A SYSTEM FOR VERIFICATION OF AUTHENTICITY OF STAMPS IN DOCUMENTS

Barbora Micenková
Master Degree Programme (3), FIT BUT
E-mail: xmicen00@stud.fit.vutbr.cz

Supervised by: Adam Herout
E-mail: herout@fit.vutbr.cz

Abstract: Classical ink stamps that are used to authenticate a content of a document have become relatively easy to forge with means of computer technology that is available to general public. Therefore an automatic system for verification of authenticity of stamps in documents is being developed. It is intended to be utilized in environments like banks and insurance companies where many documents are being processed every day. The task of stamp detection in scanned document images has not been reasonably solved till today. In this paper, a new method for color stamp detection and verification of its authenticity is outlined.

Keywords: stamp detection, image segmentation, forgery detection, computational forensics

1 INTRODUCTION

In various areas of our lives, computers have been enormously utilized but paper documents still play an important role. Contracts and invoices as well as all documents issued by formal authorities are printed on a solid paper and a signature or a stamp guarantee the authenticity of the content. Official stamped documents are often accepted without speculation, though nowadays there is an advanced computer technology misusable for falsification and available to general public. The process of a modern forgery includes photocopying or scanning the original document, digital image modification and printing. With little effort the result might appear convincingly. 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, printed invoices incoming to an insurance company are immediately digitized [3] and the scanned image remains the only evidence of possible fraud. The situation calls for automatic verification methods.

In this paper, verification of authenticity of stamps will be discussed. Classical rubber stamping is assumed which means the process of imprinting a certain type of dye applied on a pattern carved into a piece of rubber. It is our purpose to develop a system to automatically detect a stamp in a scanned document image and determine whether it is authentic. The aim is to increase efficiency in crime prevention, not to replace the human examiner. The suspicious documents would be filtered out and inspected carefully by a forensic expert.

Reliable detection of stamps in document images is not a trivial task and it has not been solved till today. The difficulty is in the fact that there is no template for a stamp – they are all different. The variations are in its shape and color, print quality or rotation. Moreover, the imprint might be overlapped by a signature or a text.

The most advanced stamp detection method so far was presented in [1] and is just applicable to stamps of red and blue color. With our approach, all color stamps can be detected and the algorithm is extensible for detection of black stamps too.

2 STAMP DETECTION METHOD

The proposed method for color stamp detection is following. To separate out background and text from color objects, all pixels whose color is close to gray levels are marked as achromatic and discarded from further process. Then, clustering is performed on the chromatic part to obtain several cluster images containing just components of the same color. The cluster images are then segmented by XY cut algorithm to obtain candidate solutions. Feature extraction is performed and the candidate is classified as stamp/non-stamp.

By this approach, only authentic stamps are aimed to be detected. If there was no stamp detected in a document where it was supposed to be, it implies that the stamp is either missing or it is fraudulent.

2.1 Conversion to YC_bC_r Color Space

Color is an important feature when detecting a stamp as stamps are monochrome objects. However, we have to take into account that color brightness can be different in different parts of the imprint. *RGB* color space is generally not convenient for segmentation due to the high correlation among the channels. For the purpose of separation of chromatic pixels and color clustering, YC_bC_r color space was chosen.

In this model, Y is the luma component and C_b, C_r are chroma components meaning the blue and red difference. The color space was developed as part of *ITU-R BT.601 standard* and was primarily designed to encode video. The conversion formulas are: In YC_bC_r color space, separation of achromatic pixels can be easily achieved by elimination of pixels with both C_b and C_r values close to 0.

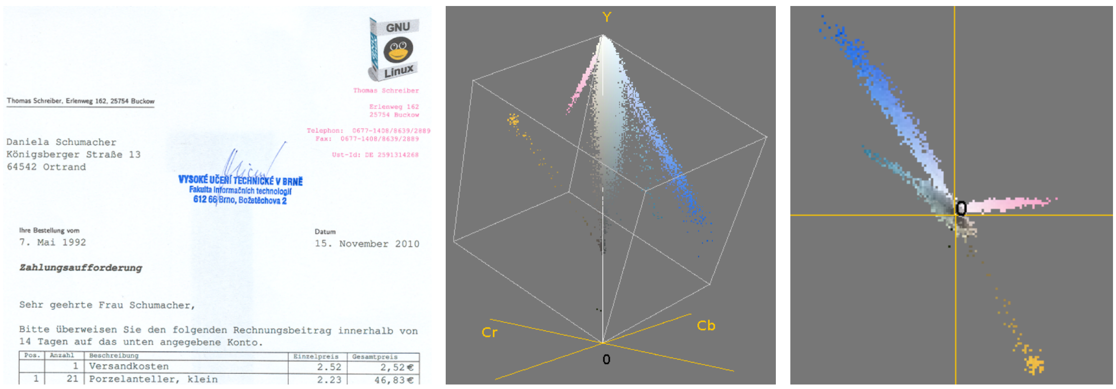

Figure 1: Ink colors in the image form clusters of elongated shapes. A scatter plot in YC_bC_r color space in the middle and a projection of pixel intensities onto the C_bC_r plane on the right.

2.2 Color Clustering

From $3D$ scatter plots of pixel intensities of various document images (a sample document in Fig. 1), one can learn that the clusters formed by inks always have elongated shapes like clouds stretching from the white color cluster. This property is utilized for clustering. For further computations, the Y component can be dropped, and all chromatic pixels are projected onto the C_bC_r color plane. Expressing the vectors in polar coordinates, $\vec{u} = (r, \theta)$, where $r = \sqrt{C_b^2 + C_r^2}$ and $\theta = \arctan \frac{C_b}{C_r}$, the radial coordinate r does not carry any meaningful information for segmentation while the angular coordinate θ is distinguishing. Therefore, k-means clustering on angular space with a redefined distance metrics is performed. For two vectors $\vec{u_1} = (r_1, \theta_1), \vec{u_2} = (r_2, \theta_2)$:

$$dist(\vec{u_1}, \vec{u_2}) = \begin{cases} |\theta_2 - \theta_1|, & \text{if } |\theta_1 - \theta_2| \leq \pi \\ 2\pi - |\theta_2 - \theta_1|, & \text{if } |\theta_1 - \theta_2| > \pi. \end{cases} \quad (1)$$

2.3 CLASSIFICATION

The derived clusters are handled as separate binary masks of the original image and segmented by XY cut algorithm. The resulting rectangles are used as bounding boxes to candidate solutions. Three types of features are extracted for binary classification of the candidate: geometrical, color-related and printing features.

Geometrical features comprise width-to-height-ratio, size of the bounding box, pixel density within the bounding box etc. They are computed relatively to the resolution of scanning. The resolution is estimated from the average connected component size. *Color-related features* include the standard deviation of hue of candidate pixels and also the difference of average hue in the close neighborhood of the candidate. With the means of *printing features* adopted from [2], it is possible to distinguish between different printing techniques - laser, inkjet or photocopy. These features are especially interesting for distinction of genuine and forged stamps.

Figure 2: On the left, stamps extracted by the algorithm are marked by red bounding boxes. On the right, correctly segmented overlapped stamps are shown.

3 CONCLUSION

A method for detection of color stamps has been designed and implemented. It is being extended for finer discrimination between original and forged stamps.

A data set of 400 scanned documents has been collected and annotated. It contains documents with logos, color text, stamps of different colors, multiple stamps on one page and also with no stamp at all. The evaluation showed that the proposed algorithm can detect original stamps with 82% precision and 82% recall. The results on a data set with forged documents are promising and new features are being employed to improve the performance.

4 ACKNOWLEDGEMENT

The research was conducted in cooperation with German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern.

REFERENCES

[1] P. Forczmański and D. Frejlichowski. Robust stamps detection and classification by means of general shape analysis. In *Computer Vision and Graphics*, volume 6374 of *Lecture Notes in Computer Science*, pages 360–367. Springer-Verlag, Berlin, Heidelberg, 2010.

[2] C. Schulze, M. Schreyer, A. Stahl, and T. M. Breuel. Using DCT features for printing technique and copy detection. In *Proc. of the 5th Int. Conf. on Digital Forensics*, pages 95–106, Orlando, FL, USA, January 2009.

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