UNIVERSITY OF THE AEGEAN

Department of Information and Communication Systems Engineering

Document Image Segmentation and Text Localization

Submitted in Total Fulfilment of the Requirements
for the Degree of Doctor of Philosophy (Ph.D.)

Nikos Vasilopoulos

May 2016

Samos, Greece

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Επίκουρος Καθηγήτρια Εργίνα Καβαλλιεράτου
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
Πανεπιστήμιο Αιγαίου

Αναπληρωτής Καθηγητής Ευστάθιος Σταματάτος
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
Πανεπιστήμιο Αιγαίου

Καθηγητής Αθανάσιος Σκόδρας
Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας Υπολογιστών
Πανεπιστήμιο Πατρών

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΣΑΜΟΣ 2016

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ
ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Επίκουρος Καθηγήτρια Εργίνα Καβαλλιεράτου
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
Πανεπιστήμιο Αιγαίου

Αναπληρωτής Καθηγητής Ευστάθιος Σταματάτος
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
Πανεπιστήμιο Αιγαίου

Επίκουρος Καθηγητής Μανώλης Μαραγκουδάκης
Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων
Πανεπιστήμιο Αιγαίου

Καθηγητής Αθανάσιος Σκόδρας
Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας Υπολογιστών
Πανεπιστήμιο Πατρών

Καθηγητής Βασίλης Αναστασόπουλος
Τμήμα Φυσικής
Πανεπιστήμιο Πατρών

Καθηγητής Νικόλαος Παπαμάρκος
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Δημοκρίτειο Πανεπιστήμιο Θράκης

Ερευνητής Β' Βασίλης Γάτος
ΕΚΕΦΕ Δημόκριτος

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΣΑΜΟΣ 2016

ΠΕΡΙΛΗΨΗ

Σε αυτή τη διατριβή διερευνώνται θέματα που σχετίζονται με την κατάτμηση των σελίδων και την εξαγωγή των πληροφοριών κειμένου από εικόνες εγγράφων. Νέες προσεγγίσεις για την αντιμετώπιση προβλημάτων παρουσιάζονται και, πιο συγκεκριμένα, μια μέθοδος ανάλυσης διάταξης σελίδας, μια τεχνική εντοπισμού κειμένου σε εικόνα, καθώς και ένα σύστημα word-spotting και ανάκτησης κειμένου. Δύο εργαλεία λογισμικού αναπτύχθηκαν με σκοπό να δοκιμαστεί η απόδοση των προτεινόμενων αλγορίθμων. Αποτελέσματα λεπτομερών πειραμάτων παρουσιάζονται αναλυτικά.

Στο Κεφάλαιο 2, παρουσιάζεται μια μέθοδος για την κατάτμηση εγγράφων με σύνθετη διάταξη (εφημερίδες, περιοδικά κλπ.). Δεν είναι απαραίτητη καμία γνώση για τη μορφή της σελίδας a priori. Μορφολογικοί τελεστές εφαρμόζονται προκειμένου να συνδεθούν γειτονικές περιοχές και να εντοπιστούν διαχωριστικές γραμμές και στήλες. Τεχνικές ανίχνευσης περιγράμματος χρησιμοποιούνται στη συνέχεια για την εξαγωγή πληροφοριών σχήματος και την ταξινόμηση των συνδεδεμένων αντικειμένων.

Στο Κεφάλαιο 3, προτείνεται μια υβριδική μέθοδος για τον εντοπισμό, σε πραγματικό χρόνο, κειμένου που είναι ενσωματωμένο σε εικόνες. Συνδυάζει ανίχνευση ακμών, μορφολογικούς τελεστές και ένα σύνολο κριτηρίων με βάση χωρικά και γεωμετρικά χαρακτηριστικά των συνδεδεμένων αντικειμένων.

Στο Κεφάλαιο 4, προτείνεται μια τεχνική κατάλληλη για την εξαγωγή όλων των πληροφοριών κειμένου από έγγραφα με σύνθετες διατάξεις. Συνδυάζει τμήματα της μεθόδου ανάλυσης διάταξης που παρουσιάζεται στο Κεφάλαιο 2 με την γρήγορη και αξιόπιστη μέθοδο για τον εντοπισμό κειμένου που παρουσιάζεται στο Κεφάλαιο 3. Η πρώτη χρησιμοποιείται για το διαχωρισμό των περιοχών της σελίδας σε κείμενο και εικόνες, ενώ η δεύτερη χρησιμοποιείται για την ανίχνευση κειμένου που μπορεί να περιέχεται μέσα στις εικόνες.

Στο Κεφάλαιο 5, προτείνεται ένα σύστημα word-spotting, κατάλληλο για την αναζήτηση κειμένου σε εκτυπωμένες εικόνες ιστορικών εγγράφων. Το σύστημα απλοποιεί αρκετά τη διαδικασία της συνηθισμένης προσέγγισης. Δεν περιλαμβάνει κατάτμηση, εξαγωγή χαρακτηριστικών ή ταξινόμηση. Αντίθετα, αντιμετωπίζει τα ερωτήματα ως συμπαγή σχήματα και χρησιμοποιεί τεχνικές επεξεργασίας εικόνας, προκειμένου να εντοπιστεί ένα ερώτημα στις εικόνες των εγγράφων.

Στο Κεφάλαιο 6, προτείνεται μια νέα τεχνική για την ανάκτηση κειμένου. Αν και είναι εμπνευσμένη από την τεχνική word-spotting που παρουσιάζεται στο Κεφάλαιο 5, εντοπίζει χαρακτήρες, γεγονός που καθιστά την ανάκτηση πιο ισχυρή και επιτρέπει τη χρήση ερωτημάτων σε μορφή κειμένου.

Στο Κεφάλαιο 7 συνοψίζονται τα συμπεράσματα από όλα τα προηγούμενα κεφάλαια, ενώ στα προσαρτήματα διερευνώνται τρεις ακόμα εφαρμογές των προτεινόμενων αλγορίθμων κατάτμησης σελίδας και ανίχνευσης κειμένου: κατάτμηση σελίδων κόμικς, εντοπισμός κειμένου σε φωτογραφίες και εντοπισμός κειμένου σε βίντεο.

ABSTRACT

In this work various issues related to the segmentation of pages and the extraction of textual information from document images are investigated. Novel approaches to challenging problems are presented, namely a layout analysis method, a text localization technique as well as a word-spotting and text retrieval system. Two software tools have been implemented in order to test the proposed algorithms. Results of detailed experiments are also shown.

In Chapter 2, a method for the segmentation of complex (newspapers, magazines etc.) layouts is presented. No a priori knowledge of the page format is needed. Morphological operations are applied to both the foreground and the background, in order to connect neighboring regions and detect separator lines and columns respectively. Contour tracing is used for the extraction of shape information and classification of the connected components.

In Chapter 3, a hybrid method for real-time localization of text embedded in images is proposed. It combines edge detection, morphological operators and a set of criteria based on spatial and geometrical features of the connected components.

In Chapter 4, a technique appropriate for extracting all the textual information from documents with complex layouts is proposed. It integrates the foreground analysis part of the layout independent method presented in Chapter 2 with the fast and reliable method for text localization presented in Chapter 3. The first one is used to segment the page in text and image blocks while the second one is used to detect text that may be embedded inside the images.

In Chapter 5, a classification-free word-spotting system, appropriate for the retrieval of printed historical document images is proposed. The system skips many of the procedures of a common approach. It does not include segmentation, feature extraction or classification. Instead it treats the queries as compact shapes and uses image processing techniques in order to localize a query in the document images.

In Chapter 6, a novel technique for text retrieval is proposed. Although, it is inspired by the word spotting technique presented in Chapter 5, it spots characters, which makes the text retrieval more robust and permits the text queries.

Chapter 7 summarizes the conclusions from all previous chapters, while in the appendixes three more applications of the proposed segmentation and text detection algorithms are explored: comic page segmentation, scene text localization and video text localization.

AKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Ergina Kavallieratou. Her contribution in major parts of this work and her kind support throughout all the phases are greatly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

Document images contain invaluable information for our past and, at the same time, they play a major role in daily life. Large collections of historical documents are constantly scanned and analyzed by historians and other researchers, while in every office huge amounts of received documents are continuously encoded for further processing every day. The automatic segmentation and classification of the document image blocks as well as the extraction of the textual information is of great importance for all those cases, especially for documents with complex layouts that cannot be directly processed by Optical Character Recognition (OCR) systems. Although many techniques have been proposed for the above tasks, the available tools are far from being fully automated. They often require to manually select or split missing or overlapping regions respectively. Moreover, the processing time of commercial software packets is not always acceptable, especially when big datasets are considered.

In this work, the document image segmentation and text localization methods found in the literature are investigated and solutions for the following problems are being explored:

- Complex layouts (newspapers, journals etc.) can hardly be analyzed by existing methods in case of multi-modal (various text and background colors in the same page) document images.
- Segmentation of a page often requires a priori knowledge of the font characteristics (sizes, spacing etc.) and the page structure (column size, inter-block gaps etc.).
- Although segmentation and localization speed is very crucial in some applications, the existing software tools don’t perform fast enough.
- Text embedded in frames (banners, logos, photographs, diagrams etc.) cannot always be extracted using page segmentation only.

In order to achieve good segmentation results, even for multi-modal pages, a layout analysis algorithm, based on local binarization, is implemented. No a priori knowledge is required. Information about the page fonts and structure is collected automatically. The input images are resized by 50% so that the processing speed is high. The ultimate goal is to extract all the text from the page, even the words and phrases that are embedded inside frames and images. For that reason, page segmentation is combined with text localization in a novel unified framework.

Sometimes, extracting all the textual information from a document image is not actually needed. For example, in case we are looking for instances of either a word or a phrase in a documents database, only the location of the closest matches between the query and each page content has to be retrieved. Moreover, if the database contains many documents, the segmentation and text localization procedure can be computationally very expensive. Systems that search for possible locations of a query image of a keyword inside a database of documents are called word-spotting systems. Modern word-spotting systems bypass the page segmentation step and provide results much faster than older systems. In this work a segmentation-free fast and reliable word-spotting tool is implemented and used for searching a collection of historical documents.

Word-spotting systems work with image queries that are manually selected from a sample of the dataset, thus limit the search results to words included in the sample page. Tools that allow for searching of any text that is entered by the user (query by text) are of great importance, for obvious reasons. Such tools should rather perform well even with handwritten documents. This is considered a very challenging task. An attempt towards this direction is made in this work as well.

Summarizing, the contribution of this work consists mainly of the:

- binarization of multi modal documents using local band thresholding,
- application of morphological operations with long lines for layout analysis,
- combination of various heuristics for the classification of text and non-text regions,
- integration of page segmentation with text localization for full text extraction,
- presentation of a query-by-text search method based on character spotting,
- development of fast and reliable software tools for the above tasks.

This chapter includes only a short introduction. Since most of the described tasks are considered separate topics, more details can be found inside the following chapters that focus at every topic independently. Each chapter consists of an introductory section, where the key concepts and the related work are presented, a major section that describes the proposed system and an experiments section, where evaluation results are shown. Most of the chapters include a “state of art” section, where the existing techniques are mentioned. None of them includes a conclusions section. Although some conclusions are reported in the experiments sections, one more chapter is added where the final conclusion is drawn. The algorithms that are proposed in this work have been coded in C# language, using the OpenCV library as well. Detailed experiments have been made, using publicly available datasets.

In Chapter 2, a method for the segmentation of complex (newspapers, magazines etc.) layouts is presented. No a priori knowledge of the page format is needed. Morphological operations are applied to both the foreground and the background, in order to connect neighboring regions and detect separator lines and columns respectively. Contour tracing is used for the extraction of shape information and classification of the connected components.

In Chapter 3, a hybrid method for real-time localization of text embedded in images is proposed. It combines edge detection, morphological operators and a set of criteria based on spatial and geometrical features of the connected components.

In Chapter 4, a technique appropriate for extracting all the textual information from documents with complex layouts is proposed. It integrates the foreground analysis part of the layout independent method presented in Chapter 2 with the fast and reliable method for text localization presented in Chapter 3. The first one is used to segment the page in text and image blocks while the second one is used to detect text that may be embedded inside the images.

In Chapter 5, a classification-free word-spotting system, appropriate for the retrieval of printed historical document images is proposed. The system skips many of the procedures of a common approach. It does not include segmentation, feature extraction or classification. Instead it treats the queries as compact shapes and uses image processing techniques in order to localize a query in the document images.

In Chapter 6, a novel technique for text retrieval is proposed. Although, it is inspired by the word spotting technique presented in Chapter 5, it spots characters, which makes the text retrieval more robust and permits the text queries.

Chapter 7 summarizes the conclusions from all previous chapters, while in the appendixes three more applications of the proposed segmentation and text detection algorithms are explored: comic page segmentation, scene text localization and video text localization.

Part of the contribution of this thesis has been presented in conferences and/or submitted for publication.

Referred International Conferences:

- Vasilopoulos Nikos and Ergina Kavallieratou. “A classification-free word-spotting system” *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, 2013
- Vasilopoulos Nikos and Ergina Kavallieratou. “Complex layout analysis based on contour classification and morphological operations”, *9th Hellenic Conference on Artificial Intelligence Proceedings*, 2016

Under Review:

- Vasilopoulos Nikos and Ergina Kavallieratou. “Complex layout analysis based on contour classification and morphological operations”, *International Journal on Document Analysis and Recognition*
- Vasilopoulos Nikos and Ergina Kavallieratou. “Real-time layout analysis of newspapers and journals”, *International Journal on Artificial Intelligence Tools*
- Vasilopoulos Nikos and Ergina Kavallieratou. “A unified layout analysis and text localization framework”, *Electronic Letters on Computer Vision and Image Analysis*
- Vasilopoulos Nikos and Ergina Kavallieratou. “Real-time text localization in complex images”, *26th International Conference on. Pattern Recognition (ICPR)*, 2016
- Vasilopoulos Nikos, Ergina Kavallieratou and Laurence Likforman-Sulem. “Word-spotting based on Character-spotting”, *Conference & Labs of the Evaluation Forum (CLEF)*, 2016

CHAPTER 2 - LAYOUT ANALYSIS

2.1 Introduction

The process of analyzing page images in order to identify physical (text, pictures etc.) and logical (titles, paragraphs etc.) structures is called layout analysis. Figure 2.1 shows the layout analysis tasks that are required before optical character recognition (OCR) takes place. The performance of layout analysis methods depends heavily on the page segmentation algorithm in use. The page segmentation methods that have been reported in the literature can be categorized into foreground analysis, background analysis and local analysis ones. Both foreground and background analysis techniques require a binarization step in order to distinguish between the foreground (black) and the background (white) pixels, while most local analysis methods are directly applied on either color or grayscale document images.

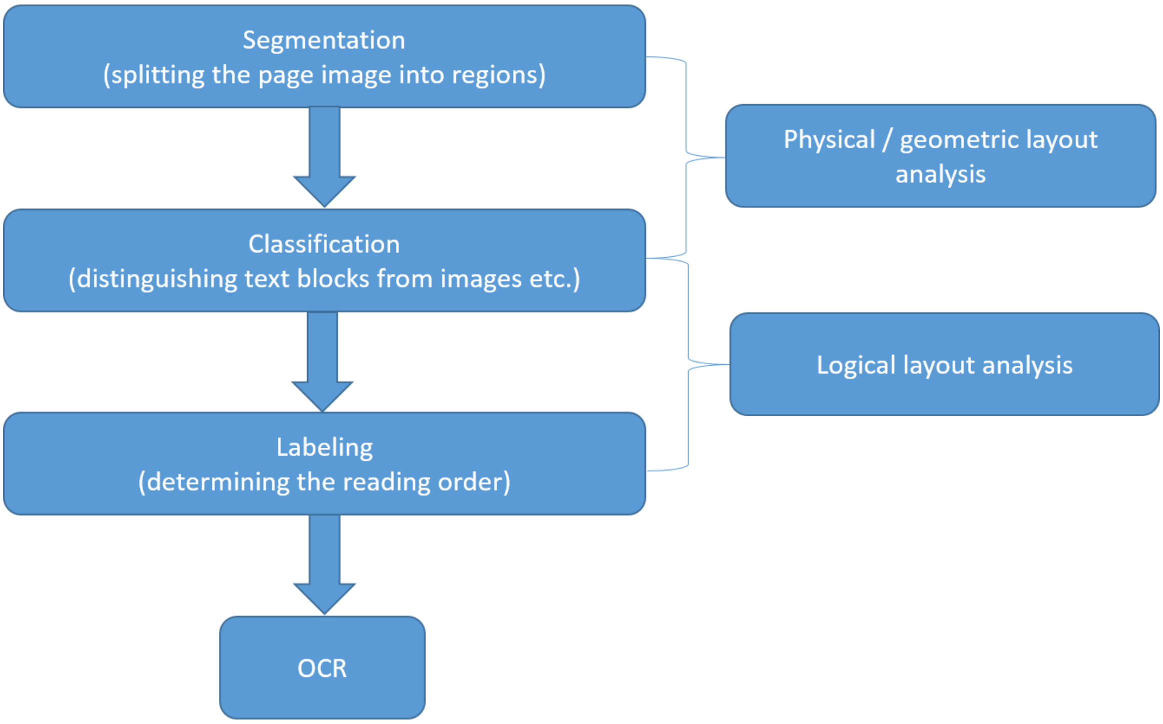

Segmentation
(splitting the page image into regions)
Classification
(distinguishing text blocks from images etc.)
Labeling
(determining the reading order)
OCR
Physical / geometric layout analysis
Logical layout analysis

Figure 2-1: A layout analysis block diagram

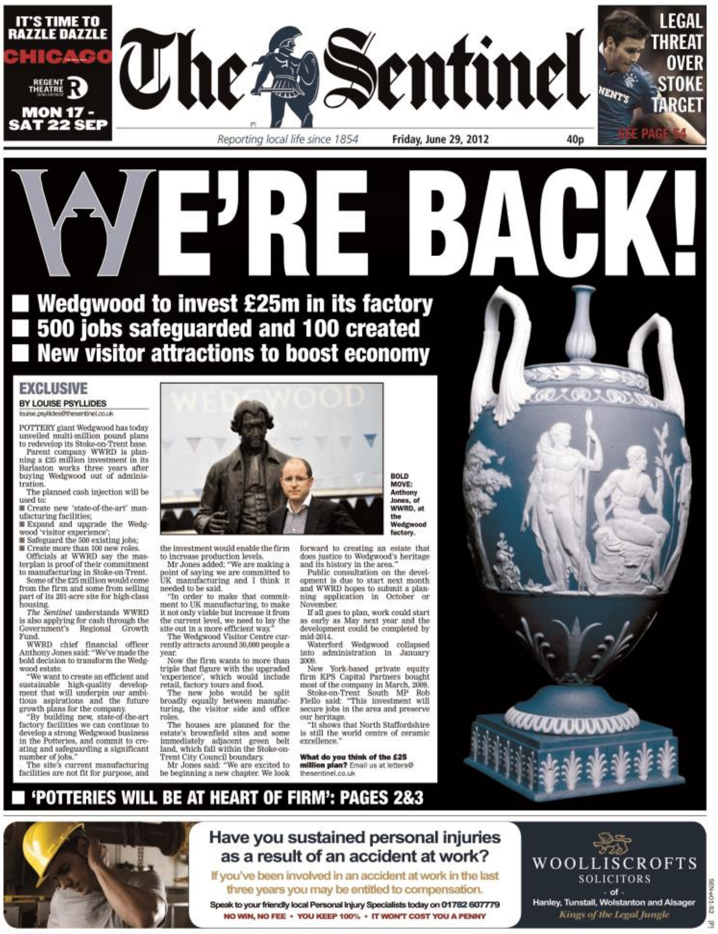
In this chapter a layout independent hybrid method for complex (newspapers, magazines etc.) layout analysis is proposed. No a priori knowledge of the page format is needed. Foreground analysis works with binary input images, therefore a binarization step is included. Morphological operations are applied to both the foreground and the background, in order to connect neighboring regions and detect separator lines and columns respectively. Contour tracing is used for the extraction of shape information and classification of the connected components.

The contribution of this work consists of the presentation of:

i. a novel and fast technique able to perform on real time

ii. subtask for complex binarization using regional band thresholding [1] and

iii. subtask to analyze the background using morphological opening with long lines.



The Sentinel

Reporting local life since 1854 Friday, June 29, 2012 40p

LEGAL THREAT OVER STOKE TARGET
SEE PAGE 54

WE'RE BACK!

- Wedgwood to invest £25m in its factory
- 500 jobs safeguarded and 100 created
- New visitor attractions to boost economy

EXCLUSIVE
BY LOUISE PSYLLIDES
louise.psyllides@thesentinel.co.uk

POTTERY giant Wedgwood has today unveiled multi-million pound plans to redevelop its Stoke-on-Trent base.

Parent company WWRD is planning a £25 million investment in its Barlaston works three years after buying Wedgwood out of administration.

The planned cash injection will be used to:

- Create new 'state-of-the-art' manufacturing facilities;
- Expand and upgrade the Wedgwood 'visitor experience';
- Safeguard the 500 existing jobs;
- Create more than 100 new roles.

Officials at WWRD say the masterplan is proof of their commitment to manufacturing in Stoke-on-Trent.

Some of the £25 million would come from the firm and some from selling part of its 281-acre site for high-class housing.

The Sentinel understands WWRD is also applying for cash through the Government's Regional Growth Fund.

WWRD chief financial officer Anthony Jones said: "We've made the bold decision to transform the Wedgwood estate.

"We want to create an efficient and sustainable high-quality development that will underpin our ambitious aspirations and the future growth plans for the company.

"By building new, state-of-the-art factory facilities we can continue to develop a strong Wedgwood business in the Potteries, and commit to creating and safeguarding a significant number of jobs."

The site's current manufacturing facilities are not fit for purpose, and the investment would enable the firm to increase production levels.

Mr Jones added: "We are making a point of saying we are committed to UK manufacturing and I think it needed to be said.

"In order to make that commitment to UK manufacturing, to make it not only viable but increase it from the current level, we need to lay the site out in a more efficient way."

The Wedgwood Visitor Centre currently attracts around 30,000 people a year.

Now the firm wants to more than triple that figure with the upgraded 'experience', which would include retail, factory tours and food.

The new jobs would be split broadly equally between manufacturing, the visitor side and office roles.

The houses are planned for the estate's brownfield sites and some immediately adjacent green belt land, which fall within the Stoke-on-Trent City Council boundary.

Mr Jones said: "We are excited to be beginning a new chapter. We look forward to creating an estate that does justice to Wedgwood's heritage and its history in the area."

Public consultation on the development is due to start next month and WWRD hopes to submit a planning application in October or November.

If all goes to plan, work could start as early as May next year and the development could be completed by mid-2014.

Waterford Wedgwood collapsed into administration in January 2009.

New York-based private equity firm KPS Capital Partners bought most of the company in March, 2009.

Stoke-on-Trent South MP Rob Flello said: "This investment will secure jobs in the area and preserve our heritage.

"It shows that North Staffordshire is still the world centre of ceramic excellence."

What do you think of the £25 million plan? Email us at letters@thesentinel.co.uk

BOLD MOVE: Anthony Jones, of WWRD, at the Wedgwood factory.

- 'POTTERIES WILL BE AT HEART OF FIRM': PAGES 2&3

SEN-e01-S2 [P]

Figure 2-2: A newspaper front page with a complex layout

In the Section 2.2, a short summary of the state of the art is given. The proposed technique and the included modules are presented in detail, in Section 2.3, while experimental and comparative results are presented on the RDCL-2015 dataset, in Section 2.4.

2.2 State of the art

Foreground analysis techniques use a bottom-up approach. They start from pixel level and merge regions together into larger components to form document structures (e.g. characters, then words, text lines, paragraphs and so on). Wong et al. classify connected components into text and non-text zones, after linking together neighboring black areas by performing a run-length smearing algorithm (RLSA) [2] (Fig. 2.3). Tsujimoto and Asada use the same smearing algorithm in order to aggregate adjacent connected components into segments by connecting two black runs separated by a small gap [3]. Strouthopoulos et al. also perform a run-length smearing operation and then classify the blocks using a principal component analyzer (PCA) and a self-organized feature map (SOFM) [4]. Sun proposed a modified smearing algorithm, called selective CRLA, capable of processing documents with non-Manhattan (Fig. 2.4), containing non-rectangular blocks, layouts [5]. The smearing algorithms can hardly be applied to skewed documents.

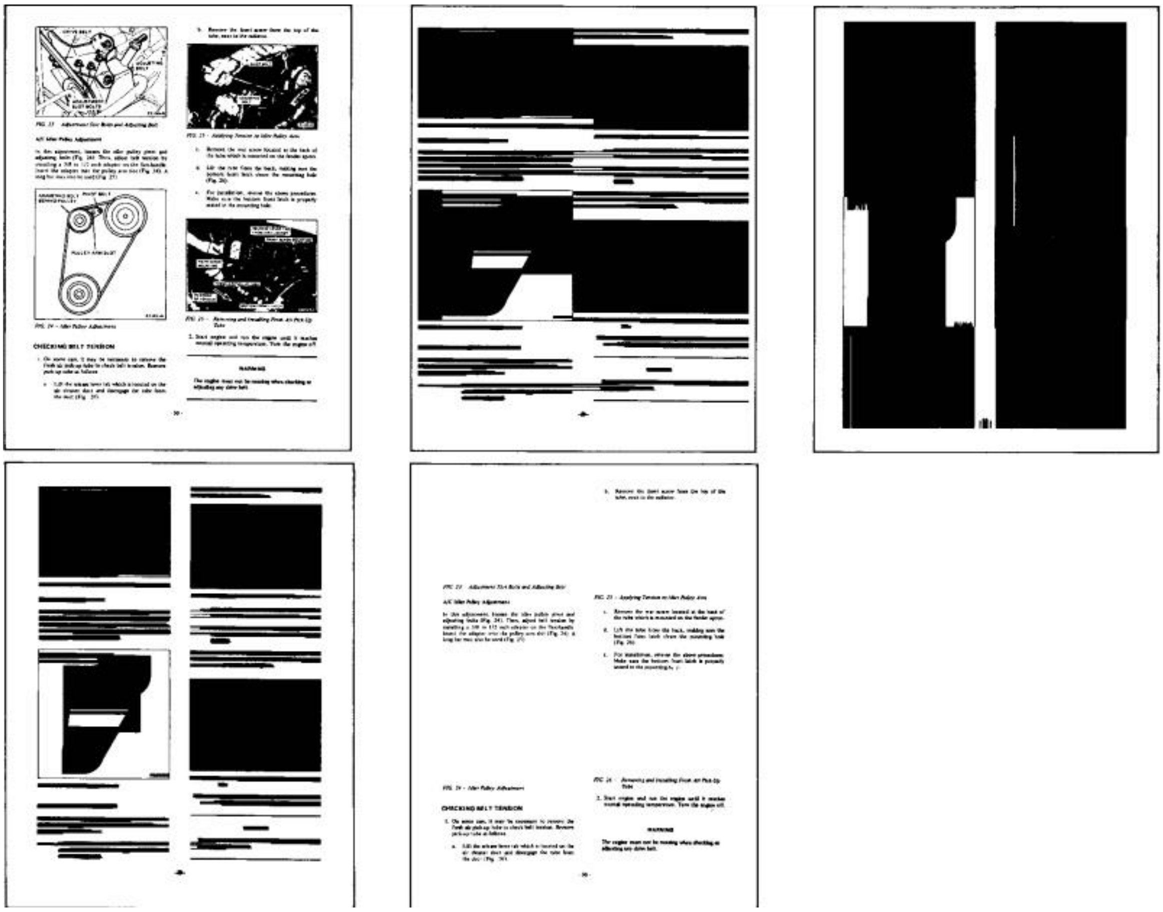


Figure 2-3: An example of page segmentation with the run-length smearing algorithm [2]

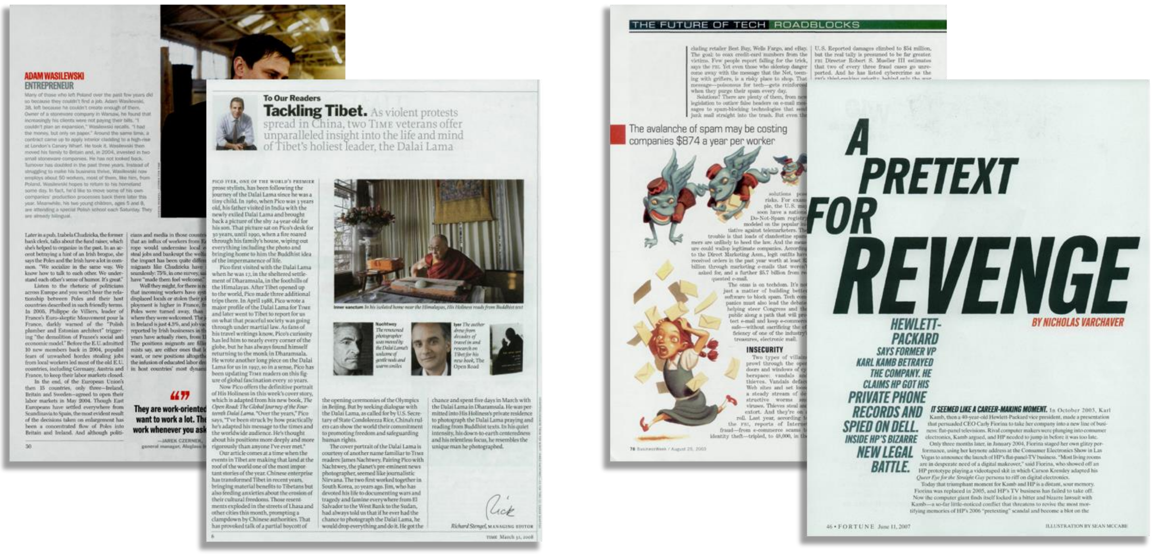


Figure 2-4: Manhattan (left) and non-Manhattan (right) layouts

Many of the foreground analysis methods proposed so far are based on connected components analysis. Fletcher et al. group together components into logical character strings using the Hough transform [6] (Fig. 2.5). Akiyama and Hagita start from extracting the field separators (solid and dotted lines) and continue with text extraction based on each connected component's height [7].

Gorman and Lawrence find the K nearest neighbors for each connected component and use distance thresholds to form text blocks [8]. Hönes and Lichter generate lines starting from triplets of neighbor components which are approximately aligned and have comparable size [9]. Zlatopolsky first groups text-like components into line segments and then into blocks, depending on their horizontal and vertical distances [10]. Déforges and Barba also merge words (extracted from a multi-resolution pyramidal representation of the image) into lines and blocks, according to their spatial relationships [11]. Wang and Yagasaki classify large (non-text) components by using characteristics of their outlines (size, thickness, density, number of holes etc.) and then cluster the remaining (textual) components into blocks by using a closeness criterion [12]. Simon and Pret also use a distance-metric between the components to construct the page structure [13]. Bukhari et al. use connected components shape and context information as a feature vector input to a self-tunable multi-layer perceptron (MLP) classifier [14]. Koo et al. start from extracting connected components and grouping them into text-lines [15]. Le et al. extract a set of features based on size, shape, stroke width and position of each connected component and label them by using Adaboosting with Decision trees [16]. Although skew independent, connected components analysis techniques require that inter-block gaps are wider than interline gaps. The performance of a number of connected components analysis algorithms has been evaluated in segmentation of newspapers [17] and other types of documents [18].

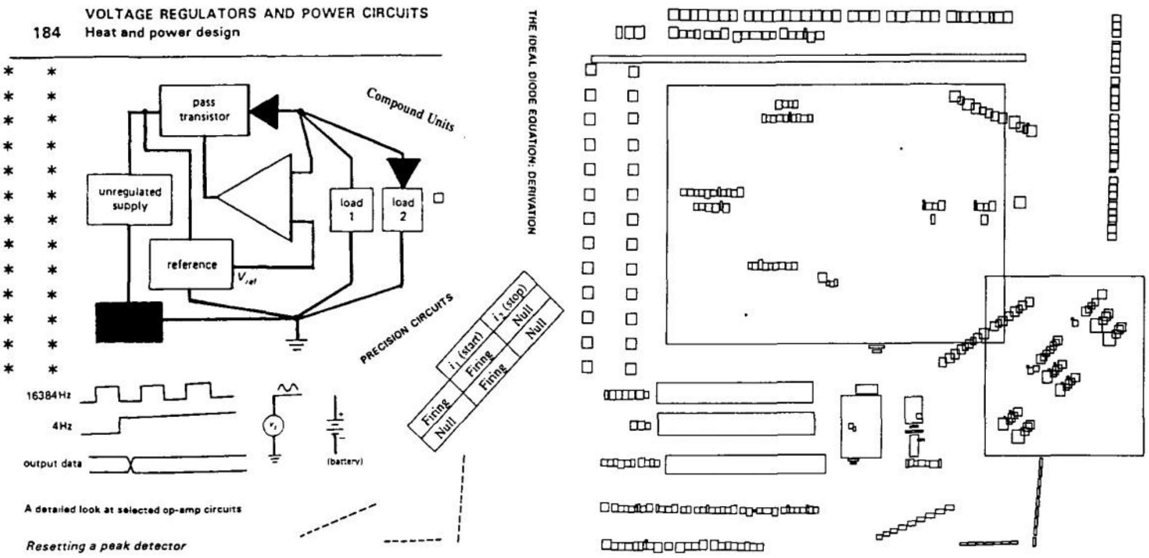
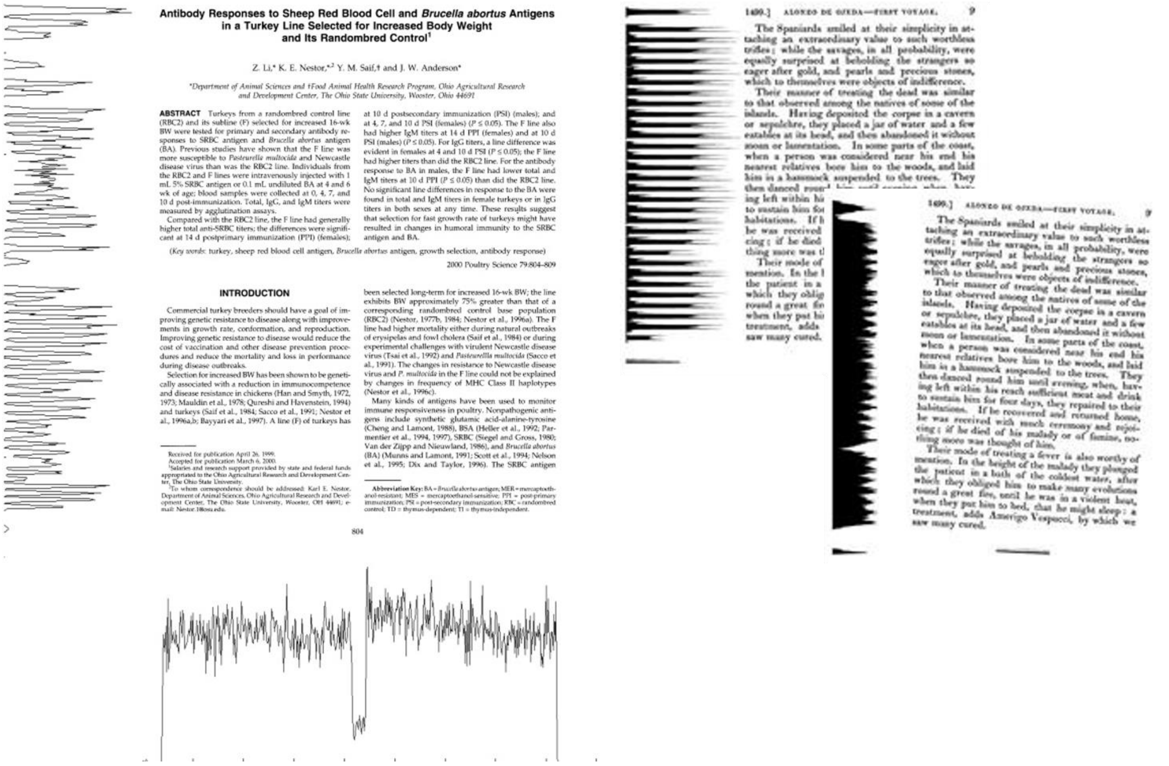


Figure 2-5: Bounding rectangles (right) of the connected components (left) [6]

Background analysis techniques use regions of white pixels to split the page into blocks which are subsequently identified and subdivided further. Nagy et al. recursively split the document at the valleys along the horizontal and vertical projection profiles [19] (Fig. 2.6). Ha et al. analyze the horizontal and vertical projections of the bounding boxes of the connected components [20]. Baird enumerates all the white rectangles covering the background which cannot be further expanded (maximal) in order to analyze the white space structure [21]. Breuel uses tall whitespace rectangles as obstacles in order to detect text-lines [22]. The above background analysis techniques are limited to Manhattan layouts (blocks surrounded by straight white streams) and are strongly affected by skew. Normand and Viard-Gaudin proposed an extension of the RLSA algorithm to two dimensions for the analysis of the document background, insensitive to the orientation of the text blocks [23]. Kise et al. proposed two more background analysis methods capable of segmenting pages with non-Manhattan layout as well as with various angles of skew: 1) thinning the white areas to form connected thin lines or chains and then finding the loops enclosing printed areas [24], 2) using the approximated area Voronoi diagram to obtain the candidates of boundaries of document components [25] (Fig. 2.7).



Antibody Responses to Sheep Red Blood Cell and *Brucella abortus* Antigens in a Turkey Line Selected for Increased Body Weight and Its Randombred Control[1]

Z. Li,* K. E. Nestor,*[2] Y. M. Saif,† and J. W. Anderson*

*Department of Animal Sciences and †Food Animal Health Research Program, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, Ohio 44691

ABSTRACT Turkeys from a randombred control line (RBC2) and its subline (F) selected for increased 16-wk BW were tested for primary and secondary antibody responses to SRBC antigen and *Brucella abortus* antigen (BA). Previous studies have shown that the F line was more susceptible to *Pasteurella multocida* and Newcastle disease virus than was the RBC2 line. Individuals from the RBC2 and F lines were intravenously injected with 1 mL 5% SRBC antigen or 0.1 mL undiluted BA at 4 and 6 wk of age; blood samples were collected at 0, 4, 7, and 10 d post-immunization. Total, IgG, and IgM titers were measured by agglutination assays.

Compared with the RBC2 line, the F line had generally higher total anti-SRBC titers; the differences were significant at 14 d postprimary immunization (PPI) (females); at 10 d postsecondary immunization (PSI) (males); and at 4, 7, and 10 d PSI (females) ($P \leq 0.05$). The F line also had higher IgM titers at 14 d PPI (females) and at 10 d PSI (males) ($P \leq 0.05$). For IgG titers, a line difference was evident in females at 4 and 10 d PSI ($P \leq 0.05$); the F line had higher titers than did the RBC2 line. For the antibody response to BA in males, the F line had lower total and IgM titers at 10 d PPI ($P \leq 0.05$) than did the RBC2 line. No significant line differences in response to the BA were found in total and IgM titers in female turkeys or in IgG titers in both sexes at any time. These results suggest that selection for fast growth rate of turkeys might have resulted in changes in humoral immunity to the SRBC antigen and BA.

(*Key words*: turkey, sheep red blood cell antigen, *Brucella abortus* antigen, growth selection, antibody response)

2000 Poultry Science 79:804–809

INTRODUCTION

Commercial turkey breeders should have a goal of improving genetic resistance to disease along with improvements in growth rate, conformation, and reproduction. Improving genetic resistance to disease would reduce the cost of vaccination and other disease prevention procedures and reduce the mortality and loss in performance during disease outbreaks.

Selection for increased BW has been shown to be genetically associated with a reduction in immunocompetence and disease resistance in chickens (Han and Smyth, 1972, 1973; Mauldin et al., 1978; Qureshi and Havenstein, 1994) and turkeys (Saif et al., 1984; Sacco et al., 1991; Nestor et al., 1996a,b; Bayyari et al., 1997). A line (F) of turkeys has been selected long-term for increased 16-wk BW; the line exhibits BW approximately 75% greater than that of a corresponding randombred control base population (RBC2) (Nestor, 1977b, 1984; Nestor et al., 1996a). The F line had higher mortality either during natural outbreaks of erysipelas and fowl cholera (Saif et al., 1984) or during experimental challenges with virulent Newcastle disease virus (Tsai et al., 1992) and *Pasteurellla multocida* (Sacco et al., 1991). The changes in resistance to Newcastle disease virus and *P. multocida* in the F line could not be explained by changes in frequency of MHC Class II haplotypes (Nestor et al., 1996c).

Many kinds of antigens have been used to monitor immune responsiveness in poultry. Nonpathogenic antigens include synthetic glutamic acid-alanine-tyrosine (Cheng and Lamont, 1988), BSA (Heller et al., 1992; Parmentier et al., 1994, 1997), SRBC (Siegel and Gross, 1980; Van der Zijpp and Nieuwland, 1986), and *Brucella abortus* (BA) (Munns and Lamont, 1991; Scott et al., 1994; Nelson et al., 1995; Dix and Taylor, 1996). The SRBC antigen

Received for publication April 26, 1999.
Accepted for publication March 6, 2000.

[1] Salaries and research support provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University.

[2] To whom correspondence should be addressed: Karl E. Nestor, Department of Animal Sciences, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH 44691; e-mail: Nestor.1@osu.edu.

Abbreviation Key: BA = *Brucella abortus* antigen; MER = mercaptoethanol-resistant; MES = mercaptoethanol-sensitive; PPI = post-primary immunization; PSI = post-secondary immunization; RBC = randombred control; TD = thymus-dependent; TI = thymus-independent.

804

1499.] ALONZO DE OJEDA—FIRST VOYAGE. 9

The Spaniards smiled at their simplicity in attaching an extraordinary value to such worthless trifles; while the savages, in all probability, were equally surprised at beholding the strangers so eager after gold, and pearls and precious stones, which to themselves were objects of indifference.

Their manner of treating the dead was similar to that observed among the natives of some of the islands. Having deposited the corpse in a cavern or sepulchre, they placed a jar of water and a few eatables at its head, and then abandoned it without moan or lamentation. In some parts of the coast, when a person was considered near his end his nearest relatives bore him to the woods, and laid him in a hammock suspended to the trees. They then danced round
ing left within hi
to sustain him fo
habitations. If h
he was received
cing; if he died
thing more was t

Their mode of
mention. In the l
the patient in a
which they oblig
round a great fi
when they put hi
treatment, adds
saw many cured.

1499.] ALONZO DE OJEDA—FIRST VOYAGE. 9

The Spaniards smiled at their simplicity in attaching an extraordinary value to such worthless trifles; while the savages, in all probability, were equally surprised at beholding the strangers so eager after gold, and pearls and precious stones, which to themselves were objects of indifference.

Their manner of treating the dead was similar to that observed among the natives of some of the islands. Having deposited the corpse in a cavern or sepulchre, they placed a jar of water and a few eatables at its head, and then abandoned it without moan or lamentation. In some parts of the coast, when a person was considered near his end his nearest relatives bore him to the woods, and laid him in a hammock suspended to the trees. They then danced round him until evening, when, having left within his reach sufficient meat and drink to sustain him for four days, they repaired to their habitations. If he recovered and returned home, he was received with much ceremony and rejoicing; if he died of his malady or of famine, nothing more was thought of him.

Their mode of treating a fever is also worthy of mention. In the height of the malady they plunged the patient in a bath of the coldest water, after which they obliged him to make many evolutions round a great fire, until he was in a violent heat, when they put him to bed, that he might sleep: a treatment, adds Amerigo Vespucci, by which we saw many cured.

Figure 2-6: Projection profiles of a document (left) and the effect of skew (right) [19]

Background analysis techniques are sometimes combined with foreground analysis algorithms, thus called hybrid methods. Pavlidis and Zhou group background column gaps into column separators after horizontal smearing of foreground pixels [26]. Antonacopoulos and Ritchings also perform smearing first and then detect streams of white tiles whose sides encircle printed regions [27]. Smith detects the tab-stops and uses them to deduce the column layout of the page [28]. Chen et al. incorporate foreground and background information in order to filter whitespace rectangles progressively so that remaining rectangles form column separators [29]. The performance of a number of hybrid methods has been evaluated in analysis of historical newspapers [30] and complex layouts [31]. Although quite high, it strongly depends on the a priori knowledge of a few document characteristics (fonts, gaps sizes etc.).

Local analysis is actually a two-step process: a set of local features is selected first and then either clustering is applied or a classifier is trained and performed. Jain and Bhattacharjee apply a small number of Gabor filters to the grayscale image in order to capture the texture of different regions [32]. Tang et al. divide the image into small non overlapping regions and classify them according to their fractal signature [33]. Sauvola and Pietikäinen start from dividing the image into small square windows and then classify them according to statistical measures (black/white pixels ratio, average length of black runs etc.) [34]. Williams and Alder use a quadratic neural network to classify features obtained from local textual characteristics as well [35]. Etemad et al. select feature vectors based on a multi-scale wavelet packet representation of the image and implement a neural network for local classification [36]. Strouthopoulos and Papamarkos use a statistical reduction procedure to select a set of local features that are further reduced by a Principal Components Analyzer (PCA) and are finally classified by a Kohonen Self Organized

Feature Map (SOFM) based neural network [37]. Acharyya and Kundu use M-band wavelets to extract features that give measures of local energies at different scales and then classify the resulting multiscale feature vectors with an unsupervised clustering algorithm [38]. Kumar et al. use globally matched wavelet filters and multiple two-class Fisher classifiers [39] (Fig. 2.8). Maji and Roy perform M-band wavelet packet analysis followed by rough-fuzzy-possibilistic c-means clustering [40]. Chen et al. apply convolutional autoencoders to learn features directly from pixel intensity values and then use these features to train a support vector machine (SVM) [41]. Local analysis techniques are very promising. They often however rely on the proper selection of feature characteristics, filters and other parameters as well as the availability of a representative training set.

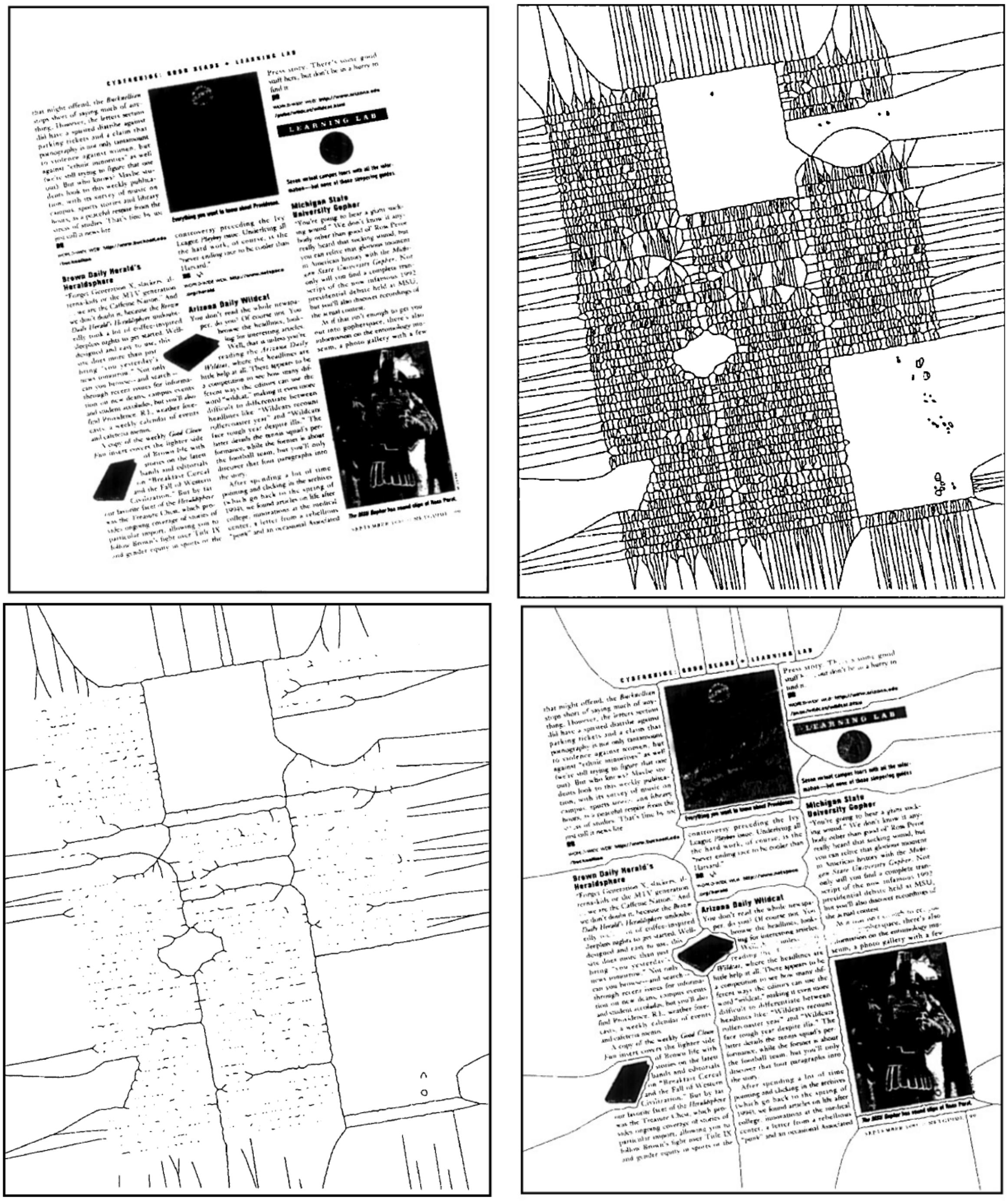


Figure 2-7: An example of background analysis [25]

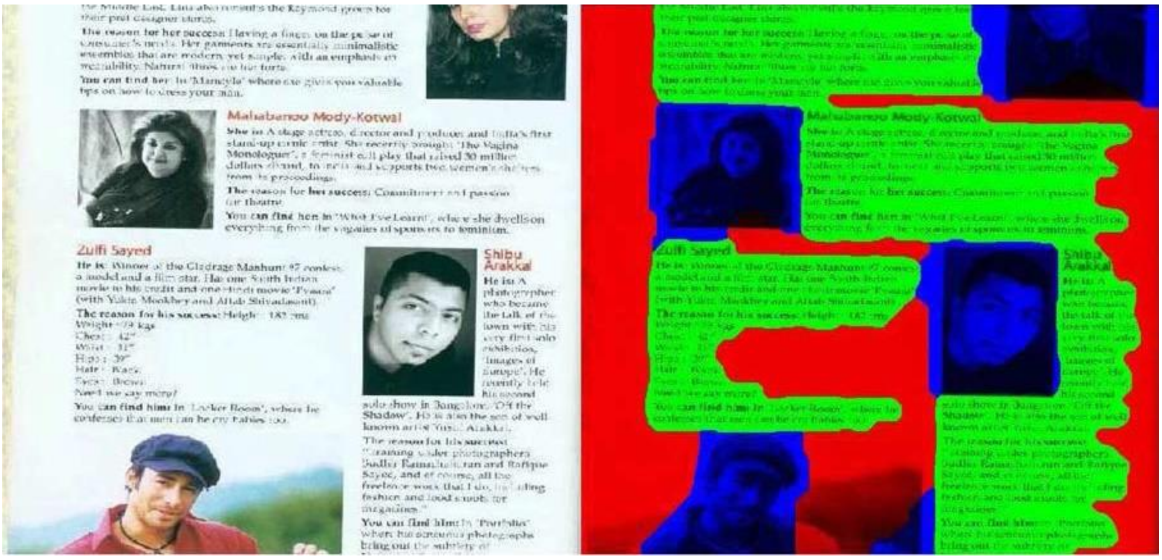


Figure 2-8: Local (texture-based) analysis example of a document image [39]

Table 2.1 contains a summary of the description as well as the restrictions of the above methods.

Method	Description	Constraints
Projection profiles	Recursively split the document at the valleys along the horizontal and vertical projection profiles	Manhattan layouts only Skew dependent
Smearing Connected components	Start from pixel level and merge regions together into larger components to form document structures (e.g. characters, then words, text lines, paragraphs and so on)	A priori knowledge of document characteristics (font sizes, spacings etc.) Inter-block gaps wider than interline gaps
Whitespace processing	Regions of white pixels split the page into blocks which are subsequently identified and subdivided further	
Local analysis	A set of local features is selected first and then either clustering is applied or a classifier is trained and performed	Proper selection of features and other parameters Representative training set

Table 2-1: Summary of the page segmentation techniques

2.3 The proposed system

In this chapter a layout-independent hybrid method, for complex layout analysis (e.g. newspapers, magazines etc.) is proposed. No prior knowledge of the document is required. Morphological operations are applied to both the foreground and the background, in order to connect neighboring components and separate lines/columns. Contour is simultaneously used for the extraction and classification of images and text blocks.

The proposed system is presented in Figure 2.9, while the detailed description of the tasks follows.

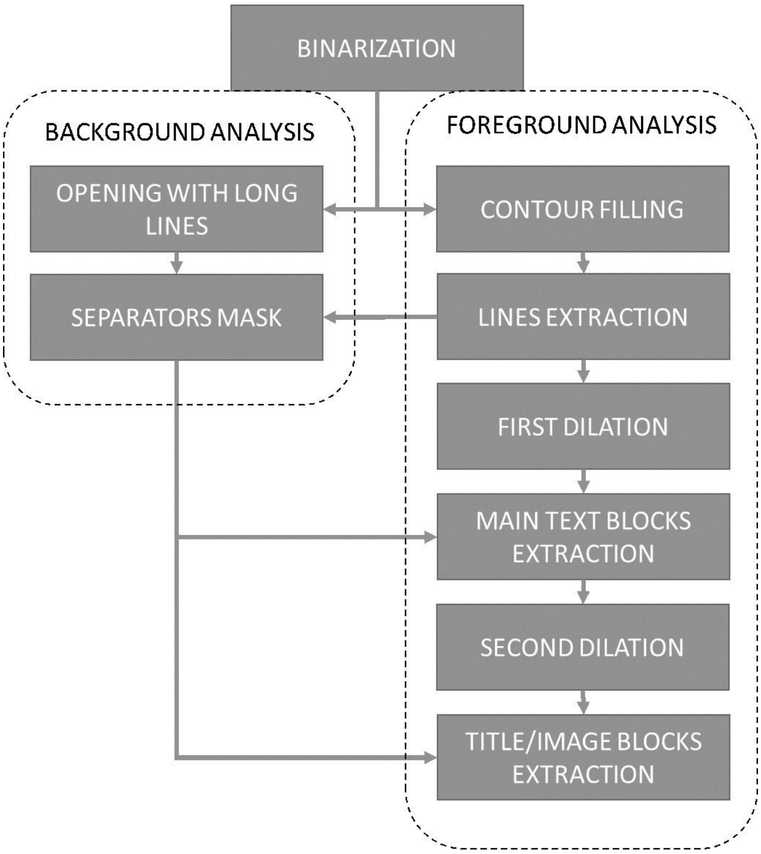


Figure 2-9: The proposed layout analysis system

2.3.1 Binarization

The method is applied on binary images. Therefore, the image is first transformed to grayscale and then binarized, so that the background is white and the foreground (text, images etc.) pixels are black. The better the discrimination between the background and the foreground, better are the results.

In a page that includes text, the background occupies the majority of the pixels. Thus, the background intensity range is used as the threshold value (maximum value). First, the gray scale histogram of the image is calculated and the highest (maximum) value is located, as the background. Then a binary image is created, where all the pixels between the proximal (on the left and the right of the peak) histogram local valleys are white and the rest of the pixels are changed to black. Then, a classical border-following algorithm [42] is applied to the image and the external contours of all the connected components are detected. The binarization procedure is repeated for all the frames (large and orthogonal contours) in the image.

This way, the method can be applied to pages with any background and foreground colors, since in the document images the majority of pixels belong to background. Moreover, if various backgrounds are used in the same collection or image, it can also be considered.

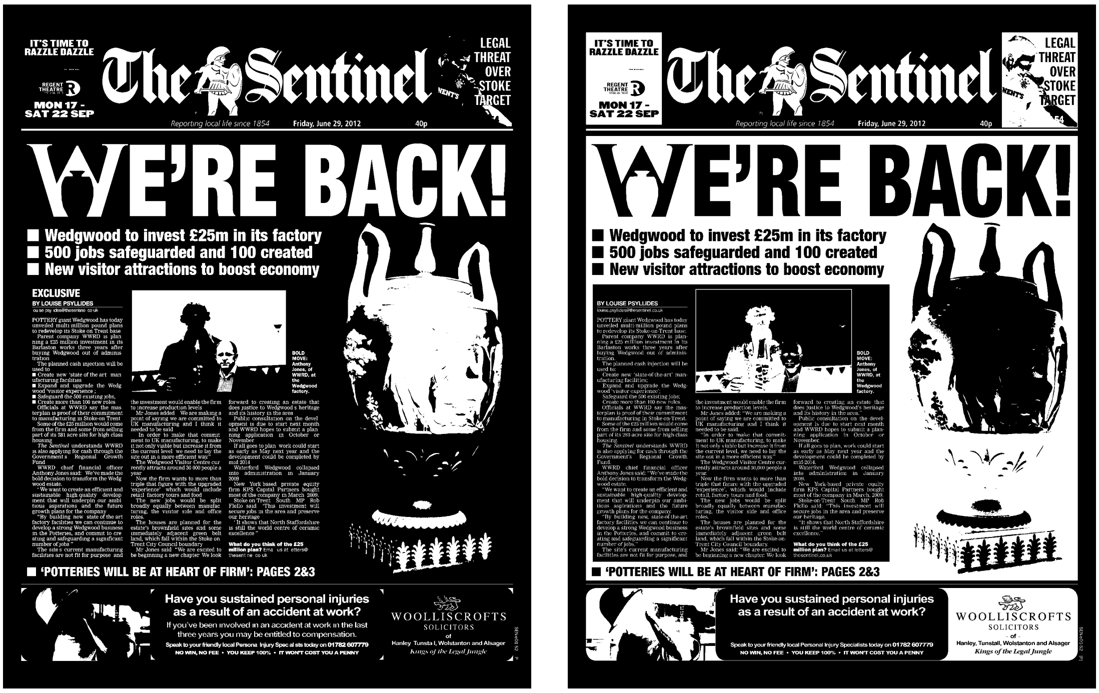

IT'S TIME TO RAZZLE DAZZLE
The Sentinel
LEGAL THREAT OVER STOKE TARGET
Reporting local life since 1854
Friday, June 29, 2012
WE'RE BACK!
Wedgwood to invest £25m in its factory
500 jobs safeguarded and 100 created
New visitor attractions to boost economy
'POTTERIES WILL BE AT HEART OF FIRM': PAGES 2&3
Have you sustained personal injuries as a result of an accident at work?
WOOLLISCROFTS SOLICITORS

Figure 2-10: Binarization by the regional band (left) and Otsu's (right) thresholding

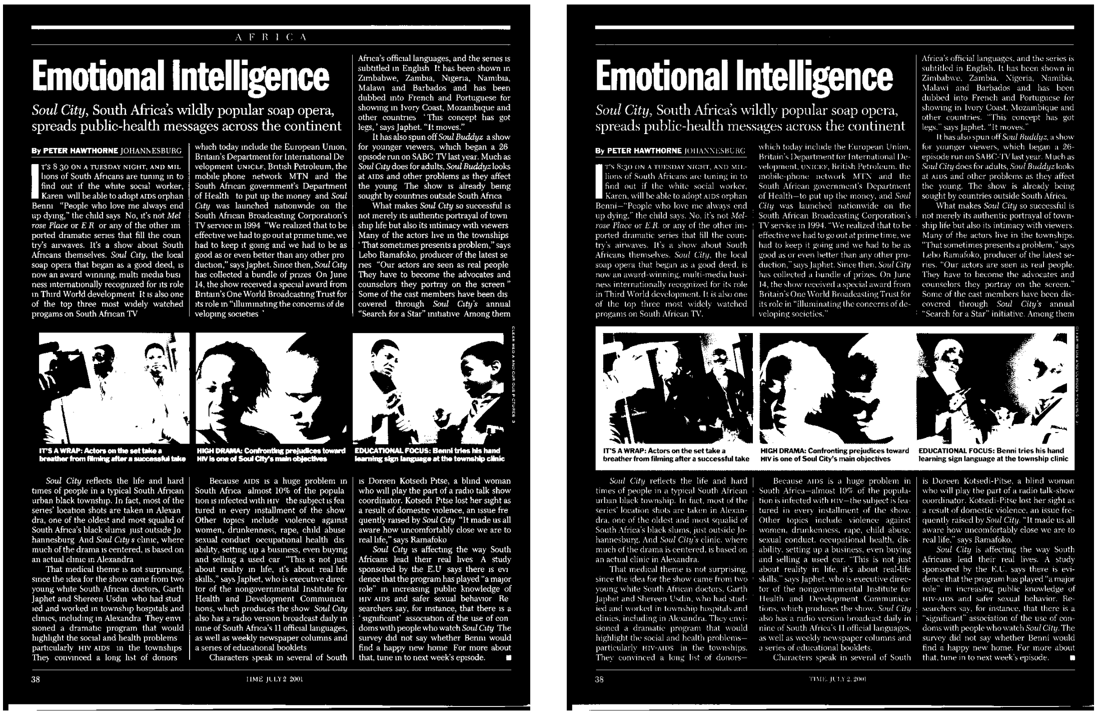

AFRICA
Emotional Intelligence
Soul City, South Africa's wildly popular soap opera, spreads public-health messages across the continent
By PETER HAWTHORNE JOHANNESBURG
IT'S A WRAP: Actors on the set take a breather from filming after a successful take
HIGH DRAMA: Confronting prejudices toward HIV is one of Soul City's main objectives
EDUCATIONAL FOCUS: Benni tries his hand learning sign language at the township clinic
38
TIME JULY 2 2001

Figure 2-11: Another example of the regional band (left) and Otsu's (right) thresholding

This is an advantage of the technique in use for the binarization. In Figure 2.10, the image of Figure 2.2 is presented binarized by the regional band thresholding [1] (left) and a typical binarization technique, Otsu's [43] (right).

Moreover, in another example (Fig. 2.11), from ICDAR2015 competition [31], it is obvious that the discrimination between neighbor images is not always possible in the case of Otsu's algorithm. In both Figures the binarized image has been inverted.

2.3.2 Background analysis

First, the background is processed in order to localize long white columns and rows. These straight white areas are classified as separators and can be used during the foreground processing to split overlapped components and improve the page segmentation results.

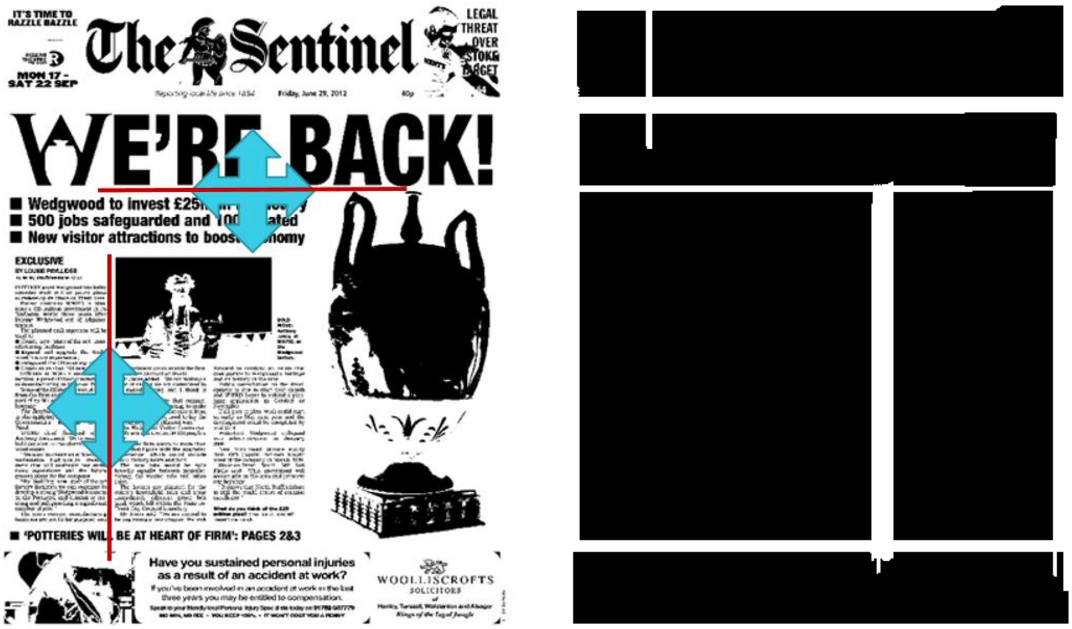

IT'S TIME TO RAZZLE DAZZLE
The Sentinel
LEGAL THREAT OVER STOKE TARGET
MON 17 - SAT 22 SEP
WE'RE BACK!
Wedgwood to invest £25m
500 jobs safeguarded and 100 created
New visitor attractions to boost economy
EXCLUSIVE
'POTTERIES WILL BE AT HEART OF FIRM': PAGES 2&3
Have you sustained personal injuries as a result of an accident at work?
WOOLLISCROFTS SOLICITORS

Figure 2-12: The image of Figure 2.2 morphologically opened with long lines

The binarized image is morphologically opened with two structuring elements: a horizontal line and a vertical line. The length of the first one is equal to half the image width while the length of the second one is equal to a quarter of the image height. A new image mask is created, containing all the column and row separators of the document (Fig. 2.12).

2.3.3 Foreground analysis

The foreground analysis is applied to the inverted image, so that foreground pixels are white while background is now black (Figs. 2.10-11). The border-following algorithm [42] is applied to the image and the external contours of all the connected components are detected. The contours are used to determine the size of the main body of small text characters (x-height).

According to Kavallieratou et al.: *By mean width of character, we consider the width of characters such as a, b, c, d etc., excluding the characters i, l, j, m, w that are either too narrow (i, j, l), or too wide (m, w). ... Although the character width differs between characters and writers, a rough estimation of the mean width could be made by accepting that excluding the ascenders and descenders the characters with mean width (as defined above), present width equal to their height*. [44]

That is the distance between the baseline and the mean-line of the lower-case letters (Fig. 2.13). The minimum size between width and height of most lower-case letters is equal to the main body.

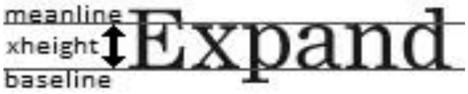


Figure 2-13: Definition of x-height

A novel technique is used here: since most of the connected components in the document images of printed text are lower-case letters, the main body can be easily calculated as follows: a) all the contours are classified according to the minimum value between their bounding rectangle width and height and b) the value of the most numerous class is considered as the x-height value.

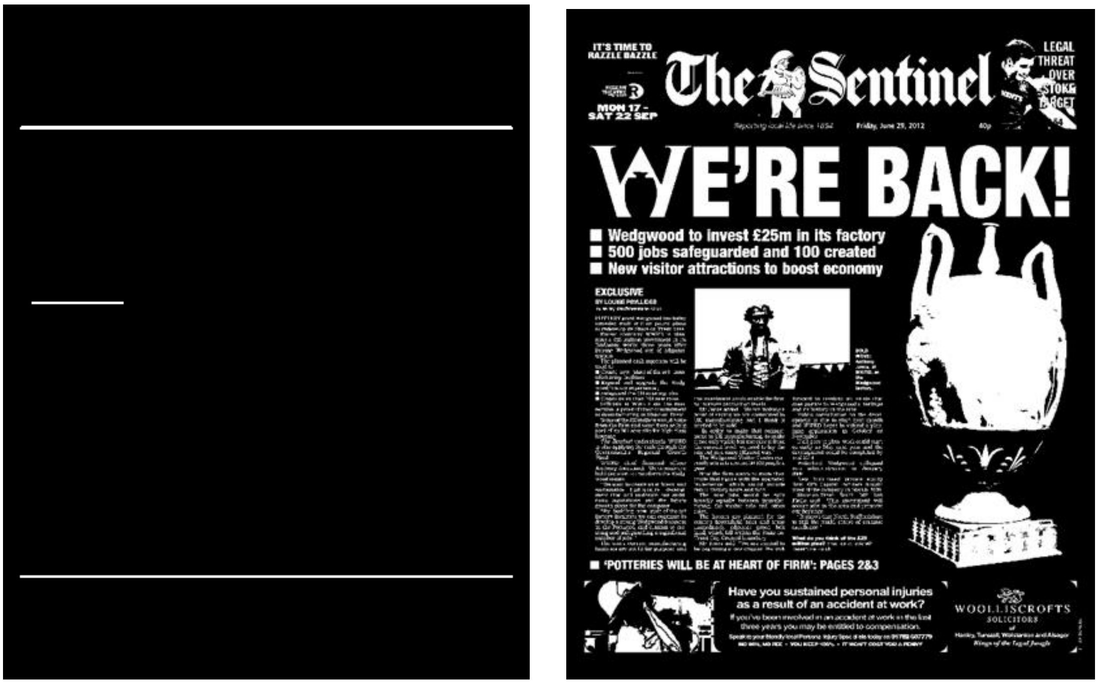

IT'S TIME TO RAZZLE DAZZLE
MON 17 - SAT 22 SEP
The Sentinel
LEGAL THREAT OVER STOKE TARGET
WE'RE BACK!
Wedgwood to invest £25m in its factory
500 jobs safeguarded and 100 created
New visitor attractions to boost economy
EXCLUSIVE
'POTTERIES WILL BE AT HEART OF FIRM': PAGES 2&3
Have you sustained personal injuries as a result of an accident at work?
WOOLLISCROFTS SOLICITORS

Figure 2-14: Separators detected by the contours (left) are removed from the image (right)

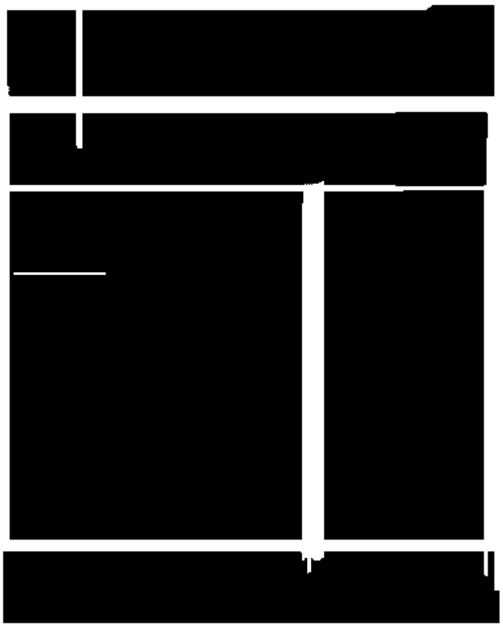


Figure 2-15: Separators added to the mask of Figure 2.12

At this point, the extraction of foreground lines follows. If their area is less than four times their perimeter, they are classified as separators. It is not required an exact threshold. This one proved to work well in most experiments. Then, they are removed from the foreground (Fig. 2.14) and are added to the separator mask (Fig. 2.15). In order for the separator mask to be complete, the outlines of the frames that were detected and thresholded during the binarization procedure are also included. The final mask (Fig. 2.16) will be used to separate the different areas of text in the next steps of the procedure.

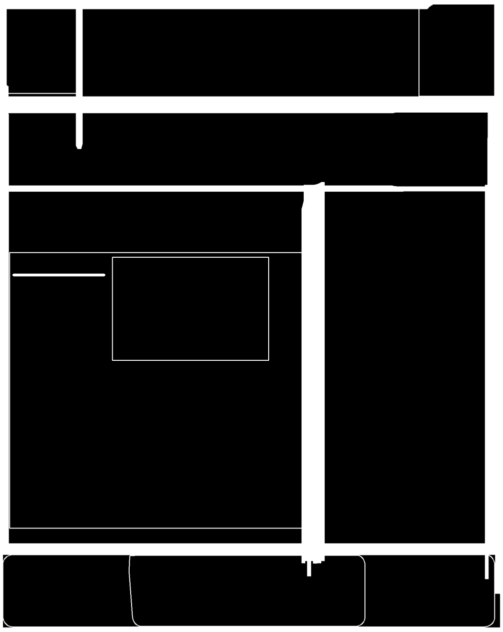


Figure 2-16: Final separator mask for the image of Figure 2.2

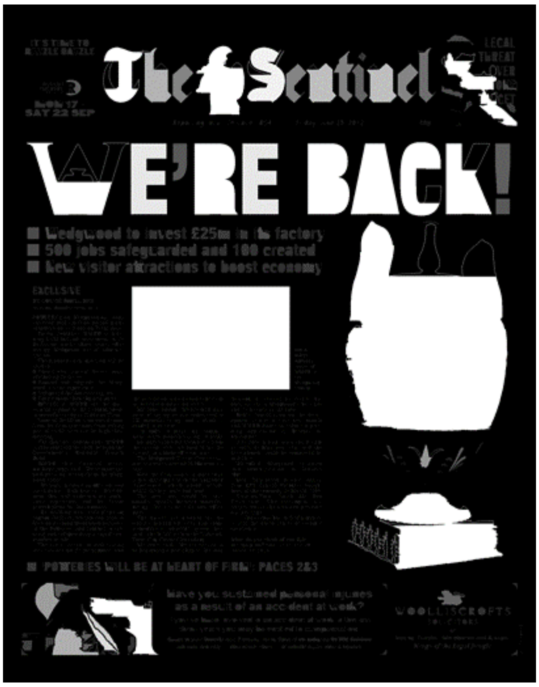

The Sentinel
WE'RE BACK!
Wedgwood to invest £25m in the factory
500 jobs safeguarded and 100 created
New visitor attractions to boost economy
POTTERIES WILL BE AT HEART OF FILM - PAGES 2&3

Figure 2-17: The components are filled with the minimum of their width-height value

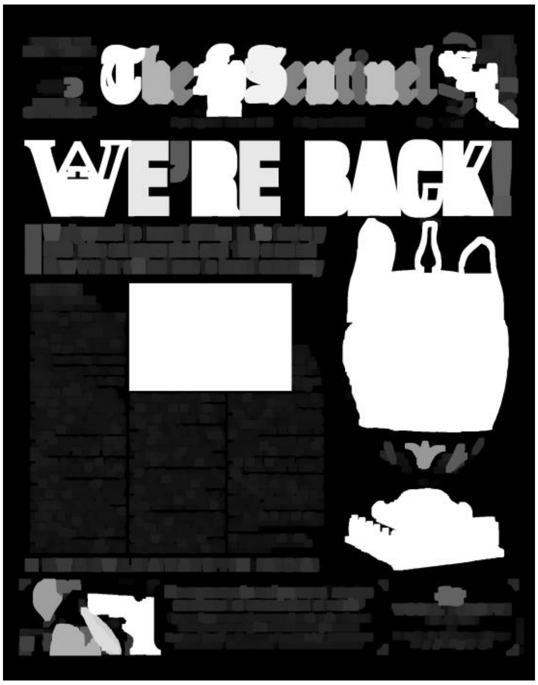


Figure 2-18: The image is dilated by an element of size one third of the x-height value

The contours of the connected components are filled before proceeding to the next step, in order to better distinguish between the small text and the large text regions. The minimum size of the bounding box will be the value of the filling color. The bigger the text, the higher this value is. Contours of more than 255 pixels bounding box minimum size are filled with the maximum available filling color that is 255. The result is a grayscale image, showing small text in dark gray and bigger text and images in lighter color (Fig. 2.17).

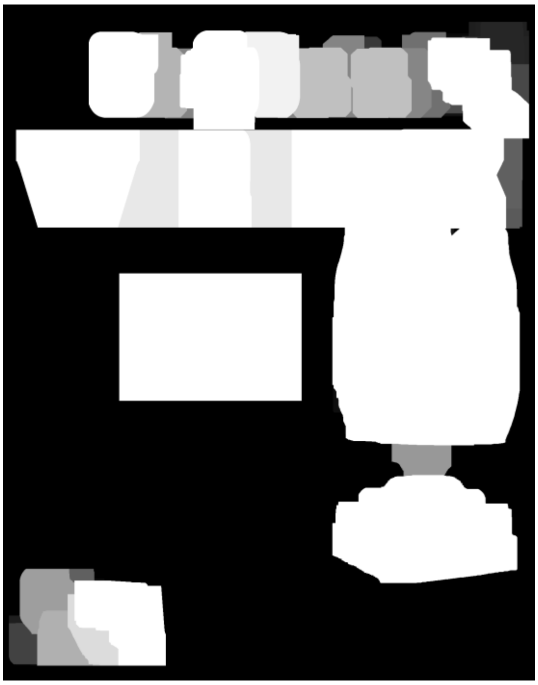


Figure 2-19: The bright regions are dilated and the larger text and image blocks are formed

Next is the extraction of main text blocks. The grayscale image containing the filled contours is dilated by a square structuring element (Fig. 2.18). The size of the structuring element is equal to one third of the main body height, selected by experimental trial. This way the small letters are connected and form text blocks, while the larger letters are not fully connected yet. Smaller structuring elements result in partially connected text blocks, while larger ones may result in overlapping blocks.

The border following algorithm [42] is applied again and the external contours of the connected elements are detected. This time the contours are classified as dark, containing small text, and lighter ones, containing large text or images. For each contour, the pixel values of the included image region are calculated. If the region mostly contains pixels with values more than two times the main body height, it's classified as large text or image. All other regions are considered as main text blocks and are extracted and added to a new mask, the layout mask.

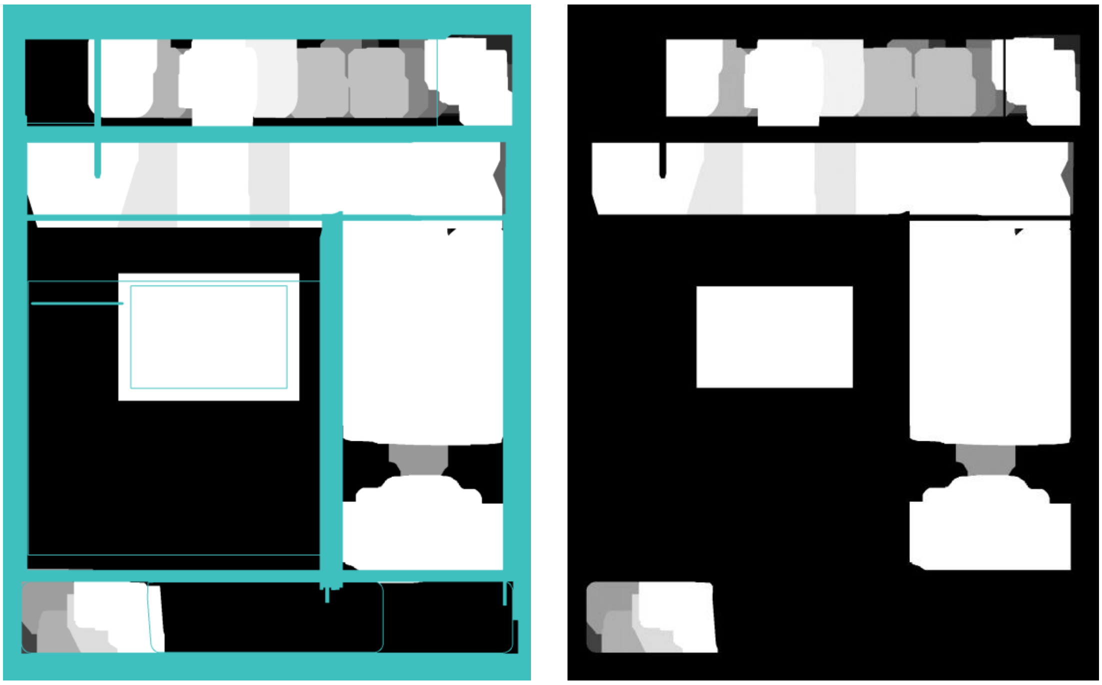


Figure 2-20. The dilated image of Figure 2.19 is ANDed with the invert separator mask

One more dilation is required, in order to cover the remaining space between the larger letters and image parts (Fig. 2.19). The size of the square structuring element is now equal to half the main body height, by experimental trial. This way large letters and image parts are also connected and form blocks.

Sometimes, the blocks may be very close to each other and dilation can merge regions from neighboring articles. For that reason, in order to split merged regions and improve segmentation results the logical AND is applied to the dilated image and the invert separator mask of Figure 2.16 (Fig. 2.20). Small parts are filtered out.

Finally, the blocks are extracted as previous and added to the layout mask as well. Very large letters and images are hard to distinguish since they are all mostly white. A shape rule has been applied in order to decide if a bright block is either title or image: long horizontal blocks are considered as titles while the rest of the bright blocks are considered as images. This simple rule proved to be effective in most cases.

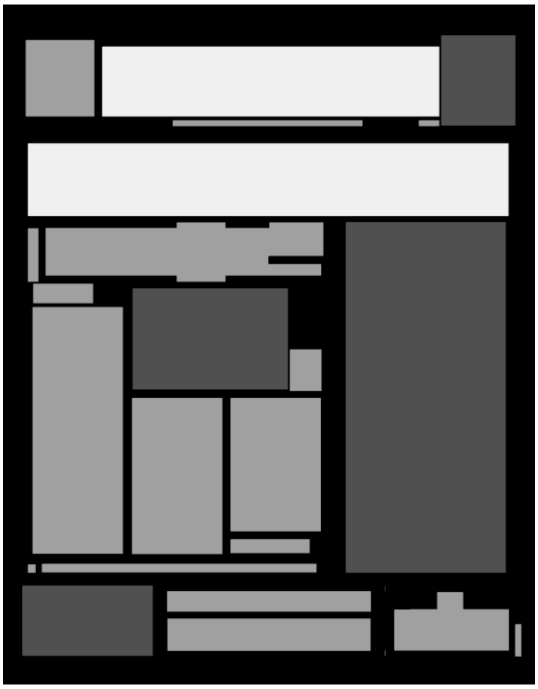


Figure 2-21: The final layout mask of image of Figure 2.2

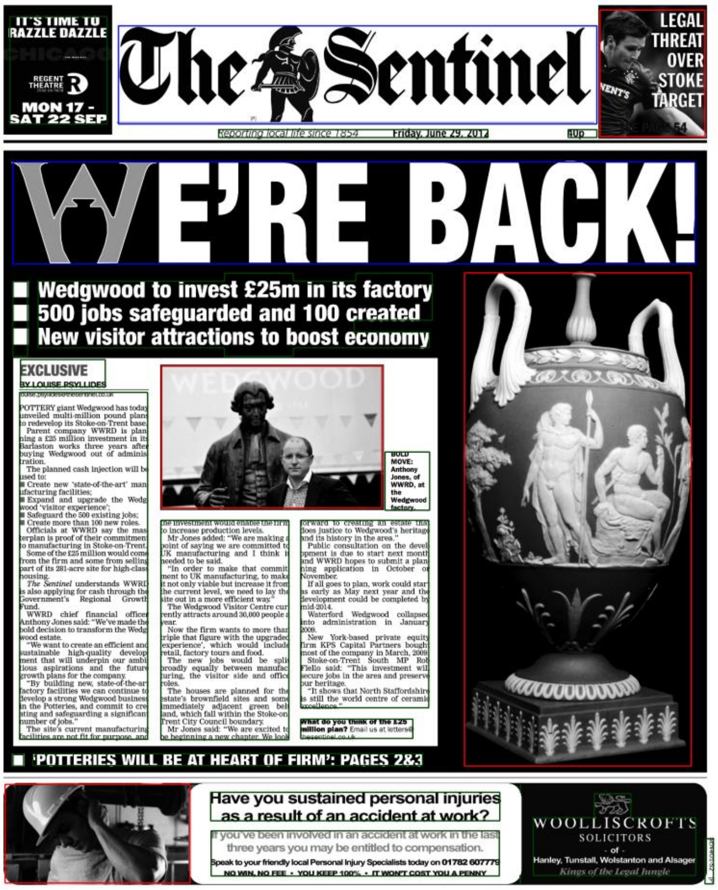

IT'S TIME TO RAZZLE DAZZLE
REGENT THEATRE
MON 17 - SAT 22 SEP
The Sentinel
LEGAL THREAT OVER STOKE TARGET
Reporting local life since 1854
Friday, June 29, 2012
WE'RE BACK!
Wedgwood to invest £25m in its factory
500 jobs safeguarded and 100 created
New visitor attractions to boost economy
EXCLUSIVE
BY LOUISE PSYLLIDES
'POTTERIES WILL BE AT HEART OF FIRM': PAGES 2&3
Have you sustained personal injuries as a result of an accident at work?
WOOLLISCROFTS SOLICITORS

Figure 2-22: The page segmentation result of image of Figure 2.2

All the blocks are classified by color to images (gray), text (light gray) and headers (white). The final layout for the image of Figure 2.2 is shown in Figure 2.21, while the classified results are shown in Figure 2.22 by the colors red (images), green (text) and blue (headers).

2.4 Experimental results

The algorithm has been coded in C# language. The morphological operations and the contour finding functions of the OpenCV library have been applied. About 2000 document images of high resolution newspaper scanned pages have been used for testing. In this case, the images were resized at 20% of their initial size before processing to reduce the computational cost. The results have shown that the method detects accurately more than 95% of the page components in less than half a second per page.

	Segmentation	OCR	Text only
The Fraunhofer Segmenter	84.1%	79.9%	85.4%
The ISPL method	83.1%	82.0%	91.0%
The MHS method	90.5%	88.3%	93.1%
The PAL method	83.2%	79.5%	87.7%
ABBYY FineReader® Engine 11	72.0%	69.3%	78.0%
Tesseract 3.03	74.2%	70.0%	76.8%
Proposed Technique	89.7%	84.3%	90.9%

Table 2-2: Comparative results for RDCL-2015 dataset

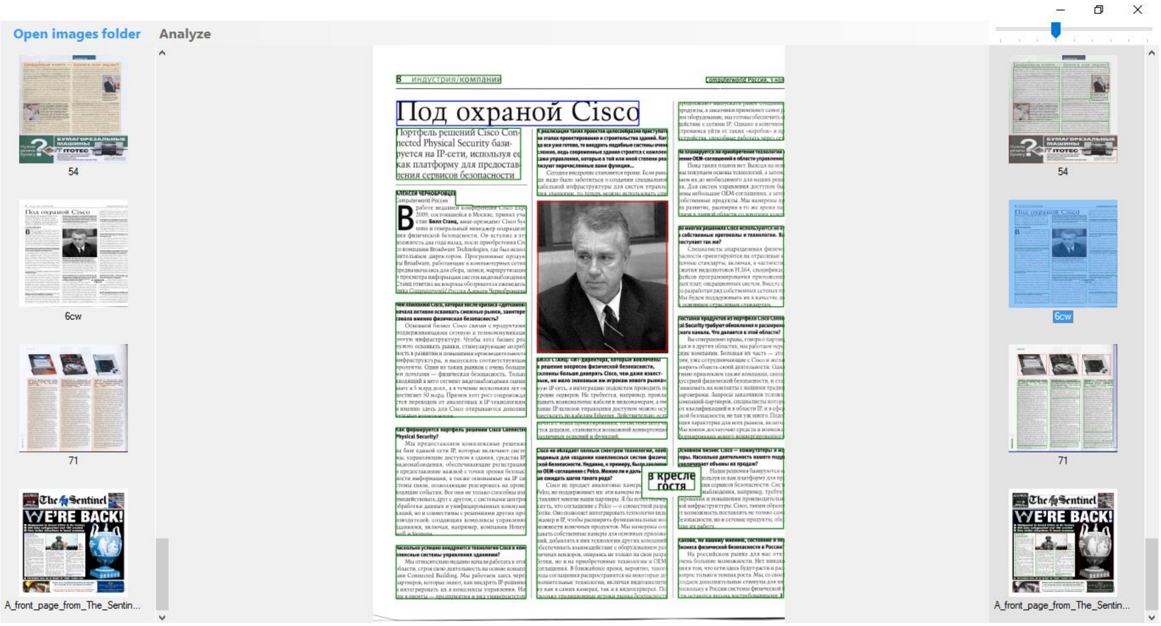
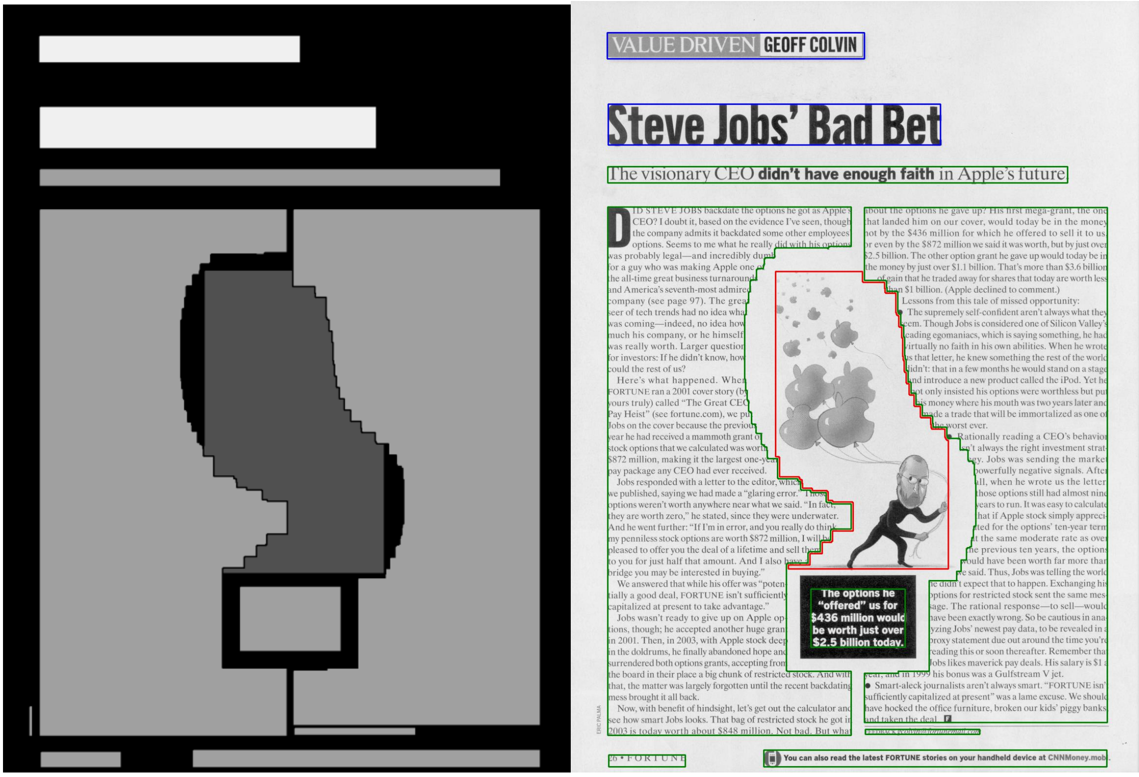
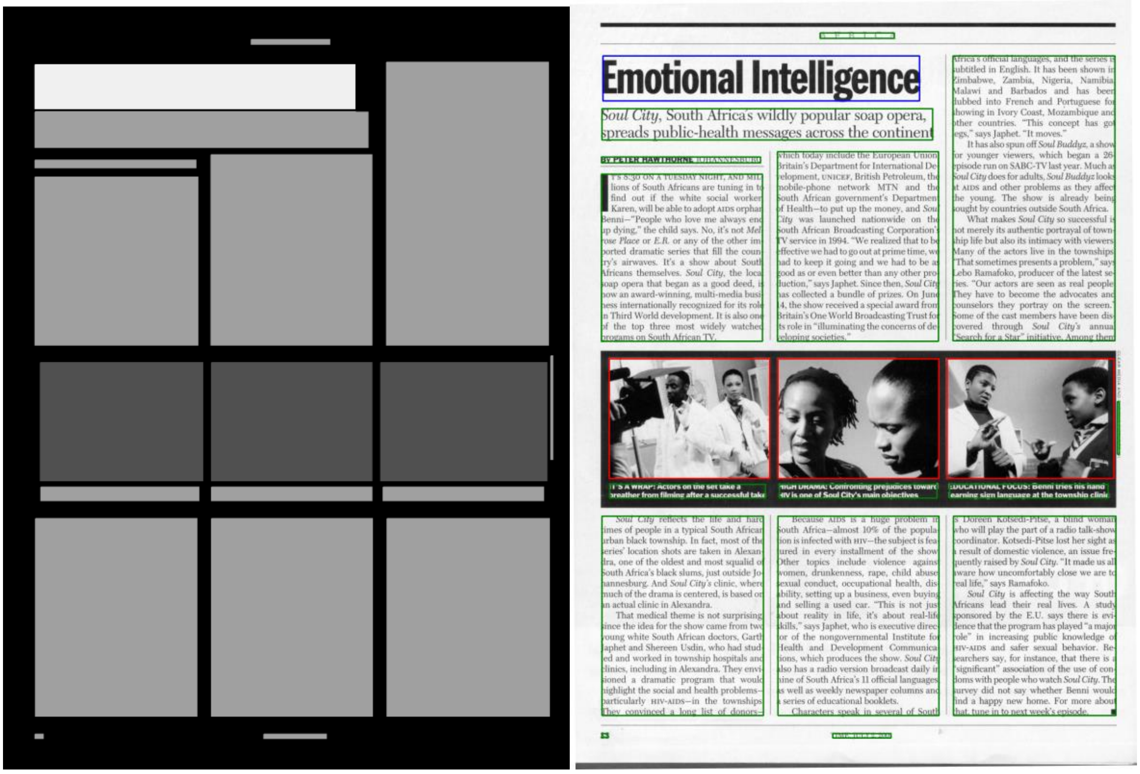
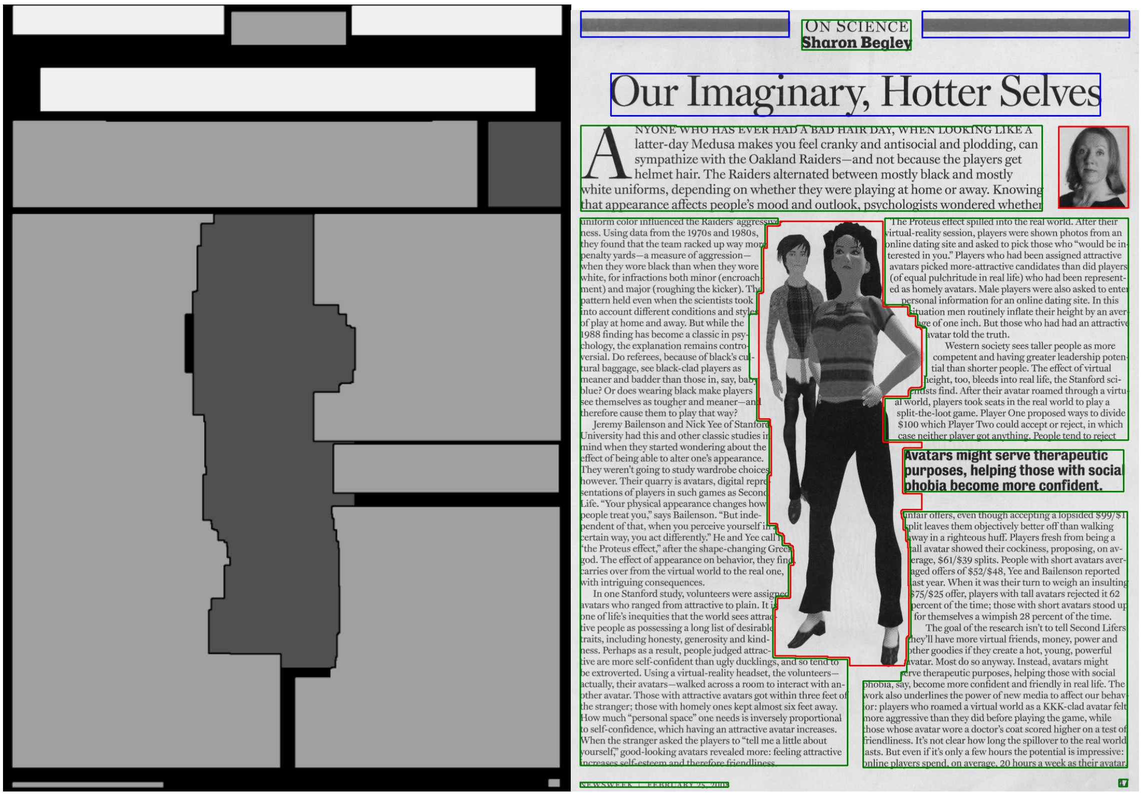

Open images folder
Analyze
54
6cw
71
A_front_page_from_The_Sentin...
Под охраной Cisco
54
6cw
71
A_front_page_from_The_Sentin...

Figure 2-23: A screenshot of the software


VALUE DRIVEN GEOFF COLVIN
Steve Jobs' Bad Bet
The visionary CEO didn't have enough faith in Apple's future
The options he "offered" us for $436 million would be worth just over $2.5 billion today.


Emotional Intelligence
Soul City, South Africa's wildly popular soap opera, spreads public-health messages across the continent



ON SCIENCE
Sharon Begley

Our Imaginary, Hotter Selves

ANYONE WHO HAS EVER HAD A BAD HAIR DAY, WHEN LOOKING LIKE A latter-day Medusa makes you feel cranky and antisocial and plodding, can sympathize with the Oakland Raiders—and not because the players get helmet hair. The Raiders alternated between mostly black and mostly white uniforms, depending on whether they were playing at home or away. Knowing that appearance affects people's mood and outlook, psychologists wondered whether uniform color influenced the Raiders' aggressiveness. Using data from the 1970s and 1980s, they found that the team racked up way more penalty yards—a measure of aggression—when they wore black than when they wore white, for infractions both minor (encroachment) and major (roughing the kicker). The pattern held even when the scientists took into account different conditions and styles of play at home and away. But while the 1988 finding has become a classic in psychology, the explanation remains controversial. Do referees, because of black's cultural baggage, see black-clad players as meaner and badder than those in, say, baby blue? Or does wearing black make players see themselves as tougher and meaner—and therefore cause them to play that way?

Jeremy Bailenson and Nick Yee of Stanford University had this and other classic studies in mind when they started wondering about the effect of being able to alter one's appearance. They weren't going to study wardrobe choices, however. Their quarry is avatars, digital representations of players in such games as Second Life. "Your physical appearance changes how people treat you," says Bailenson. "But independent of that, when you perceive yourself in a certain way, you act differently." He and Yee call it "the Proteus effect," after the shape-changing Greek god. The effect of appearance on behavior, they find, carries over from the virtual world to the real one, with intriguing consequences.

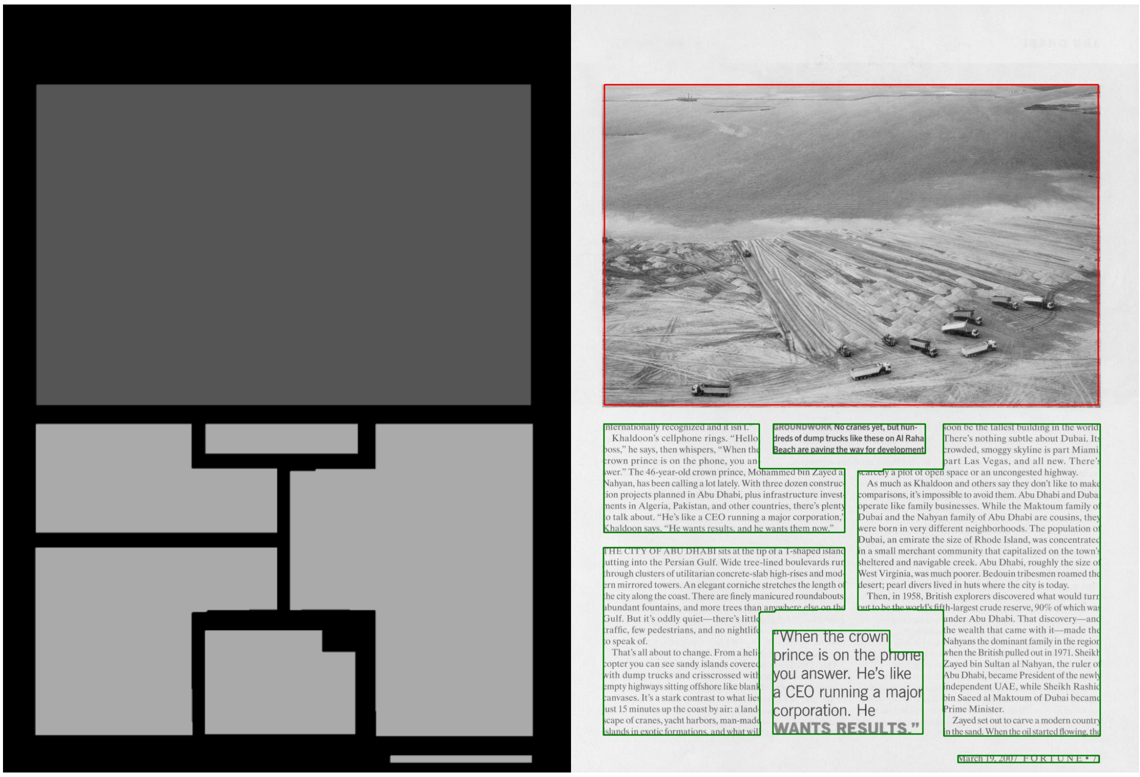
In one Stanford study, volunteers were assigned avatars who ranged from attractive to plain. It is one of life's inequities that the world sees attractive people as possessing a long list of desirable traits, including honesty, generosity and kindness. Perhaps as a result, people judged attractive are more self-confident than ugly ducklings, and so tend to be extroverted. Using a virtual-reality headset, the volunteers—actually, their avatars—walked across a room to interact with another avatar. Those with attractive avatars got within three feet of the stranger; those with homely ones kept almost six feet away. How much "personal space" one needs is inversely proportional to self-confidence, which having an attractive avatar increases. When the stranger asked the players to "tell me a little about yourself," good-looking avatars revealed more: feeling attractive increases self-esteem and therefore friendliness.

The Proteus effect spilled into the real world. After their virtual-reality session, players were shown photos from an online dating site and asked to pick those who "would be interested in you." Players who had been assigned attractive avatars picked more-attractive candidates than did players (of equal pulchritude in real life) who had been represented as homely avatars. Male players were also asked to enter personal information for an online dating site. In this situation men routinely inflate their height by an average of one inch. But those who had had an attractive avatar told the truth.

Western society sees taller people as more competent and having greater leadership potential than shorter people. The effect of virtual height, too, bleeds into real life, the Stanford scientists find. After their avatar roamed through a virtual world, players took seats in the real world to play a split-the-loot game. Player One proposed ways to divide $100 which Player Two could accept or reject, in which case neither player got anything. People tend to reject unfair offers, even though accepting a lopsided $99/$1 split leaves them objectively better off than walking away in a righteous huff. Players fresh from being a tall avatar showed their cockiness, proposing, on average, $61/$39 splits. People with short avatars averaged offers of $52/$48, Yee and Bailenson reported last year. When it was their turn to weigh an insulting $75/$25 offer, players with tall avatars rejected it 62 percent of the time; those with short avatars stood up for themselves a wimpish 28 percent of the time.

Avatars might serve therapeutic purposes, helping those with social phobia become more confident.

The goal of the research isn't to tell Second Lifers they'll have more virtual friends, money, power and other goodies if they create a hot, young, powerful avatar. Most do so anyway. Instead, avatars might serve therapeutic purposes, helping those with social phobia, say, become more confident and friendly in real life. The work also underlines the power of new media to affect our behavior: players who roamed a virtual world as a KKK-clad avatar felt more aggressive than they did before playing the game, while those whose avatar wore a doctor's coat scored higher on a test of friendliness. It's not clear how long the spillover to the real world lasts. But even if it's only a few hours the potential is impressive: online players spend, on average, 20 hours a week as their avatar.



GROUNDWORK No cranes yet, but hundreds of dump trucks like these on Al Raha Beach are paving the way for development

internationally recognized and it isn't."

Khaldoon's cellphone rings. "Hello boss," he says, then whispers, "When the crown prince is on the phone, you answer." The 46-year-old crown prince, Mohammed bin Zayed al Nahyan, has been calling a lot lately. With three dozen construction projects planned in Abu Dhabi, plus infrastructure investments in Algeria, Pakistan, and other countries, there's plenty to talk about. "He's like a CEO running a major corporation," Khaldoon says. "He wants results, and he wants them now."

THE CITY OF ABU DHABI sits at the tip of a T-shaped island jutting into the Persian Gulf. Wide tree-lined boulevards run through clusters of utilitarian concrete-slab high-rises and modern mirrored towers. An elegant corniche stretches the length of the city along the coast. There are finely manicured roundabouts, abundant fountains, and more trees than anywhere else on the Gulf. But it's oddly quiet—there's little traffic, few pedestrians, and no nightlife to speak of.

That's all about to change. From a helicopter you can see sandy islands covered with dump trucks and crisscrossed with empty highways sitting offshore like blank canvases. It's a stark contrast to what lies just 15 minutes up the coast by air: a landscape of cranes, yacht harbors, man-made islands in exotic formations, and what will soon be the tallest building in the world.

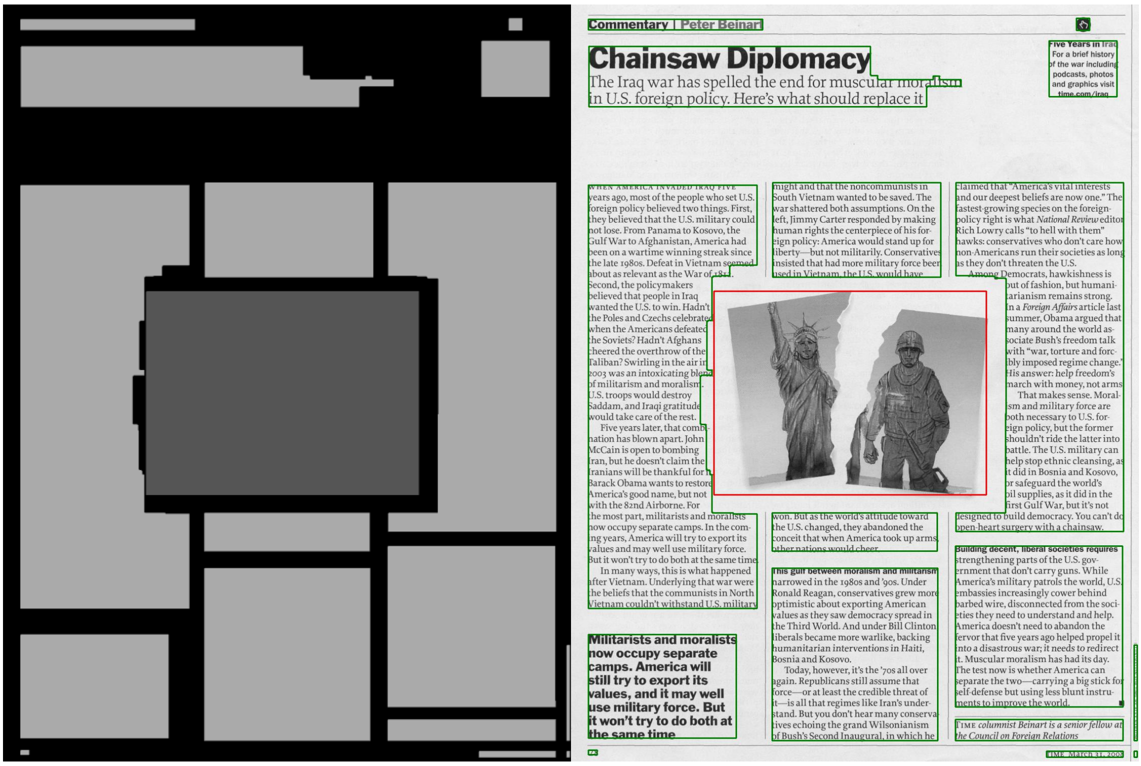
"When the crown prince is on the phone you answer. He's like a CEO running a major corporation. He **WANTS RESULTS.**"

There's nothing subtle about Dubai. Its crowded, smoggy skyline is part Miami, part Las Vegas, and all new. There's scarcely a plot of open space or an uncongested highway.

As much as Khaldoon and others say they don't like to make comparisons, it's impossible to avoid them. Abu Dhabi and Dubai operate like family businesses. While the Maktoum family of Dubai and the Nahyan family of Abu Dhabi are cousins, they were born in very different neighborhoods. The population of Dubai, an emirate the size of Rhode Island, was concentrated in a small merchant community that capitalized on the town's sheltered and navigable creek. Abu Dhabi, roughly the size of West Virginia, was much poorer. Bedouin tribesmen roamed the desert; pearl divers lived in huts where the city is today.

Then, in 1958, British explorers discovered what would turn out to be the world's fifth-largest crude reserve, 90% of which was under Abu Dhabi. That discovery—and the wealth that came with it—made the Nahyans the dominant family in the region when the British pulled out in 1971. Sheikh Zayed bin Sultan al Nahyan, the ruler of Abu Dhabi, became President of the newly independent UAE, while Sheikh Rashid bin Saeed al Maktoum of Dubai became Prime Minister.

Zayed set out to carve a modern country in the sand. When the oil started flowing, the



Commentary | Peter Beinart

Chainsaw Diplomacy

The Iraq war has spelled the end for muscular moralism in U.S. foreign policy. Here's what should replace it

Five Years in Iraq
For a brief history of the war including podcasts, photos and graphics visit time.com/iraq

WHEN AMERICA INVADED IRAQ FIVE years ago, most of the people who set U.S. foreign policy believed two things. First, they believed that the U.S. military could not lose. From Panama to Kosovo, the Gulf War to Afghanistan, America had been on a wartime winning streak since the late 1980s. Defeat in Vietnam seemed about as relevant as the War of 1812. Second, the policymakers believed that people in Iraq wanted the U.S. to win. Hadn't the Poles and Czechs celebrate when the Americans defeated the Soviets? Hadn't Afghans cheered the overthrow of the Taliban? Swirling in the air in 2003 was an intoxicating blend of militarism and moralism. U.S. troops would destroy Saddam, and Iraqi gratitude would take care of the rest.

Five years later, that combination has blown apart. John McCain is open to bombing Iran, but he doesn't claim the Iranians will be thankful for Barack Obama wants to restore America's good name, but not with the 82nd Airborne. For the most part, militarists and moralists now occupy separate camps. In the coming years, America will try to export its values and may well use military force. But it won't try to do both at the same time.

In many ways, this is what happened after Vietnam. Underlying that war were the beliefs that the communists in North Vietnam couldn't withstand U.S. military might and that the noncommunists in South Vietnam wanted to be saved. The war shattered both assumptions. On the left, Jimmy Carter responded by making human rights the centerpiece of his foreign policy: America would stand up for liberty—but not militarily. Conservatives insisted that had more military force been used in Vietnam, the U.S. would have

won. But as the world's attitude toward the U.S. changed, they abandoned the conceit that when America took up arms, other nations would cheer.

This gulf between moralism and militarism narrowed in the 1980s and '90s. Under Ronald Reagan, conservatives grew more optimistic about exporting American values as they saw democracy spread in the Third World. And under Bill Clinton liberals became more warlike, backing humanitarian interventions in Haiti, Bosnia and Kosovo.

Today, however, it's the '70s all over again. Republicans still assume that force—or at least the credible threat of it—is all that regimes like Iran's understand. But you don't hear many conservatives echoing the grand Wilsonianism of Bush's Second Inaugural, in which he claimed that "America's vital interests and our deepest beliefs are now one." The fastest-growing species on the foreign-policy right is what *National Review* editor Rich Lowry calls "to hell with them" hawks: conservatives who don't care how non-Americans run their societies as long as they don't threaten the U.S.

Among Democrats, hawkishness is out of fashion, but humanitarianism remains strong. In a *Foreign Affairs* article last summer, Obama argued that many around the world associate Bush's freedom talk with "war, torture and forcibly imposed regime change." His answer: help freedom's march with money, not arms.

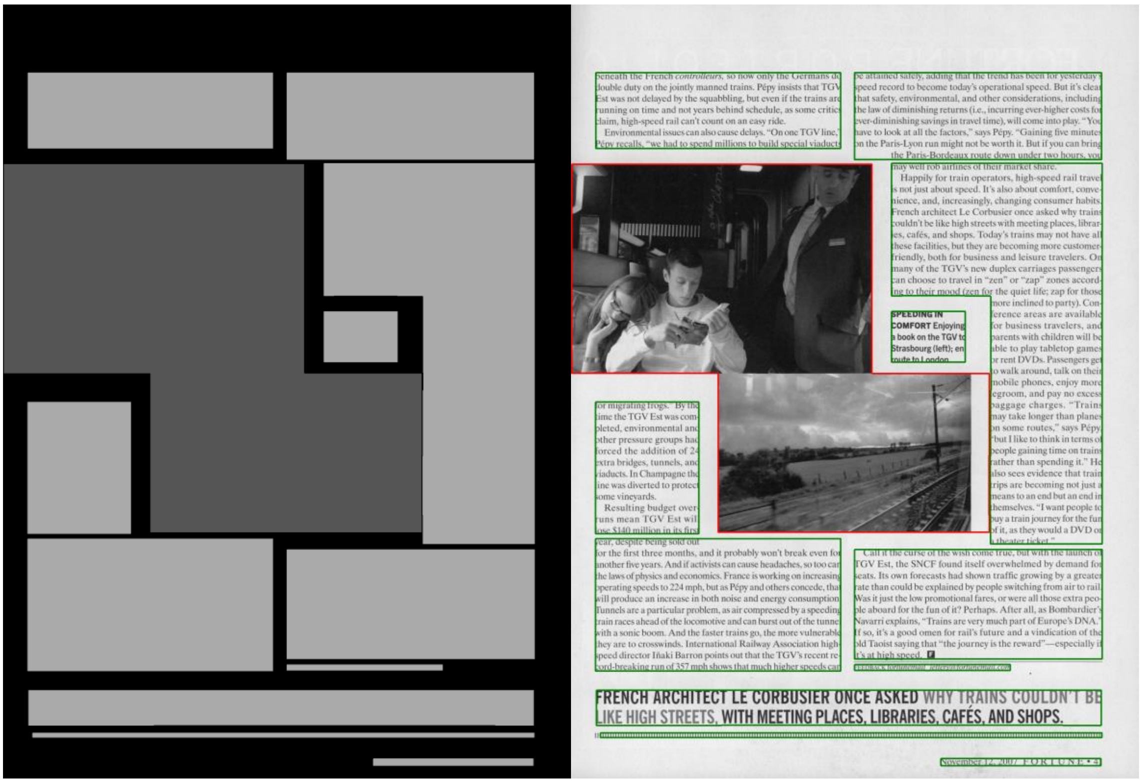
That makes sense. Moralism and military force are both necessary to U.S. foreign policy, but the former shouldn't ride the latter into battle. The U.S. military can help stop ethnic cleansing, as it did in Bosnia and Kosovo, or safeguard the world's oil supplies, as it did in the first Gulf War, but it's not designed to build democracy. You can't do open-heart surgery with a chainsaw.

Building decent, liberal societies requires strengthening parts of the U.S. government that don't carry guns. While America's military patrols the world, U.S. embassies increasingly cower behind barbed wire, disconnected from the societies they need to understand and help. America doesn't need to abandon the fervor that five years ago helped propel it into a disastrous war; it needs to redirect it. Muscular moralism has had its day. The test now is whether America can separate the two—carrying a big stick for self-defense but using less blunt instruments to improve the world. ■

Militarists and moralists now occupy separate camps. America will still try to export its values, and it may well use military force. But it won't try to do both at the same time

TIME *columnist Beinart is a senior fellow at the Council on Foreign Relations*

TIME March 31, 2008



beneath the French *controlleurs*, so now only the Germans do double duty on the jointly manned trains. Pépy insists that TGV Est was not delayed by the squabbling, but even if the trains are running on time and not years behind schedule, as some critics claim, high-speed rail can't count on an easy ride.

Environmental issues can also cause delays. "On one TGV line," Pépy recalls, "we had to spend millions to build special viaducts for migrating frogs." By the time the TGV Est was completed, environmental and other pressure groups had forced the addition of 24 extra bridges, tunnels, and viaducts. In Champagne the line was diverted to protect some vineyards.

Resulting budget overruns mean TGV Est will lose $140 million in its first year, despite being sold out for the first three months, and it probably won't break even for another five years. And if activists can cause headaches, so too can the laws of physics and economics. France is working on increasing operating speeds to 224 mph, but as Pépy and others concede, that will produce an increase in both noise and energy consumption. Tunnels are a particular problem, as air compressed by a speeding train races ahead of the locomotive and can burst out of the tunnel with a sonic boom. And the faster trains go, the more vulnerable they are to crosswinds. International Railway Association high-speed director Iñaki Barron points out that the TGV's recent record-breaking run of 357 mph shows that much higher speeds can be attained safely, adding that the trend has been for yesterday's speed record to become today's operational speed. But it's clear that safety, environmental, and other considerations, including the law of diminishing returns (i.e., incurring ever-higher costs for ever-diminishing savings in travel time), will come into play. "You have to look at all the factors," says Pépy. "Gaining five minutes on the Paris-Lyon run might not be worth it. But if you can bring the Paris-Bordeaux route down under two hours, you may well rob airlines of their market share."

Happily for train operators, high-speed rail travel is not just about speed. It's also about comfort, convenience, and, increasingly, changing consumer habits. French architect Le Corbusier once asked why trains couldn't be like high streets with meeting places, libraries, cafés, and shops. Today's trains may not have all these facilities, but they are becoming more customer-friendly, both for business and leisure travelers. On many of the TGV's new duplex carriages passengers can choose to travel in "zen" or "zap" zones according to their mood (zen for the quiet life; zap for those more inclined to party). Conference areas are available for business travelers, and parents with children will be able to play tabletop games or rent DVDs. Passengers get to walk around, talk on their mobile phones, enjoy more legroom, and pay no excess baggage charges. "Trains may take longer than planes on some routes," says Pépy, "but I like to think in terms of people gaining time on trains rather than spending it." He also sees evidence that train trips are becoming not just a means to an end but an end in themselves. "I want people to buy a train journey for the fun of it, as they would a DVD or a theater ticket."

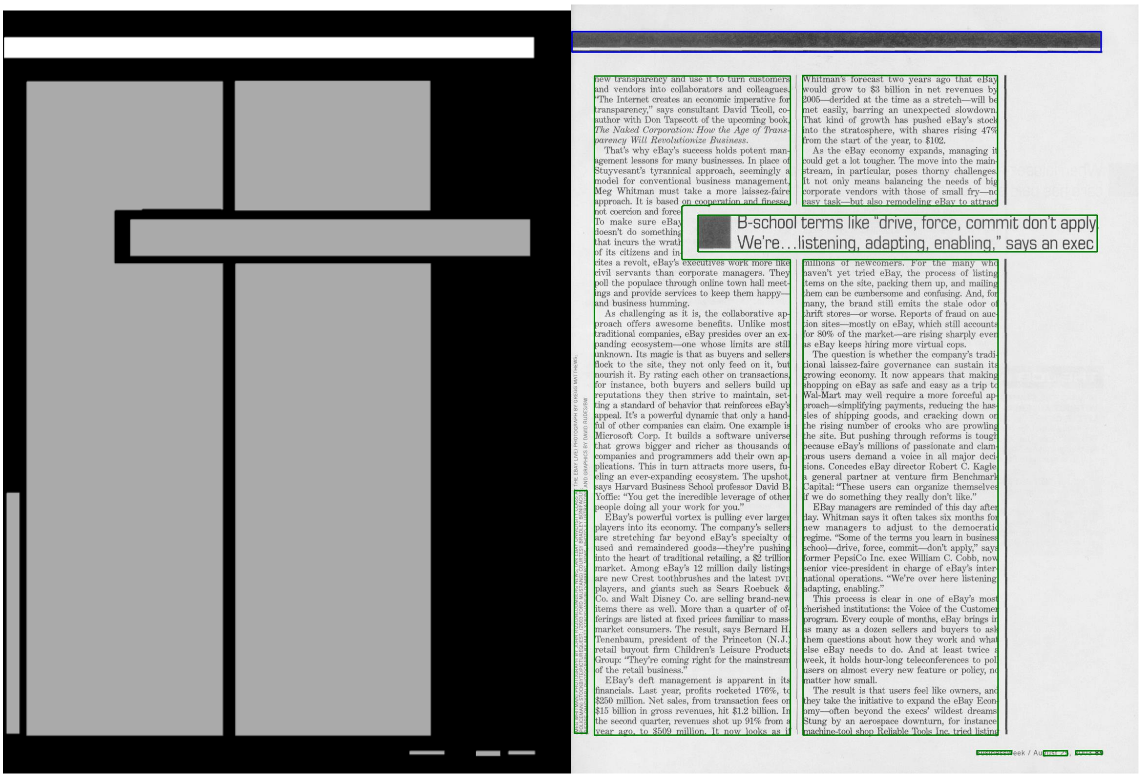
Call it the curse of the wish come true, but with the launch of TGV Est, the SNCF found itself overwhelmed by demand for seats. Its own forecasts had shown traffic growing by a greater rate than could be explained by people switching from air to rail. Was it just the low promotional fares, or were all those extra people aboard for the fun of it? Perhaps. After all, as Bombardier's Navarri explains, "Trains are very much part of Europe's DNA." If so, it's a good omen for rail's future and a vindication of the old Taoist saying that "the journey is the reward"—especially if it's at high speed. ■

FEEDBACK fortunemail_letters@fortunemail.com

SPEEDING IN COMFORT Enjoying a book on the TGV to Strasbourg (left); en route to London

FRENCH ARCHITECT LE CORBUSIER ONCE ASKED WHY TRAINS COULDN'T BE LIKE HIGH STREETS, WITH MEETING PLACES, LIBRARIES, CAFÉS, AND SHOPS.

November 12, 2007 FORTUNE • 4



new transparency and use it to turn customers and vendors into collaborators and colleagues. "The Internet creates an economic imperative for transparency," says consultant David Ticoll, coauthor with Don Tapscott of the upcoming book, *The Naked Corporation: How the Age of Transparency Will Revolutionize Business.*

That's why eBay's success holds potent management lessons for many businesses. In place of Stuyvesant's tyrannical approach, seemingly a model for conventional business management, Meg Whitman must take a more laissez-faire approach. It is based on cooperation and finesse, not coercion and force. To make sure eBay doesn't do something that incurs the wrath of its citizens and incites a revolt, eBay's executives work more like civil servants than corporate managers. They poll the populace through online town hall meetings and provide services to keep them happy—and business humming.

As challenging as it is, the collaborative approach offers awesome benefits. Unlike most traditional companies, eBay presides over an expanding ecosystem—one whose limits are still unknown. Its magic is that as buyers and sellers flock to the site, they not only feed on it, but nourish it. By rating each other on transactions, for instance, both buyers and sellers build up reputations they then strive to maintain, setting a standard of behavior that reinforces eBay's appeal. It's a powerful dynamic that only a handful of other companies can claim. One example is Microsoft Corp. It builds a software universe that grows bigger and richer as thousands of companies and programmers add their own applications. This in turn attracts more users, fueling an ever-expanding ecosystem. The upshot, says Harvard Business School professor David B. Yoffie: "You get the incredible leverage of other people doing all your work for you."

EBay's powerful vortex is pulling ever larger players into its economy. The company's sellers are stretching far beyond eBay's specialty of used and remaindered goods—they're pushing into the heart of traditional retailing, a $2 trillion market. Among eBay's 12 million daily listings are new Crest toothbrushes and the latest DVD players, and giants such as Sears Roebuck & Co. and Walt Disney Co. are selling brand-new items there as well. More than a quarter of offerings are listed at fixed prices familiar to mass-market consumers. The result, says Bernard H. Tenenbaum, president of the Princeton (N.J.) retail buyout firm Children's Leisure Products Group: "They're coming right for the mainstream of the retail business."

EBay's deft management is apparent in its financials. Last year, profits rocketed 176%, to $250 million. Net sales, from transaction fees on $15 billion in gross revenues, hit $1.2 billion. In the second quarter, revenues shot up 91% from a year ago, to $509 million. It now looks as if Whitman's forecast two years ago that eBay would grow to $3 billion in net revenues by 2005—derided at the time as a stretch—will be met easily, barring an unexpected slowdown. That kind of growth has pushed eBay's stock into the stratosphere, with shares rising 47% from the start of the year, to $102.

As the eBay economy expands, managing it could get a lot tougher. The move into the mainstream, in particular, poses thorny challenges. It not only means balancing the needs of big corporate vendors with those of small fry—no easy task—but also remodeling eBay to attract millions of newcomers. For the many who haven't yet tried eBay, the process of listing items on the site, packing them up, and mailing them can be cumbersome and confusing. And, for many, the brand still emits the stale odor of thrift stores—or worse. Reports of fraud on auction sites—mostly on eBay, which still accounts for 80% of the market—are rising sharply even as eBay keeps hiring more virtual cops.

The question is whether the company's traditional laissez-faire governance can sustain its growing economy. It now appears that making shopping on eBay as safe and easy as a trip to Wal-Mart may well require a more forceful approach—simplifying payments, reducing the hassles of shipping goods, and cracking down on the rising number of crooks who are prowling the site. But pushing through reforms is tough because eBay's millions of passionate and clamorous users demand a voice in all major decisions. Concedes eBay director Robert C. Kagle, a general partner at venture firm Benchmark Capital: "These users can organize themselves if we do something they really don't like."

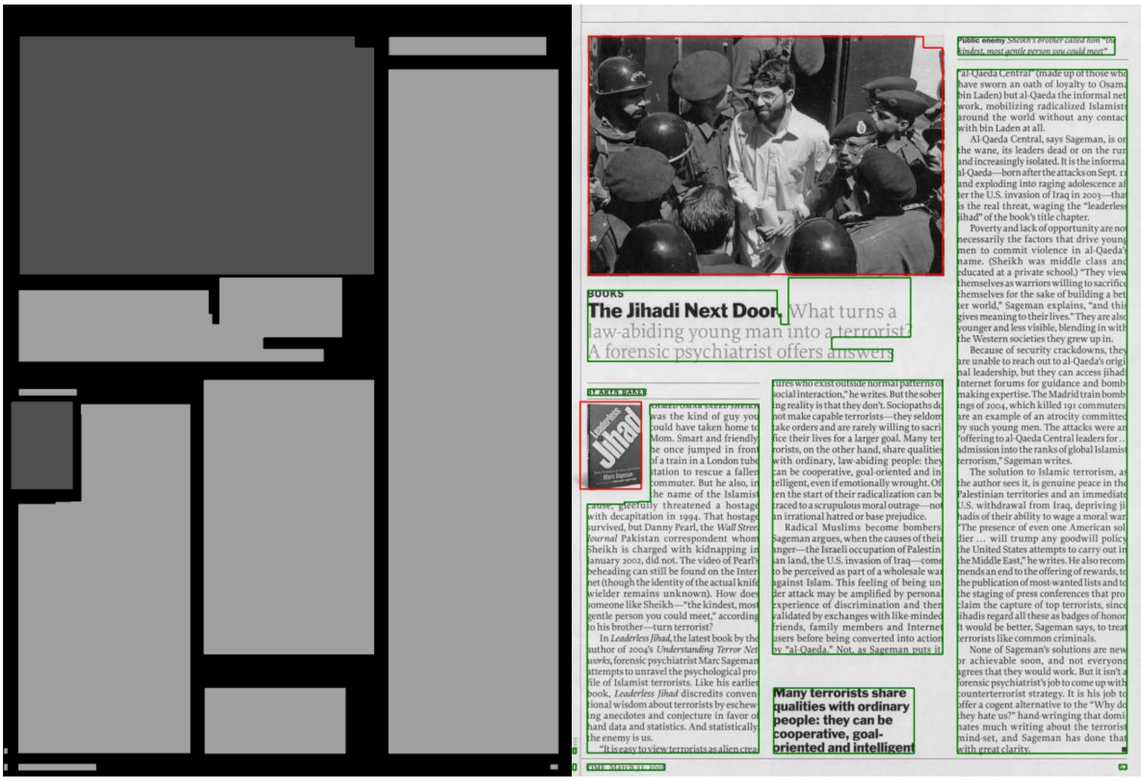
EBay managers are reminded of this day after day. Whitman says it often takes six months for new managers to adjust to the democratic regime. "Some of the terms you learn in business school—drive, force, commit—don't apply," says former PepsiCo Inc. exec William C. Cobb, now senior vice-president in charge of eBay's international operations. "We're over here listening, adapting, enabling."

This process is clear in one of eBay's most cherished institutions: the Voice of the Customer program. Every couple of months, eBay brings in as many as a dozen sellers and buyers to ask them questions about how they work and what else eBay needs to do. And at least twice a week, it holds hour-long teleconferences to poll users on almost every new feature or policy, no matter how small.

The result is that users feel like owners, and they take the initiative to expand the eBay Economy—often beyond the execs' wildest dreams. Stung by an aerospace downturn, for instance, machine-tool shop Reliable Tools Inc. tried listing

B-school terms like "drive, force, commit don't apply. We're…listening, adapting, enabling," says an exec

THE EBAY LIVE) PHOTOGRAPH BY GREGG MATTHEWS; AND GRAPHICS BY DAVID RUDES/BW



BOOKS

The Jihadi Next Door. What turns a law-abiding young man into a terrorist? A forensic psychiatrist offers answers

BY ARYN BAKER

AHMED OMAR SAEED SHEIKH was the kind of guy you could have taken home to Mom. Smart and friendly, he once jumped in front of a train in a London tube station to rescue a fallen commuter. But he also, in the name of the Islamist cause, gleefully threatened a hostage with decapitation in 1994. That hostage survived, but Danny Pearl, the *Wall Street Journal* Pakistan correspondent whom Sheikh is charged with kidnapping in January 2002, did not. The video of Pearl's beheading can still be found on the Internet (though the identity of the actual knife wielder remains unknown). How does someone like Sheikh—"the kindest, most gentle person you could meet," according to his brother—turn terrorist?

In *Leaderless Jihad*, the latest book by the author of 2004's *Understanding Terror Networks*, forensic psychiatrist Marc Sageman attempts to unravel the psychological profile of Islamist terrorists. Like his earlier book, *Leaderless Jihad* discredits conventional wisdom about terrorists by eschewing anecdotes and conjecture in favor of hard data and statistics. And statistically the enemy is us.

"It is easy to view terrorists as alien creatures who exist outside normal patterns of social interaction," he writes. But the sobering reality is that they don't. Sociopaths do not make capable terrorists—they seldom take orders and are rarely willing to sacrifice their lives for a larger goal. Many terrorists, on the other hand, share qualities with ordinary, law-abiding people: they can be cooperative, goal-oriented and intelligent, even if emotionally wrought. Often the start of their radicalization can be traced to a scrupulous moral outrage—not an irrational hatred or base prejudice.

Radical Muslims become bombers, Sageman argues, when the causes of their anger—the Israeli occupation of Palestinian land, the U.S. invasion of Iraq—come to be perceived as part of a wholesale war against Islam. This feeling of being under attack may be amplified by personal experience of discrimination and then validated by exchanges with like-minded friends, family members and Internet users before being converted into action by "al-Qaeda." Not, as Sageman puts it, "al-Qaeda Central" (made up of those who have sworn an oath of loyalty to Osama bin Laden) but al-Qaeda the informal network, mobilizing radicalized Islamists around the world without any contact with bin Laden at all.

Al-Qaeda Central, says Sageman, is on the wane, its leaders dead or on the run and increasingly isolated. It is the informal al-Qaeda—born after the attacks on Sept. 11 and exploding into raging adolescence after the U.S. invasion of Iraq in 2003—that is the real threat, waging the "leaderless jihad" of the book's title chapter.

Poverty and lack of opportunity are not necessarily the factors that drive young men to commit violence in al-Qaeda's name. (Sheikh was middle class and educated at a private school.) "They view themselves as warriors willing to sacrifice themselves for the sake of building a better world," Sageman explains, "and this gives meaning to their lives." They are also younger and less visible, blending in with the Western societies they grew up in.

Because of security crackdowns, they are unable to reach out to al-Qaeda's original leadership, but they can access jihadi Internet forums for guidance and bomb-making expertise. The Madrid train bombings of 2004, which killed 191 commuters, are an example of an atrocity committed by such young men. The attacks were an "offering to al-Qaeda Central leaders for … admission into the ranks of global Islamist terrorism," Sageman writes.

The solution to Islamic terrorism, as the author sees it, is genuine peace in the Palestinian territories and an immediate U.S. withdrawal from Iraq, depriving jihadis of their ability to wage a moral war. "The presence of even one American soldier … will trump any goodwill policy the United States attempts to carry out in the Middle East," he writes. He also recommends an end to the offering of rewards, to the publication of most-wanted lists and to the staging of press conferences that proclaim the capture of top terrorists, since jihadis regard all these as badges of honor. It would be better, Sageman says, to treat terrorists like common criminals.

None of Sageman's solutions are new or achievable soon, and not everyone agrees that they would work. But it isn't a forensic psychiatrist's job to come up with counterterrorist strategy. It is his job to offer a cogent alternative to the "Why do they hate us?" hand-wringing that dominates much writing about the terrorist mind-set, and Sageman has done that with great clarity. ■

Many terrorists share qualities with ordinary people: they can be cooperative, goal-oriented and intelligent

Public enemy *Sheikh's brother called him "the kindest, most gentle person you could meet"*

TIME March 31, 2008

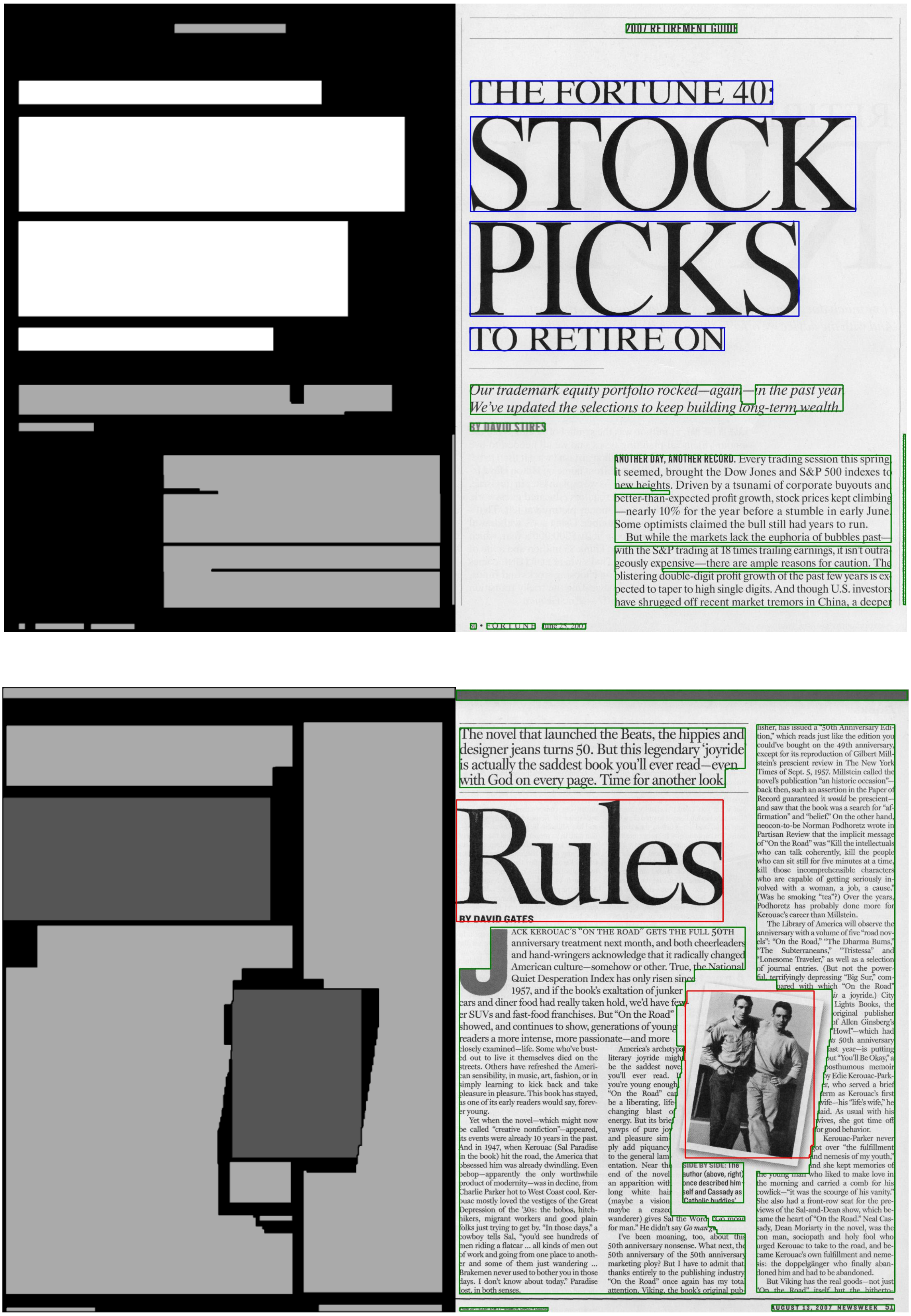


Figure 2-24: Results from the RDCL-2015 dataset of ICDAR 2015 competition

In order to give comparative results, the RDCL-2015 dataset of ICDAR 2015 competition [31] was also used. In this case the images were resized by 50% of their initial size, which increased the computational cost up to 3 seconds per page. Several results as well as a screenshot of the software tool are presented in Figures 2.24 and 2.23 respectively. Moreover, in Table 2.2 are given the results for the techniques and the measures reported in the competition paper [31] and the proposed technique. The mentioned results for the techniques are those presented in the paper, while the evaluation for the proposed technique has been done following the mentioned rules.

CHAPTER 3 - TEXT LOCALIZATION

3.1 Introduction

The text in an image or a video frame contains high-level semantic information about the content. Text in images is sometimes printed against textured backgrounds, while current OCR systems can only handle text that is printed against clean backgrounds. Therefore, text extraction from mixed content images is still considered a challenging problem. Among the difficulties are the unknown position and orientation of the characters, the varying fonts, sizes and colors, the uneven lighting, the irregular background and the embedding of text with photos, logos etc. Two kind of images can be found in documents: born digital images (logos and graphics) and scene images (photographs). Figure 3.1 shows text localization examples for both cases.

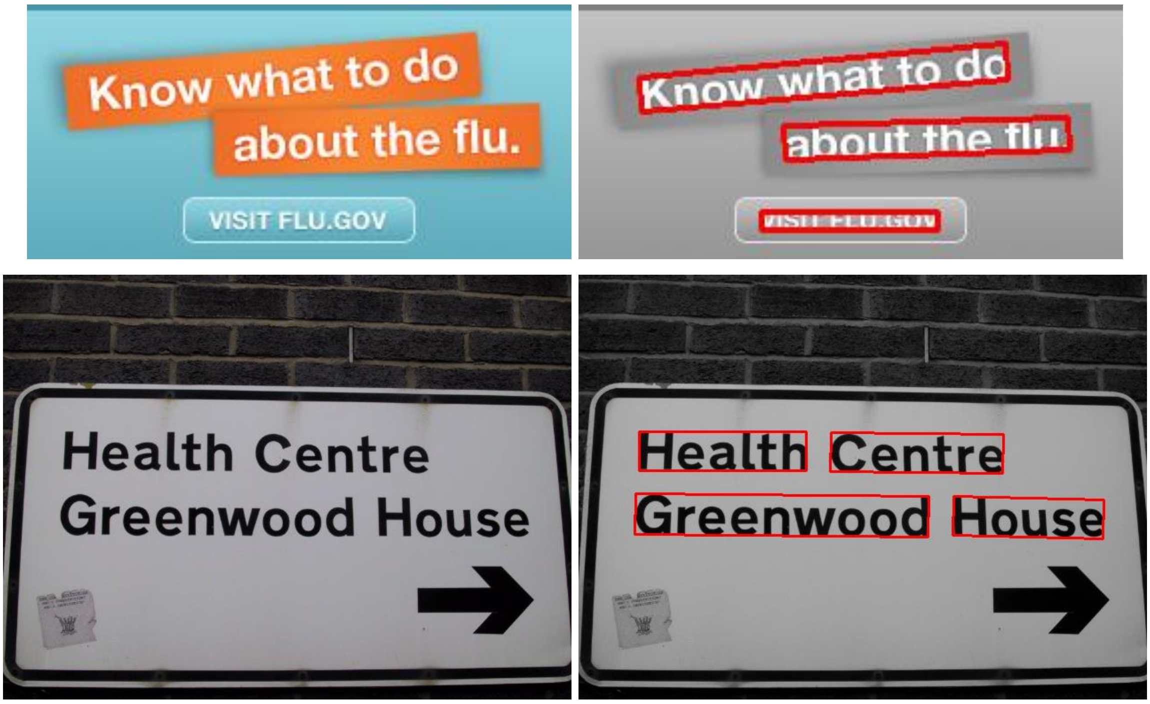


Figure 3-1: Text localization in a born digital (up) and a scene (down) image

The extraction of text information from images and video frames consists of the text localization, the text segmentation and the text recognition steps. In this chapter we focus on text localization. Text localization methods can be categorized into region-based and texture-based ones [45]. Region-based methods assume there is little or no variation of color within text and at the same time the contrast between the text and the surrounding background is high enough. Either image thresholding or edge detection is first applied in order to distinguish between the foreground and the background. Then, connected components are classified as text or non-text by using geometrical analysis and various heuristics. Non-text components are filtered out while text components are successively grouped into larger ones and form text regions. Texture-based methods assume that text and non-text regions do not have similar textural properties. They select a set of local features in order to train a classifier that is used to discriminate regions with different textures.

Region-based methods are widely used due to their simplicity. They perform better when applied on clean and high resolution images with single colored text. Some of them are applied on images that contain only horizontal or vertical text with font size in a limited range. On the other hand, texture-based methods seem to perform well with noisy and degraded images that may contain rotated text. They are however time consuming since they require an exhaustive scan of the input image and they usually involve a computationally expensive learning phase.

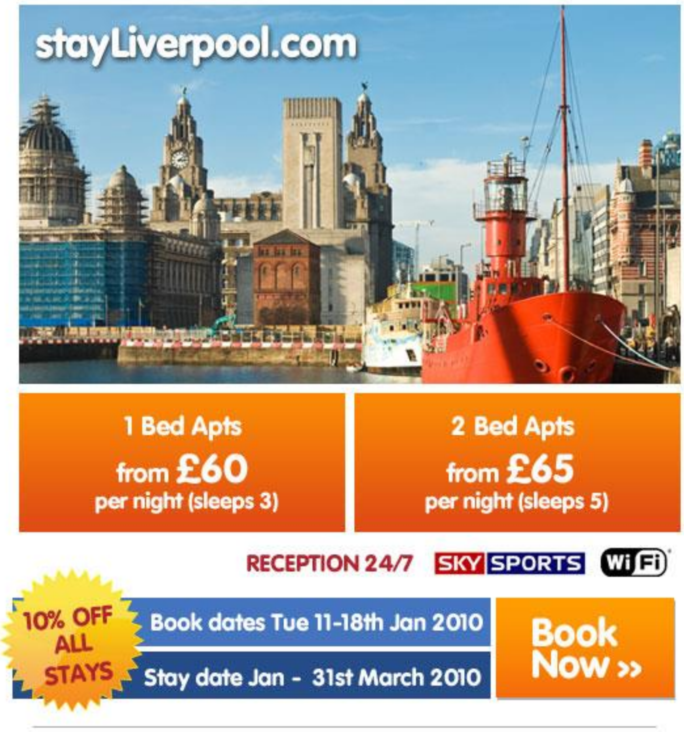


Figure 3-2: An image with text of various font size and orientation

Many authors have presented hybrid approaches to the text localization problem that combine region-based and texture-based techniques. In this chapter a fast and reliable hybrid method for text localization in complex images is proposed. No a priori knowledge of the image characteristics is required and there is no limitation on character font, size, orientation and color.

In the Section 3.2, a short summary of the state of the art is given. The proposed technique and the included modules are presented in detail, in Section 3.3, while experimental and comparative results are presented on the ICDAR 2011 Robust Reading Competition dataset, in Section 3.4.

3.2 State of the art

Most of the related works are based on connected component analysis. Ohya et al. use adaptive thresholding to binarize scene images and then detect characters by observing gray-level differences between adjacent regions [46]. Lee and Kankanhalli use edge information to generate connected components with the same gray level as well [47]. Kim calculates the color histogram of video frames and then uses the location of color peaks for segmentation [48]. Smith and Kanade extract the high contrast regions from video frames by combining edge detection and thresholding [49]. Lienhart and Stuber assume text is monochromatic, segment the image by applying a split and merge algorithm and filter out very large and very small objects [50]. An example of their method is shown in Figure 3.3: (a) original video frame; (b) image segmentation using split-and-merge algorithm; (c) after size restriction; (d) after binarization and dilation; (e) after motion analysis; and (f) after contrast analysis and aspect ratio restriction. Shim et al. merge pixels with similar gray levels into groups, then remove large groups and finally perform a contrast based region boundary analysis [51]. Jain and Yu perform color clustering to 24-bit images that are first bit-dropped to 6-bit images [52]. Messelodi and Modena classify connected components based on their geometrical features and their spatial relations [53]. Chen et al. detect edges using the Canny operator and then estimate the orientation and scale of each connected component using two groups of Gabor-type asymmetric filters [54]. Nikolaou and Papamarkos first reduce the image colors to a small number using a new color reduction technique, so as to have solid characters and uniform local backgrounds and then extract the character elements as connected components [55].

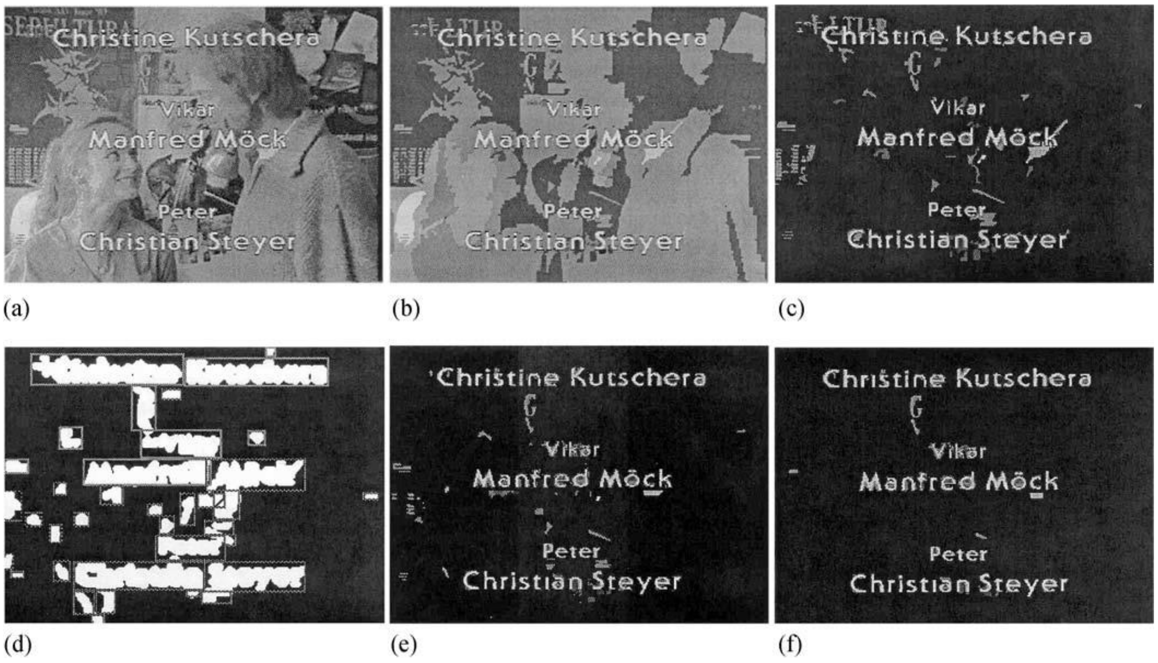


Figure 3-3: An example of Lienhart and Stuber region-based method [50]

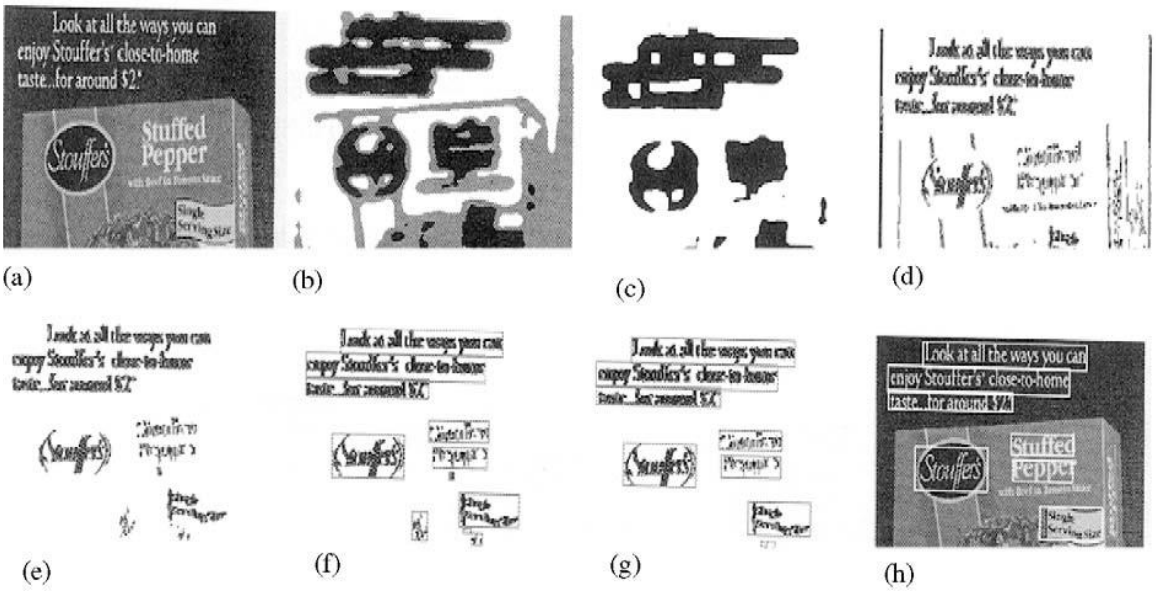

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Look at all the ways you can enjoy Stouffer's' close-to-home taste...for around $2.
Stouffer's
Stuffed Pepper
Single Serving Size

Figure 3-4: An example of Wu et al. texture-based method [56]

Texture-based techniques can also be found early in the literature. Wu et al. apply texture based segmentation and use height similarity, spacing and alignment of extracted strokes to detect text regions [56]. An example of their method is shown in Figure 3.4: (a) sub-image of input image; (b) clustering; (c) text region after morphological operation; (d) stroke generation; (e) stroke filtering; (f) stroke aggregation; (g) chip filtering and extension; (h) text localization. Li et al. use neural networks to classify wavelet features extracted from small windows in the image [57]. Jung also uses a neural network to distinguish between different textures [58]. Clark and Mirmehdi use low-level texture measures in combination with a neural network classifier [59]. Kim et al. apply a continuously adaptive mean shift algorithm (CAMSHIFT) to the texture analysis output of a support vector machine (SVM) [60]. Gllavata et al. apply a wavelet transform to the image and then extract horizontally aligned text areas using the k-means algorithm and based on the distribution of high-frequency wavelet coefficients [61]. Weinman et al. use wavelet features too and a semi-Markov model for scene text localization [62]. Jung et al. construct a stroke filter that can detect strokes of texts and use it to filter out false positives with strong edges [63]. Wang et al. also proposed a stroke-based text location scheme combined with a SVM [64]. Yin et al. use an AdaBoost classifier based on horizontal and vertical variances, stroke width, color and geometry features [65].

Method	Description	Constraints
Region-based	Either image thresholding or edge detection is first. Then, connected components are classified as text or non-text by using geometrical analysis and various heuristics. Non-text components are filtered out while text components are successively grouped into larger ones and form text regions.	Clean and high resolution images Little or no variation of color within text High contrast between the text and the surrounding background
Texture-based	A set of local features is selected first and then either clustering is applied or a classifier is trained and performed	Text and non-text regions do not have similar textural properties Proper selection of features and other parameters Representative training set Computationally expensive learning phase

Table 3-1: Summary of the text localization techniques

Quite a few hybrid methods have already been presented as well. Zhong et al. combine horizontal spatial variance and color information for segmentation [66]. Liu et al. apply edge detection and use gradient and geometrical characteristics as well as texture features derived from wavelet domain to classify contours [67]. Mancas-Thillou and Gosselin combine color or gray-level variation with spatial information by using Log–Gabor filters [68]. Pan et al. first apply a Conditional Random Field (CRF) model to distinguish between non-text and text components and then group the text components into lines using energy minimization [69]. Ephstein et al. apply an operator called the Stroke Width Transform (SWT) to transform pixel color values to stroke widths and merge neighboring pixel with similar stroke width into connected components [70]. Zhang and Kasturi use the similarity of stroke edges to compute the character energy and then calculate the link energy based on the spatial relationship between neighboring candidate character regions [71]. Zhao et al. combine wavelet-based edge detection with an adaptive run-length smoothing algorithm and projection profile analysis [72]. Yi and Tian first apply image partition to find character candidates based on local gradient features and color uniformity and then character grouping based on joint structural features [73]. In recent works, the Maximally Stable Extremal Regions (MSERs) algorithm is often used to extract character candidates [74-76].

A lot of text localization methods have been evaluated on the public datasets of ICDAR competitions [77-80]. As stated in the final report of the latest competition, there is still a significant margin for improvement.

Table 3.1 contains a summary of the description as well as the restrictions of the above methods.

3.3 The proposed system

A flowchart of the proposed system is shown in Figure 3.5. The method is applied on grayscale images. Color images are first converted to grayscale.

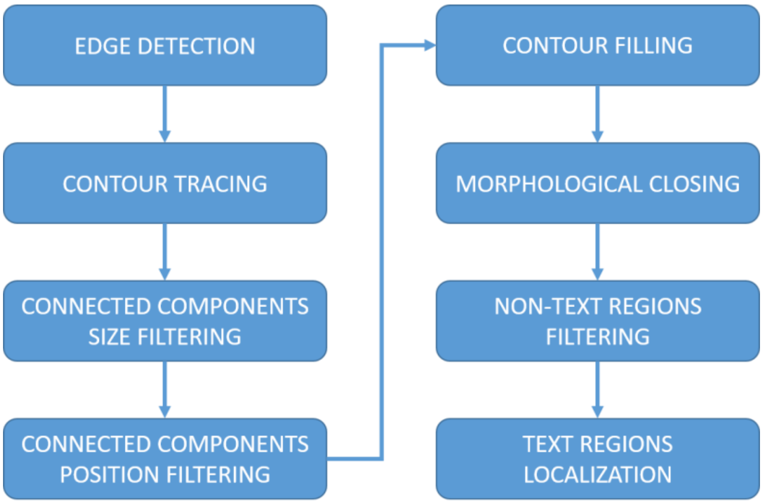

EDGE DETECTION
CONTOUR TRACING
CONNECTED COMPONENTS SIZE FILTERING
CONNECTED COMPONENTS POSITION FILTERING
CONTOUR FILLING
MORPHOLOGICAL CLOSING
NON-TEXT REGIONS FILTERING
TEXT REGIONS LOCALIZATION

Figure 3-5: Flowchart of the proposed text localization system

3.3.1 Edge detection

The Canny operator [81] is performed and the edges are detected (Fig. 3.6). In order to preserve the sharp and clean strokes of the characters, neither histogram equalization nor Gaussian smoothing are applied to the image. Both thresholds of the operator are set to the maximum value, so that only the stronger edges are selected. Those settings proved to be the most appropriate for all the dataset samples that have been used for testing.

3.3.2 Contour tracing

A classical border-following algorithm [42] is used to extract the contours of all the connected components of the image in Figure 3.6. For each contour, the following spatial and geometrical values are calculated and saved:

i. The position of the four vertices of the contour’s bounding rectangle.

ii. The orientation of the minimum area rectangle enclosing the component.

iii. The gaps between the contour and its closest left and right neighbors.

iv. The x-height (the minimum value between the component’s width and height).

As already shown in the previous chapter (Fig. 2.13), the x-height is actually the distance between the base-line and the mean-line of a character.

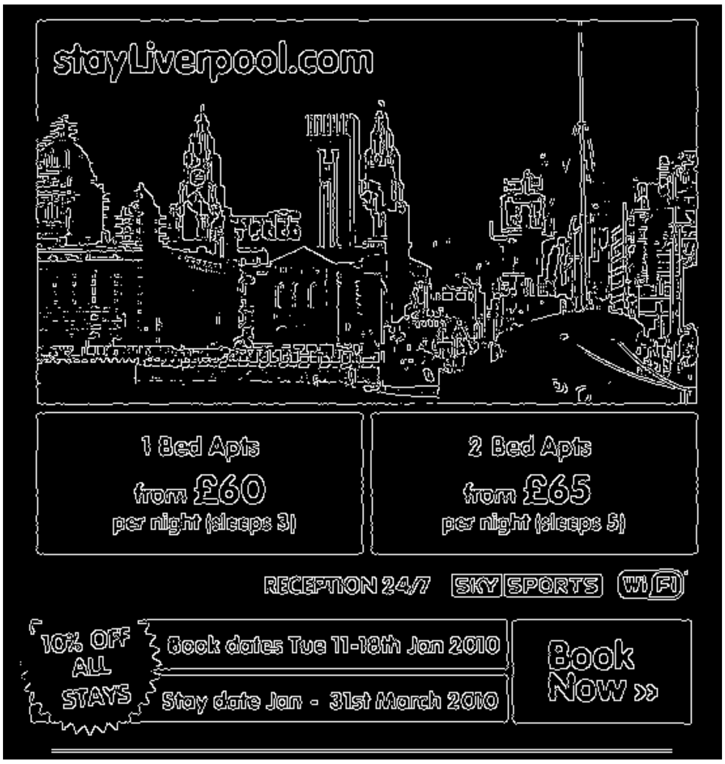


Figure 3-6: The result of Canny edge detection of the image in Figure 3.2

3.3.3 Connected components size filtering

Before proceeding further, the very small and very large components are removed (Fig. 3.7). Very small components are those who have a width or height of no more than just one pixel. On the other hand, both sizes (width and height) of those who are classified as very large are less than a third of the image width.

3.3.4 Connected components position filtering

Next, the spatial information of the remaining components is used to distinguish between aligned (at least one more component is at the same line and has the same orientation) and non-aligned ones. The position of the four vertices of the contours bounding rectangles are used to determine whether two adjacent components are aligned or not. The components are supposed to be aligned if either their bottom or their top vertices are collinear. For example, the outlined region of the image in Figure 3.7 contains aligned letters as well as a non-aligned component (highlighted with red). Non-aligned components are removed at this step.

3.3.5 Contour filling

The aligned components are filled according to their x-height values, so that small ones have a dark gray color while large ones are brighter (Fig. 3.8).

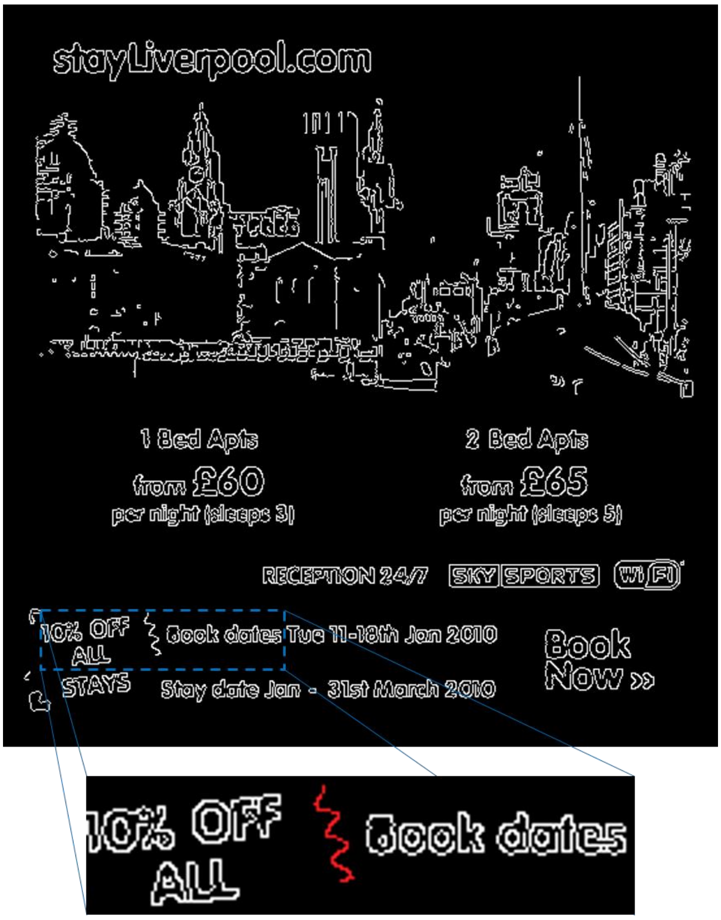


Figure 3-7: Example of a non-aligned component inside the outlined area

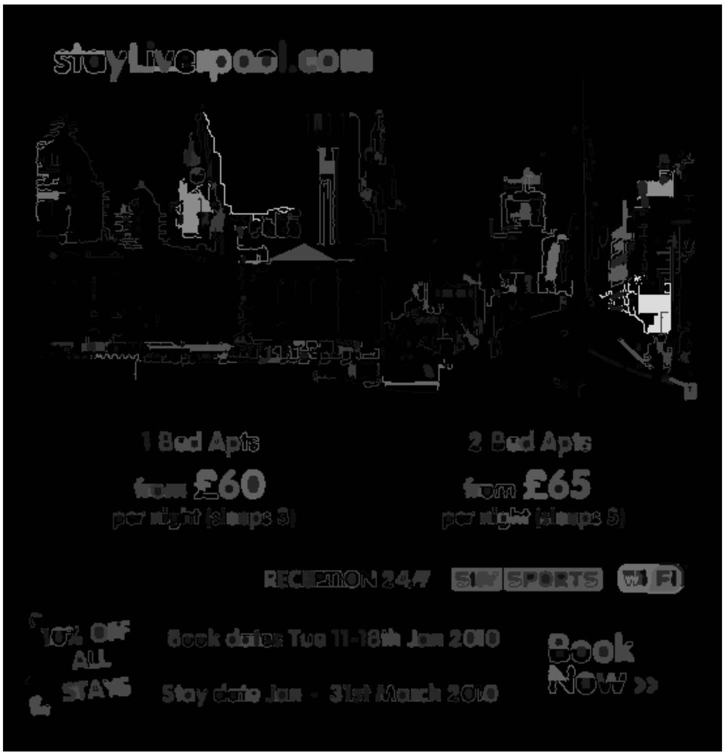


Figure 3-8: Contours are color filled with respect to their size

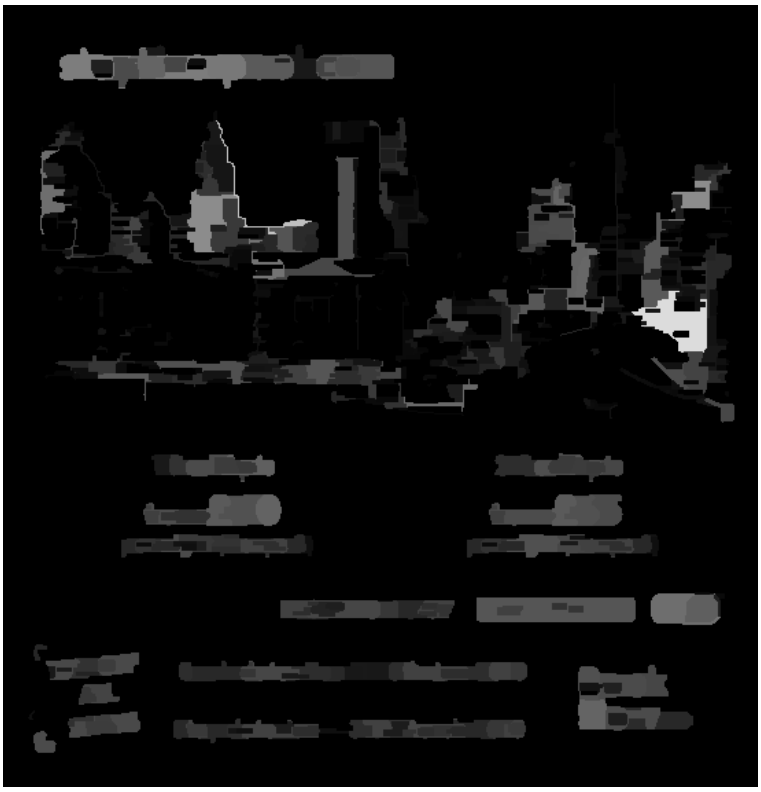


Figure 3-9: Result of morphological closing with dashes of the image in Figure 3.8

3.3.6 Morphological closing

In order to group adjacent characters into text strings, the image is morphologically closed. Closing is actually a dilation with a structuring element followed by an erosion with the same element. The element that is used is a short dash of width equal to the mean gap between neighboring connected components. The result is an image that contains merged components that can be either text or non-text (Fig. 3.9).

3.3.7 Non-text regions filtering

The candidate regions can be classified as either text or non-text assuming that adjacent characters belonging to the same word have similar sizes, similar stroke widths and regular distances between each other.

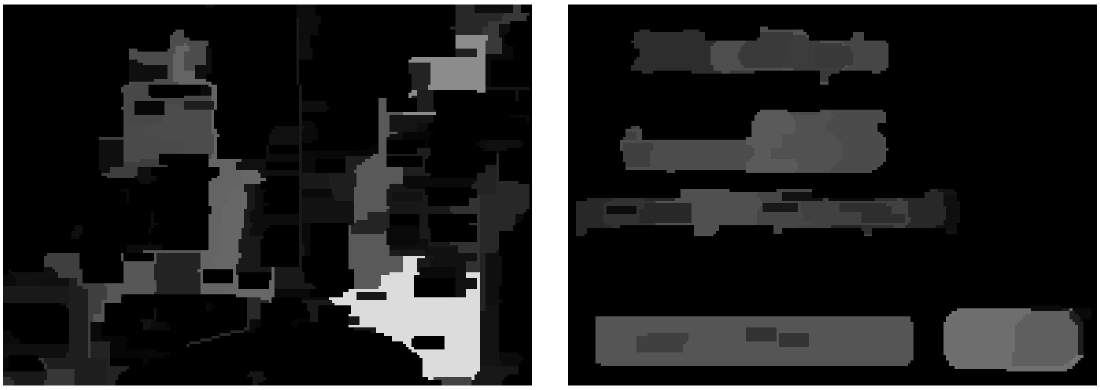


Figure 3-10: High variance non-text (left) and low variance text regions (right)

Based on the first assumption, regions with large variance are considered as non-text and are removed. Since the fill color of each character depends on the x-height and since adjacent characters of the same word have similar x-heights, text regions are expected to have little or no variance in color (Fig. 3.10).

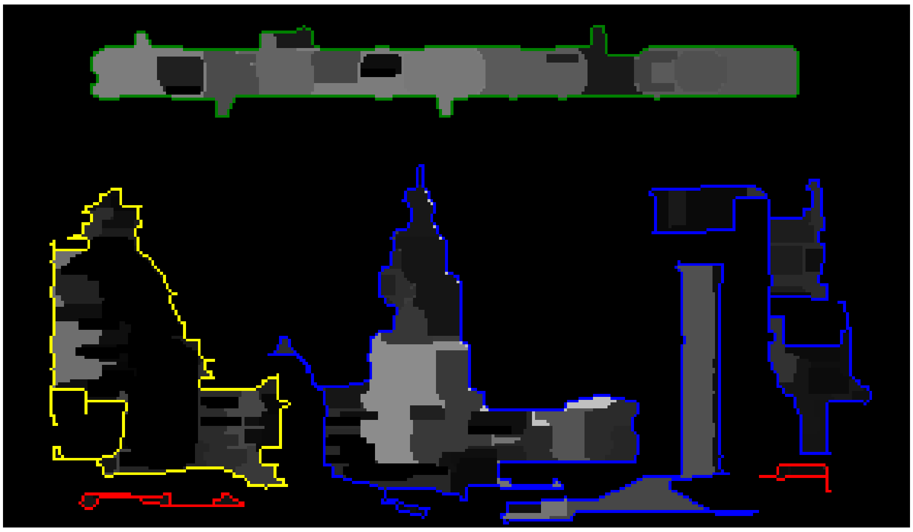


Figure 3-11: Text region (green outline) and various non-text regions (blue/red/yellow)

According to the second assumption, holes inside text regions should be enclosed by lines of uniform thickness distribution. Holes may be contained in some letters or may be formed during the closing procedure due to merging of neighboring components. In both cases, text regions are expected to have no or regularly sized holes surrounded by areas of nearly constant stroke width. In Figure 3.11 an example of a non-text region containing a large hole and a few small ones is marked with a yellow outline. Around the large hole there are areas of various thicknesses. Regions that include holes are characterized as non-text and are removed from the image if they break into several parts of different sizes after a single erosion.

The last assumption is used to detect and remove from the image less robust regions having a non-uniform shape. A text region is expected to have a uniform (almost rectangular) shape after closing. An example of a uniform text region is marked with a green outline in Figure 3.11. Irregular spacing between non-text components leads to the formation of regions of arbitrary shapes that can easily be identified. Figure 3.11 shows examples of small and large non-text regions having non-uniform shapes, marked with red and blue outlines respectively.

Defining $\mathbb{S}$ as the set of all the regions in the image, the regions are characterized as non-text if either their area (***A***) is less than two times their perimeter (***P***) or their area is less than half the area of their convex hull (***H***):

$$\boldsymbol{A}_{\mathrm{i}} < 2\boldsymbol{P}_{\mathrm{i}} \qquad \forall i \in \mathbb{S} \tag{1}$$

$$\boldsymbol{H}_{\mathrm{i}} > 2\boldsymbol{A}_{\mathrm{i}} \qquad \forall i \in \mathbb{S} \tag{2}$$

3.3.8 Text regions localization

Removing the non-text regions leaves the uniform and robust regions in the image that are supposed to include only text. The outlines of the minimum area rectangles of the text regions are drawn over the grayscale input image (Fig. 3.12). Since morphological closing has actually filled the gaps between characters of the same word or characters between adjacent words of the same text line, localization is achieved in word and text line level. In case paragraph localization is the goal, one more morphological closing is required in order for adjacent words and text lines to be grouped into paragraphs.

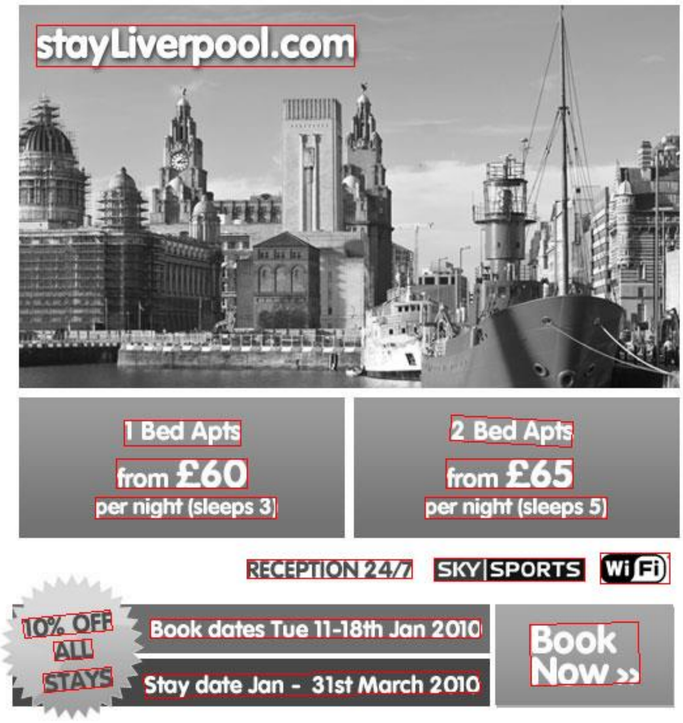


Figure 3-12: Text localization results for the image of Figure 3.2

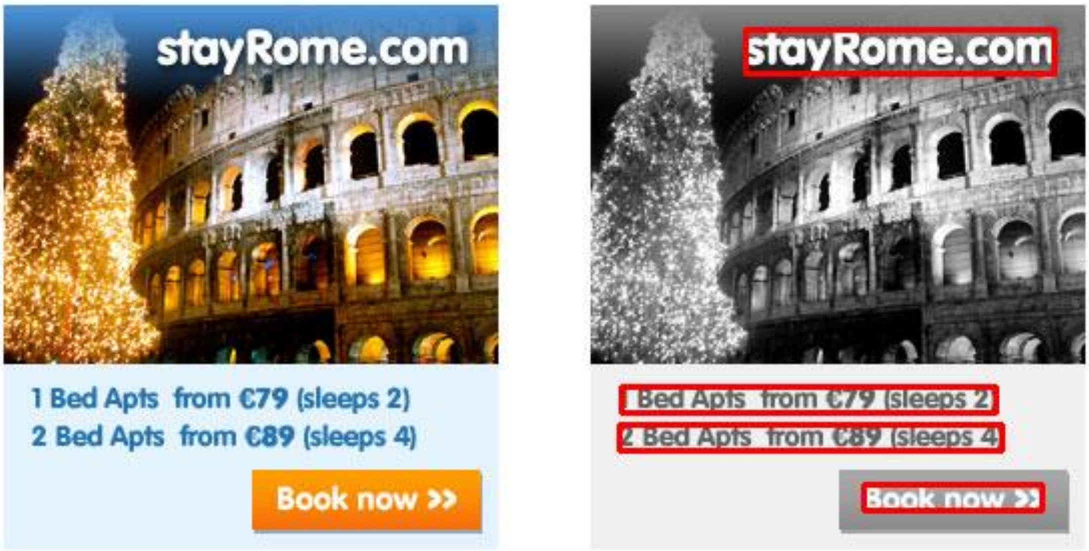
	Precision	**Recall**
Kim's Method	82.98%	62.47%
Yi's Method	67.22%	58.09%
TH-TextLoc System	66.97%	57.68%
Neumann's Method	68.93%	52.54%
TDM_IACS	63.52%	53.52%
LIP6-Retin	62.97%	50.07%
KAIST AIPR System	59.67%	44.57%
ECNU-CCG Method	35.01%	38.32%
Text Hunter	50.05%	25.96%
Proposed Technique	79.10%	64.80%

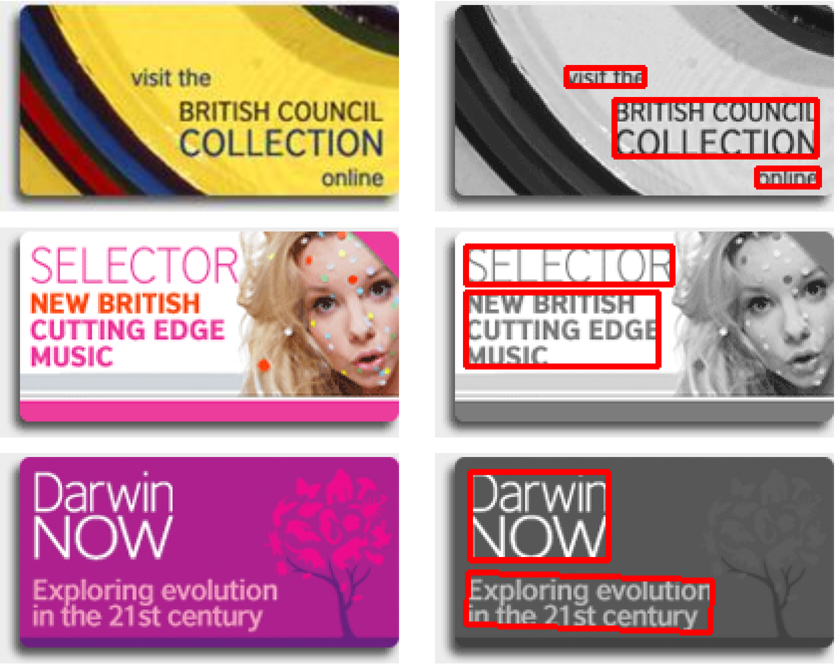
Table 3-2: Comparative results for ICDAR 2011 dataset

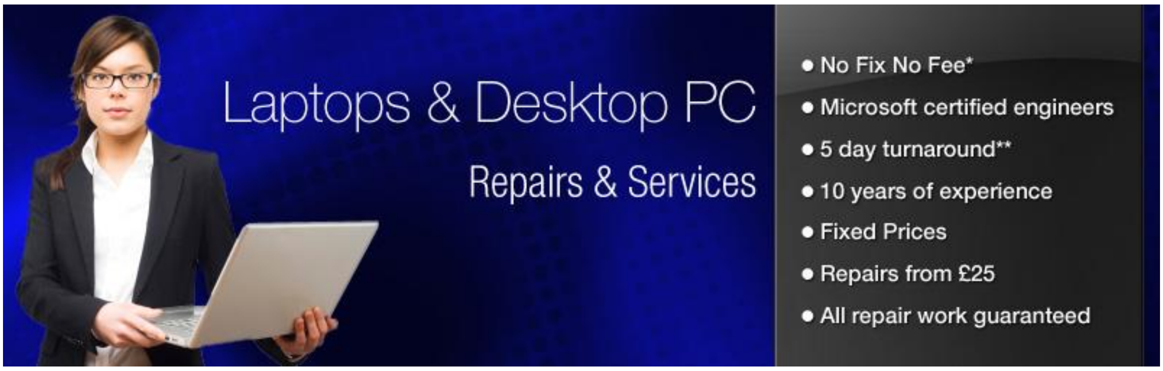
3.4 Experimental results

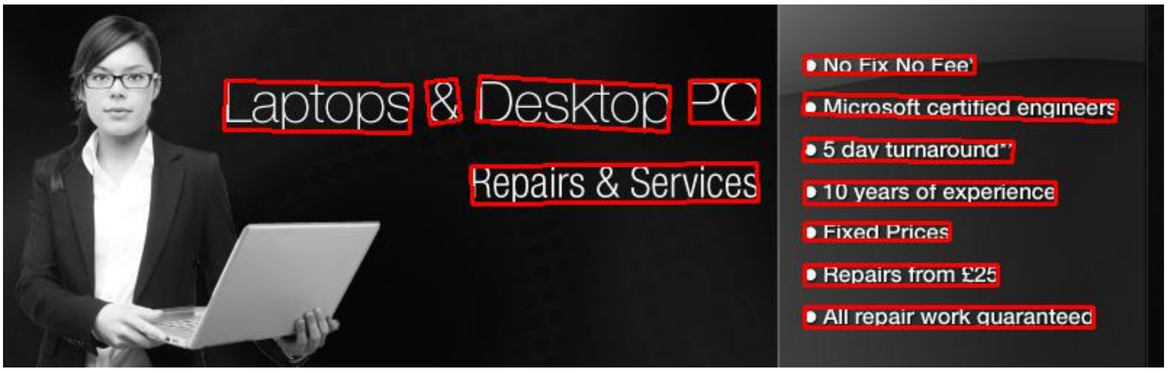
The algorithm has been implemented in C# language. The OpenCV library has also been used. The system was evaluated on the ICDAR 2011 Robust Reading Competition dataset [79] and achieved a precision of 79.1% that is close to the best reported at the competition's paper [79] and a recall of 64.8%, better than the participated methods. A much stricter methodology than the one employed at the competition, has been used: all under and over segmentation text regions, as well as the missed ones, are considered errors. The mean processing time was less than a second per image. Comparative results are shown in Table 3.2.

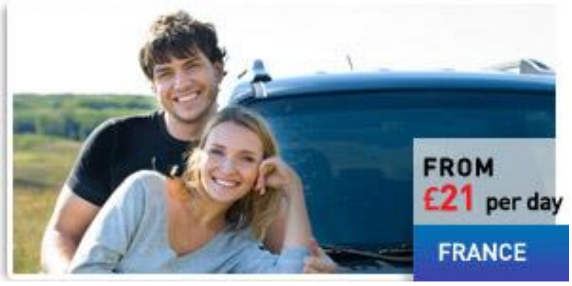
Results of the application of the proposed text localization method on various scene image samples of the ICDAR 2011 dataset are presented in Appendix II while born-digital image segmentation examples as well as a screenshot of the software are shown below.

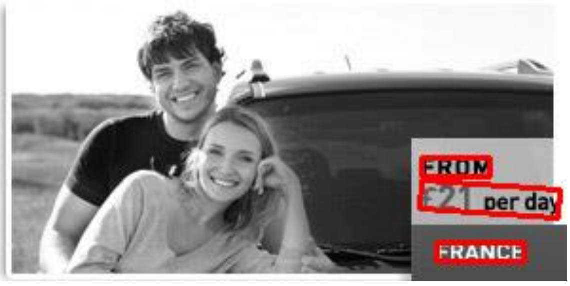

stayRome.com
1 Bed Apts from €79 (sleeps 2)
2 Bed Apts from €89 (sleeps 4)
Book now >>

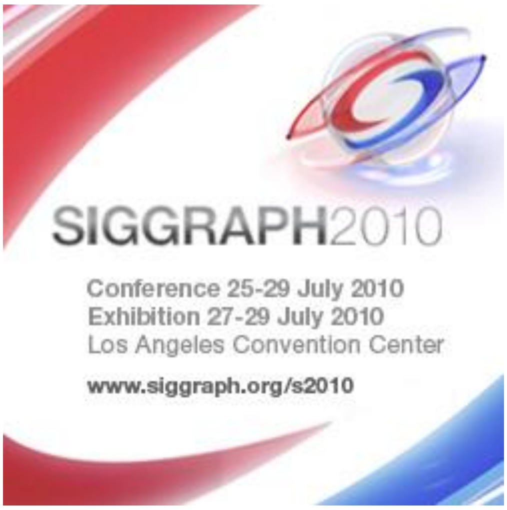


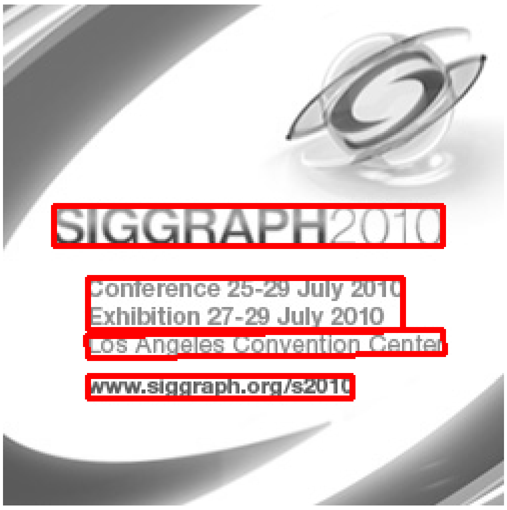

Laptops & Desktop PC
Repairs & Services
No Fix No Fee*
Microsoft certified engineers
5 day turnaround**
10 years of experience
Fixed Prices
Repairs from £25
All repair work guaranteed


Laptops & Desktop PC
Repairs & Services
No Fix No Fee*
Microsoft certified engineers
5 day turnaround**
10 years of experience
Fixed Prices
Repairs from £25
All repair work guaranteed

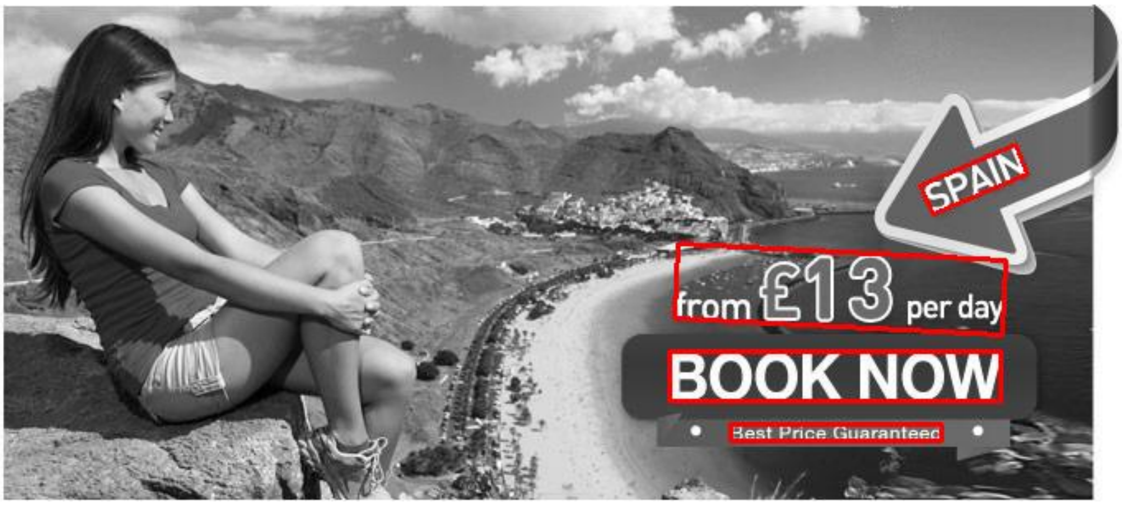

FROM
£21 per day
FRANCE


FROM
£21 per day
FRANCE

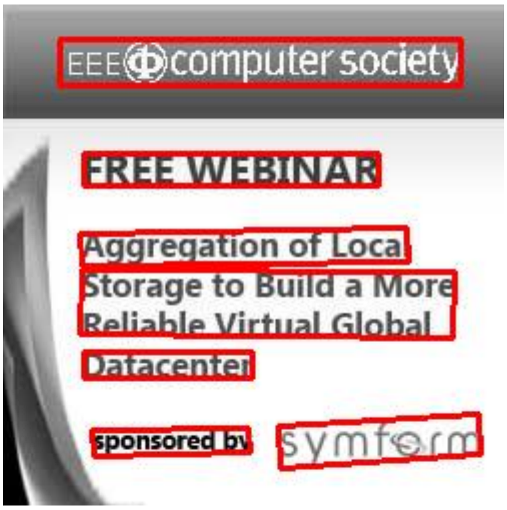

SIGGRAPH2010
Conference 25-29 July 2010
Exhibition 27-29 July 2010
Los Angeles Convention Center
www.siggraph.org/s2010

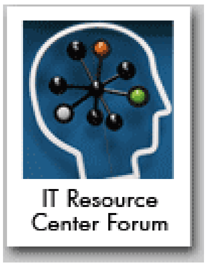

SIGGRAPH2010
Conference 25-29 July 2010
Exhibition 27-29 July 2010
Los Angeles Convention Center
www.siggraph.org/s2010

SPAIN
from £13 per day
BOOK NOW
Best Price Guaranteed


SPAIN
from £13 per day
BOOK NOW
Best Price Guaranteed

IEEE computer society
FREE WEBINAR
Aggregation of Local Storage to Build a More Reliable Virtual Global Datacenter
sponsored by symform


IEEE computer society
FREE WEBINAR
Aggregation of Local Storage to Build a More Reliable Virtual Global Datacenter
sponsored by symform


IT Resource
Center Forum

IT Resource
Center Forum

Free downloadable SPICE simulator
for Analog Devices component evaluation.

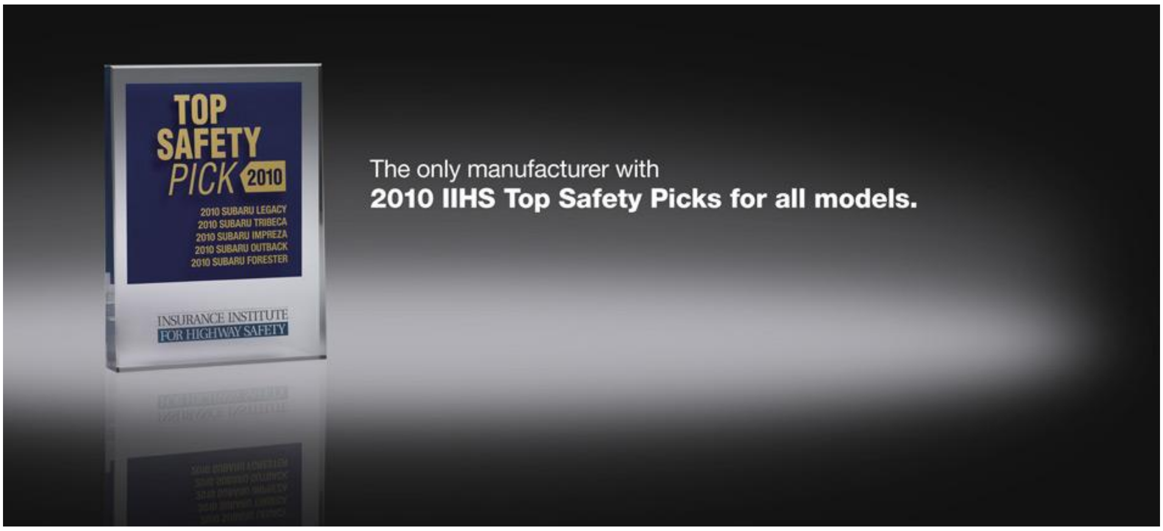
Free downloadable SPICE simulator
for Analog Devices component evaluation

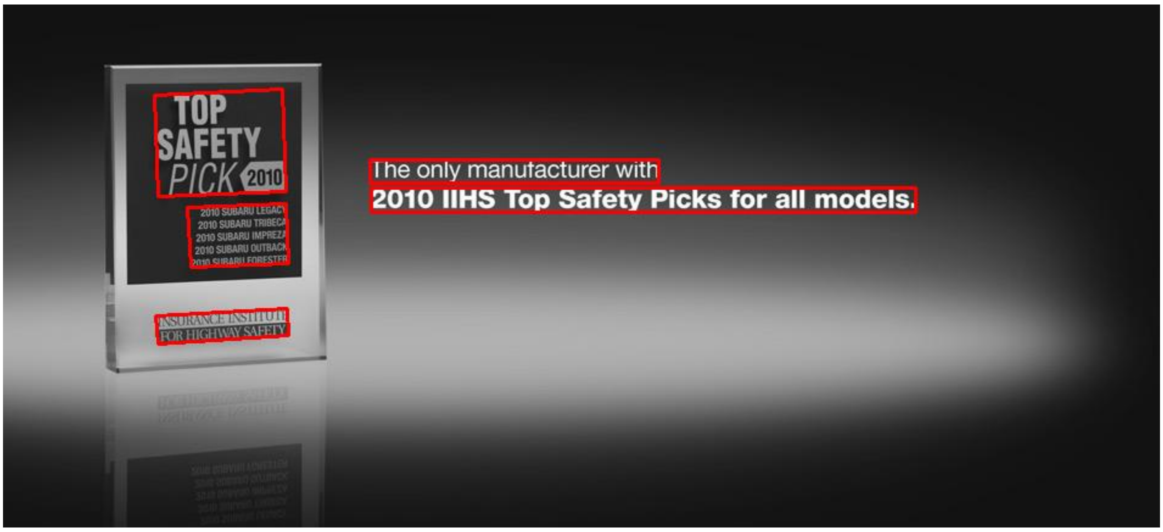
LEADING
THE TRANSFORMATION
VIEW
MWC Highlights
Today!

LEADING
THE TRANSFORMATION
VIEW
MWC Highlights
Today

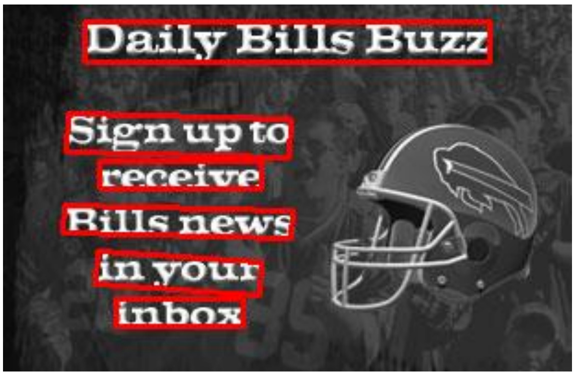
Plan your
journey,
check the
traffic
» join the Driver's Centre

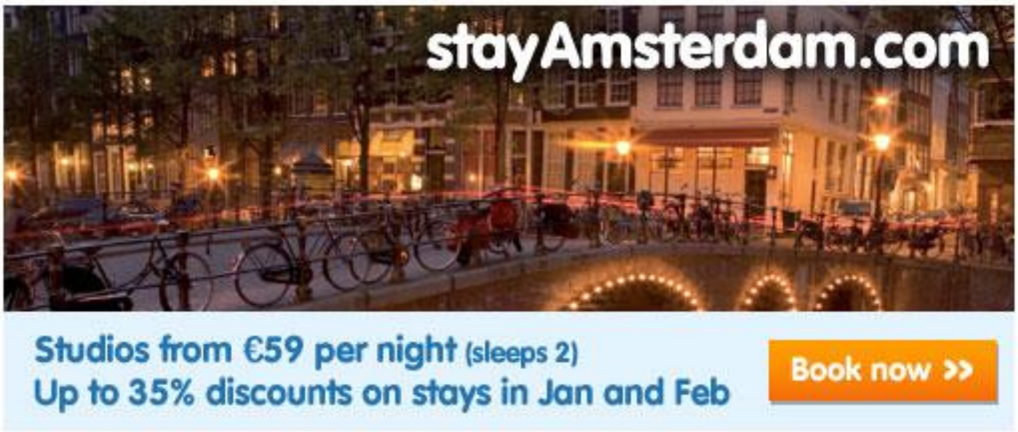
Plan your
journey,
check the
traffic
» join the Driver's Centre

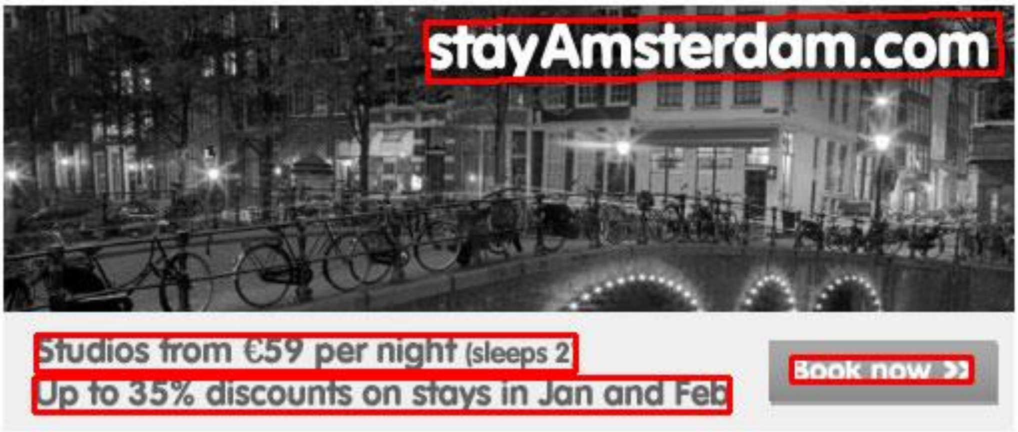

TOP
SAFETY
PICK 2010
2010 SUBARU LEGACY
2010 SUBARU TRIBECA
2010 SUBARU IMPREZA
2010 SUBARU OUTBACK
2010 SUBARU FORESTER
INSURANCE INSTITUTE
FOR HIGHWAY SAFETY
The only manufacturer with
2010 IIHS Top Safety Picks for all models.


TOP
SAFETY
PICK 2010
2010 SUBARU LEGACY
2010 SUBARU TRIBECA
2010 SUBARU IMPREZA
2010 SUBARU OUTBACK
2010 SUBARU FORESTER
INSURANCE INSTITUTE
FOR HIGHWAY SAFETY
The only manufacturer with
2010 IIHS Top Safety Picks for all models.

Daily Bills Buzz
Sign up to
receive
Bills news
in your
inbox


Daily Bills Buzz
Sign up to
receive
Bills news
in your
inbox


stayAmsterdam.com
Studios from €59 per night (sleeps 2)
Up to 35% discounts on stays in Jan and Feb
Book now >>


stayAmsterdam.com
Studios from €59 per night (sleeps 2)
Up to 35% discounts on stays in Jan and Feb
Book now >>

SALE SALE SALE SALE SALE SALE SALE SALE
Save up to 50%
SALE SALE SALE SALE SALE SALE SALE SALE

SALE SALE SALE SALE SALE SALE SALE SALE
Save up to 50%
SALE SALE SALE SALE SALE SALE SALE SALE

Book your
Repair
CLICK
HERE

Book your
Repair
CLICK
HERE

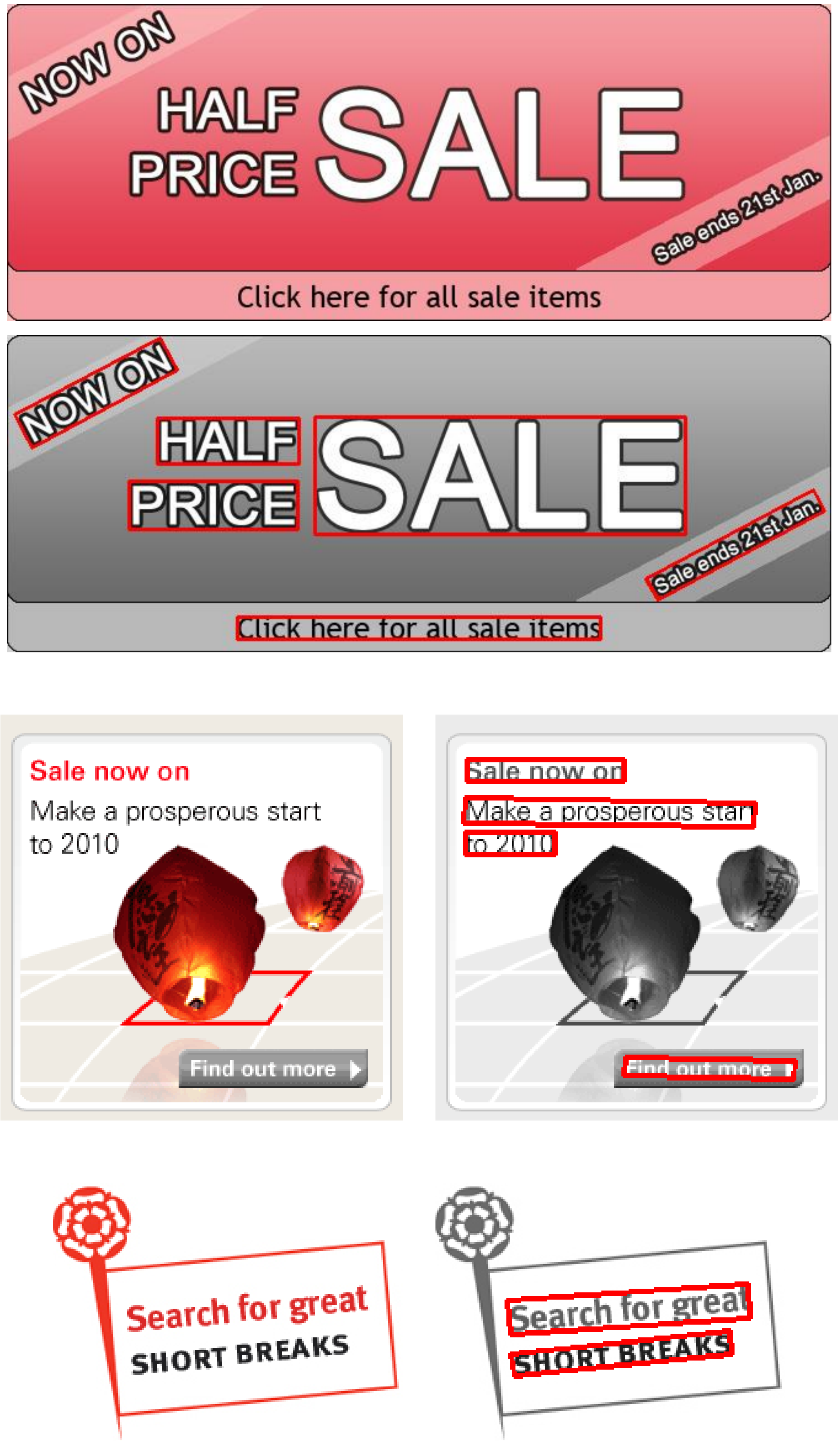


Figure 3-13: Application of the proposed method on various samples from the dataset

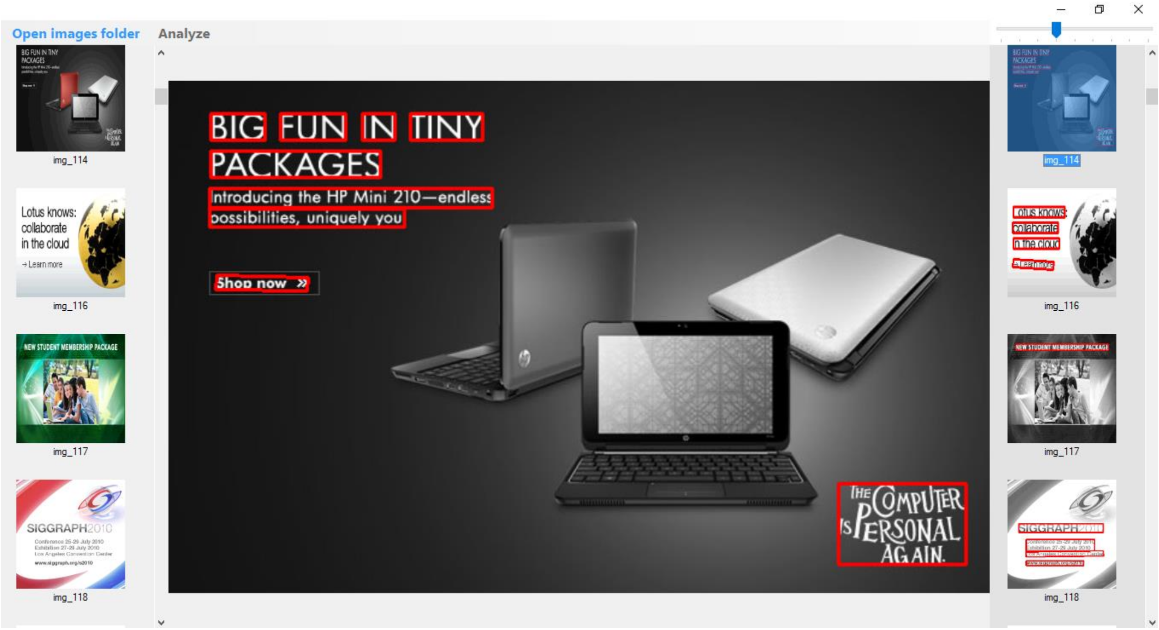
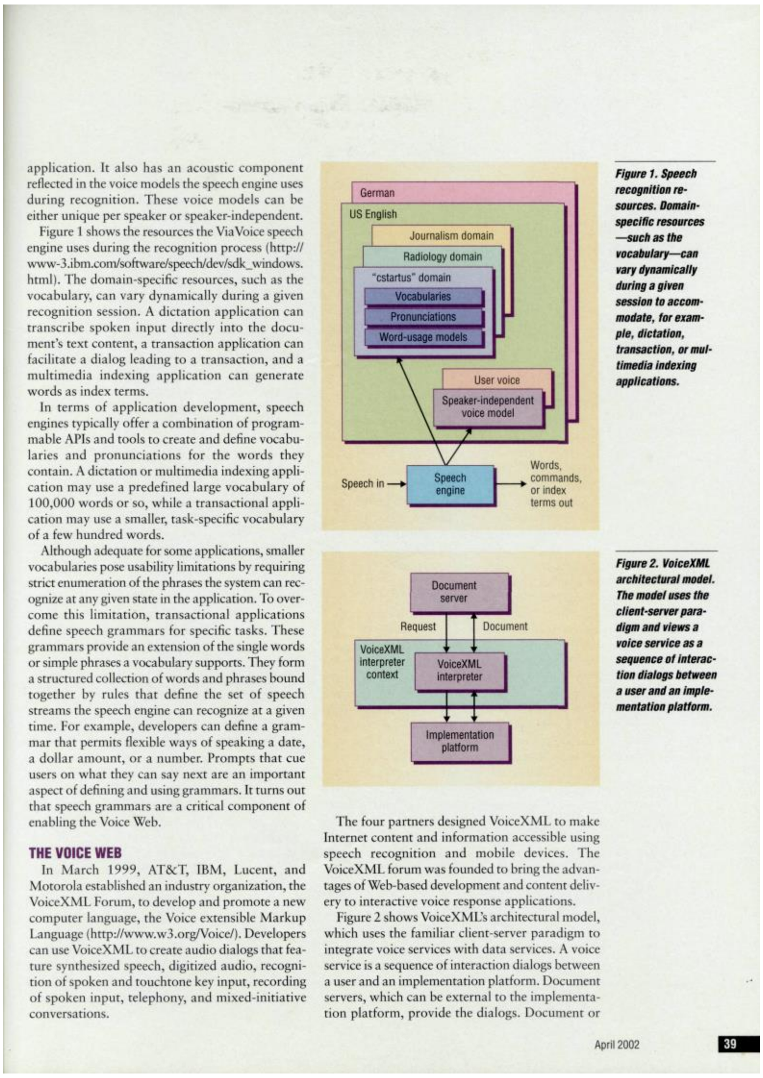

Open images folder
Analyze
img_114
img_116
img_117
img_118
BIG FUN IN TINY PACKAGES
Introducing the HP Mini 210—endless possibilities, uniquely you
Shop now »
THE COMPUTER IS PERSONAL AGAIN.
Lotus knows: collaborate in the cloud
→ Learn more
NEW STUDENT MEMBERSHIP PACKAGE
SIGGRAPH2010

Figure 3-14: The software tool

CHAPTER 4 - A UNIFIED FRAMEWORK

4.1 Introduction

Although layout analysis methods achieve to extract most of the main text blocks and the titles of the document image they often fail to detect text that may be contained in images (photos, logos, graphs, banners etc.). This information is sometimes useful and should rather be extracted as well. A text localization algorithm can be applied to the output images of the page segmentation step in order to further segment them into text and non-text regions.



application. It also has an acoustic component reflected in the voice models the speech engine uses during recognition. These voice models can be either unique per speaker or speaker-independent.

Figure 1 shows the resources the ViaVoice speech engine uses during the recognition process (http://www-3.ibm.com/software/speech/dev/sdk_windows.html). The domain-specific resources, such as the vocabulary, can vary dynamically during a given recognition session. A dictation application can transcribe spoken input directly into the document's text content, a transaction application can facilitate a dialog leading to a transaction, and a multimedia indexing application can generate words as index terms.

In terms of application development, speech engines typically offer a combination of programmable APIs and tools to create and define vocabularies and pronunciations for the words they contain. A dictation or multimedia indexing application may use a predefined large vocabulary of 100,000 words or so, while a transactional application may use a smaller, task-specific vocabulary of a few hundred words.

Although adequate for some applications, smaller vocabularies pose usability limitations by requiring strict enumeration of the phrases the system can recognize at any given state in the application. To overcome this limitation, transactional applications define speech grammars for specific tasks. These grammars provide an extension of the single words or simple phrases a vocabulary supports. They form a structured collection of words and phrases bound together by rules that define the set of speech streams the speech engine can recognize at a given time. For example, developers can define a grammar that permits flexible ways of speaking a date, a dollar amount, or a number. Prompts that cue users on what they can say next are an important aspect of defining and using grammars. It turns out that speech grammars are a critical component of enabling the Voice Web.

THE VOICE WEB

In March 1999, AT&T, IBM, Lucent, and Motorola established an industry organization, the VoiceXML Forum, to develop and promote a new computer language, the Voice extensible Markup Language (http://www.w3.org/Voice/). Developers can use VoiceXML to create audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and touchtone key input, recording of spoken input, telephony, and mixed-initiative conversations.

Figure 1. Speech recognition resources. Domain-specific resources—such as the vocabulary—can vary dynamically during a given session to accommodate, for example, dictation, transaction, or multimedia indexing applications.

Figure 2. VoiceXML architectural model. The model uses the client-server paradigm and views a voice service as a sequence of interaction dialogs between a user and an implementation platform.

The four partners designed VoiceXML to make Internet content and information accessible using speech recognition and mobile devices. The VoiceXML forum was founded to bring the advantages of Web-based development and content delivery to interactive voice response applications.

Figure 2 shows VoiceXML's architectural model, which uses the familiar client-server paradigm to integrate voice services with data services. A voice service is a sequence of interaction dialogs between a user and an implementation platform. Document servers, which can be external to the implementation platform, provide the dialogs. Document or

April 2002 39

Figure 4-1: A sample image of RCDL-2015 dataset containing block diagrams and text

In this chapter a unified technique is proposed that integrates the foreground analysis part of the layout independent method for complex layouts presented in Chapter 2 with the fast and reliable method for text localization presented in Chapter 3. The first one is used to segment the page in text and image blocks while the second one is used to detect text that may be embedded inside the images. Detailed experiments on two public datasets showed that mixing layout analysis and text localization techniques can lead to improved page segmentation and text extraction results.

No a priori knowledge of the page format and the images characteristics is required and there is no limitation on character font, size, orientation and color. Morphological operations are applied in order to connect neighboring components. Contour is simultaneously used for the extraction and classification of images and text blocks. Edge detection and a set of criteria based on spatial and geometrical features of the connected components are used for the detection of text embedded in images (photos, logos, graphs, banners etc.).

The proposed technique and the included modules are presented in detail in Section 4.2, while experimental results are presented in Section 4.3.

4.2 The proposed system

The proposed system is presented in Figure 4.2, while the detailed description of the tasks follows. The method is applied on binary images. Therefore, the image is first transformed to grayscale and then binarized using Otsu's technique [43].

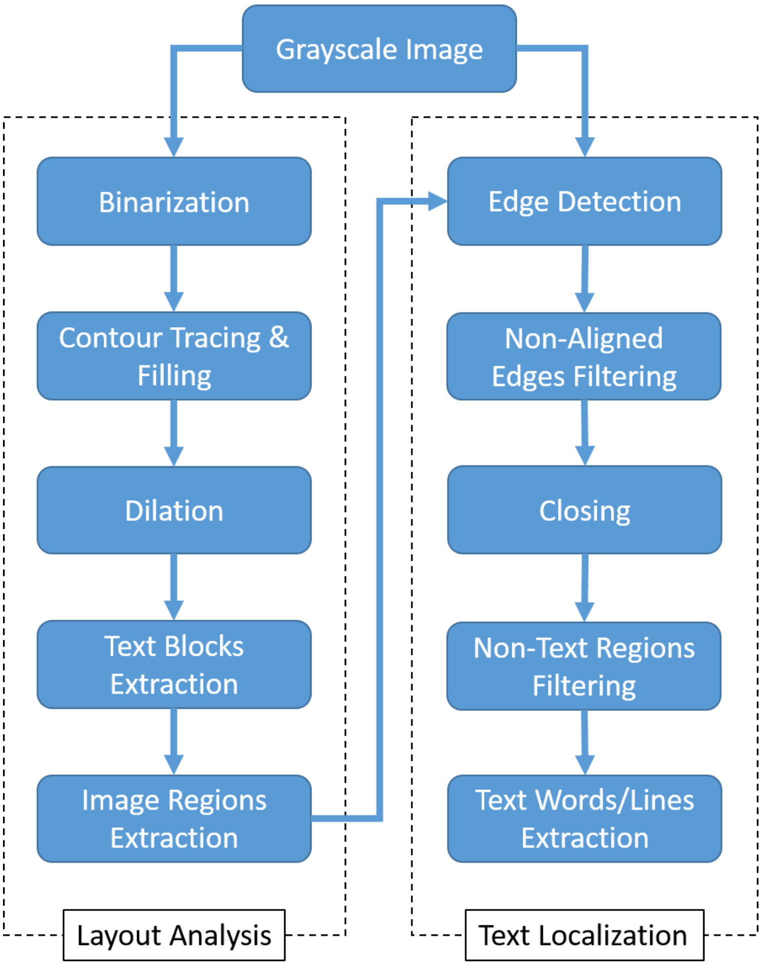


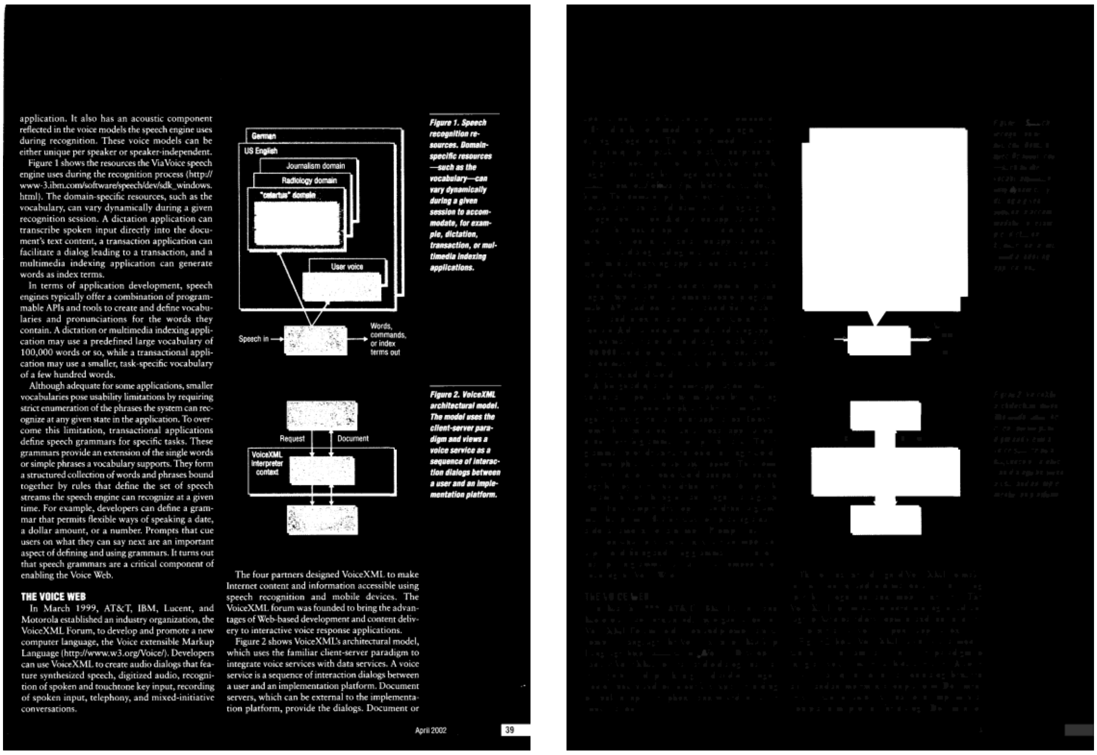
Figure 4-2: The proposed unified framework

4.2.1 Layout analysis

The binary image is inverted, so that foreground pixels are white while background is black. The border-following algorithm used in the previous chapters [42] is applied to the image and the external contours of all the connected components are detected. The contours are used to determine the size of the main body of small text characters (x-height). That is the distance between the base-line and the mean-line of the lower-case letters as already mentioned. The minimum size between width and height of most lower-case letters is equal to the main body.

Since most of the connected components in the document images of printed text are lower-case letters, the main body can be easily calculated as follows: a) all the contours are classified according to the minimum value between their bounding rectangle width and height and b) the value of the most numerous class is considered as the x-height value of the document.

The contours of the connected components are filled before proceeding to the next step, in order to better distinguish between the small text and the large text or image regions. The minimum size of the bounding box will be the value of the filling color. The bigger the text, the higher this value is. Contours of more than 255 pixels bounding box minimum size are filled with the maximum available filling color that is 255. The result is a grayscale image, showing small text in dark gray and bigger text and images in lighter color (Fig. 4.3).



application. It also has an acoustic component reflected in the voice models the speech engine uses during recognition. These voice models can be either unique per speaker or speaker-independent.

Figure 1 shows the resources the ViaVoice speech engine uses during the recognition process (http://www-3.ibm.com/software/speech/dev/sdk_windows.html). The domain-specific resources, such as the vocabulary, can vary dynamically during a given recognition session. A dictation application can transcribe spoken input directly into the document's text content, a transaction application can facilitate a dialog leading to a transaction, and a multimedia indexing application can generate words as index terms.

In terms of application development, speech engines typically offer a combination of programmable APIs and tools to create and define vocabularies and pronunciations for the words they contain. A dictation or multimedia indexing application may use a predefined large vocabulary of 100,000 words or so, while a transactional application may use a smaller, task-specific vocabulary of a few hundred words.

Although adequate for some applications, smaller vocabularies pose usability limitations by requiring strict enumeration of the phrases the system can recognize at any given state in the application. To overcome this limitation, transactional applications define speech grammars for specific tasks. These grammars provide an extension of the single words or simple phrases a vocabulary supports. They form a structured collection of words and phrases bound together by rules that define the set of speech streams the speech engine can recognize at a given time. For example, developers can define a grammar that permits flexible ways of speaking a date, a dollar amount, or a number. Prompts that cue users on what they can say next are an important aspect of defining and using grammars. It turns out that speech grammars are a critical component of enabling the Voice Web.

THE VOICE WEB

In March 1999, AT&T, IBM, Lucent, and Motorola established an industry organization, the VoiceXML Forum, to develop and promote a new computer language, the Voice extensible Markup Language (http://www.w3.org/Voice/). Developers can use VoiceXML to create audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and touchtone key input, recording of spoken input, telephony, and mixed-initiative conversations.

Figure 1. Speech recognition resources. Domain-specific resources—such as the vocabulary—can vary dynamically during a given session to accommodate, for example, dictation, transaction, or multimedia indexing applications.

Figure 2. VoiceXML architectural model. The model uses the client-server paradigm and views a voice service as a sequence of interaction dialogs between a user and an implementation platform.

The four partners designed VoiceXML to make Internet content and information accessible using speech recognition and mobile devices. The VoiceXML forum was founded to bring the advantages of Web-based development and content delivery to interactive voice response applications.

Figure 2 shows VoiceXML's architectural model, which uses the familiar client-server paradigm to integrate voice services with data services. A voice service is a sequence of interaction dialogs between a user and an implementation platform. Document servers, which can be external to the implementation platform, provide the dialogs. Document or

April 2002 39

Figure 4-3: The components of the inverted binary image (left) are filled (right)

Next, the grayscale image containing the filled contours is dilated by a square structuring element (Fig. 4.4). The size of the structuring element is equal to one third of the main body height, selected by experimental trial. This way the small letters are connected and form text blocks. Smaller structuring elements result in partially connected text blocks, while larger ones may result in overlapping blocks.

The border following algorithm [42] is applied again and the external contours of the connected elements are detected. This time the contours are classified as dark, containing small text, and lighter ones, containing large text or images. For each contour, the pixel values of the included image region are calculated. If the region mostly contains pixels with values more than two times the main body height, it's classified as large text or image. All other regions are considered as main text blocks. All blocks are extracted and added to a new mask, the layout mask (Fig. 4.5).

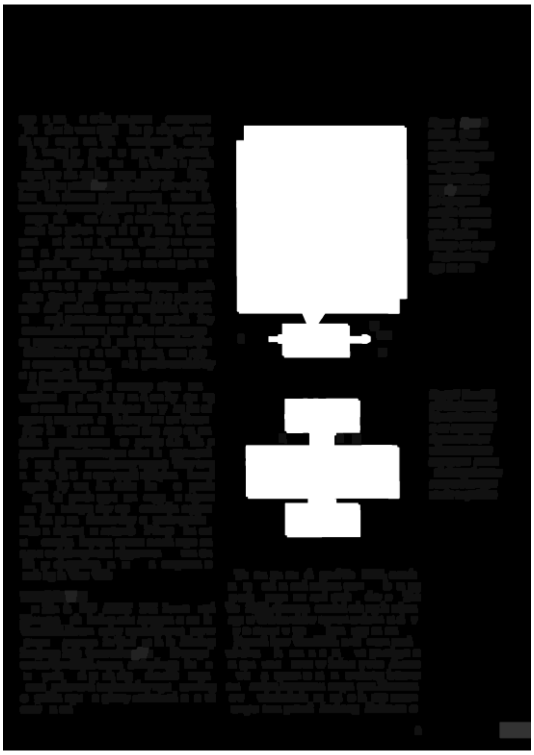


Figure 4-4: In order to connect small text, the image is dilated by a rectangular element

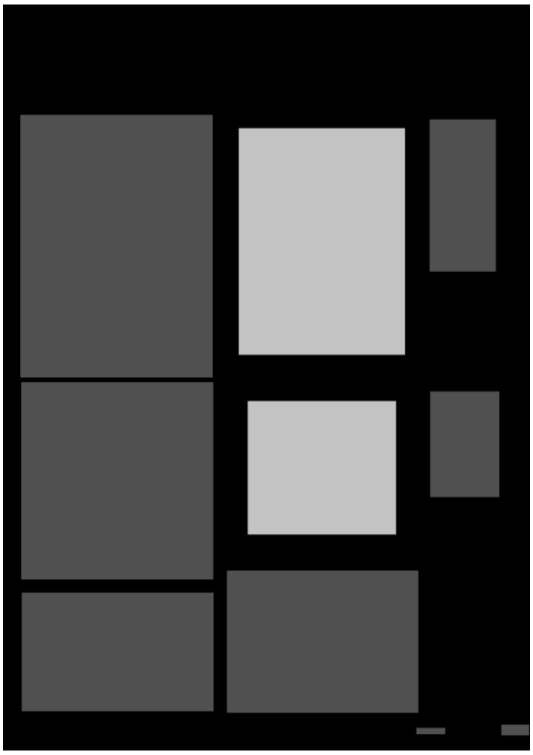


Figure 4-5: The layout contains small text (dark gray) and image (brighter) blocks

4.2.2 Text localization

The extracted images are further processed so that included or embedded text can be extracted as well. As described in Section 3.3.1, the Canny operator [81] is performed first and the edges are detected (Fig. 4.6). In order to preserve the sharp and clean strokes of the characters, neither histogram equalization nor Gaussian smoothing are applied to the image. Both thresholds of the operator are set to the maximum value, so that only the stronger edges are selected. Those settings proved to be the most appropriate for all the dataset samples that have been used for testing.

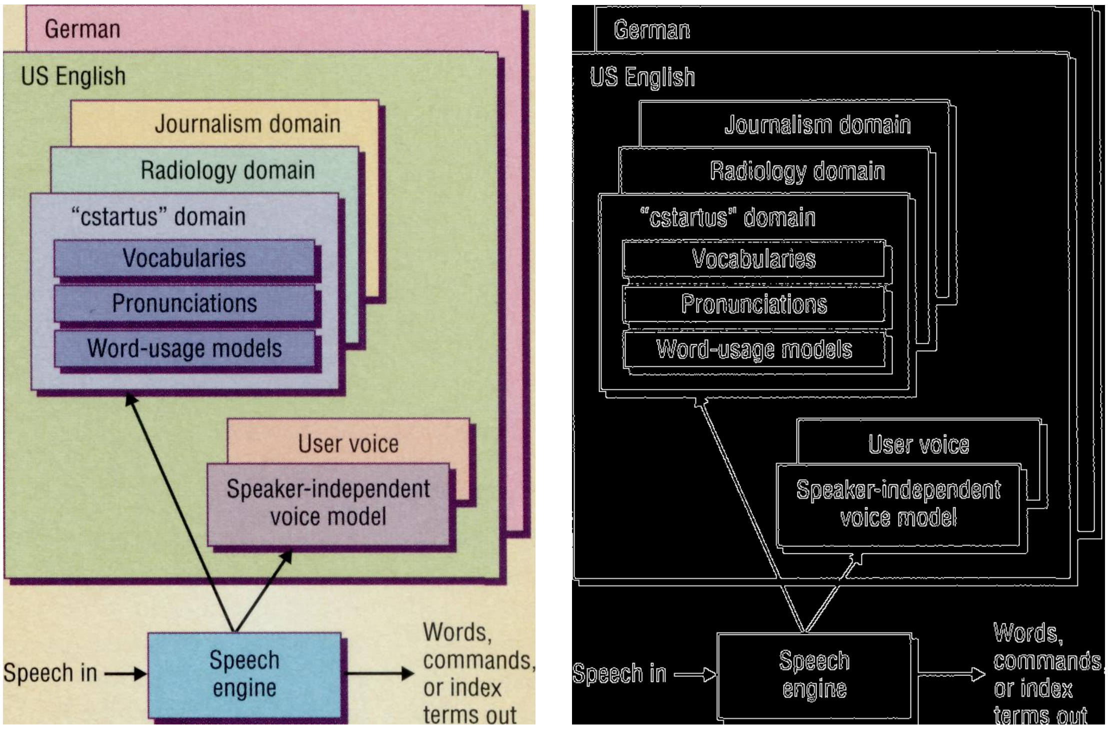


Figure 4-6: Canny edges (right) of the block diagram (left)

The border-following algorithm [42] is used to extract the contours of all the edges. For each contour, spatial and geometrical values are calculated and saved: i) position of the four vertices of the contour's bounding rectangle ii) orientation of the minimum area rectangle enclosing the component iii) gaps between the contour and its closest left-right neighbors and iv) the minimum value between the component's width and height (x-height).

Before proceeding further, the very small and very large components are removed. Very small components are those who have a width or height of no more than just one pixel. On the other hand, the maximum size of those who are classified as very large is less than a third of the image width. Moreover, the spatial information of the components is used to distinguish between aligned (at least one more component is at the same line and has the same orientation) and non-aligned ones. The position of the four vertices of the contours bounding rectangles are used to determine whether two adjacent components are aligned or not. The components are supposed to be aligned if either their bottom or their top vertices are collinear. Non-aligned components are also removed at this step (Fig. 4.7).

The aligned components are filled according to their x-height values, so that small ones have a dark gray color while large ones are brighter. In order to group adjacent characters into text strings, the image is morphologically closed. The element that is used is a short dash of width equal to the mean gap between neighboring connected components. A single erosion is performed after closing so that very thin components are filtered out. The result is an image that contains merged components that can be either text or non-text (Fig. 4.8).

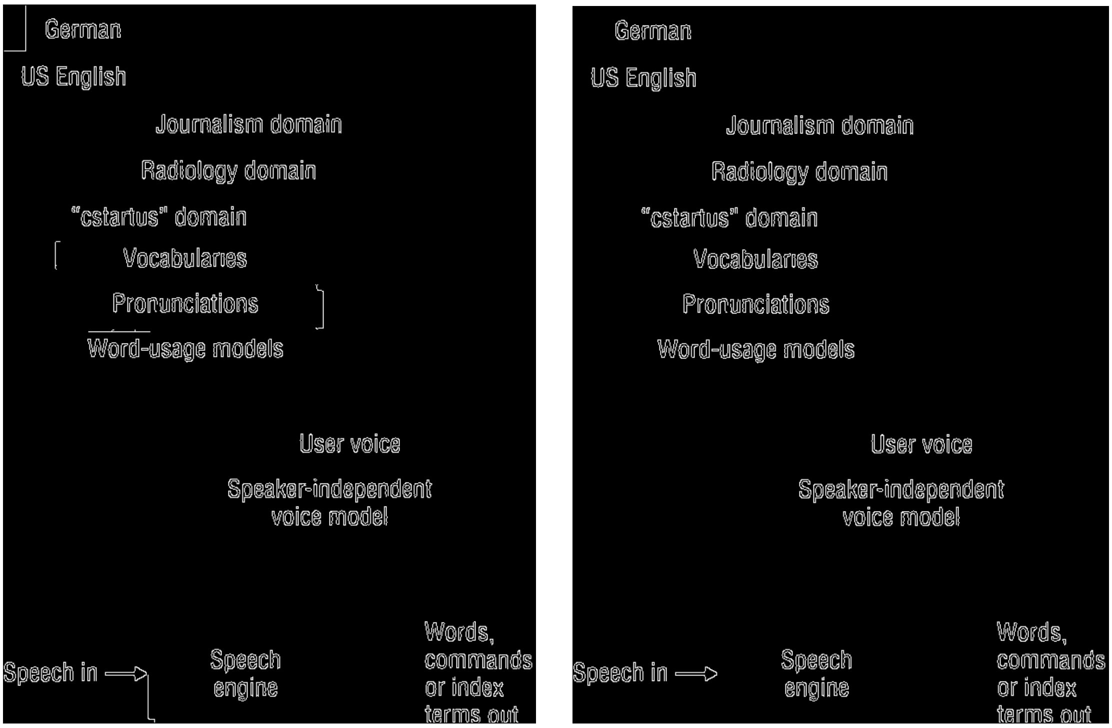


Figure 4-7: Filtering out small/large and nonaligned contours

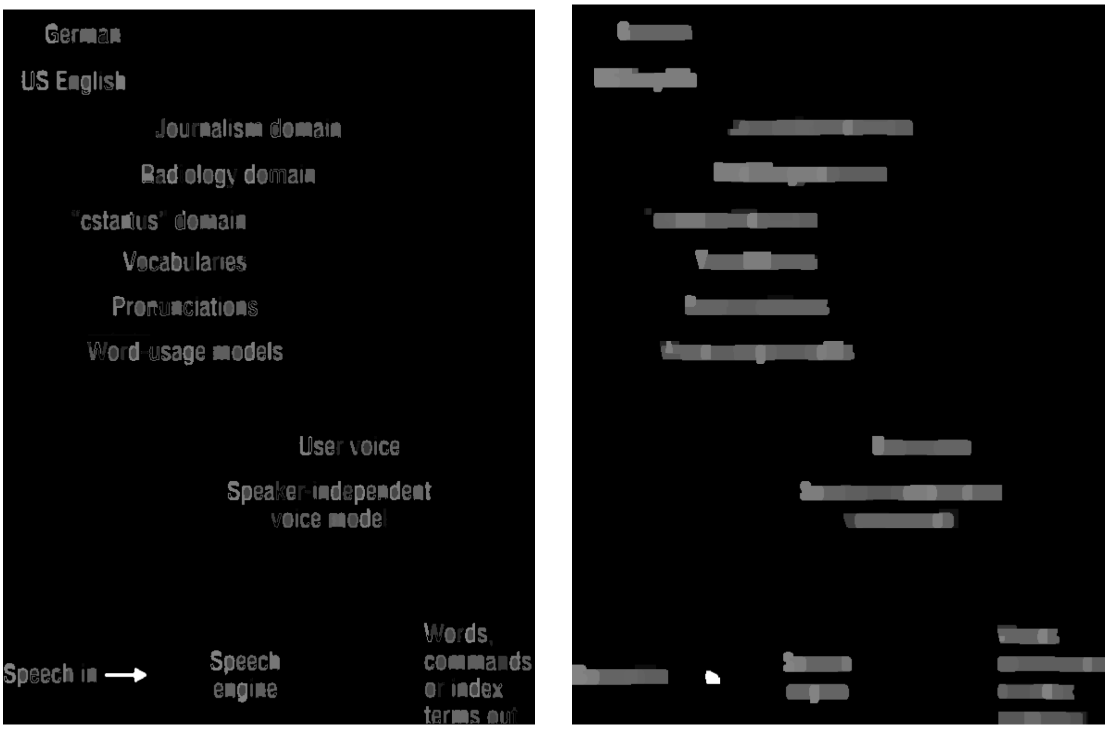
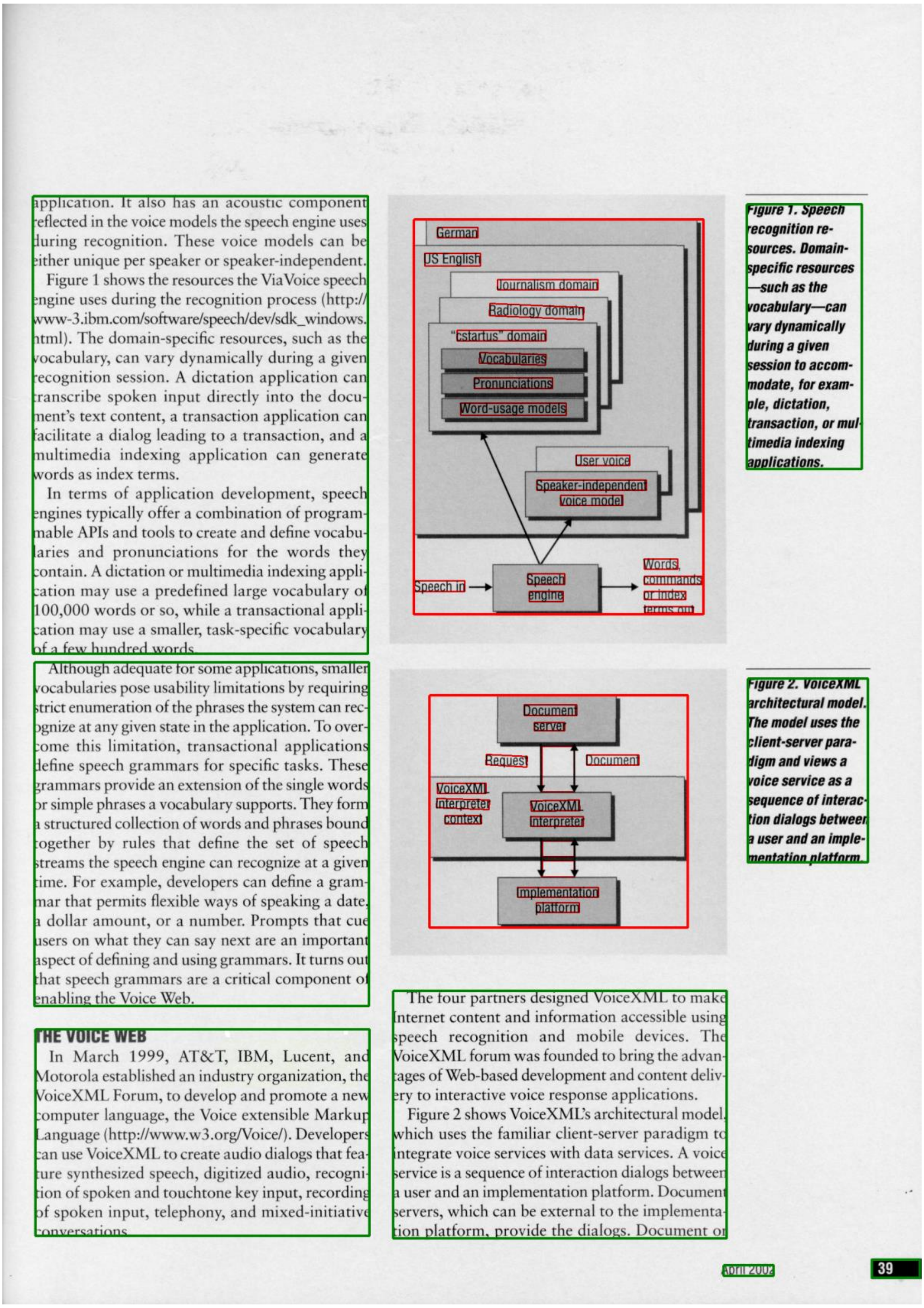


Figure 4-8: Closing with dashes followed by erosion (right) of the filled contours (left)

Assuming that adjacent characters belonging to the same word have similar sizes, similar stroke widths and regular distances between each other, less robust regions with non-uniform shapes and high color variance are classified as non-text and are removed. Bright regions with small size are also removed. Figure 4.9 shows the segmentation result of the page in Figure 4.1 using the unified approach presented in this chapter.



application. It also has an acoustic component reflected in the voice models the speech engine uses during recognition. These voice models can be either unique per speaker or speaker-independent.

Figure 1 shows the resources the ViaVoice speech engine uses during the recognition process (http://www-3.ibm.com/software/speech/dev/sdk_windows.html). The domain-specific resources, such as the vocabulary, can vary dynamically during a given recognition session. A dictation application can transcribe spoken input directly into the document's text content, a transaction application can facilitate a dialog leading to a transaction, and a multimedia indexing application can generate words as index terms.

In terms of application development, speech engines typically offer a combination of programmable APIs and tools to create and define vocabularies and pronunciations for the words they contain. A dictation or multimedia indexing application may use a predefined large vocabulary of 100,000 words or so, while a transactional application may use a smaller, task-specific vocabulary of a few hundred words.

Although adequate for some applications, smaller vocabularies pose usability limitations by requiring strict enumeration of the phrases the system can recognize at any given state in the application. To overcome this limitation, transactional applications define speech grammars for specific tasks. These grammars provide an extension of the single words or simple phrases a vocabulary supports. They form a structured collection of words and phrases bound together by rules that define the set of speech streams the speech engine can recognize at a given time. For example, developers can define a grammar that permits flexible ways of speaking a date, a dollar amount, or a number. Prompts that cue users on what they can say next are an important aspect of defining and using grammars. It turns out that speech grammars are a critical component of enabling the Voice Web.

THE VOICE WEB

In March 1999, AT&T, IBM, Lucent, and Motorola established an industry organization, the VoiceXML Forum, to develop and promote a new computer language, the Voice extensible Markup Language (http://www.w3.org/Voice/). Developers can use VoiceXML to create audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and touchtone key input, recording of spoken input, telephony, and mixed-initiative conversations.

The four partners designed VoiceXML to make Internet content and information accessible using speech recognition and mobile devices. The VoiceXML forum was founded to bring the advantages of Web-based development and content delivery to interactive voice response applications.

Figure 2 shows VoiceXML's architectural model, which uses the familiar client-server paradigm to integrate voice services with data services. A voice service is a sequence of interaction dialogs between a user and an implementation platform. Document servers, which can be external to the implementation platform, provide the dialogs. Document or

Figure 1. Speech recognition resources. Domain-specific resources—such as the vocabulary—can vary dynamically during a given session to accommodate, for example, dictation, transaction, or multimedia indexing applications.

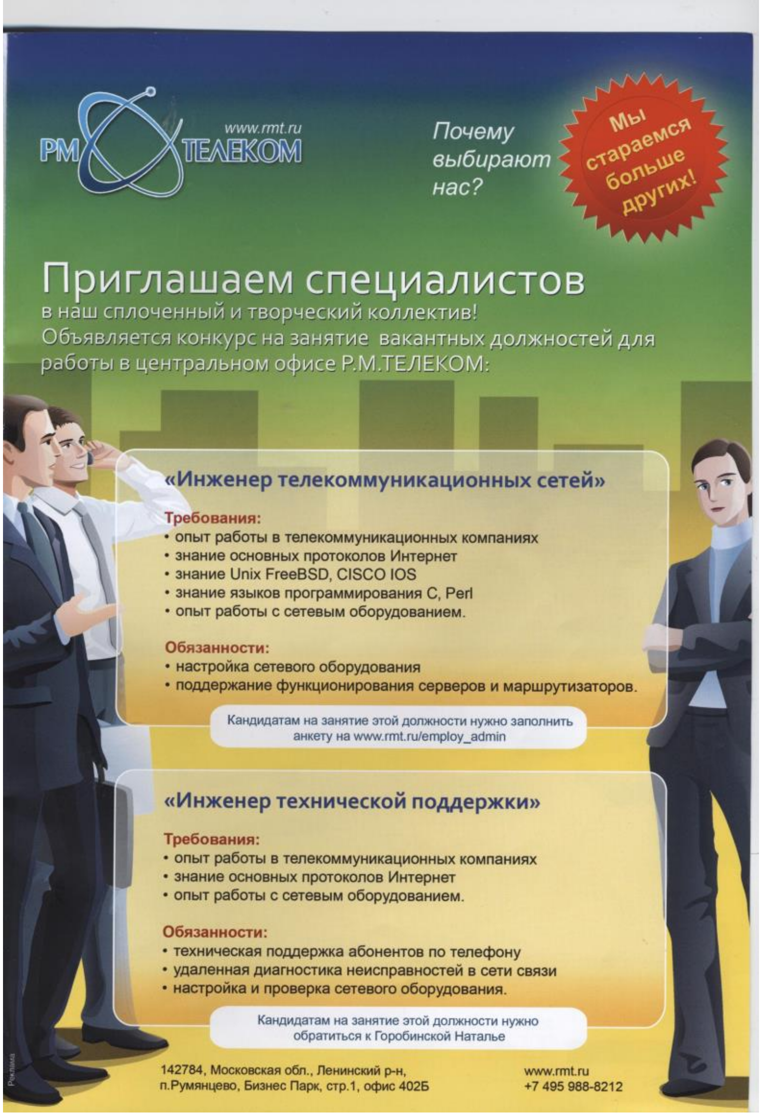
Figure 2. VoiceXML architectural model. The model uses the client-server paradigm and views a voice service as a sequence of interaction dialogs between a user and an implementation platform.

April 2002 39

Figure 4-9: Segmentation of the image in Figure 4.1 using the proposed mixed method

4.3 Experimental results

The algorithm has been coded in C# language. The morphological operations, the contour finding functions and the Canny method implementation of the OpenCV library have been applied. Scanned pages from different newspapers and magazines in Russian [82] as well as the RDCL-2015 dataset of ICDAR 2015 competition [31] were used for testing. The results have shown that the method detects accurately more than 95% of the page components in less than four seconds per page. The text localization part of the algorithm performs very well in regions where the layout analysis fails. On the other hand, the layout analysis part can extract the major text and non-text blocks of the image much faster, keeping the mean processing time low. In most of the cases the combination of layout analysis and text localization gave significantly improved results with a very small overload compared to the layout analysis method alone.



PM ТЕЛЕКОМ

www.rmt.ru

Почему выбирают нас?

Мы стараемся больше других!

Приглашаем специалистов

в наш сплоченный и творческий коллектив!

Объявляется конкурс на занятие вакантных должностей для работы в центральном офисе Р.М.ТЕЛЕКОМ:

«Инженер телекоммуникационных сетей»

Требования:

- опыт работы в телекоммуникационных компаниях
- знание основных протоколов Интернет
- знание Unix FreeBSD, CISCO IOS
- знание языков программирования C, Perl
- опыт работы с сетевым оборудованием.

Обязанности:

- настройка сетевого оборудования
- поддержание функционирования серверов и маршрутизаторов.

Кандидатам на занятие этой должности нужно заполнить анкету на www.rmt.ru/employ_admin

«Инженер технической поддержки»

Требования:

- опыт работы в телекоммуникационных компаниях
- знание основных протоколов Интернет
- опыт работы с сетевым оборудованием.

Обязанности:

- техническая поддержка абонентов по телефону
- удаленная диагностика неисправностей в сети связи
- настройка и проверка сетевого оборудования.

Кандидатам на занятие этой должности нужно обратиться к Горобинской Наталье

142784, Московская обл., Ленинский р-н,
п.Румянцево, Бизнес Парк, стр.1, офис 402Б

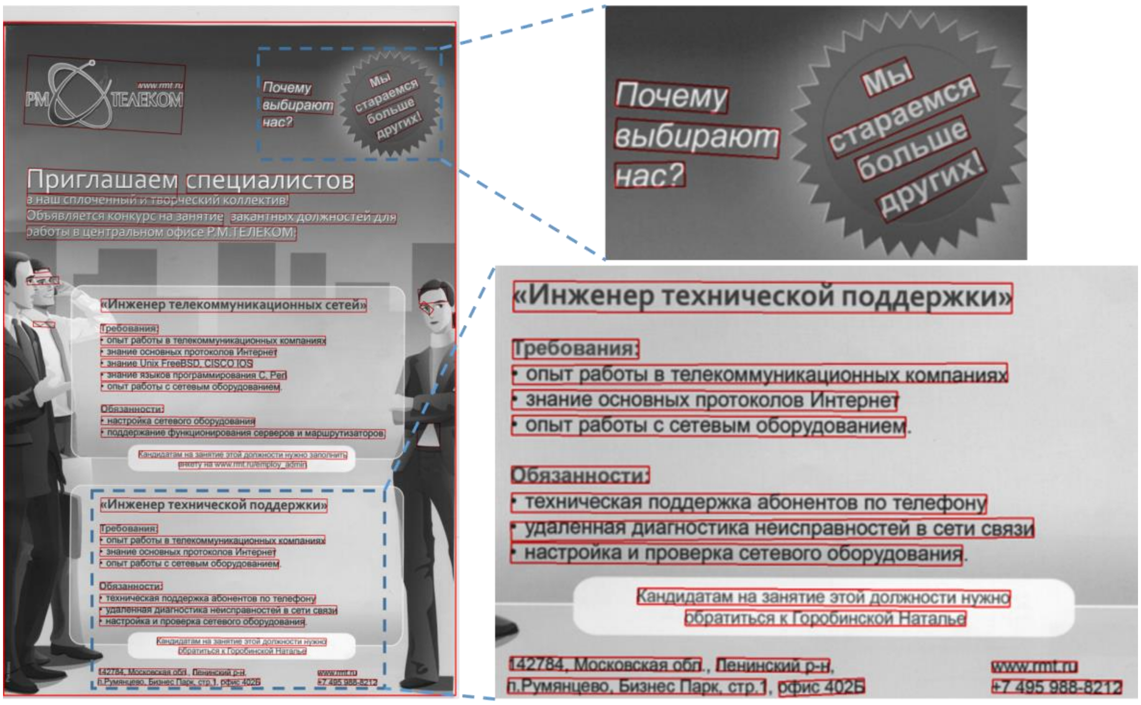
www.rmt.ru
+7 495 988-8212

Реклама

Figure 4-10: A page with gradient background and overlapping text and graphics regions

Figure 4.10 shows an example of a page with gradient background and overlapping text and graphics regions that cannot be segmented using classical layout analysis techniques. The text localization algorithm achieves however to extract all the textual information from the image (Fig. 4.11). Figure 4.12 shows another example where images with embedded text are included.

Although the layout analysis algorithm fails to detect the text inside the image frames and also misses some text lines of the main text blocks (Fig. 4.13), the text localization step extracts all the embedded and missing words (Figs. 4.14-15).



РМ ТЕЛЕКОМ
www.rmt.ru
Почему выбирают нас?
Мы стараемся больше других!
Приглашаем специалистов
в наш сплоченный и творческий коллектив!
Объявляется конкурс на занятие вакантных должностей для работы в центральном офисе Р.М.ТЕЛЕКОМ
«Инженер телекоммуникационных сетей»
«Инженер технической поддержки»
Требования:
• опыт работы в телекоммуникационных компаниях
• знание основных протоколов Интернет
• опыт работы с сетевым оборудованием.
Обязанности:
• техническая поддержка абонентов по телефону
• удаленная диагностика неисправностей в сети связи
• настройка и проверка сетевого оборудования.
Кандидатам на занятие этой должности нужно обратиться к Горобинской Наталье
142784, Московская обл., Ленинский р-н, п.Румянцево, Бизнес Парк, стр.1, офис 402Б
www.rmt.ru
+7 495 988-8212

Figure 4-11: Text detection in the page of Figure 4.10 with the proposed technique

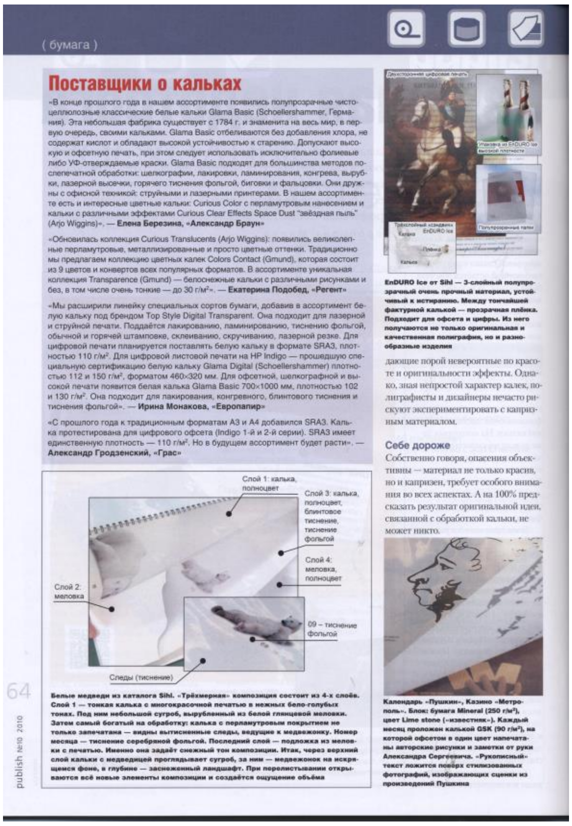
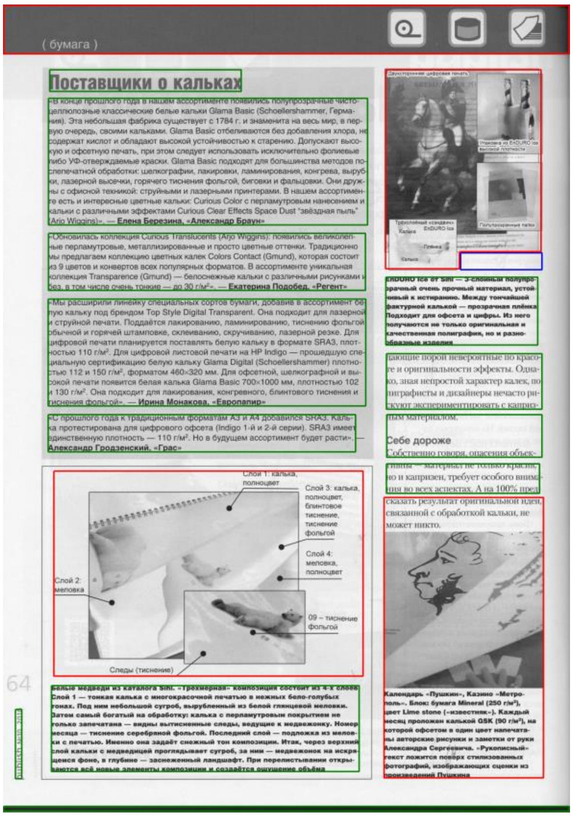


Figure 4-12. A page with images containing text



(бумага)

Поставщики о кальках

«В конце прошлого года в нашем ассортименте появились полупрозрачные чистоцеллюлозные классические белые кальки Glama Basic (Schoellershammer, Германия). Эта небольшая фабрика существует с 1784 г. и знаменита на весь мир, в первую очередь, своими кальками. Glama Basic отбеливаются без добавления хлора, не содержат кислот и обладают высокой устойчивостью к старению. Допускают высокую и офсетную печать, при этом следует использовать исключительно фолиевые либо УФ-отверждаемые краски. Glama Basic подходят для большинства методов послепечатной обработки: шелкографии, лакировки, ламинирования, конгрева, вырубки, лазерной высечки, горячего тиснения фольгой, биговки и фальцовки. Они дружны с офисной техникой: струйными и лазерными принтерами. В нашем ассортименте есть и интересные цветные кальки: Curious Color с перламутровым нанесением и кальки с различными эффектами Curious Clear Effects Space Dust "звёздная пыль" (Arjo Wiggins)». — **Елена Березина, «Александр Браун»**

«Обновилась коллекция Curious Translucents (Arjo Wiggins): появились великолепные перламутровые, металлизированные и просто цветные оттенки. Традиционно мы предлагаем коллекцию цветных калек Colors Contact (Gmund), которая состоит из 9 цветов и конвертов всех популярных форматов. В ассортименте уникальная коллекция Transparence (Gmund) — белоснежные кальки с различными рисунками и без, в том числе очень тонкие — до 30 г/м²». — **Екатерина Подобед, «Регент»**

«Мы расширили линейку специальных сортов бумаги, добавив в ассортимент белую кальку под брендом Top Style Digital Transparent. Она подходит для лазерной и струйной печати. Поддаётся лакированию, ламинированию, тиснению фольгой обычной и горячей штамповке, склеиванию, скручиванию, лазерной резке. Для цифровой печати планируется поставлять белую кальку в формате SRA3, плотностью 110 г/м². Для цифровой листовой печати на HP Indigo — прошедшую специальную сертификацию белую кальку Glama Digital (Schoellershammer) плотностью 112 и 150 г/м², форматом 460×320 мм. Для офсетной, шелкографной и высокой печати появится белая калька Glama Basic 700×1000 мм, плотностью 102 и 130 г/м². Она подходит для лакирования, конгревного, блинтового тиснения и тиснения фольгой». — **Ирина Монакова, «Европапир»**

«С прошлого года к традиционным форматам A3 и A4 добавился SRA3. Калька протестирована для цифрового офсета (Indigo 1-й и 2-й серии). SRA3 имеет единственную плотность — 110 г/м². Но в будущем ассортимент будет расти». — **Александр Гродзенский, «Грас»**

Слой 1: калька, полноцвет
Слой 3: калька, полноцвет, блинтовое тиснение, тиснение фольгой
Слой 4: меловка, полноцвет
Слой 2: меловка
09 – тиснение фольгой
Следы (тиснение)

Белые медведи из каталога Sihl. «Трехмерная» композиция состоит из 4-х слоев. Слой 1 — тонкая калька с многокрасочной печатью в нежных бело-голубых тонах. Под ним небольшой сугроб, вырубленный из белой глянцевой меловки. Затем самый богатый на обработку: калька с перламутровым покрытием не только запечатана — видны вытисненные следы, ведущие к медвежонку. Номер месяца — тиснение серебряной фольгой. Последний слой — подложка из меловки с печатью. Именно она задаёт снежный тон композиции. Итак, через верхний слой кальки с медведицей проглядывает сугроб, за ним — медвежонок на искрящемся фоне, в глубине — заснеженный ландшафт. При перелистывании открываются всё новые элементы композиции и создаётся ощущение объёма

64

Двухсторонняя цифровая печать
Упаковка из EnDURO Ice высокой плотности
Трёхслойный «сэндвич» EnDURO Ice
Калька
Плёнка
Калька
Полупрозрачные папки

EnDURO Ice от Sihl — 3-слойный полупрозрачный очень прочный материал, устойчивый к истиранию. Между тончайшей фактурной калькой — прозрачная плёнка. Подходит для офсета и цифры. Из него получаются не только оригинальная и качественная полиграфия, но и разнообразные изделия

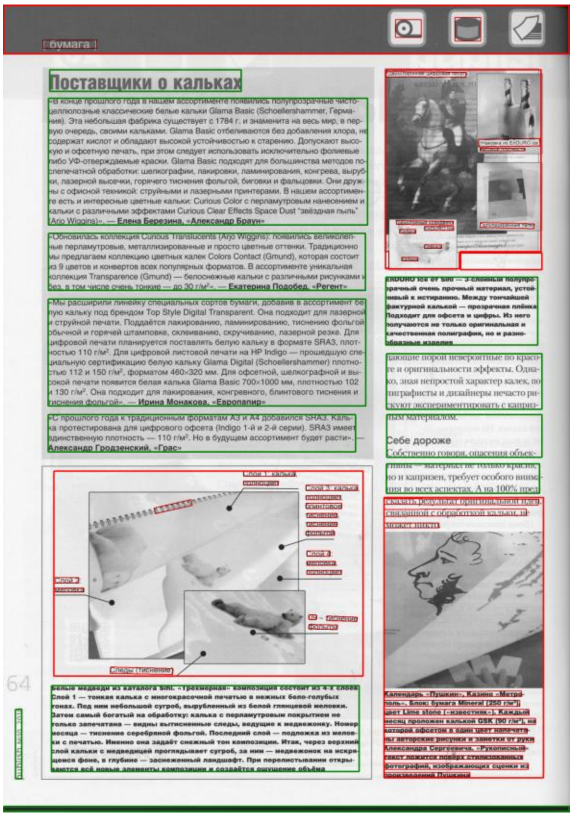
дающие порой невероятные по красоте и оригинальности эффекты. Однако, зная непростой характер калек, полиграфисты и дизайнеры нечасто рискуют экспериментировать с капризным материалом.

Себе дороже

Собственно говоря, опасения объективны — материал не только красив, но и капризен, требует особого внимания во всех аспектах. А на 100% предсказать результат оригинальной идеи, связанной с обработкой кальки, не может никто.

Календарь «Пушкин», Казино «Метрополь». Блок: бумага Mineral (250 г/м²), цвет Lime stone («известняк»). Каждый месяц проложен калькой GSK (90 г/м²), на которой офсетом в один цвет напечатаны авторские рисунки и заметки от руки Александра Сергеевича. «Рукописный» текст ложится поверх стилизованных фотографий, изображающих сценки из произведений Пушкина

Figure 4-13: Segmentation of the page in Figure 4.12 using only layout analysis



(бумага)

Поставщики о кальках

«В конце прошлого года в нашем ассортименте появились полупрозрачные чистоцеллюлозные классические белые кальки Glama Basic (Schoellershammer, Германия). Эта небольшая фабрика существует с 1784 г. и знаменита на весь мир, в первую очередь, своими кальками. Glama Basic отбеливаются без добавления хлора, не содержат кислот и обладают высокой устойчивостью к старению. Допускают высокую и офсетную печать, при этом следует использовать исключительно фолиевые либо УФ-отверждаемые краски. Glama Basic подходят для большинства методов послепечатной обработки: шелкографии, лакировки, ламинирования, конгрева, вырубки, лазерной высечки, горячего тиснения фольгой, биговки и фальцовки. Они дружны с офисной техникой: струйными и лазерными принтерами. В нашем ассортименте есть и интересные цветные кальки: Curious Color с перламутровым нанесением и кальки с различными эффектами Curious Clear Effects Space Dust "звёздная пыль" (Arjo Wiggins)». — **Елена Березина, «Александр Браун»**

«Обновилась коллекция Curious Translucents (Arjo Wiggins): появились великолепные перламутровые, металлизированные и просто цветные оттенки. Традиционно мы предлагаем коллекцию цветных калек Colors Contact (Gmund), которая состоит из 9 цветов и конвертов всех популярных форматов. В ассортименте уникальная коллекция Transparence (Gmund) — белоснежные кальки с различными рисунками и без, в том числе очень тонкие — до 30 г/м²». — **Екатерина Подобед, «Регент»**

«Мы расширили линейку специальных сортов бумаги, добавив в ассортимент белую кальку под брендом Top Style Digital Transparent. Она подходит для лазерной и струйной печати. Поддаётся лакированию, ламинированию, тиснению фольгой обычной и горячей штамповке, склеиванию, скручиванию, лазерной резке. Для цифровой печати планируется поставлять белую кальку в формате SRA3, плотностью 110 г/м². Для цифровой листовой печати на HP Indigo — прошедшую специальную сертификацию белую кальку Glama Digital (Schoellershammer) плотностью 112 и 150 г/м², форматом 460×320 мм. Для офсетной, шелкографной и высокой печати появится белая калька Glama Basic 700×1000 мм, плотностью 102 и 130 г/м². Она подходит для лакирования, конгревного, блинтового тиснения и тиснения фольгой». — **Ирина Монакова, «Европапир»**

«С прошлого года к традиционным форматам A3 и A4 добавился SRA3. Калька протестирована для цифрового офсета (Indigo 1-й и 2-й серии). SRA3 имеет единственную плотность — 110 г/м². Но в будущем ассортимент будет расти». — **Александр Гродзенский, «Грас»**

Белые медведи из каталога Sihl. «Трехмерная» композиция состоит из 4-х слоев. Слой 1 — тонкая калька с многокрасочной печатью в нежных бело-голубых тонах. Под ним небольшой сугроб, вырубленный из белой глянцевой меловки. Затем самый богатый на обработку: калька с перламутровым покрытием не только запечатана — видны вытисненные следы, ведущие к медвежонку. Номер месяца — тиснение серебряной фольгой. Последний слой — подложка из меловки с печатью. Именно она задаёт снежный тон композиции. Итак, через верхний слой кальки с медведицей проглядывает сугроб, за ним — медвежонок на искрящемся фоне, в глубине — заснеженный ландшафт. При перелистывании открываются всё новые элементы композиции и создаётся ощущение объёма

64

EnDURO Ice от Sihl — 3-слойный полупрозрачный очень прочный материал, устойчивый к истиранию. Между тончайшей фактурной калькой — прозрачная плёнка. Подходит для офсета и цифры. Из него получаются не только оригинальная и качественная полиграфия, но и разнообразные изделия

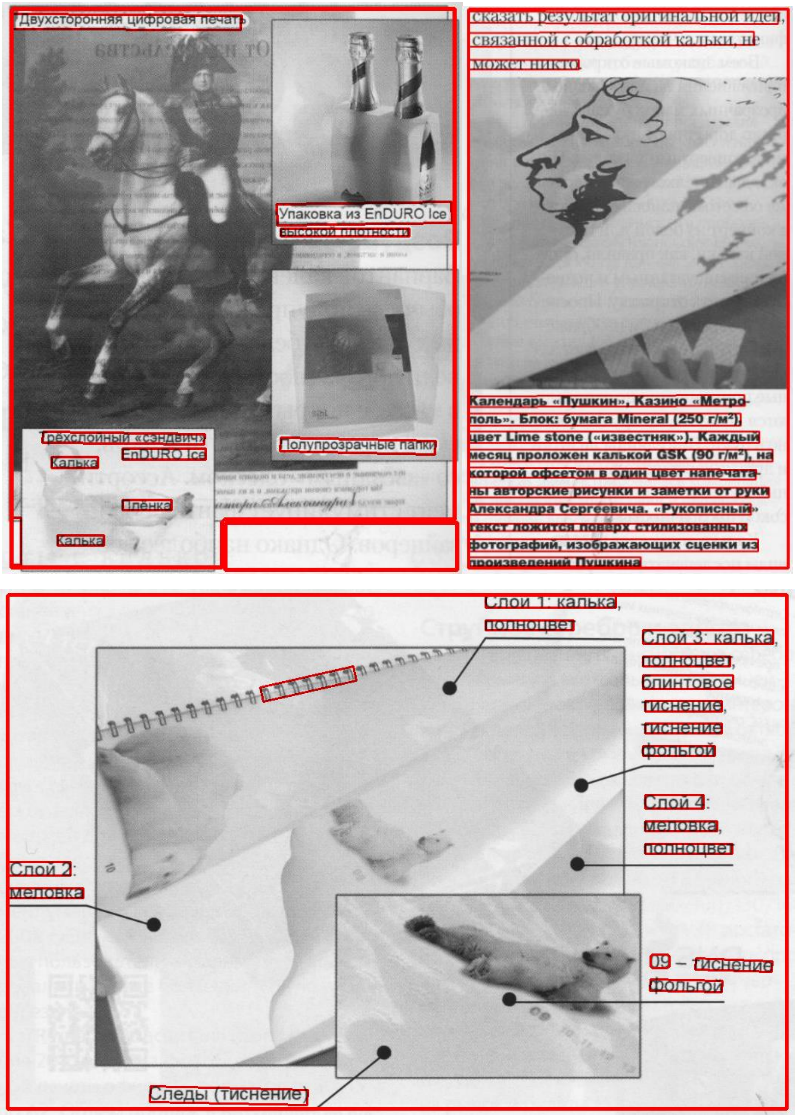
дающие порой невероятные по красоте и оригинальности эффекты. Однако, зная непростой характер калек, полиграфисты и дизайнеры нечасто рискуют экспериментировать с капризным материалом.

Себе дороже

Собственно говоря, опасения объективны — материал не только красив, но и капризен, требует особого внимания во всех аспектах. А на 100% предсказать результат оригинальной идеи, связанной с обработкой кальки, не может никто.

Календарь «Пушкин», Казино «Метрополь». Блок: бумага Mineral (250 г/м²), цвет Lime stone («известняк»). Каждый месяц проложен калькой GSK (90 г/м²), на которой офсетом в один цвет напечатаны авторские рисунки и заметки от руки Александра Сергеевича. «Рукописный» текст ложится поверх стилизованных фотографий, изображающих сценки из произведений Пушкина

Figure 4-14: Segmentation of the page in Figure 4.12 using the proposed technique



сказать результат оригинальной идеи, связанной с обработкой кальки, не может никто.

Календарь «Пушкин», Казино «Метрополь». Блок: бумага Mineral (250 г/м²), цвет Lime stone («известняк»). Каждый месяц проложен калькой GSK (90 г/м²), на которой офсетом в один цвет напечатаны авторские рисунки и заметки от руки Александра Сергеевича. «Рукописный» текст ложится поверх стилизованных фотографий, изображающих сценки из произведений Пушкина

Figure 4-15. Results of text localization applied on the image regions of page in Figure 4.12

CHAPTER 5 - WORD SPOTTING

5.1 Introduction

The word spotting procedure is inspired by speech processing and it was introduced in Document Image Processing in order to facilitate the information retrieval in cases that the perfect results of OCR cannot be achieved. Those can be the cases e.g. of documents that are degraded or they include languages or symbols too rare to warrant to worth OCR training. In the case of historical documents both can happen at the same time.

A classical word spotting methodology can include all or any of the procedures shown in Figure 5.1. In this work, a retrieval system for the Archive of the Government Gazette of the Principality of Samos, a Greek island, ex-autonomous regime under the suzerainty of the Ottoman Empire, is developed. At the General State Archives records (GSA) of Samos lies the complete set of copies of the Government Gazette of the Principality of Samos from the first year of the registration (1894) until the end of the Principality of Samos regime (1912). The Gazette was the official organ of the Administration of the Principality of Samos and therein were published laws, decrees, circulars, court actions and deeds like auctions. Apart from this official part, at that time, were also published the speeches of the liege lords, the minutes of the General Assemblies of Plenipotentiaries (i.e. the local parliament) and short reports on various topics. In total, there is one volume per year (19 volumes) and the amount of pages can vary from 250 to 750 pages per volume. Nowadays, this material is found in the GSA of Samos and this is the only existing full hard copy of this archive. Moreover, there is an already digitalized version that was used in order to build a system that will perform automatic retrieval every time that a Samian citizen wishes to look for something in that archive. The bad quality of the scanned archive (Fig. 5.2) prohibited the use of the common approach of Figure 5.1.

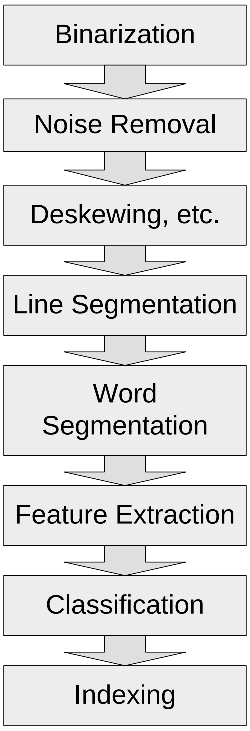


Figure 5-1: The modules of a traditional word-spotting approach

The contribution of the present work consists of:

i. The trial of a simplified word spotting system that is based mainly on picture processing techniques instead of pattern recognition, skipping segmentation and clustering and using the words as compact shapes.

ii. The creation of a small ground truth set of old Greek documents with common problems, available to the scientific community,

iii. A system that will make easier the research and study of the rare Archive of the Government Gazette of the Principality of Samos.

In the Section 5.2, a short summary of the state of the art is given. The proposed methodology is presented in Section 5.3, while in Section 5.4, retrieval results are presented for the above mentioned collection, as well as for a Google book in order to give more objective results. Some comparative results are also given for a traditional system, similar to the one presented in Figure 5.1.

5.2 State of the art

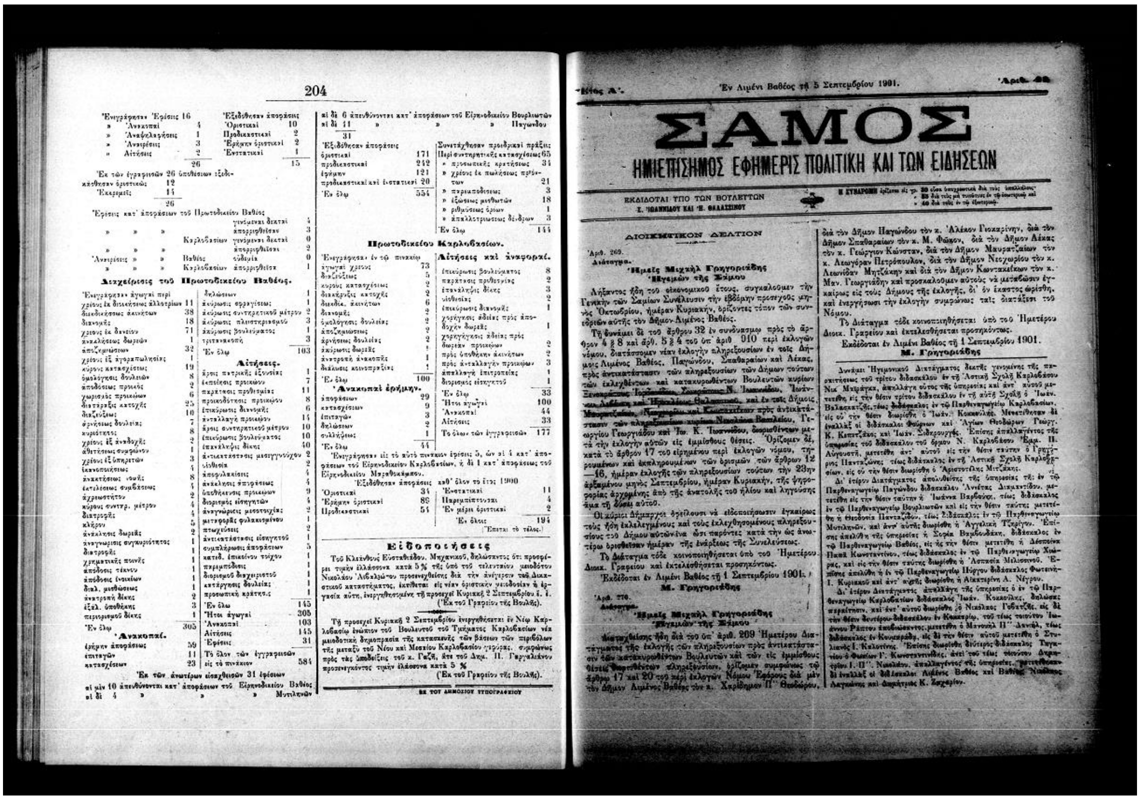
Many nice works have been proposed in the past for historical document retrieval, printed or handwritten, based on the common approach shown in Figure 5.1 or parts of it [83-85]. However, all of the above mentioned works use the segmentation stage, mostly up to word level. In the cases that the quality of the paper, the ink or the scanning is not in a perfectly well condition, the segmentation procedure can reduce the success rate of Word Spotting. Thus, several free-segmentation Word Spotting approaches have also been proposed, lately. Gatos and Pratikakis apply segmentation-free word spotting to printed Historical Documents by localizing salient areas and matching extracted features for several skews and scales [86]. Farrahi Moghaddam and Cheriet present, at the same conference, another line and word segmentation-free methodology, that is based on connected component feature extraction, DTW and Euclidean Distance [87]. Leydier et al. also present a segmentation free word retrieval technique using zones of interest and guides on which they perform cohesive matching [88]. Moreover, they allow the synthesis of the query. Rusiňol et al. present a segmentation free word spotting technique, appropriate for heterogeneous document collections, using feature extraction on patch level [89]. Zagoris et al. apply a MPEG-like descriptor containing conventional contour and region shape features [90].

5.3 The proposed system

The above mentioned archive presents some special characteristics due to its scanning that it took place several years ago by non-specialists:

- the resolution is low, just 200 dpi,
- the pages, newspaper size, are scanned two at the time (Fig. 5.2),
- unevenly lighted image (Fig. 5.2),
- lightly skewed part of the image (Fig. 5.2),
- old printing of bad quality (Fig. 5.3),
- the language in use (Fig. 5.3) is an older version of Greek with a lot of accents, that are not used at the moment and it is difficult to find appropriate OCR software.

The bad quality of printing and scanning (Fig. 5.3) proved to be a very problematic situation during the application of the known techniques, even after trying to improve them. The low performance of each task, due to the special problems, was accumulated to the whole giving a lower final result. The necessity to keep the system as simple as possible with a minimum number of modules was soon realized.



204

Ἔτος Α'. Ἐν Λιμένι Βαθέος τῇ 5 Σεπτεμβρίου 1901. Ἀριθ. 48

ΣΑΜΟΣ

ΗΜΙΕΠΙΣΗΜΟΣ ΕΦΗΜΕΡΙΣ ΠΟΛΙΤΙΚΗ ΚΑΙ ΤΩΝ ΕΙΔΗΣΕΩΝ

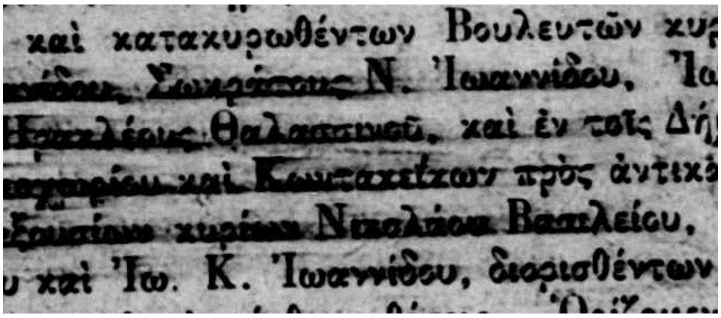
ΕΚΔΙΔΟΤΑΙ ΥΠΟ ΤΩΝ ΒΟΥΛΕΥΤΩΝ
Σ. ΙΩΑΝΝΙΔΟΥ ΚΑΙ Η. ΘΑΛΑΣΣΙΝΟΥ

ΔΙΟΙΚΗΤΙΚΟΝ ΔΕΛΤΙΟΝ

Ἀριθ. 269.
Διάταγμα.
Ἡμεῖς Μιχαὴλ Γρηγοριάδης
Ἡγεμὼν τῆς Σάμου

Λήξαντος ἤδη τοῦ οἰκονομικοῦ ἔτους, συγκαλοῦμεν τὴν Γενικὴν τῶν Σαμίων Συνέλευσιν τὴν ἑβδόμην προσεχοῦς μηνὸς Ὀκτωβρίου, ἡμέραν Κυριακήν, ὁρίζοντες τόπον τῶν συνεδριῶν αὐτῆς τὸν Δῆμον Λιμένος Βαθέος.

Figure 5-2: Page from the Government Gazette of the Principality of Samos



καὶ κατακυρωθέντων Βουλευτῶν χυ
[illegible] Σωκράτους Ν. Ἰωαννίδου, Ἰω
[illegible] Θαλασσινοῦ, καὶ ἐν τοῖς Δή
[illegible] καὶ Κονταxαίïκων πρὸς ἀντικα
[illegible] Νικολάου Βασιλείου,
υ καὶ Ἰω. Κ. Ἰωαννίδου, διορισθέντων

Figure 5-3: Detail from the archive

The proposed system appears in the Figure 5.4. It consists of simple procedures of image processing. First, adaptive thresholding is applied to the image using as threshold the mean of the 9x9 neighborhood. Figure 5.5 shows the result of adaptive thresholding of the image in Figure 5.2, compared with the result of a typical binarization technique, Otsu's thresholding [43]. Next, the main body of the text, although no segmentation is performed, is estimated by the query, using the technique mentioned in Kavallieratou et al. [91]. Unfortunately, this limits the retrieval in the words that fit the size of the query.

Instead of extracting feature vectors from the query and the document images, the whole query is kept and the document image is scanned to find the specific query, without applying any segmentation. In order to get rid of unnecessary details and smooth small differences in skew and scale, the query is transformed into more compact shape by normalization. The normalization consists of applying opening to the image with an elliptical structuring element of the size of [word main body] x [word main body x 0.5]. The same normalization is applied to all document images. This normalization that takes about 5 secs / image 3300x4500 pixels (Fig. 5.2), can be applied once to all document images and be kept stored for the future queries. Several examples of the normalization procedure are shown in Figure 5.6.

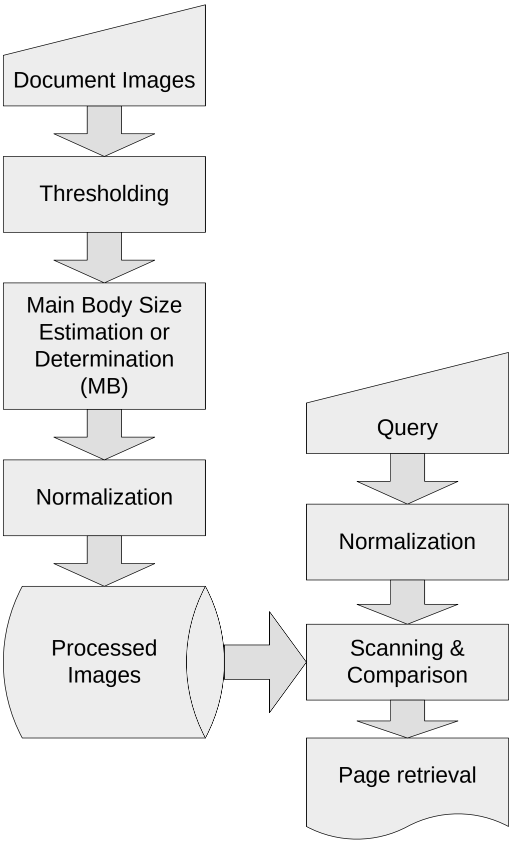


Figure 5-4: The proposed word spotting approach

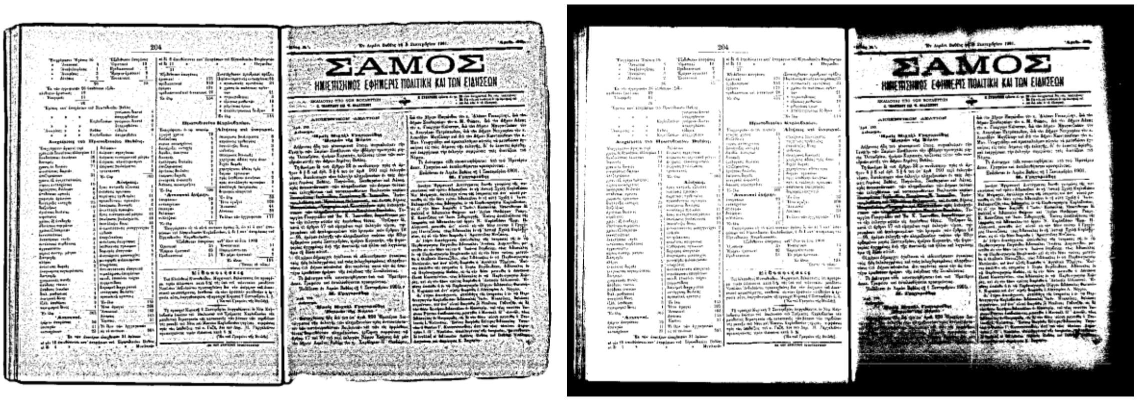


Figure 5-5: Adaptive thresholding (left) vs. Otsu’s thresholding (right)

Each query could be selected by the user or uploaded by an image file. Although synthesis was also considered, as it is described in Leydier et al. [88], the results were worse due to the bad quality of printing that results no standard relative position, in vertical direction and, the characters and the accents, that are not all present in the modern Greek. After the normalization, the query image is applied to every pixel of the image (left-upper corner) and is compared with the corresponding part of the image, using the Sum of Squared Differences (SSD) matching algorithm:

$$h[m,n] = \sum_{k,l} (q[k,l] - I[m+k, n+l])^2 \ ,$$

where I is the page image and q the query image.

In Sum of Squared Differences (SSD), the differences are squared and aggregated. Finally, if the similarity is large enough, the corresponding page is retrieved.

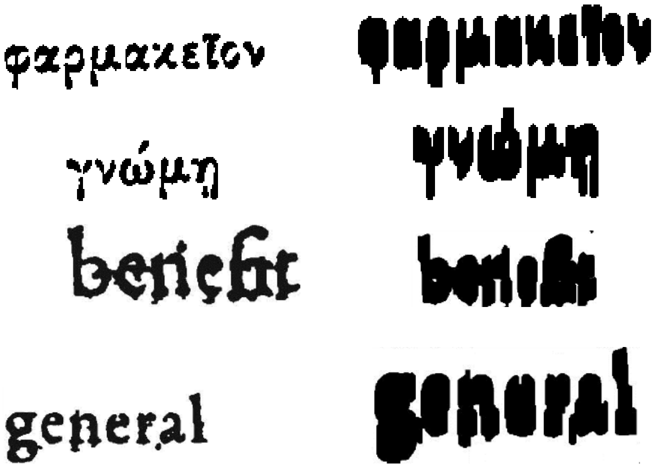


Figure 5-6: Examples of the normalization procedure

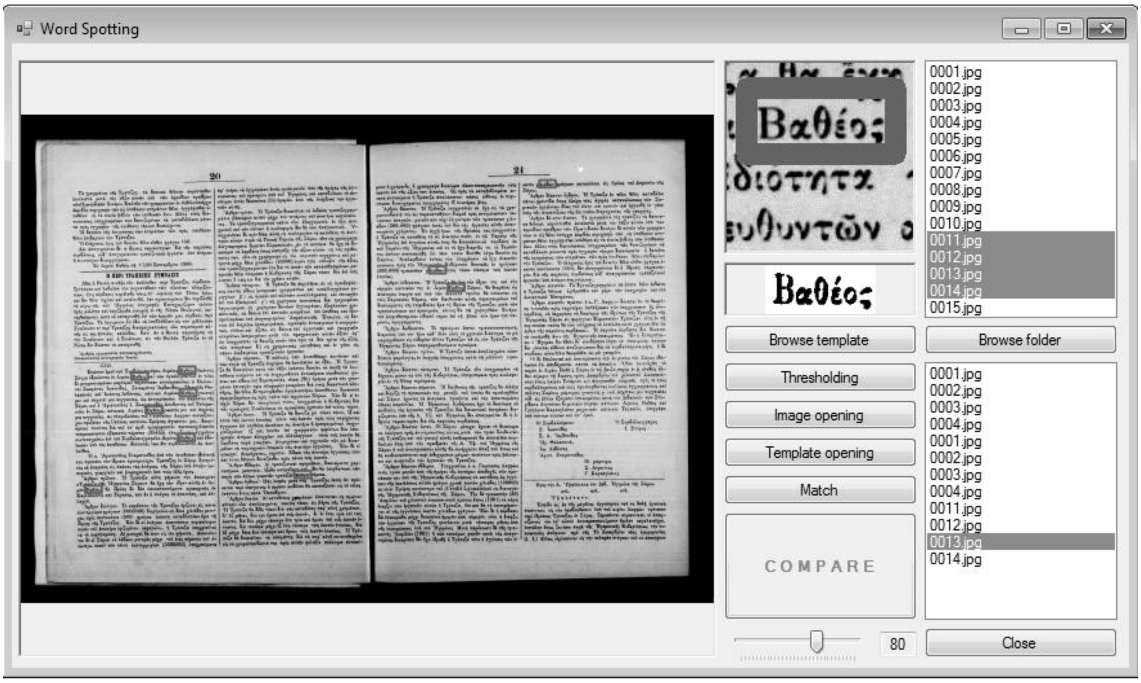


Figure 5-7: A screenshot of the implemented system

5.4 Experimental results

Visual C# and OpenCV were used to implement the thresholding, the opening filter and the SSD matching algorithm. The system shown in Figure 5.7 can look for a query (image of text), selected by the user or uploaded by file, in a collection or the selected images of it (upper list). The modules of the methodology, described above, can be applied separately or all together (compare button). The similarity accuracy can be selected by the user (slider Fig. 5.7) and the retrieved images will be presented in the lower list.

In order to perform experiments on the proposed system, ground truth results were extracted from 15 scanned images of the Greek archive, that is 30 document pages, by human reader, for 10 queries. Words of different sizes (in characters) were included. In Table 5.1, the queries are shown, next to an explanatory note and the amount of times they occur in the ground truth sample.

Moreover, a google book, *A sermon preach'd before the king* [92], was also used in order to extract more objective results with OCR text provided by Google. The book consists of 44 pages and it was published in 1675. In this case we used the book binarized images as provided by Google. A sample image is shown in Figure 5.8. The book, although carefully binarized, includes a lot of noise and an older version of alphabet symbols (Fig. 5.9).

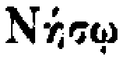
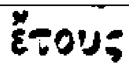
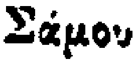
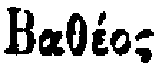
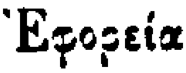
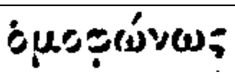
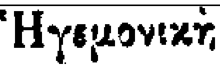
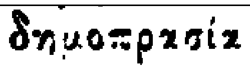
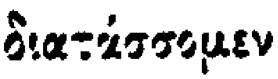
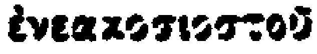
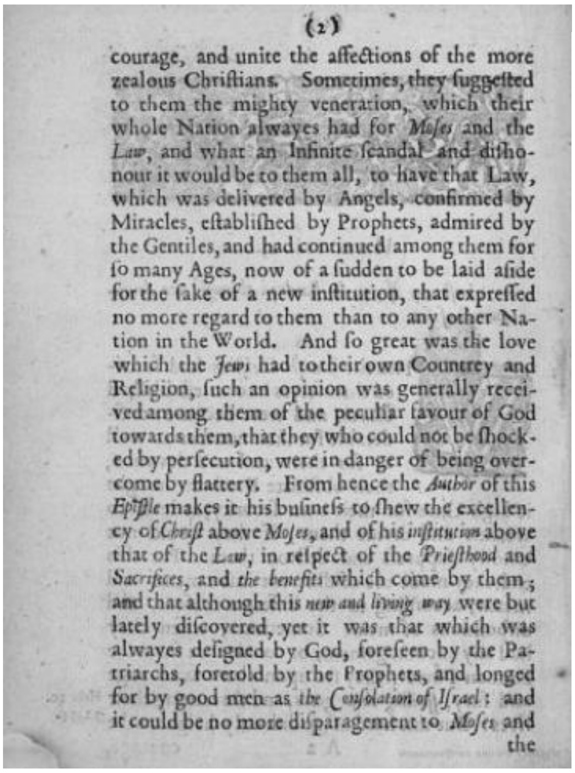
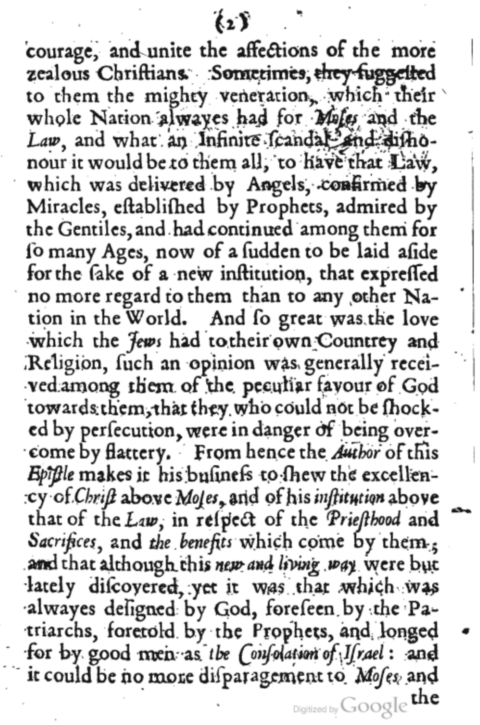
Queries	**Translation**/note	Occurrences
 Νήσῳ	**Island**, in dative	10
 ἔτους	**Year** (gen), often met in dates	33
 Σάμου	the **name of the island** (gen)	58
 Βαθέος	the **name of the capital** (gen)	73
 Ἐφορεία	**Tax office**	5
 ὁμοφώνως	**Unanimously** often met in decisions	16
 Ἡγεμονική	**Hegemonic**, reference to the island	10
 δημοπρασία	**auction**, often published	18
 διατάσσομεν	**we order**, often met in decisions	21
 ἐνεακοσιοστοῦ	**900th**, often met in dates	21

Table 5-1: Greek archive's queries, translation and occurrences in the ground truth sample



(2)

courage, and unite the affections of the more zealous Chriſtians. Sometimes, they ſuggeſted to them the mighty veneration, which their whole Nation alwayes had for *Moſes* and the *Law*, and what an Infinite ſcandal and diſhonour it would be to them all, to have that Law, which was delivered by Angels, confirmed by Miracles, eſtabliſhed by Prophets, admired by the Gentiles, and had continued among them for ſo many Ages, now of a ſudden to be laid aſide for the ſake of a new inſtitution, that expreſſed no more regard to them than to any other Nation in the World. And ſo great was the love which the *Jews* had to their own Countrey and Religion, ſuch an opinion was generally received among them of the peculiar favour of God towards them, that they who could not be ſhocked by perſecution, were in danger of being overcome by flattery. From hence the *Author* of this *Epiſtle* makes it his buſineſs to ſhew the excellency of *Chriſt* above *Moſes*, and of his *inſtitution* above that of the *Law*, in reſpect of the *Prieſthood* and *Sacrifices*, and *the benefits* which come by them; and that although this *new and living way* were but lately diſcovered, yet it was that which was alwayes deſigned by God, foreſeen by the Patriarchs, foretold by the Prophets, and longed for by good men as *the Conſolation of Iſrael*: and it could be no more diſparagement to *Moſes* and the



(2)

courage, and unite the affections of the more zealous Chriſtians. Sometimes, they ſuggeſted to them the mighty veneration, which their whole Nation alwayes had for *Moſes* and the *Law*, and what an Infinite ſcandal and diſhonour it would be to them all, to have that Law, which was delivered by Angels, confirmed by Miracles, eſtabliſhed by Prophets, admired by the Gentiles, and had continued among them for ſo many Ages, now of a ſudden to be laid aſide for the ſake of a new inſtitution, that expreſſed no more regard to them than to any other Nation in the World. And ſo great was the love which the *Jews* had to their own Countrey and Religion, ſuch an opinion was generally received among them of the peculiar favour of God towards them, that they who could not be ſhocked by perſecution, were in danger of being overcome by flattery. From hence the *Author* of this *Epiſtle* makes it his buſineſs to ſhew the excellency of *Chriſt* above *Moſes*, and of his *inſtitution* above that of the *Law*, in reſpect of the *Prieſthood* and *Sacrifices*, and *the benefits* which come by them; and that although this *new and living way* were but lately diſcovered, yet it was that which was alwayes deſigned by God, foreſeen by the Patriarchs, foretold by the Prophets, and longed for by good men as *the Conſolation of Iſrael*: and it could be no more diſparagement to *Moſes* and the

Figure 5-8: Sample page in original and binary from Google books

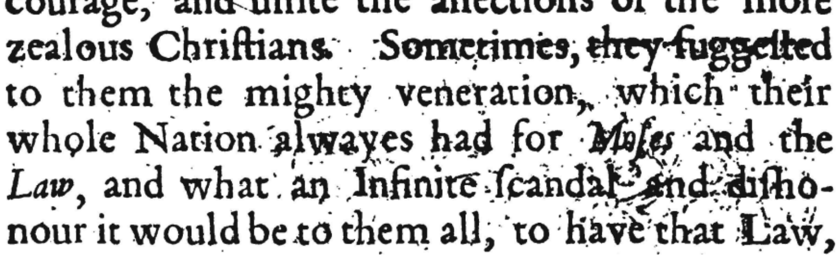
In order to evaluate the proposed system, precision (1), recall (2) and F measure (3) were used:

$$precision = \frac{correctly \quad retrieved \quad words}{total \quad retrieved \quad words} \tag{1}$$

$$recall = \frac{correctly \quad retrieved \quad words}{existed \quad words} \tag{2}$$

$$F = \frac{2 * recall * precision}{recall + precision} \tag{3}$$

The results for the Greek documents are presented in Table 5.2, while the comparative results for the Google book can be seen in Table 5.3. In the case of Google book, it should be mentioned, that from the list of occurrences in Google typed text, the occurrences in italics have been excluded, while the occurrences that include the query have been added e.g. teachers for teacher, etc.



courage, and unite the affections of the more zealous Chriſtians. Sometimes, they ſuggeſted to them the mighty veneration, which their whole Nation alwayes had for *Moſes* and the *Law*, and what an Infinite ſcandal and diſhonour it would be to them all, to have that Law,

Figure 5-9: Detail from the Google book

Since Greek is an inflectional language, for each noun several similar word forms can be found (dative, genitive, accusative, etc.). This is obvious in the results like Νήσῳ (dat.), Σάμου (gen), ηγεμονική (acc.), etc., however they were not considered correct, although it could be useful in many cases. This could be one of the reasons that the results in English text are higher, plus the fact that the Google documents are scanned in 600 dpi while the resolution of the Greek documents is just 200 dpi.

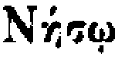
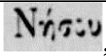
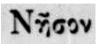
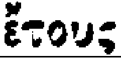
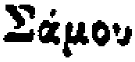
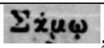
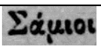
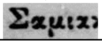
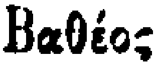
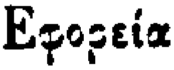
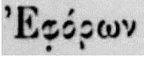
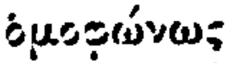
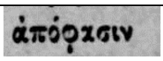
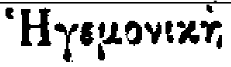
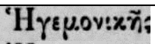
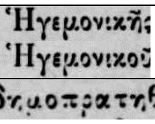
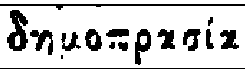
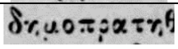
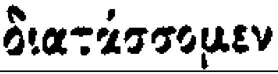
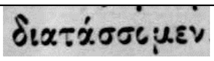
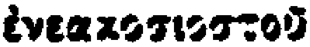
Queries	CPU time/ page (sec)	Similarity > 85%			False Positives
		Precis.%	Rec.%	F meas.%	
 Νήσῳ	4.73	22.5	90	36	 Νήσου,  Νῆσον
 ἔτους	2.59	100	62.85	77.19	-
 Σάμου	11.23	85.71	41.37	55.81	 Σάμῳ,  Σάμιοι  Σαμιαι
 Βαθέος	5.04	100	30.13	46.31	-
 Ἐφορεία	13.88	83.33	100	90.90	 Ἐφόρων
 ὁμοφώνως	15.06	94.44	100	97.14	 ἀπόφασιν
 Ἡγεμονική	126.98	18.03	100	30.55	 Ἡγεμονικῆς  Ἡγεμονικοῦ
 δημοπρασία	11.57	83.33	83.33	83.33	 δημοπρατηθ
 διατάσσομεν	11.90	94.73	100	97.29	 διατάσσομεν
 ἐνεακοσιοστοῦ	9.74	100	28.57	44.44	-

Table 5-2: Experimental results for Greek documents

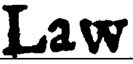
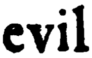
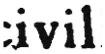
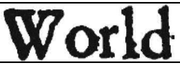
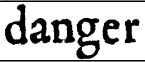
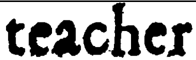
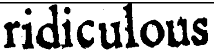
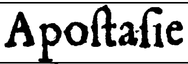
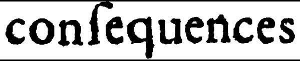
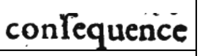
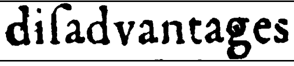
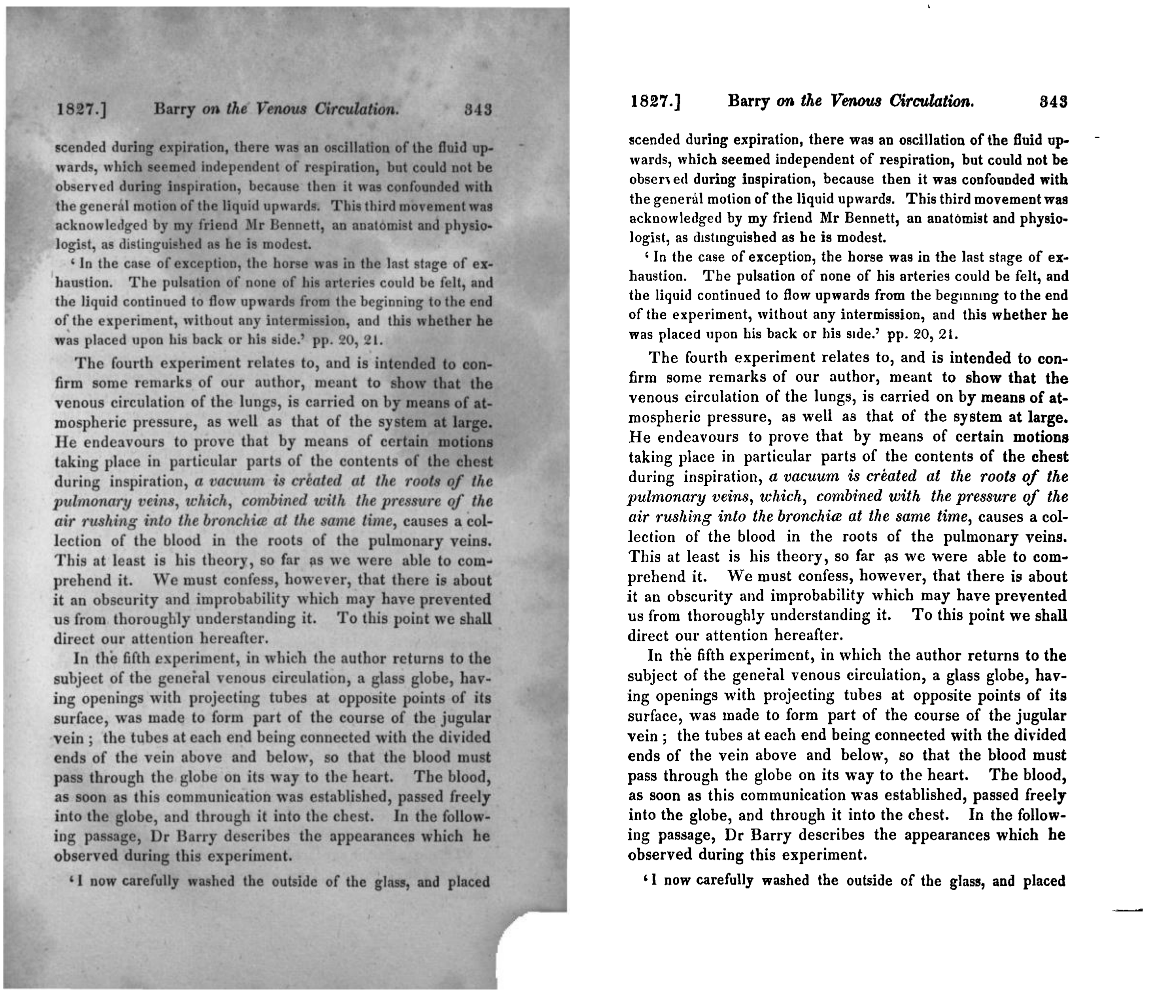
Queries	CPU time/per page (sec)	Similarity > 85%			Google OCR (occur.)	False positives
		Precision	Recall	F1		
 Law	0.12	100	66.67	80	3	-
 evil	5.56	88.89	100	94.11	8	 :ivil
 World	9.81	100	66.67	80	24	-
 danger	5.60	100	90	94.73	10	-
 teacher	6.17	100	100	100	2	-
 ridiculous	6.50	100	66.67	80	3	-
 Apoſtaſie	7.90	100	66.67	80	3	-
 conſequences	8.08	50	100	66.67	1	 conſequence
 diſadvantages	6.72	100	100	100	2	-

Table 5-3: Experimental results for Google book

The computational cost shown in Tables 5.2 & 5.3 under CPU time/per page (sec) concerns the system implemented as mentioned in Section 5.3. In the first experiments in Matlab, the computational cost could reach double or triple depending on the query size. In those cases an extra trick was introduced in order to reduce the computational cost. The images were scanned every 2-3 pixels, instead of every pixel, depending on the image resolution, in order to reduce the computational cost. However, this were not considered necessary in the implemented system (Fig. 5.7) since the computational cost is much lower. This solution is kept in mind for larger collections.

Finally, in order to give comparative results with a traditional word-spotting system, the system described in Doulgeri and Kavallieratou [93] was used. This system is very similar to the traditional ones [83-85], since it includes more of the stages described in Figure 5.1. For the experiments, the parameters, as they were proved better in the paper, were used for that system. That is, synthesized words in 300 dpi, bold Times New Roman, interpolation of 175 points and smoothing of 5 points. Since the mentioned, in that chapter, books were not available and the application of the traditional system to the ones mentioned here was impossible due to failure in segmentation, ten pages were used from the Google Book entitled *The Medico-chirurgical Review and Journal of Medical Science* that was published in London in 1826 (Fig. 5.10). Four samples were selected from the ones synthesized for the paper, and the pages were selected in order to have at least two occurrences for each of them. Both systems were applied, as they are described at the corresponding papers. The results are presented in Table 5.4.



1827.] Barry *on the Venous Circulation.* 343

scended during expiration, there was an oscillation of the fluid upwards, which seemed independent of respiration, but could not be observed during inspiration, because then it was confounded with the generàl motion of the liquid upwards. This third movement was acknowledged by my friend Mr Bennett, an anatómist and physiologist, as distinguished as he is modest.

'In the case of exception, the horse was in the last stage of exhaustion. The pulsation of none of his arteries could be felt, and the liquid continued to flow upwards from the beginning to the end of the experiment, without any intermission, and this whether he was placed upon his back or his side.' pp. 20, 21.

The fourth experiment relates to, and is intended to confirm some remarks of our author, meant to show that the venous circulation of the lungs, is carried on by means of atmospheric pressure, as well as that of the system at large. He endeavours to prove that by means of certain motions taking place in particular parts of the contents of the chest during inspiration, *a vacuum is créated at the roots of the pulmonary veins, which, combined with the pressure of the air rushing into the bronchiæ at the same time*, causes a collection of the blood in the roots of the pulmonary veins. This at least is his theory, so far as we were able to comprehend it. We must confess, however, that there is about it an obscurity and improbability which may have prevented us from thoroughly understanding it. To this point we shall direct our attention hereafter.

In the fifth experiment, in which the author returns to the subject of the genei̇al venous circulation, a glass globe, having openings with projecting tubes at opposite points of its surface, was made to form part of the course of the jugular vein; the tubes at each end being connected with the divided ends of the vein above and below, so that the blood must pass through the globe on its way to the heart. The blood, as soon as this communication was established, passed freely into the globe, and through it into the chest. In the following passage, Dr Barry describes the appearances which he observed during this experiment.

'I now carefully washed the outside of the glass, and placed

1827.] Barry *on the Venous Circulation.* 343

scended during expiration, there was an oscillation of the fluid upwards, which seemed independent of respiration, but could not be observed during inspiration, because then it was confounded with the generàl motion of the liquid upwards. This third movement was acknowledged by my friend Mr Bennett, an anatómist and physiologist, as distinguished as he is modest.

'In the case of exception, the horse was in the last stage of exhaustion. The pulsation of none of his arteries could be felt, and the liquid continued to flow upwards from the beginning to the end of the experiment, without any intermission, and this whether he was placed upon his back or his side.' pp. 20, 21.

The fourth experiment relates to, and is intended to confirm some remarks of our author, meant to show that the venous circulation of the lungs, is carried on by means of atmospheric pressure, as well as that of the system at large. He endeavours to prove that by means of certain motions taking place in particular parts of the contents of the chest during inspiration, *a vacuum is créated at the roots of the pulmonary veins, which, combined with the pressure of the air rushing into the bronchiæ at the same time*, causes a collection of the blood in the roots of the pulmonary veins. This at least is his theory, so far as we were able to comprehend it. We must confess, however, that there is about it an obscurity and improbability which may have prevented us from thoroughly understanding it. To this point we shall direct our attention hereafter.

In the fifth experiment, in which the author returns to the subject of the genei̇al venous circulation, a glass globe, having openings with projecting tubes at opposite points of its surface, was made to form part of the course of the jugular vein; the tubes at each end being connected with the divided ends of the vein above and below, so that the blood must pass through the globe on its way to the heart. The blood, as soon as this communication was established, passed freely into the globe, and through it into the chest. In the following passage, Dr Barry describes the appearances which he observed during this experiment.

'I now carefully washed the outside of the glass, and placed

Figure 5-10: Page in original from Google books and binary from Doulgeri [93]

Queries	Occurr.	Proposed System		Doulgeri	
		true positive	false positive	true positive	false positive
Close	3	2	1	3	1
Difficult	2	2	2	2	0
English	3	3	0	2	2
Habit	2	1	0	2	0

Table 5-4: Results on the Google book, for the proposed and the Doulgeri [93] systems

CHAPTER 6 - CHARACTER SPOTTING

6.1 Introduction

The document image retrieval is an easy case when the recognition problem is solved for the document e.g. successful OCR. However, this is not the case of handwritten document images, especially when it concerns historical document images. As described in the previous chapter, word-spotting is a proposed solution for such cases, where template based methods match a query word image with labeled keyword template images. Since this require the existence of the word images, more recent systems of word-spotting proposed techniques of using text queries.

In this chapter, a novel procedure is presented for text retrieval. It supports text queries while it can be used for any language or set of characters by easy training. The work is fully inspired by the word-spotting technique described in Chapter 5. In that work, queries by example are fully matched to the document images, after normalization. This technique was applied to a collection of bad quality historical printed documents with better results than the existed techniques. However, the results were not equally good for handwritten documents due to the variety in writing style. The entire shape of word can be affected dramatically by a little wider or slanted character and the simulation is not obvious anymore.

In this work, the word image is broken in characters and overlaps as well as spaces between them are allowed making the procedure much more robust. The proposed system requires text queries instead of image queries.

In the Section 6.2, a short summary of the state of the art is given. Next, in Section 6.3, the proposed system is presented, while in Section 6.4 some initial results are given on a common database for word-spotting, the letters of George Washington.

6.2 State of the art

Manmatha et al. first proposed two techniques for matching words [94]. Other features based on global image characteristics have been proposed. Zhang et al. presented an effective and efficient approach for word image matching by using gradient-based binary features [95] while Rothfeder et al. presented an algorithm that matches word images by recovering correspondences between image corners, which have been identified by the Harris detector [96] (Fig. 6.1). A set of biologically inspired features formed by a cascade of Gabor descriptors was proposed by van der Zant and Schomaker [97]. Srihari and Ball used global word shape features as the similarity measure between the query and the candidate words [98].

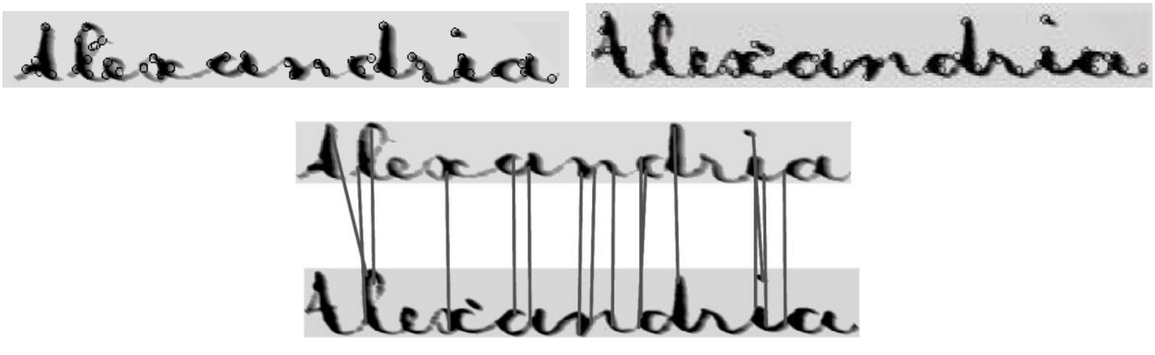


Figure 6-1: Harris corners (up) and recovered correspondences (down) [96]

Bhardwaj et al. presented a system that accepts a query in the form of text from the user, extracts moment based features from word images, stores them as index and finally uses a cosine similarity metric in order to return the word images that are most similar to the query [99]. Leydier et al. proposed another word-retrieval technique that allows the search of words entered by the user based on differential features that are compared using a cohesive elastic matching method and on zones of interest in order to match only informative parts of the words [88,100].

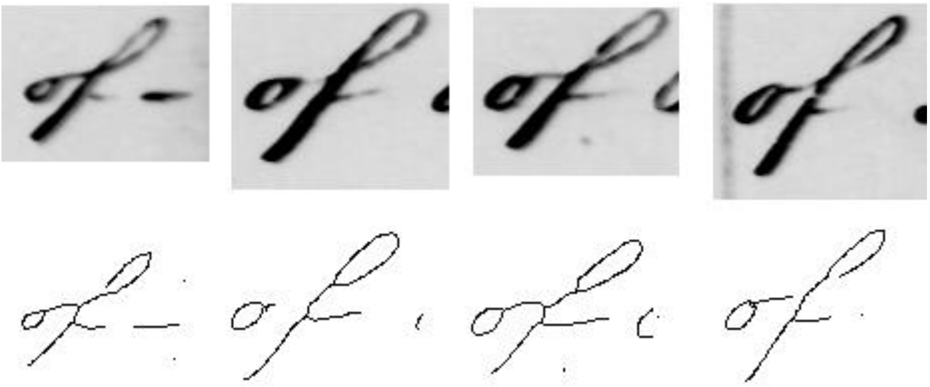


Figure 6-2: Skeleton examples for different instances of the same word class [105]

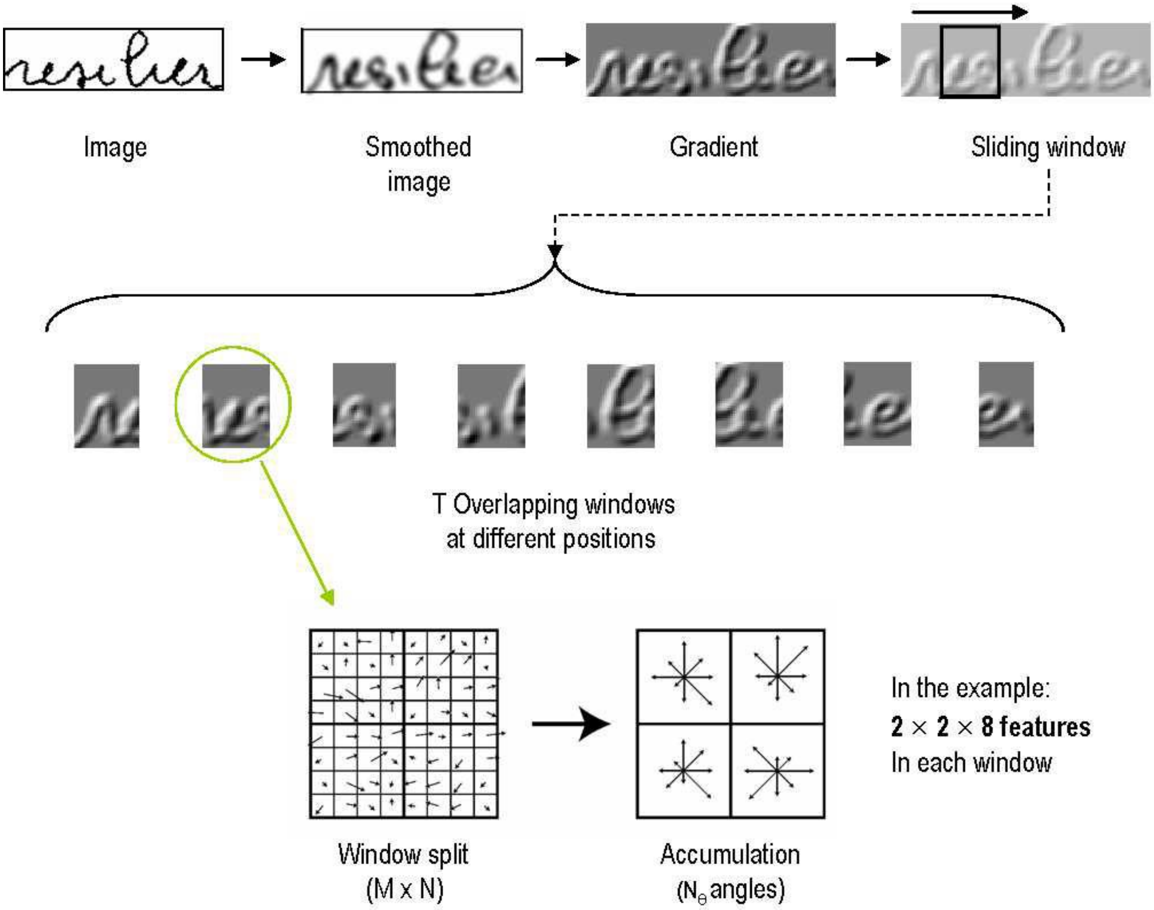


Figure 6-3: The feature extraction process of Rodriguez and Peronnin [106]

Kolcz et al. proposed the use of the Dynamic Time Warping (DTW) method (often used in speech analysis) for nonlinear sequence alignment [101]. Rath and Manmatha presented an algorithm that applies DTW to compare sets of 1-dimensional features created by segmented word images [102]. Adamek et al. proposed DTW to align convexity and concavity features extracted from contours [103]. Terasawa and Kanata proposed an extension of DTW that allows matching keyword templates with complete text lines by using a gradient distribution based feature with overlapping normalization [104]. Wang et al. presented an approach based on skeleton graph representation (Fig. 6.2) of the word images combined with DTW as the similarity measure [105].

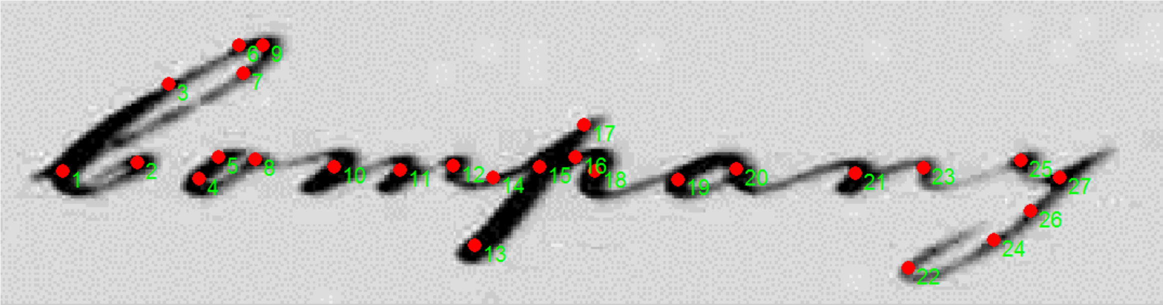


Figure 6-4: Keypoints found in a query image using the detector for SIFT [107]

Many authors from the document analysis field apply keypoint matching techniques to the problem of keyword spotting. Rodriguez and Peronnin presented a method based on local gradient histogram features (Fig. 6.3) inspired by the SIFT keypoint descriptor [106]. Zhang and Tan proposed a segmentation-free method that applies the keypoint detector for SIFT (Fig. 6.4) to locate keypoints on the document pages and the query image, then extracts Heat Kernel Signature descriptors from a local patch centered at each keypoint and finally finds local zones which contain enough matching keypoints corresponding to the query image [107].

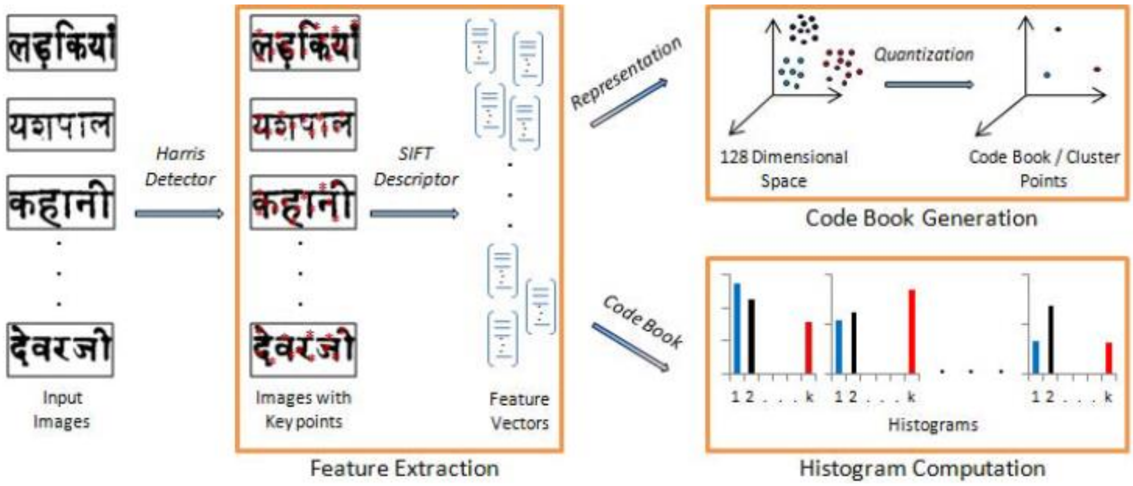


Figure 6-5: Bag of Visual Words representation by Shekhar and Jawahar [108]

In order to avoid exhaustively matching all the keypoints among them, the classic bag-of-words paradigm from the information retrieval field was reformulated as the Bag-of-Visual-Words (BoVW). Rusiňol et al. proposed a patch-based framework where patches are represented by a bag-of-visual-words model powered by SIFT descriptors [89]. Shekhar and Jawahar also presented a BoVW based approach (Fig. 6.5) that uses SIFT descriptors at interest points as feature vectors [108]. Aldavert et al. presented a query-by-string method that combines textual representation of the word images, formulated in terms of character n-grams and visual representation based on the BoVW scheme [109].

Learning-based methods employ statistical learning to train a keyword model that is then used to score query images. Choisy presented a keyword spotting system based on the NSHP-HMM that allows to dynamically create global word models from letters models and do not require any writing segmentation [110]. Perronnin and Serrano proposed the Fischer kernel framework [111] as well as a statistical framework which employs hidden Markov models (HMMs) to model keywords and a Gaussian mixture model (GMM) for score normalization [112] and one more method based on semi-continuous HMMs (SC-HMMs) [113]. Rothacker et al. proposed a segmentation-free framework based on Bag-of-Features HMMs that use statistics of local image feature representatives [114].

When the learning-based approach is applied at character level, a word spotting system obtains the capability to spot arbitrary keywords by concatenating the character models appropriately. Edwards et al. proposed a statistical model based on a generalized HMM that uses only one instance of each letter in the manuscript for training [115]. Chan et al. also presented a character-level segmentation-based approach that utilizes gHMMs with a bigram letter transition model [116]. Thomas et al. considered entire text lines as an indivisible entity and modeled them with Hidden Markov Models [117]. Fischer et al. presented a lexicon-free word spotting system based on character Hidden Markov Models where arbitrary keywords can be spotted without pre-segmenting text lines into words [118].

Hidden Markov Models are the most widely used techniques to model the keywords' sequential features although other machine learning approaches such as Neural Networks have also been used in the keyword spotting domain. Frinken et al. presented two systems based on a modification of the CTC Token Passing algorithm in conjunction with bidirectional long short-term memory neural networks (BLSTM-NNs) [119] and a recurrent neural network [120].

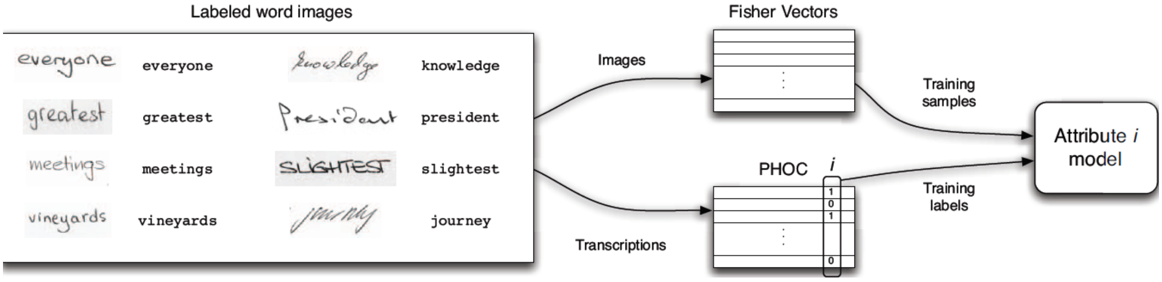


Figure 6-6: The training process for the method proposed by Almazan et al. [121]

Almazan et al. proposed a method where character attributes are used to learn a semantic representation of the word images and then a calibration of the scores with Canonical Correlation Analysis (CCA) is performed that puts images and text strings in a common subspace [121]. Figure 6.6 shows the training process for the i-th attribute model. A classifier is trained using the Fischer vector (FV) representation of the images and the i-th value of the pyramidal histogram of characters (PHOC) representation as label. Howe presented a technique where a flexible inkball generative model for word appearance, derived from the query image, allows for Gaussian random-walk deformation of the ink trace in two dimensions and fitting the query models to the target page images to find locations where there is a good (low-deformation) match [122]. Results from the above two techniques, as well as from the methods presented in [88,100] are reported in the ICFHR 2014 Competition on Handwritten KeyWord Spotting paper [123].

6.3 The proposed system

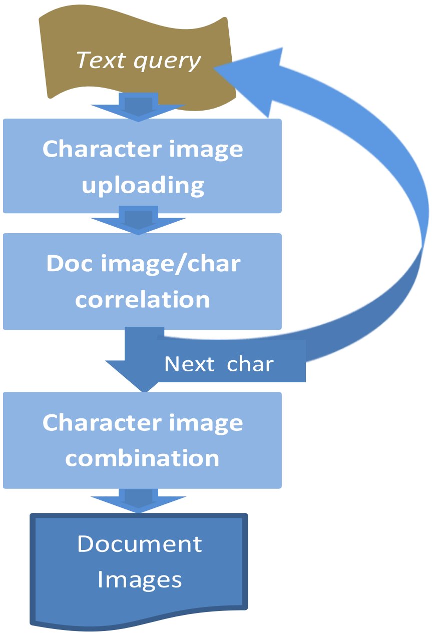

Text query
Character image uploading
Doc image/char correlation
Next char
Character image combination
Document Images

Figure 6-7: The proposed retrieval procedure

As already mentioned, the idea is fully inspired by the word-spotting procedure, described in Chapter 5. In that chapter, word-spotting was performed by matching patterns of images. The technique was applied to historical documents of printed text, considering the similarity between the normalized example-query and the corresponding area that started on every pixel of the normalized document images. One disadvantage was the query-by-example; another was the use on handwritten documents due to the big variety of writing style.

In Figure 6.7 the proposed procedure is presented. Next, the training procedures of the proposed system, as well as the rest modules are described in detail.

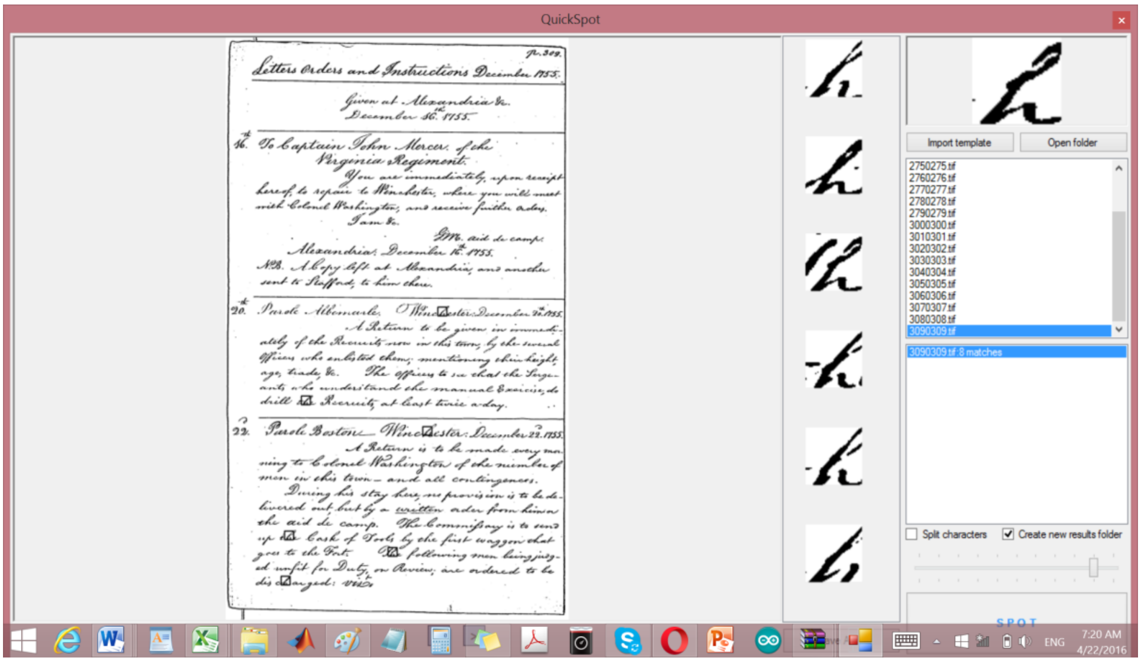
6.3.1 Training procedure

The system described in Chapter 5 was used for the training of the proposed technique, after several modifications:

- The system was spotting by example, queries of characters selected by the user.
- This time only the black pixels were considered to count similarity. Accepted were considered the areas with similarity over 70%.
- The system was modified to accept several examples in the same query, in order to include all the different styles of a character.
- The results of the query were saved altogether in bmp form, named after the symbol given by the user e.g. *h*, and an increasing number e.g. *h2.bmp*, *h3.bmp*, etc.
- The user can choose the results he wishes to keep.
- The user can clean up manually the noisy strokes e.g. of overlapped characters.

In Figure 6.8, the system for training after the modification is shown, while in Figure 6.9, examples of characters are listed.

At this stage, similarity over 70% was also taken into account, in order to keep the most similar areas.



p. 309.

Letters Orders and Instructions December 1755.

Given at Alexandria &c.
December 16th 1755.

16th. To Captain John Mercer, of the
Virginia Regiment.

You are immediately, upon receipt hereof, to repair to Winchester, where you will meet with Colonel Washington; and receive further orders.

I am &c.

G.W. aid de camp.

Alexandria: December 16th 1755.

N.B. A Copy left at Alexandria; and another sent to Stafford, to him there.

20th. Parole Albemarle. Winchester: December 20th 1755.

A Return to be given in immediately of the Recruits now in this town, by the several Officers who enlisted them; mentioning their height, age, trade, &c. The officers to see that the Sergeants who understand the manual Exercise, do drill the Recruits at least twice a-day.

22d. Parole Boston — Winchester: December 22d 1755.

A Return is to be made every morning to Colonel Washington of the number of men in this town — and all contingencies.

During his stay here, no provision is to be delivered out, but by a written order from him or the aid de camp. The Commissary is to send up the Cask of Tools by the first waggon that goes to the Fort. The following men being judged unfit for Duty, on Review; are ordered to be discharged: viz:

Figure 6-8: The training system

6.3.2 Document image / character correlation

In this step, one image is created for each character of the text query, presenting possible areas of the character in the document image.

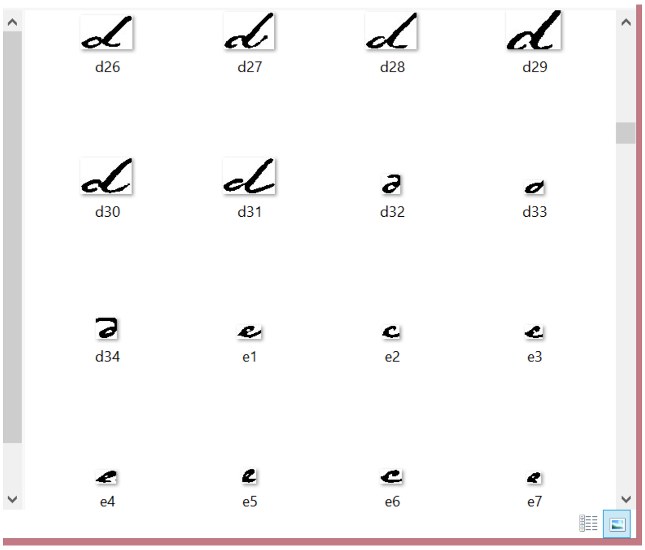
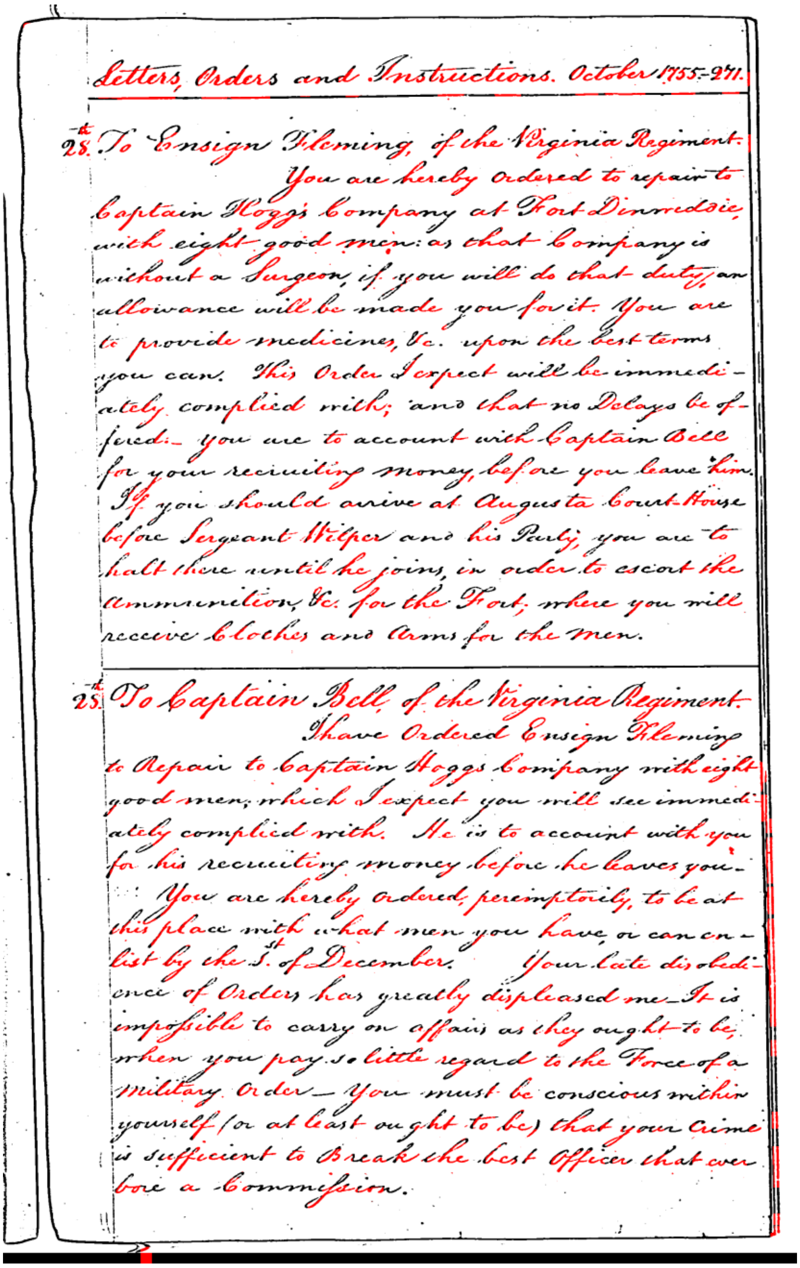


Figure 6-9: Several character samples

After the correlation of each sample (Fig. 6.9) with a document image, all the possible areas of the character are kept. Possible areas are considered the areas that after correlation with a character sample, present similarity over 70% with the specific sample. The possible areas for a certain character of the query are all the possible areas for all the samples of the characters. If a pixel is possible area for more than a sample, the sample that presents maximum similarity is considered.



Letters, Orders and Instructions. October 1755. 271.

28th. To Ensign Fleming, of the Virginia Regiment.

You are hereby ordered to repair to Captain Hoggs Company at Fort Dinwiddie, with eight good men: as that Company is without a Surgeon, if you will do that duty, an allowance will be made you for it. You are to provide medicines, &c. upon the best terms you can. This Order I expect will be immediately complied with; and that no Delays be offered: – you are to account with Captain Bell for your recruiting money, before you leave him. If you should arrive at Augusta Court-House before Sergeant Wilper and his Party, you are to halt there until he joins, in order to escort the ammunition, &c. for the Fort, where you will receive clothes and Arms for the men.

28th. To Captain Bell of the Virginia Regiment.

I have Ordered Ensign Fleming to Repair to Captain Hoggs Company with eight good men; which I expect you will see immediately complied with. He is to account with you for his recruiting money before he leaves you. – You are hereby ordered, peremptorily, to be at this place with what men you have, or can enlist by the 1st of December. Your late disobedience of Orders has greatly displeased me – It is impossible to carry on affairs as they ought to be, when you pay so little regard to the Force of a Military Order – You must be conscious within yourself (or at least ought to be) that your Crime is sufficient to Break the best Officer that ever bore a Commission.

Figure 6-10: Possible areas for character ‘t’

Although, the samples belong to the same character, they can differ in size (height or width). These dimensions are important to calculate the possible area, after the correlation, starting from a pixel of the image. These small differences also can create small mistakes in the calculation of the possible areas and the coloring of the corresponding image.

In Figures 6.10 & 6.11, the possible areas for the characters ‘t’ and ‘h’ are shown, respectively. The simplest the character the most false positives creates. The character ‘t’ that could be written in many cases as a vertical stroke, skipping the horizontal line can be matched to every vertical stroke.

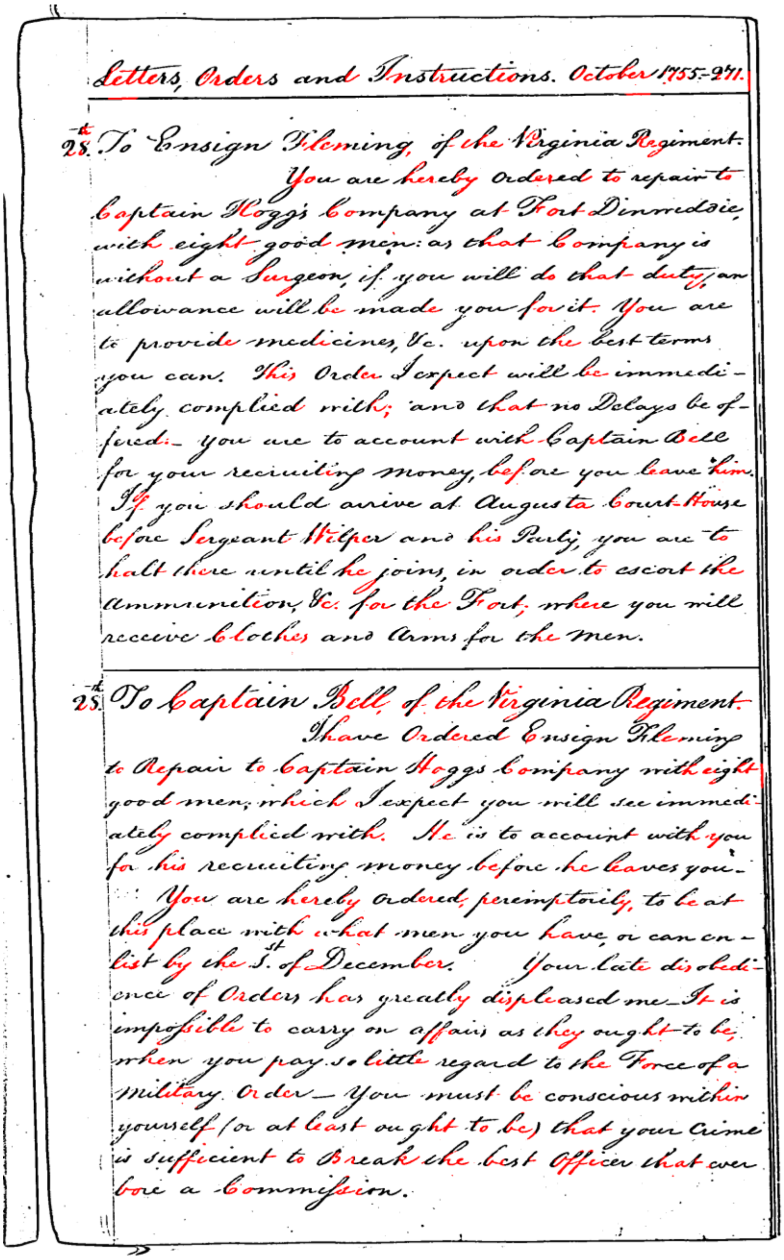
6.3.3 Character image combination

Thus, the combination of the correlation images for all the characters of the query, will give the final results. In order to succeed that, several simple rules are applied:

i. The correlation images of the characters of the query should all include consecutive possible areas in the same order as the characters in the query.

ii. The possible areas are allowed to be overlapped up to half character width with the previous character, in order to cover touching characters.

iii. The possible areas are allowed to have gap up to half character between them, in order to cover distant writing.



Letters, Orders and Instructions. October 1755-27.

28th To Ensign Fleming, of the Virginia Regiment.

You are hereby Ordered to repair to Captain Hogg's Company at Fort Dinwiddie, with eight good men: as that Company is without a Surgeon, if you will do that duty, an allowance will be made you for it. You are to provide medicines, &c. upon the best terms you can. This Order I expect will be immediately complied with; and that no Delays be offered:– you are to account with Captain Bell for your recruiting money, before you leave him. If you should arrive at Augusta Court-House before Sergeant Wilper and his Party, you are to halt there until he joins, in order to escort the Ammunition, &c. for the Fort, where you will receive Clothes and Arms for the men.

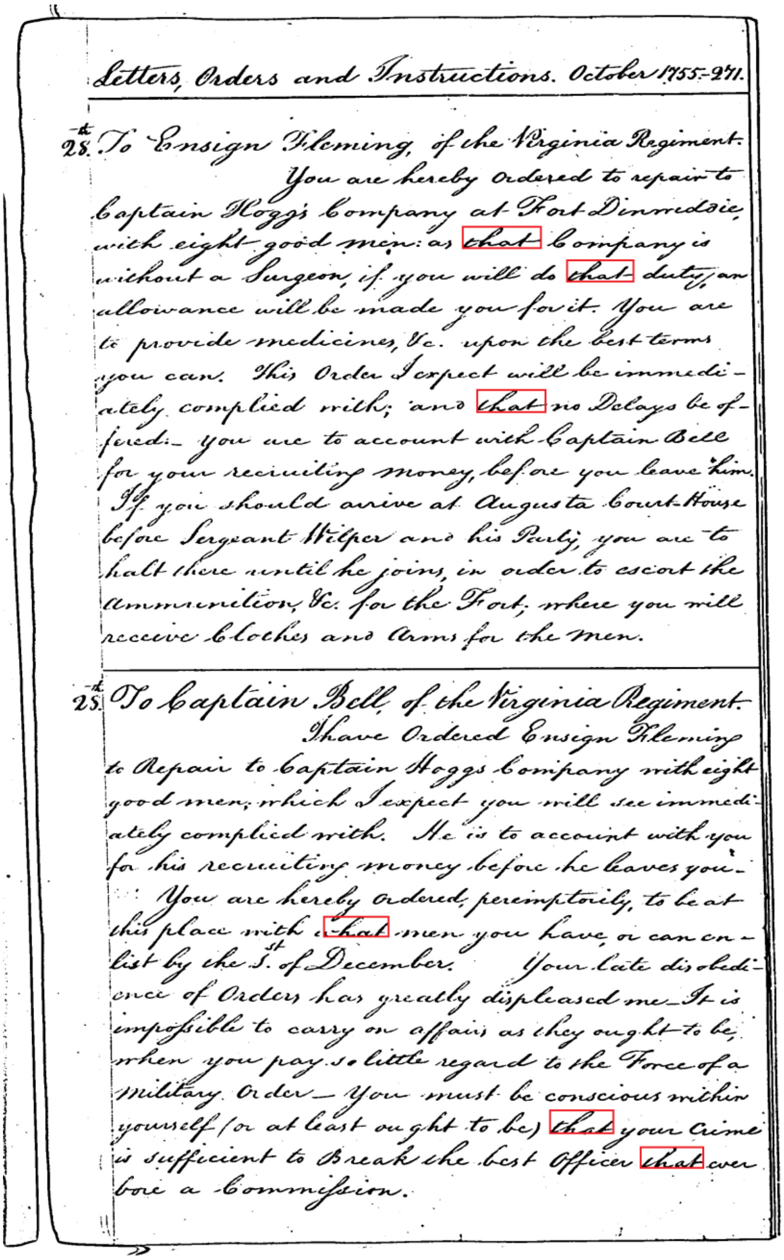
28th To Captain Bell of the Virginia Regiment.

I have Ordered Ensign Fleming to Repair to Captain Hoggs Company with eight good men, which I expect you will see immediately complied with. He is to account with you for his recruiting money before he leaves you.– You are hereby ordered, peremptorily, to be at this place with what men you have, or can enlist by the 1st of December. Your late disobedience of Orders has greatly displeased me_It is impoſsible to carry on affairs as they ought to be, when you pay so little regard to the Force of a Military Order– You must be conscious within yourself (or at least ought to be) that your Crime is sufficient to Break the best Officer that ever bore a Commiſsion.

Figure 6-11: Possible areas for character ‘h’

After localizing the areas that cover the previous rules in the document image, the borders of the words are estimated by looking for white areas around or almost white (10% black pixels can be crossed as foreign ascenders or descenders).

In Figure 6.12 a result of the retrieval of the query ‘that’ is presented. This is a good result since the technique found all the words in the specific document image and only a wrong one.



Letters, Orders and Instructions. October 1755-271.

28th. To Ensign Fleming, of the Virginia Regiment.
You are hereby Ordered to repair to
Captain Hoggs Company at Fort Dinwiddie,
with eight good men: as that Company is
without a Surgeon, if you will do that duty, an
allowance will be made you for it. You are
to provide medicines, &c. upon the best terms
you can. This Order I expect will be immedi-
ately complied with; and that no Delays be of-
fered:_ you are to account with Captain Bell
for your recruiting money, before you leave him.
If you should arive at Augusta Court-House
before Sergeant Wilper and his Party, you are to
halt there until he joins, in order to escort the
ammunition, &c. for the Fort, where you will
receive Clothes and Arms for the men.

25th. To Captain Bell, of the Virginia Regiment.
I have Ordered Ensign Fleming
to Repair to Captain Hoggs Company with eight
good men; which I expect you will see immedi-
ately complied with. He is to account with you
for his recruiting money before he leaves you._
You are hereby ordered, peremptorily, to be at
this place with that men you have, or can en-
list by the 1st of December. Your late disobedi-
ence of Orders has greatly displeased me_ It is
impoſsible to carry on affairs as they ought to be,
when you pay so little regard to the Force of a
Military Order_ You must be conscious within
yourself (or at least ought to be) that your Crime
is sufficient to Break the best Officer that ever
bore a Commiſsion.

Figure 6-12: Retrieving the query ‘that’

6.4 Experimental results

The proposed technique could be used as alternative to word-spotting, but more robust and easy to be trained for different languages and other sets of symbols. Experiments were performed on the dataset of George Washington’s letters [124]. That dataset has been used in many word-spotting systems, performing even more than 90%. Unfortunately, this is not the case of the proposed system. In an experiment were only the last letter was used for the training of the system and all the words of the letters as queries, the recall was 76,32% and the precision 64,29%. However, no normalization processing or other corrections have been applied to the document images, while the success rate for different work varies a lot. In general, longest words give better results while the presence of ascenders and descenders also improve the results even for short words as in Figure 6.12.

CHAPTER 7 - CONCLUSIONS

Although many alternative approaches can be found in the literature, the tasks of document image segmentation and text localization are still considered open problems. As stated in the Final Report of the Europeana Newspapers project [125]: […*In terms of layout analysis capabilities there is still room for improvement and any progress in this area could have a great impact on the usefulness of OCR results*…] Especially for historical newspapers and other historic documents, there is increasing interest and demand for fast and reliable analysis and text retrieval techniques.

Since most of the documents of the pre-digital era are not fully digitized yet, a huge amount of documents are expected to be available for processing the next years. The Wikipedia article for newspaper digitization [126] reports that: […*Newspapers preserve a rich record of the past, and since the advent of digital media, many institutions across the world have began to digitize them and make the digital files publicly available. However, over 90% of newspapers remained unscanned in 2015*…].

As already mentioned, newspapers commonly suffer from poor OCR outputs because OCR engines don’t like their column-style layouts. Those layouts also cause difficulties because they may contain illustrations as well as changes in font type, size and orientation. Moreover, old and partially damaged documents can hardly be segmented sometimes. This work has focused on those issues and has suggested novel page segmentation and text localization methods for the analysis of such documents. Since it is computationally very expensive to retrieve all the textual information from large collections, word-spotting and character-spotting techniques have been proposed as well. Those techniques bypass the segmentation steps and directly spot specific words in the database.

In Chapter 2, a hybrid technique for document image layout analysis has been presented. The technique is appropriate for colored and complex layouts of newspapers and journals, while no previous knowledge of the document is required. Morphological operations were applied to both the foreground and the background, in order to connect neighboring components and separate lines/columns. The contour was used for the extraction and classification of images and text blocks. A subtask for complex binarization is included, using regional band thresholding. This technique can be applied to pages with any background and foreground colors, resulting an improved binarized image, even if various backgrounds are used in the same collection or image. First, the background is processed in order to localize long white columns and rows that are used as separators during the foreground processing to split overlapped components and improve the page segmentation results. Next, the foreground analysis is applied to the inverted image. Finally, the segmented areas are classified to text, image or title blocks. The algorithm has been coded in C# language, using the OpenCV library. In order to give comparative results, the RDCL-2015 dataset of ICDAR 2015 competition was used. The results are promising, better than the commercial software mentioned in the competition and very close to the top ones.

In Chapter 3, a fast and reliable hybrid method for text localization in complex images has been proposed. No a priori knowledge of the image characteristics is required and there is no limitation on character font, size, orientation and color. First, the Canny operator is used for edge detection. Then, morphological operators are applied in order to connect neighboring components. Finally, a set of criteria based on spatial and geometrical features of the connected components are used for discriminating between text and non-text regions. The algorithm has been implemented in C# language. The OpenCV library has also been used. The system was evaluated on the ICDAR 2011 Robust Reading Competition dataset and achieved a precision close to the best reported at the competition's paper and a better recall than the participated methods. The mean processing time was less than a second per image.

In Chapter 4, a mixed layout analysis and text localization method has been presented. It is appropriate for colored and complex layouts of newspapers and journals, while no previous knowledge of the document is required. Morphological operations were applied in order to connect neighboring components. Canny edge detection and contour tracing were also used for the extraction and classification of text and non-text regions. The algorithm has been coded in C# language, using the OpenCV library. Evaluated on samples from two public datasets, the system achieved state of the art results in terms of both accuracy and speed. The text localization part of the algorithm performs very well in regions where the layout analysis fails. On the other hand, the layout analysis part can extract the major text and non-text blocks of the image much faster, keeping the total processing time low. In most of the cases the combination of layout analysis and text localization gave significantly improved results with a very small overload compared to the layout analysis method alone.

In Chapter 5, a system of word spotting has been presented and evaluated. The system proposes a simplified methodology that omits many tasks of the traditional word spotting approach. As preprocessing, it only requires a thresholding and it is a segmentation-free approach. Moreover, it does not include feature extraction and clustering or classification stages. The comparison and matching procedures are performed by image processing techniques. The proposed system was applied to a collection of Greek document images of the Government Gazette of the Principality of Samos, that are kept at the General State Archives records (GSA) of Samos. Moreover, it was also applied to a Google book of the 17th century, in order to compare results based on the OCR text provided by Google. Finally, some examples are presented for the same queries from the proposed system and an older one, that includes segmentation and classification.

In Chapter 6, a novel text retrieval technique has been presented that is inspired by word-spotting but it could make it more robust and general, as well as more easily adapted to different languages. The proposed technique is performing character spotting instead of words. This way character overlapping can be considered while samples from different writing style can be combined. Although the results of the proposed technique are not yet better than the best word-spotting systems, it can be easily improved, just applying usual technique of normalization and document image processing. In future plans, the improvement of the system includes the trial of document image processing techniques and the application to more datasets and even the consideration of more samples than just those from a single document image page.

For the future, the evaluation of the above methods on more datasets is planned. Moreover, the application and testing of the algorithms on similar computer vision problems is also considered. Two such problems are:

i. The segmentation of comic page images.

ii. The localization of text in video frames.

Appendix I includes segmentation results of scanned comic page images. The images are segmented into storyboards using a slightly modified version of the layout analysis method described in Chapter 1. Only the background analysis steps as well as the contour tracing and lines detection steps of the foreground analysis have actually been applied. The purpose of the segmentation of comic page images is mainly to produce digital comic documents for mobile devices [127-130]. More experiments are planned in order to evaluate the algorithm on more layouts and explore the possibility of text extraction from the balloons as well, in combination with the text localization method of Chapter 3.

Results of text localization in video frames are included in Appendix III. The importance of the extraction of textual information from videos as well as various proposed techniques are described in the literature [45,48-52,57,64]. The text localization method presented in Chapter 3 has been applied here in order to track text in videos. Further improvements are planned, so that the method can perform well in real-time and in low resolution videos from web cameras.

REFERENCES

[1] Sonka, Milan, Vaclav Hlavac, and Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014, pp. 179-180.

[2] Wong, Kwan Y., Richard G. Casey, and Friedrich M. Wahl. "Document analysis system." IBM journal of research and development 26.6 (1982): 647-656.

[3] Tsujimoto, Shuichi, and Haruo Asada. "Major components of a complete text reading system." Proceedings of the IEEE 80.7 (1992): 1133-1149.

[4] Strouthopoulos, C., N. Papamarkos, and C. Chamzas. "PLA using RLSA and a neural network." Engineering Applications of Artificial Intelligence 12.2 (1999): 119-138.

[5] Sun, Hung-Ming. "Page segmentation for Manhattan and non-Manhattan layout documents via selective CRLA." Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on. IEEE, 2005.

[6] Fletcher, Lloyd Alan, and Rangachar Kasturi. "A robust algorithm for text string separation from mixed text/graphics images." Pattern Analysis and Machine Intelligence, IEEE Transactions on 10.6 (1988): 910-918.

[7] Akiyama, Teruo, and Norihiro Hagita. "Automated entry system for printed documents." Pattern recognition 23.11 (1990): 1141-1154.

[8] O'Gorman, Lawrence. "The document spectrum for page layout analysis." Pattern Analysis and Machine Intelligence, IEEE Transactions on 15.11 (1993): 1162-1173.

[9] Hönes, Frank, and Jürgen Lichter. "Layout extraction of mixed mode documents." Machine vision and applications 7.4 (1994): 237-246.

[10] Zlatopolsky, A. A. "Automated document segmentation." Pattern Recognition Letters 15.7 (1994): 699-704.

[11] Olivier, Déforges, and Barba Dominique. "Segmentation of complex documents multilevel images: a robust and fast text bodies-headers detection and extraction scheme." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 2. IEEE, 1995.

[12] Wang, Shin-Ywan, and Toshiaki Yagasaki. "Block selection: a method for segmenting a page image of various editing styles." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 1. IEEE, 1995.

[13] Simon, Anikó, and Jean Christophe Pret. "A fast algorithm for bottom-up document layout analysis." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.3 (1997): 273-277.

[14] Bukhari, Syed Saqib, et al. "Document image segmentation using discriminative learning over connected components." Proceedings of the 9th IAPR International Workshop on Document Analysis Systems. ACM, 2010.

[15] Koo, Hyung Il, and Duck Hoon Kim. "Scene text detection via connected component clustering and nontext filtering." Image Processing, IEEE Transactions on 22.6 (2013): 2296-2305.

[16] Le, Viet Phuong, et al. "Text and non-text segmentation based on connected component features." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.

[17] Gatos, Basilios, S. L. Mantzaris, and Apostolos Antonacopoulos. "First international newspaper segmentation contest." Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on. IEEE, 2001.

[18] Antonacopoulos, Apostolos, Basilios Gatos, and David Bridson. "Page segmentation competition." Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on. Vol. 2. IEEE, 2007.

[19] Nagy, George, Sharad Seth, and Mahesh Viswanathan. "A prototype document image analysis system for technical journals." Computer 25.7 (1992): 10-22.

[20] Ha, Jaekyu, Robert M. Haralick, and Ihsin T. Phillips. "Document page decomposition by the bounding-box project." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 2. IEEE, 1995.

[21] Baird, Henry S., Susan E. Jones, and Steven J. Fortune. "Image segmentation by shape-directed covers." Pattern Recognition, 1990. Proceedings., 10th International Conference on. Vol. 1. IEEE, 1990.

[22] Breuel, Thomas M. "Two geometric algorithms for layout analysis." Document analysis systems v. Springer Berlin Heidelberg, 2002. 188-199.

[23] Normand, Nicolas, and Christian Viard-Gaudin. "A background based adaptive page segmentation algorithm." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 1. IEEE, 1995.

[24] Kise, Koichi, O. Yanagida, and Shinobu Takamatsu. "Page segmentation based on thinning of background." Pattern Recognition, 1996., Proceedings of the 13th International Conference on. Vol. 3. IEEE, 1996.

[25] Kise, Koichi, Akinori Sato, and Motoi Iwata. "Segmentation of page images using the area Voronoi diagram." Computer Vision and Image Understanding 70.3 (1998): 370-382.

[26] T. Pavlidis, J. Zhou, Page segmentation and classification, CVGIP: Graphical Models and Image Processing, vol. 54, pp. 484-496, 1992.

[27] Antonacopoulos, A., and R. T. Ritchings. "Flexible page segmentation using the background." Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on. Vol. 2. IEEE, 1994.

[28] Smith, Ray. "Hybrid page layout analysis via tab-stop detection." Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on. IEEE, 2009.

[29] Chen, Kai, Fei Yin, and Cheng-Lin Liu. "Hybrid page segmentation with efficient whitespace rectangles extraction and grouping." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[30] Antonacopoulos, Apostolos, et al. "Icdar 2013 competition on historical newspaper layout analysis (hnla 2013)." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[31] Antonacopoulos, A., et al. "ICDAR2015 competition on recognition of documents with complex layouts-RDCL2015." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.

[32] Jain, Anil K., and Sushil Bhattacharjee. "Text segmentation using Gabor filters for automatic document processing." Machine Vision and Applications 5.3 (1992): 169-184.

[33] Tang, Yuan Y., et al. "A new approach to document analysis based on modified fractal signature." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 2. IEEE, 1995.

[34] Sauvola, Jaakko, and Matti Pietikäinen. "Page segmentation and classification using fast feature extraction and connectivity analysis." icdar. IEEE, 1995.

[35] Williams, Paul Stefan, and Mike D. Alder. "Generic texture analysis applied to newspaper segmentation." Neural Networks, 1996., IEEE International Conference on. Vol. 3. IEEE, 1996.

[36] Etemad, Kamran, David Doermann, and Rama Chellappa. "Multiscale segmentation of unstructured document pages using soft decision integration." IEEE Transactions on Pattern Analysis & Machine Intelligence 1 (1997): 92-96.

[37] Strouthopoulos, Charalambos, and Nikos Papamarkos. "Text identification for document image analysis using a neural network." Image and Vision Computing 16.12 (1998): 879-896.

[38] Acharyya, Mausumi, and Malay K. Kundu. "Multiscale segmentation of document images using m-band wavelets." Computer Analysis of Images and Patterns. Springer Berlin Heidelberg, 2001.

[39] Kumar, Sunil, et al. "Text extraction and document image segmentation using matched wavelets and mrf model." Image Processing, IEEE Transactions on 16.8 (2007): 2117-2128.

[40] Maji, Pradipta, and Shaswati Roy. "Rough-fuzzy clustering and multiresolution image analysis for text-graphics segmentation." Applied Soft Computing 30 (2015): 705-721.

[41] Chen, Kai, et al. "Page segmentation of historical document images with convolutional autoencoders." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.

[42] Suzuki, Satoshi. "Topological structural analysis of digitized binary images by border following." Computer Vision, Graphics, and Image Processing 30.1 (1985): 32-46.

[43] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." Automatica 11.285-296 (1975): 23-27.

[44] E. Kavallieratou, N. Dromazou, N. Fakotakis, G. Kokkinakis, An Integrated System for Handwritten Document Image Processing, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 17, No. 4, pp. 101-120, 2003.

[45] Jung, Keechul, Kwang In Kim, and Anil K. Jain. "Text information extraction in images and video: a survey." Pattern recognition 37.5 (2004): 977-997.

[46] Ohya, Jun, Akio Shio, and Shigeru Akamatsu. "Recognizing characters in scene images." Pattern Analysis and Machine Intelligence, IEEE Transactions on 16.2 (1994): 214-220.

[47] Lee, Chung-Mong, and Atreyi Kankanhalli. "Automatic extraction of characters in complex scene images." International Journal of Pattern Recognition and Artificial Intelligence 9.01 (1995): 67-82.

[48] Smith, Michael A., and Takeo Kanade. Video skimming for quick browsing based on audio and image characterization. School of Computer Science, Carnegie Mellon University, 1995.

[49] Kim, Hae-Kwang. "Efficient automatic text location method and content-based indexing and structuring of video database." Journal of Visual Communication and Image Representation 7.4 (1996): 336-344.

[50] Lienhart, Rainer W., and Frank Stuber. "Automatic text recognition in digital videos." Electronic Imaging: Science & Technology. International Society for Optics and Photonics, 1996.

[51] Shim, Jae-Chang, Chitra Dorai, and Ruud Bolle. "Automatic text extraction from video for content-based annotation and retrieval." Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. Vol. 1. IEEE, 1998.

[52] Jain, Anil K., and Bin Yu. "Automatic text location in images and video frames." Pattern recognition 31.12 (1998): 2055-2076.

[53] Messelodi, Stefano, and Carla Maria Modena. "Automatic identification and skew estimation of text lines in real scene images." Pattern Recognition 32.5 (1999): 791-810.

[54] Chen, Datong, Kim Shearer, and Hervé Bourlard. "Text enhancement with asymmetric filter for video OCR." Image Analysis and Processing, 2001. Proceedings. 11th International Conference on. IEEE, 2001.

[55] Nikolaou, Nikos, and Nikos Papamarkos. "Color reduction for complex document images." International Journal of Imaging Systems and Technology 19.1 (2009): 14-26.

[56] Wu, Victor, R. Manmatha, and Edward M. Riseman. "Finding text in images." ACM DL. 1997.

[57] Li, Huiping, David Doermann, and Omid Kia. "Automatic text detection and tracking in digital video." Image Processing, IEEE Transactions on 9.1 (2000): 147-156.

[58] Jung, Keechul. "Neural network-based text location in color images." Pattern Recognition Letters 22.14 (2001): 1503-1515.

[59] Clark, Paul, and Majid Mirmehdi. "Recognising text in real scenes." International Journal on Document Analysis and Recognition 4.4 (2002): 243-257.

[60] Kim, Kwang In, Keechul Jung, and Jin Hyung Kim. "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm." Pattern Analysis and Machine Intelligence, IEEE Transactions on 25.12 (2003): 1631-1639.

[61] Gllavata, Julinda, Ralph Ewerth, and Bernd Freisleben. "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients." Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Vol. 1. IEEE, 2004.

[62] Weinman, Jerod J., Erik Learned-Miller, and Allen Hanson. "A discriminative semi-markov model for robust scene text recognition." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[63] Jung, Cheolkon, Qifeng Liu, and Joongkyu Kim. "A stroke filter and its application to text localization." Pattern Recognition Letters 30.2 (2009): 114-122.

[64] Wang, Xiufei, Lei Huang, and Changping Liu. "A video text location method based on background classification." International Journal on Document Analysis and Recognition (IJDAR) 13.3 (2010): 173-186.

[65] Yin, Xuwang, et al. "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.

[66] Zhong, Yu, Kalle Karu, and Anil K. Jain. "Locating text in complex color images." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 1. IEEE, 1995.

[67] Yangxing, L. I. U., and Takeshi IKENAGA. "A contour-based robust algorithm for text detection in color images." IEICE transactions on information and systems 89.3 (2006): 1221-1230.

[68] Mancas-Thillou, Celine, and Bernard Gosselin. "Color text extraction with selective metric-based clustering." Computer Vision and Image Understanding 107.1 (2007): 97-107.

[69] Pan, Yi-Feng, Xinwen Hou, and Cheng-Lin Liu. "Text localization in natural scene images based on conditional random field." Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on. IEEE, 2009.

[70] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.

[71] Zhang, Jing, and Rangachar Kasturi. "Character energy and link energy-based text extraction in scene images." Computer Vision–ACCV 2010. Springer Berlin Heidelberg, 2010. 308-320.

[72] Zhao, Ming, Shutao Li, and James Kwok. "Text detection in images using sparse representation with discriminative dictionaries." Image and Vision Computing 28.12 (2010): 1590-1599.

[73] Yi, Chucai, and YingLi Tian. "Text string detection from natural scenes by structure-based partition and grouping." Image Processing, IEEE Transactions on 20.9 (2011): 2594-2605.

[74] Neumann, Lukas, and Jiri Matas. "A method for text localization and recognition in real-world images." Computer Vision–ACCV 2010. Springer Berlin Heidelberg, 2010. 770-783.

[75] Yin, Xuwang, et al. "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.

[76] Shi, Cunzhao, et al. "Scene text detection using graph model built upon maximally stable extremal regions." Pattern recognition letters 34.2 (2013): 107-116.

[77] Lucas, Simon M., et al. "ICDAR 2003 robust reading competitions." null. IEEE, 2003.

[78] Lucas, Simon M. "ICDAR 2005 text locating competition results." Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on. IEEE, 2005.

[79] Shahab, Asif, Faisal Shafait, and Andreas Dengel. "ICDAR 2011 robust reading competition challenge 2: Reading text in scene images." Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011.

[80] Karatzas, Dimosthenis, et al. "ICDAR 2015 competition on Robust Reading." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.

[81] Canny, John. "A computational approach to edge detection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 6 (1986): 679-698.

[82] http://archive.ics.uci.edu/ml/datasets/Newspaper+and+magazine+images+segmentation+dataset

[83] T. Rath and R. Manmatha, “Word spotting for historical documents” International Journal of Document Analysis and Recognition, Vol. 9, No. 2-4, pp.139–152, (2007).

[84] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis, and J. Perantonis, “Keyword-guided Word Spotting in historical printed documents using synthetic data and user feedback” International Journal of Document Analysis and Recognition, Vol. 9, pp.167-177, (2007).

[85] H. Cao, A. Bhardwaj, V. Govindaraju, “A probabilistic method for keyword retrieval in handwritten document images”, Pattern Recognition, Vol.42, No 12, pp.3374-3382, (2009).

[86] B. Gatos and I. Pratikakis, “Segmentation-free word spotting in historical printed documents” Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 271–275, (2009).

[87] R. Farrahi Moghaddam. and M. Cheriet, “Application of multi-level classifiers and clustering for automatic word-spotting in historical document images”, Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 511–515, (2009).

[88] Y. Leydier, A. Ouji, F. LeBourgeois, and H. Emptoz, “Towards an omnilingual word retrieval system for ancient manuscripts”, Pattern Recognition, Vol. 42, No. 9, pp. 2089–2105, (2009).

[89] M. Rusiňol, D. Aldavert, R. Toledo, J. Llados, “Browsing Heterogeneous Document Collections by a Segmentation-Free Word Spotting Method”, International Conference on Document Analysis and Recognition (ICDAR), pp.63-67, (2011).

[90] Zagoris, Konstantinos, Kavallieratou Ergina, and Nikos Papamarkos. "Image retrieval systems based on compact shape descriptor and relevance feedback information." Journal of Visual Communication and Image Representation 22.5 (2011): 378-390.

[91] E.Kavallieratou, N.Fakotakis, and G.Kokkinakis, “An Off-line Unconstrained Handwritting Recognition System”, International Journal of Document Analysis and Recognition, no 4, pp. 226-242, (2002).

[92] http://books.google.gr/books?id=-SY3AAAAMAAJ&printsec=frontcover&dq=preach+king&hl=el&sa=X&ei=jIlQT6SkOcrP0QX79Mz0Cw&redir_esc=y#v=onepage&q=preach%20king&f=false

[93] N. Doulgeri and E. Kavallieratou, “Retrieval of historical documents by word spotting” Proceedings of SPIE, Volume 7247, Retrieval and Text Categorization, pp.06, (2009).

[94] Manmatha, Raghavan, Chengfeng Han, and Edward M. Riseman. "Word spotting: A new approach to indexing handwriting." Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on. IEEE, 1996.

[95] Zhang, Bin, Sargur N. Srihari, and Chen Huang. "Word image retrieval using binary features." Electronic Imaging 2004. International Society for Optics and Photonics, 2003.

[96] Rothfeder, Jamie L., Shaolei Feng, and Toni M. Rath. "Using corner feature correspondences to rank word images by similarity." Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on. Vol. 3. IEEE, 2003.

[97] Van der Zant, Tijn, Lambert Schomaker, and Koen Haak. "Handwritten-word spotting using biologically inspired features." Pattern Analysis and Machine Intelligence, IEEE Transactions on 30.11 (2008): 1945-1957.

[98] Srihari, Sargur N., and Gregory R. Ball. "Language independent word spotting in scanned documents." Digital Libraries: Universal and Ubiquitous Access to Information. Springer Berlin Heidelberg, 2008. 134-143.

[99] Bhardwaj, Anurag, Damien Jose, and Venu Govindaraju. "Script Independent Word Spotting in Multilingual Documents." IJCNLP. 2008.

[100] Leydier, Yann, Frank Lebourgeois, and Hubert Emptoz. "Text search for medieval manuscript images." Pattern Recognition 40.12 (2007): 3552-3567.

[101] Kolcz, Aleksander, et al. "A line-oriented approach to word spotting in handwritten documents." Pattern Analysis & Applications 3.2 (2000): 153-168.

[102] Rath, Toni M., and Raghavan Manmatha. "Word image matching using dynamic time warping." Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 2. IEEE, 2003.

[103] Adamek, Tomasz, Noel E. O’Connor, and Alan F. Smeaton. "Word matching using single closed contours for indexing handwritten historical documents." International Journal of Document Analysis and Recognition (IJDAR) 9.2-4 (2007): 153-165.

[104] Terasawa, Kengo, and Yuzuru Tanaka. "Slit style HOG feature for document image word spotting." Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on. IEEE, 2009.

[105] Wang, Peng, et al. "A novel learning-free word spotting approach based on graph representation." Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on. IEEE, 2014.

[106] Rodrıguez, José A., and Florent Perronnin. "Local gradient histogram features for word spotting in unconstrained handwritten documents." Int. Conf. on Frontiers in Handwriting Recognition. 2008.

[107] Zhang, Xi, and Chew Lim Tan. "Segmentation-free keyword spotting for handwritten documents based on heat kernel signature." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[108] Shekhar, Ravi, and C. V. Jawahar. "Word image retrieval using bag of visual words." Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on. IEEE, 2012.

[109] Aldavert, David, et al. "Integrating visual and textual cues for query-by-string word spotting." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[110] Choisy, Christophe. "Dynamic handwritten keyword spotting based on the nshp-hmm." Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on. Vol. 1. IEEE, 2007.

[111] Perronnin, Florent, and Jose A. Rodriguez-Serrano. "Fisher kernels for handwritten word-spotting." Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on. IEEE, 2009.

[112] Rodríguez-Serrano, José A., and Florent Perronnin. "Handwritten word-spotting using hidden Markov models and universal vocabularies." Pattern Recognition 42.9 (2009): 2106-2116.

[113] Rodríguez-Serrano, José A., and Florent Perronnin. "A model-based sequence similarity with application to handwritten word spotting." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.11 (2012): 2108-2120.

[114] Rothacker, Leonard, Marcal Rusinol, and Glenn A. Fink. "Bag-of-features HMMs for segmentation-free word spotting in handwritten documents." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[115] David, Jaety Edwards Yee Whye Teh, Forsyth Roger Bock Michael Maire, and Grace Vesom. "Making latin manuscripts searchable using gHMM's." Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference. Vol. 17. MIT Press, 2005.

[116] Chan, Jim, Celal Ziftci, and David Forsyth. "Searching off-line arabic documents." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.

[117] Thomas, Simon, et al. "An information extraction model for unconstrained handwritten documents." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[118] Fischer, Andreas, et al. "Lexicon-free handwritten word spotting using character HMMs." Pattern Recognition Letters 33.7 (2012): 934-942.

[119] Frinken, Volkmar, Andreas Fischer, and Horst Bunke. "A novel word spotting algorithm using bidirectional long short-term memory neural networks." Artificial Neural Networks in Pattern Recognition. Springer Berlin Heidelberg, 2010. 185-196.

[120] Frinken, Volkmar, et al. "A novel word spotting method based on recurrent neural networks." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.2 (2012): 211-224.

[121] Almazan, Jon, et al. "Handwritten word spotting with corrected attributes." Proceedings of the IEEE International Conference on Computer Vision. 2013.

[122] Howe, Nicholas R. "Part-structured inkball models for one-shot handwritten word spotting." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[123] Pratikakis, Ioannis, et al. "ICFHR 2014 competition on handwritten keyword spotting (H-KWS 2014)." Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on. IEEE, 2014.

[124] http://www.fki.inf.unibe.ch/databases/iam-historical-document-database/washington-database

[125] http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Final_Report.pdf

[126] https://en.wikipedia.org/wiki/Newspaper_digitization

[127] Yamada, Masashi, et al. "Comic image decomposition for reading comics on cellular phones." IEICE transactions on information and systems 87.6 (2004): 1370-1376.

[128] Tanaka, Takamasa, et al. "Layout Analysis of Tree-Structured Scene Frames in Comic Images." IJCAI. Vol. 7. 2007.

[129] Ishii, Daisuke, and Hiroshi Watanabe. "A study on frame position detection of digitized comics images." Proc. Workshop on Picture Coding and Image Processing, PCSJ2010/IMPS2010, Nagoya, Japan. 2010.

[130] TAKAHASHI, Atushi, et al. "Fast frame decomposition and sorting by contour tracing for mobile phone comic images." (2011).

[131] Guerin, Cyrielle, et al. "eBDtheque: a representative database of comics." Document Analysis and Recognition (ICDAR), 2013 12th international conference on. IEEE, 2013.

APPENDIX I – COMIC PAGE SEGMENTATION

In this appendix, results from the application of a slightly modified version of the layout analysis method presented in Chapter 2 are shown. The background analysis and the lines extraction steps of the proposed system are used to automatically decompose scanned comic page images into storyboards in order to produce digital comic documents that are suitable for reading on mobile devices (Fig. I.1). The reading order of the frames is determined by considering their position, starting from the upper left corner of the page and reading them row by row.

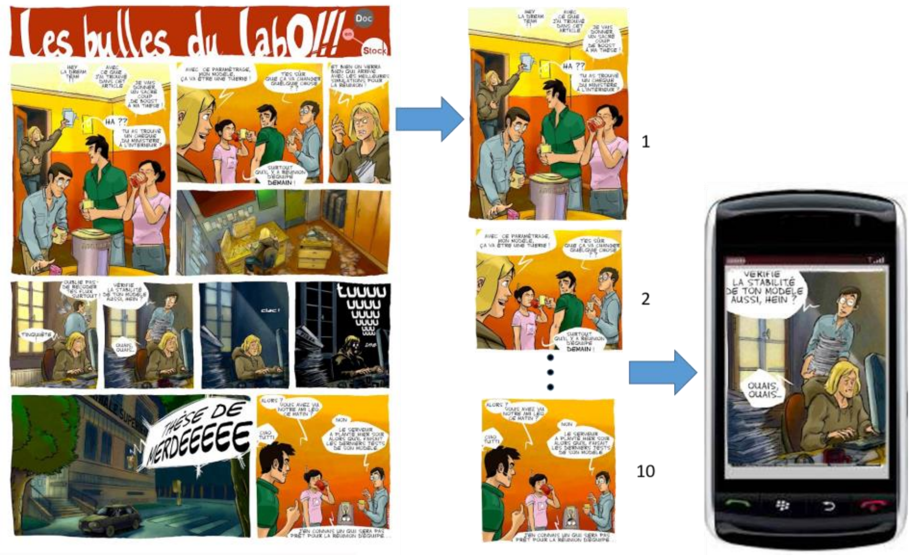
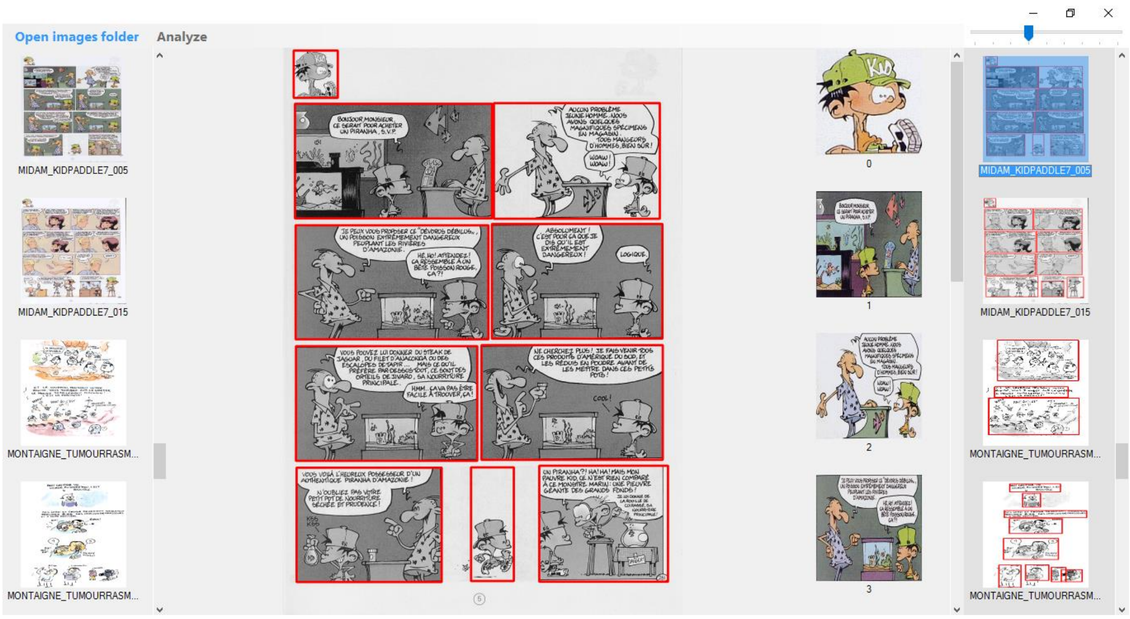
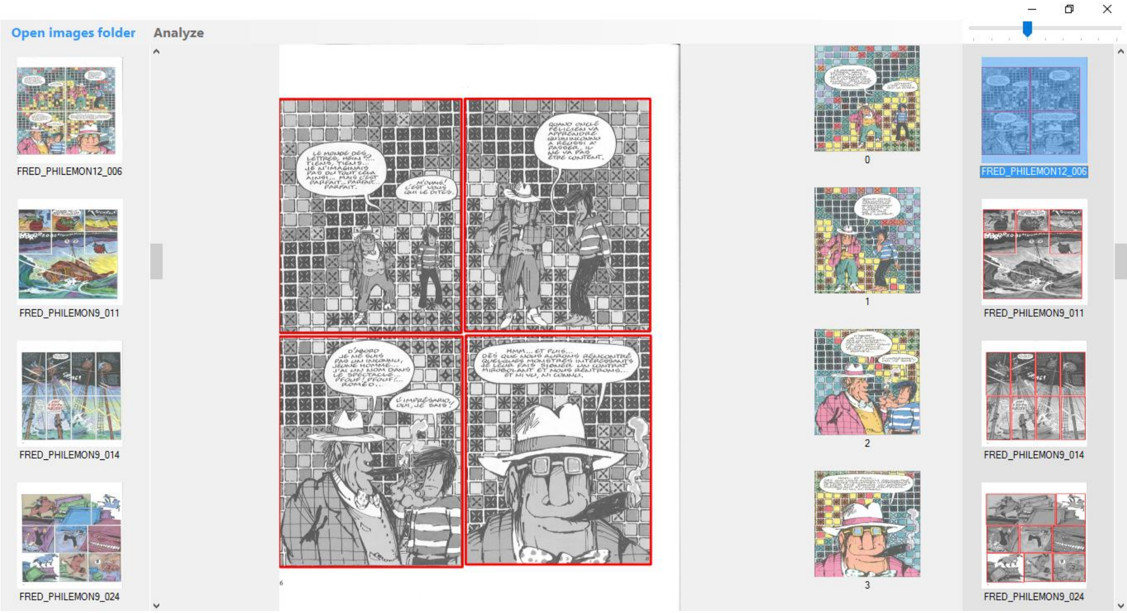
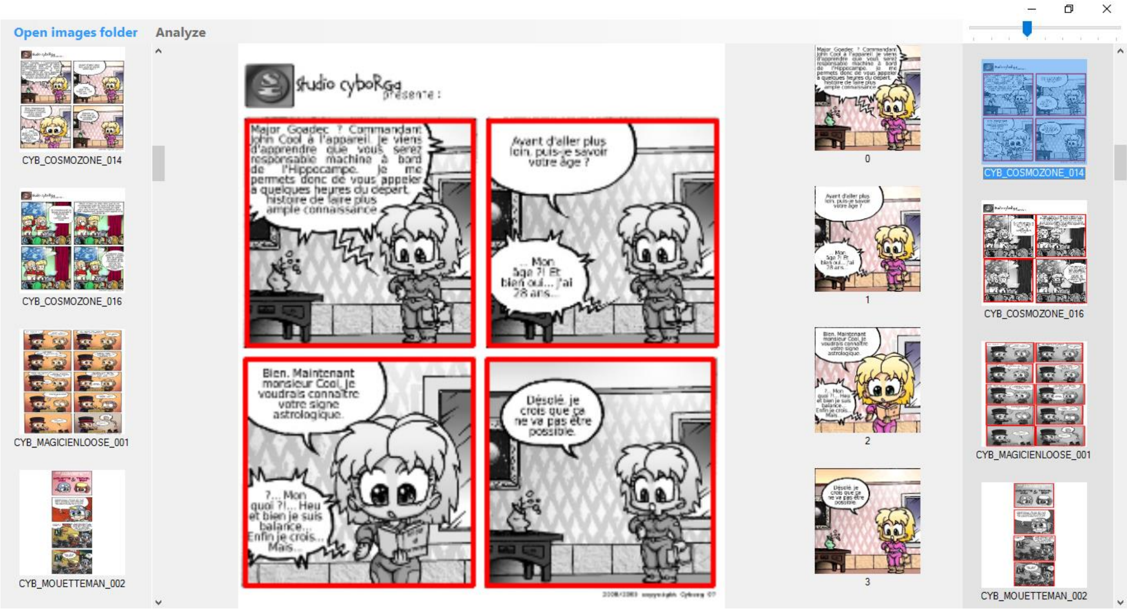


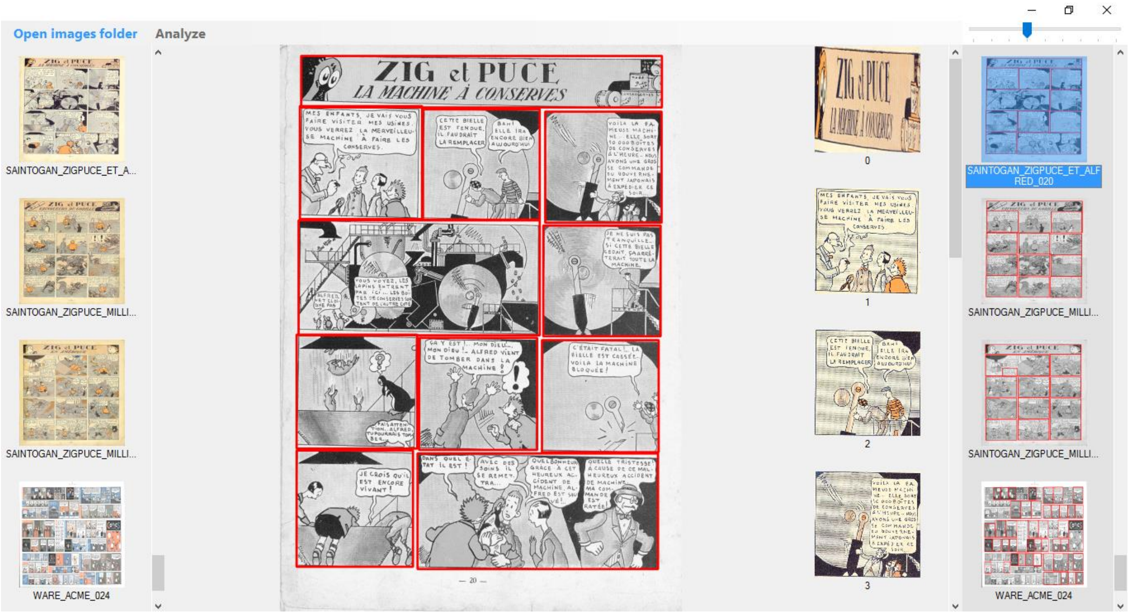
Figure I-1: Content adaptation for reading comic on mobile devices

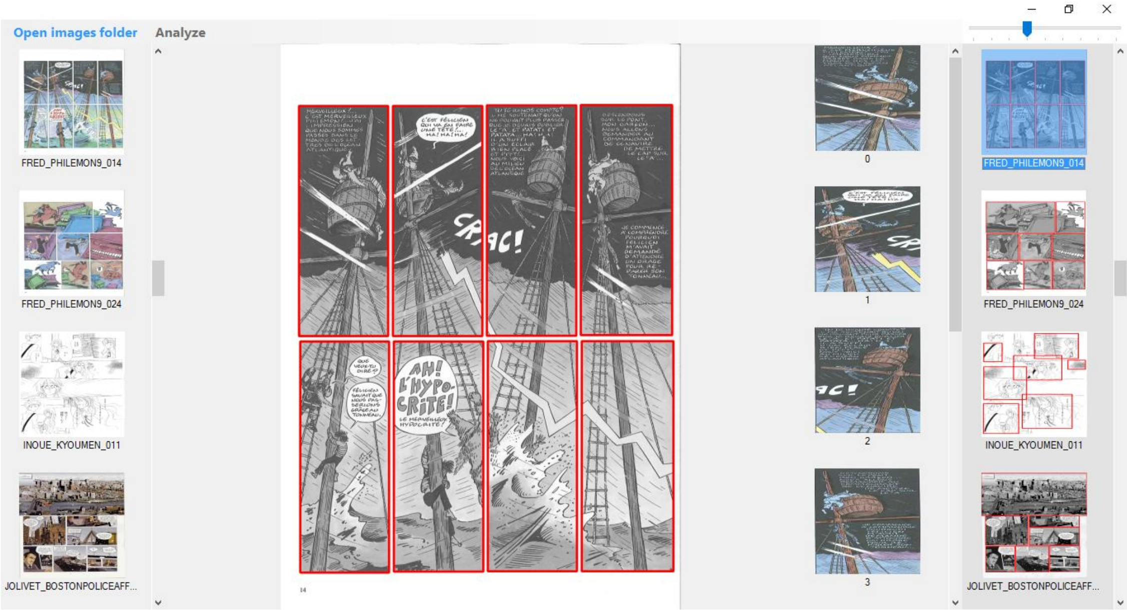
Evaluated on a public dataset [131] the technique proved to be available for segmentation of comic pages as well. Only American and European type of comics have been considered for the experiments. Japanese comics (manga) require a different approach since they have a special layout format. The following figures show screenshots of the page segmentation tool as it processes various samples from the dataset.

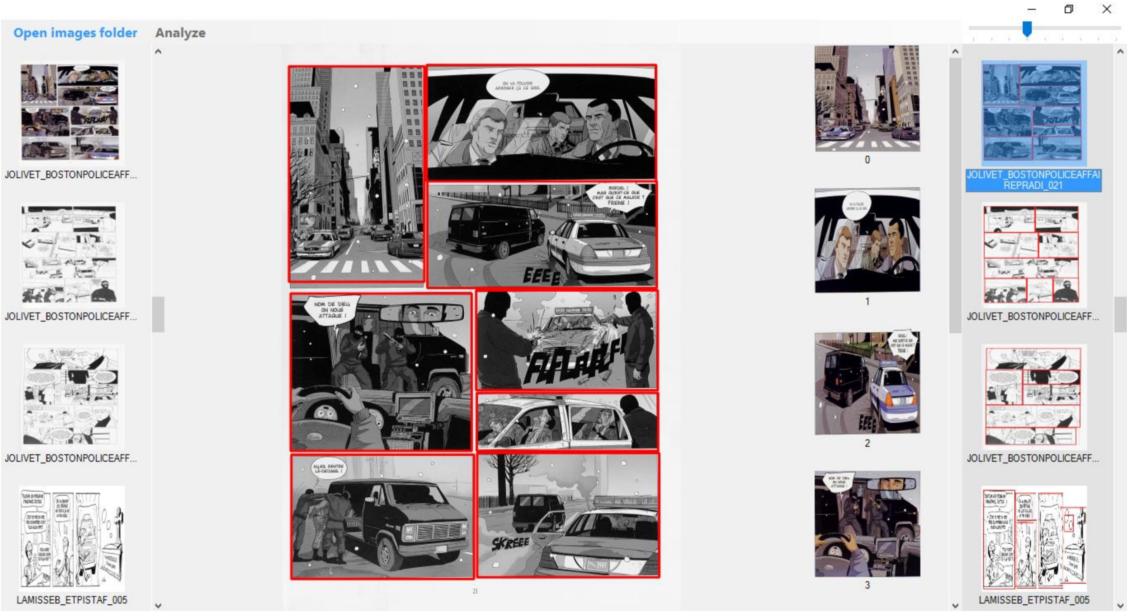


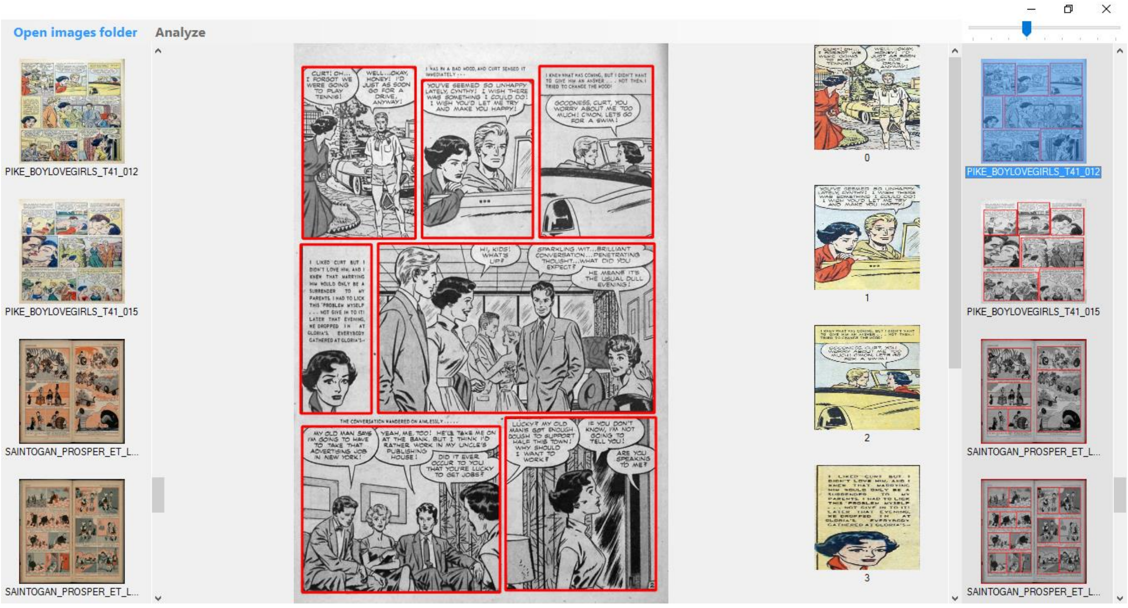

Open images folder
Analyze
FRED_PHILEMON12_006
FRED_PHILEMON9_011
FRED_PHILEMON9_014
FRED_PHILEMON9_024
0
1
2
3
FRED_PHILEMON12_006
FRED_PHILEMON9_011
FRED_PHILEMON9_014
FRED_PHILEMON9_024

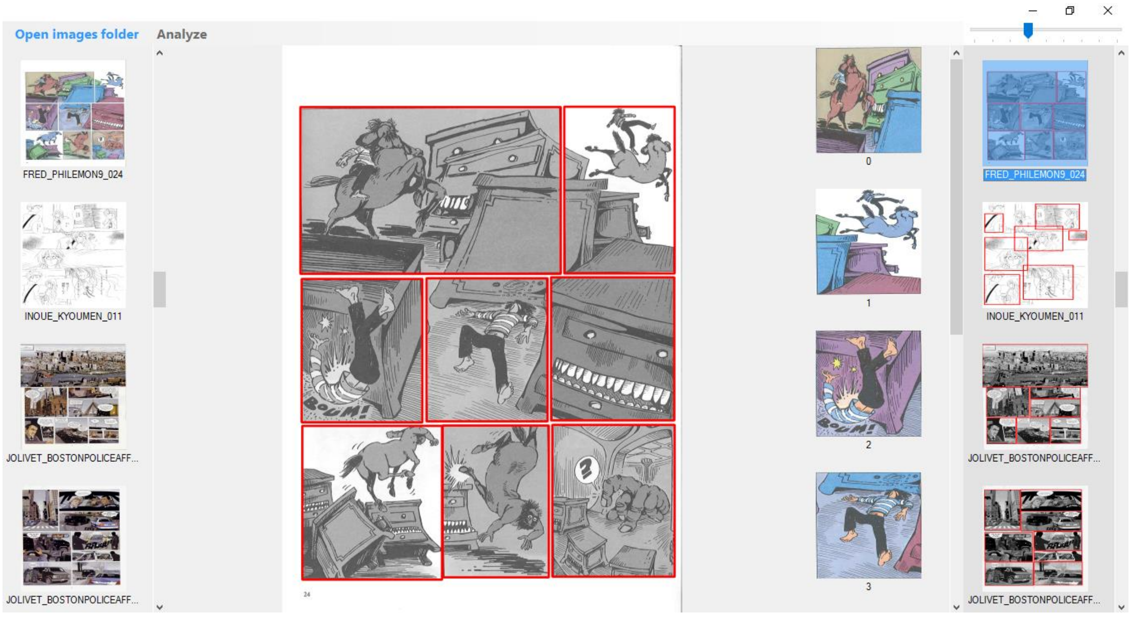

Open images folder
Analyze
CYB_COSMOZONE_014
CYB_COSMOZONE_016
CYB_MAGICIENLOOSE_001
CYB_MOUETTEMAN_002
studio cyboRga présente :
Major Goadec ? Commandant John Cool à l'appareil. Je viens d'apprendre que vous serez responsable machine à bord de l'Hippocampe. Je me permets donc de vous appeler à quelques heures du départ histoire de faire plus ample connaissance.
Avant d'aller plus loin, puis-je savoir votre âge ?
Mon âge ?! Et bien oui... j'ai 28 ans...
Bien. Maintenant monsieur Cool, je voudrais connaître votre signe astrologique.
?... Mon quoi ?!... Heu et bien je suis balance... Enfin je crois... Mais...
Désolé, je crois que ça ne va pas être possible.
0
1
2
3
CYB_COSMOZONE_014
CYB_COSMOZONE_016
CYB_MAGICIENLOOSE_001
CYB_MOUETTEMAN_002

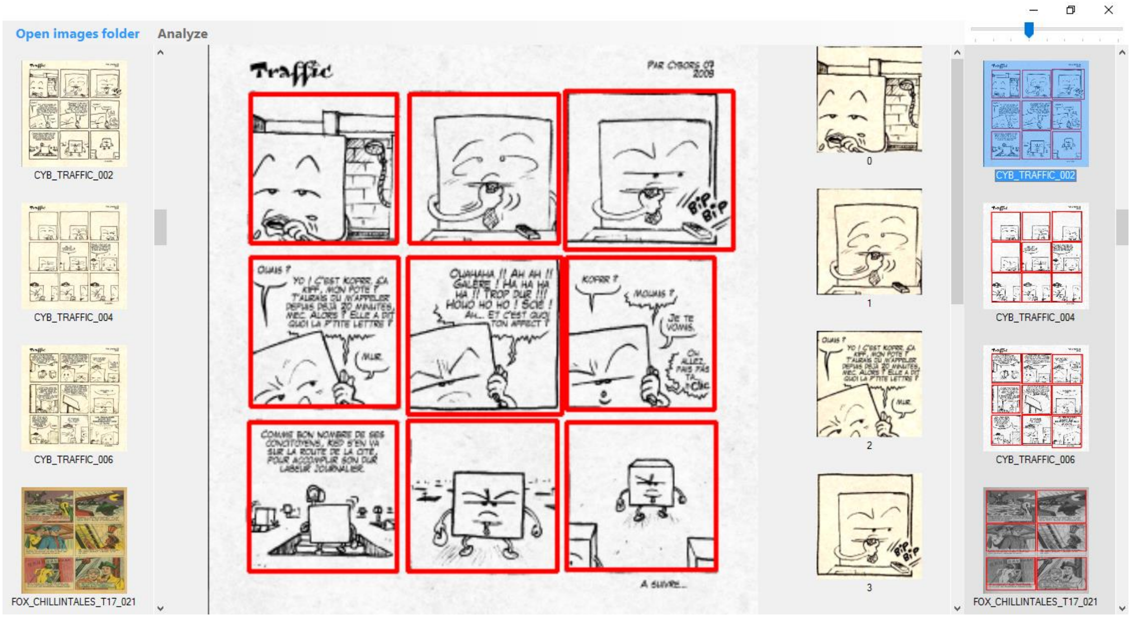

Open images folder
Analyze
SAINTOGAN_ZIGPUCE_ET_A...
SAINTOGAN_ZIGPUCE_MILLI...
SAINTOGAN_ZIGPUCE_MILLI...
WARE_ACME_024
ZIG et PUCE
LA MACHINE À CONSERVES
0
1
2
3
SAINTOGAN_ZIGPUCE_ET_ALFRED_020
SAINTOGAN_ZIGPUCE_MILLI...
SAINTOGAN_ZIGPUCE_MILLI...
WARE_ACME_024

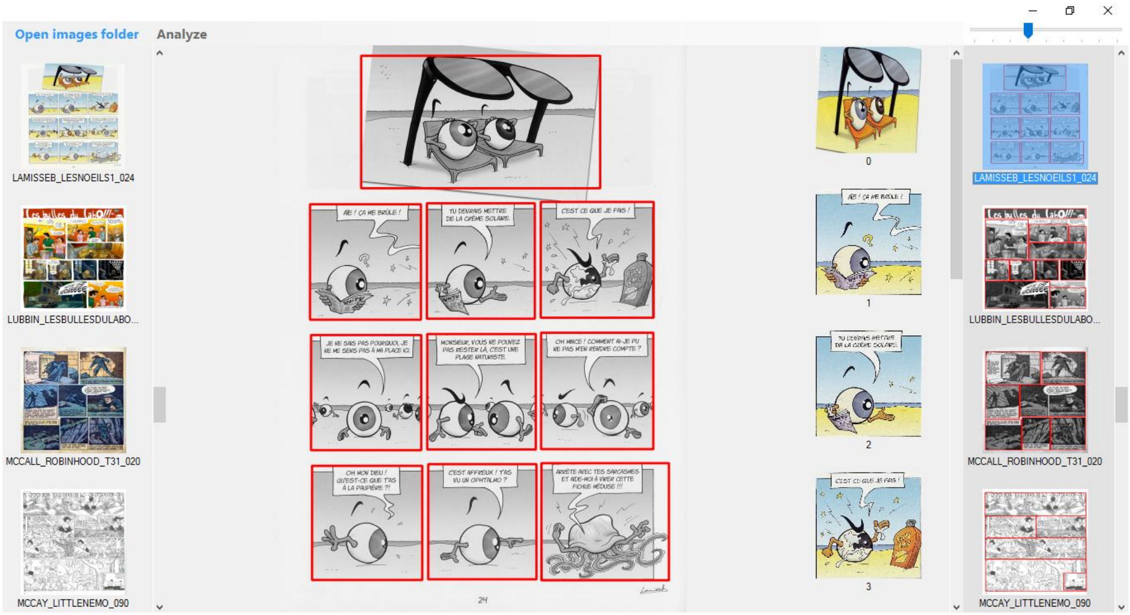

Open images folder
Analyze
FRED_PHILEMON9_014
FRED_PHILEMON9_024
INOUE_KYOUMEN_011
JOLIVET_BOSTONPOLICEAFF...
0
1
2
3
FRED_PHILEMON9_014
FRED_PHILEMON9_024
INOUE_KYOUMEN_011
JOLIVET_BOSTONPOLICEAFF...


Open images folder
Analyze
JOLIVET_BOSTONPOLICEAFF...
JOLIVET_BOSTONPOLICEAFF...
JOLIVET_BOSTONPOLICEAFF...
LAMISSEB_ETPISTAF_005
0
1
2
3
JOLIVET_BOSTONPOLICEAFFAI REPRADI_021
JOLIVET_BOSTONPOLICEAFF...
JOLIVET_BOSTONPOLICEAFF...
LAMISSEB_ETPISTAF_005


Open images folder
Analyze
PIKE_BOYLOVEGIRLS_T41_012
PIKE_BOYLOVEGIRLS_T41_015
SAINTOGAN_PROSPER_ET_L...
SAINTOGAN_PROSPER_ET_L...
0
1
2
3
PIKE_BOYLOVEGIRLS_T41_012
PIKE_BOYLOVEGIRLS_T41_015
SAINTOGAN_PROSPER_ET_L...
SAINTOGAN_PROSPER_ET_L...


Open images folder
Analyze
FRED_PHILEMON9_024
INOUE_KYOUMEN_011
JOLIVET_BOSTONPOLICEAFF...
JOLIVET_BOSTONPOLICEAFF...
0
1
2
3
FRED_PHILEMON9_024
INOUE_KYOUMEN_011
JOLIVET_BOSTONPOLICEAFF...
JOLIVET_BOSTONPOLICEAFF...


Open images folder
Analyze
Traffic
PAR CYBORG 07 2009
A SUIVRE...
CYB_TRAFFIC_002
CYB_TRAFFIC_004
CYB_TRAFFIC_006
FOX_CHILLINTALES_T17_021
0
1
2
3
CYB_TRAFFIC_002
CYB_TRAFFIC_004
CYB_TRAFFIC_006
FOX_CHILLINTALES_T17_021


Open images folder
Analyze
LAMISSEB_LESNOEILS1_024
LUBBIN_LESBULLESDULABO...
MCCALL_ROBINHOOD_T31_020
MCCAY_LITTLENEMO_090
0
1
2
3
LAMISSEB_LESNOEILS1_024
LUBBIN_LESBULLESDULABO...
MCCALL_ROBINHOOD_T31_020
MCCAY_LITTLENEMO_090

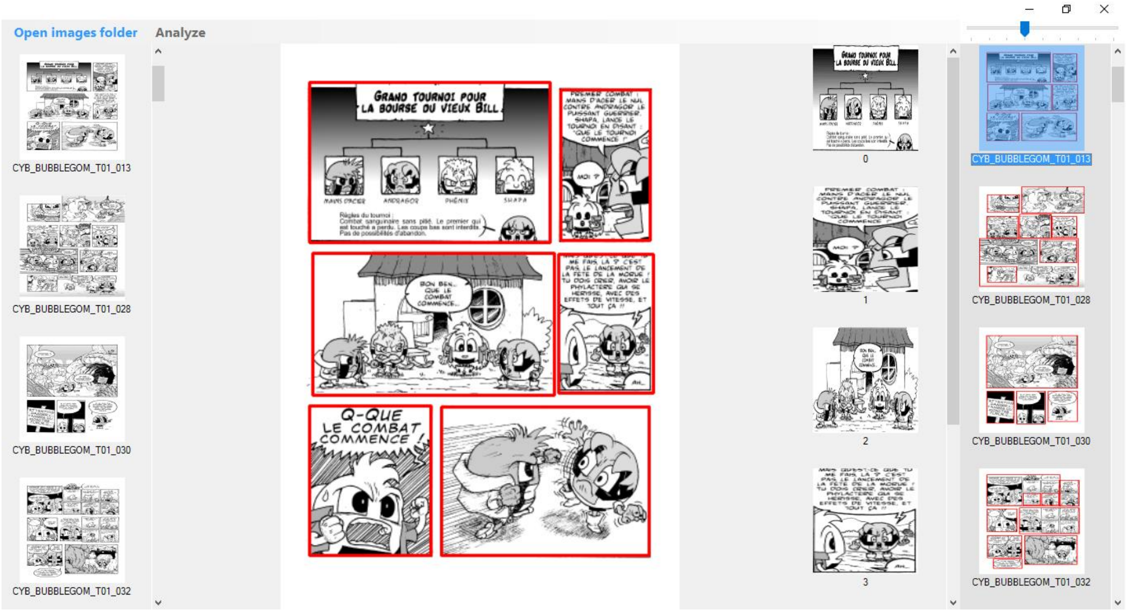
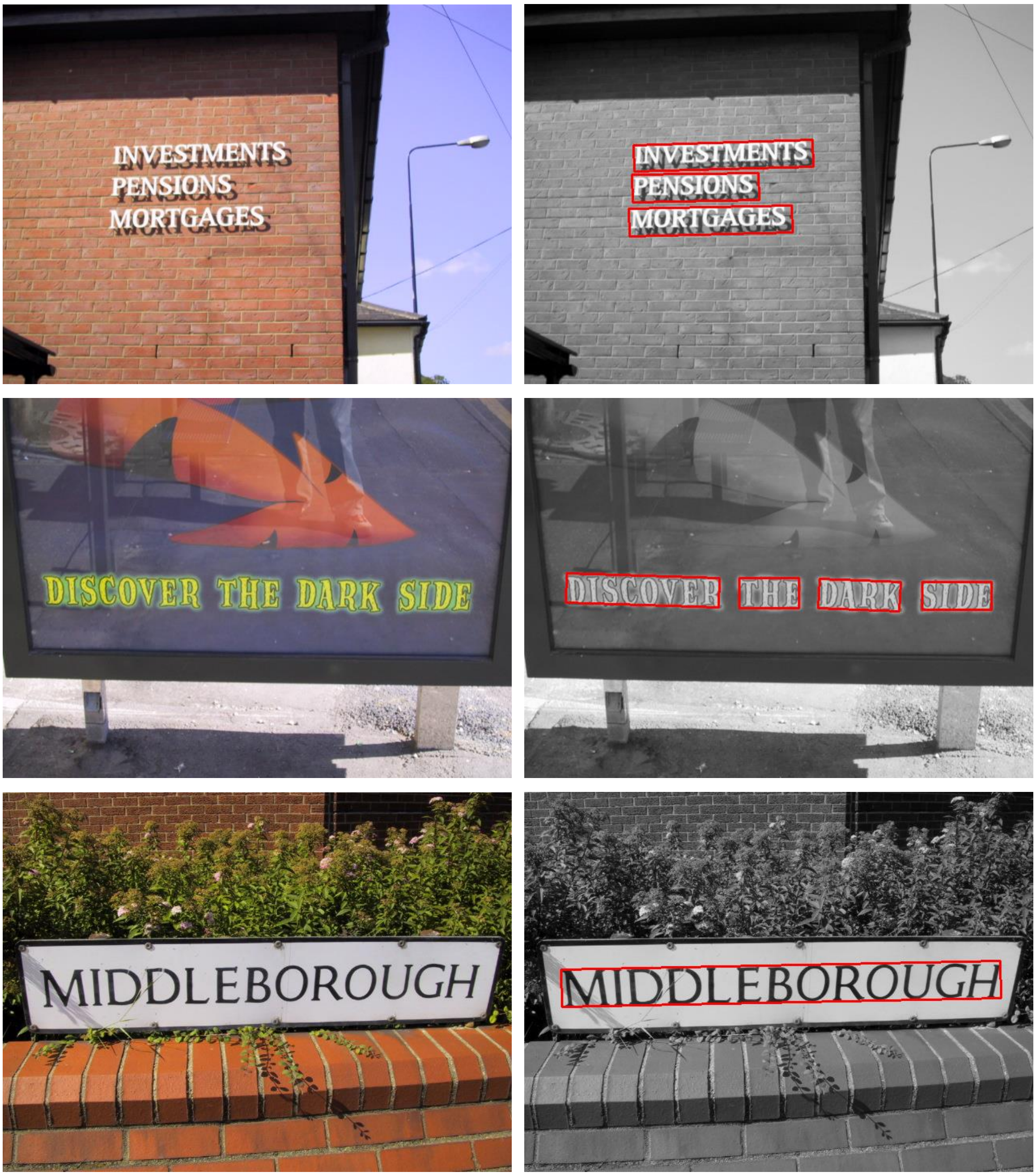


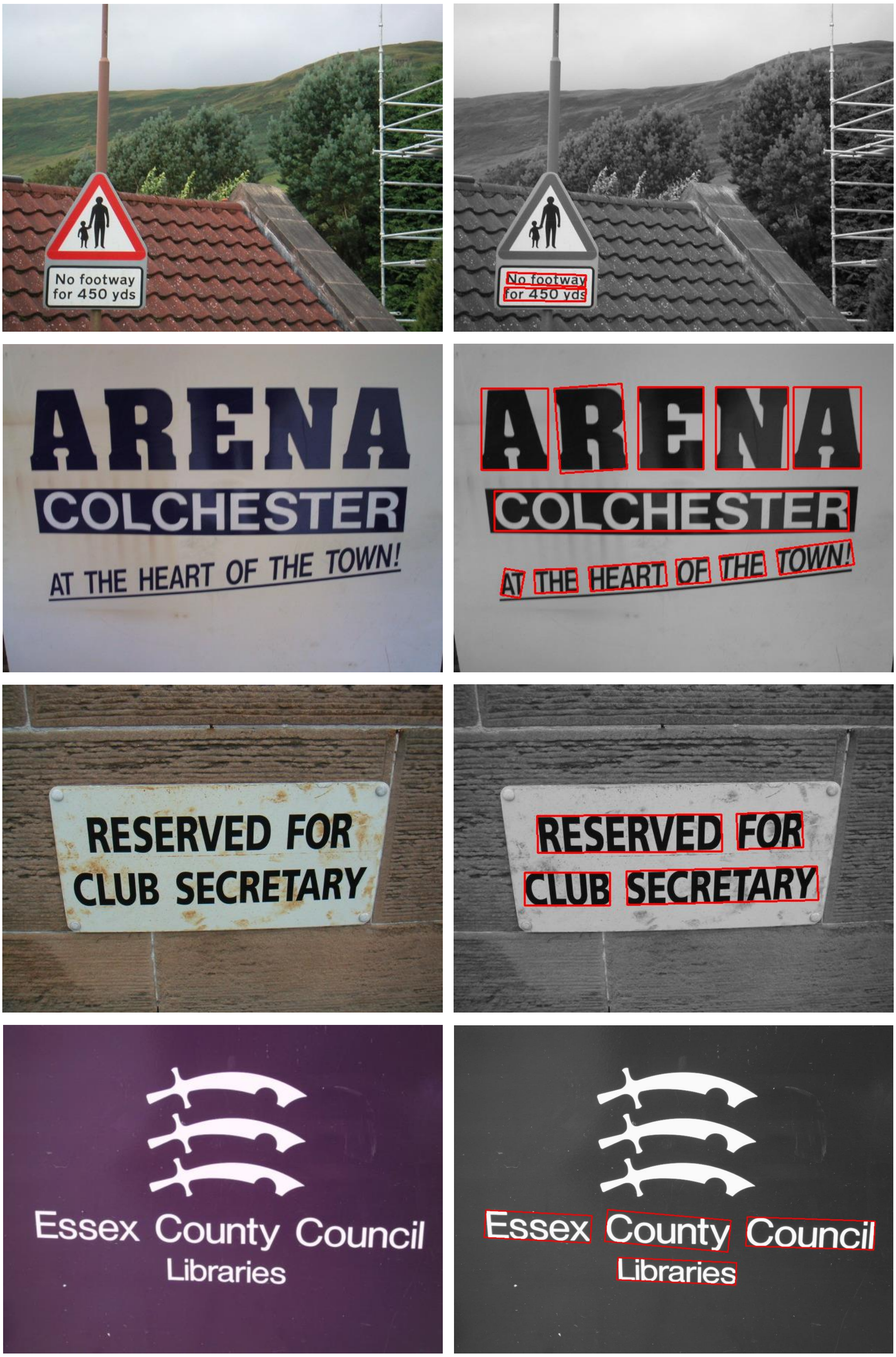
Figure I-2: Screenshots of the software

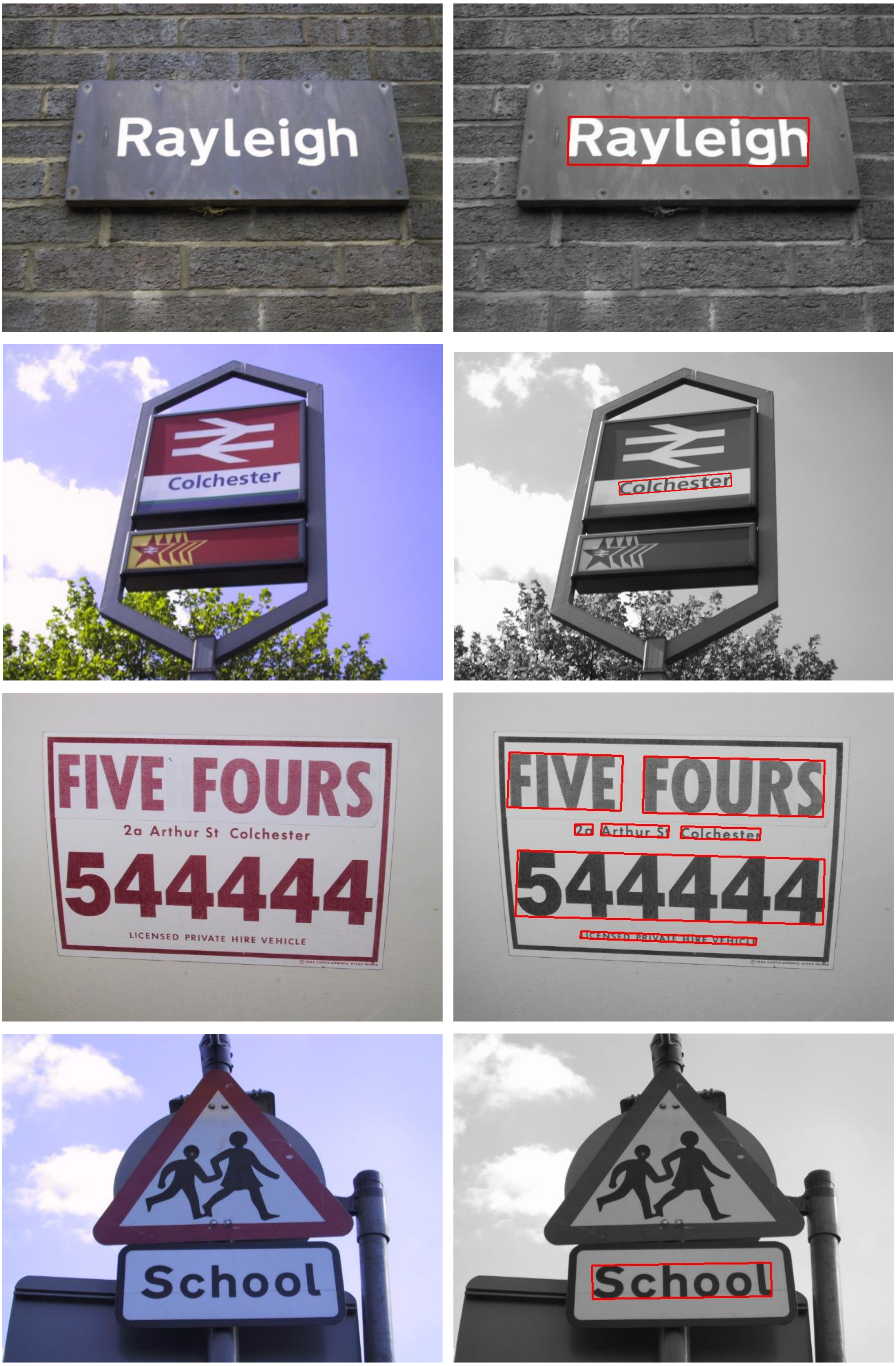
APPENDIX II – SCENE TEXT LOCALIZATION

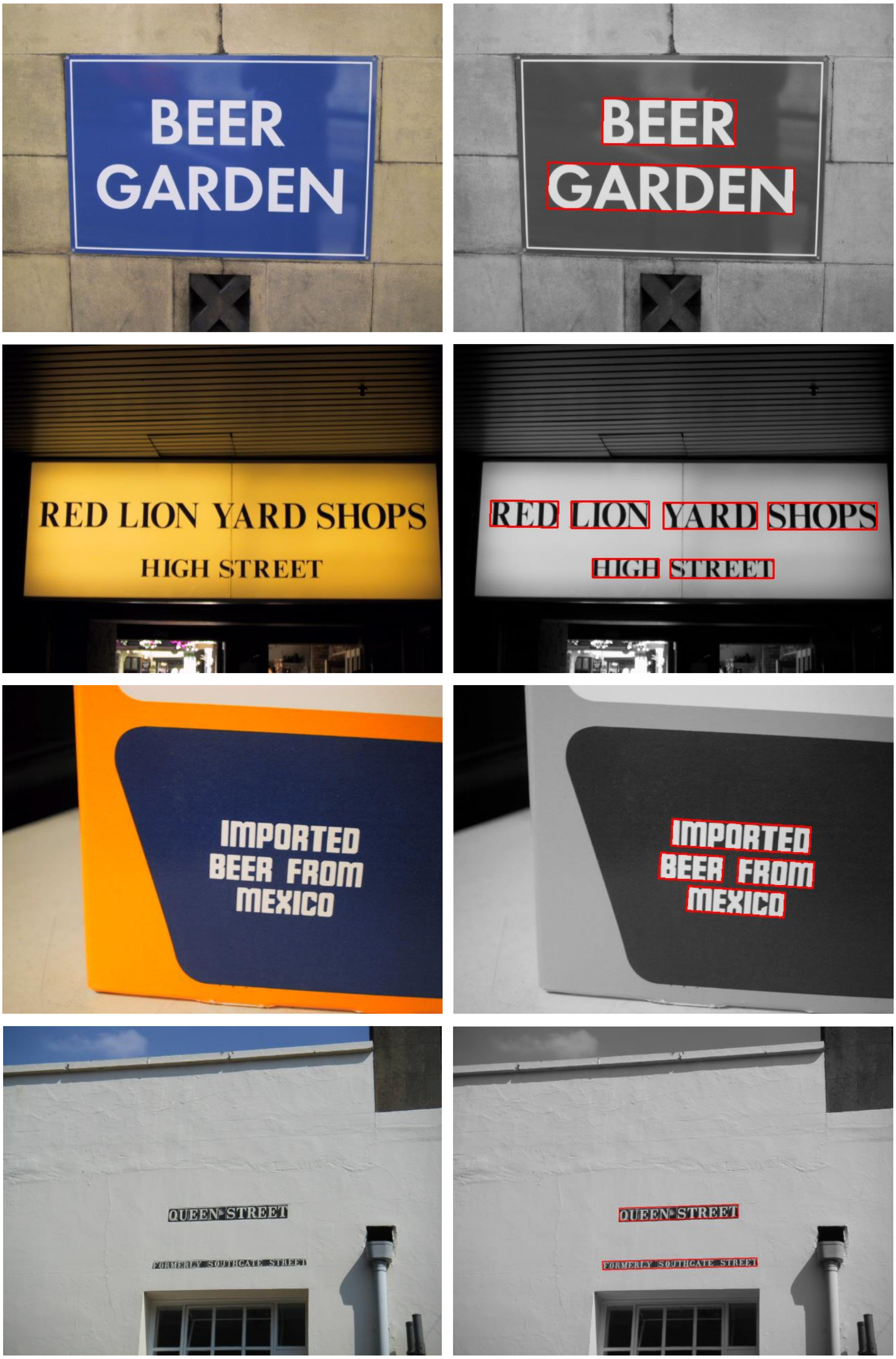
In this appendix, results from the application of the text localization method presented in Chapter 3 are shown. The proposed system is used to localize text in scene images.

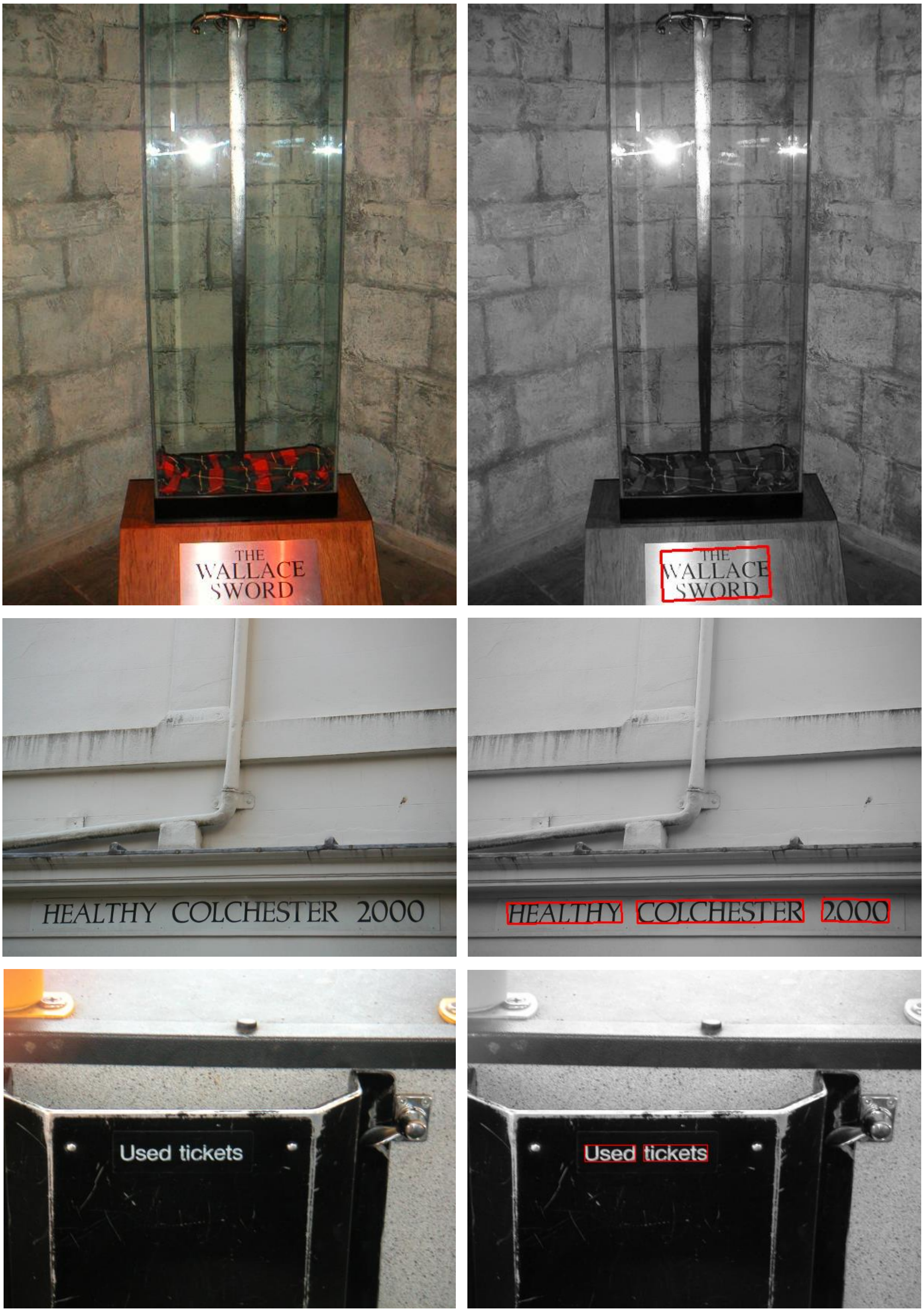
Evaluated on the challenge 2 dataset of the ICDAR 2011 Robust Reading Competition [79], the technique proved to be available for scene text localization. The following figures show results of the application of the text localization technique on various samples from the dataset.

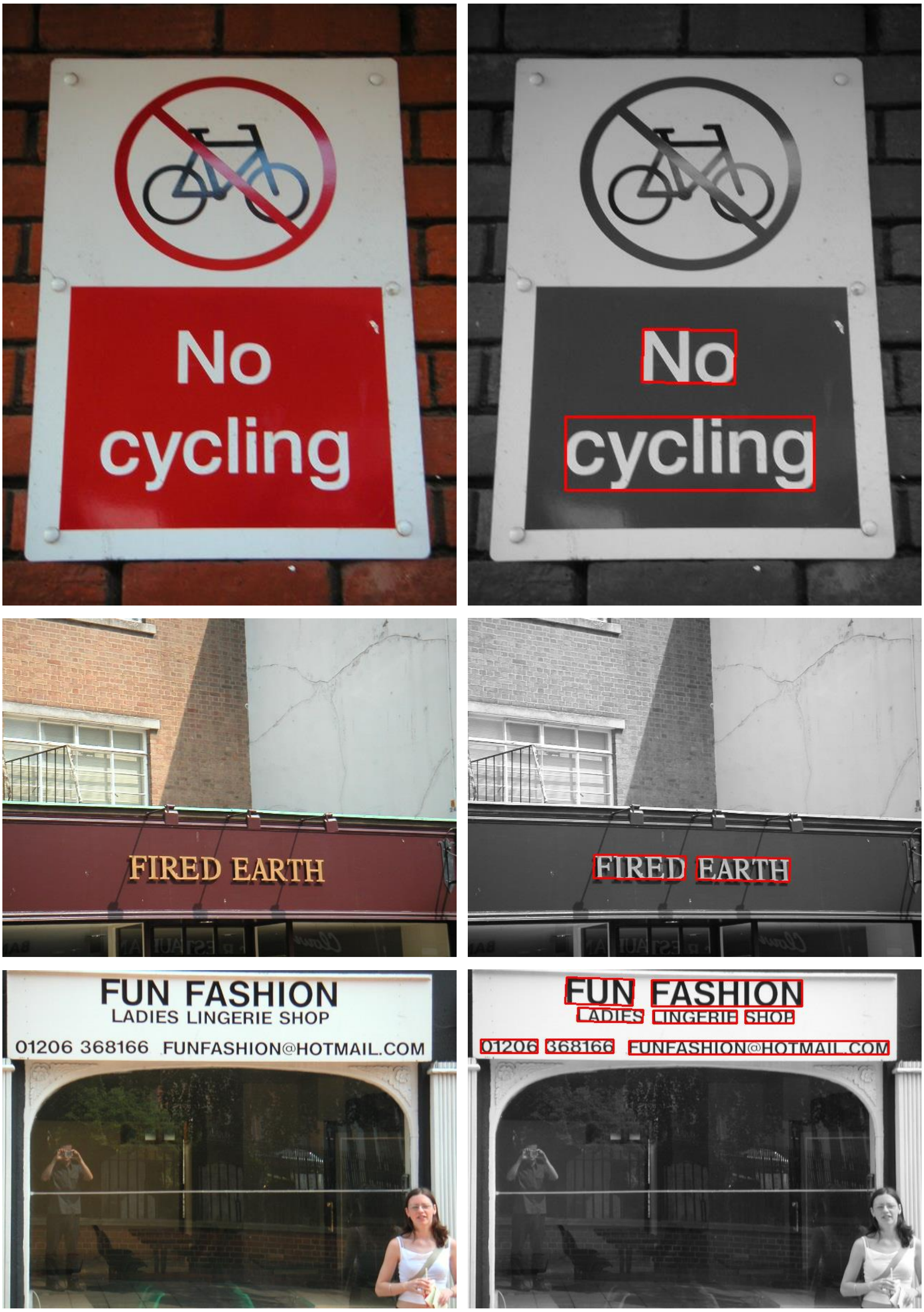


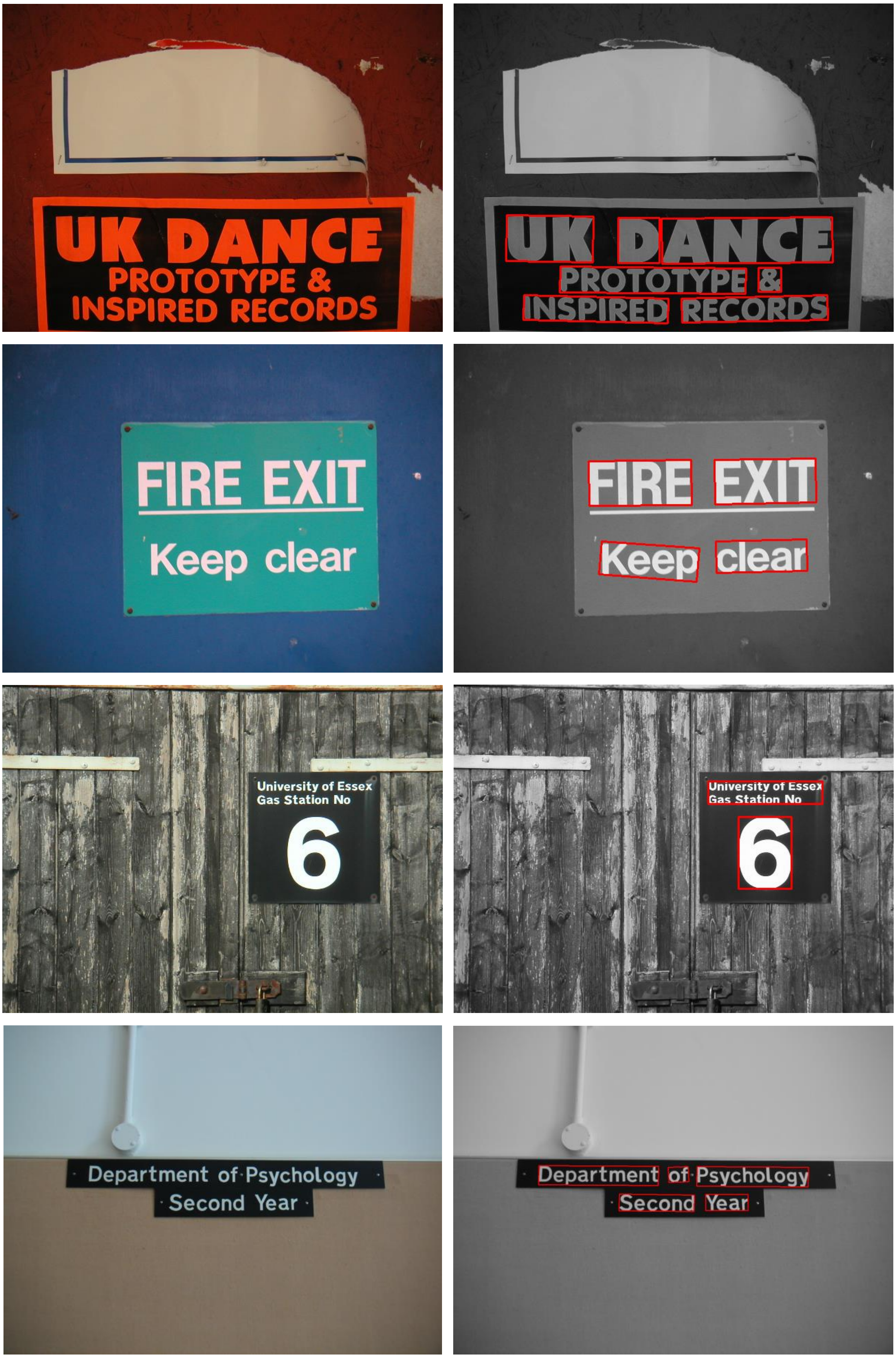

No footway
for 450 yds
ARENA
COLCHESTER
AT THE HEART OF THE TOWN!
RESERVED FOR
CLUB SECRETARY
Essex County Council
Libraries

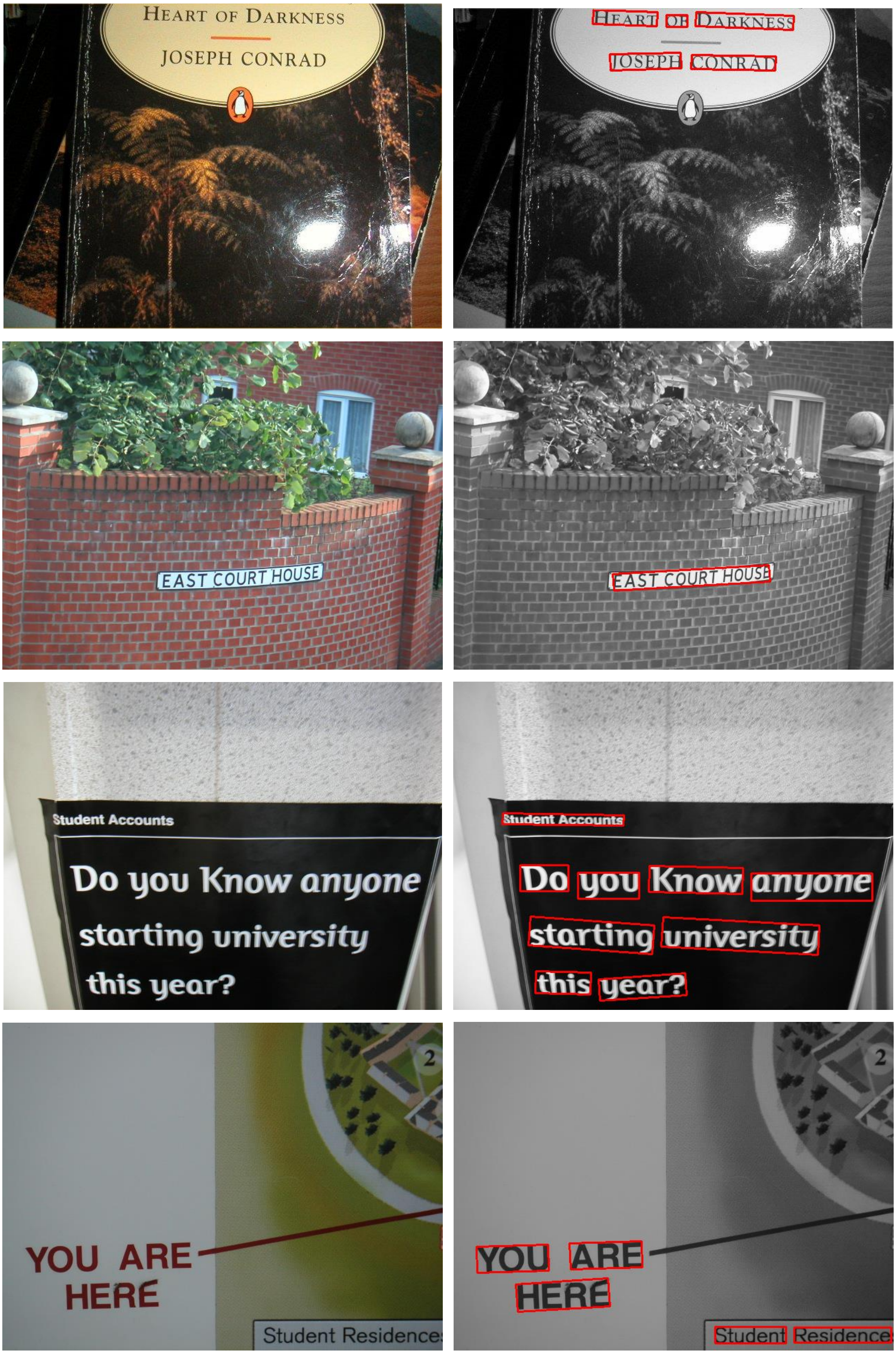

Rayleigh
Colchester
FIVE FOURS
2a Arthur St Colchester
544444
LICENSED PRIVATE HIRE VEHICLE
School


BEER
GARDEN
RED LION YARD SHOPS
HIGH STREET
IMPORTED
BEER FROM
MEXICO
QUEEN STREET
FORMERLY SOUTHGATE STREET


THE
WALLACE
SWORD
THE
WALLACE
SWORD
HEALTHY COLCHESTER 2000
HEALTHY COLCHESTER 2000
Used tickets
Used tickets


No
cycling
No
cycling
FIRED EARTH
FIRED EARTH
FUN FASHION
LADIES LINGERIE SHOP
01206 368166 FUNFASHION@HOTMAIL.COM
FUN FASHION
LADIES LINGERIE SHOP
01206 368166 FUNFASHION@HOTMAIL.COM


UK DANCE
PROTOTYPE &
INSPIRED RECORDS
FIRE EXIT
Keep clear
University of Essex
Gas Station No
6
Department of Psychology
Second Year


HEART OF DARKNESS
JOSEPH CONRAD
EAST COURT HOUSE
Student Accounts
Do you Know anyone
starting university
this year?
YOU ARE
HERE
Student Residence
2

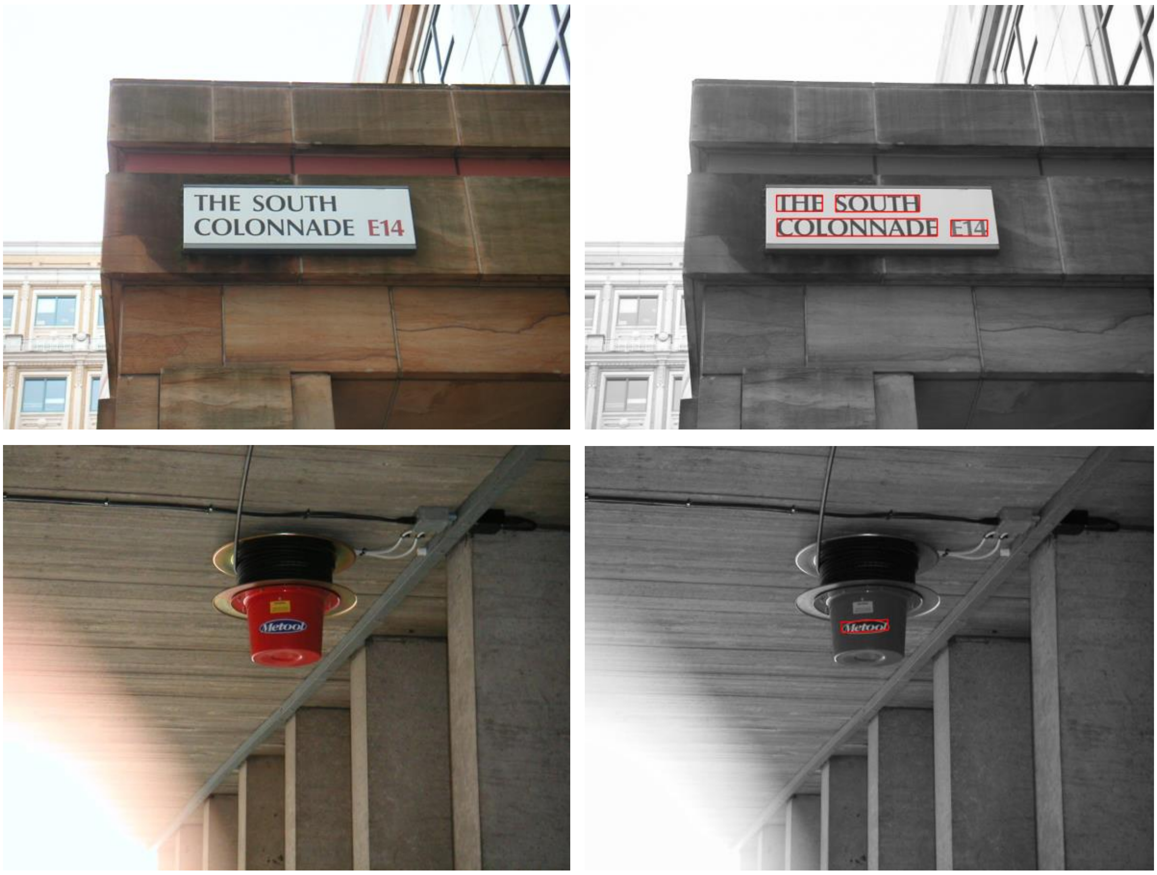


Figure II-1: Scene text localization results

APPENDIX III – VIDEO TEXT LOCALIZATION

In this appendix, results from one more application of the text localization method presented in Chapter 3 are shown. This time, the proposed system is used to localize text in video frames.

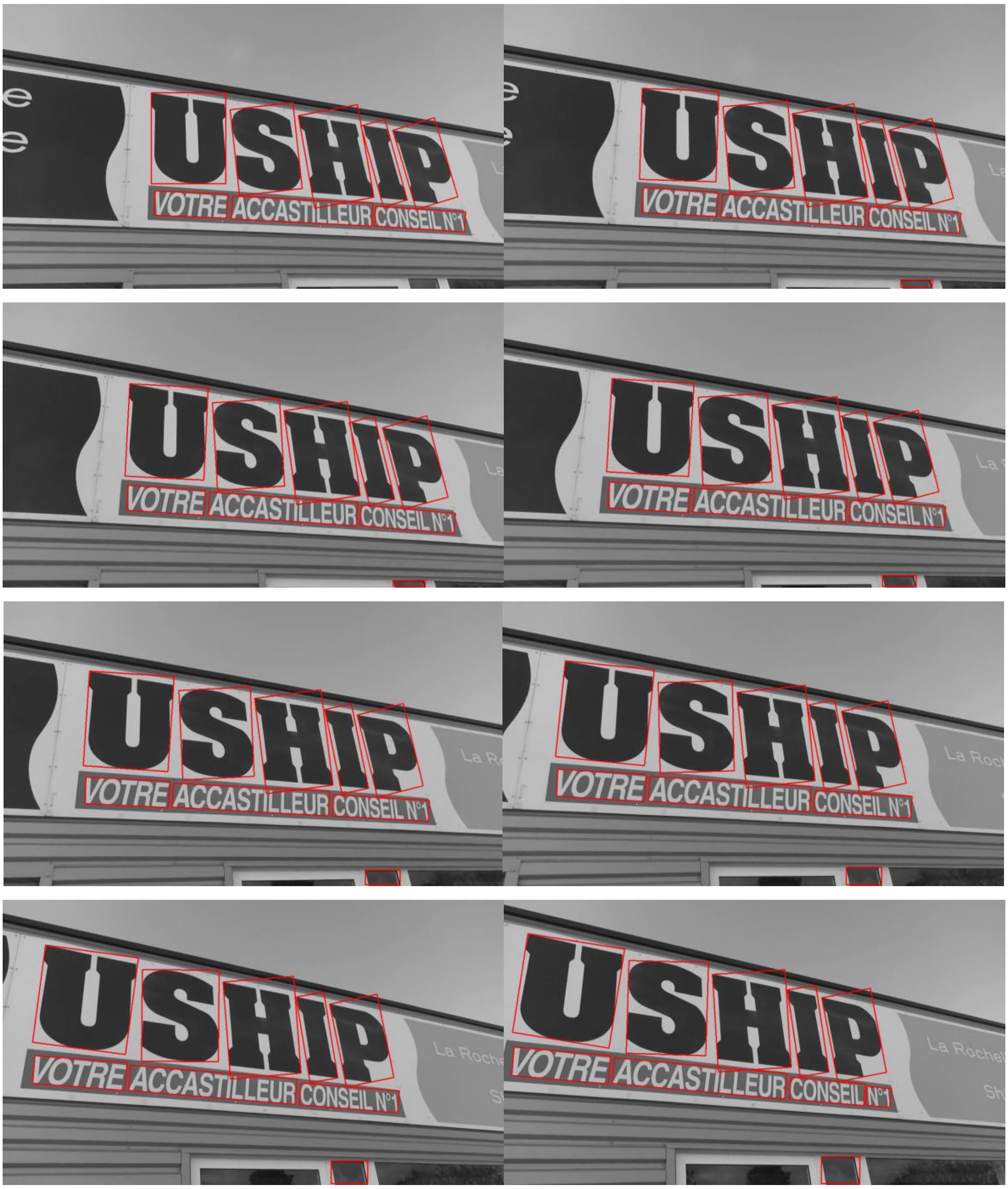


Figure III-1: Example of text localization in video captured by moving camera

Evaluated on the challenge 3 dataset of the ICDAR 2015 Robust Reading Competition [80], the technique achieved to localize most of the text in some of the videos.

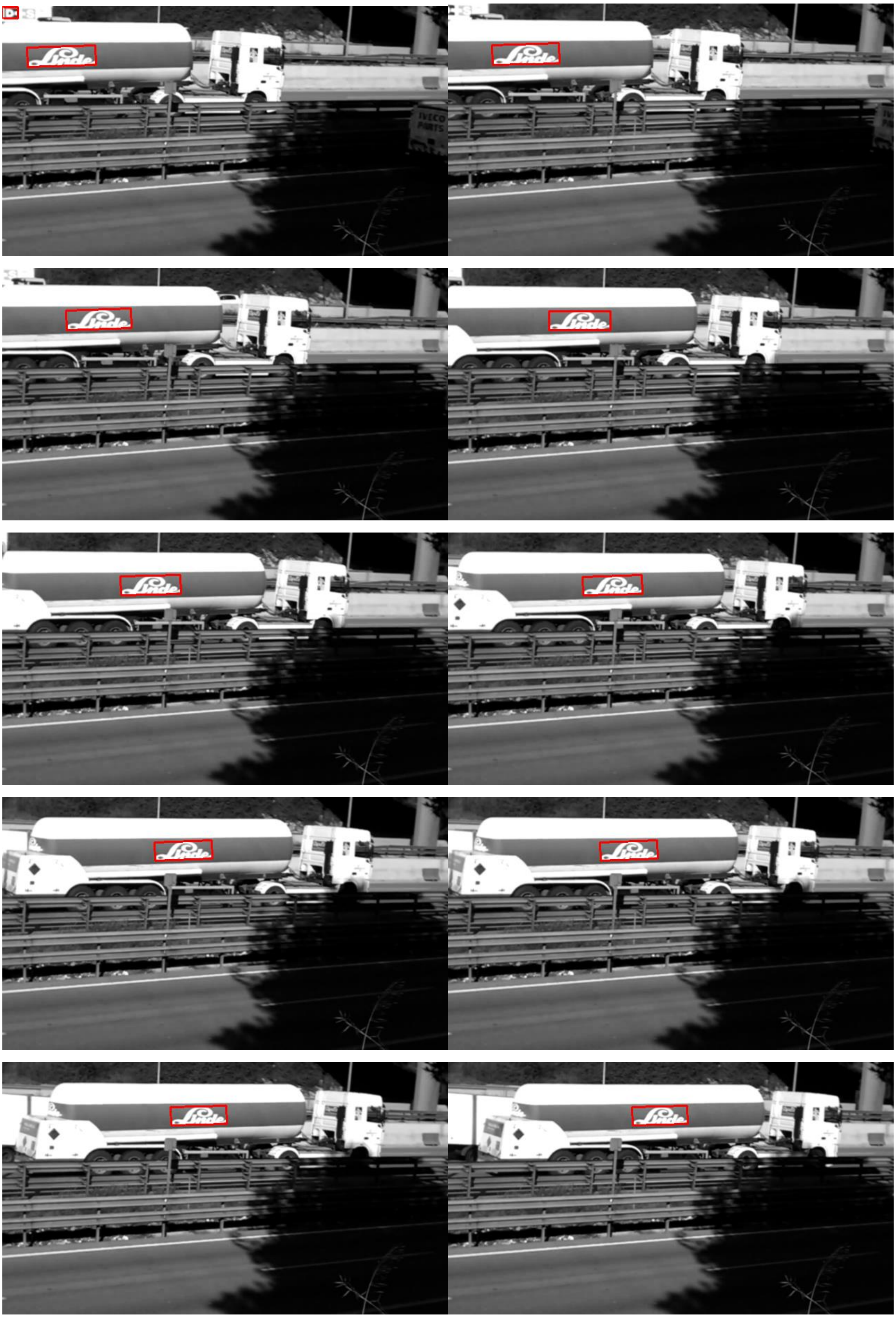


Figure III-2: Example of text localization in video captured by stable camera

Figures III.1-2 show text localization results in successive frames of two video samples from the dataset. The performance is however not so good compared to text localization in images, mainly due to the lower resolution of the video frames.

Nikos Vasilopoulos was born in Athens in 1973. He received his Diploma on Electrical Engineering and his M.Sc. on Biomedical Engineering from the University of Patras, in 1996 and 2000 respectively. He has been working as a VLSI designer in Intracom S.A., as a network operation manager in the Greek Telecommunications Organization and as a freelance software developer. The last few years, he lives in Samos island and teaches Informatics at the high school.