Pattern Recognition in Images of Counterfeited Documents

Rafael Vieira[1]
2141500@my.ipleiria.pt

Catarina Silva[1,2]
catarina@ipleiria.pt

Mário Antunes[1,3]
mario.antunes@ipleiria.pt

Ana Assis[4]
ana.assis@pj.pt

[1]School of Technology and Management, Polytechnic Institute of Leiria, Portugal

[2]Center for Informatics and Systems of the University of Coimbra, Portugal

[3]Center for Research in Advanced Computing Systems, INESC-TEC, University of Porto, Portugal

[4]Scientific Police Laboratory – Judiciary Police, Portugal

Abstract

Pattern recognition techniques are invaluable approaches to apply to forgery detection of official documents. Forgers are increasingly resorting to more sophisticated techniques to produce counterfeited documents, trying to deceive criminal polices and hamper their work. Hence, different approaches are being pursued, but seldom with real applications in real scenarios. An important challenge is the forger's *modus operandi* characterization, making it possible to obtain more information about the source of the counterfeited document.

In this paper we present a framework conceived for the Scientific Police Laboratory of the Portuguese Judiciary Police to automate counterfeit documents identification by comparing a given fraudulent document image with the images stored in a database of previously catalogued counterfeited documents.

The proposed system improves the counterfeit identification and reliefs the error prone, manual and time consuming tasks carried on by forensic experts. The framework is based on a scalable algorithm under the OpenCV framework, to compare images, match patterns and analyse textures and colours.

1 Introduction

Counterfeited documents are reproductions or imitations of the originals ones. The process of counterfeited documents identification is mostly manual and supported on expert's past experience.

The manual analysis of all the constituent elements of the questioned document is mainly based on a digital version of the original document[1,2,3] produced by using materials and printing techniques from available technologies. It is then carried out through different techniques and methodologies (physical and chemical examinations). Those elements may include printing process, watermarks, fluorescent fibers and planchettes, guilloche pattern, fluorescent and magnetic inks, optically variable inks, rainbow printing, microprinting, latent images, scrambled indicia, laser printing, photos, signatures, embossing stamps, optically variable devices, protective films, perforations, machine readable security, retro-reflective pattern, among others. This analysis provides information that may lead to the classification of the original document as genuine, false or forged.

Technical observations are based on information that may conduct to the discovery of the counterfeiting operation, i.e. associate the counterfeit with the components of its production. If a match of the counterfeited document against the database of old cases is found, the counterfeit is identified and a correlation with all the identical cases already detected in past will be successfully pursued. Otherwise, it will be created a new counterfeit number for future correlations.

The whole process of comparing a fake document with a list of previously catalogued counterfeited ones is usually made manually by the forensic experts of the Scientific Police Laboratory. Having in mind that the catalogue of documents, even for a specific document type, is potentially overwhelming, the time involved in such manual analysis may thus be prohibitive and certainly inefficient for a fast criminal investigation response. Hence, an information system based on image detection algorithms that could automate, or semi-automate such process could bring numerous advantages.

In this paper we propose a methodology to automate the comparison of a counterfeit document with an existing database of already classified counterfeit documents. The main goal is to implement an algorithm that ranks the level of similitude of an image of the questioned document being compared and thus to discard automatically those documents with less or no similitude.

In any case, the human should always remain in the loop. As such, manual verification should always be carried out in any case. However, by discarding a set of documents with less similitude probability, forensic experts' attention may be directed to the most relevant documents.

2 Image Processing Algorithms

OpenCV is a very well-known and widely used open source library for computer vision. It has tools for digital image processing and includes a set of algorithms for pattern detection and image comparison, briefly explained below.

The Harris Corner Detection algorithm[5] was developed by Chris Harris and Mike Stephens. The underpinning mathematic model used to detect corners and edges considers window in the image and then determine the average changes of image intensity. The result obtained is achieved by shifting the window by a small number of pixels in various directions. The detection of corners in digital images is made by comparing the same area in both documents.

Lowe's[6] Scale-Invariant Feature Transform algorithm (SIFT) is another algorithm to detect corners. SIFT is meant to be invariant to image scale and rotation, that is invariant when the image is zoomed out or zoomed in.

Bay et al. proposed Speeded-Up Robust Features (SURF) as a variation of SIFT, aiming to obtain an optimized version of the image to computer vision processing[7]. Rosten and Drummond developed the Fast Algorithm for Corner Detection (FAST)[8] that may have better performance in real time applications. Rublee et al. implemented Oriented Fast and Rotated Brief algorithm (ORB)[9], which is basically a mixture of SIFT and FAST algorithms.

OpenCV framework has a wide set of interesting functionalities for pattern detection in digital images, besides the core implementation. Homography is one of such functionalities that refers to the detection of an image inside another, by a matching templates.

3 Automation of Counterfeited Documents Correlation

Figure 1 depicts the image processing algorithm and the data flow used to automate the analysis of counterfeited documents. There are three distinct tasks: 1) texture analysis; 2) comparison of image areas; 3) detection of similar imperfections in text areas.

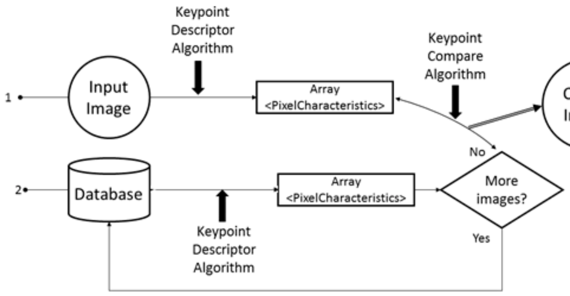
Keypoint Descriptor Algorithm
Keypoint Compare Algorithm
1
Input Image
Array <PixelCharacteristics>
Output Images
No
2
Database
Array <PixelCharacteristics>
More images?
Keypoint Descriptor Algorithm
Yes

Figure 1: Image processing algorithm data flow

In (1) a given image will be processed with a "keypoint descriptor" algorithm that is part of OpenCV SURF, SIFT or ORB implementations. The output of the image processing is an array with the following information: angles of corners, edges, pixel's intensity and directions of the most pronounced intensity changes. In (2) we apply a "descriptor" algorithm to the images stored in the database that meet the same characteristics (type of document and country) of the input document. For example, if the input document is a Portuguese driver license, only documents of this type will be processed. The information retrieved for each processed document is then compared with the array obtained in (1), using a "descriptor compare" algorithm. The output of this processing is an array with an index of similitude between the input document and each one of the stored documents that were analysed.

Texture area identification extracts the necessary parameters and is calculated by HogDescriptor texture descriptor processing algorithm, a OpenCV native algorithm.

Keypoint descriptors processing was made by SURF implementation, a non-native OpenCV algorithm. The algorithm receives an image to a keypoint descriptor for each pixel and then computes an "interestingness" function, which measures the likelihood and uniqueness of each point in another similar image. Keypoint descriptor algorithm analyzes the area around each pixel (the corners) and calculates statistical values and hashes that will be retrieved for future comparison. The algorithm continues by applying the same "keypoint descriptor" algorithm to all the other images stored on database that match the criteria (type and country) and further evaluate their level of similitude with the original input document.

3.1 Experimental Setup and Result Analysis

Our tests focused on Portuguese driver's license cards[10]. The input image is a faked stamp area, as represented in Figure 2. It was compared with images of the same type of official documents from the counterfeited images database, which in this case consists of almost 1500 entries of counterfeited driver's license cards.

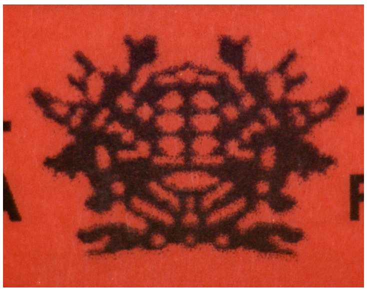

Figure 2: Input image

Using the SURF algorithm, we have calculated the closeness between the images in the dataset and the input image, by identifying the most important keypoints in each image. For each identified keypoint the respective percentage of similitude is calculated. The final score is the arithmetic average of all keypoints. Additionally, we were able to graphically represent the images side-by-side and identify the locations where images have closer keypoints, as shown in Figure 3.

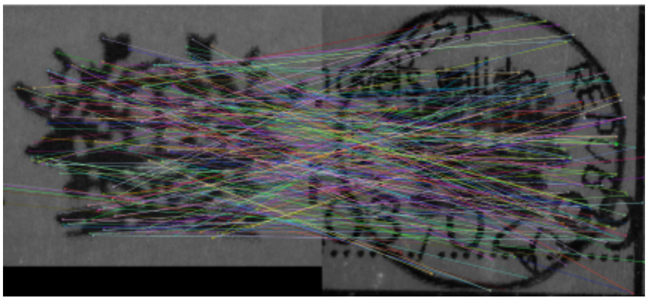

Figure 3: Graphical output representing keypoints matching

It is worth noting that, in Figure 3, both original and stored images of documents do not have perfect cuts around the area of interest to analyse. However, despite these imperfections, the results obtained by the algorithm remained uninfluenced.

With the set of experiments conducted for the Portuguese driver's license cards, we have reached the following results:

Best Candidate	2nd	3rd	4th	5th
83,2%	79,6%	59,2%	48,4%	44,8%

Table 1: Results of experimental setup (similitude degree)

The top 5 best candidates shows that the first 2 have a high degree of similitude, which means that they represented the cases with the most probability of having similar *modus operandi*.

Comparing to the correlation made by the experts of the Forensic Laboratory, the case in study had one other case correlated, and those two images belong to that case.

Therefore, not only the algorithm correctly identified candidates with high values of similitude (over 79%), but also since there was only one case associated, the remaining candidates in the top 5 had low values of similitude (under 60%).

4 Conclusions and Future Work

The purpose of this paper was to introduce a solution able to recognize similar *modus operandi* of frauds in counterfeited documents. To a criminal investigation, this relation can be useful to find the source of the production of the counterfeited documents and to avoid future falsifications of the same type. Nowadays, this recognition is carried out manually, consuming human resources and time.

The presented solution uses visual computing algorithms to compare areas of the documents where a counterfeits are identified.

The current dataset has almost 10.000 images and includes only filters by country and document type. Therefore, the current method is getting more complicated to carry out.

With a filtered dataset, the first version of the presented algorithm took about 70 minutes to process 1500 images. Optimizing the mathematical calculations and introducing parallel computing, the current processing time is about 7 minutes. Given that an expert takes at least 15 minutes to identify the easier falsifications, the current processing time is considered a really good improvement, not only time-wise, but also for relieving human resources. A new parallel computing version is being developed with Graphical Process Unit (GPU), which may improve the time used in each analysis.

Finally, this algorithm was designed to process any kind of official document or image that the forensic laboratories usually work with, e.g., stamps or banknotes.

References

[1] A. Kaur and A. K. G. Vaibhav Saran, "Digital Image Processing for Forensic Analysis of Fabricated Documents", in *International Journal of Advanced Research in Science, Engineering and Techonology*, pp. 84-89, September 2014.

[2] R. Bertrand, P. Gomez-Kramer, O. R. Terrades, P. Franco and J.-M. Ogier, "A System Based On Intrisic Features for Fraudulent Document Detection", in 12th Internation Conference on Document Analysis and Recognition, pp. 106-110, August 2013.

[3] J. Fridrich, D. Soukal and J. Lukás, "Detection of Copy-Move Forgery in Digital Images", in Forensic Science International, vol. 231, pp. 284-295, September 2013.

[4] H. Farid, "Image Forgery Detection", in *IEEE Signal Processing Magazine,* pp. 16-25, March 2009.

[5] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", in 4th Alvey Vision Conference, pp 147-151, 1998.

[6] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", in International Journal of Computer Vision, pp. 91-110, 2004.

[7] H. Bay, A. Ess, T. Tuytelaars and Luc Van Gool, "Speeded-Up Robust Features (SURF)", in Computer Vision and Image Understanding, vol. 110, pp. 346-359, June 2008.

[8] Edward Rosten and Tom Drummond, "Machine Learning for High-Speed Corner Detection", in 9th European Conference on Computer Vision, May 2006.

[9] E. Rublee, V. Rabaud, K. Kunolige and G. Bradski, "ORB: An Efficient Alternative to SIRF or SURF", in IEEE International Conference on Computer Vision, pp. 2564-2571, November 2011.

[10] R. Vieira, C. Silva, M. Antunes, A. Assis, "Information System for Automation of Counterfeited Documents Images Correlation", Procedia Computer Science, 100, 2006, pp. 421-428.