ADVANCED SEMANTIC AUTHENTICATION OF FACE IMAGES

Huajian Liu[1], *Hichem Sahbi*[2], *Lucilla Croce Ferri*[1], *Martin Steinebach*[1]

[1] Fraunhofer IPSI - Integrated Publication and Information Systems, Darmstadt, 64293, Germany
[2] IMEDIA Group, INRIA Rocquencourt, BP 105-78153, Le Chesnay Cedex, France
Email: liu@ipsi.fraunhofer.de

ABSTRACT

We propose a novel and flexible image authentication scheme in this paper, which can verify face images at different levels by digital watermarking. As the semantic content of the image is taken into account in our approach, the proposed scheme can distinguish the manipulations among faces from those in the background. It also provides more information to help inferring an attacker's motive. This capability makes it more flexible in practical applications. Experimental results demonstrate the ability of the proposed scheme to provide advanced authentication for face images.

1. MOTIVATION

One of the application fields of digital watermarking is image authentication which allows us to recognize manipulations in images. Traditional security mechanisms from cryptography, for instance hash based methods, only provide a bit-wise authentication, which can not distinguish between the image content and its digital representation. Furthermore, most of the existing watermarking schemes provide only a holistic protection of the image without taking into account the underlying semantic content. However, usually only the semantic content of the region of interest (ROI) is of interest for the user in many applications, while the other parts may change without influencing this content.

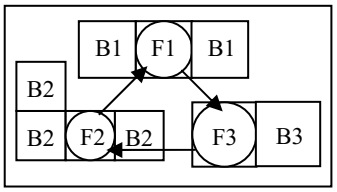
Holistic watermarking schemes, for instance [1][2], are not suitable to protect the ROIs semantically. For instance, data annotation in digital libraries may require adding visual content into images such as logos for copyright protection. These operations, including adding or cropping out logos, should not be interpreted as malicious manipulations in the context of integrity verification, as long as they are applied outside the ROIs, although these processing changes the image content. ROIs can also be famous faces provided by a photo-agency. In this case, protecting faces enables their publication with different backgrounds, possibly after format conversion. Still any illegal manipulation of the portraits can be recognized. These face manipulations can be targeted to change the identity, behavior and position of one or different persons in a scene. Another application is authenticity verification of scenes recorded by a camera-based surveillance system. In this framework, face substitution should be recognized as malicious, while changes like annotations in the background or outside the faces should leave the image authentic. This shows that additional protection mechanisms for ROIs are required, in order to distinguish between ROIs and background alterations, in contrast to current authentication solutions, which consider images manipulated only in the background as unauthentic.

We introduce an approach for integrity verification of slightly modified digital images. The modifications could produce content changes, but they are accepted only outside the ROIs. It follows our previous work [3] with more emphasis on practical applications and presenting more experimental results. We consider human faces as particular ROIs which are increasingly important for security issues and massively present in different visual contents. We provide semantic protection of images with different levels of security in different image regions. Furthermore, our approach enables partially manipulation trace and the identification of some attacks on face regions, which can help us to infer the attacker's motives. In this framework, automatic face detection is a preliminary step to the embedding process, which makes the total watermarking approach feasible. The general concept is not restricted to faces; alternative detection algorithms can be applied to identify other types of ROIs as human bodies and cars.

The paper is organized as follows: In §2, the watermarking scheme is introduced, including face and background watermark embedding and detection. The authentication process is presented in §3. Experimental results are given in §4 and we conclude in §5.

2. WATERMARKING PROCESS

In this section, we introduce the stages of the watermarking process, which includes face detection, watermark embedding and watermark detection.



F*n*: the *n*th face
B*n*: the background block assigned to the *n*th face

Figure 1 Face and Background Block Allocation

2.1 Face detection

First, faces are detected automatically. The face detection algorithm developed by [4] is utilized in our watermarking system. The applied face detection algorithm is tolerant to some manipulations including compression and geometrical transformations. At this time, no existing face detection algorithm provides exactly the same relative pose parameters (location, scale and orientation) after these manipulations. Hence, the synchronization of watermark detection can not completely rely on the face detection results. Nevertheless, when we run the face detection system [4] on the watermarked images, the relative pose parameters are very close to those obtained on the original images. The solution to the synchronization problem will be discussed in §2.4.

2.2 Face watermark embedding

The IDs (the face's number) and relative locations of the detected faces are used as face watermark information. The resulting chain of face watermarks is referred to as the "authentication loop" [3], as shown in Figure 1. This loop is modeled as a closed chain, where each node includes the face's unique ID (denoted as N_i) and the relative position of the next face (a unit vector pointing to the next face, P_i) in the chain according to the order given by *x*-*y* coordinate. The parameters of each node in this chain are embedded into a rectangular area defined by the underlying face center and scale in the image. In order to handle the slight inaccuracy of the face detection results on watermarked images, the size of the face rectangle is quantized to different fixed scales:

$$S_{Fi} = \left\lfloor \frac{S_D}{\Delta} \right\rfloor \times \Delta \qquad (1)$$

where S_D is the detected face size, S_{Fi} is the quantized size and Δ is the quantization step.

The information of each node of the authentication loop is firstly encoded using the same method as in [5] by shifting an *m*-sequence generated by a secret key, and then embedded into each corresponding face rectangle in the wavelet domain by the following equations:

$$\vec{m}_i = Encode(\vec{s}_i, N_i, P_i) \qquad (2)$$

$$c^*(x,y) = c(x,y) + \alpha\beta(x,y)m_i(k) \qquad (3)$$

where $\vec{s}_i$ is the original sequence for the face N_i, $\vec{m}_i$ is the encoded sequence. N_i and P_i are the *i*th face's ID and its unit vector pointing to the next face. $c(x,y)$ denotes the coefficient in the middle and high frequency sub-bands of the wavelet transform of the *i*th face rectangle. $c^*(x,y)$ is the watermarked one. $\beta(x,y)$ is the watermark strength at coordinate (*x*,*y*), which is determined by a visual model that takes into account the local brightness, frequency and texture [6]. α is a global control factor.

By taking into account more key management overhead, we can encode the information with different *m*-sequences using different secret keys. For simplicity, we consider the same key for one image. However, for different images, different keys are used in order to handle the substitution attacks, i.e. replacing an original face with one copied from another watermarked image but with the same face watermark information.

A malicious manipulation of the faces will produce a leak in the authentication loop. For instance, as shown in Figure 2, moving a face to another location in the same image or replacing it with another person's face will be revealed by a leak of the ID and the relative position of the next face or by the absence of the watermark. As IDs are assigned to faces according to the *x*-*y* order, deleting or adding a face will break this order and produce a mismatch in the authentication loop.

2.3 Background watermark embedding

We consider this step in order to embed extra watermark information in the background, such as the total number of detected faces in the scene (denoted as $N_{\max}$). Again, $N_{\max}$ is encoded by shifting an *m*-sequence based on the secret key of the image and then embedded into background rectangular blocks determined by the face locations and scales using the same embedding method of face watermark. We apply an allocation procedure, which assigns one or more background blocks to each detected face, as shown in Figure 1. The background blocks are tiled all over the available background. Each background block has the same size as the underlying face and is used to embed the background watermark information. In this way, every background block can be synchronized again in the authentication process by only referring to its corresponding face.

2.4. Watermark detection

As the face detection algorithm may provide slightly different face locations on watermarked images, it might be possible to retrieve wrong watermarks on authentic

faces without synchronization. Therefore, a small local search around the area of the face centers is performed.

Two search strategies can be utilized here: (1) The local search process stops when the response of the watermark detector is higher than a threshold; (2) All the possible locations in a small area are tested and the one with highest response r_i is taken as the synchronization point and it must also be higher than a threshold. The first method is more efficient, but may result in local maxima. In our experiment, the second one is applied:

$$r_{\max}^{i} = max\left(\mu\sqrt{M}/\sigma\right) \quad (4)$$

where $r_{\max}^{i}$ is the highest response of watermark detection for the ith face. μ and σ^2 are the mean and variance of $R(k)$, $R(k) = c^*(k)m_i(k)$. M is the length of m_i.

Then the underlying synchronization parameters are recovered and the face watermarks are extracted. The synchronization and retrieval of the background watermarks is derived systematically afterwards.

3. AUTHENTICATION PROCESS

Once the face and the background watermarks have been retrieved, the information is crosschecked at different levels. With the authentication loop, it will be possible to identify which face has been added, moved, replaced or deleted. All these verifications can be performed regardless of the background watermark. However, if the background watermark is not available, when the face with the highest ID is removed, it will be impossible to check the original number of faces. Furthermore, the background watermark can monitor the move of the whole face chain.

Different levels of authentication are listed below, based on the combination of the retrieved authentication loop and background watermark information:

1. **Face verification**: a face is considered verified only when the watermark is successfully retrieved with the correct secret key. However, the face's relative location will not be verified until the retrieved face ID and vector are successfully crosschecked with the watermark information and the positions of its neighbor nodes in the authentication loop. In addition, the location of the face relative to the background will be "verified" only when the corresponding background watermark is also correctly retrieved from the blocks corresponding to the current face.
2. **Image completely authenticated**: when the retrieved data of the whole authentication loop are consistent with each other and all the background watermark information is correctly retrieved and matches the number of verified faces, the whole image is declared authenticated.
3. **Face chain partially authenticated**: the data of the authentication loop are completely or partially consistent with each other. In this case, all or selected faces and their relative positions can be verified. The number of the verified faces does not match the background watermark information or the background watermark could not be retrieved. In the former case, it means that a part of the face chain has been modified. And in the latter case, it means either the background was altered or the face chain location was moved partially or together.
4. **Image completely unauthenticated**: if no face is verified (see previous definition), the whole image is declared as unauthenticated.

4. EXPERIMENTAL RESULTS

The experimental results demonstrate the feasibility and effectiveness of our approach. The face detection system [4] we used in our experiments achieves a detection rate of 89.61% with 112 false alarms on the standard CMU+MIT test set. These results are comparable to the most representative works such as [7] and [8].

During the embedding process, the quantization step Δ is set to 32. With the applied visual model, high fidelity is achieved for the watermarked image and the embedded watermark is completely transparent. All tested images show a PSNR (Peak Signal-to-Noise Ratio) above 40 dB. In the detection process, the threshold is set to 3.

In order to evaluate our approach, we consider two kinds of manipulations: (a) non-malicious manipulations, such as lossy compression, format conversion together with visual annotations or slight image cropping outside face regions, and (b) malicious manipulations, including: adding, deleting, moving and replacing faces.

As shown in Figure 2, seven kinds of manipulations are made on the watermarked image. The image shows respectively (1) no manipulation, (2) visual annotation, (3) cropping/translation/visual annotation, (4) adding face, (5) moving face, (6) replacing face, (7) deleting face, and (8) moving face chain together with respect to the background. On the manipulated images, the face detection and authentication loop verification results are drawn by dashed line and solid line with arrow respectively. The replaced or added faces are marked out by crossed rectangles. The complete authentication results, with the background watermark involved, are presented in Table 1. The experimental results demonstrate that all the above-mentioned malicious attacks to face images can be detected by the proposed scheme.

We also evaluate our scheme with common non-malicious manipulations and the results show it is robust to common image processing, e.g. JPEG compression, gamma correction, contrast adjustment, adding noise, etc.

5. CONCLUSION AND FUTURE WORK

In this paper we propose a novel and flexible image authentication scheme that provides different levels of authentication of face images. With automatic face detection, the proposed scheme can distinguish the manipulations among faces from those in the background. An authentication loop is applied in order to verify the integrity of faces, so our approach can recognize not only if the image is manipulated but also what kind of manipulations take place, which makes the scheme more flexible in some applications.

One open issue is the possibility to detect slight content-changing manipulations inside the face area. In order to tackle this problem, we plan to apply semi-fragile watermarking in order to protect the facial components.

6. REFERENCES

[1] J. Dittmann, "Content-fragile Watermarking for Image Authentication", Proceedings of SPIE, Security and Watermarking of Multimedia Contents III, Vol. 4314, 2001.

[2] E. T. Lin, C. I. Podilchuk and E. J. Delp, "Detection of Image Alterations Using Semi-Fragile Watermarks", Proc. of SPIE, Security and Watermarking of Multimedia Contents II, vol. 3971, Jan. 2000.

[3] H. Liu, H. Sahbi, L. C. Ferri, M. Steinebach, "Image authentication using automatic detected ROIs", In Proc. of WIAMIS 2004, Lisbon, Portugal, 2004.

[4] H. Sahbi, D. Geman, N. Boujemaa, "Face detection using coarse-to-fine support vector classifiers", In Proc. of the IEEE Inter. Conf. on Image Processing, pp. 925–928, 2002

[5] J. O'Ruanaidh and T. Pun, "Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking", Signal Processing 66, 1998.

[6] N. Kaewkamnerd, K.R. Rao, "Wavelet based image adaptive watermarking scheme", Electronics Letters, vol. 36, no. 2, pp. 312-313, 2000.

[7] H. Rowley, S. Baluja, T. Kanade, "Neural network-based face detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, 1998.

[8] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features", In IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001.

Table 1 Authentication Results

No	Face ID	Face Vector	Background Watermark	Authenticated?	
				Loop	Image
1 | 3	1	0	3	Yes	Yes
	2	98	3		
	3	230	3		
4	1	0	3	Yes	Partially
	2	98	3		
	3	230	3		
	X	X	X		
5	1	0	Partially	Partially	Partially
	2	98	3		
	3	230	3		
6	X	X	X	Partially	Partially
	2	98	3		
	3	230	3		
7	1	0	3	Partially	Partially
	3	230	3		
8	1	0	Partially	Yes	Partially
	2	98	Partially		
	3	230	Partially		

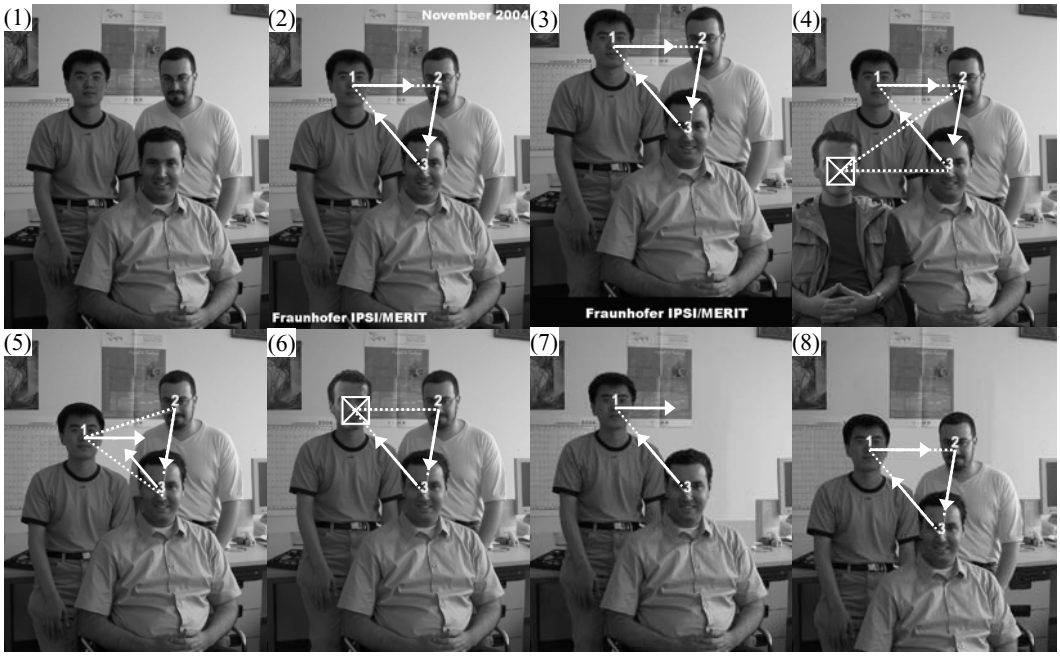


Figure 2 (1) Watermarked image and (2)-(8) Manipulated versions: dashed lines denote the face detection results and solid lines with arrow denote the authentication results