Verification of Authenticity of Stamps in Documents

Barbora Micenková*
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, 612 66 Brno, Czech Republic
barbora@cs.au.dk

Abstract

Classical ink stamps and seals used for authentication of a document content have become relatively easy to forge by the scan&print technique since the technology is available to general public. For environments where a huge volume of documents is processed, an automatic system for verification of authenticity of stamps has been developed in the scope of this work. The process of stamp authenticity verification naturally must be preceded by the phase of stamp detection and segmentation – a difficult task of Document Image Analysis. In this work, a novel method for detection and verification of stamps in color document images is proposed. It involves a full segmentation of the page, extraction of features and further classification of the segments by means of Support Vector Machines. The evaluation has shown that the algorithm is capable of differentiating stamps from other color objects in the document such as logos or text and also genuine stamps from copied ones.

Categories and Subject Descriptors

I.4.6 [**Image Processing and Computer Vision**]: Segmentation; I.5.4 [**Pattern Recognition**]: Applications

Keywords

document image analysis, document security, image segmentation, stamp detection, computational forensics

1. Motivation

A huge volume of documents is processed daily in offices such as insurance companies or banks and the degree of automation is still increasing. For example, the printed invoices incoming to an insurance company are immediately digitized and the originals are no longer stored [5]. Together with the fact that, nowadays, an advanced computer technology (such as laser color printers and photocopiers) is accessible to the general public, it makes it easier to commit a fraud by a *"modern" forgery*. It means that the document, e.g. an invoice, is scanned, digitally modified and printed again, or, alternatively, some parts are photocopied from another one. In the offices, there is usually no time and resources to check visually if the document has been tampered (digitally modified). For that reason, an automatic system for detection of potentially fraudulent documents is needed and new issues in the field of *document security* have been raised.

In this work, the problem of verification of authenticity of stamps is discussed. The most simple yet efficient technique of forgery is a photocopy. Therefore, we focus on revealing photocopied (and scanned-and-printed) stamps. A new system is proposed to detect and extract stamps from color document images and, in a second step, to verify whether the stamps are authentic or forged (copied). Thus, we have decomposed the problem into two separate tasks – stamp detection and stamp verification.

Stamp detection is not a trivial task and no general solution to the problem has been given yet. Our method intends to be as generic as possible. We want to detect stamps of different shapes – official, business as well as decorative ones. They can be placed in any position in the document and they can even be overlapped with other objects. Also the quality of the imprint can be variable. We are able to detect stamps of any colors but black, however the method allows for further extension for detection of black stamps too.

2. Related Work

All previously published methods on *stamp detection* focus on stamps of a particular type. Approaches based on shape information require a frame to be present around the stamp. Zhu *et al.* [6] presented a method being robust on degraded documents and successful for stamps overlapping with text, however it can only detect elliptical stamps. Some authors, such as Ueda [4], exploit the color information for detection. He can extract stamps from a specific Japanese bank check. The latest method was proposed by Forczmański *et al.* [1] who transform the document into YC_bC_r color space and search for areas with high C_b and C_r intensities to detect blue and red stamps. Limitations of the previous methods (on shape, background and color) motivated us to propose a new, more generic approach for stamp detection.

*Master degree study programme in Artificial Intelligence. Supervisor: Assoc. Prof. Adam Herout, Faculty of Information Technology, Brno University of Technology, Czech Republic.

Micenková, B. Verification of Authenticity of Stamps in Documents. Information Sciences and Technologies Bulletin of the ACM Slovakia, Special Section on the ACM Student Project of the Year 2011 Competition, Vol. 3, No. 4 (2011) 45-46

No work on revealing stamp copies has been published yet, but there are publications in the field of *printing techniques recognition*. Several authors (e.g. [2]) have tried to classify laser-printed and photocopied text and they have introduced feature extraction techniques for that purpose. We have adopted some of the features in order to verify the authenticity of stamps.

3. Approach

The system for stamp detection and verification presented in the paper is multi-stage and learning-based. Providing images with annotated authentic stamps, the system can learn to recognize stamps in new, previously unseen documents. Furthermore, providing examples of copied stamps, the system can be finely tuned to decide about the authenticity of the newly presented stamps.

To summarize the approach briefly: The hardcopy of the document is scanned in color and a full segmentation of the image is done. Candidate solutions to the stamp detection problem are identified in the form of rectangular segments and features are extracted out of them. Binary classification is performed on the candidates to decide between stamps and non-stamps (logos, text etc.). Segments identified as stamps are further classified to distinguish between authentic and forged stamps. A more detailed description follows.

3.1 Image segmentation

Segmenting the image, we take an advantage of the fact that stamp is a plain-color object. Since we focus on color stamps so far, only chromatic parts of the image are separated and further processed. RGB color space is not convenient for segmentation, therefore we first convert the image to YC_bC_r color space.

An interesting property can be observed from 3D color histograms of various scanned document images. Clusters formed by inks always have an elongated shape stretching from the cluster corresponding to the background and they are quite narrow. This property was taken into account when designing a fast clustering algorithm to separate objects of different colors in page. More details about the algorithm can be found in our previous publication [3].

Clusters containing objects of a very similar color are treated as separate binary images that are further processed. Stamps are not connected objects, they consist of several small components, and so they need to be grouped first. The XY-cut algorithm is applied to recursively partition the page into rectangles, resulting in minimum bounding boxes of candidate solutions. The process is shown in Fig.1.

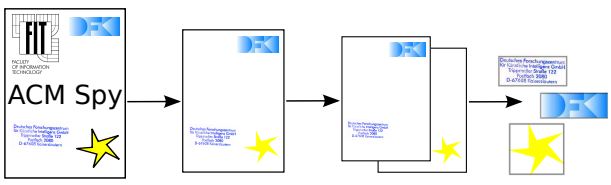

Figure 1: Segmentation algorithm.

3.2 Classification

Having the candidate solutions segmented, those containing stamps need to be differentiated from those containing other objects such as logos, color texts or parts of pictures. For that reason, three kinds of features are computed over each segment – *geometrical*, *color-related* and *print-related*. Sample geometrical features are relative area of the bounding box, width-to-height ratio or rotation of text-lines (if any are found). A color-related feature is e.g. standard deviation of hue. Print-related features concern the sharpness and roughness of edges or uniformity of the printed area.

Binary classification of the candidates is performed by means of Support Vector Machines. In the next step, candidates identified as stamps are classified as *authentic* or *forged*. A subset of color and print-related features is selected and a second classification is performed.

4. Evaluation and Results

For the purpose of evaluation we collected and annotated a new data set[1] since there was none available to public during our work on the project. 400 documents were printed, stamped, and scanned at different resolutions. A pixel-wise groundtruth had to be created manually. Apart from stamps, the documents contain black text and color objects such as logos and texts. Forged documents were obtained by photocopying 14 randomly chosen documents from the data set on 5 different copy-machines, which totally equals to 70 copies.

Cross-validation by the *leave-one-out* method was performed for stamp detection on the 400 documents, and precision as well as recall of approximately 89% (in terms of pixel accuracy) were achieved. Separately, 317 authentic and 78 forged stamps were segmented from our data sets and a *10-fold* cross validation was performed. As a result, 90% of the copied stamps and 95% of the authentic stamps were recognized correctly.

5. Conclusion

In the scope of this work, a new method for detection and extraction of stamps from color document images has been proposed. On the top of it, the first approach to verification of authenticity of stamps by revealing photocopies has been introduced.

References

[1] P. Forczmański and D. Frejlichowski. General shape analysis applied to stamps retrieval from scanned documents. In *Proc. of the 14th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications*. Springer-Verlag, 2010.

[2] N. Khanna, A. K. Mikkilineni, G. T. Chiu, J. P. Allebach, and E. J. Delp. Survey of scanner and printer forensics at Purdue University. In *Proc. of the 2nd Int. Workshop on Computational Forensics*. Springer-Verlag, 2008.

[3] B. Micenkova and J. van Beusekom. Stamp detection in color document images. In *Proc. of the 11th Int. Conf. on Document Analysis and Recognition*. IEEE, 2011.

[4] K. Ueda. Extraction of signature and seal imprint from bankchecks by using color information. In *Proc. of the 3rd Int. Conf. on Document Analysis and Recognition*, volume 2. IEEE, 1995.

[5] J. van Beusekom. *Optical Document Security in High Volume Office Environments*. PhD thesis, University of Kaiserslautern, Germany, 2010.

[6] G. Zhu, S. Jaeger, and D. Doermann. A robust stamp detection framework on degraded documents. In *Proc. of SPIE Document Recognition and Retrieval XIII*, 2006.

[1] The data set has been made public and it is available at `http://madm.dfki.de/downloads-ds-staver`.