



Embedding biometric identifiers in 2D barcodes for improved security

Afzel Noore^{a,*}, Nikhil Tungala^a, Max M. Houck^b

^a*Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, United States*

^b*Forensic Science Initiative, West Virginia University, Morgantown, WV 26506, United States*

Received 18 March 2004; revised 22 August 2004; accepted 1 September 2004

KEYWORDS

Digital watermarking;
Biometrics;
Data security;
Authentication;
User identification

Abstract Two-dimensional (2D) barcode symbology is an emerging technology used for compactly storing and retrieving information. These barcodes can be found on the back of drivers' licenses and are encoded with secure text data. Standard 2D barcode such as PDF417 uses upper and lowercase alphabets, numeric digits and special characters for encoding. Some barcodes also include a compressed photo of the individual. The visual quality of the compressed image is usually poor and occupies a large amount of space which greatly reduces the capacity needed for encoding text. This paper presents a novel approach for embedding uncompressed images in a standard PDF417 2D barcode using a blind digital watermarking technique. The text is encoded in the standard PDF417 format with error correction, while the face and fingerprint images are watermarked in the encoded 2D barcode. Experimental results show that the proposed technique effectively increased the standard capacity of the PDF417 2D barcode without altering the contents of the encoded data. The results also show that the visual quality of the extracted photo image is high. The extracted fingerprint image when compared with the original fingerprint using an AFIS system yielded a high matching score.

© 2004 Elsevier Ltd. All rights reserved.

Introduction

The use of machine readable identity cards is proliferating to allow efficient transition from manual authentication to an automated process.

For security purposes, the cards need to be carefully scrutinized and the time taken to read the data and verify the accuracy should be kept as practical as possible (O'Gorman and Rabinovich, 1998). Identity cards are used in critical applications such as financial and medical transactions, access to restricted areas, border crossing, and in airlines (Ross, 2003; Vaidya, 1995; Chow et al., 1993). These documents are available in various

* Corresponding author. Tel.: +1 304 293 4821x2547; fax: +1 304 293 8602.

E-mail address: afzel.noore@mail.wvu.edu (A. Noore).

forms such as drivers' license, passports, credit cards and badges. The text information and biometric identifiers such as facial image, signature, fingerprint template, or iris template are typically stored on the card to enable verification of the identity of the owner. Recently 2D barcodes with extended data capacity have been used to compactly store information on ID cards. Currently there are over 20 2D barcode technologies available (Pavlidis, 2000). The PDF417 (Portable Data File 417) 2D barcode is widely adopted by major industries and various national and international organizations such as ANSI, International Civil Aviation Organization, and AIM because of its extended data capacity and error correction capabilities (AIM, 1994).

As the number of applications increases, there is a growing demand for embedding more data in 2D barcodes (O'Gorman and Pavlidis, 1999). The maximum capacity is largely constrained by the ability of the scanner to reliably read the encoded data and by the physical area available for encoding. For identification card applications such as a driver's license, a 2D barcode symbol on the back of the card stores the person's personal information to prevent tampering and identity theft. Fig. 1 shows a sample driver's license with the photograph and demographic text data of an individual and the 2D barcode on the back of the card. In addition, a photo or a fingerprint image can be encoded into a 2D barcode. Since the maximum data capacity of PDF417 is 1 KB, the images have to be compressed. However, as shown in Fig. 2, the quality of the image is poor. Also, the amount of available space for encoding the text data is considerably reduced.

One approach to improve the security of an identification document and increase the data capacity of the 2D barcode is digital watermarking. Watermarking is used to hide known information imperceptibly in digital data for preserving its integrity (Liu and Qiu, 2002; Meerwald and Uhl,

2001). The hidden information can also be retrieved from the printed and scanned images. Dittmann et al. (2001) presented a document authentication technique using hologram watermarks. The hologram watermarks are computer generated holographic images comprising personal identifiers and biometric attributes. The hologram watermarks are embedded into the personal data printed on the identification card. The embedded watermarks have weaker degree of robustness due to high redundancy in the encoded data. Another watermarking method related to identification card and document security is proposed by Wong et al. (1997). The image present on the document is protected by embedding a self-generated signature extracted from the image itself. The signature value embedded should be stored in order to identify the correctness of the data.

This paper proposes a novel approach to embed uncompressed face and fingerprint images without affecting the maximum text capacity of the PDF417 barcode. A blind watermarking technique is used for embedding and retrieving images. The proposed watermarking technique provides authenticity, integrity and verification of the 2D barcode image and its contents. The face and fingerprint images are used as watermarks and embedded into the dark regions of a barcode image. The following sections describe the architecture of PDF417 2D barcode symbology, the watermark embedding and decoding algorithm, the computation of pixels available for embedding a face and fingerprint image, and the results of validation. The watermarking process described in this paper is adapted for 2D barcode applications already in use. It represents a novel approach for increasing security by using biometric technologies. The minimal increase in overhead offsets the potential high return on investment due to decrease in costs emanating from fraud, tampering, or identity theft.

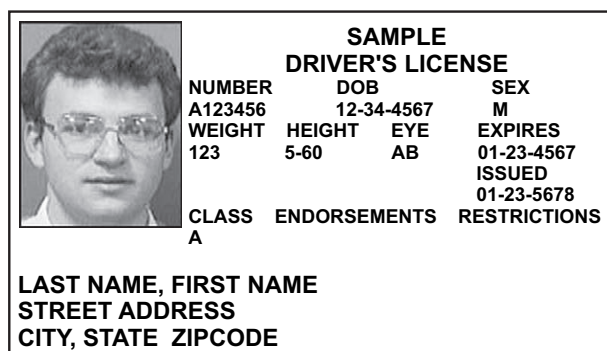


Figure 1 (a) Driver's license, (b) back of driver's license.



Figure 2 JPEG face image (a) uncompressed, 30 KB; (b) 30% compression, 950 bytes; (c) 10% compression, 719 bytes; and (d) 5% compression, 587 bytes.

Architecture of 2D barcode

Fig. 3 shows the structure of a PDF417 barcode (AIM, 1994). It is comprised of bars and spaces represented by dark and light regions, respectively. These bars and spaces are stacked into rows and columns. There is a start pattern on the left and a stop pattern on the right of the barcode indicating the beginning and ending of the symbol. All rows use the same stop and start patterns. Immediately after the start pattern and before the stop pattern are the left and right row indicators. These contain the information on the number of rows, the number of columns and the desired error correction level. In PDF417, data are stored in codewords that reside between the left row indicator and the right row indicator. Each codeword contains four pairs of alternate bars and spaces. The width of bars and spaces are unique to the codeword encoded. The minimum width of a bar or a space is referred to as the X dimension or module. Each codeword has a maximum of 17 modules.

Data codewords in the barcode are arranged in the form of rows and columns. The number of rows varies from 3 to 90 and the number of columns varies from 1 to 30. The first codeword is called the

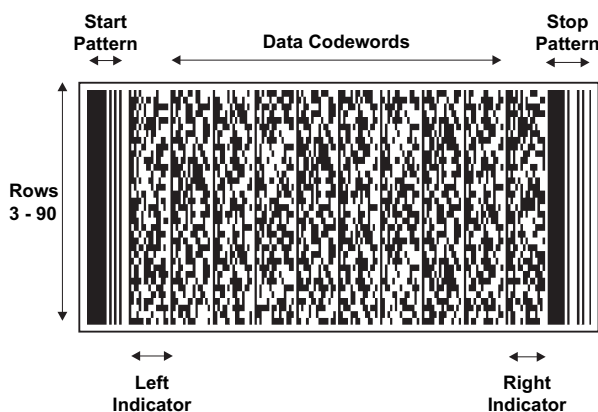


Figure 3 PDF417 2D barcode structure.

symbol length descriptor and encodes the total number of data codewords in the symbol. The actual data codewords to be encoded immediately follows the symbol length descriptor. The lower rows of the symbol are reserved for error correction codewords. There are a total of nine levels of error correction from 0 to 8. The higher the error correction level, the more the number of error correction codewords and space needed.

Watermarking algorithm

The 2D barcode is first generated with the demographic text data found in the front of the ID card. This ensures the efficient use of 2D barcode space for encoding data codewords and error correction codewords. We illustrate the digital watermarking technique by embedding the person's face and fingerprint images in the 2D barcode. For best visual quality, uncompressed images are used for embedding. The specifications of the fingerprint image and the face image are obtained from the NIST data interchange standard (INCITS/M1 03-0114, 2003; INCITS/M1 03-0116, 2003).

In the proposed watermarking algorithm, the face and the fingerprint images are used as watermarks and are embedded into the black bars of the 2D barcode image. The grayscale values of the watermark images vary from 0 to 255 and are divided into three different bit-streams, L , I , and M representing the least significant, intermediate, and most-significant integer values, respectively. The range of values for L , I and M are given by $0 \leq (L, I) \leq 9$ and $0 \leq M \leq 2$. Let I_{barcode} be the 2D barcode image of size $m \times n$. The pixels in the 2D barcode image that have a threshold $\leq T$, where, $T = L_{\text{max}}$, are selected for embedding. $O(i, j)$ are pixels of this image selected for embedding. The embedding process replaces the intensity value of the selected barcode pixels with the L , I and M bits, one at a time.

$$O(i, j) = \sum_{i=1}^m \sum_{j=1}^n l_{\text{barcode}}(i, j) < T \quad (1)$$

Let the face watermark image, W_f , be of size $p \times q$ where,

$$W_f(i, j) = 100 \sum_{i=1}^p \sum_{j=1}^q M_f(i, j) + 10 \sum_{i=1}^p \sum_{j=1}^q l_f(i, j) + \sum_{i=1}^p \sum_{j=1}^q L_f(i, j) \quad (2)$$

Let the fingerprint watermark image, W_{fp} , be of size $r \times s$ where,

$$W_{fp}(i, j) = 100 \sum_{i=1}^r \sum_{j=1}^s M_{fp}(i, j) + 10 \sum_{i=1}^r \sum_{j=1}^s l_{fp}(i, j) + \sum_{i=1}^r \sum_{j=1}^s L_{fp}(i, j) \quad (3)$$

The integer bit-streams L , I , and M are embedded into the pixel locations of $O(i, j)$ to obtain the watermarked barcode, $\hat{l}_{\text{barcode}}$.

$$\begin{aligned} O(i, j) &= \sum_{i=1}^p \sum_{j=1}^q L_f(i, j) \\ O(i+p, j+q) &= \sum_{i=1}^p \sum_{j=1}^q M_f(i, j) \\ O(i+2p, j+2q) &= \sum_{i=1}^p \sum_{j=1}^q l_f(i, j) \end{aligned} \quad (4)$$

$$O(k, l) = \sum_{i=1}^r \sum_{j=1}^s L_{fp}(i, j)$$

$$O(k+r, l+s) = \sum_{i=1}^r \sum_{j=1}^s M_{fp}(i, j)$$

$$O(k+2r, l+2s) = \sum_{i=1}^r \sum_{j=1}^s l_{fp}(i, j) \quad (5)$$

where $k = i + 3p$ and $l = j + 3q$.

This embedding method along with a secret key to select the pixel locations randomly from $O(i, j)$ will ensure the security of the watermark. A secret key is a random permutation of integers from 1 to N , where N is the width of the 2D barcode image. The random permutation is obtained by providing an input seed to a random number generator. The watermark is embedded in these randomly selected locations of the black bars. For added security, the seed is encrypted.

The extraction process of the watermark involves collecting all pixel values below the threshold T using the secret key and combining them into bit-streams, L , I , and M to generate the original watermark face image, W_f , and the fingerprint image, W_{fp} . Fig. 4 shows the embedding and decoding of face image and fingerprint image in a 2D barcode.

Our proposed watermarking approach is capable of detecting attacks such as data substitution, data alteration, and duplication of the ID card. Using sophisticated printers and scanners that are commonly available along with software tools for digital

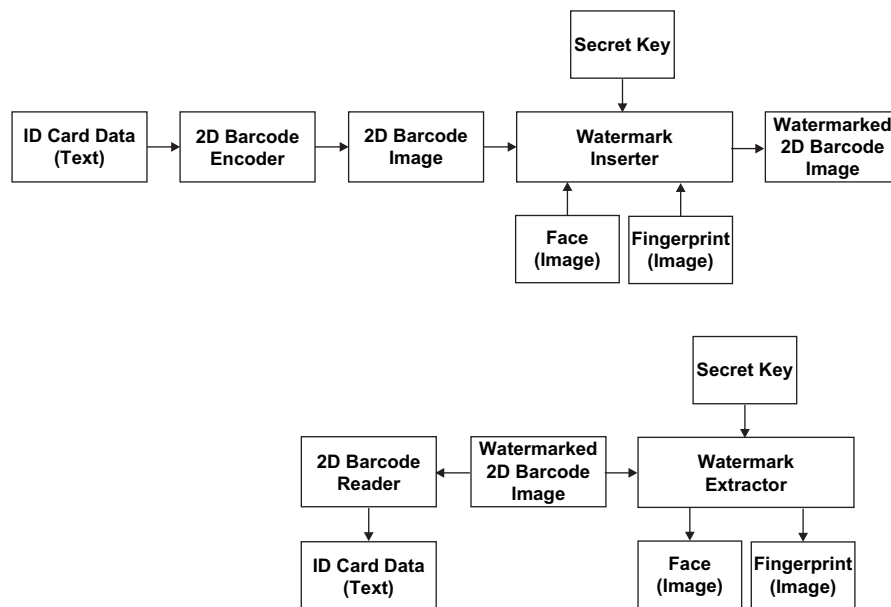


Figure 4 Block diagram of the proposed watermarking technique.

image editing and processing, counterfeiting and forgery can be easily performed. The encoded data in 2D barcodes can be substituted or data can be altered using barcode readers and encoders. However, the embedded watermarks provide authenticity, integrity and verification of the barcode image. Lack of watermarks or mismatch of the watermarks can reveal any attempt to tamper. The embedded watermarks are fragile. Any alteration in encoded data is manifested by changes in the embedded watermark locations. Also, the copying or duplication of identification cards will make the extraction of biometric attributes impossible without the encrypted secret key used for embedding. In addition, the real-time verification of the watermarked 2D barcode biometric identifiers with the actual biometric data obtained from the individual will determine if the owner of the identification card is genuine or an impersonator.

Computing the availability of sufficient embedding locations

In this section we compute the available embedding locations in a given PDF417 symbol to determine the feasibility of watermarking the face and fingerprint images. Let W and H represent the width and height of the 2D barcode symbol, respectively. The size of the barcode depends on four parameters – number of rows, R ; number of columns, C ; minimum width of a module (bar or space), X ; and height of the module, Y . From the structure of the PDF417 layout, the width of the barcode symbol is given by

$$W = (C17 + 69)X + 2Q \quad (6)$$

where $1 \leq C \leq 30$. The height of the barcode symbol is given by

$$H = (RY) + 2Q \quad (7)$$

where $3 \leq R \leq 90$

$$Q = 2X \quad (8)$$

The term Q in Eqs. (7) and (8) represents the minimum white space around the 2D barcode. This is referred to as the quiet zone and has a minimum value of $2X$. The height of the module, Y , can vary such that $X \leq Y \leq 5X$. The aspect ratio of each module $X:Y$ is chosen to be at least 1:3 to ensure that the scanner is able to decode the contents reliably. The PDF417 barcode specifications for drivers license and ID cards by ANSI and AAMVA

(American Association of Motor Vehicles Administrators) are as follows (AAMVA DL/ID, 2003):

$$H = 1.5 \text{ in}, \quad W = 2.975 \text{ in}, \quad X_{\min} = 0.0066 \text{ in}, \\ X_{\max} = 0.015 \text{ in}, \quad Y = 3X \text{ and } Q = 1X$$

The number of modules in a given row = $W/X_{\min} = 450$ and the number of rows, $R = H/Y = 75$. The total number of modules in the 2D barcode symbol is 33,750.

From the codewords generated for all possible characters, upper and lower case alphabets, numeric digits, special characters and ASCII characters, we observe that there exists a minimum of seven bars (dark modules) from a total of 17 combination of bars and spaces for each character. The minimum number of dark modules available in a given symbol is 13,897. If the printer resolution is 600 dpi, then the number of pixels in a module = $600X_{\min}600Y = 46$ pixels. The minimum number of pixels available for embedding in the 2D barcode symbol is 639,262 pixels. Increasing the error correction level from 1 to 8 will increase the number of error correction codewords from 4 to 512 and will reduce the number of codewords for encoding the text characters. Since the size of the barcode image is fixed, the total number of codewords remains the same. The minimum number of bars generated for any possible codeword is 7 and therefore the number of black intensity pixels available for embedding the watermarks is also the same.

According to the INCITS (International Committee for Information Technology Standards) interchange format a face image of size 240×320 (30 KB) and a fingerprint image of 300×300 pixels (58 KB) are chosen. Using the proposed embedding process the grayscale watermark images are divided into three integer bit-streams, L , I and M , representing the least significant, intermittent and the most-significant integers of the pixel intensity values. The total number of black pixels on the barcode image required for embedding is three times the size of each watermark image. The number of pixels needed for embedding the facial image is 230,400 pixels and the number of pixels needed for embedding the fingerprint image is 270,000 pixels. The total number of 500,400 pixels of watermark information can be easily embedded in the 639,262 black pixel locations of the given barcode image.

Experimental results

The specifications for PDF417 2D barcode were selected from NIST and AAMVA standard for

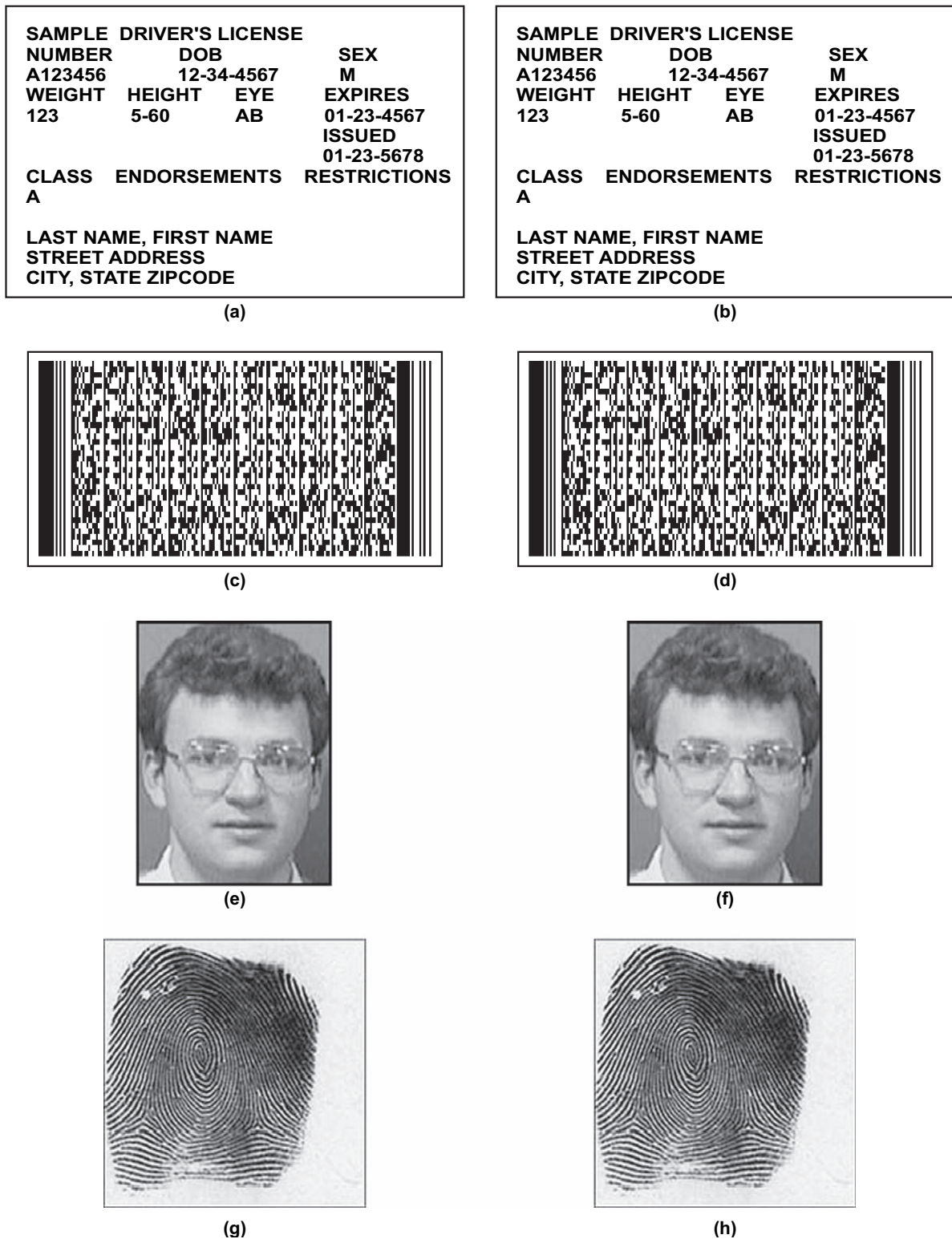


Figure 5 (a) Original ID card data, (b) decoded ID card data, (c) 2D barcode encoded with text, (d) 2D barcode watermarked with face and fingerprint images, (e) original face image, (f) extracted face image, (g) original fingerprint image, and (h) extracted fingerprint image.

driver's license and ID card design. The width of the barcode, $W = 2.975$ in and the height of the barcode, $H = 1.5$ in. An error correction level of 5 is used. The maximum number of text characters for all required fields is 183. There are several optional fields that could be embedded. We have designed a sample driver's license with 186 characters comprising both the required fields and some optional fields.

The 2D barcode image is watermarked with the face image, W_f , and fingerprint image, W_{fp} , using the proposed watermarking algorithm. The watermarked barcode image is decoded with a PDF 2D barcode decoder and the encoded text is extracted. There was 100% match between the 256 codewords that were generated during encoding and the codewords extracted after decoding. There was also 100% match between the original text in the ID card and the decoded text extracted from the watermarked barcode image. The results show that the accuracy and readability of the 2D barcode was not affected when the face image and the fingerprint image were watermarked in the 2D barcode. The two watermark images W_f and W_{fp} are next extracted from the watermarked barcode image using the proposed algorithm and compared with the original images. Fig. 5 shows that the extracted images are of high quality and closely resemble the original image.

Table 1 summarizes the visual quality metrics of the barcode and the watermark images. Measures such as the peak signal to noise ratio (PSNR) and the correlation between the original and the extracted watermark images are computed to compare the quality. These traditional image quality measures are pixel-based. However, recent measures are based on the human visual system and are better suited for visual perception (Wang et al., 2002, 2004; Van der Weken et al., 2003). The Structural Similarity Measure (SSIM) (Wang et al., 2004) and the Universal Image Quality Index (UIQI) (Wang et al., 2002) were computed for the original images and the corresponding watermarked 2D barcode, face image, and fingerprint image. In all cases, the quantitative measures indicate that there is a high correlation and integrity between the original and processed images.

Besides computing the visual quality measures, the original fingerprint and the extracted fingerprint were included in an AFIS database and the matching score of the second best fingerprint was obtained. Fig. 6 shows that the matching scores for the original and extracted fingerprints were high, verifying that besides the high visual quality of the image, the watermarking and extraction process did not alter the matching characteristics of the fingerprint.

Table 1 Visual quality of original watermarks and extracted watermarks

Image	Pixel-based metrics		Human visual system based metrics	
	PSNR	Correlation	SSIM	UIQI
2D barcode	86.47	0.9999	0.9686	0.9563
Face image	64.33	0.9861	0.7384	0.9091
Fingerprint image	58.87	0.972	0.8326	0.9408

Conclusions

In this paper a novel watermarking technique is used to embed uncompressed face and fingerprint images in 2D barcodes of an identity card for improved security. This technique does not affect the original utility and storage capacity of the 2D barcode. The embedding of face and fingerprint images in the ID card provides additional security, deters tampering and prevents misuse of identity cards. This additional level of security is transparent to the holder of the bar-coded item. The secure information is also imperceptible and not readily apparent to a person who wants to tamper. Additionally, without the proper key, the potential for abuse of personal information is very low. This could address the concerns of many citizens and privacy advocates about the misuse or loss of personal information through the distribution of biometric technology. It is observed that the original text of the 2D barcode is unaltered when the face and fingerprint images are embedded in the watermarking process. Experimental results show that the extracted face and fingerprint images resemble the original images. The visual quality metrics computed show that there is a close correlation between the extracted and original images. The fingerprint matching performed on an AFIS also verifies that the embedding and extraction process did not alter the characteristics of the

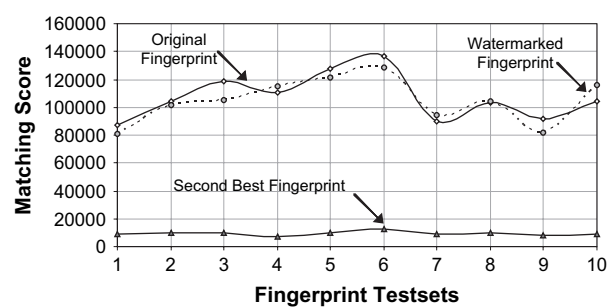


Figure 6 Matching scores of original, watermarked and the second best fingerprints.

fingerprint needed to perform matching. The ubiquity of the 2D barcode system makes it potentially more vulnerable to tampering and fraud. The watermarking process described in this paper reinforces the security of this system while retaining the logistical systems already in use. It would therefore increase security while keeping costs steady or even reducing them through the prevention of inappropriate or illegal activities.

Acknowledgments

This research (Award No. 2003-RC-CX-K001) was supported by the Office of Science and Technology, National Institute of Justice, Office of Justice Programs, US Department of Justice. The authors wish to thank Sagem Morpho for donating the AFIS system used in performing this research.

References

- AAMVA DL/ID-2003. American Association of Motor Vehicle Administrators (AAMVA) International Specification for the Drivers License and Identification Card Design; 2003.
- AIM 1994. Uniform Symbology Specification: PDF417; 1994.
- Chow S, Serinken N, Shlien S. Forgery and tamper-proof identity document. In: Proceedings of IEEE international carhahan conference on security technology; October 1993. p. 11–4.
- Dittmann J, Ferri LC, Vielhauer C. Hologram watermarks for document authentications. In: IEEE international conference on information technology: coding and computing; 2001. p. 60–4.
- INCITS/M1 03-0114. Face recognition format for data interchange; 2003.
- INCITS/M1 03-0116. Finger minutiae format for data interchange; 2003.
- Liu T, Qiu Z-D. The survey of digital watermarking-based image authentication techniques. In: 6th International conference on signal processing, vol. 2. 2002. p. 1556–9.
- Meerwald P, Uhl A. A survey of wavelet-domain watermarking algorithms. Proceedings of SPIE, electronic imaging, security and watermarking of multimedia contents III, vol. 4314. January 2001. San Jose, CA.
- O’Gorman L, Pavlidis T. Auto ID technology: from barcodes to biometrics. IEEE Robotics and Automation Magazine 1999; 6(1):4–6.
- O’Gorman L, Rabinovich I. Secure identification documents via pattern recognition and public-key cryptography. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998;1097–102.
- Pavlidis T. A new paper/computer interface: two-dimensional symbolgies. In: IEEE international conference on pattern recognition, vol. 2. Barcelona, Spain. 2000. p. 145–51.
- Ross D. Back on the cards. IEE Review November 2003;49(11): 22–3.
- Vaidya AW. Keeping card data secure at low cost. European Convention on Security and Detection May 1995:212–5.
- Van der Weken D, Nachtegaal M, Kerre EE. A new similarity measure for image processing. Journal of Computational Methods in Sciences and Engineering 2003;3(2):209–22.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. A universal image quality index. IEEE Signal Processing Letters 2002;9(3): 81–4.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error measurement to structural similarity. IEEE Transactions on Image Processing 13(1); 2004.
- Wong KK, Tse CH, Ng KS, Lee TH, Cheng LM. Adaptive water marking. IEEE Transactions on Consumer Electronics 1997; 43(4):1003–9.
- Afzel Noore** received his Ph.D. in Computer Engineering from West Virginia University. He is an Associate Professor in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. From 1996 to 2003, he served as the Associate Dean for Academic Affairs and Special Assistant to the Dean in the College of Engineering and Mineral Resources at West Virginia University. His research interests include biometrics, fuzzy and neural systems, fault-tolerant computing, software reliability, and consumer electronics. His research has been funded by Westinghouse, GE, NSF, Electric Power Research Institute, Department of Energy, and the US Department of Justice.
- Nikhil Tungala** is a graduate research assistant in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. He is currently completing his M.S. degree in Electrical Engineering. His research interests are in digital watermarking, biometrics and image processing.
- Max M. Houck** is the Director of the Forensic Science Initiative, a program that develops research and professional training for forensic scientists and related professionals. Houck is a trace evidence expert and forensic anthropologist assigned to the Trace Evidence Unit at the FBI Laboratory from 1992 to 2001. Just before he joined WVU, he was assigned to Dover Air Force Base, Delaware, to assist with the examination and identification of the victims of the 9/11/01 Pentagon attack victims. He was the recipient of a Quality Award in 1999 from the FBI Laboratory and the American Society of Testing and Materials Forensic Science Award in 2000. He is a graduate of Michigan State University, and a Fellow of the American Academy of Forensic Sciences.

Available online at www.sciencedirect.com

