



ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ (Τ.Ε.Ι.) ΣΕΡΡΩΝ

Τμήμα ΠΛΗΡΟΦΟΡΙΚΗΣ & ΕΠΙΚΟΙΝΩΝΙΩΝ

ΕΡΕΥΝΗΤΙΚΑ ΠΡΟΓΡΑΜΜΑΤΑ ΑΡΧΙΜΗΔΗΣ

ΕΝΙΣΧΥΣΗ ΕΡΕΥΝΗΤΙΚΩΝ ΟΜΑΔΩΝ ΣΤΟ ΤΕΙ ΣΕΡΡΩΝ

Ενέργεια. 2.2.3.στ

ΘΕΜΑ ΕΡΕΥΝΑΣ: ΔΙΑΡΘΡΩΣΗ ΠΕΡΙΕΧΟΜΕΝΟΥ ΕΧΡΩΜΩΝ ΕΓΓΡΑΦΩΝ

ΥΠΟΕΡΓΟ 3

ΕΠΙΣΤΗΜΟΝΙΚΟΣ ΥΠΕΥΘΥΝΟΣ

ΧΑΡΑΛΑΜΠΟΣ ΣΤΡΟΥΘΟΠΟΥΛΟΣ - ΚΑΘΗΓΗΤΗΣ

ΑΝΑΠΤΥΞΗ ΤΕΧΝΙΚΗΣ ΜΕΙΩΣΗΣ ΤΩΝ ΧΡΩΜΑΤΩΝ ΕΙΚΟΝΑΣ

ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΤΗΣ ΔΟΜΗΣ ΣΕΛΙΔΑΣ

ΠΑΚΕΤΟ ΕΡΓΑΣΙΑΣ 2

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	3
2. Εφαρμογές ΑΔΣ.....	5
3. Κατηγορίες μεθόδων ΑΔΣ.....	6
4. Εξαγωγή κειμένου από εικόνες.....	7
4.1. Διαδικασία εξαγωγής κειμένου.....	8
4.2. Εξομάλυνση Κατά Μήκος μίας Διαδρομής.....	9
4.3. Προβολές πλαγίων όψεων	9
4.4. Γενική μεθοδολογία για εξαγωγή κειμένου	11
4.5. Σύντομη περιγραφή εργασιών με θέμα την εξαγωγή κειμένου	12
4.5.1. Locating text in complex color images.....	12
4.5.2. Adaptive page segmentation for color technical journals' cover images ..	12
4.5.3. Text extraction from colored book and journal covers.....	13
4.5.4. A word extraction al Pattern Recognition algorithm for machine-printed documents using a 3D neighborhood graph model.....	14
4.5.5. Hybrid approach to efficient text extraction in complex color images.....	14
5. Γενικές πληροφορίες.....	15
5.1. Διεθνή περιοδικά συσχετιζόμενα με την ΑΔΣ.....	15
5.2. Σύνδεσμοι διαδικτύου	15
6. Βιβλιογραφία	15

ΑΝΑΛΥΣΗ ΔΟΜΗΣ ΣΕΛΙΔΑΣ (PAGE LAYOUT ANALYSIS)

1.Εισαγωγή

Η γρήγορη ανάπτυξη των ψηφιακών συστημάτων και της τεχνολογίας υπολογιστών έχει δημιουργήσει ένα μεγάλο φάσμα εφαρμογών που πραγματεύονται την προσομοίωση ανθρώπινων λειτουργιών. Η σχετικά νέα αυτή επιστήμη καλείται «Τεχνητή Νοημοσύνη» (Artificial Intelligence) και αποτελεί ένα από τα πιο ενδιαφέροντα και ενεργά ερευνητικά πεδία στην επιστημονική κοινότητα.

Μία κατηγορία εφαρμογών που ανήκει στο ευρύ πεδίο της τεχνητής νοημοσύνης, είναι η «Όραση Μηχανής» (Computer Vision). Ως απώτερο σκοπό έχει την αντικατάσταση της ανθρώπινης όρασης από μηχανές. Αναλογιζόμενοι τη σημερινή κατάσταση της τεχνολογίας σε σύγκριση με τις ικανότητες της ανθρώπινης όρασης εύκολα οδηγούμαστε στο συμπέρασμα ότι θα χρειαστεί πολύς χρόνος για φτάσουμε στο επιθυμητό αποτέλεσμα, με ισχυρή αμφιβολία αν είναι καν εφικτό.

Ένα παράδειγμα εφαρμογής της όρασης μηχανής αποτελεί η προσπάθεια να υλοποιηθούν τεχνικές οι οποίες να είναι σε θέση να εξάγουν από ένα έγγραφο το κείμενο, τις εικόνες, τα σχήματα κ.λ.π. Αυτή η προσπάθεια σχετίζεται απόλυτα με το θέμα που πραγματεύεται η Ανάλυση Δομής Σελίδας.

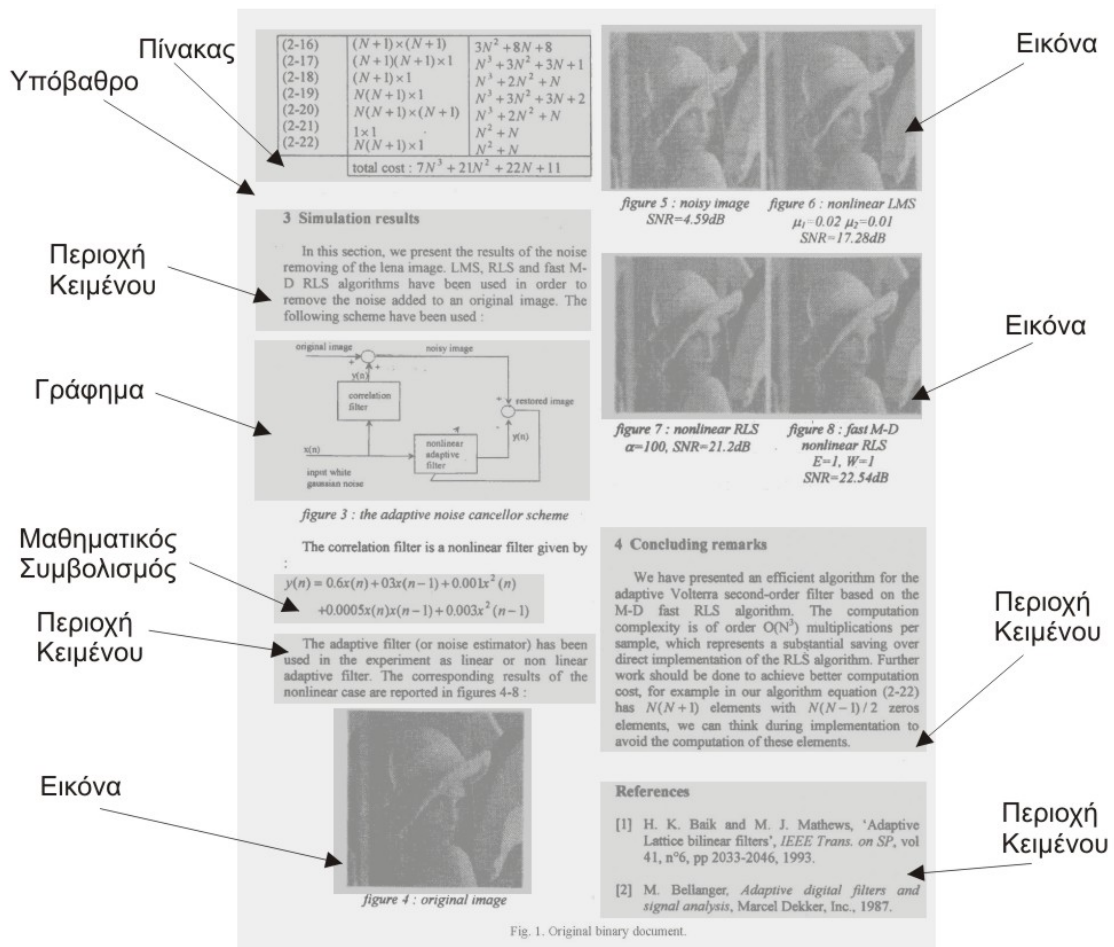
Ως **Ανάλυση Δομής Σελίδας – ΑΔΣ** (Page Layout Analysis - PLA) θα μπορούσε να οριστεί η ανίχνευση (detection), ο εντοπισμός (localization), η εξαγωγή και η περιγραφή των δομικών στοιχείων (Structure Elements) ενός εγγράφου σε μορφή ψηφιακής εικόνας (Bitmap). Δομικά στοιχεία μέσα σε ένα έγγραφο μπορούν σε γενικές γραμμές να ονομαστούν οι παρακάτω οντότητες.

- Κείμενο
- Εικόνες
- Γραφήματα
- Πίνακες
- Μαθηματικοί συμβολισμοί
- Υπόβαθρο (background).

Η εξαγωγή του κειμένου μπορεί να γίνει βρίσκοντας μεμονωμένους χαρακτήρες, λέξεις, γραμμές κειμένου, ολόκληρες παραγράφους ή απλά περιοχές κειμένου.

Το κείμενο αποτελεί πολύ σημαντική πληροφορία τόσο στα έγγραφα όσο και σε άλλου τύπου εικόνας και εικονοσειράς (video). Σε αυτό έχει επικεντρωθεί άλλωστε και το μεγαλύτερο μέρος της ερευνητικής δραστηριότητας. Αυτό οφείλεται κυρίως στη σχετικά μεγαλύτερη διακριτικότητα που διαθέτει έναντι των υπολοίπων δομικών στοιχείων, στον συγκριτικά ευκολότερο εντοπισμό του και φυσικά στο γεγονός ότι βάσει αυτού είναι δυνατή η περιγραφή του περιεχομένου της εικόνας. Στην ενότητα 4

αναλύεται λεπτομερώς η διαδικασία της εξαγωγής κειμένου και οι ιδιαιτερότητες που παρουσιάζει γενικότερα το κείμενο κατά τη διαδικασία της ΑΔΣ.



Σχήμα 1. Δυαδικό έγγραφο και η επίδειξη των δομικών στοιχείων του.

Σε σχέση με την τελευταία κατηγορία (υπόβαθρο), σημειώνεται ότι σε ένα έγγραφο είναι δυνατή η συνύπαρξη πολλών τέτοιων περιοχών τα οποία χαρακτηρίζονται ως τοπικά υπόβαθρα (local backgrounds). Αν το υπόβαθρο του εγγράφου καλύπτει όλη την επιφάνεια ή μεγάλο μέρος του μπορεί να χαρακτηριστεί ως καθολικό υπόβαθρο (global background). Όλα τα υπόλοιπα δομικά στοιχεία ανήκουν χωρικά σε ένα ή περισσότερα υπόβαθρα.

Το Σχήμα 1 δείχνει μία περίπτωση δυαδικού εγγράφου όπου περιέχονται όλες οι κατηγορίες δομικών στοιχείων που προαναφέρθηκαν.

Οι δυσκολίες που έχουν να αντιμετωπίσουν οι ερευνητές σε σχέση με το παρόν θέμα είναι πολλές και αυτό οφείλεται κυρίως στην εξαιρετικά πολύπλοκη σύνθεση των εγγράφων καθώς και την ιδιομορφία των δομικών τους στοιχείων. Μερικά παραδείγματα τέτοιων δυσκολιών είναι:

- Οι ομοιότητες που παρουσιάζουν μεταξύ τους τα δομικά στοιχεία με αποτέλεσμα τη δυσκολία διαχωρισμού τους.
- Η χωρική επικάλυψη των δομικών στοιχείων. Ένα παράδειγμα είναι η ύπαρξη κειμένου μέσα σε εικόνες.
- Η γεωμετρική ανομοιογένεια ίδιου τύπου δομικών στοιχείων. Παράδειγμα αποτελεί η πληθώρα γραμματοσειρών του κειμένου.
- Η χαμηλή αντίθεση της εικόνας και σύνθετα backgrounds.
- Η πολυπλοκότητα των αλγορίθμων που οδηγεί σε ανάγκη μεγάλης επεξεργαστικής ισχύος λόγω της ίδιας της σύνθετης δομής του προβλήματος.

Μερικές από τις επιθυμητές ιδιότητες των τεχνικών ΑΔΣ είναι:

1. Να παρουσιάζει σχετική ανεξαρτησία ως προς τα γεωμετρικά χαρακτηριστικά των χαρακτήρων (π.χ μέγεθος, στυλ, γραμματοσειρά κ.τ.λ.).
2. Να είναι σε θέση να εντοπίζει τα δομικά στοιχεία του εγγράφου ανεξαρτήτως της κλίσης που ορίζουν οι είτε τα ίδια τα στοιχεία είτε το ίδιο το έγγραφο.
3. Να μην έχει μεγάλες απαιτήσεις τόσο σε επεξεργαστική ισχύ όσο και χρονικό διάστημα εκτέλεσης.

2.Εφαρμογές ΑΔΣ

Η ΑΔΣ αποτελεί μία σημαντική προεργασία για πολλές εφαρμογές. Η πιο συνήθης είναι η οπτική αναγνώριση χαρακτήρων (OCR – Optical Character Recognition). Τα τελευταία χρόνια υπάρχει μια συνεχώς αυξανόμενη ανάγκη για συστήματα ικανά να μετατρέπουν αυτόματα μεγάλο όγκο έντυπων εγγράφων σε κατάλληλη ηλεκτρονική μορφή, έτσι ώστε η πληροφορία να είναι άμεσα επεξεργάσιμη. Για να είναι εφικτή αυτή η μετατροπή θα πρέπει το σύστημα να μπορεί να διαχωρίζει τις περιοχές κειμένου από το έγγραφο και να δίδεται το αποτέλεσμα ως είσοδος σε μία μονάδα αυτόματης αναγνώρισης κειμένου. Πολλά εμπορικά λογισμικά πακέτα έχουν αναπτυχθεί τα οποία υλοποιούν την τεχνική του OCR. Μερικά από τα πιο γνωστά είναι: ABBYY FineReader, ReadIris, OmniPage.

Συμπίεση και μετάδοση. Ο διαχωρισμός της εικόνας σε επιμέρους επίπεδα βελτιώνει σημαντικά το λόγο συμπίεσης. Ως παράδειγμα μπορεί να αναφερθεί η απομόνωση του κειμένου και η ξεχωριστή αποθήκευσή του.

Δημιουργία ψηφιακών αρχείων και ανάκτηση εικόνων εγγράφων. Η κατηγορία αυτών των τεχνικών ανήκει στο γενικότερο πεδίο της ανάκτησης εικόνων με βάση το περιεχόμενό τους (Content Based Image Retrieval). Η περιγραφή των εγγράφων γίνεται από χαρακτηριστικά που έχουν εξαχθεί από αυτό και έχουν αποθηκευτεί σε μία βάση δεδομένων. Η αναζήτηση γίνεται συγκρίνοντας τα χαρακτηριστικά της υπό εξέταση εικόνας (query image) με αυτά που είναι αποθηκευμένα στη βάση.

Υπάρχει επίσης μια πληθώρα από εξειδικευμένες εφαρμογές που περιλαμβάνουν ανίχνευση αριθμών και χαρακτήρων σε χάρτες, μηχανικά σχέδια, φακέλους αλληλογραφίας (address blocks on letters), εξώφυλλα CD κ.λ.π.

Σημειώνεται ότι, καμία από τις μέχρι τώρα τεχνικές ΑΔΣ δεν έχει τη δυνατότητα επιτυχίας σε όλων των ειδών τις εφαρμογές. Η κάθε τεχνική επικεντρώνεται σε μία εξειδικευμένη κατηγορία εφαρμογών (ad hoc applications ή application oriented techniques).

3.Κατηγορίες μεθόδων ΑΔΣ

Γενικά, το περιεχόμενο μίας εικόνας μπορεί να διαχωριστεί σε δύο κύριες κατηγορίες [1]

1. Δομικό περιεχόμενο (Perceptual/Structural content). Ονομάζεται επίσης και φυσικό ή γεωμετρικό περιεχόμενο.
2. Λειτουργικό περιεχόμενο (Semantic/Functional content). Ονομάζεται επίσης και συντακτικό ή λογικό περιεχόμενο.

Η πρώτη κατηγορία αφορά τα «χαμηλού επιπέδου» χαρακτηριστικά (Low level features). Το χρώμα, η φωτεινότητα, η γεωμετρία, η υφή κ.τ.λ. ανήκουν στην ομάδα αυτών των χαρακτηριστικών.

Το λειτουργικό περιεχόμενο μιας εικόνας σχετίζεται με τα «υψηλού επιπέδου» χαρακτηριστικά (High level features). Αυτή η κατηγορία περιλαμβάνει τα αντικείμενα της εικόνας (π.χ κείμενο, πρόσωπα, πινακίδες αυτοκινήτων κ.τ.λ.) και τη σχέση που αυτά έχουν μεταξύ τους. Το λειτουργικό περιεχόμενο ταυτίζεται με τον τρόπο της ανθρώπινης αντίληψης και κατανόησης (human understanding).

Με βάση τα παραπάνω, και οι τεχνικές ΑΔΣ μπορούν να ταξινομηθούν σε δύο κατηγορίες ανάλογα με τον τρόπο που επεξεργάζονται το περιεχόμενο του εγγράφου:

1. Ανάλυση Δομικής Διάταξης – ΑΔΔ. Το αποτέλεσμα των τεχνικών αυτών είναι ο χωρικός διαχωρισμός των αντικειμένων βασισμένες στα χαρακτηριστικά χαμηλού επιπέδου.
2. Ανάλυση Λειτουργικής Διάταξης - ΑΛΔ. Το αποτέλεσμα των τεχνικών ΑΛΔ είναι ο χαρακτηρισμός των αντικειμένων με βάση το διαχωρισμό που προέκυψε από μία διαδικασία ΑΔΔ. Έτσι, σε ένα έγγραφο μπορούν να χαρακτηριστούν οι περιοχές κειμένου, εικόνων, μαθηματικών συμβολισμών κ.τ.λ. Επιπλέον, το κείμενο μπορεί να ταξινομηθεί εκ νέου σε τίτλους, επικεφαλίδες, παραγράφους κ.τ.λ.

Ένας άλλος διαχωρισμός που μπορεί να γίνει στις τεχνικές ΑΔΣ είναι το αν αυτές εφαρμόζονται σε δυαδικά έγγραφα (binary documents) [6,7,21,51,43,45] ή σε έγχρωμα έγγραφα [5,13,23,25,26,35,38], που είναι σαφώς ένα πρόβλημα με μεγαλύτερη πολυπλοκότητα. Στα δυαδικά έγγραφα τα επίπεδα φωτεινότητας είναι δύο, το άσπρο (255) και το μαύρο (0). Οι τεχνικές ΑΔΣ υποθέτουν ότι τα αντικείμενα (printed regions) ανήκουν στην κλάση των μαύρων pixel ενώ το background στην

κλάση των άσπρων pixel. Στην περίπτωση των έγχρωμων εγγράφων συνήθως εφαρμόζεται μία τεχνική μείωσης χρωμάτων [14] πριν από την κύρια επεξεργασία ή απλά μετατρέπεται σε gray scale [18,35]. Έτσι μειώνεται σημαντικά η πολυπλοκότητα της μιας και οι πραγματικού χρώματος εικόνες μπορούν να περιέχουν πολλές χιλιάδες χρώματα.

Μία επιπλέον κατηγοριοποίηση σχετίζεται με τη σειρά ομαδοποίησης της πληροφορίας του κειμένου (textual objects information grouping). Έτσι, έχουμε

1. Τις τεχνικές ανάλυσης «από κάτω προς τα πάνω» (bottom-up ή data driven). Διαδοχικά ενώνονται συστατικά μικρού μεγέθους και προοδευτικά σχηματίζονται δομές μεγαλύτερης κλίμακας, π.χ ομαδοποίηση γραμμών για να τη δημιουργία λέξεων, λέξεις για τη δημιουργία γραμμών κειμένου κ.τ.λ. Είναι συνήθως πιο ευέλικτες από τις τεχνικές της επόμενης κατηγορίας, προκαλούνται όμως συσσωρευμένα σφάλματα κατά τη μετάβαση από τις μικρές κλίμακες στις μεγαλύτερες.
2. Τις τεχνικές ανάλυσης «από πάνω προς τα κάτω» (top-down ή pattern driven). Εκκινούν την ανάλυση ανιχνεύοντας μεγάλης κλίμακας αντικείμενα (π.χ στήλες κειμένου) και με διαδοχικές διαιρέσεις καταλήγουν σε μικρής κλίμακας αντικείμενα (π.χ λέξεις ή μεμονωμένοι χαρακτήρες). Για να έχουν επιτυχία οι συγκεκριμένες τεχνικές, είναι αναγκαία η εκ των πρότερων γνώση κάποιων δεδομένων για τη δομή της σελίδας. Είναι λοιπόν αναγκαίο να γίνουν παραδοχές όσον αφορά τη χωρική διάταξη των δομικών στοιχείων του εγγράφου.
3. Τις υβριδικές τεχνικές όπου η φιλοσοφία των δύο πρώτων τεχνικών συνδυάζεται για να βελτιωθεί το αποτέλεσμα.

Τέλος, ανάλογα με τις παραδοχές που γίνονται όσον αφορά τον τύπο του εγγράφου, υπάρχουν δύο ειδών τεχνικές όπου στην πρώτη ανιχνεύεται μόνο κάθετα ή οριζόντια τοποθετημένο κείμενο ενώ στη δεύτερη είναι δυνατή η ανίχνευση του κειμένου ανεξάρτητα από την κλίση ή την τοποθέτησή του. Στην πρώτη περίπτωση, το έγγραφο θα πρέπει να μην έχει κλίση ή αυτή να έχει διορθωθεί πριν από την ανάλυσή του (skew correction).

4.Εξαγωγή κειμένου από εικόνες

Όπως έχει αναφερθεί στην εισαγωγή, το κείμενο σε οποιονδήποτε τύπο εικόνας ή εικονοσειράς, αποτελεί ισχυρή μορφή πληροφορίας. Μπορεί να δώσει άμεσα στο θεατή το θέμα με το οποίο σχετίζεται το έγγραφο που διαβάσει. Είναι επίσης πολύ σημαντικό γιατί μπορεί να εξαχθεί σχετικά εύκολα συγκριτικά με άλλα αντικείμενα και καθιστά εφικτή τη δημιουργία ενός ευρύτατου φάσματος εφαρμογών.

Τονίζεται επίσης ότι το κείμενο ανήκει στην κατηγορία του λειτουργικού περιεχομένου (προηγούμενη ενότητα) καθώς με αυτό είναι δυνατή η περιγραφή της λειτουργικής διάταξης της εικόνας μέσα στην οποία περιέχεται.

4.1. Διαδικασία εξαγωγής κειμένου

Σε γενικές γραμμές, μία διαδικασία εξαγωγής κειμένου περιλαμβάνει τα εξής βήματα:

1. Ανίχνευση κειμένου (text detection), επιβεβαίωση δηλαδή για το αν όντως υπάρχει κείμενο μέσα στην εικόνα. Οι περισσότερες εφαρμογές υποθέτουν ότι η ύπαρξη κειμένου είναι δεδομένη, κυρίως σε εφαρμογές στατικών εικόνων εγγράφων. Αυτό το βήμα αφορά εφαρμογές επεξεργασίας video όπου ο αριθμός των frames με κείμενο είναι συνήθως πολύ μικρός σε σχέση με τον αριθμό των συνολικών frames.
2. Εντοπισμός κειμένου (text localization). Ανιχνεύονται περιοχές οι οποίες βάσει κάποιου κριτηρίου είναι πολύ πιθανόν να περιέχουν κείμενο. Το αποτέλεσμα είναι κάποια ορθογώνια παραλληλόγραμμα (bounding boxes) που μέσα περιέχουν τις εν λόγω περιοχές. Ο τρόπος με τον οποίο εκτελείται το παρόν βήμα καθορίζει και τον χαρακτηρισμό της τεχνικής σε *από κάτω προς τα πάνω* ή *από πάνω προς τα κάτω*.
3. Επιβεβαίωση - επικύρωση περιοχών κειμένου (text block verification). Απομάκρυνση των εσφαλμένα εντοπισμένων περιοχών κειμένου του προηγούμενου βήματος. Το βήμα αυτό περιλαμβάνει την εξαγωγή χαρακτηριστικών και τη χρήση τους για την ταξινόμηση των περιοχών.
4. Απομόνωση κειμένου (text data layer isolation). Δυαδική απεικόνιση του αποτελέσματος με δύο κλάσεις, τα pixel κειμένου και τα υπόλοιπα pixel της εικόνας. Για τις περισσότερες τεχνικές ΑΔΣ, αυτό είναι συνήθως και το τελικό στάδιο.
5. Βελτίωση αποτελέσματος (Enhancement, post-processing). Τελικό φιλτράρισμα της εικόνας για καλύτερη ποιότητα των χαρακτήρων του κειμένου με σκοπό την αύξηση του ποσοστού επιτυχίας την μονάδας αυτόματης αναγνώρισης κειμένου.
6. Οπτική αναγνώριση χαρακτήρων (OCR). Δημιουργία ASCII χαρακτήρων για άμεση ηλεκτρονική επεξεργασία.

Δύο πολύ δημοφιλείς αλγόριθμοι πάνω στους οποίους βασίζονται πολλές από τις «από πάνω προς τα κάτω» τεχνικές είναι

1. ο αλγόριθμος *Εξομάλυνσης Κατά Μήκος μίας Διαδρομής* - ΕΚΜΔ (Run Length Smoothing/Smearing Algorithm RLSA) [47] και
2. ο αλγόριθμος *προβολών πλαγίων όψεων κατά τομές* (projection profile cut ή recursive X-Y cuts) [46].

Και οι δύο αυτοί αλγόριθμοι εφαρμόζονται μόνο σε δυαδικές εικόνες και αφορούν το βήμα 2 όπου γίνεται το αρχικό ξεκαθάρισμα για τις πιθανές περιοχές κειμένου.

4.2.Εξομάλυνση Κατά Μήκος μίας Διαδρομής

Η μέθοδος ΕΚΜΔ χρησιμοποιείται για την εξομάλυνση της δυαδικής μορφής του εγγράφου με σκοπό τη συνένωση τμημάτων της εικόνας που ανήκουν στην ίδια κατηγορία.

Έστω ότι τα μαύρα pixel δηλώνονται με τον αριθμό 1 και τα άσπρα pixel (το υπόβαθρο) δηλώνονται με τον αριθμό 0. Αν έχουμε μια τυχαία ακολουθία αποτελούμενη από τους δύο αυτούς αριθμούς και ένα συντελεστή C, τα άσπρα pixel μεταξύ δύο μαύρων των οποίων η απόσταση είναι μικρότερη ή ίση του C αντικαθίστανται με την τιμή 1. Για παράδειγμα, θεωρούμε την παρακάτω δυαδική ακολουθία:

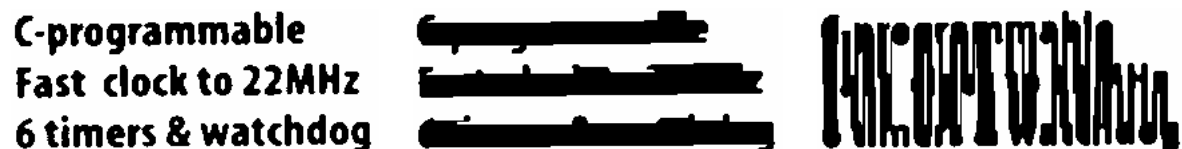
1 1 0 0 1 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 1 1

και υποθέτουμε ότι η τιμή του C είναι 3. Η σειρά των χαρακτήρων μετατρέπεται ως εξής

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

Η τιμή του C καθορίζει το βαθμό εξομάλυνσης της εικόνας και μπορεί να ληφθεί πειραματικά ή από στατιστική εξέταση της εικόνας για τον υπολογισμό των αποστάσεων μεταξύ των χαρακτήρων.

Ο αλγόριθμος ΕΚΜΔ μπορεί εφαρμοστεί τόσο κατά την οριζόντια όσο και κατά την κάθετη διεύθυνση με αποτέλεσμα δύο νέες εικόνες. Συνδυαζόμενες με κάποιο τελεστή (συνήθως με το λογικό AND) δίνουν το τελικό αποτέλεσμα της τεχνικής.



(α) Αρχική εικόνα

(α) Οριζόντια εξομάλυνση (C=20)

(α) Κάθετη εξομάλυνση (C=30)

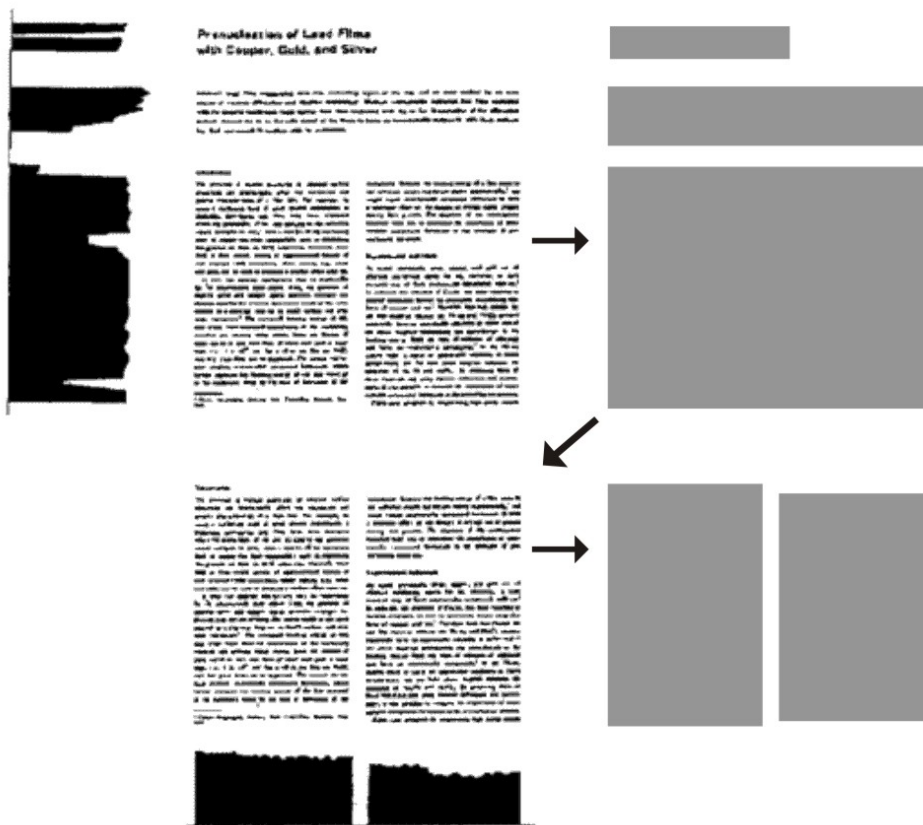
Σχήμα 2. Η μέθοδος ΕΚΜΔ.

4.3.Προβολές πλαγίων όψεων

Η μέθοδος στηρίζεται στον επαναληπτικό υπολογισμό των οριζόντιων και κάθετων προβολών της εικόνας για τη διαδοχική τμηματοποίησή της. Η προβολή υπολογίζεται αθροίζοντας τον αριθμό των pixel της ίδιας κλάσης (συνήθως τα μαύρα) σε καθορισμένη διεύθυνση και τη δημιουργία μίας συνάρτησης μίας μεταβλητής.

Σε μια τυπική εφαρμογή της τεχνικής, τα βήματα που ακολουθούνται είναι

1. Εύρεση προβολής στην οριζόντια διεύθυνση.

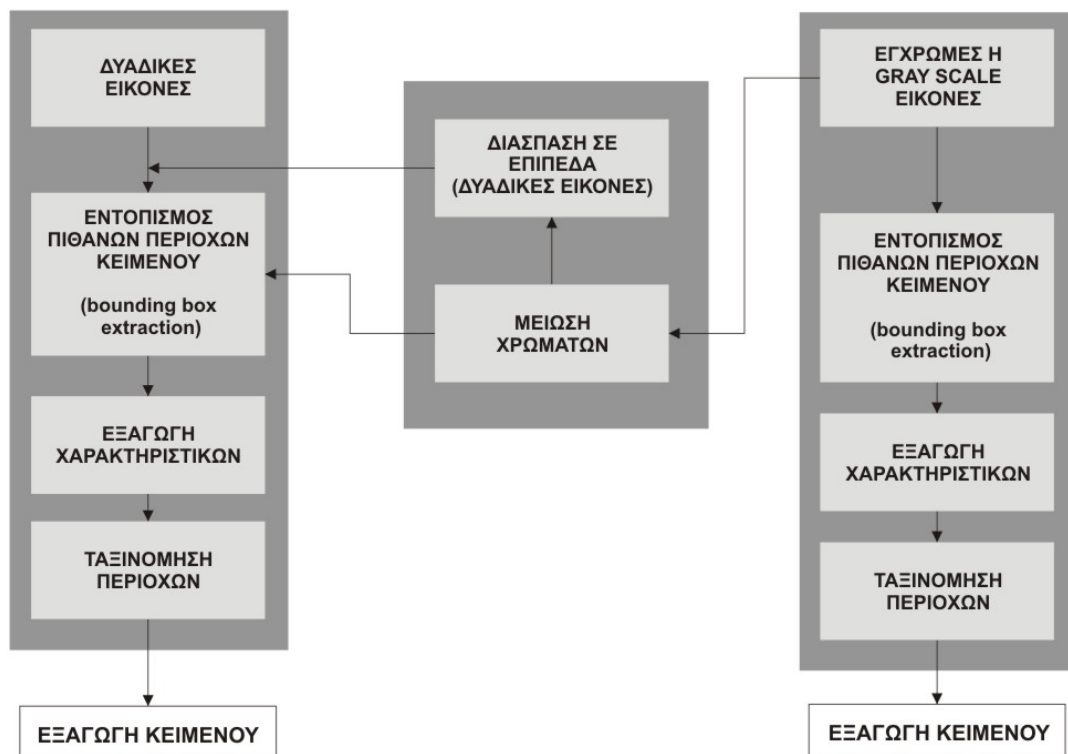


Σχήμα 3. Παράδειγμα εφαρμογής της μεθόδου recursive X-Y cuts στην οριζόντια και κάθετη διεύθυνση.

2. Εντοπισμός των διαδοχικών μηδενικών της οριζόντιας προβολής. Αν ο αριθμός τους είναι πάνω από κάποιο όριο γίνεται διαχωρισμός της εικόνας σε επιμέρους τμήματα.
3. Εύρεση κάθετης προβολής για κάθε επιμέρους τμήμα που προέκυψε από το βήμα 2. Εντοπισμός των διαδοχικών μηδενικών της κάθετης προβολής και εκ νέου διαχωρισμός με το ίδιο κριτήριο.
4. Επανάληψη των 1-3 μέχρι τη μη περαιτέρω δυνατότητα διαχωρισμού ή μέχρι ο αριθμός των περιοχών να ξεπεράσει κάποιο μέγιστο όριο.

Η μεθοδολογία αυτή σε πολλές περιπτώσεις μπορεί να εφαρμοστεί παραλλαγμένη ως προς την προβολή που υπολογίζεται αρχικά και ως προς το κριτήριο διαχωρισμού της εικόνας. Για παράδειγμα, σε περιπτώσεις ύπαρξης θορύβου, αντί του εντοπισμού διαδοχικών μηδενικών θα ήταν δυνατό να ανιχνεύονται σημεία της προβολής που παρουσιάζεται μικρός αριθμός pixels. Στο Σχήμα 3 φαίνεται ένα παράδειγμα εφαρμογής της μεθόδου.

Σημειώνεται επίσης ότι ο συγκεκριμένος διαδοχικός διαχωρισμός της εικόνας μπορεί να αναπαρασταθεί με τη μορφή δυαδικών δέντρων (X-Y tree) και η πληροφορία να χρησιμοποιηθεί για την περιγραφή της δομής του εγγράφου. Μια τέτοια προσέγγιση είναι πολύ χρήσιμη στην ανάκτηση εγγράφων από βάσεις δεδομένων.



Σχήμα 4. Διάγραμμα γενικής μεθοδολογίας εξαγωγής κειμένου από έγγραφα.

4.4.Γενική μεθοδολογία για εξαγωγή κειμένου

Στο Σχήμα 4 παρουσιάζονται με γενικό τρόπο τα βασικά βήματα που ακολουθούνται για την εξαγωγή κειμένου, είτε πρόκειται για δυαδικές είτε έγχρωμες/gray scale εικόνες. Ο σκοπός αυτής της παρουσίασης είναι να δείξει τις διαφορετικές προσεγγίσεις της διαδικασίας που περιγράφει η ενότητα 4.1, ιδιαίτερα για το πώς ανιχνεύονται οι πιθανές περιοχές κειμένου όταν έχουμε έγχρωμα ή gray scale έγγραφα (για δυαδικά έγγραφα έχουν αναφερθεί δύο ενδεικτικές τεχνικές, ενότητες 4.2-4.3).

Οι περισσότερες από τις τεχνικές που έχουν υλοποιηθεί μέχρι σήμερα ακολουθούν τρεις κυρίως κατευθύνσεις.

1. Απ' ευθείας εντοπισμός των πιθανών περιοχών κειμένου, εξαγωγή χαρακτηριστικών από αυτές, ταξινόμηση και εξαγωγή τελικού κειμένου. Χαρακτηριστικά υφής (texture based features), ακμών (edge based features) και τεχνικές βασισμένες σε συνδεδεμένα στοιχεία (Connected Components based methods) είναι τα κύρια μέσα για την επίτευξη της ανίχνευσης κειμένου.
2. Εφαρμογή της μεθόδου πάνω σε εικόνα που έχει υποστεί μείωση χρωμάτων. Μια τέτοια περίπτωση αποτελεί η εργασία [25].

3. Μείωση χρωμάτων, δημιουργία μίας δυαδικής εικόνας για κάθε επίπεδο χρωματικής κλάσης και εύρεση κειμένου σε κάθε επίπεδο ξεχωριστά. Σε αυτή τη περίπτωση μπορεί να υπάρξει και ένα τελικό στάδιο συνδυασμού των αποτελεσμάτων.

Για λεπτομερή ανασκόπηση της βιβλιογραφίας με θέμα την εξαγωγή κειμένου από εικόνες και video δείτε τις εργασίες [1,2,3,4].

4.5.Σύντομη περιγραφή εργασιών με θέμα την εξαγωγή κειμένου

4.5.1.Locating text in complex color images

Y. Zhong, K. Karu, A.K. Jain [38]

Οι Zhong κ.α. [38] πρότειναν ένα υβριδικό σύστημα εξαγωγής κειμένου από σύνθετες εικόνες με ειδικευση σε εξώφυλλα CDs. Η πρώτη μέθοδος διαιρεί την έγχρωμη εικόνα σε συνδεδεμένα στοιχεία με χρωματική ομοιομορφία. Με τη χρήση μιας ακολουθίας συλλογισμών κατατάσσει τα στοιχεία σε περιοχές κειμένου ή όχι. Η μέθοδος αυτή χρησιμοποιεί μια μέθοδο ομαλοποίησης ιστογράμματος για την διαδικασία της ταξινόμησης των χρωμάτων που μπορεί να παράγει από 5 έως 500 πρωτότυπα χρώματα. Η δεύτερη μέθοδος εξάγει τις περιοχές κειμένου αξιοποιώντας τις τοπικές μεταβολές της φωτεινότητας σε ασπρόμαυρης διαβάθμισης έγγραφα. Ο συνδυασμός των δύο παραπάνω μεθόδων συντελεί στην αποτελεσματικότερη ανίχνευση κειμένου. Η μέθοδος πάντως αποτυγχάνει σε περιπτώσεις που το κείμενο περιέχεται σε εικόνες.

4.5.2.Adaptive page segmentation for color technical journals' cover images

W.Y. Chen and S.Y. Chen [35]

Προτείνεται μία μέθοδος εξαγωγής κειμένου από έγχρωμα έγγραφα προερχόμενα κυρίως από εξώφυλλα τεχνικών περιοδικών (technical journal's cover images). Οι παραδοχές που γίνονται για την επιτυχία της μεθόδου είναι: η κλίση του εγγράφου να μην είναι πολύ μεγάλη, οι χαρακτήρες θεωρούνται ότι έχουν ομοιόμορφη χρωματική κατανομή και ότι το κείμενο είναι σε οριζόντια διάταξη.

Αρχικά η εικόνα υφίσταται μείωση χρωμάτων για να απλοποιηθεί η δομή της. Χρησιμοποιείται ένας αλγόριθμος μείωσης χρωμάτων βασισμένος στο μοντέλο YIQ και το αποτέλεσμα αυτής της φάσης είναι ένα έγγραφο που περιέχει 42 ή λιγότερα χρώματα.

Ακολουθώς, ανιχνεύονται οι ακμές με τη χρήση του τελεστή Sobel. Οι τιμές τους αντιστοιχούν στο άθροισμα των ακμών και των τριών επιπέδων RGB. Η εικόνα που προκύπτει μετατρέπεται σε δυαδική με βάση ένα κατώφλι για να απομονωθούν οι ισχυρές ακμές και στη συνέχεια εντοπίζονται τα κύρια τμήματα του εγγράφου (primary blocks extraction). Η διαδικασία αυτή βασίζεται στον αλγόριθμο Εξομάλυνσης Κατά Μήκος μίας Διαδρομής (RLSA – Run Length Smoothing

Algorithm) ο οποίος εφαρμόζεται πάνω στην εικόνα που προέκυψε μετά την δυαδικοποίησή της.

Το αν θα επεξεργαστεί περαιτέρω κάποιο από τα κύρια τμήματα που εξάχθηκαν, αποφασίζεται από την ανάλυση της πολυπλοκότητάς της. Αν προκύψει ότι θα πρέπει να γίνει επιπλέον ανάλυση τότε ξεχωριστά σε κάθε χρωματικό επίπεδο (color plane) εφαρμόζεται ο προηγούμενος αλγόριθμος για να εξαχθούν τα κύρια τμήματα.

Το τελικό στάδιο σχετίζεται με την ταξινόμηση των blocks που έχουν μέχρι στιγμής ανιχνευθεί, σε blocks κειμένου και μη κειμένου. Προτείνεται ένα σύνολο από 9 χαρακτηριστικά, εκ των οποίων τα 5 είναι συμβατικά, και με ένα σύστημα από κανόνες ταξινόμησης εκτελείται η τελική διαδικασία της ταξινόμησης.

4.5.3. Text extraction from colored book and journal covers

K. Sobottka, H. Kronenberg, T. Perroud, H. Bunke [25]

Παρουσιάζονται δύο μέθοδοι για την εξαγωγή πιθανών περιοχών κειμένου από έγγραφο, εγγραφο και αφορούν εφαρμογές για εξώφυλλα βιβλίων και περιοδικών. Η πρώτη μέθοδος βασίζεται στην «από πάνω προς τα κάτω ανάλυση» (top-down analysis) ενώ η δεύτερη στην «από κάτω προς τα πάνω ανάλυση» (bottom-up analysis). Τα αποτελέσματα και των δύο μεθόδων συνδυάζονται με σκοπό την τελική εξαγωγή του κειμένου. Οι εικόνες που χρησιμοποιούνται είναι 24 bit βάθους χρώματος, έχουν σαρωθεί με ανάλυση 200 dpi και το μέγεθός τους είναι της τάξης των 1300x1800 pixels. Η προϋπόθεση είναι το κείμενο να είναι βρίσκεται σε οριζόντια διάταξη.

Κατά το αρχικό στάδιο, το έγγραφο εγγραφο υφίσταται μείωση των χρωμάτων του με μια χωρίς επίβλεψη διαδικασία τμηματοποίησης του χρωματικού χώρου (unsupervised color space clustering) βασισμένη στον αλγόριθμο graph-theoretical clustering ή όπως αναφέρεται στην εργασία «histogram-based clustering technique».

Στη συνέχεια εφαρμόζονται οι δύο αλγόριθμοι τμηματοποίησης ώστε να εξαχθούν οι πιθανές περιοχές κειμένου.

Ο πρώτος (top-down) εφαρμόζει μία διαδοχική διαδικασία κατάτμησης του εγγράφου, γνωστή ως XY-tree decomposition (ή απλά XY-cut), όπου για κάθε σάρωση γραμμής (ή στήλης) υπολογίζεται ο αριθμός των χρωμάτων και στην περίπτωση που είναι μικρότερος από ένα κατώφλι τότε η γραμμή (ή η στήλη) θεωρείται ότι ανήκει στο background, αλλιώς ανήκει σε πιθανή περιοχή κειμένου.

Ο δεύτερος (bottom-up) αφορά μια τεχνική ανάπτυξης περιοχών όπου αρχίζοντας από μια ομοιογενή (στην ίδια κλάση) περιοχή 1x3 ή 3x1 pixels τα οποία βρίσκονται μέσα σε μια γειτονιά 3x3 ομαδοποιούνται διαδοχικά αν ανήκουν στην ίδια κλάση. Η διαδικασία τερματίζεται μέχρι να εξεταστούν όλα τα pixel ή αν δεν μπορεί να βρεθεί άλλη περιοχή εκκίνησης.

Ακολούθως, τα δύο αποτελέσματα συνδυάζονται ώστε να δώσουν τις τελικές περιοχές κειμένου με βάση τα γεωμετρικά και τα χωρικά χαρακτηριστικά των περιοχών που ανιχνεύθηκαν. Τα τελικά βήματα που συνθέτουν την εξαγωγή κειμένου

περιλαμβάνουν αρχικά την ομαδοποίηση των περιοχών και στη συνέχεια την δημιουργία των γραμμών κειμένου (text lines).

4.5.4.A word extraction al Pattern Recognition algorithm for machine-printed documents using a 3D neighborhood graph model

Hwan-Chul Park, Se-Young Ok, Young-Jung Yu, Hwan-Gue Cho [21]

Προτείνονται δύο μέθοδοι για τον εντοπισμό μεμονωμένων χαρακτήρων και στη συνέχεια την ομαδοποίησή τους σε λέξεις (word grouping). Η διαδικασία αφορά δυαδικές εικόνες εγγράφων όπου μαζί με το κείμενο συνυπάρχουν και άλλα στοιχεία, όπως εικόνες και γραφήματα (mixed text/graphic machine printed documents). Όσον αφορά τον τύπο του κειμένου, η μέθοδος δεν περιορίζεται σε οριζόντιες και κάθετες διευθύνσεις.

Αρχικά, ανιχνεύονται όλα τα συνδεδεμένα στοιχεία 8^{th} γειτονιάς και απορρίπτονται όσα έχουν μέγεθος κάτω από ένα κατώφλι T (6 pixel). Στη συνέχεια κατηγοριοποιούνται τα υπόλοιπα στοιχεία σε τρεις κατηγορίες: γενικοί χαρακτήρες, αρνητικοί χαρακτήρες (άσπρα γράμματα πάνω σε μαύρο φόντο) και εικόνες.

Για να εντοπιστούν οι χαρακτήρες από το έγγραφο γίνεται χρήση γεωμετρικών καθώς και στατιστικών χαρακτηριστικών όπως μέγεθος, επιμήκυνση και πυκνότητα.

Η εξαγωγή των λέξεων με τη χρήση της πληροφορίας του προηγούμενου βήματος γίνεται με ένα τρισδιάστατο μοντέλο γραφήματος (3D Graph model). Μέσα σε έναν τρισδιάστατο ευκλείδειο χώρο, τα x, y των κόμβων (vertex) απεικονίζονται βάσει των συντεταγμένων που είχαν στο δισδιάστατο χώρο (αρχική εικόνα) και το z αναφέρεται στο μέγεθος της περιοχής που περικλείει ένα γράμμα (bounding box size). Η τελική τμηματοποίηση (σύνδεση γραμμμάτων) βασίζεται σε μέτρηση των αποστάσεων μεταξύ των σημείων στο χώρο.

4.5.5.Hybrid approach to efficient text extraction in complex color images

Keechul Jung, JungHyun Han [13]

Προτείνεται μία υβριδική μέθοδος για την ανίχνευση και εξαγωγή κειμένου σε σύνθετες έγχρωμες εικόνες εγγράφων και video. Συνδυάζονται δύο τεχνικές, η μία βασίζεται στην προσέγγιση της υφής (texture-based methods) και η άλλη στα συνδεδεμένα στοιχεία (connected component based methods). Με τον τρόπο αυτό γίνεται προσπάθεια να αποφύγει τα μειονεκτήματα που παρουσιάζουν οι δύο αυτές μέθοδοι όταν εφαρμόζονται αυτόνομα.

Τα γενικά βήματα που ακολουθούνται στην παρούσα εργασία είναι:

1. Δημιουργία εικόνας όπου σε κάθε pixel προσδίδεται μία τιμή που αντιπροσωπεύει την πιθανότητα να ανήκει στο κείμενο (TPI – Text Probability Image). Για την απόδοση των τιμών της πιθανότητας γίνεται χρήση νευρωνικών δικτύων τύπου MLP (Multi-layer Perceptron).

2. Μείωση χρωμάτων της εικόνας σε 9-15 κλάσεις και ανίχνευση συνδεδεμένων στοιχείων τα οποία φιλτράρονται ώστε να απορριφθούν τυχόν εσφαλμένες περιοχές του προηγούμενου βήματος.
3. Επεξεργασία του αποτελέσματος του βήματος 2 ανάλογα με τον τύπο της εικόνας. Αν η εικόνα είναι έγχρωμο ή gray scale έγγραφο τότε ακολουθείται η επεξεργασία του με τον αλγόριθμο X-Y recursive cut και αν είναι video τότε εφαρμόζεται ο αλγόριθμος CAMshift (adaptive mean shift algorithm) που αποτελεί μια προσαρμοστική παραλλαγή του mean shift.

5.Γενικές πληροφορίες

5.1.Διεθνή περιοδικά συσχετιζόμενα με την ΑΔΣ

1. Pattern Recognition - Elsevier
2. Pattern Recognition Letters - Elsevier
3. Computer Vision and Image Understanding - Elsevier
4. Image and vision computing - Elsevier
5. Engineering Applications of Artificial Intelligence - Elsevier
6. Lecture Notes in Computer Science - Springer
7. International Journal on Document Analysis and Recognition (IJ DAR) - Springer
8. Transactions on Pattern Analysis and Machine Intelligence (TPAMI) - IEEE

5.2.Σύνδεσμοι διαδικτύου

CEDAR <http://www.cedar.buffalo.edu>
Center of Excellence for Document Analysis
and Recognition

Document Image Understanding Information <http://documents.cfar.umd.edu>
Server

Keith Price Bibliography Contents for <http://iris.usc.edu/Vision-Notes/bibliography/contentschar.html>
Document Analysis and Character
Recognition Systems

Imadoc: Electronic Reading and Writing <http://www.irisa.fr/imadoc/HTML/Welcome.en.html>

Document Processing and Understanding <http://research.microsoft.com/dpu/>
Microsoft Research

6.Βιβλιογραφία

1. Jung, K., Kim, K.I., Jain, A.K., Text information extraction in images and video: a survey, Pattern recognition, No. 5, May 2004, pp. 977-997.

2. D. Chen, J. Luetttin, K. Shearer, A survey of text detection and recognition in images and videos, Institut Dalle Molle'Intelligence Artificielle Perceptive (IDIAP) Research Report, IDIAP-RR 00-38, August 2000.
3. A.K. Jain, B. Yu, Document representation and its application to page decomposition, IEEE Trans. PAMI 20 (3) (1998) 294–308.
4. Y.Y. Tang, S.W. Lee, C.Y. Suen, Automatic document processing: a survey, Pattern Recognition 29 (12) (1996) 1931–1952.
5. C. Strouthopoulos, N. Papamarkos and A. Atsalakis, Text extraction in complex color documents, Pattern Recognition, Vol. 35, No. 2 , pp. 1743-1758, 2002.
6. C. Strouthopoulos, N. Papamarkos and C. Chamzas, PLA using RLSA and a neural network, Engineering Applications of Artificial Intelligence 12, 119-138 (1999).
7. C. Strouthopoulos and N. Papamarkos, Text identification for document image analysis using a neural network, Image and Vision Computing-Special Issue on Document Image Processing and Multimedia Environments 16(12-13), 879-896 (1998).
8. C. Strouthopoulos, N. Papamarkos and C. Chamzas, Identification of text-only areas in mixed type documents, Engineering Applications of Artificial Intelligence 10(2), 387-401 (1997).
9. Dihua Xi and Seong-Whan Lee, Extraction of reference lines and items from form document images with complicated background, Pattern Recognition, Volume 38, Issue 2, February 2005, Pages 289-305.
10. Qixiang Ye, Wen Gao, Qingming Huang, Automatic Text Segmentation from Complex Background. IEEE International Conference on Image Processing (ICIP 2004), Singapore, October 24-27, 2004.
11. Phillip E. Mitchell and Hong Yan, Newspaper layout analysis incorporating connected component separation, Image and Vision Computing, Volume 22, Issue 4, 1 April 2004, Pages 307-317.
12. Yi Xiao and Hong Yan, Location of title and author regions in document images based on the Delaunay triangulation, Image and Vision Computing, Volume 22, Issue 4, 1 April 2004, Pages 319-329.
13. Keechul Jung, JungHyun Han, Hybrid approach to efficient text extraction in complex color images, Pattern Recognition Letters, v.25 n.6, p.679-699, 19 April 2004.
14. Hiroyuki Hase, Masaaki Yoneda, Shogo Tokai, Jien Kato and Ching Y. Suen, Color segmentation for text extraction, International Journal on Document Analysis and Recognition 6(4): 271-284 (2003).
15. Yi Xiao and Hong Yan, Text region extraction in a document image based on the Delaunay tessellation, Pattern Recognition, Volume 36, Issue 3, March 2003, Pages 799-809.
16. Hui Cheng, Zhigang Fan, Background Identification Based Segmentation and Multilayer Tree Representation of Document Images. IEEE International Conference on Image Processing (ICIP 2002), New York, September 22-25, 2002.
17. L. Cinque, S. Levialdi, L. Lombardi and S. Tanimoto, Segmentation of page images having artifacts of photocopying and scanning, Pattern Recognition, Volume 35, Issue 5, May 2002, Pages 1167-1177.
18. Paul Clark, Majid Mirmehdi, Recognising text in real scenes, International Journal on Document Analysis and Recognition 4(4): 243-257 (2002).
19. Hideaki Goto and Hiroto Aso, Character pattern extraction from documents with complex backgrounds, International Journal on Document Analysis and Recognition 4(4): 258-268 (2002).
20. Jie Xi, Jianming Hu and Lide Wu, Page segmentation of Chinese newspapers, Pattern Recognition, Volume 35, Issue 12, December 2002, Pages 2695-2704.
21. Hwan-Chul Park, Se-Young Ok, Young-Jung Yu, Hwan-Gue Cho, A word extraction algorithm for machine-printed documents using a 3D neighborhood graph model, International Journal on Document Analysis and Recognition 4(2): 115 - 130 (2001).
22. Shulan Deng, Shahram Latifi and Emma Regentova, Document segmentation using polynomial spline wavelets, Pattern Recognition, Volume 34, Issue 12, December 2001, Pages 2533-2545.

23. H. Hase, T. Shinokawa, M. Yoneda, C.Y. Suen, Character string extraction from color documents, *Pattern Recognition* 34 (7) (2001) 1349–1365.
24. Seong-Whan Lee and Dae-Seok Ryu, Parameter-Free Geometric Document Layout Analysis, *IEEE Trans. PAMI* 23 (11) (2001) 1240-1256.
25. K. Sobottka et al, Text Extraction from Colored Book and Journal Covers, *International Journal on Document Analysis and Recognition*, Vol. 2, No. 4, 2000, pp. 163-176.
26. C. Garcia, X. Apostolidis, Text detection and segmentation in complex color images, *International Conference on Acoustics, Speech and Signal Processing*, 2000, pp. 2326–2329.
27. George Nagy, Twenty Years of Document Image Analysis in PAMI, *IEEE Trans. PAMI* 22 (1) (2000) 38–62.
28. V. Wu, R. Manmatha, E.M. Riseman, TextFinder: an automatic system to detect and recognize text in images, *IEEE Trans. PAMI* 21 (11) (1999) 1224–1229.
29. P. Parodi and R. Fontana, Efficient and Flexible Text Extraction from document Pages, *International Journal on Document Analysis and Recognition* 2, 67-79 (1999).
30. S. Messelodi, C.M. Modena, Automatic identification and skew estimation of text lines in real scene images, *Pattern Recognition* 32 (1999) 791–810.
31. Apostolos Antonacopoulos, Page Segmentation Using the Description of the Background, *Computer Vision and Image Understanding*, Volume 70, Issue 3, June 1998, Pages 350-369.
32. L. Cinque, L. Lombardi and G. Manzini, A multiresolution approach for page segmentation, *Pattern Recognition Letters*, Volume 19, Issue 2, February 1998, Pages 217-225.
33. C. L. Tan and P. O. Ng, Text extraction using pyramid, *Pattern Recognition*, Volume 31, Issue 1, January 1998, Pages 63-72.
34. A.K. Jain, B. Yu, Automatic text location in images and video frames, *Pattern Recognition* 31 (12) (1998) 2055–2076.
35. W.Y. Chen and S.Y. Chen, Adaptive page segmentation for color technical journals' cover images, *Image and Vision Computing* 16, 855-877 (1998).
36. J. Zhou and D. Lopresti. Extracting text from WWW images. In *International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 248-252, August 1997
37. A.K. Jain, Y. Zhong, Page segmentation using texture analysis, *Pattern Recognition* 29 (5) (1996) 743–770.
38. Y. Zhong, K. Karu, A.K. Jain, Locating text in complex color images, *Pattern Recognition* 28 (10) (1995) 1523–1535.
39. M. Suen and J.F. Wang, Text string extraction from images of color printed documents, *Proc. IPPR Conf. Computer Vision, Graphics Image Processing, Taiwan*, 534-541 (1995).
40. Yi Lu, Machine printed character segmentation — An overview, *Pattern Recognition*, Volume 28, Issue 1, January 1995, Pages 67-80.
41. K.C. Fan, C.H. Liu and Y.K. Wang, Segmentation and classification of mixed text/graphics/image documents, *Pattern Recognition Letters* 15, 1201-1209 (1994).
42. J. Ohya, A. Shio, S. Akamatsu, Recognizing characters in scene images, *IEEE Trans. PAMI* 16 (1994) 214–224
43. L. O' Gorman, The Document Spectrum for Page Layout Analysis, *IEEE Trans. PAMI* 15, 1162-1173 (1993).
44. K. Jain, S. Bhattacharjee, Text segmentation using Gabor Filters for automatic document processing, *Mach. Vision Appl.* 5 (1992) 169–184.
45. L. Fletcher, R. Kasturi, A robust algorithm for text string separation from mixed text/graphics images, *IEEE Trans. PAMI* 10 (1988) 910–918.
46. G. Nagy, S.C. Seth, S.D. Stoddard, Document analysis with an expert system, in: E.S. Gelsema and L.N. Kanal (Eds.), *Pattern Recognition in Practice II*, Elsevier, Amsterdam, 1986, pp. 149–197.
47. E.G. Johnston, Printed text discrimination, *Comp. Graphics Image Process* 3 (1) (1974) 83–89.