Journal of Korea Multimedia Society Vol. 11, No. 6, JUNE 2008(pp. 796-806)

Video Watermarking Scheme for Scalable Video Coding using ROI

Ji-Sun Yoon[†], Seong-Geun Kwon[††], Suk-Hwan Lee[†††], Yoon-Chul Song[††††], Min-Hwan Kim[†††††], Ki-Ryong Kwon[††††††]

ABSTRACT

This paper presents a blind video watermarking algorithm that has the robustness against spatial, temporal, and SNR scalability and transcoding for the copyright protection of video contents in heterogeneous multimedia service. The proposed process of watermark embedding and detecting is accomplished on base layer for considering spatial scalability. The watermark consists of the string and the ordering number of string for considering temporal scalability. Thus, each of frames has the bitstream of one character and a ordering number of its character. To robust against FGS, the proposed algorithm quantizes low and middle frequency coefficients in ROI region of each of frames and embeds its watermark bitstream into the specific bits of the quantized coefficients. Experimental results verified that the proposed algorithm satisfies the invisibility of watermark and also has the robustness against spatial scalability, temporal scalability and FGS.

Key words: Video Watermarking, Scalable Video Coding

1. INTRODUCTION

Recently, the concept of universal multimedia access (UMA) has been on the issue as the service of multimedia contents in heterogeneous environment linked with broadcasting network, wired/wireless internet and mobile communication network has been requested rapidly to many people. UMA addresses the delivery of multimedia resources under different and varying network conditions, diverse terminal equipment capabilities, specific user or creator preferences and needs and usage environment conditions.

In heterogeneous multimedia service with UMA, scalable video coding (SVC) [1-5] and multimedia transcoding [6-8] are well known as the core video processing technique. The former, SVC, is standardized as MPEG-4 SVC or H.264 Scalable Extension (SE) by Joint Video Team (JVT) that MPEG (moving picture experts group) under ISO/IEC and VCEG (video coding experts group) under ITU-T organized together. SVC can provide the service of multimedia video contents according to conditions of various devices since it performs spatial, temporal, and SNR scalabilities to a video

※ Corresponding Author : Ki-Ryong Kwon, Address : (608-737) 599-1 Daeyeon-dong, Nam-gu, Busan, Korea, TEL : +82-51-620-6495, FAX : +82-51-620-6350, E-mail : krkwon@pufs.ac.kr

Receipt date : Jan. 9, 2008, Approval date : June 20, 2008

† Dept. of Computer Engineering, Pusan National University
(E-mail : yjs4383@nate.com)
†† Samsung Electronic Company
(E-mail : seonggeunkwon@hanmail.net)
††† Dept. of Information Security, Tongmyong University
(E-mail : skylee@tu.ac.kr)
†††† Korea Association of Information & Telecommunication
(E-mail : uckysong@kait.or.kr)
††††† Dept. of Computer Engineering, Pusan National University
(E-mail : mhkim@pusan.ac.kr)
†††††† Div. of Elec. Comp., and Telecomm. Eng., Pukyong National University

※ This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-521-D00461).

content. The other, transcoding, is a must due to the diversity of mobile devices. This diversity requires an intermediate state of content adaptation in order to make sure that the source content will adequately present on the target device it is sent to. Both SVC and transcoding aim to achieve UMA. But SVC is to transmit or store a source video content with a variety of spatial or temporal resolutions or qualities which is the purpose of video bit-stream scalability for adapting dynamically the characteristics of modern video transmission systems. On the other hand, transcoding is to convert a source video content into an optimized format of another user's device environment.

As mentioned above, with developing technique of UMA, the necessity of unified security system has been raised to provide the security, reliability and authentication of contents in heterogenous multimedia service. As this need for copyright protection in video system, video watermarking techniques had been researched rapidly[9-11]. However, most of video watermarking have been proposed for the copyright protection of single content that can be transferred in singular network and for having the robustness against MPEG compression and various video processing algorithms. There has not been researched about robust video watermarking against MPEG-4 SVC (H.264 SE) and transcoding in heterogeneous multimedia service.

This paper proposed a blind video watermarking with the robustness against spatial scalability, temporal scalability and FGS of MPEG-4 SVC. The proposed algorithm generates a watermark string that consists of copyright string and ordering number of its string and embeds bitstream that one character and an ordering number are converted to ASCII code into DCT frequency of each frame. Thus, each frame has one character and an ordering number among watermark string for the robustness against temporal scalability. The base layer of spatial scalability is used as the reference layer in the watermark embedding and extracting process to be robust against spatial scalability. The embedding targets are selected to specific bits of low and middle frequency coefficients in ROI (region of interest) region of base layer considering the robustness against FGS. From experimental results, we confirmed that the proposed algorithm has the robustness against spatial scalability, temporal scalability and FGS of MPEG-4 SVC and the quality degradation has not been perceived since average PSNR of various FGS layers is above about 30dB.

2. Related Works

2.1 Previous video watermarking

Hartung, et al. [9] suggested the remarkable video watermarking algorithm in 1998 that the watermark is embedded by using secure spread spectrum method on the 1-dimensional line scanned signal of 3-dimension signal consists of spatial and temporal factors in decoded video frame. This method can embed watermark without bit stream increment while compressing but has high complexity because it decodes the encoded signal for embedding watermark. In 2001, Kong, et al. [10] suggested a method that performed 8x8 block DCT and quantization on original video followed by embedding watermark in DC coefficient of each block by referencing the quantized values. This method embeds watermark in all frames, therefore, is robust on frame-rate change and typical compression but, is weak on scaling or resolution change because the watermark is detected from least significant bit (LSB) of quantized coefficients. In 2006, Wang, et al. [11] suggested a method that the watermark is embedded in AC coefficients nearby DC coefficients of I-frames with referencing DCT coefficients of before and after frames after performing DCT by group unit on selected I-frame and its before and after frames while encoding MPEG-2. This method is robust on

MPEG-2 compression and typical geometrical attacks but, is weak on dropping of reference frame or frame rate change. From the above paragraphs, we can conclude that many researches about watermarking techniques on typical video sources were suggested until now. But the most techniques are for copyright protection on single content which is transferred in single network that do not consider about multiple contents which are transferred in the different multimedia services. Therefore, there are some problems that these are weak on SVC and transcoding.

2.2 MPEG-4 SVC

MPEG-4 SVC supports spatial, temporal, and resolution scalable coding with minimizing the coding efficiency decline. SVC decoder can decode various spatial and temporal resolution video from one bit stream using a part of bit stream selectively depend on the capacity of network and display device [1-5]. Fig. 1 shows the structure of MPEG-4 SVC that offers this scalability.

2.3 Problems in view of watermarking

As mentioned above, MPEG-4 SVC supports spatial, temporal, SNR scalability and FGS. First, in spatial SVC, it is difficult to detect watermark because the spatial resolution can be changed by data loss or change of frequency coefficients after embedding watermark. Second, in temporal SVC, it is difficult to detect watermark because the embedded frame can be dropped by temporal SVC that offers several different frame rate and the connection between frames is cut off. Third, when SNR SVC and FGS encode by assigning different QP values to each layer, as quantization step size is grown, quantization error is also grown and degrades the image. Therefore, when SNR or FGS layer is changed, it is difficult to detect watermark because of increasing of data loss at high frequency. Lastly, video transcoding can degrade the video contents by converting format, resolution of video data and by dropping frames. In this case, if we embed watermark in whole frame, it is difficult to detect accurate watermark because the watermark information in decoded video frame can be transformed or damaged. Hence, considering above problems on the view of watermarking system, watermark should be detected regardless the number of layers of decoded video contents and resolution and also should be detected in case of frame dropping of watermark embedded frame or resolution lowering when FGS layer is changed. Also, frame degrading with temporal connectivity of frames and bitstream increment caused by watermark embedding should not be produced. Moreover, watermark should be detected in spite of change of frame rate or frame size caused by format change and in case of video sequence that

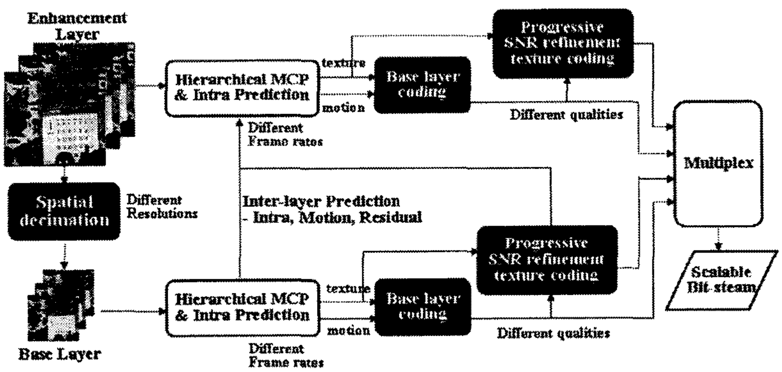

Fig. 1. The structure of MPEG-4 SVC.

can not be reused with specific CODEC.

3. Proposed Video Watermarking

In this paper, we propose robust blind watermarking technique against MPEG-4 SVC and multimedia transcoding for providing copyright protection of video contents on different multimedia services. Fig. 2 shows the proposed video watermarking system in MPEG-4 SVC.

Spatial SVC in MPEG-4 SVC is separated into several enhancement layers $f_{Ln}(x,y)$, $0 \le x < NH$, $0 \le y < NV$and a base layer $f_{L0}(x,y)$, $0 \le x < NH/2^{(L\_MAX-n)}$, $0 \le y < NV/2^{(L\_MAX-n)}$ depending on spatial resolution. NH and NV means horizontal and vertical resolution of original video respectively, n has a value $0 \le n \le L\_MAX$, and if $n=0$, it represents base layer. L_MAX represents the highest resolution layer in spatial scalability. $2^{(L\_MAX-n)}$ means the resolution is reduced in terms of power of 2 when the original video resolution is down sampled to produce the base layer. Watermark embedding and detecting is performed based on spatial base layer with considering spatial scalability. $f^*_{L0}(x,y)$ can be obtained by embedding watermark in $f_{L0}(x,y)$ using proposed method. $f^*_{L0}(x,y)$ in here is a video sequence of watermarked base layer. For obtaining video sequence of watermarked enhancement layer $f^*_{Ln}(x,y)$, we calculate weighted sum of 16 neighbor pixels of pixel $f_{i,j}$ in $f_{L0}(x,y)$ with interpolation function as shown in eq. (1) and up-sample it.

$$I(f_{L0}(x,y)) = \begin{cases} \frac{1}{2}|f_{i,j}|^3 - |f_{i,j}|^2 + \frac{2}{3}, & 0 \le |i|,|j| < 1 \\ -\frac{1}{6}|f_{i,j}|^3 + |f_{i,j}|^2 - 2|f_{i,j}| + \frac{4}{3}, & 1 \le |i|,|j| < 2 \\ 0, & 2 \le |i|,|j| \end{cases} \tag{1}$$

During watermark embedding, for minimizing the information difference, we obtain the information difference $r_{Ln}(x,y)$ between the original video frame and the up-sampled one as eq. (2).

$$r_{Ln}(x,y) = f_{Ln}(x,y) - I(f_{L0}(x,y)) \tag{2}$$

After that, we can obtain the video sequence of watermarked enhancement layer $f^*_{Ln}(x,y)$ by up-sampling $f^*_{L0}(x,y)$ using eq. (1) followed by adding $r_{Ln}(x,y)$ as eq. (3).

$$f^*_{Ln}(x,y) = I(f^*_{Ln}(x,y)) + r_{Ln}(x,y) \tag{3}$$

Spatial SVC in MPEG-4 SVC is encoded this

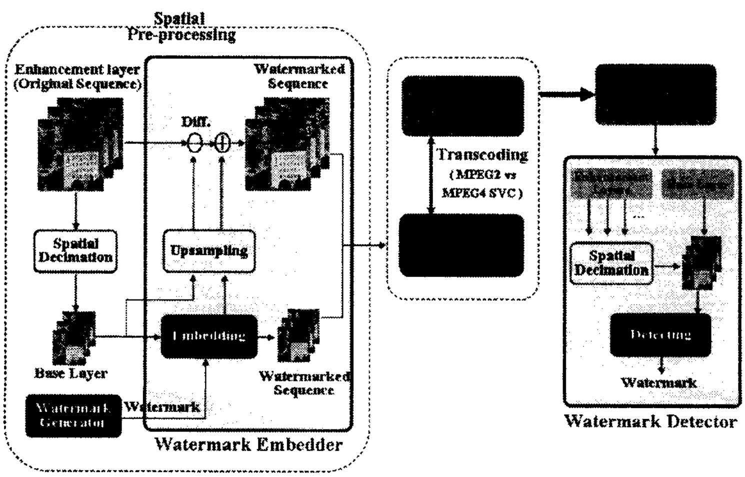

Fig. 2. Proposed video watermarking system in MPEG-4 SVC.

obtained $f^{*}_{Ln}(x,y)$, $0 \le n < L\_MAX$ by layer. On the other hand, MPEG-2 is the compression technique for single content transferred in single network. Therefore, it does not support several layers, each $f^{*}_{Ln}(x,y)$ is encoded by MPEG-2. When watermark detecting, the watermark is detected in decoded video with arbitrary resolution after converting it to base layer resolution.

3.1 Watermark generation algorithm

Fig. 3 shows the block diagram of watermark generation procedure using ASCII code, and the description of each step is as follows.

Watermark string is defined as WS= $\{WS_i | i \in [1,N]\}$, $N = \text{SVC GOP} \times n$, $n > 0$ and N means the length of the string. The watermarked frame can be eliminated because the frame rate of temporal SVC is changeable. Because it is impossible to detect watermark from eliminated frames, we generate watermark WC= $\{WC_i = i\,WS_i | i \in [1,N]\}$ including the order $i \in [1,N]$ of each character of watermark string WS. The generated WC_i is ASCII code that each is 8bit and total 16bit(2 characters). For embedding WC in the original video sequence repeatedly, the original video sequence is grouped with N which is same with the length of watermark string, i. e. a group GF= $\{F_i | i \in [1,N]\}$ is made from N frames F_i. After that, we can obtain P(WC)= $\{p(WC)_i | \ i \in [1,N]\}$ by random permutation of WC for embedding it by unit of group GF repeatedly and convert this to bit stream. At this point, the result of permuting function P(WC) is represented as pseudo-random number and the number of cases is N! In here, the embedded bit stream of watermark per frame is defined as W= $\{W_i | i \in [1,N_w]\}$, $N_w = R_w \times 16$. N_w means the number of embedded watermark bits per frame and R_w is the number of times of WS. We can obtain encrypted bit stream P(W)= $\{p(W)_i | i \in [1,N_w]\}$ that will be embedded in each frame by random permuting of bit stream W.

3.2 Key generation algorithm for watermark embedding

We introduce a method to generate proposed watermark embedding key for blind watermarking. We extend the 8x8 quantization table Q_o= $\{q_{i,j} | 0 \le i,j \le 7\}$ to ROI size using horizontal and vertical interpolator as shown in eq. (4) and eq. (6).

$$q^{ROI}_{i,j} = \begin{cases} q_{i,m}, & j = l_H \times m \\ round((q_{i,m} + q_{i,m+1})/2), & j = round((l_H \times m + l_H \times (m+1))/2) \\ & \vdots \\ q_{i,7}, & j = s \times (R_H / 2^{(L\_MAX)})/8 - 1 \end{cases} \quad (4)$$

$$l_H = round(s \times (R_H / 2^{(L\_MAX)})/8),\ 0 \le m \le 7 \quad (5)$$

$$q^{ROI}_{i,j} = \begin{cases} q_{m,j}, & i = l_V \times m \\ round((q_{m,j} + q_{m+1,j})/2), & i = round((l_V \times m + l_V \times (m+1))/2) \\ & \vdots \\ q_{7,j}, & i = s \times (R_V / 2^{(L\_MAX)})/8 - 1 \end{cases} \quad (6)$$

$$l_V = round(s \times (R_V / 2^{(L\_MAX)})/8),\ 0 \le m \le 7 \quad (7)$$

The watermark embedding key Q_ℓ is generated by zigzag scanning except $q^{ROI}(0,0)$, $q^{ROI}(0,1)$,

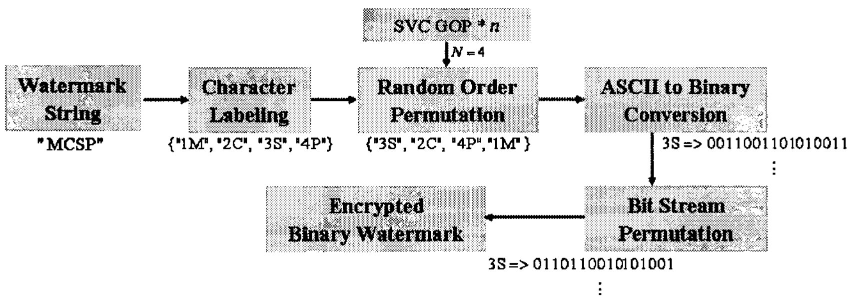

Fig. 3. The block diagram of proposed watermark generation procedure.

$q^{ROI}(1,0)$, $q^{ROI}(1,1)$of the extended quantization table $Q_{ROI}=\{q^{ROI}_{i,j} | 0 \le i \le s \times (R_V/2^{(L\_MAX)}-1), 0 \le j \le s \times (R_H/\}$ and averaging the three adjacent neighbors of these values. In this paper, we uses ℓ for 1-dimensional declaration and means ℓ^{th} quantization coefficient or ℓ^{th} DCT coefficient.

3.3 Watermark Embedding Algorithm

Transcoding converts a high quality resource to a low quality resource by reducing a bit stream or changing spatial or temporal axis of contents. In this case, if we embed watermark in whole frame, it is difficult to detect the watermark because the watermark information can be transformed or damaged. Considering above problem, ROI $R(i,j) \in f_{L0}(x,y)$ in each frame of base layer $f_{L0}(x,y)$ is extracted as shown Fig. 4.

ROI is defined as $R(i,j)$, $a \le i < (a+R_H)$, $b \le j < (b+R_V)$. In here, $R_H = s \times (NH/\ 2^{(L\_MAX)})$, $R_V = s \times (NV/2^{(L\_MAX)})$ and s, a factor that determines the size of ROI, is a value between $0 < s \le 1$ and a, b is the start point of the ROI. Generally, the images of video are consider as interest region because of the center of frame which has a lot of motion area. IF size of ROI will be large, The watermarks are embedding in more information of the

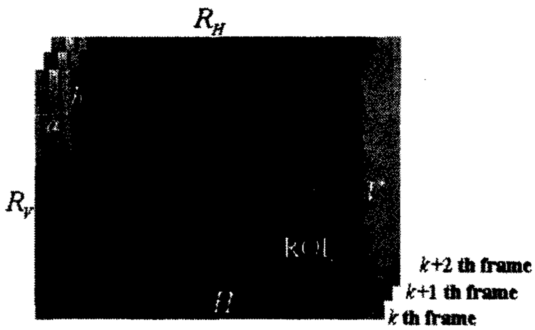

Fig. 4. ROI in each frame of base layer

frames of video. While loss possibility of watermark is grow up and invisibility will be drop which transcoding such as geometric attacks are due to change re-scaling, rotation, cropping. In contrast to size of ROI will be small, loss possibility of watermark is small and invisibility will be grow in geometric attacks, but embedding quantity of watermark is decrease. The size of ROI is experimentally determined to $s = \frac{7}{8}$ which is from trade-off relation. Fig. 5 shows the block diagram of watermark embedding per frame, and the description of each step is as follows.

For embedding watermark robust on FGS of MPEG-4 SVC, the proposed method performs DCT in selected ROI in each frame as shown in eq. (8).

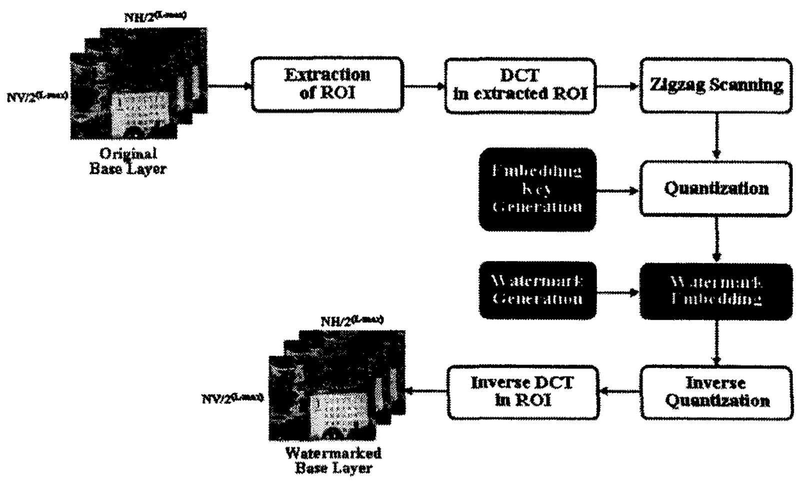

Fig. 5. The block diagram of proposed watermark embedding per frame.

$$C(u,v) = \frac{2}{\sqrt{R_H \cdot R_V}} C(u)C(v) \sum_{i=0}^{R_H-1} \sum_{j=0}^{R_V-1} \cos\left[\frac{(2i+1)u\pi}{2R_H}\right] \cos\left[\frac{(2j+1)v\pi}{2R_V}\right] R(i,j) \quad (8)$$

$$\text{where } C(\gamma) = \begin{cases} \frac{1}{\sqrt{2}} & if\ \gamma = 0 \\ 1 & else \end{cases}$$

Among the converted DCT coefficient $C(u,v)$, we except DC coefficient $C(0,0)$ which means the average brightness of an image and its neighbors $C(0,1)$, $C(1,0)$, $C(1,1)$. After that, we perform zigzag scan for the same number of times N_w which is the embedding number of watermark bit per frame generated in chapter 3.2 on the low·mid frequency coefficients that are less sensitive about quantization error. At this point, we embed watermark using the DCT coefficient ZC_{N_w} contained as ordering of zigzag scan and the embedding key Q_ℓ of watermark generated in chapter 3.2 as shown in eq. (9).

$$ZQC_n^* = \begin{cases} ZQC_n \mid (1 << s), & if(w_i = 1) \\ ZQC_n \,\&inv\ (1 << s), & otherwise \end{cases} \quad (9)$$

$$\text{where } ZQC_n = floor\left[\frac{ZC_{N_w}}{Q_\ell \times \alpha}\right]$$

In eq. (9), s means the embedding strength of watermark and n has a value between $0 \le n \le N_w$ and α is quantization scale vector that used 1 in this paper. In here, | is OR bit operator, & is AND bit operator. In other words, of watermark bit $w_i = 1$, set up the third LSB bit of quantized coefficient ZQC_n 1, otherwise 0. As this procedure, we embed watermark in each DCT coefficient for N_w. We can obtain watermarked video frame by performing inverse quantization and inverse DCT as eq. (10) and eq. (11) to watermarked frequency coefficient ZQC_n^*.

$$C_n^* = ZQC_n^* \times Q_\ell \times \alpha \quad (10)$$

$$R^*(i,j) = ROI\_IDCT(C_n^*) \quad (11)$$

3.4 Watermark Detecting Algorithm

The bit stream of watermarked video sequence which is encoded by MPEG-2 or MPEG-4 SVC can be contained to restored video sequence by decoder. In procedure of detecting watermark, cause each layer of the restored video by MPEG-4 SVC decoder has different resolution, we spatially down-sample the random encoded layer until we get the resolution of base layer. After that, we perform DCT and zigzag scan to ROI of each frame which is selected when embedding watermark. Through eq. (12), we can detect the watermark bit from quantized DCT coefficient $\widehat{C_n}$.

$$\widehat{w_n} = \begin{cases} 1, & if(\widehat{C_n} \& (1 << s)) = TRUE \\ 0, & otherwise \end{cases} \quad (12)$$

$$\text{where } \widehat{C_n} = floor\left[\frac{\widehat{C_{N_w}}}{Q_\ell \times \alpha}\right]$$

In eq. (12), $\widehat{w_i}$ means the detected watermark. That is, if the third LSB bit of $\widehat{C_n}$ is 1, we can detect $\widehat{w_i} = 1$, otherwise $\widehat{w_i} = 0$. The bit number N_w of the detected watermark bit stream W consists of watermark string WS repeated by R_w. Therefore, for converting W to WS by using ASCII code, we use statistical probability through R_w. We can obtain $p(WC)_i'$ by converting this detected bit to watermark string. Finally, after sorting by the included ordering character of $p(WC)'$ which is embedded by the unit of group GF, we can detect watermark string WS.

4. EXPERIMENTAL RESULTS AND ANALYSIS

For performance evaluation of the proposed method, we used Crew CIF(352x288)@30fps and Football CIF@30fps test video. In our experiments, the watermark we used is 35 characters consist of alphabets and numbers. we encoded them to 16bit (2 characters) code by randomly permuting these characters and converting them to binary bit stream of 8bit extended ASCII code. For experiments on MPEG-4 SVC, we used Joint Scalable

Table 1. The conditions of SVC used in experiments.

	SVC	Scalable Parameter
MPEG-4 SVC	Spatial SVC	CIF/QCIF
	Temporal SVC	15/30 (frame/sec)
	FGS	Layer 0-3

Video Model(JSVM) [3] provided by JVT for reference software of SVC and for transcoding, MPEG-2 v12. The down-sampling about size converting of video screen and of video frame rate used DownConvertStatic provided by JSVM 7.0. This program down-samples video frame of YUV format without resolution re-calculation. For experiments on spatial scalability, we used 2 layers one for low resolution with quantization value 30 and one for high resolution with quantization value 20. The interval of I-frame is 96 and GOP for SVC is determined to have 4 which is same with the number of watermark character. The conditions of SVC in this experiments are in table 1.

4.1 Invisibility experimental

For evaluation of invisibility of the proposed method, we compared average PSNR between watermarked video and original video. Fig. 6 and 7 show the first frames of watermarked base layer

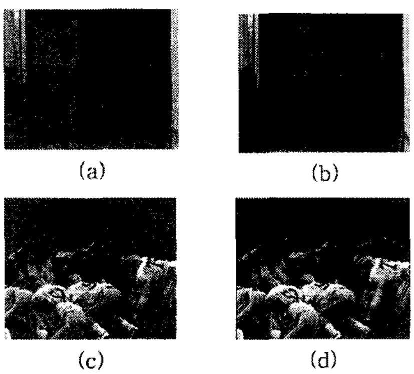

Fig. 6. The base layer of (a) original frame and (b) watermarked frame in Crew video, (c) original frame and (d) watermarked frame in Football video.

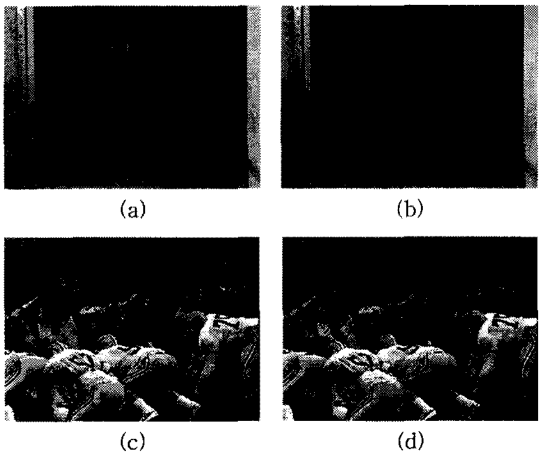

Fig. 7. The enhancement layer of (a) original frame and (b) watermarked frame in Crew video, (c) original frame and (d) watermarked frame in Football video.

and enhancement layer of each video data by using the proposed method. Table 2 shows average PSNR of each SVC layer of the 30fps video data that are applied MPEG-4 SVC and MPEG-2 before and after embedding watermark.

As you can see in Fig. 6 and 7, the watermark is embedded invisibly because the difference of original and watermarked video frame is very small. Also, from Table 2, each watermarked video encoded by MPEG-2 and MPEG-4 SVC has about 32.08~39.22dB average PSNR which prove that it shows good resolution.

4.2 Robustness experimental

For robustness evaluation of the proposed method, we compared bit error rate(BER) of detected watermark string after transcoding between MPEG-2 and MPEG-4 SVC and the proposed method and the previous methods. Fig. 8 shows the embedded watermark characters P(WC) which are randomly permuted by the unit of group GF and Fig. 9 shows the detected watermark string from each SVC when we performed transcoding from MPEG-2 to MPEG-4 SVC in Crew video. We used "MCSP" for WS(N=4). The shadowed blocks in

Table 2. average PSNR(dB) of each SVC at the 30fps Crew and Football video.

SVC Parameter \ Test Video		CREW		FOOTBALL	
Resolution	FGS Layer	Transcoding from MPEG-2 to MPEG-4 SVC before watermark is embedded	Transcoding from MPEG-2 to MPEG-4 SVC after watermark is embedded	Transcoding from MPEG-4 SVC to MPEG-2 before watermark is embedded	Transcoding from MPEG-4 SVC to MPEG-2 after watermark is embedded
QCIF	0	40.67	37.56	37.19	35.38
	1	44.12	38.45	41.05	37.26
	2	47.35	38.93	43.86	38.11
	3	51.38	39.22	45.89	38.50
CIF	0	40.23	35.49	35.27	32.08
	1	41.84	35.71	36.60	32.51
	2	42.96	35.81	37.19	32.64
	3	43.73	35.88	37.41	32.69

Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Watermark	3S	2C	4P	1M	4P	2C	3S	1M	1M	2C	3S	4P	4P	3S	2C	1M	3S	2C	4P	1M	3S	2C	1M	4P	1M	2C	4P	3S	1M	2C

Fig. 8. The embedded watermark characters.

Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
FGS L3	3S	2C	4P	1M	4P	2C	3S	1M	1M	2C	3S	4P	4P	3S	2C	1M	3S	2C	4P	1M	3S	2C	1M	4P	1M	2C	4P	3S	1M	2C
FGS L2	3S	2C	4P	1M	4P	2C	3S	1M	1M	2C	3S	4P	4P	3S	2C	1M	3S	2C	4P	1M	3S	2C	1M	4P	1M	2C	4P	3S	1M	2C
FGS L1	3S	2C	4P	1M	4P	2C	3R	1M	1M	2C	3S	4P	4P	3S	2C	1M	3S	2C	4P	1M	3S	2C	1M	4P	1M	2C	4P	3S	1M	2C
FGS L0	3S	2C	4P	1M	4P	2C	3R	1M	1M	2A	3S	4P	4P	3S	2C	1M	3S	2C	4P	1M	3S	2C	1M	4P	1M	2C	4P	3S	1M	2C

(a)

Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
FGS L3	3S	4P	4P	3S	1M	3S	4P	2C	3S	4P	3S	1M	1M	4P	1M
FGS L2	3S	4P	4P	3S	1M	3S	4P	2C	3S	4P	3S	1M	1M	4P	1M
FGS L1	3S	4P	4P	3R	1M	3S	4P	2C	3S	4P	3S	1M	1M	4P	1M
FGS L0	3S	4P	4P	3R	1M	3S	4P	2C	3S	4P	3S	1M	1M	4P	1M

(b)

Fig. 9. Detected watermark from each FGS layer of (a) 30fps and (b) 15fps that is performed transcoding from MPEG-2 to MPEG-4 SVC in Crew video with QCIF size.

Fig. 9 mean the watermark is detected.

Fig. 10 shows the detected watermark characters P(WC)' and BER of 2-layer resolution of CIF/QCIF, 30/15fps, and FGS layer 0-3 when transcoding from MPEG-2 to MPEG-4 SVC. Through the results of resolution and frame rate, we can see that the previous methods can not detect the watermark in CIF or low frame rate, but the proposed method can detect 80%(FGS layer 0)~95%(FGS layer 3) watermark among 60 embedded watermark characters with 2~16% BER. In the result of each FGS layer, over 95% of watermark is detected in FGS layer 3 which shows the best resolution and over 85% in FGS layer 1~2 which shows medium resolution. The detection error rate in FGS layer 0 was a little higher than other FGS layers, but we can see that the watermark is detected over 60~70%. After re-arranging the detected watermark by the unit of group GF and sorting by the order, we can confirm the embedded watermark string WS. Through experiments, we can verify that the proposed method is robust against SVC of spatial, temporal, and SNR.

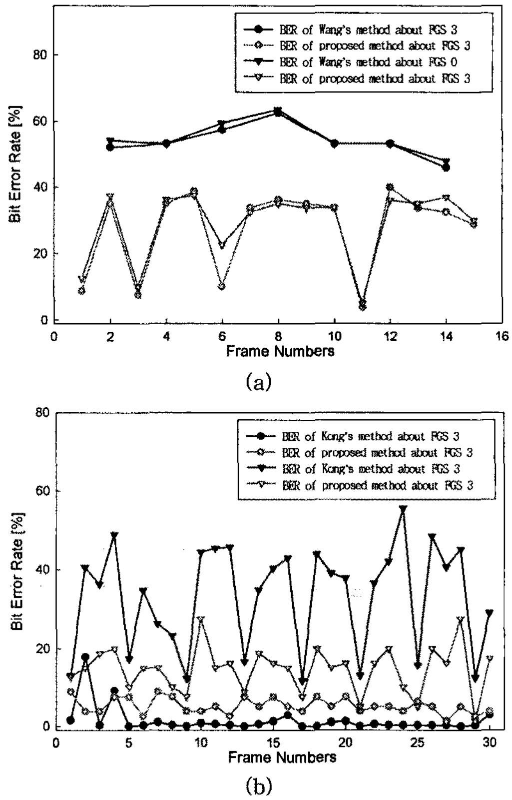

Fig. 10. BER from each FGS layer of (a) QCIF@15fps (b) CIF@30fps in Crew video.

5. CONCLUSIONS

A blind video watermarking scheme for providing safety, authenticity, and copyright protection is proposed in this paper, which is robust to MPEG-4 SVC. A watermark is first permutated and each permutated character with its original order number is sequentially embedded at each frame of a frame group whose size is equal to length of the watermark. The robustness of temporal SVC is achieved by embedding differently permutated watermark at each frame group. On the one hand, the permutated character is repeatedly embedded only at ROI of base layer in each frame. Thereby the proposed scheme is robust to the spatial SVC. Robustness to the FGS is achieved by embedding the permutated character adaptively at low-middle frequency band of DCT result for the ROI. The proposed scheme is also robust to multimedia transcoding, because it may not give rise to a serious distortion beyond one caused by the spatial SVC, temporal SVC, and FGS. Through experiment with four typical videos, robustness of the watermark against the spatial SVC, temporal SVC, FGS is confirmed and robustness against to multimedia transcoding between MPEG-2 and MPEG-4 SVC is also verified. Watermarking techniques for transferring video contents to different multimedia services are in the first stage and continuous researches are necessary for applying in various applications. Considering various high-definition platforms as DVD, PC, and HDTV, algorithm security and real-time processing are required and researches for solving these are necessary.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11/N5231, Applications and requirements of scalable video coding N6830,Palma de Mallorca, Spain, Oct. 2004.

[2] ITU-T document, "Joint draft 10 of SVC amendment," Joint video team JVT-W201, JVT 23rd meeting, San Jose, USA, Apr. 2007.

[3] ISO/IEC JTC1/SC29/WG11, "Joint scalable video model (JSVM) 7.0 reference encoding algorithm description," N7556, Klagenfurt, Austria, July 2006.

[4] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. on Circuits and Systems for Video Tech*, Vol.11, No.3, pp. 301-317, Mar. 2001.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," *Proceedings of ICME'06*, Toronto, Canada, July 2006.

[6] A. Vetro, C. Christopoulos, and H. Sun. "Video transcoding architectures andtechniques: An overview," *IEEE Signal Processing Mag.*, Vol.20, pp. 18-29, Mar. 2003.

[7] J. Xin, C. W. Lin, and M. T. Sun, "Digital video transcoding," *Proceeding of IEEE*, Vol.93, No.1, pp. 84-97, Jan. 2005.

[8] P. Yin, M. Wu, and B. Lui, "Video transcoding by reducing spatial resolution," *Procedding of IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Vol.1, pp. 972-975, Oct. 2000.

[9] Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, Vol.66, No.3, pp. 283-301, May 1998.

[10] X. Kong, Y. Liu, and H. liu, "Adaptive video watermarking scheme," *IEEE Pacific-Rim Conference On Multimedia (PCM)*, LNCS 2195, pp. 933-939, Oct. Springer-Verlag 2001.

[11] Y. Wang and Al. Pearmain, "Blind MPEG-2 video watermarking robust against geometric attacks: A set of approaches in DCT domain," *IEEE Trans. on Image Processing*, Vol.15, No.6, pp. 1536-1543, June 2006.

Ji-Sun Yoon

She received a B.S. degree from Pusan University of Foreign Language and a M.S. degree in the Department of Computer Engineering from Pusan National University, Busan, Korea, in 2006 and 2008, respectively. Her research interests include multimedia security and watermarking.

Seong-Geun Kwon

He received a B.S., a M.S., and a Ph. D. degree in Electrical Engineering from Kyungpook National University, Korea in 1996, 1998, and 2002 respectively. He is currently in senior engineering in R&D Group, Mobile Communication Division, Samsung Electronics Co. His research interests include multimedia security, visual communication, and image processing.

Suk-Hwan Lee

He received a B.S., a M.S., and a Ph. D. degree in Electrical Engineering from Kyungpook National University, Korea in 1999, 2001, and 2004 respectively. He is currently an assistant professor in department of Information Security at Tongmyong University. His research interests include multimedia security, digital image processing, and computer graphics.

Yoon-Chul Song

He received his B.A from Kyung- nam University in 1988 and his M.S. degree from Chungang University in 1994 and completed the course of Ph.D. in business administration from Konkuk University in 2007. He is currently a Director General of Korea Association of Information & Telecommunication. He conducted five technological projects from Government in the field of internet appliance, next generation PC, speech & language information.

Min-Hwan Kim

He received his B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, in 1980, 1983, and 1988, respectively. He is currently a professor of the Dept. of Computer Engineering in Pusan National University, Korea. His research interests include intelligent surveillance system, multimedia information retrieval, color engineering, and computer vision.

Ki-Ryong Kwon

Ki-Ryong Kwon received the B.S., M.S., and Ph.D. degrees in electronics engineering from Kyungpook National University in 1986, 1990, and 1994 respectively. He worked at the Pusan University of Foreign Language form 1996-2006. In 2006, he joined the Pukyong National University, where he is currently a professor. His research interests are in the area of digital image processing, multimedia security and watermarking.