# Image Watermaking With Biometric Data For Copyright Protection

Barbier, Morgan*[1], Le Bars, Jean-Marie†[1], and Rosenberger, Christophe‡[1]

[1]ENSICAEN-UNICAEN-CNRS, GREYC, F-14032 Caen, France

March 16, 2015

**Abstract**

In this paper, we deal with the proof of ownership or legitimate usage of a digital content, such as an image, in order to tackle the illegitimate copy. The proposed scheme based on the combination of the watermarking and cancelable biometrics does not require a trusted third party, all the exchanges are between the provider and the customer. The use of cancelable biometrics permits to provide a privacy compliant proof of identity. We illustrate the robustness of this method against intentional and unintentional attacks of the watermarked content.

**Keywords:** Proof of ownership, biometrics, watermarking, BioHashing, biometric commitment.

## 1 Introduction

In the last decade, intensive work led to different methods to manage the author rights of images using cryptographic protocols [3, 7]. However, these schemes ensure a low security guarantee of the data owner. Indeed, these systems cannot link the owner identity to its use rights, this is intrinsic to cryptography where the security is related to the knowledge of a secret, not to an identity. In order to tackle this problem, some researchers thought to use biometric systems. The embedding of biometric data into an image has been proposed for the first time in 2004 to link the photographer iris to taken pictures [2]. The same technique

*morgan.barbier@ensicaen.fr
†jean-marie.lebars@unicaen.fr
‡christophe.rosenberger@ensicaen.fr

was also proposed for the purpose of multi-biometric applications consisting in embedding fingerprint into visage pictures [9, 4], but not any to prove the ownership of an image. The previous solutions embed very sensitive data into a shared image, which affect hugely to the user privacy and appears a significant problem. Recent schemes have been proposed in order to obtain cancelable biometric data [8]. This process is called BioHashing. The knowledge of this data does not provide any information of the original biometric data, which ensures the user privacy. However, it implies to be carefull with the seed initialization [5]. Indeed, if an attacker knows both BioCode and this seed, he can forge a pretty a good estimate of the FingerCode.

In this article, we present a new scheme of proof of ownership exploiting cancellable biometrics. With the help of cryptographic protocols, the proposed method ensures the watermark security and preserve the owner privacy. A particular value is shared between these two actors and can be seen as a biometric commitment. As far we know, it is the first time that a biometric commitment is done, and this pretty property avoids a trusted third party; which is a significant improvement for some applications. In this setting, this article is a direct improvement of [1].

First, we define the requirements that a such system has to respect for the different actors in terms of security and privacy. Second, we detail the proposed method. Experimental results show the benefit of this solution. We perform in the following a security and privacy analysis, where we discuss about the respect of the previous requirements. Finally, we conclude this paper and give different perspectives of this study.

## 2 Security and privacy requirements

A biometric system handles, by definition, very sensitive personal data of users. These data aim to be protected for prevent their usurpation, modification or falsification. In our context, we define two main actors. The first one is the **customer** who wishes to acquire the copyright of a digital data proposed by a **provider** or **owner**. Afterwards, the provider should be able to check that a customer has the copyright for a particular data. Moreover, a customer should be able to prove he personally had previously obtained the right to use this digital content. In this context, we introduce the main security and privacy requirements of the previous system:

$R_1$: **Proof of ownership** assures that the legitimate customer can prove, at any moment, his right to use the data.

$R_2$: **Proof of paternity** assures that the provider can prove, at any moment, its paternity of the digital content.

$R_3$: **Unlinkability** of the watermarks for a same user in different data. An attacker should not be able to link the different watermarks of a customer.

$R_4$: **Confidentiality of customer data** ensures to not be known by anybody.

$R_5$: **Confidentiality of provider data** ensure to not be known by anybody.

$R_6$: **Customer data sovereignty** implies that the copyright checking can be done only with the agreement of the customer.

$R_7$: **Provider data sovereignty** implies that the copyright checking can be done only with the agreement of the provider.

$R_8$: **Non-falsification** prevents a customer/attacker to make a legitimate mark.

$R_9$: **Non-repudiation** prevents a provider to deny the sale of a media.

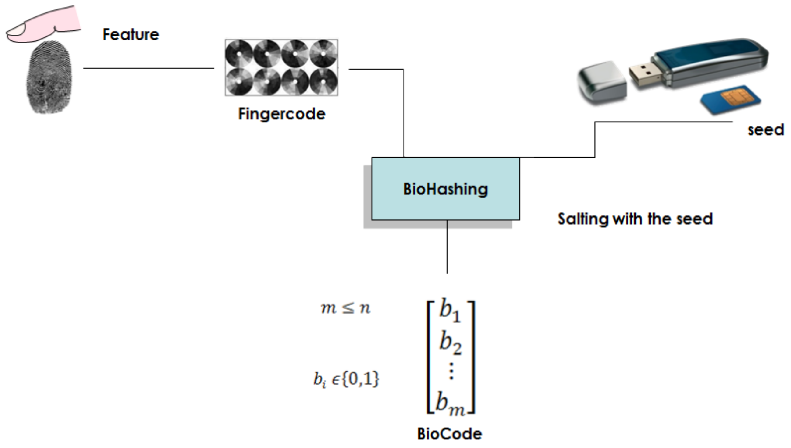


Figure 1: Use of the BioHashing algorithm

# 3 Proposed method

The proposed method has for objective to protect at the same time the ownership of an image and the right of use for a customer. As biometrics is the only technology that really guarantees the user identity, we use this kind of information in the proposed method. We embed a mark into an image in order to prove these two aspects. The watermark is computed from a cancelable biometric data to permit a secure and privacy compliant identity verification. We detail in the next section the used watermarking method.

### 3.1 Image watermarking

In the proposed approach, we use the watermarking method introduced by Wenyin and Shih [10]. This method uses image texture parameters so-called Local Binary Patter (LBP), to select pixels to embedded the mark. It ensures a good robustness to different distortions like compression and cropping simulating intentional and unintentional attacks of the watermarked image. Figure 2 an example of watermarked image with this approach, the difference between the original and the watermarked images multiplied by 10 is displayed as illustration.

Figure 2: From left to right: original image, watermarked, difference (multiplied by 100)

### 3.2 Image identifier computation

The copyright protection must be verified easily and be related to the image. We propose to define an image identifier that is easy for anybody to compute. The image is divided in blocks where the number is related to the size in bits of the image identifier. As for example, Figure 3 shows the method to compute an identifier with 16 bits. We compute first the average gray level value called $E[I]$ where $I$ is the image. Second, we compute the average grey level value of each block (denoted as $E[Block]$) and we compare it to $E[I]$. Based on the result, we assign a value 1 or 0 to each of the bit of the image identifier.

Figure 3: Image identifier computation

The BioHashing algorithm is used to compute the cancellable biometric data,

it is detailed in the next section.

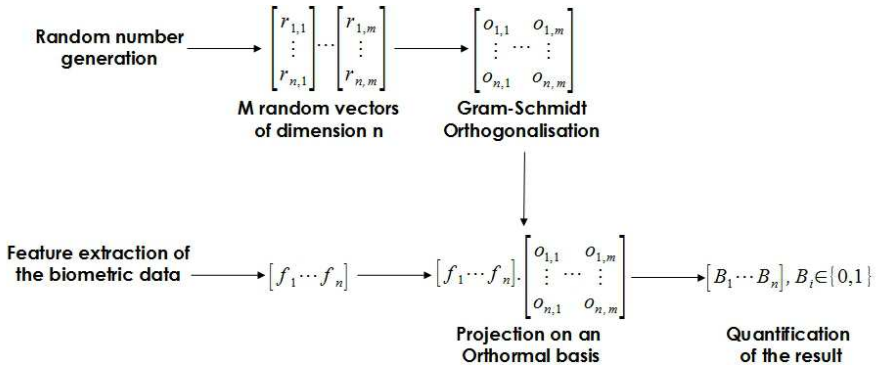


Figure 4: Principle of the BioHashing algorithm

### 3.3 BioHashing algorithm

The BioHashing algorithm transforms a real-valued vector of length $n$ (i.e. the FingerCode, resulting from a feature extraction method) into a binary vector of length $m \leq n$ (i.e. the BioCode), as first defined by Teoh *et al.* in [8].

It consists in projecting the FingerCode on an orthogonal basis defined by a random seed (considered here as a secret), to generate the BioCode. The template transformation uses the following algorithm, where the inputs are the random seed and the FingerCode $F$; the output is the BioCode $B$:

1. For $i = 1, \ldots, m$, $m \leq n$ pseudorandom vectors $v_i$ of length $n$ are generated (from the random seed) and are gathered in a pseudorandom matrix.

2. The Gram-Schmidt algorithm is applied on the $m$ vectors $v_i$ of the matrix, for the generation of $n$ orthonormal vectors $V_1, \ldots, V_m$.

3. For $i = 1, \ldots, m$, $m$ scalar products $p_i =< F, V_i >$ are computed using the FingerCode $F$ and the $m$ orthonormal vectors $V_i$ .

4. The $m$-bit biocode $B = (B_0, \ldots, B_m)$ is finally obtained thanks to the following quantization process:

$$B_i = \begin{cases} 0 & \text{if} \quad p_i < t \\ 1 & \text{if} \quad p_i \geq t, \end{cases}$$

   where $t$ is a given threshold, generally equal to 0.

When used for authentication the *Reference BioCode* (computed from the FingerCode after enrollment and after exhibiting the secret) is compared with the *Capture BioCode* (computed from the FingerCode computed after a new

capture with the secret) with the Hamming distance. If this value is lower than a specified threshold set by the system administrator, the identity of the user is checked. Roughly speaking, the first part of the algorithm, including the scalar products with the orthonormal vectors, is used for the performance requirements and the last step of the algorithm is used for the non-invertibility requirements of the BioHashing algorithm. As mentioned before, the random seed guarantees the diversity and revocability properties.

The user authentication protocol applies multiple times the BioHashing algorithm.

### 3.4 Mark insertion

Figure 5 describes the proposed method to insert in an image some information related to the image's owner and customer. The mark we embed is defined as the repetition the global Biocode (16 times of a 256 bits BioCode to generate the 64x64 pixels to fulfill the requirements of the used watermarking method). The seed is a secret defined by the owner. The value $H(H^k(seed) \oplus \ image\ identifier)$ can be seen as a cryptographic commitment shared between the image's owner and customer. Each time a BioCode is issued, the value of $k$ is incremented (to guarantee different commitments for different applications of the watermarking approach). The term $H^k(seed)$ corresponds to an OTP (dynamic value) generated from a seed. The use of the BioHashing algorithm ensures to not be able to obtain the data at a lower level as it is a non invertible function. Note that the customer should provide his/her fingerprint (as identity proof) and a password (that can be the same for different images).

The embedded mark in the image biometric data from both the image's owner and user (concatenation of two BioCodes). The verification of property or right to use is possible by considering fingerprint as identity proof for owner or user side. Note that at the first level, the owner's BioCode is not related to the image as the right of user could be provided multiple times. For user side, the BioCode is related to the image as an user will buy only once the right of use for a specific image.

### 3.5 Mark verification

If the owner finds its image on internet, the mark can be extracted (if present). First, based on the image identifier that can be computed from this image, the owner can look at existing BioCodes for this image. If this BioCode exists in his/her database, the $k$ value is determined. Second, the right of use must be checked. To achieve this goal, the webmaster of the website where the image is used can be contacted in order to verify this point. In this case, the user's fingerprint should be given and the user's BioCode generated (with the $k$ value set previously). If the Hamming distance between this BioCode and the one present in the image matches is lower a certain threshold (set by the owner),

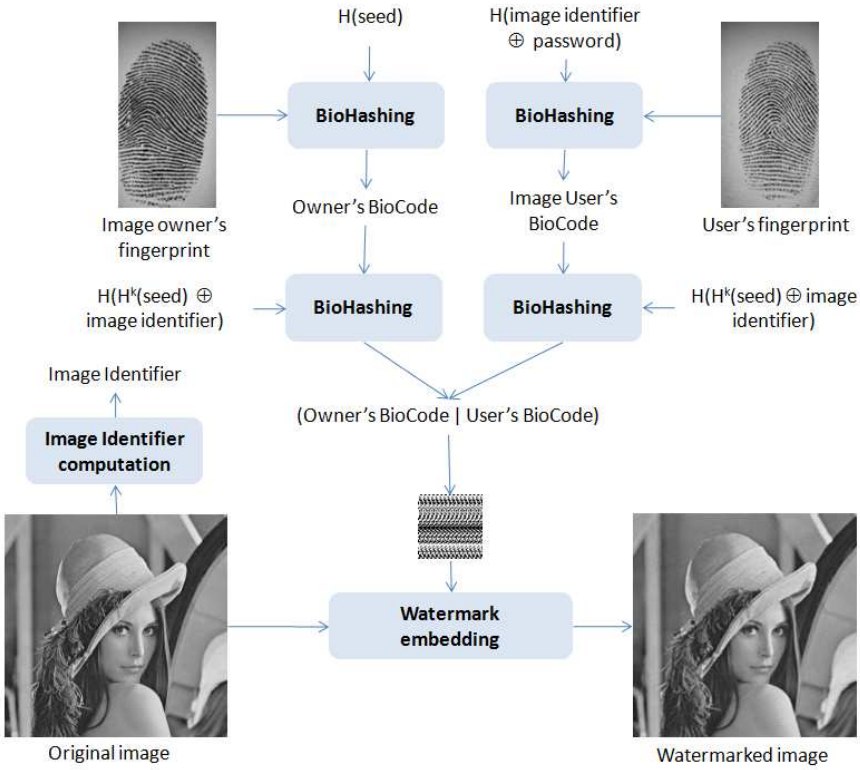


Figure 5: Inserting the mark in an image

the user's right of use is validated.

The proposed method permits to verify at the same time the ownership and the right of use. An attacker is not able to generate a valid user's BioCode as he/her does not know the cryptographic commitment.

# 4 Experiments

In this section, we analyze first the robustness the watermarking method face to attacks. Second, we analyze the performance of the biometric recognition (owner or user side) even if the watermarked image has been altered. In the next section, we define the experimental protocol used in this study.

## 4.1 Protocol

In this study, we used three fingerprint databases, each one is composed of 800 images from 100 individuals with 8 samples from each user:

- FVC2002 benchmark database DB2: the image resolution is 296 × 560 pixels with an optical sensor "FX2000" by Biometrika ;
- FVC2004 benchmark database DB1: the image resolution is 640 × 480 pixels with an optical Sensor "V300" by CrossMatch ;
- FVC2004 benchmark database DB3: the image resolution is 300 × 480 pixels with a thermal sweeping Sensor "FingerChip FCD4B14CB" by Atmel.

Figure 6 presents one image from each database. We can see that fingerprints are quite different and representative of the different types of fingerprint (acquired with sensors using different technologies).

(a) (b) (c)

Figure 6: One fingerprint example from each database: (a) FVC2002 DB2, (b) FVC2004 DB1, (c) FVC2004 DB3

As FingerCode, we used Gabor features (GABOR) [6] of size n=512 (16 scales and 16 orientations) as template. These feature are very well known and permit a good texture analysis of a fingerprint. For each user, we used the first FingerCode sample as reference template. Others are used for testing the proposed scheme. BioCodes are of size m=256 bits. In order to quantify the performance of the proposed approach, we computed 1400 comparisons (with the Hamming distance) between the BioCode embedded in the watermarked image and the computed one for each user.

The evaluation process is detailed in Figure 7. We apply many alterations on the watermarked image (illustrated on Figure 8) simulating active or passive attacks.

The robustness of the watermarking algorithm is estimated for the Error Bit Rate metric (EBR) defined as follows. The performance of the biometric recognition is determined by the Equal Error Rate (EER). This metric computes

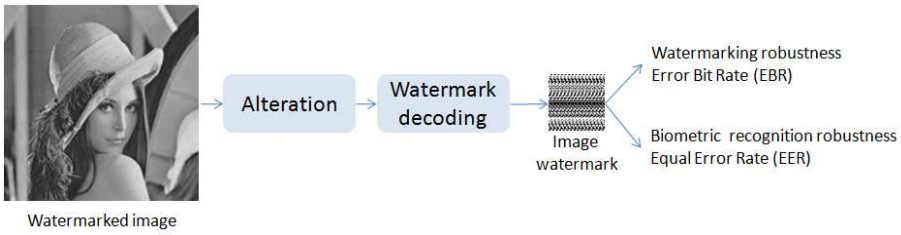


Figure 7: Evaluation process of the proposed approach

the performance of the biometric system when the threshold is set to have a compromise between the false acceptance rate and false rejection rate.

$$EBR = \frac{\sum\sum C(x,y) \oplus \tilde{C}(x,y)}{M.N}$$

Where $C(x,y)$ is the initial value of the mark at pixel $(x,y)$, $\tilde{C}$ is the decoded mark, M and N are respectively the number of lines and columns of the mark (in our case, N=M=64), the symbol $\oplus$ corresponds to a logical XOR.

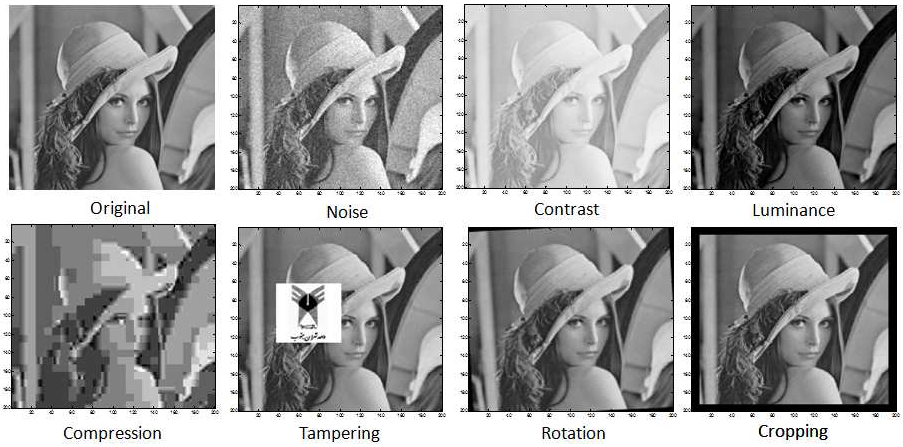


Figure 8: Alterations of the watermarked image

# 5 Results

First, we study the robustness othe image watermarking method face to different alterations illustrated in Figure 8. Figure 9 presents the evolution of the EBR for each alteration. We can see that the watermarking method is completely invariant to the contrast alteration. Some alterations have as impact a low modification of the mark such as the luminance, cropping or tampering.

Other alterations have strong impact of the mark.

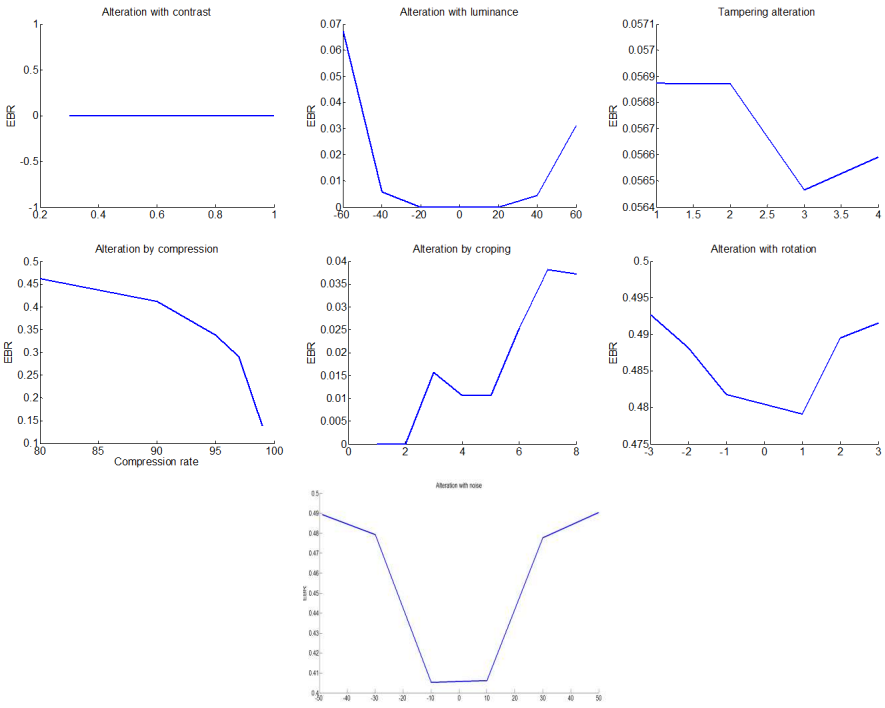


Figure 9: EBR evolution for different distortions

Figure 10 gives the biometric recognition performance (of the owner or user) when the watermarked image is attacked considering different alterations. We can see that for invisible alteration, the EER value is near 0%. For more visible alterations, EER values can be very high but there is no great benefit for an attacker to be able to suppress the copyright with such an alteration of the image. These results shows the robustness of the proposed approach (that can be of course improved).

# 6 Security and privacy analysis

## 6.1 Attack model

We consider the strongest attack model where the attacker is able to modify data. However, its actions will be limited thanks to using SSL channel to preserve communication confidentiality, actor authentications and data integrity.

| No alteration | Crop 1 line | Crop 2 lines | Crop 3 lines | Crop 4 lines |
|---|---|---|---|---|
| 0% | 0% | 0% | 0% | 0% |
| Crop 5 lines | Crop 8 lines | Noise -60 | Noise -40 | Noise -20 |
| 0% | 0% | 39.2% | 30.3% | 0.2% |
| Noise +20 | Noise +40 | Noise +60 | Contrast 1 | Contrast 0.7 |
| 0% | 30.8% | 40.5% | 0% | 0% |
| Contrast 0.5 | Contrast 0.3 | Luminance -60 | Luminance -40 | Luminance -20 |
| 0% | 0% | 0% | 0% | 0% |
| Luminance +20 | Luminance +40 | Luminance +60 | Compression 99% | Compression 97% |
| 0% | 0% | 0% | 0% | 0% |
| Compression 95% | Compression 90% | Compression 80% | Tampering | Rotation -3° |
| 0% | 0.2% | 11% | 0% | 48.5% |
| Rotation -2° | Rotation -1° | Rotation +1° | Rotation +2° | Rotation +3° |
| 48% | 42.1% | 29% | 47.6% | 47.1% |

Figure 10: Biometric recognition performance for the FVC2002 DB2 dataset. Results are similar for the two other datasets.

Without lost in generality, we may assume that the attacker is passive for communications and active to attempt to create legal documents, to usurp identities and to link the customer sales.

## 6.2 Analysis

In this section, we discuss about the proposed system and the security and privacy requirements introduced in Section 2.

$R_1$ In order to prove that a customer is legitimate, the court begin to check the provider BioCode, then OTP value is now known, and the provider is engaged on this value. Finally, the court computes the customer BioCode with the hash of the previous OTP. If the provider tries to cheat with the OTP in order to harm the legitimate customer, the provider BioCode computed cannot be checked. Moreover, our system is resilient to some modifications carry out by the legitimate customer, as previously seen Section 4. Thus as long as the customer perform some modifications on its digital content listed in Section 4 like crop, contrast, luminance and decent noise and compression; the customer will be recognized as legitimate, so the requirement is respected.

$R_2$ The first part of the embedded mark is dedicated only to the provider; this computation is based on the provider fingercode, some image characteristics and the seed. All these previous values can be supply by the provider, to proof the paternity of the digital content.

$R_3$ Since the BioHashing is based on a projection of vector into a space with less dimension, this operator is not bijective and is not inversible. Moreover, for each digital content, the BioCode will be completely different, then as far we know, it is impossible to make the link from different marks to a customer; this requirement is satisfied.

$R_4$ To access to the customer data, an attacker has to inverse twice the Bio-Hashing, which is impossible. However, estimations can be computed if and only if the seed is known; which is not the case up to customer's fingercode; $R_4$ is fulfilled.

$R_5$ For the same argues that for the requirement $R_4$, we deduce that $R_5$ is satisfied.

$R_6$ Since provider data used is only its FingerCode by the BioCode, it cannot be manipulated without its permission. The requirement $R_4$ is fulfilled.

$R_7$ For the same argues that for the requirement $R_6$, we deduce that $R_7$ is satisfied.

$R_8$ The embedded mark is generated thanks to the provider's FingerCode and an ID related to the digital content. Hence the marks are totally different and an attacker cannot forge a legitimate mark. The requirement $R_8$ is satisfied.

$R_9$ Finally, if the provider suspect an user to use a digital content without to be the right owner, the provider will take this user to court. Then the justice will make the whole computation asking provider and customer's FingerCodes to tell the final decision.

# 7 Conclusion and perspectives

This paper contains a new method to achieve a proof of ownership of a digital content with a biometric scheme preserving the privacy. The resiliency of various distortions simulate an intentional and unintentional attacks of the watermarked content shows a good robustness of the introduced scheme for biometric checking. Our main goal was to avoid a trusted third party and then our scheme is entirely satisfactory.

However, some parts of the scheme could improved. Indeed further investigations will be conducted to make the watermark scheme more efficient. We may apply other watermark insertion algorithms and replace the repetition code to a more adapted binary codes. These modifications should lead to a more robust scheme. We could also consider color image; since such a image contains more information, we can expect to improve the robustness.

# References


[1] Morgan Barbier and Christophe Rosenberger. Tatouage d'images avec des données biométriques révocables pour la preuve de propriété. In *Conférence sur la Sécurité des Architectures Réseaux et des Systèmes d'Information (SAR SSI)*, 2014.

[2] Paul Blythe and Jessica Fridrich. Secure digital camera. In *in Proceedings of Digital Forensic Research Workshop (DFRWS)*, pages 17–19, 2004.

[3] Nelly Fazio. *On Cryptographic Techniques for Digital Rights Management*. PhD thesis, New York University, September 2006.

[4] Nikos Komninos and Tassos Dimitriou. Protecting biometric templates with image watermarking techniques. In Seong-Whan Lee and StanZ. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 114–123. Springer Berlin Heidelberg, 2007.

[5] Patrick Lacharme, Estelle Cherrier, and Christophe Rosenberger. Reconstruction attack on biohashing. In *International Conference on Security and Cryptography (SECRYPT)*, pages 8–p, 2013.

[6] B. S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:37–42, 1996.

[7] Saraju P. Mohanty. A secure digital camera architecture for integrated real-time digital rights management. *Journal of Systems Architecture*, 55(10-12):468 – 480, 2009.

[8] A.B.J. Teoh, D. Ngo, and A. Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern recognition*, 40, 2004.

[9] M. Vatsa, R. Singh, A. Noore, M. Houck, , K. Morris, Mayank Vatsa, Richa Singh, Afzel Noore, and Keith Morris. Robust biometric image watermarking for fingerprint and face template protection, 2006.

[10] Zhang Wenyin and Frank Y. Shih. Semi-fragile spatial watermarking based on local binary pattern operators. *Elsevier journal on Optics Communications*, pages 3904–3912, 2011.