International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:1, 2016

Anonymous Editing Prevention Technique Using Gradient Method for High-Quality Video

Jiwon Lee, Chanho Jung, Si-Hwan Jang, Kyung-Ill Kim, Sanghyun Joo, Wook-Ho Son

Abstract—Since the advances in digital imaging technologies have led to development of high quality digital devices, there are a lot of illegal copies of copyrighted video content on the Internet. Also, unauthorized editing is occurred frequently. Thus, we propose an editing prevention technique for high-quality (HQ) video that can prevent these illegally edited copies from spreading out. The proposed technique is applied spatial and temporal gradient methods to improve the fidelity and detection performance. Also, the scheme duplicates the embedding signal temporally to alleviate the signal reduction caused by geometric and signal-processing distortions. Experimental results show that the proposed scheme achieves better performance than previously proposed schemes and it has high fidelity. The proposed scheme can be used in unauthorized access prevention method of visual communication or traitor tracking applications which need fast detection process to prevent illegally edited video content from spreading out.

Keywords—Editing prevention technique, gradient method, high-quality video, luminance change, visual communication.

I. Introduction

In the past, tangible assets were much more important than intangible assets. However, as content providers and owners started to be concerned about their copyrights, intangible assets are getting more important than tangible assets, so this tendency would be continued. At the same time, the infringement of copyright is also increased. It is easy to find illegally recorded or edited content on the black market or the Internet helps to distribute lots of high quality digital content and recording devices. In this situation, it is natural that they are interested in copyright protection.

Most of illegally made still images contain geometric distortions such as rotation, scaling, and translation (RST). Accordingly, many editing protection schemes have been designed to resist geometric distortion in still images, using such methods as invariant transforms [1], [2], image features [3], [4], template insertion [5], and periodical sequences [6], [7]. Lots of illegally recorded video content contain also some distortions introduced by format conversion (signal-processing distortions), resizing, or capturing by a digital camcorder. Since video content can be treated as a sequence of still images, the image oriented schemes mentioned above can be applied for video content and the video content which is signal embedded by these approaches guarantee good fidelity. However, since these approaches have time-consuming preprocessing phase for each frame, the embedding process takes long time if a HQ video is provided as input signal. Also, these schemes cannot handle video processing attacks such as format conversion and camcorder capturing. Therefore, these image oriented approaches are difficult to use practically for protecting video from illegal editing.

Several papers have addressed video editing prevention schemes robust to geometric distortions and signal-processing distortions. Leest et al. [8] proposed a scheme that exploits temporal axis to embed the signal by changing the mean luminance value of each frame, thereby achieving robustness against geometric distortions. However, the detection performance is affected by luminance signal changes inside the given video. For example, the detection performance is not good on music show because innate high frequency of luminance signal fluctuates too much. Lee et al. [9] exploited local auto-correlation function (LACF) to restore geometric distortions and constructed a mathematical model for synchronization of the embedded signal. Since this scheme has to restore geometric distortions before detection, detection process needs quite large computing time. However, since the target of these schemes is video content, they have good robustness compared with image oriented schemes.

In this paper, we propose a video editing prevention scheme that is robust to both geometric distortions and signal-processing distortions. The proposed scheme have high imperceptibility by using spatial and temporal gradient luminance changing method. Since the proposed scheme exploits only the difference of mean luminance value of each frame in detection process, the detector is very simple, fast, and accurate. Therefore, this scheme can be used in broadcast monitoring or traitor tracking applications which need fast detection process to prevent illegally recorded video contents from spreading out.

The remainder of the paper is organized as follows. In Section II, we propose embedding and detecting process of the proposed scheme. In Section III, we present experimental results to prove the effectiveness of the proposed method. Finally, Section IV provides the conclusion.

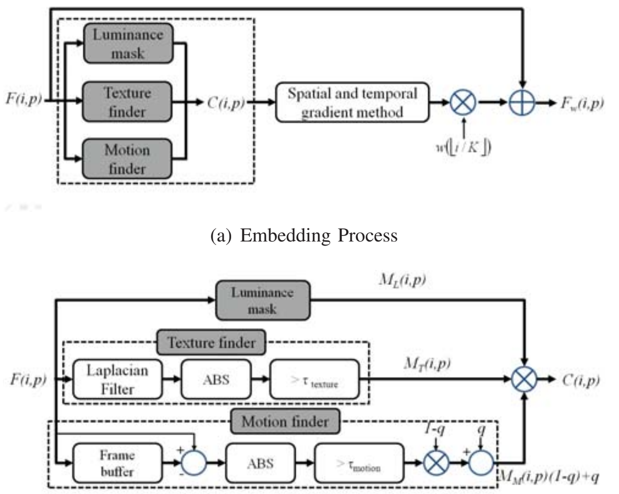
II. Proposed Editing Prevention Scheme

Since the proposed scheme changes the mean luminance value smoothly using spatial and temporal gradient methods, the signal embedded video has better both fidelity and detection performance than previous techniques using luminance changes.

A. Signal Embedding

Fig. 1 (a) shows the proposed embedding process. It consists of two steps:

Jiwon Lee, Chanho Jung, Si-Hwan Jang, Kyung-Ill Kim, Sanghyun Joo and Wook-Ho Son are with the Mobile Content Section, SW·Content Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon, Republic of Korea (e-mail: {ez1005, peterjung, jjangshan, kki, joos, whson}@etri.re.kr).



(a) Embedding Process

(b) Overview of estimating maximum changeable value $C(i, p)$

Fig. 1 Signal embedding process

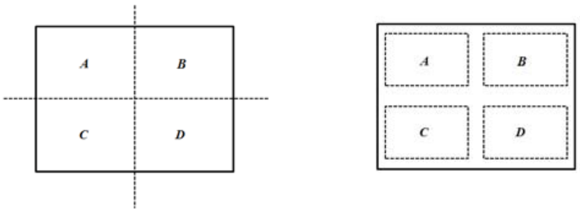
- Estimating the maximum changeable luminance value
- Applying the gradient method

The detail of the embedding process is as follows:

The embedded signal pattern W is a pseudo-random sequence with length L and each signal bit w_l can be 1 or -1. Then, we modify the embedded signal pattern W to make it imperceptible using two finders and one mask: texture finder, motion finder, and luminance mask. More mathematically we have

$$\begin{aligned} C(i,p) &= M_L(i,p)M_M(i,p)M_T(i,p) \\ &\quad +qM_L(i,p)[1-M_M(i,p)]M_T(i,p) \qquad (1) \end{aligned}$$

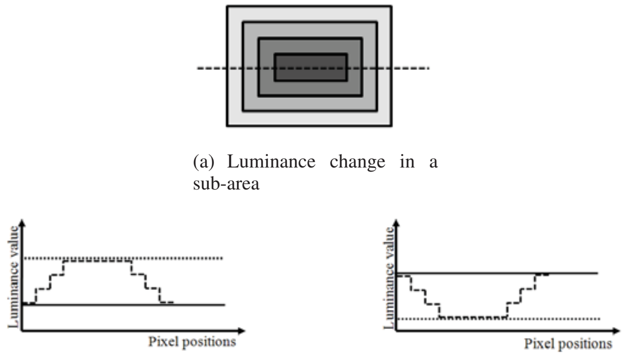
where $C(i,p)$ is the maximum changeable luminance value in pixel position p of i^{th} frame that human eyes cannot perceive the difference from the original luminance value. $M_L(\cdot)$, $M_M(\cdot)$, and $M_T(\cdot)$ are the outputs of the luminance mask, motion finder and the texture finder. The output of the luminance mask uses the same luminance mask table of [8]. $q(0 \leq q < 1)$ is no motion scaling parameter which makes it possible to embed the signal even if there is not enough motion factor. Then, this estimated C is used to embed the signal bit into consecutive K frames. Fig. 1 (b) describes the overview of estimating C.



(a) Embedding area

(b) Detection area

Fig. 2 Signal embedding and detection area



(a) Luminance change in a sub-area

(b) Embedded signal is 1

(c) Embedded signal is -1

Fig. 3 Spatial gradient G_s

Next, we split a frame into 4 sub-areas to embed 4 bits per each frame. Fig. 2 (a) shows the embedding sub-areas of a frame. The proposed scheme achieves high capacity because it has four sub-areas, but spatially visible artifacts between two sub-areas may occur. For example, if the embedded bit of sub-area **A** in Fig. 2 (a) is 1 and the embedded bit of sub-area **B** is -1, we could find the difference of luminance at the boundary between **A** and **B** due to the difference of the luminance value between two sub-areas. These artifacts could be made at the dashed lines in Fig. 2 (a). To solve this problem, we apply the spatial gradient method which changes the luminance value of each sub-area smoothly to reduce the visible artifacts as shown in Fig. 3. The spatial gradient factor $G_s(\cdot)$ is calculated by

$$G_s(p) = \begin{cases} 0.1 & \text{if } 0 \leq abs(B_p - p) < D_p \\ 0.4 & \text{if } D_p \leq abs(B_p - p) < 2 \cdot D_p \\ 0.7 & \text{if } 2 \cdot D_p \leq abs(B_p - p) < 3 \cdot D_p \\ 1 & \text{if } 3 \cdot D_p \leq abs(B_p - p) \end{cases} \qquad (2)$$

where B_p is the closest border of the current pixel p in a embedded area and $abs(\cdot)$ is the absolute value, and D_p is pre-defined constant. Figs. 3 (b) and (c) are cross sections of dashed line of Fig. 3 (a) when the embedded signal is 1 and -1, respectively. In the figure, the solid line represents the original luminance value, the dotted line means the maximum changeable value, and the dashed line is actual changed value of each pixel. Since the proposed embedding process changes the luminance value of each pixel gradiently, the visible artifacts is alleviated. In addition, since the average luminance values of consecutive frames at temporal axis decrease or increase rapidly at each K frames when we embed the different bits into frames, the proposed scheme has a flickering problem. To alleviate this flickering problem, we apply the temporal gradient method. The output of luminance value before and after applying the temporal gradient method is illustrated dashed line and dotted line each, and the original mean luminance is drawn as solid line in Fig. 4. The temporal gradient method changes the luminance value smoothly in temporal axis, so we can eliminate the flickering problem. The temporal scaling factor $G_t(\cdot)$ is calculated by using the magnitude of a sinusoidal signal whose frequency is $1/(2K)$.

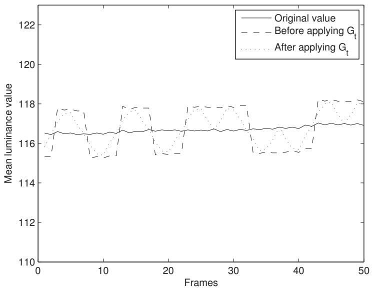


Fig. 4 Temporal gradient G_t

More mathematically we have

$$G_t(i) = sin(\pi/K{\cdot}i) \tag{3}$$

where i is the frame number.

Finally, the signal embedded frame is calculated as:

$$F_w(i,p) = \begin{cases} F(i,p) + G_s(p){\cdot}G_t(i){\cdot}C(i,p) & \text{if } w_{\lfloor i/K \rfloor} = 1 \\ F(i,p) - G_s(p){\cdot}G_t(i){\cdot}C(i,p) & \text{if } w_{\lfloor i/K \rfloor} = -1 \end{cases} \tag{4}$$

where F_w is a signal embedded frame, F is an original frame, and $G_s(\cdot)$, $G_t(\cdot)$ are the spatial and temporal scaling factors come from gradient methods, respectively.

B. Signal Detection

Since the proposed scheme changes the luminance of each sub-area in the embedding process, we exploit the difference of the mean luminance values of each frame to detect the embedded signal. Fig. 2 (b) represents the detection area. Since the signal embedded video content may contain geometric distortions, the detection area is smaller than embedding area. When we detect the signal from the signal embedded video clip, it is only needed the sign of the embedded signal value because the signal pattern is a sequence of 1 and -1. Since the embedder embeds the signal in sinusoidal embedding power temporally, the mean luminance values of K consecutive frames that embed one signal bit form sinusoidal shape. It means that the mean luminance values of the first and the last few frames (side frames) of K consecutive frames are changed not so much, but the mean luminance values of the center frames are changed a lot. The proposed scheme exploits this property to detect the embedded signal. If 1 is embedded in one sub-area of K consecutive frames, the subtraction the mean luminance values of the side frames from the mean luminance values of the center frames would be positive value. On the contrary, if -1 is embedded, the subtraction would be negative value. Fig. 5 describes this concept. Therefore, we can detect the embedded signal $\tilde{\nu}(k,n)$ as:

$$\tilde{\nu}(k,n) = \begin{cases} 1 & \text{if } \mu_{LSF}(n) < \mu_{CF}(n) > \mu_{RSF}(n) \\ -1 & \text{if } \mu_{LSF}(n) > \mu_{CF}(n) < \mu_{RSF}(n) \end{cases} \tag{5}$$

where k is an element number of the embedded signal in K consecutive frames, n is a sub-area number and $\mu_{LSF}(\cdot)$, $\mu_{CF}(\cdot)$, $\mu_{RSF}(\cdot)$ are the mean luminance values of left side frames, center frames and right side frames. Since the signal bits are embedded circularly, $\tilde{w}_r$ of the signal w is estimated by accumulating $\tilde{\nu}(k,n)$:

$$\begin{aligned} \tilde{w}_0(k,n) &= 0 \\ \tilde{w}_r(k,n) &= \tilde{w}_{r-1}(k,n) + \tilde{\nu}(k,n) \end{aligned} \tag{6}$$

where r is the number of detected signal pattern $\tilde{\nu}(k,n)$. Then, we can get the final signal pattern $\tilde{w}$ by using the sign of $\tilde{w}_r(k,n)$.

$$\tilde{w}(k,n) = sign\Big(\tilde{w}_r(k,n)\Big) \tag{7}$$

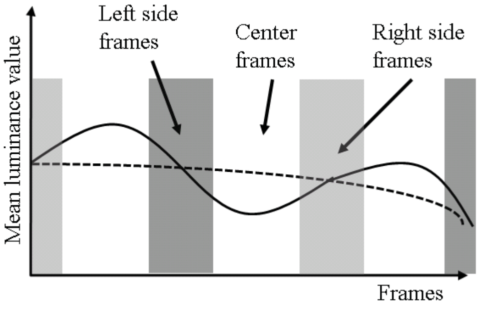
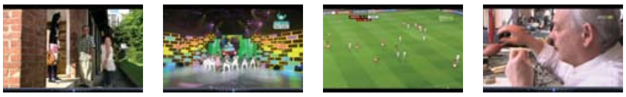


Fig. 5 Original (dashed line) and signal embedded (solid line) mean luminance value



(a) Drama (b) Show (c) Sports (d) Documentary

Fig. 6 Test video clips

TABLE I
RESULTS OF PSNR AND SSIM VALUES FOR TEST VIDEO CLIPS

	Leest's scheme [8]		Proposed scheme	
	PSNR	SSIM	PSNR	SSIM
Drama	41.3271	0.9982	46.0353	0.9989
Show	41.2670	0.9977	44.5087	0.9992
Sports	42.7912	0.9964	45.5723	0.9985
Documentary	41.8229	0.9971	45.4273	0.9987

III. EXPERIMENTAL RESULTS

We carried out experiments to measure the fidelity and robustness of the proposed scheme. The sequence of size 1024

TABLE II
RESULTS OF NCC OF TWO SCHEMES UNDER VARIOUS DISTORTIONS ()

Distortions		No attack	R	S	T	Conv.
Leest's scheme [8]	Drama	0.471	0.464	0.461	0.464	0.465
	Show	0.206	0.201	0.201	0.201	0.156
	Sports	0.331	0.291	0.291	0.295	0.331
	Docu.	0.607	0.560	0.557	0.599	0.560
Proposed scheme	Drama	0.718	0.703	0.702	0.715	0.682
	Show	0.434	0.426	0.441	0.411	0.412
	Sports	0.851	0.816	0.816	0.815	0.826
	Docu.	0.830	0.810	0.801	0.811	0.779

R: rotation 5°, S: scaling to 720×480, T: translation by 20 pixels in each axis, Conv.: format conversion from MPEG-2 to MPEG-4 bits was used as watermark signal W and four 5-minutes full-HD (1920×1080) video clips shown in Fig. 6 were tested. We used the same textured and motion finders thresholds with the compared scheme [8] in the experiments, and K was 5. Also, we set the threshold of 10^{-7} for the probability that a false positive would occur is around 0.14. In detection process, D_p for the spatial gradient method was 30. Moreover, the mean luminance values of the recorded footage were upsampled from 24 samples per second to 30 samples per second before detecting when it is needed. Also, we assume that the detector knows the exact start position of embedded signal.

Since the embedding method of Leest's scheme is similar with the proposed scheme, we compared the performance of two schemes. In order to show the performance of our spatial and temporal gradient methods, we compared peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of two embedding processes to measure the fidelity. The high values of PSNR and SSIM specifically indicate that the video clips are nearly lossless after the signal embedding process. Table I shows the results of PSNR and SSIM values for test video clips. The average PSNR and SSIM values in compared scheme were 41.8020 dB and 0.9973, and the average PSNR and SSIM values in the proposed scheme were 45.3859 dB and 0.9988. Thus, in average 4 dB in PSNR and 0.0015 in SSIM are enhanced. Since human eyes feel uncomfortable and unsatisfactory under 42 dB PSNR in full-HD video clips, the difference between PSNR value of two schemes is meaningful. In addition, we computed the normalized cross correlation (NCC) of two schemes to measure the robustness after various attacks. The signal detector was performed on the signal embedded video clips which are undergone RST attacks and format conversion. The results of NCC of compared scheme and the proposed scheme under certain distortions are described in Table II. In terms of NCC values, the robustness of the proposed scheme is improved from 37% to 157% from this result. Generally, it is difficult to have both good robustness and good fidelity in watermarking applications because they have tradeoff relation. The proposed scheme achieves NCC improvement in spite of high PSNR values.

IV. Conclusion

In this paper, we proposed the HQ video editing prevention scheme which achieves both high imperceptibility and robustness. We exploit temporal axis to embed signal and apply the spatial and temporal gradient method to improve the fidelity and the robustness. Experimental results show the proposed scheme is robust to geometric distortions and signal-processing distortions in terms of NCC values while keeping the embedding distortion lower in terms of PSNR and SSIM values. In future research, we will investigate the effectiveness of our proposed approaches in visual communication (VC) systems using editable visual objects (EVO). Since the relation map of VC systems is automatically evolved by EVO usage of designated user, blocking the illegal access of unauthorized users is extremely important. In this situation, the proposed scheme can be a good solution.

Acknowledgment

This work was supported by the Industrial Strategic Technology Development Program, (10044312, Developments of Visual Communication Platform and Authoring Tool for Emotional Expression) funded by the Ministry of Science, ICT and Future Planning (MSIP, Korea).

References

[1] J. O. Ruanaidh and T. Pun, "Rotation, scale, and translation invariant spread spectrum digital image watermarking," *Signal Processing*, vol. 66, no. 3, pp. 303–317, May. 1998.

[2] C. Lin, J. Bloom, I. Cox, M. Miller, and Y. Lui, "Rotation, scale, and translation-resilient watermarking for images," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, May. 2001.

[3] M. Alghoniemy and A. Tewfik, "Geometric distortion correction in image watermarking," in *Proc. SPIE*, vol. 3971, pp. 82–89, May. 2000.

[4] P. Bas, J. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1014–1028, Sep. 2002.

[5] S. Pereira and T. Pun, "Robust template matching for affine resistant image watermarks," *IEEE Trans. Image Processing*, vol. 9, no. 6, pp. 1123–1129, Jun. 2000.

[6] M. Kutter, "Watermarking resisting to translation, rotation, and scaling," in *Proc. SPIE*, vol. 3528, pp. 423–431. Jan. 1999.

[7] S. Voloshynovskiy, R. Deguillaume, and T. Pun, "Multibit digital watermarking robust against local nonlinear geometrical distortions," in *Proc. ICIP*, vol. 3, pp. 999–1002, Oct. 2001.

[8] A. V. Leest, J. Haitsma, and T. Kalker, "On digital cinema and watermarking," in *Proc. SPIE*, vol. 5020, pp.526–535, Jun. 2003.

[9] M. J. Lee, K. S. Kim, H. Y. Lee, T. W. Oh, Y. H. Suh, and H. K. Lee, "Robust watermark detection against d-a/a-d conversion for digital cinema using local auto-correlation function," in *Proc. ICIP*, vol. 37, pp.425–428. Oct. 2008.

Jiwon Lee received the B.S. degree in Computer Engineering from Kyungpook National University, Republic of Korea, in 2008, and the Ph. D. degree in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea, in 2013. Since 2013 he has been a senior researcher in the Mobile Content Section, SW·Content Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea. His research interests include mobile content processing, image/video watermarking, and image/video processing.

Chanho Jung received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2013. From 2006 to 2008, he was a Research Engineer with the Digital Television Research Laboratory, LG Electronics, Seoul. Since 2013, he has been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, where he is currently a senior researcher. His current research interests are computer vision, pattern recognition, and image processing.

Si-Hwan Jang received the B.S. degree in Industrial Engineering from Kangwon National University, Republic of Korea, in 2010, and the M.S. degree in information security from Graduate School of Information Security, Korea University, Republic of Korea, in 2013. Since 2013 he has been a researcher in the Mobile Content Section, SW·Content Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea. His research interests include Mobile Content Processing, Steganography,Watermarking, Heuristic Algorithm and Standardization of MPEG.

Kyung-Ill Kim received the Ph. D. degree in Computer Science from ChungNam NationalUniversity, Daejeon, Korea, in 2014 and the M.S. degree in Computer Information System from Korea University, Chochiwon, Rep. of Korea, in 2001, respectively. Since 1983, he has worked as a principal researcher at Electronics and Telecommunications Research Institute (ETRI), Daejeon, Rep. of Korea. His research interests are cloud computing, visual communication and multimedia system.

Sanghyun Joo received his BS and MS degrees in electronic engineering from Dongguk University, Seoul, Rep. of Korea, in 1989 and 1994, respectively. He received his PhD degree in computer science, from Niigata University, Japan, in 1999. From 1994 to 1996, he worked as a research engineer for KaiTech, Seoul, Rep. of Korea. From 1999 to 2001, he worked as a research associate at Niigata University. Since then, he joined ETRI, where he is now a Director in Mobile Content Research Lab. Since 2012, he has worked for MPEG-21 UD (ISO/IEC 21000-22) as both a chairman and an editor. His research interests include media context and control, user description, visual communication technologies.

Wook-Ho Son received the B.S. degree in computer science from Yonsei University, Seoul, Korea, in 1987 and the M.S. and Ph.D. degrees from Texas A&M University, College Station, in 1996 and 2001, respectively. Currently, he is in charge of the Content Platform Research Department, ETRI, Daejeon, Korea. His research interests include digital holography, virtual reality, augmented reality, haptic interaction, physically based dynamic simulation, and robotics.