A NEW STUDY ON WOOD FIBERS TEXTURES: DOCUMENTS AUTHENTICATION THROUGH LBP FINGERPRINT

Francesco Guarnera⋆*, Dario Allegra*⋆*, Oliver Giudice*⋆†*, Filippo Stanco*⋆ *and Sebastiano Battiato*⋆

⋆ Department of Mathematics and Computer Science, University of Catania, Catania, Italy
†Banca d'Italia, Rome, Italy
{*allegra, giudice, fstanco, battiato*}*@dmi.unict.it, francesco.guarnera@unict.it*

ABSTRACT

The authentication of printed material based on textures is a critical and challenging problem for many security agencies in many contexts: valuable documents, banknotes, tickets or rare collectible cards are often targets for forgery. This motivates the study of low-cost, fast and reliable approaches for documents authenticity analysis. In this paper, we present a new approach based on the extraction of translucent patterns from paper sheet by means of a specific-built framework. A fingerprint is obtained by computing a Local Binary Pattern descriptor on the digital image. To validate the robustness of the proposed method for authentication analysis, we introduce a novel dataset and perform retrieval tests under both, ideal and noisy conditions. Experimental results prove the validity of the proposed strategy.

Index Terms— Paper fingerprint, Retrieval, Secure document.

1. INTRODUCTION

The manufacturing process of paper involves the use of wood particles as a base; subsequently, the application of other compounds, results in what we know as a paper sheet. This process introduces random imperfections, which makes the paper sheet unique and allow to build a fingerprint. The massive demand of robust authentication methods in many context [1, 2, 3, 4, 5], makes the fingerprint extraction an attractive and challenging problem. Although several techniques have been proposed, most of them require industrial and expensive devices. This drives researchers in the field, to find cheaper solution which does not require high-end hardware. Inspired by the use of wood fibers pattern for fingerprint extraction [6], we propose a new method for document authentication that takes advantage of a fast feature extraction strategy and a robust fingerprint description. To this aim, we employ Local Binary Pattern features (LBP) [7, 8] and prove it outperforms results obtained by Toreini et al. [6] in terms of efficiency and effectiveness.

Since the paper texture is unique, in ideal conditions any sufficiently descriptive approach performs almost perfectly in terms of accuracy results. This led us to further investigate by performing retrieval tests on data where some alterations are applied. For example, one alteration consists of artificially removing parts of the input data (e.g., by introducing black blocks) in order to simulate torn paper or holes. These experiments, allowed to evaluate the robustness of the proposed fingerprint, by simulating real-case scenarios where part of the original information is missing. Moreover, since no wood fibers pattern dataset is publicly available, the dataset presented in this paper will be available to researchers to test their own approaches. At best of our knowledge, this dataset is the first publicly available one to address this problem. The following points summarize the contributions of this paper:

1. a new framework for acquisition of a digital image of paper sheets;
2. a new public dataset which includes images showing wood fibers patterns.
3. a new fingerprint extraction method, based on LBP, which outperforms state-of-the-art;
4. a fingerprint robustness evaluation by using altered paper patterns;

1.1. Related Works

The use of a fingerprinting technique for documents authentication was proposed for the first time by Buchanan et al. [9] in 2005. Their finding was that the surface of a paper sheet presents unique microscopic imperfections. This fingerprint makes the forgery unfeasible, given that it is unique and virtually impossible to be modified controllably. To extract the fingerprint from paper structure, they employed laser irradiation from four different angles and acquired the reflected energy. Inspired by Buchanan et al., van Beijnum et al. [10] proposed an improvement based on correlation metrics between the acquired energy signals. In 2008, Cowburn introduced the use of laser speckle for products authentication [11]. In 2009, Clarkson et al. [12] proposed to extract 3D paper structure by scanning the paper in four different orientations. Then, they employed Voronoi distribution features to built a robust fingerprint. In 2010, Samsul et al. [13] proposed a fingerprint extraction method, which exploits CCD sensors and laser speckle, namely a pattern of bright and dark spots caused by interference of two or more light beams with different phases. A similar approach, has been proposed by Sharma et al. in 2011 [14]. Differently by [13], they employed a microscope to acquire the speckle pattern. Although the aforementioned approaches works well for paper fingerprint extraction, they require industrial and specific expensive equipment. In 2017, this limitation was surpassed by the works of Wong et al. [15] and Toreini et al. [6]. Wong et al. [15] proposed a strategy to extract paper surface imperfections by exploiting multiple shots taken by a mobile camera under semi-controlled light conditions; in 2019 they furtherly investigate candidate mathematical models for camera captured images [16]. Differently from previous works, Toreini et al. do not detect surface imperfections, rather capture the texture related to the random disposition of the wood fibers inside the paper sheet. To extract paper pattern, they exploited a consumer camera and a backlit surface. However, they print a bounding box on the analysed paper to simplify the automatic texture registration. Since in real scenario this registration strategy is not applicable, we propose a different acquisition framework to easily register the acquired document. After acquisition and registration, Toreini et al. rescale the extracted patch to 640×640 and use 100×100 Gabor filter for feature extraction. Then, they create a 2048 bits fingerprint by processing filtering output. Nonetheless, convolution with 100×100 Gabor filter is very time consuming.

As already demonstrated in [17], the random disposition of wood fibers on paper sheets makes possible the construction of a fingerprint virtually impossible to tamper; hence, given the limits of the previous works in terms of costs, acquisition constraints and robustness, we propose a novel fingerprint extraction strategy by using specific low-cost image acquisition equipment and a simpler and faster method namely, Local Binary Pattern [7]. Although in recent years, CNN-based methods achieved great performance in image retrieval and classification, they have a higher complexity and require GPUs to perform quickly; for this reason, we have not explored these approaches in the present study.

The remainder of this paper is organized as follows: Section 2 describes the acquisition procedure and the organization of the new dataset; Section 3 details the proposed fingerprinting extraction strategy; in Section 4, we report experimental results with a discussion on the validity of the proposed approach. Conclusions are summarized in Section 5.

2. ACQUISITION AND DATASET

Starting from the idea described in [6], data acquisition process required relatively cheap devices, namely an overhead projector as light source and a consumer RGB camera. We built an acquisition framework in which a camera is hanged on a projector arm to acquire a top view picture of the paper sheet crossed by light. Although the technique described in [6], introduces a printed rectangle in the document to easily register the captured pattern, in real scenarios it is not possible to print anything on the analysed document. Hence, we added an angle support for the paper, which allows to identify the image top-left corner for an easier registration and guarantees the same area of the paper sheet is acquired each time. The proposed acquisition framework is shown in Figure 1. System settings are the followings:

- Xerox overhead projection with 24V/250W lamp;
- Camera Nikon D3300; Lens Nikon DX VR 15mm-55mm 1:3.5-5.6 GII.
- Distance between paper and lens is 7.4cm.

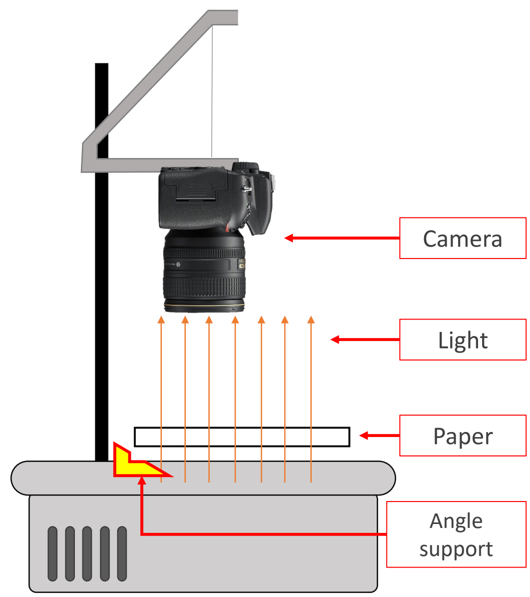


Fig. 1. Acquisition framework consisting of an overhead projector, a consumer RGB camera and a fixed metallic support to control paper positioning.

In Figure 2, an example crop of the 6000×4000 image acquired by the proposed system is shown. The dark bands on the top and on the left, are related to the area outside the paper (i.e., the metallic support). A simple luminance threshold is used to distinguish paper pixels from external area; then, let (x_0, y_0) the position of the top-left paper pixel, we extract a 5000×1000 subimage anchored in position $(x_0 + m, y_0 + m)$ with $m = 20$. Subimage size has been chosen to have a spatial resolution of $11 \times 55mm$, which is compatible with the one in [6]. The dataset has been created by selecting 55 different A4 sheets of paper with grammage $80g/m^2$; then, we have extracted the related 5000×1000 patterns. Secondly, 25 samples have been removed from the original set and the remaining 30 papers have been acquired

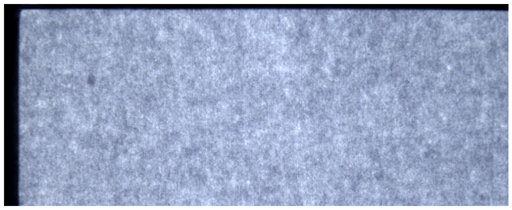


Fig. 2. A raw picture crop acquired through the proposed hardware.

again multiple times. Overall, we performed 167 new acquisitions in order to obtain a dataset of $55 + 167 = 222$ sample patterns. Finally, all images have been converted to grayscale. The dataset is publicly available online to encorauge the research on the field [1].

3. DESCRIBING PAPER FINGERPRINTS WITH LBP

The unique disposition of fibers in a paper sheet can be extracted exploiting visible light through the framework described in previous section. The obtained image, is a visible representation of the complex random texture that have to be mathematically described in order to be employed in automatic processing. To this aim, the objective fingerprint descriptor should possess the following properties: easy to implement, low complexity, encoding capabilities for all the information with robustness to absence of paper parts.

Local Binary Pattern (LBP) [7, 8] is demonstrated to satisfy all the properties described above: it guarantees high quality results in the presence of small differences due to input image variabilities. LBP is a local descriptor computed by comparing a pixel, called pivot, to its neighbours. Comparison result can be represented with an histogram (LBPH). Computing LBP on an entire image I means computing it for each pixel of I and thus building the LBPH by counting color occurrences on $LBP_{r,n}(I)$ values. r, n are LBP parameters: r is the radius of pixels around the pivot, n is the number of circularly symmetric neighbour points. However, using this approach on the entire image produces an histogram where most of the spatial information is lost. Thus, the image is divided into p non-overlapping square patches of $L \times L$ pixels. Then, the LBPH for the entire image I is obtained as a concatenation of each LBPH on patches as described in Figure 3. In other words:

$$LBPH_{r,n}(I) = LBP_{r,n}(\{patch_k(I, L))k \in [1..p]\}) \quad (1)$$

In order to reduce histograms length and to introduce rotational invariance, uniform binary patterns have been used. Given two paper sheet images I_1, I_2 for which the contents can be described by LBPs computed as in 1, the authenticity test can be done by means of comparison. A simple distance $d(LBPH_{r,n}(I_1), LBPH_{r,n}(I_2))$ can be employed.

[1] iplab.dmi.unict.it/paperFingerprint

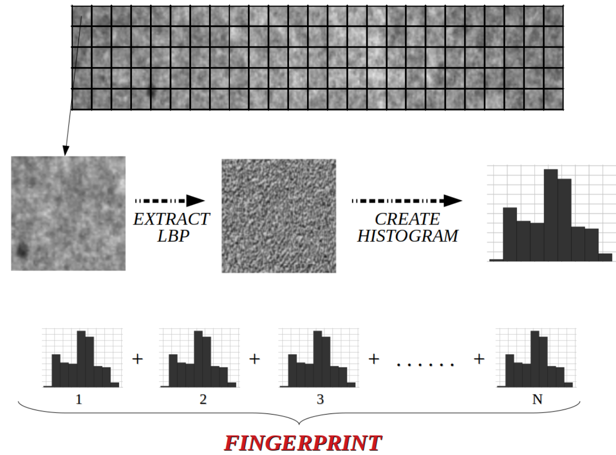
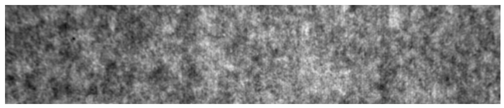


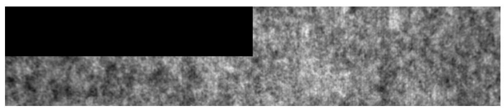
Fig. 3. Pipeline for fingerprint extraction.

4. EXPERIMENTAL SETTINGS AND RESULTS

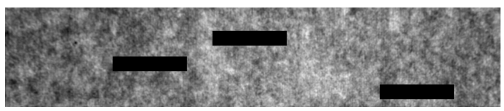
To assess the performance of the proposed approach, document authentication is treated as a retrieval problem. For a fair comparison with [6], all image patterns have been rescaled to 640×128, in order to have a similar dpi resolution. The 55 patterns related to 55 different papers are used for querying against the remaining 167. Note that only 30 of 55 papers pattern have a correct match with at least one query. The 25 unique patterns are included to make the task more challenging. The paper fingerprints have been extracted with both analysed methods. In accordance with Toreini et al.[6], the parameters for both, LBP and Gabor filter, have been found through a grid search to maximize the performances. Hence, in this study, LBP features are computed with *radius* $r = 12$, *neighbour* $n = 48$ and patch size $p = 32$. Concerning the Gabor filter employed in [6], we use *frequency*=1, θ=π and σ=10. Ranking for LBP fingerprint retrieval are based on Bhattacharyya Distance [18] between the query pattern and the ones within the database; whereas, fingerprint extraction proposed by [6] exploits Hamming distance. The retrieval performances are measured using accuracy and mean average precision (mAP) [19]. The accuracy is a measure for the rate of queries which have a correct match with the first retrieved document; whereas, mAP allows to measure the overall performance of a retrieval system by considering the whole ranking of queries result. In this work, tests have been performed under different conditions. The first experiment, aims to evaluate the performance under optimal conditions, namely when no alterations occurs on query documents. Then, in order to evaluate real scenarios were the original paper texture can be altered by physical damages (e.g., tears), two kind of artifi-



(a)



(b)



(c)

Fig. 4. (a) Genuine texture; (b) Texture with Tear noise, factor $k = 0.5$; (c) Texture with Stain noise, factor $s = 0.15$.

cial alterations with 10 different intensity degrees have been introduced: Tear and Stain.

Tear: it simulates a tear and a consecutive loss of document portion. When this alteration is introduced with degree $k \in [0, 1]$ in a $W \times H$ image, we remove information from all the pixels (x, y) with $0 \leq x \leq k * W$ and $0 \leq y \leq k * W$ (Figure 4(b)).

Stain: it introduces random black blocks on the texture to simulate stains or holes. When this artifact is introduced with factor $s \in [0, 1]$, information from three boxes in random positions is removed. In a $W \times H$ image, each box sizes w and h, are randomly chosen in the range $[1, s * W]$ and $[1, s * H]$. (Figure 4(c)).

Note that, the synthetic artifacts are not detected and not removed, hence they are completely included in the descriptor. All the experiments have been coded in Python language and run on CPU Intel i5-3210M 2.50GHz, RAM 6 GB, SO Ubuntu 18.04 LTS.

4.1. Results

Achieved results prove that proposed LBP fingerprint approach is more robust, effective and even faster than Toreini et al. method. The average time to extract a fingerprint have been computed with both methods and a time ratio of 1.80 has been obtained: the proposed technique is about 2 times faster. Plot in Figure 5, shows accuracy of the retrieval tests, when the introduced alteration degree is increasing. For Tear and Stain artifacts, 10 values in $[0.15, 0.6]$ and $[0.05, 0.5]$ ranges respectively have been taken. For each artifact, the alteration degree in $[0, 1]$ is normalized in order to improve plots readability. The obtained result confirms that the proposed technique is able to encode a more robust descriptor in real-case scenarios. Some words are needed for the Stain case. Stain noise is the most realistic one and if the noise level

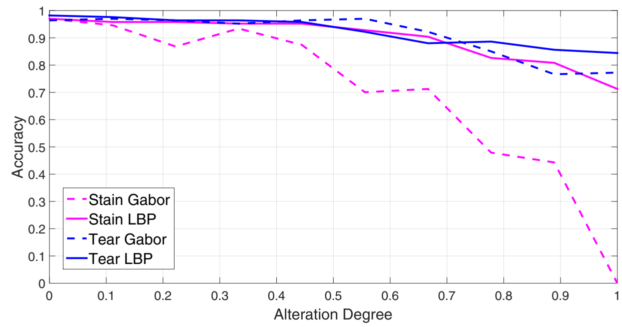


Fig. 5. Retrieval accuracy at different alteration degrees, for different artifacts.

surpasses a little degree, Toreini et al. approach is heavily outperformed by the proposed one. This is because LBP is able to describe the paper texture fingerprint, even if most of the information is torn apart, given its properties of local descriptor and the patching technique employed.

Finally, mAP measures for retrieval tests are shown in Table 1. As regards the tests on altered patterns, we have 10 different values of mAP (i.e., one for each alteration degree), hence the average on them is reported. Since mAP describes the overall performances of a retrieval engine, these results furtherly confirm that the proposed fingerprint achieves a better ranking even when the first match is not correct. As expected, in ideal condition both the approaches perform very well, however the proposed one demonstrates better retrieval performance.

	No artifacts	**Tear**	**Stain**
Proposed	98.68%	93.95%	91.94%
Toreini et al.[6]	97.98 %	93.49%	75.82%

Table 1. Mean Average Precision for documents retrieval under different conditions.

5. CONCLUSION

Starting from the translucent patterns extracted from the paper sheet through a specific low-cost acquisition framework, an LBP-based technique has been proposed to describe these patterns. Results demonstrated the effectiveness of the strategy in both, ideal conditions and altered/noisy scenarios, by outperforming state-of-the-art works in terms of accuracy and efficiency. For future works, robustness tests in very stressing conditions, in conjunction with LBP variants [20, 21] and new handcrafted descriptors, will be studied. To this aim the dataset will be expanded by including more samples and also different kind of translucent papery material.

6. REFERENCES

[1] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt, "Combating digital document forgery using new secure information hiding algorithm," in *International Conference on Digital Information Management*, 2008, pp. 922–924.

[2] Amr G. H. Ahmed and Faisal Shafait, "Forgery detection based on intrinsic document contents," in *International Workshop on Document Analysis Systems*, 2014, pp. 252–256.

[3] Albert Berenguel, Oriol R. Terrades, Josep Lladós, and Cristina Cañero, "Banknote counterfeit detection through background texture printing analysis," in *International Workshop on Document Analysis Systems*, 2016, pp. 66–71.

[4] Arcangelo R. Bruna, Giovanni M. Farinella, Giuseppe C. Guarnera, and Sebastiano Battiato, "Forgery detection and value identification of euro banknotes," *Sensors*, vol. 13, no. 2, pp. 2515–2529, 2013.

[5] Navpreet K. Gill, Ruhi Garg, and Er A. Doegar, "A review paper on digital image forgery detection techniques," in *International Conference on Computing, Communication and Networking Technologies*, 2017, pp. 1–7.

[6] Ehsan Toreini, Siamak F. Shahandashti, and Feng Hao, "Texture to the rescue: Practical paper fingerprinting based on texture patterns," *ACM Transactions on Privacy and Security*, vol. 20, no. 3, pp. 1–29, 2017.

[7] Timo Ojala, Matti Pietikainen, and David Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *International Conference on Pattern Recognition*, 1994, vol. 1, pp. 582–585.

[8] Timo Ojala, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[9] James D. Buchanan, Russel P. Cowburn, Ana-Vanessa Jausovec, Dorothee. Petit, Peter Seem, Gang Xiong, Del Atkinson, Kate Fenton, Dana A. Allwood, and Matthew T. Bryan, "Forgery: Fingerprinting documents and packaging," *Nature*, vol. 43, no. 28, 2005.

[10] Frerik van Beijnum, EG van Putten, KL Van der Molen, and AP Mosk, "Recognition of paper samples by correlation of their speckle patterns," *arXiv preprint physics/0610089*, 2006.

[11] Russell Cowburn, "Laser surface authentication - natural randomness as a fingerprint for document and product authentication," in *Optical Document Security Conference*, 2008.

[12] William Clarkson, Tim Weyrich, Adam Finkelstein, Nadia Heninger, J Alex Halderman, and Edward W Felten, "Fingerprinting blank paper using commodity scanners," in *Symposium on Security and Privacy*, 2009, pp. 301–314.

[13] Wiwi Samsul, Henri P Uranus, and MD Birowosuto, "Recognizing document's originality by laser surface authentication," in *International Conference on Advances in Computing, Control and Telecommunication Technologies*, 2010, pp. 37–40.

[14] Ashlesh Sharma, Lakshminarayanan Subramanian, and Eric A Brewer, "Paperspeckle: microscopic fingerprinting of paper," in *ACM Conference on Computer and communications security*, 2011, pp. 99–110.

[15] Chau-Wai Wong and Min Wu, "Counterfeit detection based on unclonable feature of paper using mobile camera," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1885–1899, 2017.

[16] Chau-Wai Wong, Min Wu, and Runze Liu, "Enhanced geometric reflection models for paper surface based authentication," *SigPort*, 2019.

[17] Tobias Haist and Hans J Tiziani, "Optical detection of random features for high security applications," *Optics Communications*, vol. 147, no. 1-3, pp. 173–179, 1998.

[18] Anil Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

[19] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Computer Vision and Pattern Recognition*, 2007.

[20] Sheryl Brahnam, Lakhmi C. Jain, Loris Nanni, and Alessandra Lumini, Eds., *Local Binary Patterns: New Variants and Applications*, Springer-Verlag Berlin Heidelberg, 2014.

[21] Alessandro Torrisi, Giovanni M. Farinella, Giovanni Puglisi, and Sebastiano Battiato, "Selecting discriminative CLBP patterns for age estimation," in *International Conference on Multimedia & Expo Workshops*, 2015.