PRINTED DOCUMENT AUTHENTICATION USING TWO LEVEL QR CODE

Iu. Tkachenko[1,2], *W. Puech*[1], *O. Strauss*[1]

[1]Laboratory LIRMM, UMR CNRS 5506
University of Montpellier
860 rue de St Priest,
34090 Montpellier Cedex 5, France

C. Destruel[2], *J.-M. Gaudin*[2]

[2]Authentication Industries
CAP OMEGA,
Rond-point Benjamin Franklin
34960 Montpellier Cedex 05, France

ABSTRACT

The availability of high quality copy machines provides a large amount of printed document counterfeits. Numerous authentication techniques based on security printing, graphical codes, hashing or local hashing have been suggested earlier. In this paper, we propose a novel printed document authentication system based on sensitivity of a new two level QR (2LQR) code to copying process. This 2LQR code contains specific textured patterns, which are sensitive to printing and copying processes. Therefore, it can be used to detect unauthorized document duplication. Experimental results show the efficiency of this 2LQR code for copy detection.

Index Terms— QR code, private message, printed document authentication, print-and-scan process

1. INTRODUCTION

Document protection is a wide, important and quickly developed field of multimedia security. The nowadays challenges are the detection of fake invoices, bank checks, diplomas, tax forms and other valuable documents. Due to technical improvements in printing and scanning devices, the number of such fake documents increases. That is why, numerous new methods for document authentication have been suggested by multimedia security researchers.

Printed document authentication approaches can be divided in two categories. The first is intended to authenticate the document content by using forensics methods [1, 2] or perceptual hashes [3]. The authors in [4] suggest to encode the document local hash in the barcodes used for document tamper proofing. Nevertheless, the authentication using bar codes can be easily copied if it has been printed with ordinary inks [5].

The second approach is based on considering that each time an image is printed or scanned, some information is lost [6]. This technique consists of using specific graphical codes called copy detection patterns [6] that allow distinguishing original document from its copies. These techniques are used to fight against counterfeits, as well as for product and document authentication [7, 8].

In this paper, we suggest to use a new two level QR (2LQR) code [9] for document authentication. This 2LQR code has two storage levels, where the second level is constructed using specific textured patterns. These textured patterns can be chosen to be sensitive to Print-and-Scan (P&S) process. Therefore, we suggest to use this 2LQR code to detect unauthorized document duplication.

The rest of the paper is organized as follows. The document authentication system we propose is described in Section 2. Section 3 presents a short overview of the 2LQR code and the choice of sensitive textured patterns. Section 4 presents the authentication test we use. The experimental results, that include database description, authentication test evaluation as well as pattern detection results for P&S 2LQR codes and copied 2LQR codes, are indicated in Section 5. Finally, we conclude and discuss future work in Section 6.

2. PRINTED DOCUMENT AUTHENTICATION SYSTEM

Let's consider a valuable document and a 2LQR code with public and privates messages that have been generated by a legitimate source. This 2LQR code is inserted into the considered numerical valuable document to ensure its genuineness. After generation, this valuable protected document is printed using a legitimate printer. Further, during legitimate document verification, this printed document is scanned using an authorized device. The P&S image of 2LQR code is then used for document authentication. If the authentication test reaches a positive note, then the private level of 2LQR code can be extracted. Fig. 1 presents the illustration of considered authentication system.

In order to provide also the document content authentication, a document hash (perceptual or cryptographic) can be stored in the private 2LQR code level (or in both 2LQR code levels). Nevertheless, this paper does not focus on document content authentication.

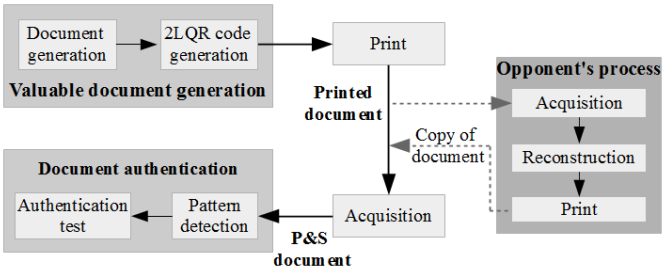


Fig. 1. Considered authentication system for valuable documents.

Most of counterfeits are produced in the interval between print process and scan process, since an opponent has only access to a printed version of the protected document. The changes made by P&S process are considered as a physical unclonable function [10]. Due to the physical changes and their stochastic nature, the P&S process impact can also be used for document authentication [8]. The P&S process adds different distortions as gamma tweaking, digital halftoning [11], dot gain, print-to-print instability, scanner gamma compensation, digitization, geometric transformations [12] and random noise addition. The noise is usually considered as being an additive or a multiplicative Gaussian process [13, 14]. The aforementioned distortions produce irreversible modifications in the input image.
The counterfeits imply the scanning and printing operation using unknown devices. Therefore, the proposed 2LQR code needs to be sensitive to copying process.

3. TWO LEVEL QR CODE FOR AUTHENTICATION

The 2LQR code proposed in [9] has two storage levels. In the first level, the information is stored in a standard way [15]. The second (private) storage level is created by replacing the black modules with specific textured patterns. The generation process of this 2LQR code is described in details in [9]. This process consists of 1) a public message storage in standard QR code, 2) a choice of specific textured patterns, 3) a private message encoding using an Error Correction Code (ECC), 4) the encoded message (codeword) scrambling process and, 5) a replacement of black modules with textured patterns respecting the codeword. In the referenced paper, the authors suggest to choose textured patterns that can be detected after the impact of P&S process, since the main idea is to increase the storage capacity of QR code.
The document authentication of our approach is based on the sensitivity of the textured patterns to the P&S process. Thus, we propose a new method for selecting a textured pattern combination that ensures this sensitivity to P&S process (Section 3.1). In addition, we propose to slightly modify the 2LQR code reading process (Section 3.2).

3.1. Textured pattern choice

The textured patterns used are images $P_i, i = 1, \ldots, q$ of size $p \times p$ pixels. Each textured pattern is a binary image with fixed ratio of black and white pixels. Textured patterns with the same characteristics were used in the textured image generation proposed in [16]. The reading capacity of the private level depends on the pattern density: a large density value can disable the reading process of private level, as well as the authentication process.
Here, we propose to refine the criteria proposed in [9]: the correlation value between textured module P_i and its degraded version S_i must be bigger than any other correlation value (with original patterns $P_j, i \neq j$ and with degraded versions $S_j, i \neq j$). Our goal is to enhance the original idea to support authentication. To fulfill this goal, we propose to impose the gap between the correlation score between P_i and S_i and the best second correlation score (i.e. $\max_{i \neq j}(cor(P_i, S_j))$) to be higher than a given threshold ε:

$$\forall i, j \in [1, q], \quad cor(P_i, S_i) - \max_{i \neq j}(cor(P_i, S_j)) \geq \varepsilon. \quad (1)$$

Therefore, the textured patterns used for the 2LQR code generation should respect the criterion (1).

3.2. 2LQR code reading process

The reading process of the 2LQR code consists of 1) an image pre-processing (where the P&S 2LQR code is re-scaled and rotated), 2) a standard QR code reading process [15] including an error correction process (where the public information is retrieved and the positions of black modules are determined), 3) a second level authentication and, 4) a retrieving of the private information, if the 2LQR code is authentic.
The positions of the black modules, determined in stage 2), are used to generate the class of black modules (BC) that contains the textured patterns $BP_i, i = 1, \ldots, B \times n$, where $B \times n$ is the total number of digits, B is the number of codewords and n is the number of digits in the codeword. Therefore, we have $B \times n$ textured patterns, which belong to q classes.
The textured pattern recognition is based on comparing the P&S textured modules BP_i with the original numerical patterns $P_l, l = 1, \cdots, q$ using the Pearson correlation coefficients $cor(BP_i, P_i)$. We have the textured modules $BP_i, i = 1, \ldots, B \times n$ from black module class BC, as an input to the pattern recognition algorithm. We calculate the correlation values between each textured module BP_i and the original textured modules $P_1, \cdots P_q$: $cor_j^i = cor(BP_i, P_j), i = 1, \ldots, B \times n, j = 1, \cdots, q$. The maximum correlation values $cor_r = \max_{j=1,\cdots,q} \{cor_j^i\}$ define each bit value: $c_i' = r, r = 1, \cdots, q$. In the end of this process, we obtain the scrambled codeword C'. Before both unscrambling and decoding operations, an authentication test has to be performed, which is described in Section 4. If the authentication test approves the

2LQR code authenticity, the unscrambling with key K and error correction decoding process will be applied for the private message extraction. The parity-check bits of ECC are used for error detection and correction.

4. AUTHENTICATION PROCESS

A copy attack is a reproduction chain that implies two successive P&S processes. The modifications induced on the document are cumulative. The authentication step aims at evaluating the pattern degradation in order to acknowledge for successive P&S functions applied to the original pattern. We propose to use the Pearson correlation coefficients for measuring the degradation. Fig. 2 shows this evaluation process.

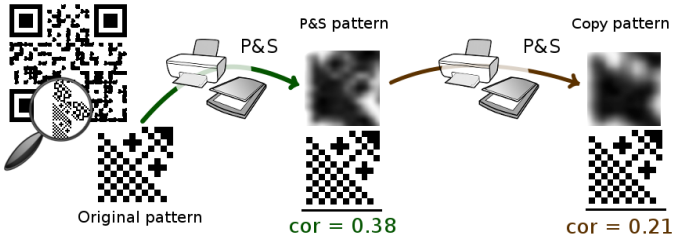


Fig. 2. An example of textured pattern changes during copying process.

An authentication threshold Th has been determined during the pre-trial phase. The authentication test consists of comparing the mean value of $cor_r^i, i = 1, \ldots, B \times n$ with threshold Th. The document is said to be authentic if this mean value is bigger than Th ($mean(cor_r^1, \cdots, cor_r^{B \times n}) \geq Th$).

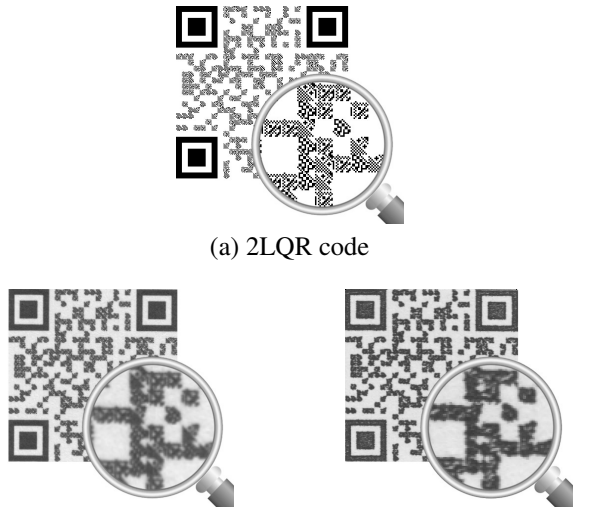
5. EXPERIMENTAL RESULTS

In this section, we present a printed document authentication scenario, describe the database we use and show the pattern recognition results after P&S and copying processes. Additionally, the proposed authentication test has been provided in order to show its effectiveness against unauthorized 2LQR code duplication.

This authentication scenario is as follows. The authority center creates a valuable document. The public and private information is stored in the first and second levels of the 2LQR code. Finally, the generated 2LQR code is inserted into the document, and the document is printed using a desktop printing device.

In the verification step, the valuable document is scanned using a desktop scanner device and a verification of 2LQR code authenticity is applied. If the 2LQR code is authentic, both public and private messages can be read.

In these experiments, the QR code version 3 in Low error correction level is used. This version has 29×29 modules and



(a) 2LQR code

(b) P&S 2LQR code (c) Copy of P&S 2LQR code

Fig. 3. An example of a) 2LQR code, b) P&S 2LQR code and c) copy of P&S 2LQR code.

can store a 440-bit message. The private message is encoded using the ternary Golay code $[11, 6, 5]$, where each of B codewords has $n = 11$ digit length, with $k = 6$ informative digits. With error correction code, that we have chosen, we can store 174 ternary digits on the private level of the 2LQR code, that corresponds to nearly 275 message bits. The textured patterns have been chosen with respect to criterion (1) with threshold $\varepsilon = 0.25$.

The database, that we use, is composed of 120 P&S 2LQR codes with different public and private messages printed with a Brother HL-4150CDN printer and scanned with a Canon LIDE210 scanner. The copies of P&S 2LQR codes are made with a copy machine Toshiba e355 (in standard mode (CM1) and with a maximal contrast (CM2)), a RICOH Aficio MP C2050 (CM3), a Toshiba e456 (CM4) and a Toshiba e256 (CM5). In total, we have 600 copied samples. The printing, scanning and copying processes have been performed using a 600 dpi resolution. Note that we want to highlight the effect of the document copy process: we fix the production chain (that represents a real case) and test several copy machines. Of course, the production and control chain can be improved (for example up to 1200 dpi) to increase the protection of the system. Fig. 3 presents an example of the 2LQR code, as well as an example of P&S and copy 2LQR codes.

Table 1 gathers all obtained results: the first line shows the results obtained to P&S 2LQR codes, the lines from second to sixth represent the results obtained to copies using different copy machines and, the last line illustrates the mean values of lines from 2 to 6 (copies results). For each P&S 2LQR code and each copy, we apply the proposed module recognition method. In the first column of Table 1, we note, that the probability of a wrong pattern detection equals 1.03% for P&S 2LQR codes (original). Nevertheless, the mean probability

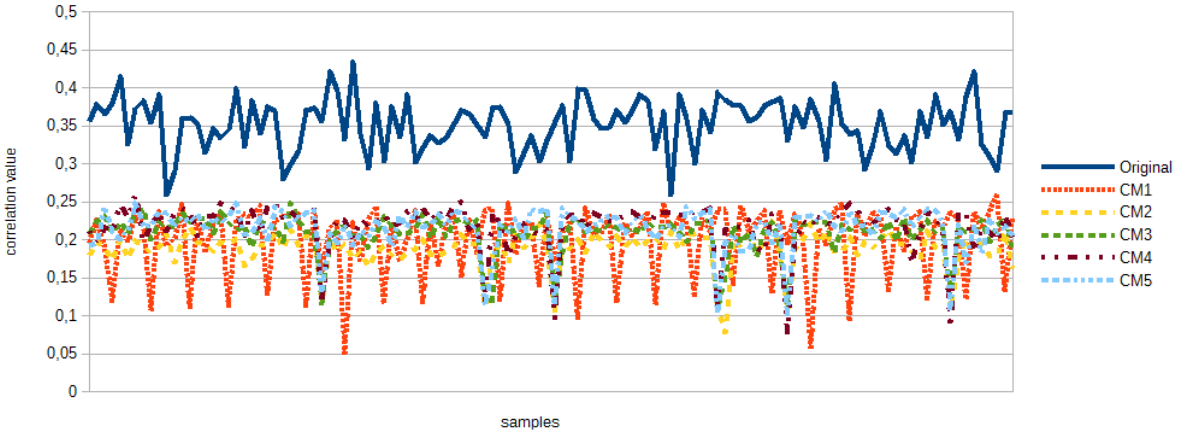


Fig. 4. The mean correlation values for P&S 2LQR codes (blue line) and for copy P&S 2LQR codes using different printers and scanners (orange, yellow,green, vinous and light blue lines).

	Error probability of pattern detection	Error probability of digit decoding after ECC	Positive authentication test ($Th = 0.3$)
Original	1.03%	0.30%	98.33%
CM1	20.30%	21.32%	0.00%
CM2	26.29%	26.24%	0.00%
CM3	23.92%	22.48%	0.00%
CM4	20.12%	15.77%	0.00%
CM5	20.76%	17.01%	0.00%
Mean	22.28%	20.56%	0.00%

Table 1. Error probability of pattern detection and error probability of message decoding after ternary Golay error correction algorithm.

of a wrong pattern detection of copies using different copy machines equals 22.28%.

After the unscrambling using key K, error detection and correction algorithm of ternary Golay code, the private message M_{priv} is retrieved. The error probabilities of incorrect digit decoding are presented in the second column of Table 1. We can see that the private message could not always be decoded in the copy samples, as the pattern detection operation is not performed successfully.

We evaluate the *authentication test* with authentication threshold $Th = 0.3$. The last column of Table 1 shows the probability of code authentication. All copy codes and less than 2% of original 2LQR codes do not pass the authentication test. Our goal is to detect all unauthorized duplication of documents (false positive rate has to be zero) even if false negative rate is non-zero. In reality, all documents, which failed the authentication test, undergo a deeper analysis (high time consuming process), that will distinguish the false negative from real fakes. Therefore, we suppose that the 2% false positive error is acceptable, when the false negative error is equal to 0%. In Fig. 4, we can see the mean correlation values for P&S 2LQR codes (blue line) and for copy 2LQR codes using different copy machines (orange, yellow, green, vinous and light blue line). Fig. 4 shows that the threshold between original and copy 2LQR code can be calculated experimentally, and then this threshold can be used for authentication test performance. The pattern detection results show that when the authentication test failed, we cannot extract the correct private message in most cases.

6. CONCLUSION

In this paper, we have presented a new authentication process dedicated to printed documents that uses a two level QR code. The 2LQR code has two levels of information storage: the public level, where information is stored in standard way, and the private level, where the black modules are replaced by specific textured patterns, that are sensitive to copying process. Thanks to characteristics of textured patterns, the private level of 2LQR code cannot be read into overprinted document copy.

The authentication test, that we propose, aims at verifying the printed document authenticity. If the maximal correlation values between P&S textured patterns and original numerical patterns are smaller than a predetermined authentication threshold, the printed document is considered as being counterfeited.

In future work, we will deeply study the characteristics of textured patterns, as well as the combination criteria. We also plan to find the limit number of different textured patterns that can be used in the 2LQR code. Finally, the limits of document authentication by smartphone cameras have to be studied.

7. REFERENCES

[1] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of copy-move forgery in digital images," in *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003.

[2] H.-J. Lin, C.-W. Wang, and Y.-T. Kao, "Fast copy-move forgery detection," *WSEAS Transactions on Signal Processing*, vol. 5, no. 5, pp. 188–197, 2009.

[3] J. Fridrich, "Visual hash for oblivious watermarking," in *Electronic Imaging*. International Society for Optics and Photonics, 2000, pp. 286–294.

[4] R. Villán, S. Voloshynovskiy, O. Koval, F. Deguillaume, and T. Pun, "Tamper-proofing of electronic and printed text documents via robust hashing and data-hiding," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 65051T–65051T.

[5] S. Voloshynovskiy, O. Koval, R. Villan, E. Topak, J. E. V. Forcén, F. Deguillaume, Y. Rytsar, and T. Pun, "Information-theoretic analysis of electronic and printed document authentication," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 60721D–60721D.

[6] J. Picard, "Digital authentication with copy-detection patterns," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 176–183.

[7] C. Baras and F. Cayre, "2d bar-codes for authentication: A security approach," in *Signal Processing Conference (EUSIPCO), Proceedings of the 20th European*, 2012, pp. 1760–1766.

[8] A. T. P. Ho, B. A. M. Hoang, W. Sawaya, and P. Bas, "Document authentication using graphical codes: Reliable performance analysis and channel optimization," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 9, 2014.

[9] Iu. Tkachenko, W. Puech, O. Strauss, C. Destruel, J-M. Gaudin, and C. Guichard, "Rich QR code for multimedia management applications," in *Image Analysis and Processing (ICIAP) 2015*, pp. 383–393. Springer, 2015.

[10] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proceedings of the 44th annual Design Automation Conference*. ACM, 2007, pp. 9–14.

[11] R. Ulichney, *Digital halftoning*, MIT press, 1987.

[12] K. Solanki, U. Madhow, B. S. Manjunath, S. Chandrasekaran, and I. El-Khalil, "Print and scan resilient data hiding in images," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 4, pp. 464–478, 2006.

[13] L. Yu, X. Niu, and S. Sun, "Print-and-scan model and the watermarking countermeasure," in *Image and Vision Computing*. 2005, vol. 23, pp. 807–814, Elsevier.

[14] C. Baras and F. Cayre, "Towards a realistic channel model for security analysis of authentication using graphical codes," in *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*. IEEE, 2013, pp. 115–119.

[15] ISO/IEC 18004:2000, "Information technology - Automatic identification and data capture techniques - Bar code symbology - QR Code," 2000.

[16] Iu. Tkachenko, W. Puech, O. Strauss, J.-M. Gaudin, C. Destruel, and Guichard C., "Fighting against forged documents by using textured image," in *Signal Processing Conference (EUSIPCO), Proceedings of the 22th European*, 2014.