Filtering Based ROI Coding Using Dynamic Range Compression and Updating Source Picture Filter

KAZUHITO SAKOMIZU* Member, TAKAO ONOYE† Non-member

(Received July 19, 2017, revised October 08, 2017)

Abstract: The region of interest (ROI) coding, which compresses background region more strongly than foreground region, is valuable technology in restricted communication environments such as limited bandwidth and/or storage capacity. This paper proposes a filtering scheme which reduces the amount of bits by filtering background region in order to support efficient ROI coding on ordinary devices without equipping ROI coding functionality. Specifically, dynamic range compression (DRC) filter and updating source picture (USP) filter are added to a low-pass filter, which has been used for conventional filtering scheme. The DRC filter compresses dynamic range of all pixels by down-scaling. The USP filter measures distance between current and previous picture by block basis, and then overwrites the block of the current picture when the distance is smaller than predetermined threshold. Experimental results show that the proposed filtering scheme reduces bit-rate by half compared to the conventional low-pass filter in the condition with the PSNR-Y of background region kept to 27 dB. Compared to uniform quality coding which does not incur the filters, there is the case that the proposed filtering scheme reduces bit-rate by 58 %. These results promise to cut down on cost of communication and storage.

Keywords: Region of interest, Filtering scheme, Video coding

1. Introduction

Recently video surveillance systems become popular rapidly amid growing concern about security. Development of image sensors is also active and omnidirectional network cameras have been released. As the number of network cameras increases, the cost for transmitting and recording data often becomes a problem because the amount of video data is larger than that of audio data generally. When communication environments is restricted such as limited bandwidth and/or storage capacity is not sufficient, we have to take measures such as shortening recording time, decreasing frame rate and/or narrowing view angle and decreasing resolution. However these are not ideal solution because these measures impair security performance. To address this problem, region of interest (ROI) coding have attracted attention[1][2].

The ROI coding is technology to ensure picture quality in important region on limited bandwidth by allocating more bit amount to ROI than background region [3]. This idea has developed to improve subjective picture quality mainly for video telephones and video conferences. New researches such as combining with saliency and combining with new video coding standards are active in recent years [4]~[10].

In case of a part of video surveillance system, it is sufficient if a security officer can identify who are in field angle. For instance, face regions must be sent finely to identify the individual adequately, while in other regions rough picture quality which is enough to understand monitoring location is allowed. In this situation, ROI coding can reduce the amount of bits in background region, and then bit-rate reduction is realized[1]. Recently computer vision technique such as human / face detection has been developed to a practical level [11]~[13], and commercialization is also in progress[14]. We expect that combination of computer vision technique and ROI coding resolves cost problem of video surveillance systems.

The simplest method for ROI coding is quantization parameter (QP) control scheme, which uses built-in function for quality control [5]~[7][15]~[18]. For example, the H.264 / MPEG4 AVC[19], which is mainstream coding standard in video surveillance systems, has syntax to control quality of a macro block (MB). By using this syntax and assigning lower QP meaning narrow quantization step to ROI than background region, ROI coding is realized. This scheme has the advantage that coding efficiency of background region is high, while there is restriction that we have to select video coding standard and implementation which have built-in function to control quality of a MB[20]. In video surveillance systems, hardware accelerator is often used for video encoding because video encoding is commonly known as heavy process. However ordinary encoders may not support the built-in function, and therefore this scheme is not feasible. In addition, this scheme

* Smart Network Technologies R&D Department, Corporate Research & Development Center, Information & Technologies Planning Group, Oki Electric Industry Co., Ltd.

2-6-8 Bingomachi Chuo-ku, Osaka, 541-0051, Japan
sakomizu688@oki.com

† Department of Information Systems Engineering, Osaka University
1-1 Yamadaoka, Suita 565-0871, Japan
onoye@ist.osaka-u.ac.jp

has also a disadvantage that range of trade-off between picture quality and bit-rate is limited within a used video coding standard. In case of the H.264 / MPEG4 AVC, range of the quality parameter is defined as value from 0 to 51. There are other researches for ROI coding by utilizing built-in functions of codecs or by alteration of the codec, such as controlling the number of candidate coding modes[21], suppressing high frequency components[2] and making use of the ROI scalability[9] [22] of the scalable video coding extension of the standards[23] [24], but these technologies also have the same restrictions as above.

Another method for ROI coding is filtering scheme which separates background region of a picture by image processing to reduce the amount of bits before encoding [1] [4] [15] [20] [25]. For instance, a low-pass filter such as the Gaussian filter can reduce the amount of bits in background region. In general, a video encoder compresses the amount of bits by removing spatial and temporal redundancy, and therefore the amount of bits is expected to be reduced by the low-pass filter in regions where redundancy is increased, in other words energy of the AC component is reduced. As a similar research, although there is a drawback that post-processing after decoding is essential, it is studied that transmitting only the ROI and reconstructing the background region from the already transmitted picture in the post-processing[26]. An advantage of filtering scheme is that it can also be used in the situation where using video coding implementation does not support ROI coding, because filters are independent from the codec. It also eliminates the need for a costly adaption to new coding standards and implementations. Moreover this scheme is also an advantage that range of trade-off between picture quality and bit-rate is not limited. This paper considers these features to be important, and discusses the filtering system in the following.

Filtering scheme has the advantages as described above, while coding efficiency is low compared to the QP control scheme. This is because discussion in terms of filter design to reduce the amount of bits of background region efficiently for state-of-the-art standards was not seriously done since the H.264 / MPEG4 AVC supports syntax for quality control of a MB. For instance, in case that the low-pass filter is used as a prefilter before encoding, high-frequency components are suppressed. This is workable to reduce the amount of bits. As for the QP control scheme, the amount of bits of background region is reduced by controlling degree of entropy reduction in quantization process and possibility that skip mode is selected in prediction process, because both of quantization step and basis of coding mode selection also depend on the QP. Specifically, ratio of skip mode MBs increases in background region. However, in a filtering scheme, it is assumed that identical QP is assigned to whole regions, and therefore it is impossible to obtain the effect of reducing the amount of bits by above two controls. Especially absence of controlling skip mode selection basis has a profound effect because skip mode contributes coding efficiency so much when there is little motion. Since measures for these differences are not sufficient, coding ef-

ficiency of a filtering scheme in background region becomes lower than the QP control scheme.

This paper proposes a new filtering scheme that consists of dynamic range compression (DRC) filter and updating source picture (USP) filter in addition to a conventional low-pass filter. The DRC filter compresses dynamic range of all pixels by down-scaling. The filter realizes the function which varies the amount of bits by changing dynamic range, which is similar to entropy, of the source signal in QP control scheme. Down-scaled pixels are restored to original scale at a postfilter after decoding so as to avoid matter that dynamic range is different depending on region when watching recorded video. The USP filter measures distance between current and previous picture by block basis, and then overwrites the block of the current picture to induce skip mode when that distance is smaller than predetermined threshold. This filter realizes the function by which skip mode is likely to be selected in background region. The proposed filtering scheme reduces the amount of bits of background region efficiently by combination of these filters. If filtering scheme can sufficiently reduce the amount of bits, it is expected to dramatically increase the applicable range of ROI coding, since filtering scheme can be introduced on ordinary devices without equipping ROI coding functionality.

This paper is organized as follow. Section 2 describes a proposed system for ROI coding using the proposed filtering scheme. Section 3 explains detail of proposed filters such as the DRC filter and the USP filter, which is main contribution of this paper. Section 4 presents performance of the proposed filtering scheme. Finally, Sec. 5 concludes this paper.

2. The proposed system for ROI coding

The proposed system filters background region in a source picture based on ROI information detected by object detection engine, and encodes the filtered picture. Figure 1 is overall image of the proposed system. Detail is shown in following paragraphs.

The object detection engine in Fig. 1 searches object regions from the input picture, and outputs rectangular coordinates of object regions. In this paper, face and human are detected as the object. Several object detection methods have been proposed as previously described. In this paper, a detection method using the HOG feature[11] is employed for human detection, and a detection method using the Haar-like feature[12] [13] is employed for face detection,

The object detection engine outputs coordinates of rectangular areas including a human or a face by the pixel. In contrast, the encoder divides a source picture into MBs, and encodes the MBs. When there is a boundary between ROI and background region in a MB, energy of AC components is rather increased because the boundary causes a sharp edge. For this problem, it is proposed to prevent an edge due to region boundary in a MB by introducing multiple levels of background quality and changing quality smoothly[20]. In our paper, the system adopts to adjust pixel coordinates

Figure 1: Overall of the proposed system for ROI coding.

Figure 2: ROI coordinate adjustment

of the ROI from the object detection engine by separating ROI and other regions by boundary of MB on the ROI coordinate adjustment in Fig. 1, because sharp edge due to region boundary has a positive side effect that it attracts attention of a viewer. Figure 2 shows an image of adjustment process. This process is formulated below.

$$X_{L,T} = \lfloor x_{L,T}/N \rfloor \times N, \quad (1)$$

$$Y_{L,T} = \lfloor y_{L,T}/N \rfloor \times N, \quad (2)$$

$$X_{R,B} = \lfloor (x_{R,B} + N - 1)/N \rfloor \times N - 1, \quad (3)$$

$$Y_{R,B} = \lfloor (y_{R,B} + N - 1)/N \rfloor \times N - 1, \quad (4)$$

where $x_{L,T}$ and $y_{L,T}$ are coordinates of top left pixel of ROI. $x_{R,B}$ and $y_{R,B}$ are those of bottom right pixel. $X_{L,T}$ and $Y_{L,T}$ are those of top left pixel of adjusted ROI. $X_{R,B}$ and $Y_{R,B}$ are those of bottom right pixel. N is size of MB. In this paper, N is set to 16 because the H.264 / MPEG4 AVC is assumed as described later.

The adjusted coordinates are transmitted to the pre-filter. The pre-filter reduces the amount of bits of regions belonging to MBs determined as background. Detail of design about the pre-filter is described in following section.

Finally a filtered picture, of which the amount of bits in background region is reduced, is input to the video encoder. Although there is no particular limitation on the video encoder for filtering scheme, the H.264 / MPEG4 AVC encoder Baseline Profile, which is a mainstream standard currently in the surveillance camera market, is assumed in this

Figure 3: Block diagram of pre-filter

paper. Therefore assumed bit depth of pictures is 8, in other words assumed range of pixel value is from 0 to 255.

In the decoder side, the H.264 / MPEG4 AVC decoder receives and decodes the transmitted stream. Then the decoded picture is input into the post-filter. Detail of design about the post-filter is described in following section. Output of the post-filter is that of the proposed decoder system.

3. Proposed filtering scheme

This section describes detail of the pre-filter and post-filter those are the core of the proposed filtering scheme. First, a structure of pre-filter is shown in Fig 3. The pre-filter is composed of a low-pass filter, the DRC filter and the USP filter. Purpose and behavior of each module are described in following subsections.

3.1 Low-pass filter Low-pass filter has been conventionally used in filtering scheme[20] [22]. This filter reduces the amount of bits in background region by suppressing the high-frequency components. In this paper the general Gaussian filter is used as this filter. Definition of the used Gaussian filter is following.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (5)$$

$$\sigma = 0.3 \times \left(\frac{k}{2} - 1\right) + 0.8, \quad (6)$$

Figure 4: A flow chart of the DRC filter

where σ is the variance parameter of the Gaussian filter, and the value is obtained from the kernel size k .

3.2 DRC filter The QP control scheme varies the amount of bits by changing dynamic range, similar to entropy, of the source signal. In particular, quantization step of background region is larger than that of ROI. The DRC filter replicates the capability on a filtering scheme, in which it is assumed that quantization step is same between ROI and background region. This filter compresses dynamic range by performing the down-scaling processing for all pixels in pixel domain. Entropy of picture signal becomes low by compressing dynamic range, and therefore the amount of bits is also reduced.

Figure 4 shows behavior of the DRC filter. First subtraction 128 from the pixel value p_i is performed to make center value of scaling process 128. Then the value is down-scaled by using a predetermined scaling factor R_e , and is rounded to integer value. A range of R_e is from 0 to 1.0. Finally compressed value p_o is produced by adding 128 to the integer value. By using this filter, background region becomes dim since dynamic range of the region is narrow.

By the way, since center value of scaling process is set to 128, following effects are expected.

- Under an assumption that scene conditions can not be specified, pixel value of input pictures is distributed around 128. In particular, chroma components have this tendency. Therefore setting the center value to 128 prevents that residual entropy increases wastefully because it is possible to minimize the expected amplitude of residual pixel value at switching point between ROI and background region.
- When playing without the post-filter, a sense of discomfort about color is small subjectively. For instance, if center value of scaling process is set to 0, color is biased towards green such as Fig. 5(a), because the prefilter processes pictures in YUV space as well as many video codecs. On the other hand, when scaling centering on 128, color closes to natural as shown in Fig. 5(b).

3.3 USP filter In the QP control scheme, skip mode is likely to be selected in background region. This has influence on allocation of the amount of bits. The USP filter realizes the similar function in a filtering scheme. That measures distance between current and previous picture by block basis, and then overwrites the block of the current picture to induce skip mode when the distance is smaller than predetermined threshold,

Figure 6 shows behavior of the USP filter. In this filter, it is determined whether the overwriting process is performed or not by threshold processing of following.

$$\text{MAD}(B_t(m, n), B_{t-1}(m, n)) < R_e \times T, \quad (7)$$

where $B_t(m, n)$ represents a block composed of 16×16 pixels of input pictures to be filtered, and $B_{t-1}(m, n)$ represents the co-location block of $B_t(m, n)$ in the filtered picture transmitted to the encoder just before time. t is picture index of the block such as timestamp, and m, n represents position of the block in the picture. MAD is a function to calculate the mean absolute difference (MAD) of the two blocks. T is the

(a) Center value is set to 0

(b) Center value is set to 128

Figure 5: Difference of center value in scaling process

Figure 6: A flow chart of the USP filter

parameter which is used to calculate the threshold. In the proposed filtering scheme, dynamic range of background region is compressed by the DRC filter, and the lower scaling factor R_e becomes, the lower the MAD of $B_t(m, n)$ and $B_{t-1}(m, n)$ are. Therefore multiplying R_e to the threshold T is performed to determine whether the overwriting process is performed or not in a stable manner regardless of R_e . If the condition Eq.7 is true, update processing to $B_t(m, n)$ is applied by copying the filtered block $B_{t-1}(m, n)$. This process encourages that $B_t(m, n)$ is encoded by skip mode in the H.264 / MPEG4 AVC encoder.

3.4 The post-filter In the proposed system, it is necessary to decompress dynamic range to the original scale at the decoding side, since the pre-filter generates the picture of which dynamic range in background region is compressed as described above. The post-filter is a module for this function. The post-filter performs inverse operation of the DRC filter to return the compressed pixel value the original scale. Even without this filter, it is possible to view a decoded picture, while color of background region becomes dim.

Figure 7 shows behavior of the post-filter. q_i and q_o represents pixel value. The basic behavior is the same as the DRC filter. The difference is that the scaling factor is inverse number of R_e , $R_d = 1/R_e$, and that cropping process is added to the end.

As described above, q_i is determined by rounding off in calculation process of p_o and by deteriorating mixing by quantization noise on the video encoder. Therefore, the result of the inverse operation using R_e is likely to take a value less than 0 or a value of 256 or more. However, since it is known that range of p_i , which is pixel value before compression, is $[0, 255]$, it is reasonable to keep q_o within $[0, 255]$. Therefore the negative pixel values are shifted to 0, and 256 or more pixel values are shifted to 255 in the cropping process. In this paper, range of pixel value is assumed as $[0, 255]$ in a typical pattern, while it is necessary to adjust the range as appropriate in an application such as high dynamic range compression.

4. Simulation

In this section, experiments are performed in order to evaluate how close the proposed filtering schemes show perfor-

Figure 7: A flow chart of the post-filter

mance to the QP control scheme. The comparison is conducted in terms of the amount of reduced bits by the ROI codings and picture quality of decoded pictures. We also compare it with the Gaussian filter as a conventional filter scheme.

4.1 Evaluation method To evaluate performance of the ROI codings, the graph of which horizontal axis is reduction rate, and vertical axis is picture quality of decoded pictures in background region is drawn. In this paper, reduction rate Q is defined by following equation.

$$Q = \frac{F_B - F_R}{F_B}, \quad (8)$$

where F_B is a reference bit-rate calculated by uniform quality coding which does not incur the filters. F_R is a bit-rate of ROI coding. As picture quality of decoded pictures, the PSNR-Y and the SSIM[27] are used. The PSNR-Y is calculated by using pixels belonging to background region because the ROI coding schemes of this paper maintain picture quality of ROI while reducing the amount of bits. It is possible to evaluate pixels belonging only to background region since the deterioration is measured for each pixels independently as for the PSNR-Y.

4.2 Test sequences Test sequences used to draw the graph are Hall_monitor created by Rensselaer Polytechnic Institute and Bowling created by FUJITSU LABORATORIES LTD. We select test sequences which are taken by the fixed camera since we assume surveillance systems. Resolution of these sequences is 352×288 , and that frame rate is 30 fps.

The Hall_monitor is a video sequence that monitors the corridor and pedestrians appear, and therefore we use a human detection detector using the HOG feature for the object detection engine. On the other hand, since the Bowling is a video sequence in which a face toward the camera appears, a face detector using the Haar-like feature is used for the object detection engine.

4.3 Parameter configuration QP of ROI is fixed to 26. In the QP control scheme, QP of background region is

Figure 8: Relationship between reduction rate and picture quality of decoded pictures in case of the Hall_monitor

Figure 9: Relationship between reduction rate and picture quality of decoded pictures in case of the Bowling

varied from 30 to 51 to reduce the amount of bits of background region. In filtering schemes, QP of both ROI and background region is fixed to 26, and the amount of bits is reduced by the filters.

In the filtering schemes, there are three parameters, the kernel size k of the Gaussian filter, the scaling factor R_e of the DRC filter, and the parameter T of the USP filter. A number of combinations are assumed for these parameters. Therefore we perform encoding and decoding with multiple combinations of these parameters, and envelope curves are drawn to collected data in terms of the reduction rate and picture quality of decoded pictures in background region.

4.4 Results Figure 8 and 9 show simulation results of relationship between reduction rate and picture quality of

decoded pictures. When compared to the proposed filtering scheme and the conventional filtering scheme, the proposed filtering scheme reduces the amount of bits more efficiently than the conventional filtering scheme does at all the points. For example, in case of the Hall_monitor, when the PSNR-Y is approximately 27 dB, reduction rate of the conventional filtering scheme is approximately 22 %, while that of the proposed filtering scheme is about 58 %. In other words, on the basis of bit-rate of the conventional filtering scheme, the proposed filtering scheme can reduce bit-rate to half roughly while maintaining picture quality. On the other hand, comparison of the proposed filtering scheme and the QP control scheme shows that the proposed filtering scheme is slightly inferior than the QP control scheme. When the

(a) Uniform quality coding

(b) QP control scheme (Reduction rate is about 60 %)

(c) Conventional filtering scheme (Reduction rate is about 22 %)

(d) Proposed filtering scheme (Reduction rate is about 58 %)

Figure 10: Comparison of a decoded picture in case of the Hall_monitor when the PSNR-Y of each decoded pictures are approximately 27 dB.

PSNR-Y is approximately 27 dB, the reduction rate of the QP control scheme is about 60 %, and difference between two approaches is about 2 %. These tendency are the same in case of the Bowling.

Even when comparing in terms of the SSIM, the proposed filtering scheme is superior to the conventional filtering scheme and inferior to the QP control scheme as in the terms of the PSNR-Y. Focusing on low reduction rate range in the graph, the difference between the proposed filtering scheme and the QP control scheme shrinks. It indicates that the proposed filtering scheme can achieve almost the same performance as the QP control scheme.

Figure 10 and 11 show comparison of decoded pictures when performing the uniform quality coding (a), the QP control scheme (b), the conventional filtering scheme (c) and the proposed filtering scheme (d). The PSNR-Y of each decoded pictures are approximately 27 dB. Comparing ROI, each ROI coding schemes can achieve fine picture quality enough to identify the individual. In background

region, tendency of the degradation between each scheme is different although the same picture quality is achieved in terms of the PSNR-Y. For the QP control scheme, since the amount of bits is reduced by coarse quantization, block noise is conspicuous. On the other hand, the conventional filtering scheme reduces the amount of bits just by the low-pass filter, and therefore block noise is not observed while blur is observed over the picture instead. In case of the proposed filtering scheme, tendency is similar to the QP control scheme. The QP control scheme is better than the proposed filtering scheme with respect to reproducibility of color, and the proposed filtering scheme is fine with respect to reproducibility of detail such as edges.

5. Conclusion

In this paper, a new filtering scheme is proposed to reduce the amount of bits in background region efficiently. The proposed filtering scheme is designed at the thought of the way that the QP control scheme reduces the amount of bits.

Figure 11: Comparison of a decoded picture in case of the Bowling when the PSNR-Y of each decoded pictures are approximately 27 dB.

It is composed of the DRC filter and the USP filter in addition to the conventional low-pass filter. The simulation results show that the proposed filtering scheme can reduce bit-rate to half roughly while maintaining picture quality when the PSNR-Y is approximately 27 dB. Compared to the uniform quality coding, 58 % of bit-rate can be reduced on the simulation in this paper, and we expect a significant reduction in storage cost and communication cost of a wireless environment.

References

- [1] T. Hata, N. Kuwahara, T. Nozawa, D. Schwenke and A. Vetro, "Surveillance System with Object-aware Video Transcoder", *IEEE 7th Workshop on Multimedia Signal Processing*, 2005, pp.1-4, 2005. DOI: 10.1109/MMSP.2005.248636
- [2] Y. Y. Zheng, X. Tian and Y. W. Chen, "Adaptive Frequency Coefficient Suppression for ROI-based H.264/AVC Video Coding", *IEEE International Conference on Networking, Sensing and Control*, 2008, pp.714-718, 2008. DOI: 10.1109/ICNSC.2008.4525309
- [3] S. Daly, K. Matthews and J. R. Corbera, "Face-based Visually-optimized Image Sequence Coding", *Proceedings 1998 International Conference on Image Processing, 1998. ICIP98*, Vol.3, pp.443-447, 1998. DOI: 10.1109/ICIP.1998.727233
- [4] L. Itti, "Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention", *IEEE Transactions on Image Processing*, Vol.13, No.10, pp.1304-1318, 2004. DOI: 10.1109/TIP.2004.834657
- [5] Z. Li, S. Qin and L. Itti, "Visual Attention Guided Bit Allocation in Video Compression", *Image and Vision Computing*, Vol.29, No.1, pp.1-14, 2011. DOI: 10.1016/j.imavis.2010.07.001
- [6] H. Hadizadeh and I. V. Bajic, "Saliency-aware Video Compression", *IEEE Transactions on Image Processing*, Vol.23,

- No.1, pp.19-33, 2014. DOI: 10.1109/TIP.2013.2282897
- [7] Z. Xu, M. Yu, S. Fang and S. Xu, "A New Saliency Based Video Coding Method with HEVC", *Proceedings of the 2015 International Conference on Intelligent Systems Research and Mechatronics Engineering*, pp.672-679, 2015. DOI: 10.2991/isrme-15.2015.135
- [8] B. Boyadjis, C. Bergeron, B. P. Popescu and F. Dufaux, "Background Simplification for ROI-oriented Low Bitrate Video Coding", *IEEE 18th International Workshop on Multimedia Signal Processing, 2016. MMSP-16*, No.1, pp.1-6, 2016. DOI: 10.1109/MMSP.2016.7813365
- [9] D. H. Lee, B. Jin, J. Ryu and N. I. Cho, "A Scalable Coding Based on Structure-preserving Filter and Saliency-ordered Transmission", *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2016. APSIPA-16*, pp.1-4, 2016. DOI: 10.1109/APSIPA.2016.7820851
- [10] L. Zhang, J. Chen and B. Qiu, "Region-of-interest Coding Based on Saliency Detection and Directional Wavelet for Remote Sensing Images", *IEEE Geoscience and Remote Sensing Letters*, Vol.14, No.1, pp.23-27, 2017. DOI: 10.1109/LGRS.2016.2623670
- [11] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR-05*, Vol.1, pp.886-893, 2005. DOI: 10.1109/CVPR.2005.177
- [12] P. Viola and M. Jones, "Rapid Object Detection Using A Boosted Cascade of Simple Features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR-01*, Vol.1, pp.1-511-I-518, 2001. DOI: 10.1109/CVPR.2001.990517
- [13] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", *Proceedings. International Conference on Image Processing, 2002. ICIP-02*, Vol.1, pp.1-900-I-903, 2002. DOI: 10.1109/ICIP.2002.1038171
- [14] C. Zhang and Z. Zhang, "A Survey of Recent Advances in Face Detection", *Microsoft Research*, 2010. DOI: 10.1.1.167.5270
- [15] M. J. Chen, M. C. Chi, C. T. Hsu and J. W. Chen, "ROI Video Coding Based on H.263+ with Robust Skin-color Detection Technique", *IEEE Transactions on Consumer Electronics*, Vol.49, No.3, pp.724-730, 2003. DOI: 10.1109/TCE.2003.1233810
- [16] Y. Liu, Z. Li, Y. Soh and M. Loke, "Conversational Video Communication of H.264/AVC with Region-of-interest Concern", *International Conference on Image Processing, 2006. ICIP-06*, pp.3129-3132, 2006. DOI: 10.1109/ICIP.2006.312936
- [17] Y. Sun, I. Ahmad, D. Li and Y. Q. Zhang, "Region-based Rate Control and Bit Allocation for Wireless Video Transmission", *IEEE Transactions on Multimedia*, Vol.8, No.1, pp.1-10, 2006. DOI: 10.1109/TMM.2005.861296
- [18] V. Sanchez, F. A. Llinas, and R. Vanam and J. B. Rapesta, "Rate Control for Lossless Region of Interest Coding in HEVC Intra-coding with Applications to Digital Pathology Images", *IEEE International Conference on Acoustics, Speech and Signal Processing, 2015. ICASSP-15*, pp.1250-1254, 2015. DOI: 10.1109/ICASSP.2015.7178170
- [19] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of The H. 264/AVC Video Coding Standard", *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol.13, No.7, pp.560-576, 2003. DOI: 10.1109/TCSVT.2003.815165
- [20] L. S. Karlsson and M. Sjöström, "Improved ROI Video Coding Using Variable Gaussian Pre-filters and Variance in Intensity", *IEEE International Conference on Image Processing, 2005. ICIP-05*, pp.560-576, 2005. DOI: 10.1109/ICIP.2005.1530054
- [21] Y. Liu, Z. G. Li and Y. C. Soh, "Region-of-interest Based Resource Allocation for Conversational Video Communication of H.264/AVC", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.18, No.1, pp.134-139, 2008. DOI: 10.1109/TCSVT.2007.913754
- [22] D. Grois and O. Hadar, "Complexity-aware Adaptive Bit-rate Control with Dynamic ROI Pre-processing for Scalable Video Coding", *IEEE International Conference on Multimedia and Expo, 2011*, pp.1-4, 2011. DOI: 10.1109/ICME.2011.6012110
- [23] C. H. Wu, C. W. Lo and J. F. Wang, "Overview of The Scalable Video Coding Extension of The H. 264/AVC Standard", *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol.17, No.9, pp.1103-1120, 2007. DOI: 10.1109/TCSVT.2007.905532
- [24] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramanian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.26, No.1, pp.20-34, 2016. DOI: 10.1109/TCSVT.2015.2461951
- [25] L. S. Karlsson, M. Sjöström and R. Olsson, "Spatio-temporal Filter for ROI Video Coding", *European Signal Processing Conference, 2006*.
- [26] H. Meuel, M. Munderloh, F. Kluger and J. Ostermann, "Codec Independent Region of Interest Video Coding Using A Joint Pre- and Postprocessing Framework", *IEEE International Conference on Multimedia and Expo, 2016. ICME-16*, pp.1-6, 2016. DOI: 10.1109/ICME.2016.7552876
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: from Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, Vol.13, No.4, pp.600-612, 2004. DOI: 10.1109/TIP.2003.819861

Kazuhito Sakomizu (Member) received the bachelor and master degrees from Ritsumeikan University, Japan, in 2007 and 2009, respectively. He joined the Corporate Research & Development Center, Oki Electric Industry Co., Ltd, Osaka, Japan. His research interests include video processing and coding, and multimedia transmission.

Takao Onoye (Non-member) received the B.E. and M.E. degrees in electronic engineering, and the Dr.Eng. degree in information systems engineering from Osaka University, Osaka, Japan, in 1991, 1993, and 1997, respectively. He is currently a Professor with the Department of Information Systems Engineering, Osaka University. His research interests include media-centric lowpower architecture and its SoC implementation.