

Extraction of Signatures from Document Images for Real World Applications

Sheraz Ahmed^{1,2}, Muhammad Imran Malik^{1,2}, Andreas Dengel^{1,2}, and Marcus Liwicki^{2,3}

¹German Research Center for Artificial Intelligence, Kaiserslautern, Germany

²Technische Universität Kaiserslautern, Germany

³University of Fribourg, Switzerland

Various automatic methods for signature verification have been reported in the recent past. A common issue with nearly all of these methods is that they are built on the assumption that signatures are available pre-segmented or pre-extracted from document images. Accordingly, such systems are refined and tested on data containing signatures on the foreground and very little, to ideally no other, information in the background. The authors argue that these settings are not realistic and, in reality, experts encounter cases where signatures are written on documents having a lot of other information, than just the signatures, e.g., machine-printed text, ruling lines, logos, etc. We believe, to better assist forensic experts, a system should have the capability to automatically extract/segment signatures from documents like bank checks, forms, bills, wills, etc. This paper identifies that automatic signature extraction/segmentation from document images is a relevant problem faced by forensic document examiners and compares the various approaches currently available. Furthermore, we present an alternative method for extracting signatures from document images. The proposed method is capable of distinguishing machine-printed text from signatures and is based on a well known local feature descriptor, Speeded Up Robust Features (SURF). We evaluate the proposed method on the publicly available Tobacco-800 dataset in order to compare it to the currently available extraction methods. The research found that the proposed method was able to find all of the signatures in the document images tested. The authors see the application of the proposed method as a tool for document examiners to automatically extract signatures from document images.

1 Introduction

Today, automatic systems are used in almost every field to facilitate different processes in our daily life. These uses range from simple vending machines through to sophisticated systems for automatically diverting postal mail. Research in the field of document image analysis has enabled the development of various automatic systems to analyze documents on the fly and extract different types of information for different purposes, e.g., optical character recognition for sorting of postal mails based on zip code [1], automatic extraction of names, addresses, numbers and dates from documents [2]. Along with the extraction of different types of information, researchers are aiming to develop automatic systems to verify the authenticity of questioned documents [3–5].

Signatures are a widely used authentication mechanism in many industries such as banking

and law [6]. In the last three decades researchers have developed various offline (using only spatial information, such as scanned signature images) and online (using both spatial and temporal/dynamic information) signature verification systems. However, in most of the existing systems, it is assumed that signatures are readily available (segmented/extracted) from the document that originally contained them. In addition, existing publicly available datasets for the development and evaluation of signature verification systems also contain only signatures and not the complete documents containing signatures and/or other information. Contrary to this, in reality, signatures are contained on documents which also contain other background information. Such documents include bank checks, invoices, contracts, credit card pay slips and wills. In these real world scenarios, due to the lack of segmentation, the ex-

isting signature verification and identification systems cannot be used as is. This paper focuses on the challenges faced by researchers when developing a complete automatic document analysis system that is capable of performing signature segmentation/extraction from documents and then performing signature verification.

The remainder of this paper is organized into the following six sections labelled 2 through 7. Section 2 provides an overview of the systems currently used for segmentation of signatures from document images. Section 3 provides details on the proposed alternative method for signature segmentation. Section 4 presents details about the dataset used for evaluation of the segmentation method. Section 5 reports on the evaluation protocols and results of the proposed signature segmentation method. Section 6 discussed areas requiring further research and Section 7 concludes this paper.

2 Existing Segmentation Systems

To date, there has been very little research in the area of segmentation and extraction of signatures from documents, especially from document images. However, some research has been undertaken that focused on the segmentation of handwritten text from machine-printed text. In this section we will provide an overview of both cases, i.e., the existing systems available for segmentation of handwriting from machine-printed text, and those used for the segmentation of signatures from documents and bank checks.

2.1 Printed vs. Handwritten Text Segmentation

Imade et al. [7] proposed a method for segmentation and classification of printed characters, handwritten characters, photographs, and painted image regions using feed-forward neural networks. Similarly, Kuhnke et al. [8] proposed a classification system which reads a raster image of a character and outputs confidence values for machine-written and hand-written character

classes. Features from machine-written and handwritten text are extracted and passed through a feed forward neural network to obtain the confidence score. Guo and Ma [9] addressed the problem of separating handwritten annotations from machine-printed text within a document. Here Hidden Markov Models (HMMs) were used to distinguish between machine-printed and handwritten materials. Zheng et al. [10] proposed a system for detection of machine-printed and handwritten text in documents containing background noise. To achieve this, a Fisher classifier⁴ initially separates the machine-printed and handwritten text from background noise. After this noise filtering process, Markov Random Field (MRF⁵) is used to segment the machine-printed text from handwritten text. Peng et al [12] used a two step approach to separate handwriting from machine-printed annotated text using MRF. In the first step, patches/blocks of machine-printed, handwritten, and overlapped text (handwritten on machine printed) were extracted from the entire document using G-means⁶. In the second step, MRF based relabeling was performed to separate overlapped text into machine-printed and handwritten text using shape context based pixel level features. Similarly, Chanda et al. [14] used a chain-code feature with a Support Vector Machine (SVM⁷) classifier for segmentation of machine-printed text from handwritten text. Mozaffari and Bahar [16], Banerjee and Chaudhuri [17], and Banerjee [18] proposed systems for the segmentation of handwritten from machine-printed text in Farsi/Arabic and Bangla, respectively. More recently, Awal et al. [19] proposed a method for handwritten and machine-printed text separation using pseudo-lines and contextual relabeling.

2.2 Signature Segmentation from Bank-checks

Jayadeven et al. [20] undertook a comprehensive survey of the systems used for processing of bank checks. However, this survey does not cover the area of proposed signature segmentation methods for bank checks. On the other hand, research by Djeziri et al. [21] and Madasu et al. [22] did

⁴Fisher classifier is a well known linear discriminant classifier. For further details, please refer to https://en.wikipedia.org/wiki/Linear_discriminant_analysis

⁵MRF is a well known graphic model for joint probability distribution. For further details, please refer to https://engineering.purdue.edu/~bouman/publications/tutorials/mrf_tutorial/view.pdf and [11]

⁶G-means is a modified version of the well known K-means clustering algorithm. For further details, please refer to [13]

⁷SVM is a well known and commonly used classification method. For further details, please refer to [15].

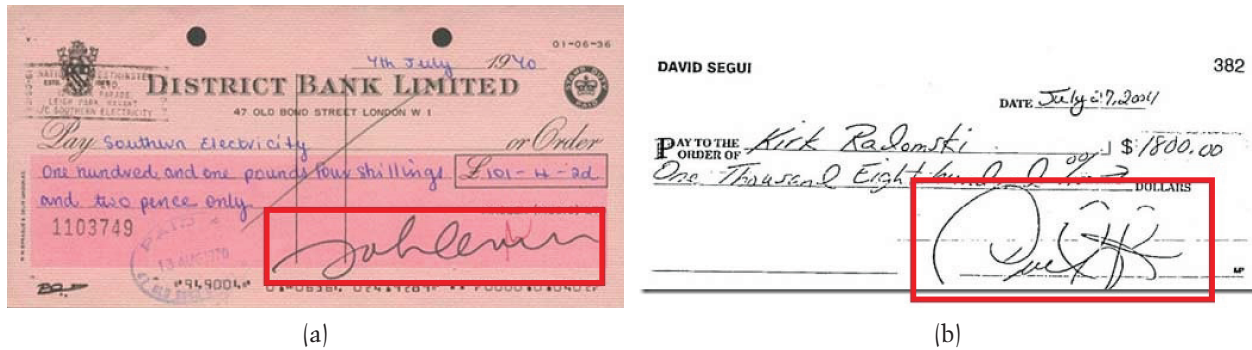


Figure 1. Bank check images [24]

specifically propose methods for extraction of signatures from bank checks. Djeziri et al. [21] proposed a method inspired from human visual perception that is based on filiformity criteria. Using these criteria, contour lines of objects are differentiated from handwritten lines. Madasu et al. [22] used the sliding window method to calculate entropy and to fit the window to signature blocks. Sankari et al. [23] proposed an approach for the segmentation of bank check account numbers and account holders' signatures from check images using prior knowledge about the possible location of account numbers and signatures in the Cartesian coordinate space. They used these segmented regions for training and verification by Hamming distance measure (number representing the difference between two binary strings). Note that in case of bank checks, prior information about the location of signatures is generally already available which makes the segmentation process comparatively easy. Therefore, most of the existing systems for signature segmentation can only be applied directly to bank checks if this prior information is available. In the cases where prior information regarding the particular location/position of signatures on a document is not known (as may be the case with contracts, wills, suicide notes, etc.), the application of the current automatic signature segmentation systems is still challenging.

2.3 Signature Segmentation from Document Images

There are many documents apart from bank checks that also contain signatures, e.g., contracts, invoices, and wills. Figure 2 shows that segmentation in such documents becomes more challenging as signatures can be located in different areas on the documents depending upon the

document's content. To deal with these situations, Zhu et al. [25] proposed a method for segmenting signatures from a complete document using a saliency map. In addition to a signature segmentation method, Zhu et al. [25] also introduced a publicly available dataset called Tobacco-800 which consists of complex document images containing information about signatures on printed text documents. Along with other information, Tobacco-800 contains patch level information (about which blocks contain signatures, without specific marking of signature pixels/strokes so that signatures overlapping with machine-printed text are not fine separated) for 900 signatures on complete documents. A complete document retrieval system is presented by Zhu et al. [26] where signature matching is combined with the detection framework discussed in the Zhu et al. 2007 [25] paper. Mandal et al. [27] proposed an approach for signature segmentation using conditional random fields. Results are reported on a subset of the Tobacco-800 dataset, i.e., 105 images out of the total 1290 documents contained in the dataset. One of the main problems with this approach is that it requires a large number of training samples. Furthermore, Mandal et al. [27] segments signatures on patch level, whereas in some cases some machine-printed text is touching the signature components which may cause problems in signature verification. To rectify this problem, Mandal et al. [28] recently proposed an alternative approach for segmentation of signatures by also segmenting those characters which are touching the signature strokes. Estaban et al. [29] detected the position of signatures in document images using the accumulative evidence technique. This method works by searching for regions that maximize the measure of correspondence (i.e., the slope of the strokes in the signature) with a given reference signature.

Mar-01-98 13:47 From: +3222470505 T-658 P-04/97 F-544
 WACHTELL, LIPTON, ROSEN & KATZ
 The Honorable John F. Kerry
 The Honorable Richard H. Bryan
 February 27, 1998
 Page 3

I certainly understand your concern about the remarks attributed to me in the newspaper. I hope that this letter clarifies what I actually said.

Very truly yours,

Meyer G. Koplow

MGR:lv
 Enclosure

cc: The Honorable John McCain, Chairman
 The Honorable Ernest P. Hollings, Ranking Member
 The Honorable John Ashcroft
 The Honorable John B. Breaux
 The Honorable Sam Brownback
 The Honorable Conrad Burns
 The Honorable Byron L. Dorgan
 The Honorable Wendell H. Ford
 The Honorable Bill Frist
 The Honorable Slade Gorton
 The Honorable Kay Bailey Hutchison
 The Honorable Daniel K. Inouye
 The Honorable Trent Lott
 The Honorable John D. Rockefeller IV
 The Honorable Olympia J. Snowe
 The Honorable Ted Stevens
 The Honorable Ron Wyden

86502300

Inq. BARZANO P. ZANARDO
 dell'Avv. Ing. Umberto Anzani di Brondello
 Roma - Milano

582, 77

ITALY

NO LEGALIZATION REQUIRED

LETTER OF AUTHORIZATION

The undersigned Philip Morris Incorporated _____
 a corporation organized under the _____
 laws of the State of Virginia, U.S.A. _____

domiciled in 100 Park Avenue, New York, New York 10017, U.S.A. _____

confirm(s) having authorized your Office since _____

to file in Italy at one of the Chambers of Commerce, Industry, Craftmanship and Agriculture or at the Central Patent office, an application for _____

and to do everything necessary for the allowance of said application and for the acknowledgment of the rights and powers derived therefrom in respect of the Ministry of Industry, Commerce and Craftmanship (presentation of documents, payments of taxes, requests for certificates and certified copies, requests for repayments and cashing of the sums repaid, filing of applications for patents of addition, modification or withdrawal of the applications, collection of granted patents and relative documents, filing and documentation and withdrawal of the appeals to the Commission established by Art. 71 of the Royal Decree of June 29, 1939, No. 1127). For which purpose the undersigned elect(s) domicile at your address, to whom therefore there may be sent directly from the Ministry of Industry, Commerce and Craftmanship any communications and requests, and may be made, if necessary, any request and delivery of documents. Dated this 30 day of September 1974.

Seal

PHILIP MORRIS INCORPORATED

By John T. Landry _____

LETTERA D'INCARICO

_____ sottoscritt _____

domiciliat _____ in _____

conferma di aver incaricato codesto Ufficio sin dal _____

di depositare in Italia presso una delle Camere di Commercio, Industria, Artigianato e Agricoltura o presso l'Ufficio Centrale Brevetti la domanda di _____

e di fare quant'altro occorra per l'accoglimento della domanda stessa e per il riconoscimento dei diritti e delle facoltà che da essa derivano nei rapporti col Ministero dell'Industria, del Commercio e dell'Artigianato (presentazione di documenti, pagamento di tasse, richieste di certificati e di copie autentiche, richieste di rimborso e riscossioni delle somme ammesse al rimborso, presentazione di domande di brevetti completivi, modificazione o ritiro delle domande, ritiro dei brevetti rilasciati e dei documenti relativi, presentazione e documentazione e ritiro dei ricorsi alla Commissione di cui all'Art. 71 del Decreto reale 29 Giugno 1939, No. 1127).

A tale scopo _____ sottoscritt _____ elegg _____ domicilio presso codesto Ufficio al quale pertanto verranno fatte direttamente dal Ministero dell'Industria, del Commercio e dell'Artigianato tutte le comunicazioni e tutte le richieste e le consegne di documenti che si rendessero necessarie.

2026385760

(a)

(b)

Figure 2. Documents having signatures at different positions.

The main problem with this method is that it requires one signature sample from each author. This limits its applicability in real world scenarios where it is not always possible to have signatures from only the already known persons, e.g., signatures on invoices received from an international supplier.

A major limitation with the majority of the above mentioned approaches of signature segmentation is that they have not been applied on real life problems. Furthermore, none of them has been tested on complete datasets such as the Tobacco-800 or any other publicly available dataset. Subsets of Tobacco-800 dataset have been used in different studies but it is not mentioned which images are included in any particular subset. This makes it difficult to analyze and compare the performance of these methods. Also for the methods tested, none of the researchers re-

ported the efficiency of their systems in terms of time and complexity.

Furthermore, some commercial systems are available. These systems are capable of finding one or two signatures in bank checks as well as (IRD⁸) images and snippets, and applying signature verification on these segmented signatures, e.g., SignatureXpert-2⁹ by Parascript. The problem with commercial systems is that the data on which they are trained are never made publicly available. Also, details about how these systems work and the algorithms applied is proprietary information.

3 Dataset

Currently, to the best of the authors' knowledge, there are two publicly available datasets that con-

⁸An Image Replacement Document (IRD) is a replacement check on paper, generated from the electronic image of an authentic paper check.

⁹<http://www.parascript.com/recognition-products/forms-processing/signaturexpert-2>.

tain information about signature patches/zones. These are the Tobbaco-800 dataset [30] and the Maryland Arabic dataset [31]. The Tobbaco-800 dataset contains 1290 images containing the handwritten and machineprinted text in English as well as machine-printed logos. There are 900 labeled signatures in this dataset. The Maryland Arabic dataset contains 169 images containing handwritten English and Arabic text along with 149 labeled signatures.

To generate results comparable to the other approaches, such as those undertaken by Zhu et al. [30], we performed evaluations of our proposed system on the Tobbaco-800 dataset. This dataset contains only the ground truth information about the logos and signatures contained in a document on the patch level (i.e., which block in the image contains the signature and which block the logos). However, if the signature is overlapping printed text in that patch, only general information about the patch is supplied. As a result, it is not possible to evaluate the individual pixels in the image to determine which pixels are a part of signature and which are part of the printed text. As mentioned in Section 1, the document analysis community has only recently started considering the problem of signature segmentation. As a result, the datasets currently used only have the patch level ground truth information about signatures available. For these datasets to be useful to the current research being undertaken, the stroke/pixel level information is also required.

To compare our method with the method proposed by Mandal et al. [27], we have used a subset of images from the Tobbaco-800 dataset containing only machine-printed text and signatures.

4 Proposed Method for Signature Segmentation

This section provides an insight into the authors' proposed method for signature segmentation from document images. This method is based on local features, i.e., Speeded Up Robust Features (SURF¹⁰). SURF is a part based approach that rep-

resents the image as a set of key points. As part based approaches extract key points/points of interest from the parts of image (which represent local features), it brings robustness against different variations in the image [34, 35].

For each of the SURF key points, a 128 bit descriptor is extracted which represents that particular key point. This descriptor is used to find similarities between different parts of the image. For extraction of SURF, we used a Hessian threshold of 400, i.e., all the key points having a Hessian threshold of less than 400 were ignored. This filtering removes unimportant features from the images.

For training purposes¹¹, we used ten documents from the Tobaaco-800 dataset containing machine-printed text and signatures. To ensure the performance of the system, it is required that all the machine-printed text is separated from the signatures. As the Tobbaco-800 dataset does not include ground truth information for machine-printed text, the authors manually generated two new images for each document dividing it into the printed text and signatures image components. These generated images were then used for training the program. Pixels containing overlaid areas were extracted for both the printed text as well as signature image of the training set.

For all of the printed text components, the extracted printed text key points and their respective descriptors were added to a printed text features database. Similarly, for all of the signature components, the extracted signature key points and their respective descriptors were added to a signature features database. These two databases serve as the reference for matching the signature and printed-text features during testing. Figure 4 shows the training procedure used.

To segment a signature from a document containing both signatures as well as printed text, the connected components¹² are extracted (as shown in Figure 3). Here, the SURF features are extracted for each of the connected component (full signature component, full printed text component, and the overlapping signature and print-

¹⁰SURF is a well known computer vision method for detecting and describing important areas in images. For details, please refer to [33].

¹¹In computer science, programs are initially trained to learn the different underlying properties of the subjects. This process is called training. Contrary to this, testing is the process when a trained computer program is run on unknown data to evaluate its performance.

¹²Extraction of connected components is a well known image processing technique.

Note that connected components here do not mean the overlaying signature and machine-printed components. For details about connected components, please refer to [36].

In the meantime, I hope you and your friends are continuing to enjoy your Marlboro Lights cigarettes. We don't think you can find a better cigarette on the market. Thank you.

Sincerely,


JOHN T. Landry

Ms. Angela Ober
221 Farnham Avenue
New Haven, Connecticut 06515

JTL:bh

1005063374

Figure 3. Original document

ed text components). The descriptor of the key points are then compared with the descriptors of printed text keypoints and signature keypoints from the two reference databases. The Euclidean distance metric is used as a distance measure.

Finally for the classification of connected components, a majority voting approach is applied. If a connected component's keypoint has less Euclidean distance to the signature keypoints reference database as compared to the printed text keypoints reference database, one vote is added to the signatures class and vice versa. The pro-

cess is repeated until all of the connected components are assigned to one of the two classes (See Figure 4). Once all of the connected components are marked as printed text or signature, separate image for signature is generated. To segment the signature from the test document, the original image is cloned and bounding boxes of all connected components of printed text are filled with white color on that image, which in turn results in a segmented signature image.

As a post processing step, horizontal run length smearing is performed on the segmented signa-

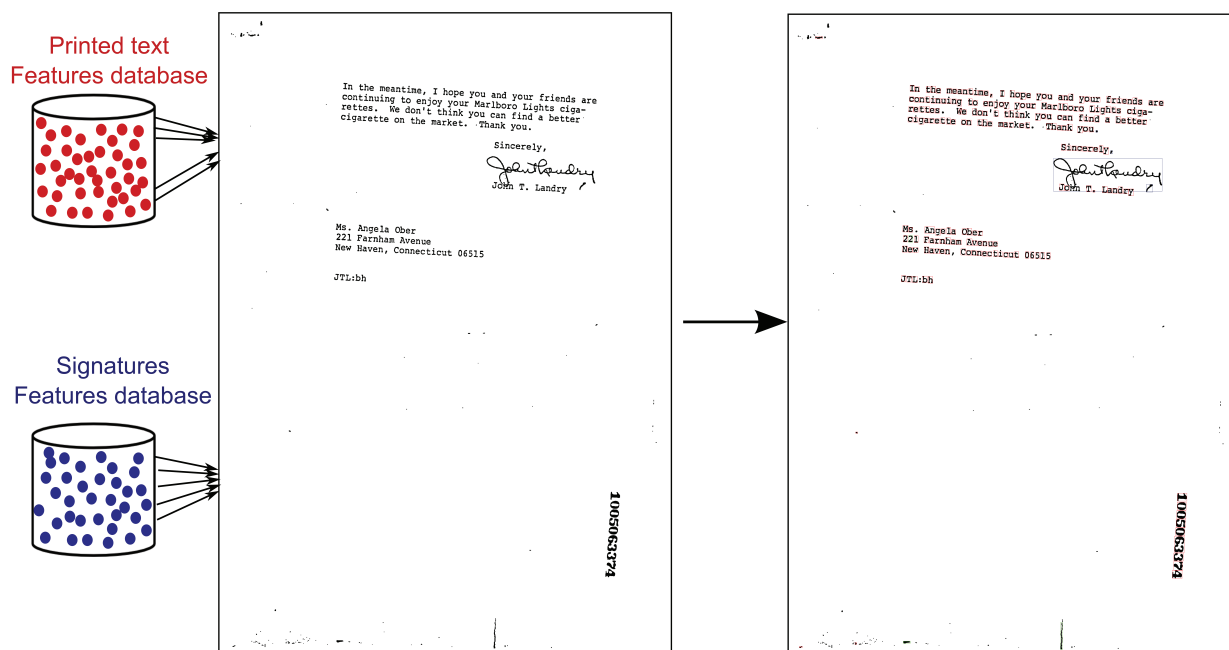


Figure 4. Extracted and marked connected components from question document image.

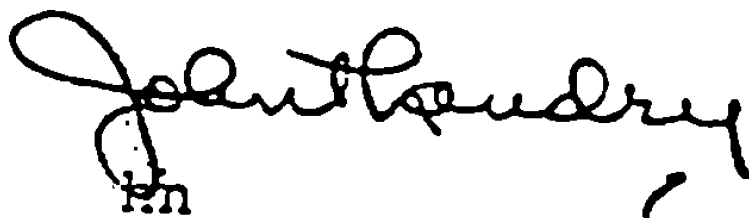


Figure 5. Extracted Signature.

ture image. Applying smearing merges all of the neighboring components. Connected components are extracted from smeared images and all of the small connected components are neglected. The remaining components are considered as signature patches. Figure 5 shows the extracted signatures from the document shown in Figure 3. One of the main advantages of our approach is that it requires a very limited number of training samples.

5 Evaluation

To evaluate the performance of the authors' proposed method the precision and recall measures are used. As mentioned in Section 3, the ground truth of the available dataset contains only patch level information of the signatures. Therefore,

the authors also calculated the precision and recall on the patch level information for the images used. The signature is considered detected if there is at least a 40% of overlap between the ground truth and the detected signature patch.

The evaluation results of the proposed method are presented in Table 1. This method has a recall rate of 100%, which means that all the signatures were extracted successfully. A minor drawback of this method is, however, that the precision is currently quite low. One reason for this is that the images containing logos were also tested. The proposed method sometimes incorrectly marked logos as signature patches during the segmentation process. The authors however consider this drawback to be removable by adding the class "logo".

Figure 7 shows some of the segmentation results of the proposed method. Qualitatively, the correctly segmented signatures are comparable to

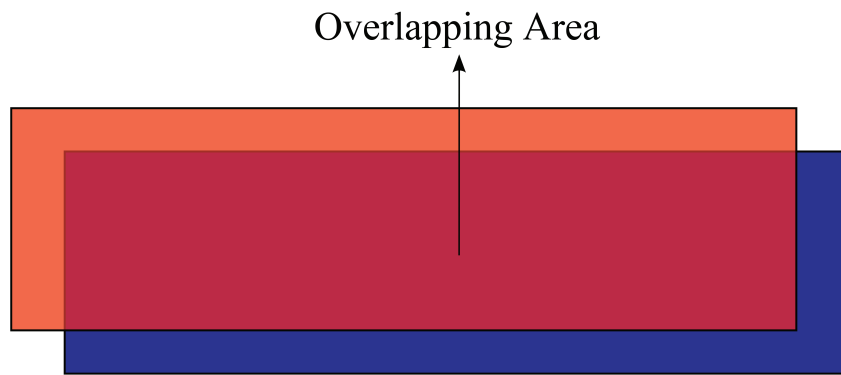


Figure 6. Overlapping area between ground truth (RED) and detected (BLUE) signature patch.



Figure 7. Examples of correctly segmented signatures (a,b) and false positives (c,d).

Table 1. Signature Segmentation results on patch level.

Method	Precision%	Recall%
Proposal method	56.52	100
Mandal et al. (105 images) [27]	not reported by authors	98.56
Cuangyu et al. [37, 30]	not reported by authors	92.8

manually cropped signatures. Figure 7 also shows some examples of false positives.

As can be seen, the proposed method preforms quite well on a difficult subset. More than every second extracted patch is a signature, with all the signatures contained in the images successfully extracted. This outcome is very positive suggesting it will be a useful technique for document examiners.

6 Scope for Future Research

In order to automate the complete document analysis process, various methods need to be developed to extract different types of information from documents in a format usable for different applications. Likewise, reliable automatic segmentation methods that can be integrated with signature verification systems need to be devel-

oped to make it a viable resource for document examiners. To develop such methods, benchmark datasets on which signature segmentation systems can be evaluated in terms of recall (how many signatures/items are extracted from the document), precision (how many of the extracted items are actually signatures) and efficiency (both in terms of speed and complexity), need to be compiled. As discussed in Section 2, currently most of the existing signature segmentation systems are evaluated on subsets of Tobacco-800 dataset and (to the best of authors' knowledge) there is no publicly available dataset specifically designed for signature segmentation. As discussed, the major disadvantage with the Tobacco-800 dataset is that it only contains patch level information about signatures, i.e., which block contains primarily signatures (meaning that some other information may also be there). This non-availability of datasets shows the current lack of research interest in signature segmentation. Since 2011, some researchers have considered this problem but still there is a lot more that needs to be done in this area.

The authors are currently working on developing a large dataset for signature segmentation and verification. Along with patch level information, this dataset will have also signature stroke information (i.e., information about each pixel that belongs to signatures or printed text) and would be usable for testing complete signature segmentation and verification frameworks for analysis of documents containing signatures. Furthermore, in this dataset multi-spectral information about signatures and other parts of documents will also be incorporated.

In addition to developing datasets specific to signature segmentation, another important area requiring research is in the development of a program capable of performing layout-free segmentation of signatures. As previously discussed, signatures are not always located in the same place on a document (as shown in Figure 2). This means that an automatic system needs to be capable of finding signatures without using any prior information about the layout of a document and probable location of a signature. Some efforts to achieve this have already been made by Man-

Mar-02-98 13:47 From: +12023470505 T-635 P.04/07 F-544

WACHTELL, LIPTON, ROSEN & KATZ

The Honorable John F. Kerry
The Honorable Richard H. Bryan
February 27, 1998
Page 3

I certainly understand your concern about the remarks attributed to me in the newspaper. I hope that this letter clarifies what I actually said.

Very truly yours,

Meyer G. Koplow

MGK:lv
Enclosure

cc: The Honorable John McCain, Chairman
The Honorable Ernest F. Hollings, Ranking Member
The Honorable Spencer Abraham
The Honorable John Ashcroft
The Honorable John B. Breaux
The Honorable Sam Brownback
The Honorable Conrad Burns
The Honorable Byron L. Dorgan
The Honorable Wendell H. Ford
The Honorable Bill Frist
The Honorable Slade Gorton
The Honorable Kay Bailey Hutchison
The Honorable Daniel K. Inouye
The Honorable Trent Lott
The Honorable John D. Rockefeller IV
The Honorable Olympia J. Snowe
The Honorable Ted Stevens
The Honorable Ron Wyden

86592300

(a)

(b)

Figure 8. Example of signatures overlaying printed text.

dal et al. [27, 28] and Ahmed et al. [38], but there is still a lot more to be done in terms of quality of extraction, precision, recall, and efficiency.

Another important aspect is to tune signature verification systems in such a way so that they are capable of distinguishing between genuine and forged signatures, even in the presence of some noise in signatures, e.g., overlapping characters or missing parts of signature. The existing signature verification systems assume that the questioned signature image contains no other information than the signature itself, which is not always the case (see Figure 2 for reference).

Figure 8 (b) shows a very common scenario where most of the existing signature verification systems will misclassify this signature as a forgery simply due to the presence of text in the signature image. This extraneous text is considered as a part of signature during verification by current automatic systems. Therefore, further development is required to tune the existing signature verification systems so as to make them more robust to noise and touching components. A probable solution could be to use only parts of the signature for verification rather than using the complete signatures [39, 40]. This, however, will raise further questions about which parts of signatures are to be used and why. In addition, a possibility is to use hyper-spectral imaging technology for distinguishing the signatures from background noise. Hyper-spectral imaging provides wavelength information about the underlying text/signatures represented from visible to near infrared regions. This information may allow the program to very accurately segment the signature from its surrounding, even at the pixel level [41].

7 Conclusions and Future Work

In this paper, state-of-the-art automatic signature segmentation/extraction methods with the potential to be integrated with verification to perform authentication were discussed. The paper outlined the limitations of the currently available automatic signature segmentation systems in real world scenarios. The main drawback of the current systems is that documents often contain information other than just the signatures, e.g., background text, lines, and logos. In order to authenticate these documents by performing signature verification, the signature must firstly be segmented and extracted from the document. The paper assesses the various approaches that

have been proposed for signature segmentation from document images and outlines their limitations. The paper also notes the current lack of databases suitable for the development and testing of complete document authentication systems involving signature segmentation and verification.

Furthermore, the authors have presented a part-based method for the extraction of signatures from documents based on the SURF key points method. The proposed method requires very limited training as compared to other contemporary approaches discussed in this paper. The experiments were performed on the Tobacco-800 dataset where all of the available signatures were successfully extracted. However, some false positives were also detected mostly due to presence of other class information in the images such as logos. This limitation can be overcome by extending the method to identify other classes and to use advanced classifiers such as SVM. By incorporating these modifications into the system, the precision of the method will be vastly increased. Moreover, using the approach of removing the less distinctive features from all the classes (proposed by Ahmed et al. [42]) will hopefully further improve the results.

It is anticipated that the authors' dataset containing signature patch and stroke level information will be completed in the near future. Once completed, it will be made publicly available. Emphasis will be given to incorporating examples of signatures overlapping text and graphics in the dataset to make it more reflective of real world situations. The patch information will be recorded in both the visible and near infrared spectra as recommended by Khan et al. [41]. Further, testing the methods proposed by Ahmed et al. [42] for separation of overlapping text/graphics to finely extract the signature strokes is also planned.

References

1. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, Dec 1989.
2. Hamish Cunningham, Diana Maynard, and Kalina Bontcheva. *Text Processing with GATE*. Gateway Press CA, 2011.
3. Marcus Liwicki and Muhammad Imran Malik. Surprising? power of local features for automated

- signature verification. In 15th International Graphonomics Society Conference (IGS2011), pages 18–21, Live Aqua Cancun, Mexico, June 2011.
4. Muhammad Imran Malik, Sheraz Ahmed, Andreas Dengel, and Marcus Liwicki. A signature verification framework for digital pen applications. In 10th IAPR International Workshop on Document Analysis Systems (DAS), pages 419–423. IEEE, 2012.
 5. M. I. Malik, Marcus Liwicki, Andreas Dengel, Seiichi Uchida, and Volkmar Frinken. Automatic signature stability analysis and verification using local features. In 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 621–626. IEEE, 2014.
 6. M. I. Malik, Marcus Liwicki, Andreas Dengel, and Bryan Found. Man vs. machine: A comparative analysis for forensic signature verification. In 16th Biennial Conference of the International Graphonomics Society (IGS), pages 9–13. International Graphonomics Society, 2013.
 7. S. Imade, S. Tatsuta, and T. Wada. Segmentation and classification for mixed text/image documents using neural network. In 2nd International Conference on Document Analysis and Recognition (ICDAR), pages 930–934, Oct 1993.
 8. K. Kuhnke, L. Simoncini, and Zs.M. Kovacs-V. A system for machine-written and hand-written character distinction. In 3rd International Conference on Document Analysis and Recognition (ICDAR), volume 2, pages 811–814, Aug 1995.
 9. J.K. Guo and M.Y. Ma. Separating handwritten material from machine printed text using hidden Markov models. In 6th International Conference on Document Analysis and Recognition (ICDAR), pages 439–443, 2001.
 10. Yefeng Zheng, Huiping Li, and D. Doermann. Machine printed text and handwriting identification in noisy document images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(3):337–353, Mar 2004.
 11. Anand Rangarajan, Rama Chellappa, and Anand Rangarajan. Markov random field models in image processing. The Handbook of Brain Theory and Neural Networks. Pages 564–567, MIT Press, 1995.
 12. Xujun Peng, Srirangara j Setlur, Venu Govindaraju, and Ramachandruna Sitaram. Handwritten text separation from annotated machine printed documents using Markov random fields. International Journal on Document Analysis and Recognition (IJDAR), 16(1):1–16, 2013.
 13. Greg Hamerly and Charles Elkan. Learning the k in k-means. Advances in neural information processing systems 16:281–288, MIT Press, 2004.
 14. Sukalpa Chanda, Katrin Franke, and Umapada Pal. Structural handwritten and machine print classification for sparse content and arbitrary oriented document fragments. In Symposium on Applied Computing (SAC’ 10), pages 18–22. ACM, 2010.
 15. Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov., 2(2):121–167, June 1998.
 16. Saeed Mozaffari and Parnia Bahar. Farsi/Arabic handwritten from machine-printed words discrimination. In 13th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 698–703. IEEE, 2012.
 17. Purnendu Banerjee and Bidyut Baran Chaudhuri. A system for hand-written and machine-printed text separation in Bangla document images. In 13th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 758–762. IEEE, 2012.
 18. Sandipan Banerjee. Identification of handwritten text in machine printed document images. In Natarajan Meghanathan, Dhinaharan Nagamalai, and Nabendu Chaki, editors, Advances in Computing and Information Technology, volume 177 of Advances in Intelligent Systems and Computing, pages 823–831. Springer Berlin Heidelberg, 2013.
 19. Ahmad Montaser Awal, Abdel Belaid, and Vincent Poulain d’Andecy. Handwritten/printed text separation using pseudo-lines for contextual re-labeling. In 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 29–34, 2014.
 20. R. Jayadevan, S.R. Kolhe, P.M. Patil, and U. Pal. Automatic processing of handwritten bank cheque images: a survey. International Journal on Document Analysis and Recognition, 15:267–296, 2012.
 21. S. Djeziri, F. Nouboud, and R. Plamondon. Extraction of signatures from check background based on a filiformity criterion. IEEE Transactions on Image Processing, 7(10):1425–1438, Oct 1998.

22. Vamsi Krishna Madasu, M Hafizuddin, Mohd Yusof, M. Hanmandlu, and Kurt Kubik. Automatic extraction of signatures from bank cheques and other documents. In 7th International Conference on Digital Image Computing: Techniques and Applications (DICTA), pages 591–600, 2003.
23. M. Sankari, M. Benazir, and R. Bremananth. Verification of bank cheque images using hamming measures. In 11th International Conference on Control Automation Robotics and Vision (ICARCV), pages 2531–2536, Dec 2010.
24. Historical and interesting bank checks. <http://www.smartmoneydaily.com/personalfinance/15-historical-interesting-bank-checks.aspx>. Accessed: June 15, 2015.
25. Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Multi-scale Structural Saliency for Signature Detection, pages 1–8. Minneapolis, MN, 2007.
26. Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Signature detection and matching for document image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2015–2031, 2009.
27. R. Mandal, P. Pratim. Roy, and U. Pal. Signature segmentation from machine printed documents using conditional random field. In 11th International Conference on Document Analysis and Recognition (ICDAR), pages 1170–1174. 2011.
28. Ranju Mandal, Partha Pratim Roy, and Umapada Pal. Signature segmentation from machine printed documents using contextual information. *International Journal on Pattern Recognition and Artificial Intelligence*:1–25. 2012.
29. José L Esteban, José F Vélez, and Ángel Sánchez. Off-line handwritten signature detection by analysis of evidence accumulation. *International Journal on Document Analysis and Recognition (IJДАР)*, 15(4):359–368, 2012.
30. Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Multi-scale structural saliency for signature detection. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.
31. Yi Li, Yefeng Zheng, David S. Doermann, and Stefan Jaeger. Script-independent text line segmentation in freestyle handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1313–1329, 2008.
32. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, Jun 2008.
33. Duy-Nguyen Ta, Wei-Chao Chen, Natasha Gelfand, and Kari Pulli. Surftrac: Efficient tracking and continuous object recognition using local feature descriptors. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2937–2944. IEEE, 2009.
34. Wang Song, Seiichi Uchida, and Marcus Liwicki. Comparative study of part-based handwritten character recognition methods. In 11th International Conference on Document Analysis and Recognition (ICDAR), pages 814–818, 2011.
35. Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing* (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
36. Guangyu Zhu, Yefeng Zheng, David Doermann, and Stefan Jaeger. Signature Detection and Matching for Document Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2015–2031, November 2009.
37. Sheraz Ahmed, Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Signature segmentation from document images. In 13th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 423–427. IEEE, 2012.
38. Muhammad Imran Malik, Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel. Freak for real time forensic signature verification. In 12th International Conference on Document Analysis and Recognition (ICDAR), pages 971–975. IEEE, 2013.
39. Muhammad Imran Malik, Marcus Liwicki, and Andreas Dengel. Part-based automatic system in comparison to human experts for forensic signature verification. In 12th International Conference on Document Analysis and Recognition (ICDAR), pages 872–876. IEEE, 2013.
40. Zohaib Khan, Faisal Shafait, and Ajmal Mian. Hyper-spectral imaging for ink mismatch detection. In 12th International Conference on Document Analysis and Recognition (ICDAR), pages 877–881. IEEE, 2013.
41. Sheraz Ahmed, Marcus Liwicki, and Andreas Dengel. Extraction of text touching graphics using surf. In 10th IAPR International Workshop on Document Analysis Systems (DAS), pages 349–353, Mar 2012.
42. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and