Private Image Region Protection and Reconstruction Scheme for Cloud-based Video Monitoring System using Transform Domain Wyner-Ziv Video Coding

Jongbin Park, *Member, IEEE*, Kyung-Won Kim, Jae-Won Moon, Seung-Woo Kum, Jong-Jin Jung,
Tae-Beom Lim, Myung-Hyun Yoon, and Byeungwoo Jeon, *Member, IEEE*

***Abstract*—Cloud-based video monitoring system has many advantages and interesting applications. However, it has a critical problem of invading privacy. For example, if cloud servers or transmission channels are attacked, their private data can be leaked to the attackers. If we use well-known encryption schemes such as AES or DES, the security lifetime will increase but it requires additional computational operations and algorithm modules. Therefore, we propose a novel video content protection scheme which can protect private event regions in a video sequence for cloud-based video monitoring services. When specific events have occurred or the high-resolution video needs to be stored, region adaptive transform domain Wyner-Ziv coding scheme is applied.**

***Index Terms*—Video Coding and Processing, Video Encryption, Transform Domain Wyner-Ziv, Cloud Monitoring**

I. Introduction

In this paper, we propose a novel video content protection scheme in a video sequence for cloud-based video monitoring services. The proposed scheme mainly uses a transform domain Wyner-Ziv (TDWZ) video coding for the content protection and fast encoding of the private event image regions. The TDWZ is a lossy distributed video coding (DVC) scheme which is developed based on the Slepian-Wolf and the Wyner-Ziv theorems published in 1970's [1], [2].

Cloud-based video monitoring system has many advantages such as computing resource sharing, storage sharing, easy system update, and mobile device accessibility [3]. It can also provide a lot of interesting applications such as young baby monitoring, video surveillance of home or office. However, if the number of cameras increases, the required transmission bandwidth also increases. Moreover, it has a critical problem of invading privacy. For example, if cloud servers or transmission channels are attacked, their private data will be leaked to the attackers. Some video encryption techniques can solve the problem through protecting the private image regions by full bit-stream encryption [4], partial bit-stream encryption [4], or perceptual video scrambling [5]. However, due to the massive data size of video, the full bit-stream encryption method [4] requires additional computational operation in encoding and decoding and the video scene information cannot be used at all without the encryption key. If we use the partial bit-stream encryption [4] or the perceptual video scrambling [5], not only faster processing is possible but also the scrambled image can be recognized by human although it is incomplete. However, their scrambled image quality is insufficient for video analysis and visual quality of the protected image regions cannot be controlled because they manipulate some critical bits of video bit-stream such as header bits or motion vectors [4], [5].

Therefore, we propose an alternative method without using well-known encryption schemes such as AES (Advanced Encryption Standard) or DES (Data Encryption Standard) [6]. All monitoring images are encoded with low quality so the subjective quality is degraded significantly for privacy protection. Despite low image quality, not only some major context extraction was possible, but also the amount of transmission bits can be reduced.

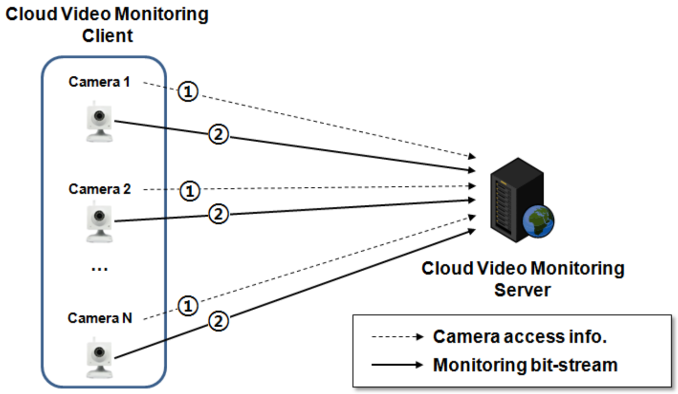


Fig. 1. Structure of the conventional cloud-based video monitoring system

J. Park (corresponding author), K.-W. Kim, J.-W. Moon, S.-W. Kum, J.-J. Jung, T.-B. Lim, and M.-H. Yoon are with the Korea Electronics Technology Institute (KETI), 9FL, Electronics Center, #11, World cup buk-ro 54-gil, Mapo-gu, Seoul, Korea, e-mail: {jpark, kwkim, jwmoon, swkum, tblim, and mhyoon}@keti.re.kr.

B. Jeon is with the School of Electronic and Electrical Engineering, Sungkyunkwan University, 300 Chunchun-dong, Jangan-gu, Suwon, Korea, e-mail: bjeon@skku.edu.

II. Cloud Based Video Monitoring System

Fig.1 shows the structure of conventional cloud-based video monitoring (CVM) system. There are two major parts: CVM client and CVM server. The CVM client typically has multiple cameras for monitoring. In order to acquire monitoring images, the client must send camera access information to the CVM server such as uniform resource identifier (uri), user name, password, or location information of cameras (①). After this camera registration step, CVM server controls the registered cameras remotely. The cameras take monitoring images or videos, then compresses and sends it the CVM servers (②). The server analyzes the transmitted images or videos using image processing algorithms to figure out the context information such as number of persons, moving objects, and various events. Here, context information is usually categorized according to location, identity, activity, and time [8].

III. Proposed Cloud-based Video Monitoring System

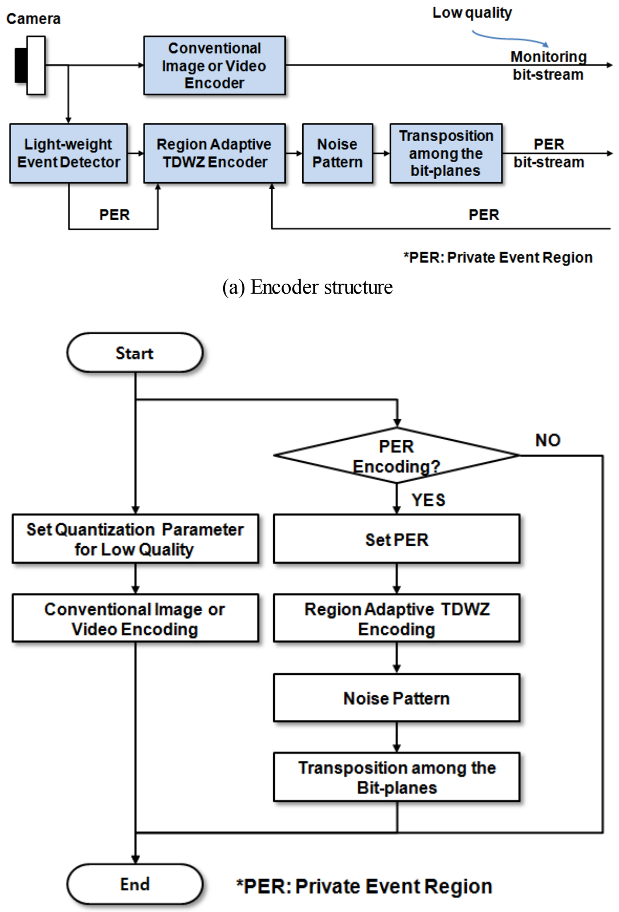
Despite these advantages, cloud-based video monitoring approach has a critical problem of potentially invading privacy. For example, if cloud servers or transmission channels are attacked, their private data can be leaked to the attackers. Moreover, if the camera access information (e.g. uri, user id, and password) is leaked, attacker might freely control camera and monitor it under user permission level. Therefore, a major requirement of our system is that the monitoring images should be protected from external attacks in transmission channel or cloud servers. To provide these capabilities, all monitoring images are encoded with very low quality by appropriate quantization parameter setting.

A. Encoding

Fig. 2 shows the proposed encoder for protection of private event region. Input original image is compressed by conventional image or video coding schemes such as JPEG, MPEG2, or H.264/AVC. The quantization parameters are selected for low quality encoding. It is configured by user. However, if some events are detected from "Light-weight Event Detector" in encoder or the high-resolution video needs to be stored, it is compressed by TDWZ encoder employing known bit filling scheme for region adaptive TDWZ coding as Fig. 3 [7]. This event area is called a private event region (PER) in this paper. The zero filling bits ratio r is defined as (1).

$$r = \frac{\text{the number of zero filling bits}}{\text{the length of message bits}} \qquad (1)$$

Most of the TDWZ bit-stream is composed with systematic parity bits of LDPCA [9] encoder. If proper side information is unavailable, it is difficult to restore the PER by only using the parity information except full systematic parity case. If the TDWZ decoder has full systematic parity bits, then it can easily decode the PER without side information [9]. On the other hand, if some attacker somehow estimates side information, the PER might be decoded through repeated attempts.


Camera
Conventional Image or Video Encoder
Low quality
Monitoring bit-stream
Light-weight Event Detector
Region Adaptive TDWZ Encoder
Noise Pattern
Transposition among the bit-planes
PER bit-stream
PER
PER
*PER: Private Event Region
Start
PER Encoding?
NO
YES
Set Quantization Parameter for Low Quality
Set PER
Conventional Image or Video Encoding
Region Adaptive TDWZ Encoding
Noise Pattern
Transposition among the Bit-planes
End
*PER: Private Event Region

(a) Encoder structure

(b) Encoding procedure

Fig. 2. Proposed encoder for protection of private event region

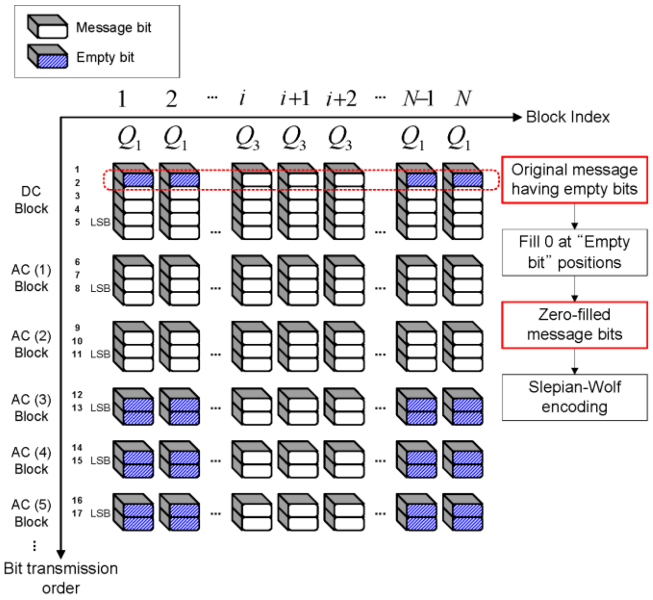


Fig. 3. Example of region adaptive TDWZ coding

To prevent these situations, noise pattern is added for the TDWZ parity bit-planes with bit error probability 0.5. The transposition (or permutation) process is carried out for all bit-planes after adding the noise. The transposition result can enhance security without well-known encryption process. The noise pattern and the transposition mapping table are randomly generated at CVM client, and its information is properly communicated with CVM server. Equation (2) and (3) are a mask image and quantization matrix for region adaptive TDWZ coding [7].

$$MASK_i = \begin{cases} 255, & PER \\ 0, & Non-PER \end{cases}, \; i \in \{1,2,...,N\} \tag{2}$$

where N is the number of 4×4 blocks in a picture and i is a 4×4 block index under processing.

$$Q_i = \begin{cases} Q_M, & PER \\ 0, & Non-PER \end{cases}$$

$$, \; i \in \{1,2,...,N\}, M \in \{1,2,...,8\} \tag{3}$$

where Q_M is quantization matrix index [7] which is selected in advance by user.

B. *Decoding*

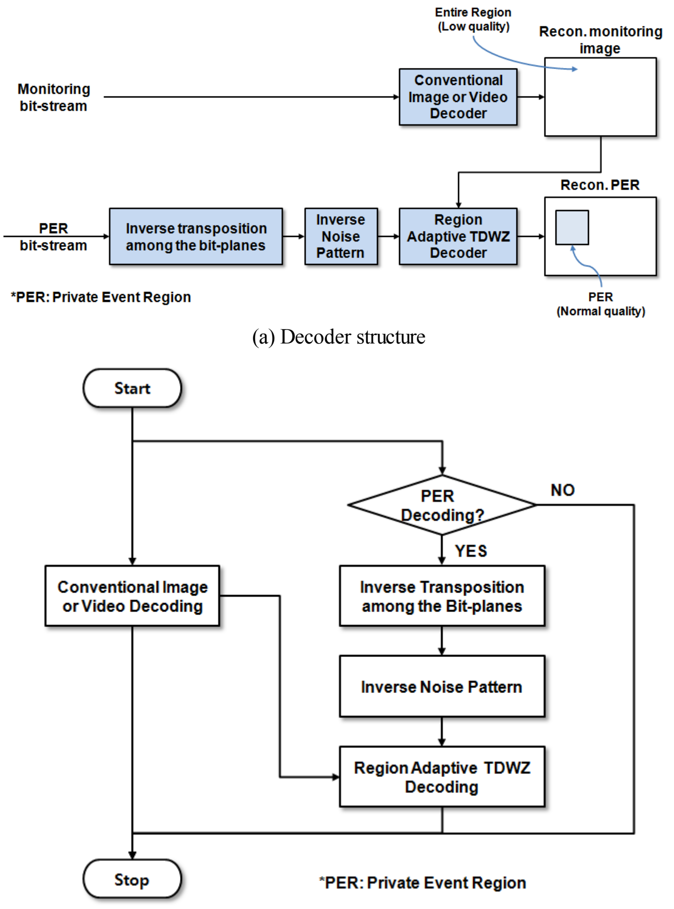
Fig. 4 shows the proposed decoder for the reconstruction. The conventional image or video decoder decodes the monitoring images. After inverse transposition and inverse noise pattern process, the TDWZ decoder reconstructs the PER, when the PER bit-stream is available. Decoded monitoring images are used as the side information of TDWZ decoder. In our system, TDWZ decoding overhead is not a significant problem because decoding process is not always necessary.

IV. Implementation And Results

The proposed CVM server was implemented using the *node.js* software platform with the *express* web application framework to provide services for multiple clients at the same time [10], [11]. Our image processing algorithms were implemented using the *OpenCV* library [12]. Simulations are performed with the proposed encoder and decoder in Fig. 2 and 4. The monitoring bit-stream is generated using MPEG-4 H.264/AVC intra coding (JM17.2) with baseline profile, and the quantization parameter is 50 for encoding in monitoring images. The PER are encoded using the region adaptive TDWZ coding scheme with quantization matrix number 8 [7]. Fig. 5 shows the normalized number of requested parity bits according to various bit error probabilities under different zero filling ratio r. The "r=1.0" means that all message bit-planes are zero filled. The PER ratio is a ratio between PER and entire image area. Fig. 6 shows the rate distortion results of video sequences in PER and entire picture area. Foreman with Siemens logo and Hall monitor sequences are used for test (148 pictures, 15Hz, QCIF). X-axis is a bit-rate [kbps@15Hz] and y-axis is an average PSNR (Peak signal-to-noise ratio) [dB]. As the area of PER becomes smaller, bit-rate decreases because zero filling bits are increased. Fig. 7 compares the subjective image quality between normal low-quality monitoring image and its enhanced version with reconstructed PER.

V. Conclusion

In this paper, we proposed a novel video content protection scheme which could protect private event regions in a video sequence for cloud-based video monitoring services without full encryption. Despite low quality of decoded images, not only some major context extraction was possible, but also the amount of transmission bits could be reduced. Future studies will be required to improve rate-distortion performance for PER.


Entire Region
(Low quality)
Recon. monitoring image
Monitoring bit-stream
Conventional Image or Video Decoder
Recon. PER
PER bit-stream
Inverse transposition among the bit-planes
Inverse Noise Pattern
Region Adaptive TDWZ Decoder
*PER: Private Event Region
PER
(Normal quality)
Start
PER Decoding?
NO
YES
Conventional Image or Video Decoding
Inverse Transposition among the Bit-planes
Inverse Noise Pattern
Region Adaptive TDWZ Decoding
Stop
*PER: Private Event Region

(a) Decoder structure

(b) Decoding procedure

Fig. 4. Proposed decoder for the reconstruction

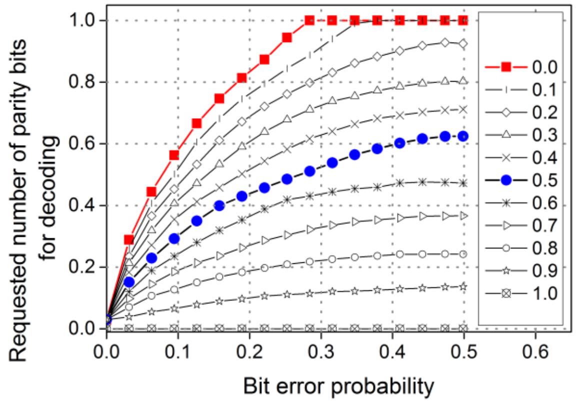


Fig. 5. The normalized number of requested parity bits (y-axis) according to various bit error probabilities under different zero filling ratio

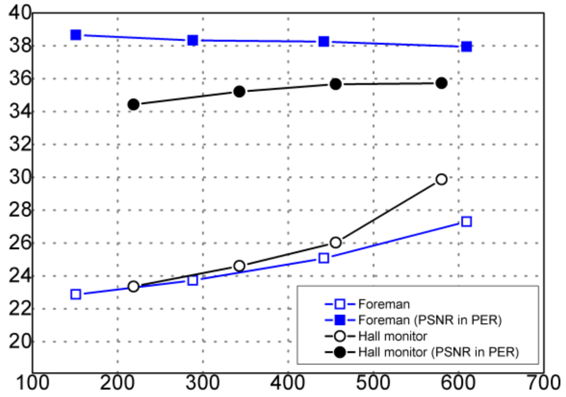


Fig. 6. Rate distortion performance; x-axis is bit-rate [kbps@15hz], y-axis is an average PSNR [dB].

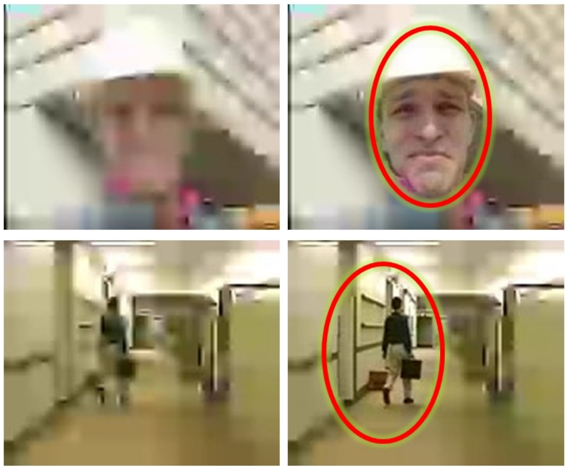


Fig. 7. Example of PER reconstruction of frame #25; the red circles means the detected PER and the PER ratio is 0.2.

REFERENCES

[1] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol.19, no.4, pp.471–480, Jan.1973.

[2] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility,” *Future Generation computer systems*, vol. 25(6), pp.599-616, June 2009.

[4] J. Shah and V. Saxena, “Video Encryption: A Survey,” *IJCSI International Journal of Computer Science Issues*, vol. 8, pp.525-533, Mar. 2011.

[5] J. Ahn, H.-J. Shim, B. Jeon, and I. Choi, “Digital Video Scrambling Method Using Intra Prediction Mode,” *Lecture Notes in Computer Science*, vol. 3333, pp.386-393, 2005.

[6] Bose, Ranjan, *Information theory, coding and cryptography*, Tata McGraw-Hill Education, 2002.

[7] J. Park, B. Jeon, D. Wang, and A. Vincent, “Wyner-Ziv video coding with region adaptive quantization and progressive channel noise modeling,” in *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2009. (BMSB 2009)*, May 2009, pp. 1–6.

[8] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, “Towards a better understanding of context and context-awareness,” *Lecture Notes in Computer Science*, vol. 1707, pp. 304-307, 1999.

[9] D. Varodayan, A. Aaron, and B. Girod, “Rate-adaptive codes for distributed source coding,” *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, no. 11, pp.3123-3130, Nov. 2006.

[10] *Node.js software platform*, Available: http://www.nodejs.org/

[11] *Express web application framework for node*, Available: http://expressjs.com/

[12] *Opencv image processing library*, Available: http://opencv.org/