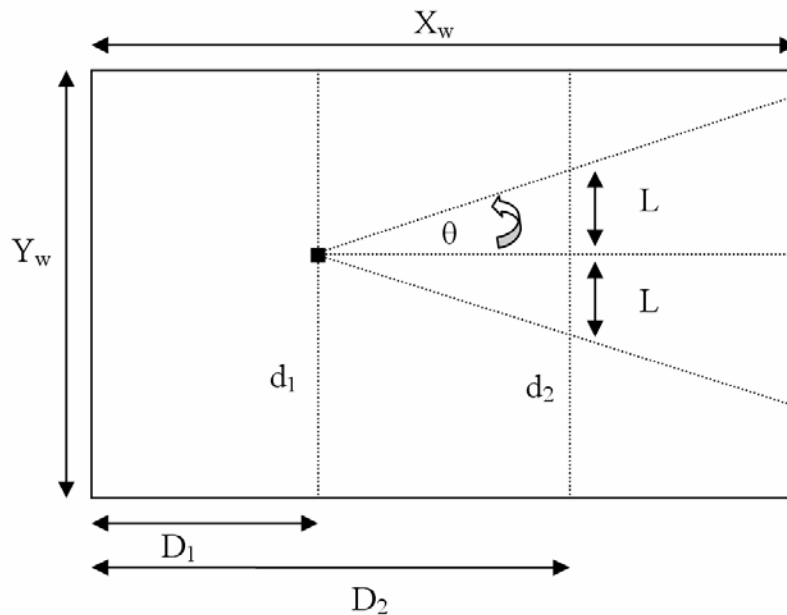
$$-L < \lambda < L, \text{ όπου } L = (D_2 - D_1) \tan(2\pi \theta_{\max} / 360) \quad (4.3)$$

όπου L ακέραιος που εκφράζεται σε αριθμό κάθετων pixels.

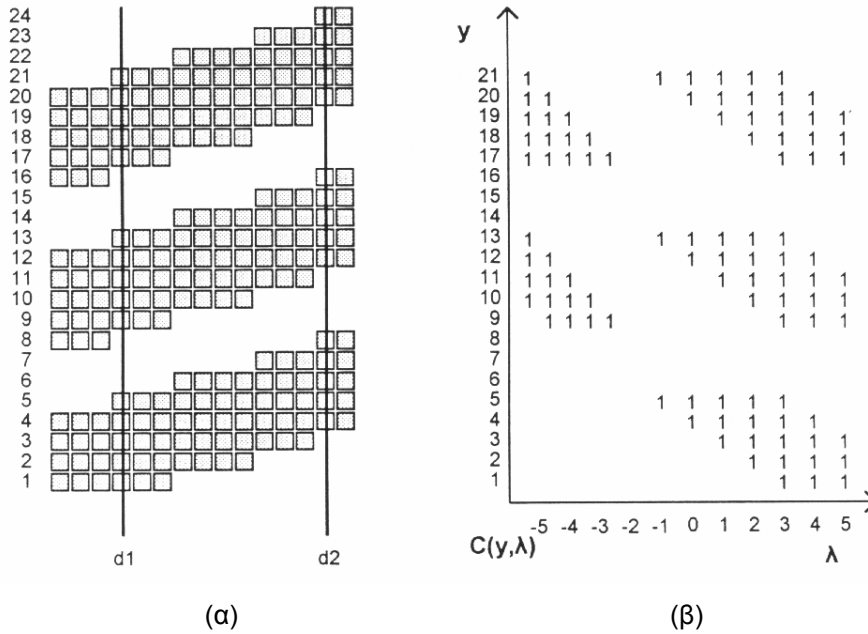
Για κάθε σημείο κειμένου $f(D_1, y)$ της κάθετης ευθείας d_1 ανιχνεύουμε τα σημεία κειμένου στην κάθετη ευθεία d_2 με περιοχή της κατακόρυφης μετατόπισης τα όρια $[-L, L]$, δηλαδή τα pixels στην κάθετη ευθεία d_2 τέτοια ώστε $f(D_2, y + \lambda) = 1$ και $\lambda \in [-L, L]$. Στη συνέχεια αποθηκεύουμε αυτές τις πληροφορίες σε έναν πίνακα διασυσχέτισης $C(y, \lambda) \in \{0, 1\}$ ο οποίος ορίζεται ως εξής:

$$C(y, \lambda) = f(D_1, y) f(D_2, y + \lambda), \text{ όπου } 1 \leq y \leq Y_w \text{ και } -L \leq \lambda \leq L \quad (4.4)$$

Στο παράδειγμα του σχήματος 4.16 φαίνεται ο πίνακας διασυσχέτισης για δύο κάθετες ευθείες σε μία απλή εικόνα. Έχουμε $C(1, 3) = 1$ επειδή υπάρχει pixel στην ευθεία d_1 με $y = 1$, δηλαδή $f(D_1, 1) = 1$, καθώς και στην ευθεία d_2 για $y = 1 + 3 = 4$, δηλαδή $f(D_2, 4)$.



Σχήμα 4.15. Χρήση δύο κάθετων ευθειών d_1 και d_2 .



Σχήμα 4.16. Υπολογισμός του πίνακα διασυσχέτισης των πληροφοριών της γραμμής d_1 ως προς την γραμμή d_2 . (α) Η εικόνα και οι δύο κάθετες γραμμές. (β) Ο πίνακας διασυσχέτισης.

Αν η εικόνα έχει γωνία στροφής θ , τότε κάθε γραμμή κειμένου αποτυπώνεται στις δύο ευθείες με κατακόρυφη απόσταση $(D_2 - D_1)\tan(\theta)$. Έτσι, ο πίνακας διασυσχέτισης C θα έχει μέγιστη συσσώρευση σημείων κατά τον y -άξονα για την τιμή $\lambda = (D_2 - D_1)\tan(\theta)$. Άρα προσδιορίζοντας το καθολικό μέγιστο της κάθετης προβολής του πίνακα διασυσχέτισης C εντοπίζουμε και την στροφή του εγγράφου.

Η προβολή του πίνακα διασυσχέτισης δίδεται από τον τύπο:

$$P(\lambda) = \sum_{k=1}^{Y_w} C(k, \lambda), \quad \lambda \in [-L, L] \quad (4.5)$$

Αν το καθολικό μέγιστο της συνάρτησης $P(\lambda)$ είναι στο σημείο $\lambda = \lambda_{\max}$, τότε η στροφή του κειμένου δίδεται από τον τύπο:

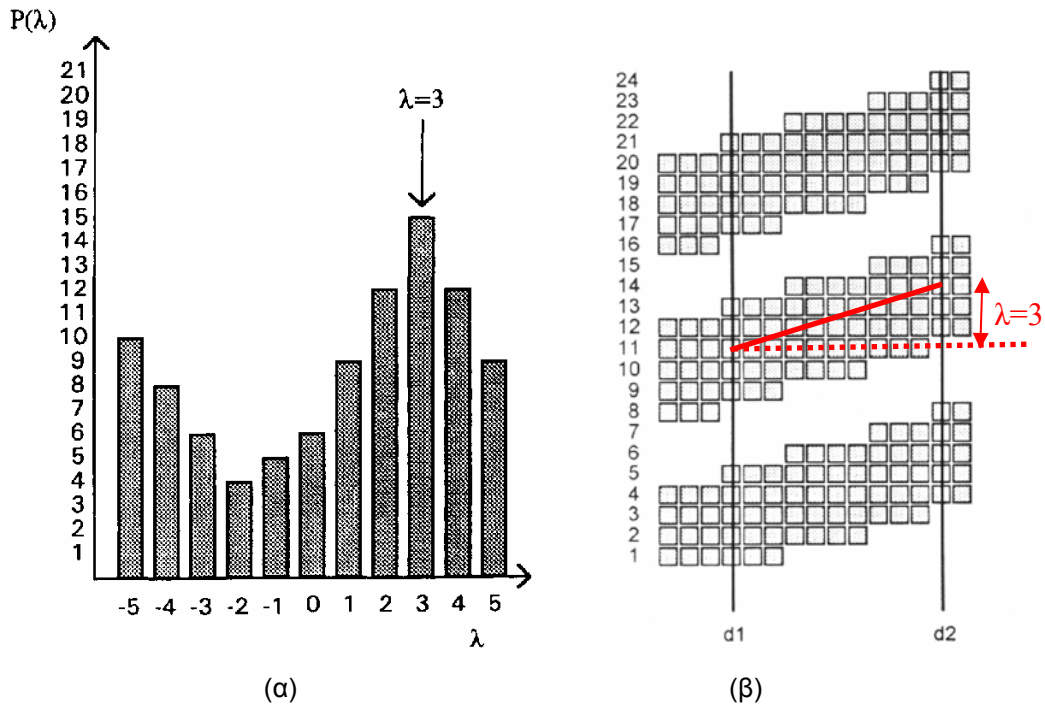
$$\theta = \tan^{-1}\left(\frac{\lambda_{\max}}{D_2 - D_1}\right) \quad (4.6)$$

Στο σχήμα 4.17 δίδεται η προβολή του πίνακα διασυσχέτισης του σχήματος 4.16β. Στο παράδειγμα αυτό το καθολικό μέγιστο της προβολής του πίνακα διασυσχέτισης δίδεται για $\lambda = 3$, το οποίο αντιστοιχεί στην γωνία στροφής του σχήματος 4.17β.

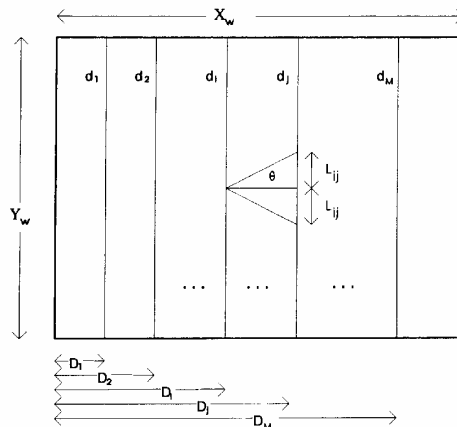
Για να αυξήσουμε την ακρίβεια της μεθόδου, μπορούμε να χρησιμοποιήσουμε τις πληροφορίες των pixels κειμένου που βρίσκονται σε περισσότερες από δύο κάθετες ευθείες. Αν χρησιμοποιήσουμε M κάθετες ευθείες, τότε αυτές κατανέμονται ομοίμορφα στο παράθυρο της εικόνας με αποστάσεις από το αριστερό όριο $D_i = (i X_w)/(M+1)$, $i = 1 \dots M$ (σχήμα 4.18). Για κάθε ζευγάρι ευθειών d_i και d_j ανιχνεύουμε όλα τα pixels με κατακόρυφη απόσταση λ τέτοια ώστε:

$$-L_{ij} < \lambda < L_{ij}, \quad \text{όπου } L_{ij} = (D_j - D_i)\tan(2\pi \theta_{\max}/360) \quad (4.7)$$

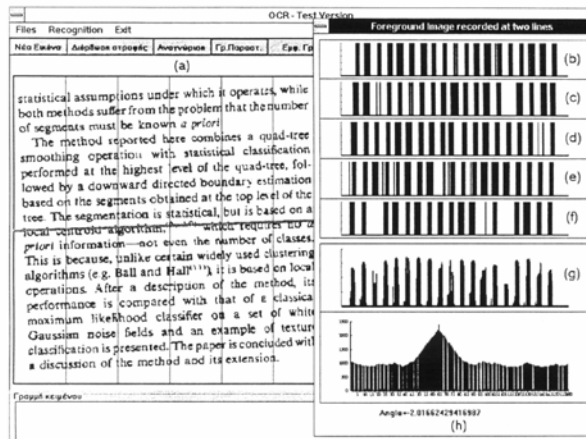
Στη συνέχεια υπολογίζουμε όλους τους πίνακες διασυσχέτισης C_{ij} , οι οποίοι αφορούν όλα τα ζευγάρια ευθειών d_i και d_j , και αφού τους κανονικοποιήσουμε ώστε η μεταξύ τους απόσταση να θεωρείται πάντα ίδια, δημιουργούμε έναν συνολικό πίνακα διασυσχέτισης CG ο οποίος μας δίνει την στροφή του κειμένου ακολουθώντας την μεθοδολογία της προσέγγισης με τις δύο κάθετες ευθείες (σχήμα 4.19).



Σχήμα 4.17. Η προβολή $P(\lambda)$ του πίνακα διασυσχέτισης του σχήματος 4.13β (α) και η αντίστοιχη γωνία κλίσης που αντιστοιχεί στην μέγιστη τιμή του $P(\lambda)$ (β).



Σχήμα 4.18. Εικόνα $X_w \times Y_w$ με κάθετες γραμμές σε αποστάσεις D_1, D_2, \dots, D_m .



Σχήμα 4.19. Εύρεση στροφής με χρήση 5 κάθετων ευθειών.

4.6 Διόρθωση της στροφής

Αφού έχουμε εντοπίσει την γωνία στροφής του εγγράφου θ , θα πρέπει να στρέψουμε το έγγραφο κατά $-\theta$ μοίρες ώστε να επανέλθει στις 0° . Για τον σκοπό αυτό μπορούμε να χρησιμοποιήσουμε δύο μεθόδους, την άμεση μέθοδο και την έμμεση μέθοδο (Kwag 2002).

Σύμφωνα με την άμεση μέθοδο, κάθε μαύρο pixel p της αρχικής εικόνας μετασχηματίζεται στο σημείο p' πολλαπλασιάζοντας τις συντεταγμένες του pixel p σύμφωνα με τον παρακάτω πίνακα στροφής:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \quad (4.8)$$

όπου (x,y) οι συντεταγμένες του pixel p της αρχικής εικόνας και (x',y') οι συντεταγμένες του pixel στην τελική εικόνα. Ένα από τα βασικά προβλήματα της άμεσης μεθόδου στροφής της εικόνας είναι ότι λόγω της στρογγυλοποίησης που γίνεται για να παραχθεί η τελική ψηφιακή εικόνα, γειτονικά τμήματα εικόνας ενώνονται μεταξύ τους και ενιαία τμήματα διαχωρίζονται. Επιπλέον, στο τελικό αποτέλεσμα εμφανίζονται κενά ή μικρές τρύπες σε ενιαία τμήματα εικόνας τα οποία στράφηκαν.

Η έμμεση μέθοδος είναι η αντίθετη της άμεσης. Για κάθε σημείο (x',y') της τελικής εικόνας, αναζητούμε αν υπάρχει αντίστοιχο σημείο (x,y) στην αρχική εικόνα. Αν υπάρχει, το σημείο (x',y') της τελικής εικόνας παίρνει την τιμή του σημείου (x,y) της αρχικής. Η αντιστοιχία υπολογίζεται εφαρμόζοντας τον αντίστροφο πίνακα στροφής σε κάθε σημείο (x',y') της τελικής εικόνας:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (4.9)$$

Με την έμμεση μέθοδο λύνεται το πρόβλημα της στρογγυλοποίησης οπότε δεν έχουμε τα προβλήματα που εμφανίζονται με την άμεση μέθοδο, όμως έχουμε μεγάλο υπολογιστικό κόστος γιατί θα πρέπει να εφαρμόσουμε τον παραπάνω τύπο σε όλα τα σημεία της τελικής εικόνας. Για καλύτερα αποτελέσματα, στην αρχική εικόνα ελέγχουμε τον αριθμό των pixels σε μία γειτονία γύρω από κάθε σημείο (x,y) (Amin 2000).

4.7 Διόρθωση της κλίσης των γραμμών

Εκτός από την εύρεση της γωνίας στροφής του κειμένου, ιδιαίτερο ενδιαφέρον έχει ο εντοπισμός και η διόρθωση της κλίσης των γραμμών του κειμένου. Οι λέξεις που έχουν γράμματα χωρίς κλίση μπορούν πιο εύκολα να καταταμηθούν και να αναγνωριστούν.

Επειδή τα γράμματα μπορεί να έχουν διαφορετική κλίση σε κάθε λέξη, το στάδιο αυτό προϋποθέτει τον εντοπισμό των λέξεων του κειμένου. Στη συνέχεια, για κάθε εντοπισμένη λέξη βρίσκεται και διορθώνεται η κλίση των γραμμών της. Μια ενδιαφέρουσα και αποτελεσματική μέθοδος περιγράφεται στο (Bozinovic 1989). Τα διάφορα βήματα της μεθόδου έχουν ως εξής:

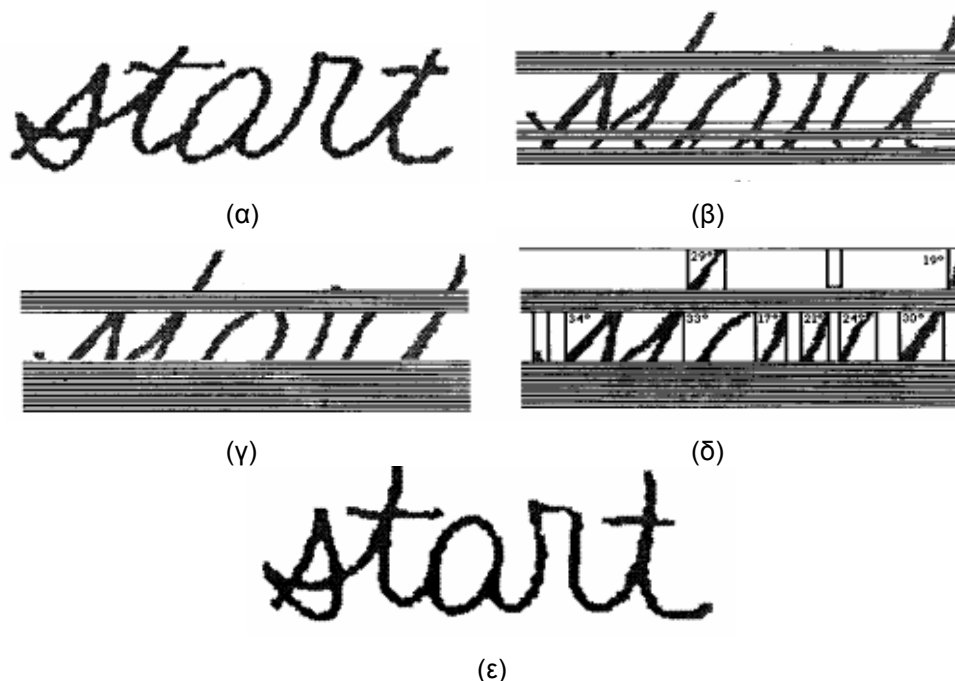
Βήμα 1: Απομάκρυνση όλων των οριζόντιων γραμμών που περιέχουν ένα τουλάχιστον τμήμα με διαδοχικά 1 μήκους $> th$ (σχήμα 4.20β).

Βήμα 2: Αφαίρεση και όλων των οριζόντιων τμημάτων με ύψος $< h_{th}$ (σχήμα 4.20γ).

Βήμα 3: Τα κομμάτια κάθε τμήματος που διαχωρίζονται με κάθετες γραμμές απομονώνονται σε παράθυρα (σχήμα 4.20δ). Για κάθε παράθυρο υπολογίζουμε τα κέντρα βάρους των πάνω και κάτω τμημάτων και τα συνδέουμε. Η κλίση των γραμμών της λέξης θ είναι η μέση τιμή των κλίσεων των συνδεδεμένων γραμμών.

Βήμα 4: Η διόρθωση της κλίσης (σχήμα 4.20ε) επιτυγχάνεται εφαρμόζοντας για κάθε pixel (x,y) της αρχικής εικόνας τον παρακάτω μετασχηματισμό:

$$x' = x - y \tan(\theta), \quad y' = y \quad (4.9)$$



Σχήμα 4.20. Εύρεση και διόρθωση της κλίσης των γραμμάτων της λέξης. (α) Αρχική λέξη. (β) Απομάκρυνση οριζόντιων γραμμών. (γ) Αφαίρεση οριζόντιων τμημάτων μικρού ύψους. (δ) Απομόνωση τμημάτων σε παράθυρα. (ε) Τελική εικόνα με διορθωμένη την κλίση των γραμμάτων.

Βιβλιογραφία

- (Amin 2000)** Amin, A., Fischer S.: A Document Skew Detection Method Using the Hough Transform. *Pattern Analysis & Applications* 3 (2000) 243-253
- (Baird 1987)** Baird, H.S.: The skew angle of printed documents. *Proc. SPSE 40th Conf. Symp. Hybrid Imaging Systems*, Rochester, N.Y. (1987) 21-24
- (Bozinovic 1989)** Bozinovic, R.M., Srihari, S.N.: Off-Line Cursive Script Word Recognition. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 11, No 1 (1989) 68-83
- (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 1997)** Gatos, B., Papamarkos, N., Chamzas, C.: Skew detection and text line position determination in digitized documents. *Pattern Recognition*, Vol. 30, No. 9 (1997) 1505-1519
- (Hinds 1990)** Hinds, S.C., Fisher, J. L., D'Amato, D.P.: A document skew detection method using run-length encoding and the Hough transform. *Proc. 10th Int. Conf. On Pattern Recognition* (1990) 464-468
- (Kwag 2002)** Kwag, H.K., Kim, S.H., Jeong, S.H., Lee, G.S.: Efficient skew estimation and correction algorithm for document images. *Image and Vision Computing* 20 (2002) 25-35
- (Thoma 1994)** Thoma, G.R., Wechsler, H.: Automated page orientation and skew angle detection for binary document images. *Pattern Recognition*, vol. 27, No. 10 (1994) 1325-1344
- (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
- (Yin 2001)** Yin P. Y.: Skew detection and block classification of printed documents. *Image and Vision Computing* 19 (2001) 567-579