A NEW TECHNIQUE FOR AUTHENTICATION OF IMAGE/VIDEO FOR MULTIMEDIA APPLICATIONS

Chih-Hsuan Tzeng and Wen-Hsiang Tsai

Department of Computer and Information Science, National Chiao Tung University
1001 Ta Hsueh Rd., HsinChu, Taiwan 300, R.O.C.
Tel.+886-3-5720631
{chtzeng, whtsai}@cis.nctu.edu.tw

ABSTRACT

In this study, we propose a new image/video authentication system based on the digital signature technique. The proposed digital signature is composed of visually important and compression-tolerant features in images/videos. In the authentication process, the features of each block being evaluated are compared with the corresponding features of the block recorded in the digital signature. Authenticity is granted if the similarity is high enough. The experimental results show that the proposed techniques are not only efficient to protect security and integrity of multimedia data, but also are flexible and feasible for real applications.

Keywords

Digital signature, image/video authentication, JPEG/MPEG compression, compression tolerance.

1. INTORDUCTION

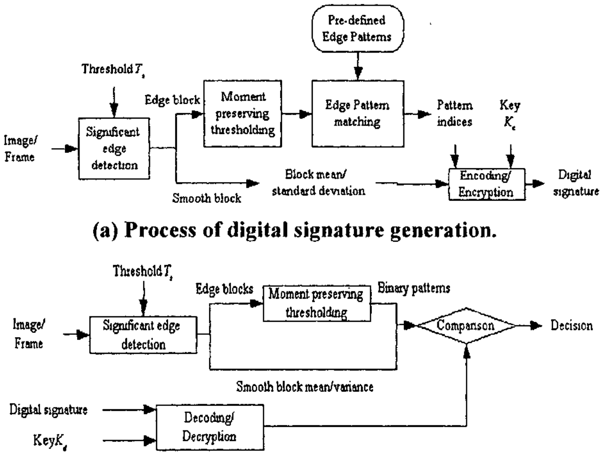
The study of image/video authentication falls into two categories: the data hiding approach [1] and the digital signature approach [2-6]. Our proposed method belongs to the category of the digital signature approach. The digital signature approach is based on the idea of extracting invariant features from image/video data and encoding them to form digital signatures. The extracted invariant features are preserved after acceptable image/video manipulations. To certify a received image/video, two feature sets, with one being extracted from the received image/video and the other from the digital signature, are compared. A confident measure is given to indicate the degree of authenticity. Our proposed method is belonged to the approach.

Schneider and Chang [2] proposed a scheme based on the use of the histograms of image blocks. Though the proposed method is viable, the generated digital signature requires a large amount of storage space. Besides, an image can be manipulated without altering its histogram. Dittmann et al. [3] proposed a method that is based on the use of features extracted by the Canny edge detector [7]. The extracted edge features are matched with pre-defined binary patterns and coded by variable run-length codes. In the authentication process, the generated feature codes of an image to be inspected are directly compared with those of the corresponding digital signature. Because of the use of image compression that results in distorted edges in the reconstructed image, some extracted edge features might be slightly moved or shifted, affecting the correctness of authentication.

Many other invariant features were proposed [4-6]. Most methods reported so far are based on only one invariant feature. In our experience, image/video data are composed of both color and geometric information. As a result, previously proposed methods, though viable, still encounter problems when image/video data are highly compressed. In this paper, we present a new authentication scheme based on the use of a new type of digital signature, which exploits both color and geometric visual features, and prevents explosion of the signature size in the mean time. In the following sections, we describe the idea of the proposed authentication technique in Section 2. In Section 3, we show some experimental results. Finally, we make a brief conclusion in Section 4.

2. SIGNATURE GENERATION AND AUTHENTICATION

The proposed technique is composed of two processes, namely, signature generation and signature authentication, as illustrated in Figure 1.



(a) Process of digital signature generation.

(b) Process of authentication.

Figure 1. Digital signature generation and authentication processes.

Multimedia and Security Ottawa Canada
Copyright ACM 2001 1-58113-393 -6/01/10...$5.00

2.1 Digital signature generation

An image/video is composed by color and geometrical information. By using these two types of information, an image/video can be well represented. Because telltale operations intend to cause a different interpretation from the original image/video by people, these operations result in greater modification in color or edge information than image/video compression. As a result, to design a digital signature-based authentication technique that is robust against image/video compression but sensitive to telltale operations, color and edge information is the best choice of information to be contained in a digital signature.

2.1.1 Block classification

Figure 1(a) illustrates the proposed digital signature generation process. In the beginning, a significant edge detection method was applied to an input image/video. The purpose of significant edge detection is to classify each non-overlapping block into two types: smooth blocks and edge blocks. For simplicity, we use the Sobel operator, which consists of a pair of 3x3 convolution kernels (G_x and G_y), and a pre-defined threshold T_s, to construct a temporary binary image. If the magnitude of the gradient G at a pixel is larger than T_s, then the value of the corresponding location of a temporary binary image is set to "1"(white), indicating a significant edge pixel; otherwise, the value is set to "0"(black). The pre-defined threshold T_s can be computed automatically by some histogram thresholding techniques like [8] or manually. It can be contained in the beginning part of the digital signature or be embedded into the image or video frame as part of the annotation, so that it can be retrieved during the authentication process.

Referring to the temporary binary image, we classify each non-overlapping block b_k with the same size $m \times m$ into two categories: smooth block and edge block. The block size $m \times m$ is set as the basic unit of region for tampering detection. We use 8x8 blocks in the proposed method because this is moderate for noticeable difference by people. For each block b_k in the input image/video, if there is at least one connected component with size larger than 4 in the corresponding binary image I', then b_k is classified as an edge block; otherwise, b_k is classified as a smooth block.

2.1.2 Feature extraction

As stated in the previous section, it is sufficient to represent smooth blocks by their mean values. To be more effective, we also use the standard deviation of pixel values in a smooth block as the second feature. To encode the features into a digital signature, a precedent bit with value "0" was set to indicate the existence of a smooth block, followed by its mean and standard deviation encoded with α and β bits, respectively.

Because edge blocks contain more details and have larger color variances than smooth blocks, it is neither efficient nor sufficient to represent edge blocks using only color information. In fact, the color information is implicitly conveyed in the geometric information in an edge block. For instance, if an edge block contains a strong edge that separates the block into two visually smooth regions, it implies that the color histogram of the block is bimodal and the variance of the histogram is large. Consequently, binary edge patterns are used as features for edge blocks in this study. One reason of using binary edge patterns instead of color ones is to prevent the size of the digital signature from explosion. In addition, it is difficult for a tampering to modify image/video data semantically by only changing their color information without affecting the edge information.

In the proposed method, moment-preserving thresholding [8] is applied to transform each 8x8 edge block b_k into a binary one b_k'. Instead of encoding b_k' directly, each of the four non-overlapping 4x4 sub-blocks of b_k' are matched with a set of 4x4 pre-defined binary edge patterns p_j where $j = 1, 2, \ldots, s$. And for each sub-block of b_k', the index $i \in \{1, 2, \ldots, s\}$ of the best match pattern p_i is used to generate the digital signature. Let b' be a 4x4 binary block, f a matching function, i the output of f, and d a similarity measure function. Then the operations of f and d can be expressed as follows:

$$f(b') = i \qquad (1)$$

and

$$d(b', p_i) = \min_j d(b', p_j) \cdot \qquad (2)$$

The similarity function d measures the similarity between two input patterns, for which we use the Hamming distance in this study. Figure 2 shows an example of 4x4 edge patterns used in our study. To encode features into digital signatures, a precedent bit with value "1" was set to indicate the existence of an edge block and followed by four pattern indices, each being encoded with the same length, ω bits, to represent the indices of best matching patterns.

A reason of using 4x4 sub-blocks for pattern matching rather than 8x8 blocks is that 4x4 edge blocks can be well represented by a small set of 4x4 binary patterns. Because 4x4 blocks are small enough, most of 4x4 blocks in an image/frame contain only one edge. We can design a small set of patterns, each containing only one edge. This simplifies the candidate binary pattern set design process and reduces the size of the matching set; nevertheless, it keeps the codeword used to encode pattern indices relatively short. These facts stated above were confirmed experimentally in [9] where the authors selected 16 binary matching patterns as shown in Figure 2 to represent binary patterns in an input image and showed that the probability for the Hamming distance between the original pattern and its representative to be greater than 4 is smaller than 0.1 in average.

At last, we construct the digital signature by cascading all the encoded features extracted from the blocks in the input image/frame, followed by an encryption key K_e for security purposes, which further encrypts the cascaded bit stream. Because the result of block classification can be taken as a binary image, instead of using precedent bits to indicate whether a block is a smooth block or an edge one, we use binary image compression techniques such as [10] and [11] to further reduce the size of the digital signature.

2.2 Process of authentication

In the proposed authentication method, the features of an image/video are compared directly with the features recorded in the corresponding digital signature to test whether the image/video has been tampered or not. The authentication process of the proposed method is illustrated in Figure 1(b).

After receiving an image/video and its corresponding digital signature, the proposed authentication process decrypts the digital signature using the decryption key K_d. Next, like what is done in the process to generate the digital signature stated in the last section, each block in the input image/frame is classified as a smooth block or an edge block. Due to the possibility of distortion made by acceptable operations or tampering by oblivious attacks, the block classification result B' might not be the same as the block type B recorded in the digital signatures. It is too strict to conclude that a block is tampered if B is found to be different from B'. As a result, the proposed authentication scheme take four conditions into consideration, i.e., four possible combinations of (B, B') block-type pairs.

Condition 1 (B and B' are smooth blocks): In this situation, the difference of the mean m and the standard deviation σ recorded in the digital signature, and the mean m' and the standard deviation σ' of the block b_k in question are compared with pre-defined thresholds T_1 and T_2 to see if the block was tampered. The rule for authentication in this condition is expressed as follows:

$$b_k \text{ is} \begin{cases} \textit{tampered,} & |m - m'| > T_1 \text{ and } |\sigma - \sigma'| > T_2; \\ \textit{authentic,} & \textit{otherwise.} \end{cases} \tag{3}$$

Condition 2 (B and B' are edge blocks): In this situation, moment-preserving thresholding [8] is applied to the block in question and the resulting four binary patterns (p_1', p_2', p_3' and p_4') of the four 4×4 sub-blocks are compared with the four representatives (p_1, p_2, p_3 and p_4) recorded in the digital signature. The rule for authentication is expressed in the following rule:

$$b_k \text{ is} \begin{cases} \textit{tampered,} & \sum_{i=1}^{4} d(p_i, p_i') > T_3; \\ \textit{authentic,} & \textit{otherwise,} \end{cases} \tag{4}$$

where d is the Hamming distance function.

Condition 3 (B is a smooth block and B' is an edge block): This situation may result from acceptable operations that make the gradient stronger in the original blocks to become weak, or from intended operations that modify the geometric property of the blocks. In our experience, if a threshold T_s is properly set to classify blocks, the possibility of occurrence of this condition is very low and the blocks B' in the received image/video data are isolated. The possibility of more than two adjacent blocks B' in this condition is very low in image/video data. As a result, we also take the neighboring blocks $b_{k(j)}$ into consideration. Because the input image/frame is available, the mean m' and the standard deviation σ' of b_k are computed and compared with the mean m and the standard deviation σ recorded in the digital signature, respectively. With pre-defined thresholds T_4 and T_5, the decision rule of this condition is as follows:

$$b_k \text{ is} \begin{cases} \textit{tampered}\ , & |m - m'| > T_4 \text{ and } |\sigma - \sigma'| > T_5 \\ & \text{and for some } j, b_{k(j)} \text{ is tampered };\\ \textit{authentic}\ , & \textit{otherwise}\ . \end{cases} \tag{5}$$

Condition 4 (B is an edge block and B' is a smooth block): This situation may result from compression operations that cause distorted edges, so that the gradients of strong edges in original blocks become too weak to be detected by a global threshold. It also might result from telltale operations that tamper the geometric property. To determine whether the block b_k in question is authentic, moment-preserving thresholding [8] was applied to b_k and the resulting four binary patterns (p_1', p_2', p_3' and p_4') of the four 4×4 sub-blocks are compared with the four representatives (p_1, p_2, p_3 and p_4) recorded in the digital signature. If the sum of the four Hamming distances between the corresponding patterns is smaller than a pre-defined threshold T_5, then block B in question is decided to be authentic; otherwise, tampered. In addition, as stated in *Condition 3*, the occurrence of the block is often isolated, so we also take its neighboring block into consideration. This can be expressed in the following rule:

$$b_k \text{ is} \begin{cases} \textit{tampered}\ , & \sum_{i=1}^{4} d(p_i, p_i') > T_6 \\ & \text{and for some } j, b_{k(j)} \text{ is tampered };\\ \textit{authentic}\ , & \textit{otherwise}\ , \end{cases} \tag{6}$$

where d is the Hamming distance function.

Besides, though smooth blocks might be tampered maliciously by preserving its mean and standard deviation, it is not easy to tamper semantically without causing the change of the edge activities in the region to be tampered. The modification of smooth blocks may result in *Condition 3* as stated above, so we can protect smooth blocks from this type of attack by deciding a region to be tampered if more than 3 connected blocks belonging to *Condition 3* are found.

2.3 Authentication of videos using temporal information

The proposed method can be used to authenticate video sequences. Each frame in the video can be taken as a still image and the digital signature can be generated using the proposed method. Because some operations will cause insertion or drop of frames, it will result in lost of synchronization between the frames and the digital signatures. To deal with this problem, a unique identifier can be embedded as a watermark in the frame to be protected, and record the identifier in the beginning of the corresponding digital signature to link the frame and the signature together. Later in the authentication process, the two identifiers from the frame and the digital signature are compared to check the authenticity of the frame.

In addition, video frames are temporally related and common video editing operations are applied to a sequence of temporally connected frames, so we may take advantage of this property to improve the proposed authentication method. The way we adopt is if a block in the location (x, y) of the i-th frame is checked to be tampered using the proposed authentication process, then those blocks in the same location (x, y) of the previous m preceding frames and their eight neighboring blocks are checked to see if the region is tampered consecutively. In this study, m is taken to be 2. If the number of tampered blocks around the specific location (x, y) in the three consecutive frames is larger than 3, then the block at the specific location (x, y) is decided to be tampered.

3. EXPERIMENTAL RESULTS

We have evaluated the performance of the proposed authentication method using the well-known test images and videos, and a variety of images and videos randomly selected from multimedia databases. For color images and videos, the proposed method can be used to generate the digital signature for the luminance channel (Y) and the chrominance channels (U and

V). In the block classification process, the value of the threshold T_s is set to 150 to obtain moderate classification results for image/video. In the digital signature generation, the mean and the standard deviation of a smooth block are coded using 5 bits and 4 bits. The pattern index of the best representative among the predefined patterns is coded using 6 bits. The probabilities of the occurrence of a smooth block in a general image/video are in the range of 0.2 to 0.5. The average signature length is about 15 to 20 bits for an 8x8 block. In the authentication process, the threshold T_1, T_2, T_3, T_4, T_5, and T_6 are set to 10, 3, 16, 10, 3, and 16, respectively. In fact, different threshold values can be set in accordance with different images. These image-dependent threshold values can be recorded in the beginning of the digital signature or embedded in the image/video as annotations to increase the flexibility of the authentication method. In the experiments, we set these values globally for simplicity.

The authentication results of still images are shown in Figure 3. We simulated the incidental operations by adding a small amount of random noise to the shoulder and blurring the hair moderately. And we simulated the malicious operations by adding a text in the top-left corner. Figure 3(b) shows the image after a sequence of operations stated above including the JPEG compress at quality factor 40. The authentication result is given in Fig. 3(c), where only the text region is considered as maliciously tampered. This output corresponds to our desired result before authentication. The video authentication result is also shown in Figure 4. The test video sequence was compressed using MPEG at 1.2Mbps and tampered locally by removing the small boat. The authentication results in Fig. 4(c) shows that the tampered region can be well located and the proposed method is also tolerant against the MPEG compression. Moreover, because the thresholds of the proposed method are related to the measurement of distortion, different values can be set to tighten or to relax the authentic degree. This gives the flexibility to alter the authentic policy.

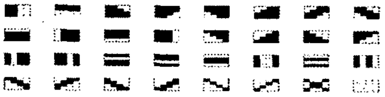
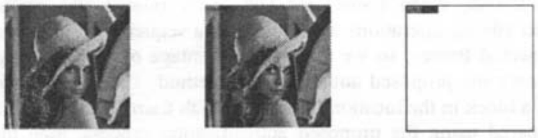
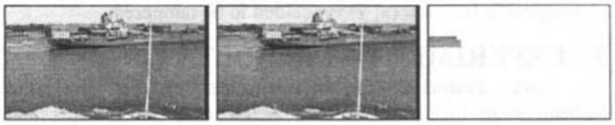


Figure 2: The predefined 64 edge patterns (the reversed are not shown).



(a) (b) (c)

Figure 3: Image authentication result of the proposed method: (a) Original image; (b) Tampered and JPEG compressed; (c) Result.



(a) (b) (c)

Figure 4: Video authentication result: (a) the 45th frame of the original video; (b) the MPEG compressed at 1.2Mbps and tampered; (c) the authentication result.

4. CONCLUSIONS

In this study, we have proposed a new technique for image/video authentication. Experiment results prove that the proposed method can detect malicious tampered regions and can tolerate JPEG/MPEG compression. The proposed method is also simple and computational efficient. Because the proposed authentication technique is digital signature-based, it inherits the merits of using the digital signature such as asymmetric authentication which is more flexible than the existing schemes that require the original embedding key for authentication. In addition, the proposed authentication technique provides better flexibility than other authentication techniques that the permitted degree of image/video manipulations is bounded or fixed in advance.

5. References

[1] E. T. Lin and E. J. Delp, "A Review of Fragile Image Watermarks," Multimedia and Security Workshop in ACM Multimedia '99, Orlando, Florida, USA, 1999.

[2] M. Schneider and S. F. Chang, "A robust content based digital signature for image authentication," 1996 International Conf. on Image Processing, Lausanne, Switzerland, 1996.

[3] J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," IEEE International Conf. on Multimedia Computing and Systems, Italy, 1999.

[4] C. S. Lu and H. Y. M. Liao, "Structural digital signature for image authentication," ACM International Conf. on Multimedia, Los Angeles, CA, USA, 2000.

[5] D. C. Lou and J. L. Liu, "Fault resilient and compression tolerant digital signature for image authentication," *IEEE Trans. Consumer Electronics*, vol. 46, pp. 31-39, 2000.

[6] C. Y. Lin and S. F. Chang, "Generating robust digital singnature for image/video authentication," ACM Multimedia Workshop, Bristol, U.K., 1998.

[7] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.

[8] W. H. Tsai, "Moment-preserving thresholding: a new approach," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 377-393, 1985.

[9] C. K. Yang and W. H. Tsai, "Improving block truncation coding by line and edge information and adaptive bit plane selection for gray-scale image compression," *Pattern Recognition Letters*, vol. 16, pp. 67-75, 1995.

[10] C. H. Tzeng and W. H. Tsai, "Compression of Chinese Character Patterns in Document Images Based on Rectangular Region Partitioning Using Contour Information and Huffman Coding," *International Journal of Computer Processing of Oriental Languages*, vol. 13, pp. 177-202, 2000.

[11] W. Kuo, *Digital Image Compression Algorithms and Standards*, vol., ed. Norwell, Massachusetts: Kluwer Academic Publishers, 1995.