A NEW VIDEO AUTHENTICATION TEMPLATE BASED ON BUBBLE RANDOM SAMPLING

Fabrizio Guerrini, Riccardo Leonardi, Pierangelo Migliorati

SCL - DEA University of Brescia, Via Branze, 38, 25123, Brescia, Italy
e-mail: {fabrizio.guerrini, riccardo.leonardi, pierangelo.migliorati}@ing.unibs.it

ABSTRACT

The rapid growth of digital video distribution has highlighted new important issues in digital rights management, as well as in other important applications such as video authentication. Digital watermarking offers a promising solution against piracy and it is therefore a very active area of research. However, robustness to video manipulations, either malicious or not, is a demanding task because there are many different types of possible attacks that can be envisioned. Among these, geometric and temporal distortions play the major roles. The countermeasures against these specific attacks are still an open challenge. In this paper we propose the use of a video authentication template based on bubble random sampling. The authentication template is introduced in order to ensure temporal synchronization and to prevent content tampering. The simulation results are encouraging and this approach is therefore worth further development efforts.

1. INTRODUCTION

The recent innovations in the field of digital technology, in particular with the introduction of the Digital Versatile Disc (DVD), and the widespread diffusion of the Internet network, have made possible the large scale diffusion of digital video.

However, although digital video offers high performance to the final user, content owners have been concerned with the problem of digital piracy caused by the simplicity with which digital data can be copied and distributed. Encryption offers only a first stage solution because, once decrypted, the data are not protected anymore.

Video watermarking has been proposed as a complementary solution. Video watermarking consists of the embedding of some hidden copyright information, called watermark, inside the original data.

The principal conflicting requirements for watermarking are capacity, which is the quantity of information associated to the watermark, imperceptibility, that requires the watermark to produce no visible artifacts in the original video, and robustness, that is the ability of the watermark to resist to attacks, intended as any possible video manipulation, malevolent or not. Real-time feature may be an addictional requirement.

Video watermarking was initially employed to enforce copyright protection and to provide an effective copy control mechanism. It has also been proposed in other contexts. In fingerprinting, a watermarking system can be used to trace traitors transmitting copyrighted data across Internet via a peer-to-peer system. In active broadcast monitoring, digital watermarking could be used to embed some data to ensure the payment of royalties. It could also improve video coding techniques by replacing the channel coding step with the insertion of a watermark consisting of the error correcting codes.

Image watermarking has been developed first, but nowadays many other fields have been explored, including text, audio, and video. The robustness of image watermarking is often studied applying transcoding between different formats, by adding some noise, or performing a geometric attack, such as, e.g., StirMark [1], that tries to break the consistence between the embedder and the detector coordinates. Virtually all types of image processing techniques have been used in still image watermarking, each one with a different grade of robustness against the possible attacks, and its own pros and cons.

In case of digital video watermarking, new attacks may be envisioned. For example, we mention temporal desynchronizations, statistical attacks, including collusion and average attacks [2] that aim to estimate the hidden data using frames or watermark correlation properties, and video editing, either in the pixel domain or in the temporal domain. All types of denoising [3] should also be considered as statistical attacks.

Early video watermarking techniques rely on still image watermarking results and inherit all their nice properties. However, video frames are not usual still images because they possess peculiar temporal correlation caused by motion, and simple extension of still image watermarking schemes to video is not advisable [4][5].

Consequently, other approaches have been explored, including the combined use of DCT and drift compensation to reduce visible artefacts [6], temporal wavelet decomposition [7], 3-D DFT [8] and local activity measures [9].

To prevent geometric and/or temporal distorsions, video watermarking schemes might include proper templates, as proposed for example in [9], to invert these operations and to preserve the detector ability to recover the watermark.

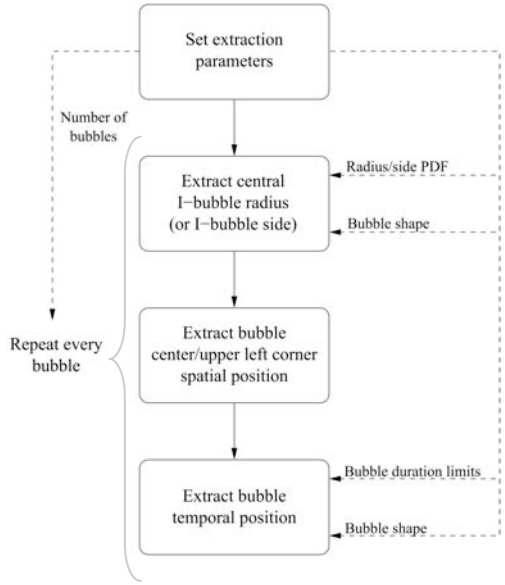

Figure 1. The process of bubble random sampling.

In this paper we propose the use of a video authentication template based on bubble random sampling to recover the exact frame position within the original video. In this way it is possible to reverse eventually occurred temporal desynchronizations.
The method is based on bubble random sampling, described in Section 2. The proposed authentication template is presented in Section 3, whereas experimental results are reported in Section 4. Conclusions are drawn in the final section.

2. BUBBLE RANDOM SAMPLING

Bubble random sampling of a still image consists of a random selection of circular (or square) areas of the considered image. This approach has been originally proposed for content-based retrieval, in which case these areas were used to generate a fingerprint of the considered image [12].
In this paper, we use video bubbles that include also the temporal dimension. Every 3-D bubble consists of a sequence of 2-D bubbles, which are called intra-frame bubbles, or I-bubbles, of identical spatial position and belonging to adjacent frames. The size of the I-bubbles could be set or could be varied along the temporal dimension. The process of bubble random sampling is illustrated in Fig. 1. First of all, the radius and the spatial position of each bubble is randomly extracted, accordingly to a probability distribution function set a priori. Then, the temporal position of the bubble is determined by extracting, from a uniform probability distribution function, the position in time of the first frame of the bubble, and the number of consecutive associated frames.
In general, the temporal length of the bubble should be enough to guarantee a significant histogram boundaries evaluation, as explained in the following section.
Some of the parameters of this process can be modified by the user, namely: the number of bubbles to be considered, the probability distribution function of the radius/side of the bubbles, and the limits of the bubble temporal length.
Figure 2 shows an example of an ellipsoidal video bubble. Frames are depicted as rectangles; the lighter areas represent the I-bubbles.

3. THE PROPOSED AUTHENTICATION TEMPLATE

In this section we describe the application of bubble random sampling to the extraction of an effective video authentication template.
Specifically, the video is first decoded, then the bubble random sampling is carried out. Finally a list of extracted features, which is called the video hash, is stored. The decoding process is necessary because the proposed technique is based on uncompressed video.
Two different versions of the authentication template have been considered, both using some features based on the luminance histogram. Figure 3 depicts the processing steps involved in the video hash generation. The process of random bubble extraction needs the introduction of an additional user defined parameter. It is in fact necessary to specify how many bubbles are forced to include the first and the last frame, that have for statistical reason less probability of being extracted.

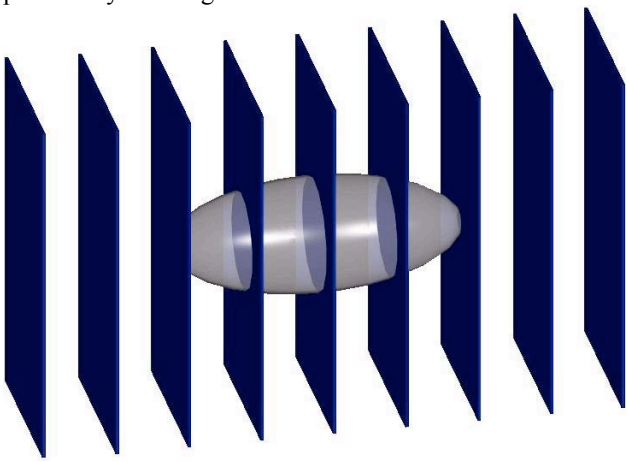

Figure 2. Ellipsoidal video bubble example.

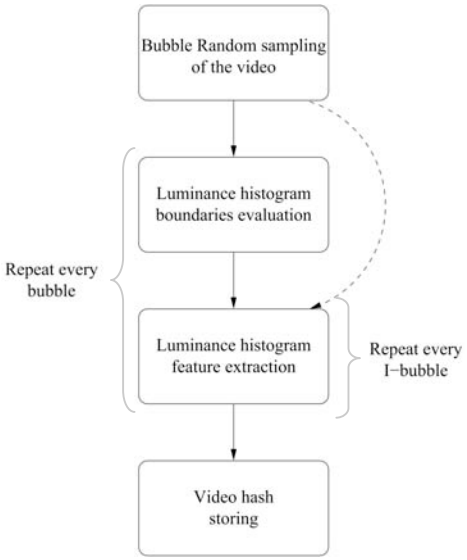

Figure 3. Procedure for video hash generation.

The luminance histogram is the feature selected for video frame authentication.
The storing and matching of the complete histogram is too cumbersome, and therefore an histogram quantization is necessary.
This operation has been already evaluated in the framework of content-based image retrieval [10]. In that application, the problem is to find the perceptually similar images, and so the quantization is aimed to match at best the human perception [11].
In our case, the template uses two very simple histogram quantizations laws, in order to keep a reduced overall complexity.
In the first case, it consists in a uniform quantization of the 256 luminance levels into 16 bins; histogram matching is performed via standard L_1 metric evaluation, as follows:

$$D' = \sum_{i=1}^{16} (H_1[i] - H_2[i])$$

In the second case, the histogram quantization is again a uniform quantization with 16 bins, but it considers only the luminance level portion (in multiplies of 16) of the histogram in which it is present at least the 90% of the total luminance values.
The new histogram distance *D''* is computed as follows:

$$D'' = \sum_{i=1}^{16} (H_1[i] - H_2[i]) \cdot (1 + \alpha_i)$$

where the factor α_i is the semisum of adjacent bins histogram difference, except for the first and the last bins, for which it evaluates the successive and precedent bin difference respectively.
The inclusion of this factor accounts for occasional spilling of luminance values into adjacent bins due to noise adding.
The histogram distance *D''* is essentially a L_1 metric in which every bin contributes to the sum in dependance of a factor that weights the adjacent bins differences.
In Figure 3 the dashed line represents the first type template that doesn't use any luminance histogram boundaries evaluation step.
It should be noted that the occurence of value-metric distorsions, e.g. increasing the image luminosity, are likely to drastically modify the luminance histograms. If robustness to such attacks has to be granted, other approaches, already presented in the literature, need to be explored.
In the synchronization phase, the proposed algorithm reads the original video hash, that could be sent along with the video in an encrypted form, and whose absence should prevent correct video reproduction, and compares it with that obtained by the processing of the received video.
Luminance histograms are then computed in spatial coordinates; for every I-bubble, the frames that have scored the minimum distance values are identified.
To decide the correct frame localization, a common output on all the I-bubbles pertaining to the same frame have to be detected; an exception is made for zero distance frames, that are likely to occur between contiguous frames.
A text output file lists the frames recovered correctly, that is to say correctly authenticated.

4. EXPERIMENTAL RESULTS

The performance of the proposed authentication template were evaluated considering the sequences "Mobile and Calendar" and "Flowers garden". Both of them last nearly 10 seconds, and were MPEG-2 encoded with a frame rate of 40 frames per second, with a 4:2:0 pattern.
The sequences were decoded, and the frames stored in YUV format. Several video hash generations were performed, with different overall number of bubbles and temporal bubble length. As expected, the template could easily recognize temporal video editing operations as frame dropping or frame shuffling.
Then the frames were manipulated by adding a moderate Gaussian noise to simulate transmission errors. The first template, that uses uniform histogram quantization and standard L_1 metric, fails to recover the correct frame synchronization on about 15% of frames; this is mainly caused by the coarse quantization, especially in very dense histograms.
The second, more sophisticated template decreases this error rate to about 6%.

These values have been obtained considering 20 experiments. As a final test, a central frame was edited by inserting a black zone covering the 10% of the total frame extension. If the number of bubbles were sufficient (more than 20 per frame), these manipulated frames were recognized as not original, 19 times out of 20. For the system to perform correctly it is in fact necessary that the zone tampered with to be covered by at least one bubble. Increasing the number of bubbles helps to decrease false positive occurrence, but increases also the computational cost. These results are described in Fig. 4.

TEMPLATE NO. / ATTACKS	1 (*D*')	2 (*D*'')
Single or Multiple Frame Dropping	1.00	1.00
Frame Shuffling	1.00	1.00
Gaussian Noise Addiction	0.85	0.94
Single Frame Cropping	0.95	0.95

Figure 4. Experimental successful temporal synchronization percentages (out of 20 tries).

5. CONCLUSIONS

In this paper we have presented a bubble random sampling approach applied to the construction of a video authentication template used for synchronization and content verification in the context of video watermarking. More specifically, a suitable feature extraction is performed in randomly selected areas of the video, that are the union of adjacent frames square areas. The positions, side and temporal localization of these bubbles result from a random extraction, with user-defined parameters. The reliability and the effectiveness of this scheme have been proven through simulation on real data. Further efforts are to be addressed in this application to improve the feature extraction process, for instance including a time-stamping to helps the synchronization process, and a geometric template to make the system more robust with respect to geometric intra-frame distorsions. Moreover, the template itself should be implemented in a complete video watermarking system.

6. REFERENCES

[1] F. Petitcolas, R. Anderson, M. Kuhn, "Attacks on Copyright Marking Systems", Proc. of the Second International Workshop on Information Hiding, Vol. 1525, Springer, Berlin, pp. 218-238, 1999.

[2] G. Doerr, J.-L. Dugelay, "A Guide Tour of Video Watermarking", IEEE Signal Processing: Image Communication, Vol. 18, No. 4, pp. 263-282, 2003.

[3] F. Deguillaime, G. Csurka, T. Pun, "Countermeasures for Unintentional and Intentional Video Watermarking Attacks", Proc. of the IS&T/SPIE Electronic Imaging 2000: Security and Watermarking of Multimedia Content II,Vol. 3971, pp. 346-357, 2000.

[4] J. Dittman, M. Stabenau, R. Steinmetz, "Robust MPEG Video Watermarking Technologies", Proc. ACM Multimedia '98, pp. 71-80, 1998.

[5] J. C. Oostven, A. A. C. Kalker, J. A. Haitsma, "Visual Hashing of Digital Video: Applications and Techniques", Proc. of the SPIE Applications of Digital Image Processing XXIV, Vol. 4472, pp. 121-131, 2001.

[6] F. Hartung, J. K. Su, B. Girod, "Spread-Spectrum Watermarking: Malicious Attacks and Counterattacks", Proc. of the IS&T/SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 147-158, 1999.

[7] M. Swanson, B. Zhu, A. Tewfik, "Multiresolution Scene-Based Video Watermarking Using Perceptual Models", IEEE JSAC, Vol. 16, No. 4, pp. 540-550, 1998.

[8] F. Deguillaime et al., "Robust 3D DFT Video Watermarking", Proc. of the IS&T/SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents, pp. 113-124, 1999.

[9] T. Kalker, G. Depovere, J. Haitsma, M. Maes, "A Video Watermarking System for Broadcast Monitoring", Proc. of the IS&T/SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 103-112, 1999.

[10] L. Cinque, S. Levialdi, A. Pellicano, "Color-Based Image Retrieval Using Spatial-Chromatic Histograms", IEEE Multimedia Systems, Vol. 2 , pp. 969-973, 1999.

[11] R. B. Wolfgang, C. I. Podilchuk, E. J. Delp, "Perceptual Watermarks for Digital Images and Video", Proc. IEEE, Special Issue on Identification and Protection of Multimedia Information, Vol. 87, No. 7, pp. 1108-1126, 1998.

[12] F. Guerrini, R. Leonardi, P. Migliorati, "Image Retrieval with Random Bubbles", Proc. of the European Signal Processing Conference 2004, pp. 1035-1038, 2004.