Video Watermarking Based on Interactive Detection of Feature Regions

Asma Kerbiche, Saoussen Ben Jabra, Ezzedine Zagrouba, Vincent Charvillat

Video Watermarking Based on Interactive Detection of Feature Regions

Asma Kerbiche, Saoussen Ben Jabra, Ezzeddine Zagrouba
Lab. RIADI - Team of Research SIIVA
Higher Institute of Computer Science
University Tunis El Manar
Ariana, Tunisia
asma.kerbiche@gmail.com saoussen.bj@laposte.net
ezzeddine.zagrouba@fsm.rnu.tn

Vincent Charvillat
Lab. IRIT - Team of Research VORTEX
ENSEEIHT - INP TOULOUSE
University of Toulouse
Toulouse, France
vincent.charvillat@enseeiht.fr

***Abstract*—Video watermarking is very important in many areas of activity and especially in multimedia applications. Therefore, security of video stream has recently become a major concern and has attracted more and more attention in both the research and industrial domains. In this perspective, several video watermarking approaches are proposed but, based on our knowledge, there is no method which verified the compromise between invisibility and robustness against all usual attacks. In our previous work, we proposed a new video watermarking approach based on feature region generated from mosaic frame and multi-frequential embedding. This approach allowed obtaining a good invisibility and robustness against the maximum of usual attacks. In our future work, we propose to optimize the choice of the region of interest by using crowdsourcing technique. This last one is an emerging field of knowledge management that involves analyzing the behavior of users when they view a video to automatically deduct the regions of interest.**

Keywords-watermarking; crowdsourcing; robustness; invisibility.

I. INTRODUCTION

The fast development of the Internet in recent years has eased the process of coping, transmitting, and distributing digital data such as image, 3D meshes and video. In fact, the recent decade has seen the emergence of video-based application technologies, such as wireless video, video conferencing, and videophones. As with other types of multimedia data, it was necessary to find new techniques for visualization, compression, indexing, and also the security of this type of media to enable the protection of rights, authentication and data integrity during transfer on a given communication channel. Therefore, the need for protection from piracy and illegal use rises more and more. The best technique to protect digital video from this manipulation is watermarking. It consists of embedding a signature into data and to try to detect it after any manipulation done on marked data. Usually, signature must be imperceptible and should resist to malicious attacks which try to destroy or to remove it. Several studies have been made to develop robust and invisible video watermarking methods [18]. These methods differ mainly in the insertion area [27] (spatial or frequency) and the type of the treated stream [26] (compressed or uncompressed). To improve the invisibility and the robustness of the watermark, some methods are based on the choice of region of interest where the embedding will be done. This choice can be based on many tests applied on different feature regions [17] or using a specific technique [13]. In our work, we propose to use crowdsourcing technique that involves analyzing the behavior of users when they view a video to automatically deduct the regions of interest. The remainder of this paper is organized as follows: in Section 2, video watermarking and crowdsourcing technique are presented while Section 3 describes the proposed watermarking method based on interactive detection of regions of interest using crowdsourcing technique. Finally, conclusion is drawn in Section 4.

II. RELATED WORK

A. Watermarking

Several techniques of watermarking video have been proposed in the literature. We chose to classify them according to two criteria. The first one is the original video format which can be compressed or uncompressed. In fact, watermarking can be applied to compressed stream where the insertion is done during the compression process or after compression [1],[2],[3]. For the uncompressed video, four classes of embedding methods can be applied (some of them can also be used on compressed video). The first class is derived from frame by frame watermarking that consider video as a succession of images and consists of applying still images watermarking algorithms [4],[5]. The second one is the spatio-temporal schemes where video is defined as a 3D signal considering the temporal dimension in video sequences. These schemes decompose the video by performing spatial 2D transform on individual frames followed by 1D transform in the temporal domain [6],[7],[8]. The third class is the temporal schemes that insert the signature in the temporal domain by modifying only the low spatial frequencies [9],[10],[11]. Finally, the last class is based on mosaic frame generated from the original video [12],[13]. This last one selects an interesting area where the mark should be embedded. In fact, mosaïcing allows the insertion of the same mark into the same pixels which represent the same physical point.

TABLE I. ADVANTAGES AND INCONVENIENCES OF EACH CLASS ACCORDING TO VIDEO FORMAT

Methods	Advantages	Inconveniences	Invisibility
Frame by frame watermarking	Minimum insertion time and Avoids compression and decompression steps which can degrade the images of the sequence.	May be ineffective against a MPEG-4 compression that will destroy the inserted signature. Possibility of artifacts	+
Spatio-temporal schemes	Robustness against temporal changes	Unblinded detection and low insertion capacity	++
Temporal Schemes	The mark value is constant for a given image, but vary from one image to another.	Fragile watermark against temporal changes.	+
Compressed Video	If the encoding is complex detection is not.	Fragile watermark against motion estimation and may disappear at re-coding.	++
Mosaic Schemes	Robust against MPEG-4 compression, collusion attacks and deleting images	Not robust against MPEG-2 compression.	++

The second criterion is embedding domain where signature can be inserted directly on video by modifying its pixels [14] or by modifying some video transformations like DCT (Discrete Cosine Transform**)**, DWT (Discrete Wavelet Transform) and SVD transform (Singular Value Decomposition) [15],[16]. The scheme disperses the watermark in the spatial domain of the video frame, hence making it very difficult to remove the embedded watermark.

TABLE II. COMPARISON OF EMBEDDING DOMAIN CLASSES

	Methods	Advantages	Invisibility
Frequency Domain	DCT	The most robust against compression	++
	DWT	The most robust against compression	++
	SVD	Robust against compression, rotation, noise and deleting frames	++
Spatial Domain	LSB	Not robuste against collusion, compression, noise...	+
	Correlation-Based techniques		

In our previous work, after a comparative study of these classes [17], we proposed a new watermarking schema based on mosaic and multi-frequentiel embedding algorithm using wavelet, DCT and SVD transforms. In fact, it consists to generate mosaic from original video. Then, the region where the objects move is selected to be marked. Finally, signature is inserted using wavelet, DCT and SVD transforms. This choice allows obtaining robustness against various types of attacks, such as geometric transformations (rotation, zooming), cropping vertices, MPEG4 part X (h.264 Advanced Video Coding (AVC)) compression, noise, frame suppression and collusion.

B. Crowdsourcing

With the sweeping progress of Web 2.0 technologies and capabilities, many socio-technical systems have attracted attention from both practitioners and scholars. Crowdsourcing is a new emerging Web 2.0 based phenomenon and becomes a recognized sourcing mechanism for problem solving in organizations and societies by outsourcing problems to an undefined entity or the 'crowd'. For that, crowdsourcing research has become a dynamic and vibrant research area, and has been steadily growing over the years.

The term crowdsourcing was first coined by Howe, in a Wired Magazine article in June 2006 [19]: "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers".

In essence, crowdsourcing is based on a simple, but powerful, concept: virtually everyone has a potential to plug in valuable information [20]. It seeks to mobilize competence and expertise, which are distributed among the crowd and has different forms [21].

Crowdsourcing is not exclusive for business purposes. In fact, many non-profit organizations have adopted it as an effective model for problem-solving [22],[23]. In addition to having gained great attention and interest from the industry, crowdsourcing has also gained attention from the academic community.

Indeed, several recent studies are based on crowdsourcing technique. Xie et al. [24] propose new method to detect user interest maps and extract user attention objects from the image browsing log using crowdsourcing where ten subjects were selected to take part in this study, and the criterion for their selection was that they should be very familiar with the use of computers and cell phones before the study. A smart image viewer was then developed based on user interest analysis and a second experiment was carried out to study how users behave with such a viewer. This approach is more efficient than image-analysis based methods and can better represent users' actual interest. Based on the fact that the viewing experience on the mobile devices can be improved by determining important and interesting

regions within the video (regions of interest, or ROIs) and displaying only the ROIs to the viewer, Carlier et al [25] propose an alternative paradigm to infer ROIs from a video by crowdsourcing from a large number of users through their implicit viewing behavior using a zoom and pan interface, and infer the ROIs from their collective wisdom. A retargeted video, consisting of relevant shots determined from historical users' behavior, can be automatically generated and replayed to subsequent users who would prefer a less interactive viewing experience. A user study with 48 participants shows that this automatically retargeted video is of comparable quality to one handcrafted by an expert user.

III. PROPOSED METHOD

The study of watermarking and crowdsourcing recent works shows that these two areas can be combined to propose a new watermarking approach which presents a high level of robustness against the most important attacks. In fact, our work aims to develop new robust approaches to introduce signatures in videos. Our idea consists in a first step to understand the visual content of the original video and then to select feature regions to embed signature. To achieve our goal, crowdsourcing technique will be used. Although, the concept of crowdsourcing is based on sharing media to the public, this can cause confidentiality problems and can damage watermarking process. To avoid this problem, we thought to generate a video summary and to share it to a fixed number of selected users.

The proposed approach is decomposed to three main stages: video summarization, interactive detection of feature regions using crowdsourcing technique and signature insertion. General architecture of the proposed approach is presented by Figure 2.

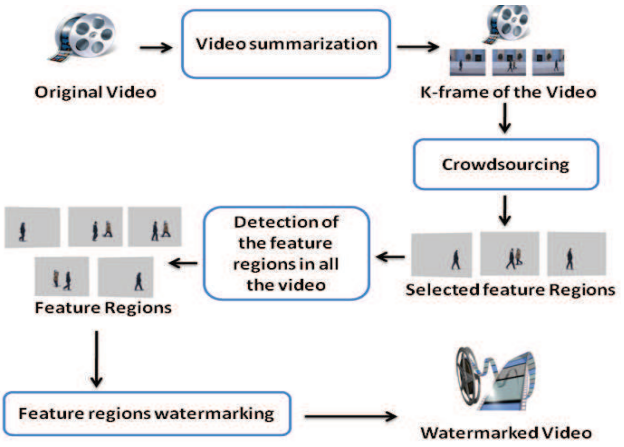

Figure 1. General architecture our proposed method.

A. 1st Step: Video Summarization

To avoid confidentiality problems that can be caused by sharing the original video, its summary will be generated using the approach proposed in [29]. This last one is an object-based technique which allows extracting a relatively small number of still key-frames in order to summarize the salient visual content of a video. This method is based on spatial segmentation of each frame in order to detect the important events. Indeed, the extraction of key-frames will be facing a much more semantic criterion so that each extracted key-frame present an important event such as the appearance and/or the disappearance of significant objects.

B. 2nd Step: Feature Regions Detection

After video summarization, feature regions must be selected to embed signature. To detect these regions, crowdsourcing technique is chosen. In fact, this technique is an emerging field of knowledge management which allows analyzing the behavior of users when they watch a video to automatically deduct the regions of interest. Indeed, the combination of the analysis of visual content and interactive use of the identified data can improve the detection of interest visual objects in a video. The method proposed in [26] and explained in the previous section, has been used for this step. To ensure the confidentiality of our video and to avoid video hacking, we choose to interact with an X number of selected users in our laboratory.

C. 3rd Step: Watermarking

After extracting user attention objects, the signature will be inserted in the N feature regions selected from second step. To embed signature, our multi-frequential watermarking scheme [17] based on DWT, DCT and SVD transforms will be used. In fact, the regions selected by users are the most important regions in the video and their destruction will cause the destruction of the whole video. Thereby, the insertion in these selected regions will ensure a greatest robustness of our watermark. General architecture of the embedding step is presented by Figure 2.

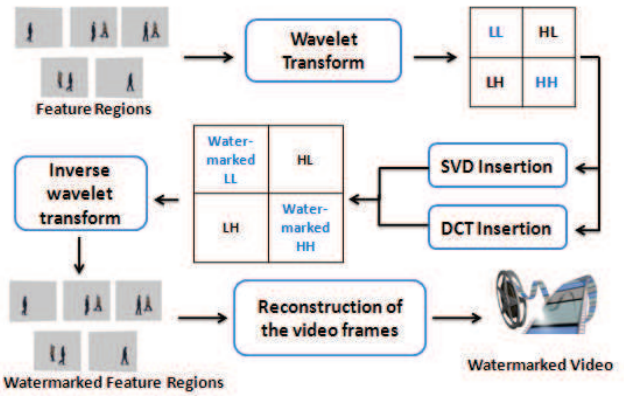

Figure 2. General architecture of the embedding step.

D. Test

The reliability of the developed method will be tested by evaluating the two main criterions: robustness and invisibility. In fact, for the robustness we will test the resistance of the watermark against compression phases, collusion, Cropping and several main types of video attacks that could destroy the watermark. For invisibility criterion

and in order to enhance our tests, crowdsourcing process will be used. A user study will be carried.

IV. Conclusion And Futur Work

Crowdsourcing has become a dynamic and vibrant research area, and has been steadily growing over the years. In fact, it can be applied to a wide variety of problems and ever more classes of applications. For this reason, we proposed in this paper a new video watermarking based on crowdsourcing technique to select the most interesting feature regions. These last ones will be used to insert signature. In fact, interactive detection of feature regions using Crowdsourcing technique will guaranty a high level of robustness and invisibility of marked video. The robustness of the proposed method will be verified after application of various types of attacks such as geometric transformations (rotation, zooming), cropping vertices, MPEG4 compression, noise, frame suppression and collusion. For invisibility, we will use three metrics (peak signal to noise ratio (PSNR), Haussdorff distances and correlation) and a user study will be done to measure marked video quality.

This method can be extended to 3D video and can be used by all watermarking applications such as video indexing.

References

[1] N. Mohaghegh and O. Fatemi, "H.264 copyright protection with motion vector watermarking," in Proc. Int'l Conf. on Audio, Language and Image Processing, July 2008, pp. 1384-1389.

[2] P. Bas, J. M. Chassery, and B. Macq, "Image watermarking; An evolution to content based approaches," Pattern Recognition, Special Issue on Image/Video Communication, vol 35, march 2002, pp 545-561.

[3] H. Joumaa and F. Davoine, "Performance of an application video watermarking scheme using informed techniques," in IEEE International Conference on Image Processing, vol 1, Sept. 2005, pp. 261-365.

[4] J. J. Chae and B. S. Manjunath, "Data hiding in video," in 6th IEEE International Conference on Image Processing (ICIP'99), Kobe, Japan, vol 1, Oct. 1999, pp. 311-315.

[5] S. N. Merchant, A. Harchandani, S. Dua, H. Donde, and I. Sunesara, "Watermarking of video data using integer-to-integer discrete wavelet transform," in Proc. IEEE TENCON Conference on Convergent Technologies for the Asia-Pacific Region, vol 3, Oct 2003, pp. 939-943.

[6] F. Deguillaume, G. Csurka, J. J. O'Ruanaidh, and T. Pun, "Robust 3D DFT video watermarking," in Proc. Security and Watermarking of Multimedia Contents, SPIE, vol. 3657, no. 1, 1999, pp. 113-124.

[7] J. H. Lim, D. J. Kim, H. T. Kim, and C. S. Won, "Digital video watermarking using 3D-DCT and intracubic correlation," in Proc. SPIE Security and Water- marking of Multimedia Contents III, vol. 4314, no. 1, 2001, pp. 64-72.

[8] S. J. Kim et al., "A new digital video watermarking using the dual watermark images and 3D DWT," in Proc. IEEE Region 10 TENCON, vol. 1, 2004, pp. 291-294.

[9] P. Vinod and P. K. Bora, "Motion-compensated inter-frame collusion attack on video watermarking and a countermeasure," in IEEE Proceedings on Information Security, vol. 153, no. 2, June 2006, pp. 61-73.

[10] P. Vinod, G. Doerr, and P. K. Bora, "Assessing motion-coherency in video watermarking," in Proc. ACM Multimedia and Security, 2006, pp. 114-119.

[11] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion resistant digital video watermarking," in IEEE Trans. Multimedia, vol. 7, no. 1, Feb 2005, pp. 43-51.

[12] G. Doerr and J. Dugelay, "Secure background watermarking based on video mosaicing," in Proc. SPIE 5306 Electronic Imaging, 2004, pp. 304-314.

[13] M. Koubaa, M. Elarbi, C. Ben Amar, and H. Nicolas, "Collusion, MPEG4 compression and frame dropping resistant video watermarking," in International Journal of multimedia tools and applications MTAF, Springer Netherlands, Vol. 56, 2012, pp. 281-301.

[14] H. Seddik, M. Sayadi, F. Fnaiech, and M. cheriet, "A New Spatial Watermarking Method, based on a Logarithmic transformation of An Encrypted embedded Mark", in 17th IMACS World Congress on Scientific Computation, Applied Mathematics and Simulation, Paris, France, July 2005.

[15] L. S. Liu, R. H. Li, and Q. Gao, "A robust video watermarking scheme based on DCT," in Machine Learning and Cybernetics, Proceedings of 2005 International Conference, Vol. 8, Aug. 2005, pp. 5176-5180.

[16] P. P. W. Chan, and M. R. Lyu. "A DWT-based Digital Video Watermarking Scheme with Error Correcting Code," in ICICS, Vol. 2836 Springer, 2003, pp. 202-213.

[17] A. Kerbiche, S. B. Jabra, and E. Zagrouba, "A robust video watermarking based on image mosacing and multi-frequential embedding," in IEEE International Conference on Intelligent Computer Communication and Processing ICCP, Aug. 2012, pp 159-166.

[18] R. T. Paul, "Review of Robust Video Watermarking Techniques," in IJCA Special Issue on Computational Science New Dimensions and Perspectives, Vol. 3, 2011, pp. 90-95.

[19] J. Howe, "The rise of crowdsourcing," in Wired Magazine 14(6), June 2006, pp. 1-4.

[20] S. Greengard, "Following the Crowd," in Communications of the ACM 54(2), Feb. 2011, pp 20–22.

[21] Y. Zhao and Q. Zhu, "Evaluation on crowdsourcing research: Current status and future direction." in Information Systems Frontiers, 2012, pp 1-18.

[22] D. C. Brabham, "Crowdsourcing as a model for problem solving: an introduction and cases," in The International Journal of Research into New Media Technologies 14(1), 2008, pp 75-90.

[23] D. C. Brabham, "Moving the crowd at threadless: motivations for participation in a crowdsourcing application," in Information, Communication and Society 13(8), 2010, pp 1122–1145.

[24] X. Xie , H. Liu , S. Goumaz , and W. Y. Ma, "Learning User Interest for Image Browsing on Small-formfactor Devices" in SIGCHI Conference on Human Factors in Computing Systems, 2005, pp. 671-680.

[25] A. Carlier, V. Charvillat, W. T. Ooi, R. Grigoras, and G. Morin, "Crowd-sourced Automatic Zoom and Scroll for Video Retargeting," in ACM Multimedia, 2010, pp. 201-210.

[26] P. Campisi and A. Neri, "Video watermarking in the 3D-DWT domain using perceptual masking," in Proc. IEEE ICIP, vol. 1, 2005, pp. 997-1000.

[27] L. Rajab, T. Al-Khatib, and A. Al-Haj, "Video Watermarking Algorithms Using the SVD Transform," in European Journal of Scientific Research, 2009, pp. 389-401.

[28] N. Mlik, W. Barhoumi, and E. Zagrouba, "Object-based event detection for the extraction of video key-frames," in International Conference on Multimedia Computing and Systems, Tangier, Morocco, May 2012.