

Improving Picture Quality for Regions of Interest

Soon-kak Kwon

Dept. of Computer Software Engineering, Dongeui University, Korea

Abstract:- In order to improve the picture quality, the region of interest within video sequence should be handled differently to regions which are of less interest. Generally the region of interest is coded with smaller quantization step-size to give higher picture quality, compared with the remainder of the image. However the abrupt change picture quality variation on the boundary of regions of interest can cause the some degradation of the overall subjective picture quality. This paper presents a method where the local picture quality is varied smoothly between the region of interest and the remainder of the image. We first classify the type of subjectively region of interest which can be determined by using motion vectors estimated in the coding process. Then the region of interest is coded by decreasing the quantization step-size according to a gradual linear change from the other regions within a video sequence.

Keywords:- ROI, Quantization, Subjective picture quality

I. INTRODUCTION

Various multimedia services, such as pictures and videos, have been supplied due to the recent development in digital signal processing, storage media and device, and transmission technologies including the wide distribution of computers and Internet. In particular, the applications of video services have been significantly increased due to the development of coding technology for large volumes of video data. The video coding standards for supplying various compression bitrates to various communication networks and its terminals were established as MPEG-1, 2, 4, H.261, H.262, H.263, and H.264 [1].

The large compressions required by video coding can be achieved through appropriately transforming the video data to concentrate the energy into a few coefficients, followed by a quantization algorithm to reduce the volume of data, hence bitrate. The two main classes of quantization algorithm are scalar and vector quantizers [2-5]. The MPEG-2 coding method uses a scalar quantizer with quantization step-sizes ranging from 1 to 32 [6]. The recently established H.264 and MPEG-4 Part 10 use a total of 52 quantization step-sizes [1-3]. Most of the conventional research into scalar quantizers [7,8] has focused on the relationship between the quantization step-size and the bitrate, where the primary role of the quantizer is to control the data volume. Quantization causes certain loss due to the restriction of the floating point coefficients to a series of discrete values. In particular, a large quantization step-size can reduce the large data volume to small size but with a consequent decrease in picture quality. Therefore the quantization step-size should be determined at a range to give the best picture quality subject to the restricted bitrate and the displayed resolution. The different resolutions utilized by various multimedia devices require that the quantization step-size can be selected in an adaptable manner by considering the picture resolution and the subjective picture quality for the regions of interest.

Several methods have been researched for coding based on regions of interest (ROI). ROI coding can be characterized according to ROI extraction method, quantization and bitrate control method, and the application area. The ROI was extracted in [9] by the segmenting moving regions through picture differences between successive frames. In [10], the ROI was determined by human face detection and tracking. The direct picture difference and skin-tone information were combined to indicate the ROI [11]. In [12], the human attention information was defined through the visual rhythm analysis to detect the video content such as object and camera motions. The visual rhythm is a representation of a video that captures the temporal information of content status from a coarsely spatially sampled video sequence. In [13], the three-dimensional depth sensation of the human visual system for multiview video was used to determine the ROI. It utilizes the temporal correlation among the frames and the inter-view correlation from view by view tracking. In [14], the two-level neural network classifier was used : the first level for providing an approximate classification of video object planes in foreground / background and the second level for implementing the object connectivity criteria and an automatic retraining procedure of neural network.

The bitrate for encoding the ROI was controlled by the assignment of highest quantization parameter for non-ROI [15], the adaptive perceptual quantization algorithm based on the masking effects of human visual system [16], and the fuzzy logic control algorithm for adjustment of quantization parameter [17]. All of the conventional methods aim to improve the picture quality within the ROI compared at the expense of the rest of

the image. However, the abrupt picture quality variation between adjacent the regions of ROI and non-ROI is open very noticeable and can degrade the subjective quality.

This paper proposes a method that improves the subjective picture quality by applying a quantization step-size differentially within a video sequence without abrupt change in picture quality between the ROI and remainder of the image. Examination of a range of video sequences reveals there are primarily four different types of ROI: those that show a lot of motion in the central region of a picture, at the upper region, at the lower region, and in the periphery. Then this paper presents a method that automatically allocates the quantization step-size differentially according to the location of the ROI.

The remainder of this paper consists of four sections. Section 2 presents the H.264 scalar quantizer in a coding process. Section 3 describes a new method for improving the subjective picture quality by varying the quantizer step-size throughout the image. Section 4 shows the simulation results for the analysis of objective and subjective picture quality. Finally, section 5 concludes this paper and describes future research.

II. H.264 SCALAR QUANTIZER

Quantization replaces the coefficient values with a much smaller number of approximate values. H.264 uses the discrete cosine transform to decorrelate the pixel values, and concentrate the energy into a few components. The coefficients are divided by the quantization step-size and then rounded to the nearest integer. The reconstructed coefficient has a small error relative to original because the original exact real value cannot be determined from the rounded value. In general, the quantizer can be formulated as follows [1].

$$FQ = \left[\frac{X}{Qstep} \right] \quad (1)$$

$$Y = FQ \cdot Qstep \quad (2)$$

where X is an input coefficient, $Qstep$ is the quantization step-size, $[\]$ is a round operator, FQ is the value obtained through quantization, and Y is a reconstructed value through inverse quantization.

The quantization performed by H.264 must be selected from one of 52 step-sizes, with the actual quantization step-size determined from a redefined table indexed by the quantization parameter. Within the table, the quantization step-sizes form a geometric sequence, with each increase of the parameter by six doubling the values of the quantization step-size. The wider range of the quantization step-size compared with MPEG-2 enables the quantizer to flexibly and precisely to control the balance between the bitrate and picture quality.

III. IMPROVING SUBJECTIVE PICTURE QUALITY WITHIN THE REGION OF INTEREST

The quantization step-size can be controlled for each macroblock or slice in order to achieve the desired bitrate. In the case of slice control, the step-size index for current slice will be increased if more than the desired number of bits are generated for the previous slice. On the other hand, the step-size index will be decreased if more bits are allowed after coding the previous slice.

In general, the bitrate is controlled according to only the sufficiency or insufficiency compared with an amount of the generated bits on a unit by unit basis. This manner of selecting the quantization step-size is acceptable if the whole area within each picture of the video sequence is region of interest. However, in many videos, the interest is not uniform across each frame.

Thus, this paper proposes a method that applies a variable quantization step-size differentially by considering the region of interest within a video. For performing this method, each frame within a video is classified according to the region of interest and then the quantization step-sizes are differentially allocated according to these regions.

3.1. Criterion for the Region of Interest

For all of video sequences including a scene of face within background and a scene of camera movement, the primary region of interest can be extracted from the global and local motion vectors. In the literature, the global motion vectors have been estimated in several ways. In [18], the global motion was achieved by removing the outlier motion vectors from the motion vector field. In [19], the feature macroblocks were extracted and then the global motion was determined using the only pixels in the feature macroblocks. In [20], the pixel- and motion vector-based global motion was estimated. In [21], the global motion flow field was generated. In [22], the camera zoom operation was detected through geometric constraint of the motion vector field. In [23], the gray-coded bit-planes were extracted to match the global motion.

While extraction of accurate global motion can give the effective classification for the region of interest, it also has large computational complexity to estimate perfectly. Instead, we use the motion vectors estimated during the process of picture coding to classify the region of interest without the extra complexity of searching for global motion.

First each frame is divided vertically and horizontally to 12 regions as shown in Figure 1. Then for each region, a representative motion vector is determined as the motion vector having highest frequency. According to the distribution of the representative motion vectors, each frame can be classified according to its region of interest. Because a slice, the unit for bitrate control, is generally divided in the vertical direction (although a horizontally divided slice can be acceptable) we classify the subjective region of interest vertically according to the following four cases.

(1) Central focus video: The focus is on the central regions rather than the whole frame in the case where object movement is primarily in the central regions of a picture. The background may either be fixed or moving uniformly through camera panning. The representative motion vector values at the central regions (R21, R22, R23, R31, R32, R33) of a picture are non-zero while those of the top and bottom regions (R11, R12, R13, R41, R42, R43) are either zero or approximately uniform. Also this central focus includes the case of that the number of non-zero representative vectors of the central regions is greater than the sum of both the top and bottom regions while all of the central regions have not the non-zero vectors. Central focus is also used if all of the representative motion vectors have non-zero values within all of regions. This can be considered a default type if a frame is not classified into the other three types because humans tend to concentrate their viewpoint on the central region of a video.

(2) Peripheral focus video: The region of interest is focused on the periphery in the case where the scene is being zoomed-out. The representative motion vector values of the top and bottom regions have non-zero values while the central regions are approximately zero.

(3) Upper focus video: The region of interest is in the upper regions in the case of the movement of the object located at the upper regions of a picture. The representative motion vector values in the upper regions (R11, R12, R13, R21, R22, R23) of a frame have non-zero values with zero values in the lower regions (R31, R32, R33, R41, R42, R43). Also this upper focus includes the case of that the number of non-zero representative vectors of the upper regions is greater than the lower regions while all of the upper regions have not the non-zero vectors.

(4) Lower focus video: The region of interest is in the lower regions in the case of the movement of the object located at the lower regions of a picture. The representative motion vector values in the lower regions (R31, R32, R33, R41, R42, R43) of a frame have non-zero values with zero values in the upper regions (R11, R12, R13, R21, R22, R23).

R11	R12	R13
R21	R22	R23
R31	R32	R33
R41	R42	R43

FIGURE 1: Picture division for classification of concerned region.

3.2. Differential Allocation of a Quantization Step-size

In a coding process according to the subjective importance of videos, the differential allocation of the quantization step-size can improve the subjective picture quality. The quantization step-size can be differentially applied according to the detected region of interest. However, the abrupt quality variation among the ROI and non-ROI regions should be avoided.

In all cases, the step-size is changed gradually according to the slice, vertically within a picture. First, it is necessary to find an average of the quantization step-size, \bar{Q}_s , for all of slices within a picture.

$$\bar{Q}_s = \frac{\sum_{i_s=0}^{N_s-1} Q_s(i_s)}{N_s} \quad (3)$$

where i_s is the slice order, and N_s is the number of total slices in a picture.

Then the new value of the quantization step-size, \tilde{Q}_s for each slice according to the classified ROI can be determined as follows:

(1) Central and peripheral focus videos

$$\widehat{Q}_s(i_s) = \alpha \left(i_s - \frac{\lfloor \frac{N_s}{2} \rfloor - 1}{2} \right) + \overline{Q}_s, \quad 0 \leq i_s \leq \frac{N_s}{2} \quad (4)$$

$$\widehat{Q}_s(i_s) = -\alpha \left(i_s - \frac{\lfloor \frac{N_s}{2} \rfloor - 1}{2} - \left\lfloor \frac{N_s}{2} \right\rfloor \right) + \overline{Q}_s, \quad \frac{N_s}{2} < i_s < N_s \quad (5)$$

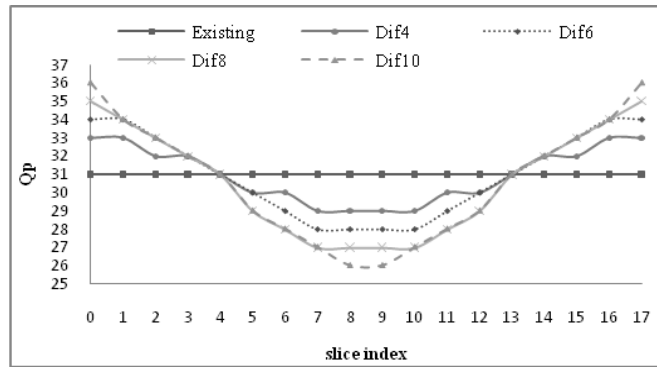
where $\lceil \cdot \rceil$ is a rounding up operator, $\lfloor \cdot \rfloor$ is a rounding down operator, and α is a constant. The value is determined as $\alpha < 0$ and $\alpha > 0$ for the central and peripheral focus videos, respectively.

(2) Upper and lower focus videos

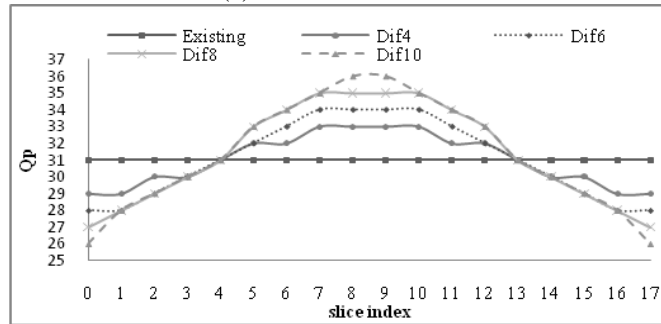
$$\widehat{Q}_s(i_s) = \alpha(i_s - (N_s - 1)/2) + \overline{Q}_s, \quad 0 \leq i_s < N_s \quad (6)$$

where α is a constant. The value is determined as $\alpha < 0$ and $\alpha > 0$ for the upper and lower focus videos, respectively.

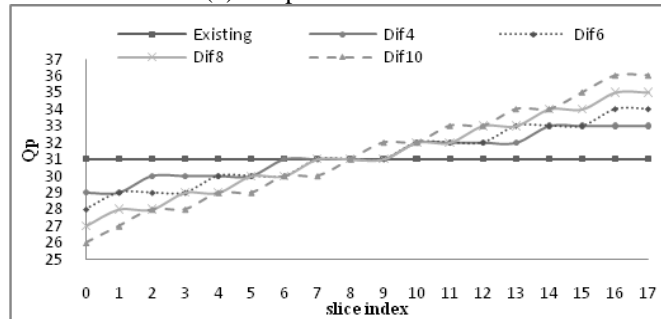
Figure 2 shows the examples of quantization parameter (step-size) allocated by Eqs. (4-6) for the four types of ROI in the case of $N_s=18$. The calculated quantization step-size may not exactly match those shown in Table 1. In this case, the nearest quantization step-size from the predefined allowed values is used.



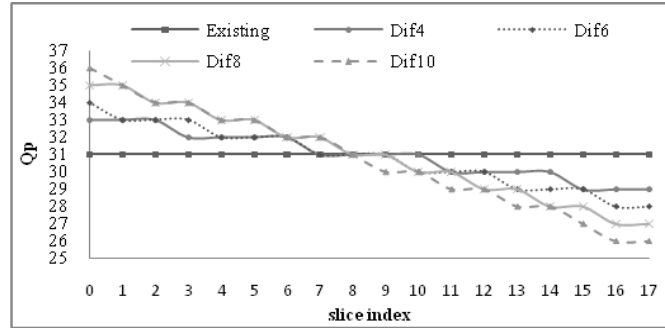
(a) Central focus video



(b) Peripheral focus video



(c) Upper focus video



(d) Lower focus video

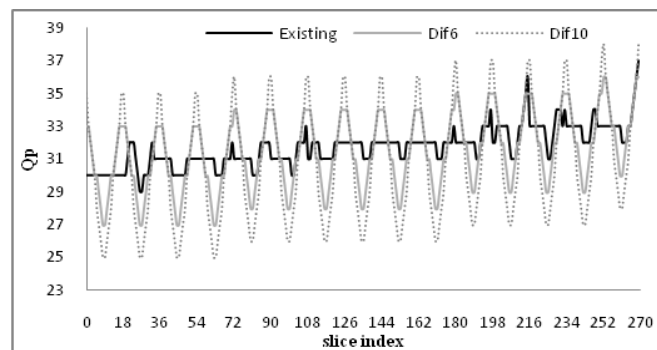
 FIGURE 2: Assigning of the quantization parameter for each ROI type (Dif n = difference between the maximum and minimum values of the quantization parameter was set to n).

As shown in Figure 2, in assigning the quantization parameter for the central focus video, the small and large values of the quantization step-size are assigned to the slices at the central and peripheral regions, respectively. However, for the periphery focus video, the small and large values of the quantization step-size are assigned to the peripheral and central regions, respectively. In the application of Eqs. (4-6), the quantization parameter between adjacent slices should not differ by more a factor of two. This avoids the abrupt degradation of the subjective picture quality. If the difference in the value is greater than a factor of two, it may cause the deterioration of picture quality due to the blocking effect at the border of slices.

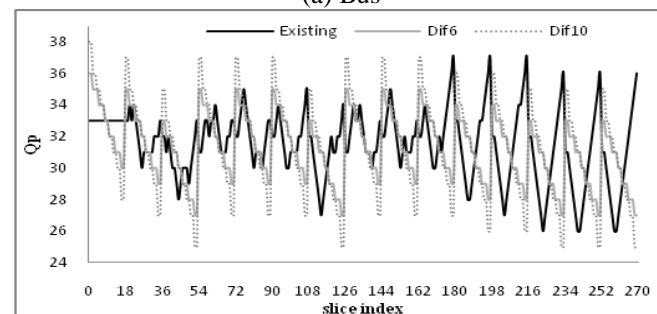
IV. SIMULATION RESULTS

The coding was performed using the JM encoder [24]. Four kinds of video sequence were selected, Bus, Flower, Foreman, and Waterfall videos with 352×228 picture resolution. The B-picture is not included because of the use of H.264 baseline profile. Each video contains 45 frames, with the GOP size set to 15 frames. The number of slices is the same as for each picture at 18 in the vertical direction. For the proposed method, the quantization step-size was differentially applied for the 18 slices for each picture. Four levels of α were used in order to make a difference between the maximum and minimum values of the quantization parameter as 4, 6, 8, and 10. Each video is classified as corresponding ROI from the estimated motion vectors. Foreman and Bus videos were classified as central focus videos, Flower video as lower focus video, and Waterfall video as peripheral focus video.

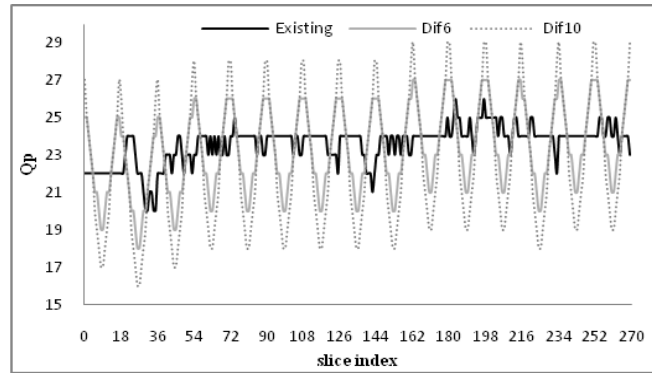
Figure 3 shows the assigned quantization parameters within the first GOP for the existing method and the proposed method. The target bitrate for each of these examples was 1Mbps.



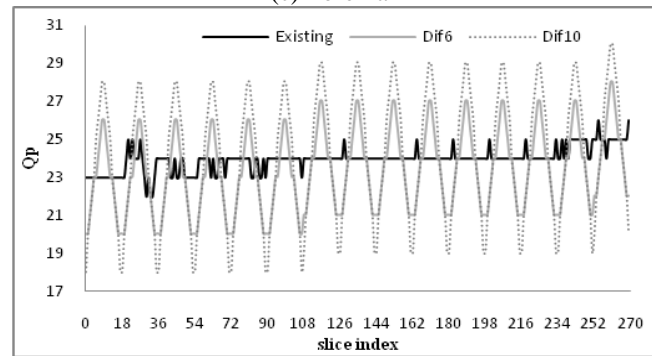
(a) Bus



(b) Flower



(c) Foreman



(d) Waterfall

FIGURE 3: Results of the assigning of the quantization step-size.

We can see that the quantization parameter of the proposed method is differentially assigned on basis of slice and slice to fit ROI.

4.1. Measurement of the Objective Picture Quality

For an objective picture quality criterion, the PSNR (peak signal to noise ratio) of reconstructed picture was investigated. Table 1 shows the average value of the PSNR for 45 pictures. Although the objective picture quality for each slice can be varied within a picture according to the differential allocation of quantization parameter, the average PSNR has the similar value for all of video sequence independently of the region of interest. However, if the difference between the maximum and the minimum values of the quantization parameter is assigned to more than 10, the PSNR can be lower than the existing method.

TABLE 1: Average PSNR (dB) while coding was performed as 1Mbps bitrate.

Video	Existin g method	Proposed method			
		Dif4	Dif6	Dif8	Dif10
Bus	32.36	32.36	32.34	32.33	32.33
Flower	31.05	31.09	31.02	31.01	31.10
Foreman	39.70	39.70	39.68	39.67	39.65
Waterfal l	38.24	38.25	38.23	38.23	38.12

4.2. Measurement of the Subjective Picture Quality

For measuring of the subjective picture quality, we used DSCQS (Double Stimulus Continuous Quality Scale) testing [25]. According this method, the assessors are positioned at a distance from the monitor equal to three times the diagonal length of the monitor used to display the videos. They then observe two videos in sequence on the monitor; one is an original video and the other is a video either using the existing method or the proposed method. The presentation order of the original and processed videos was random. The presentation of the test material:

The presentation of the test material:

- 1) Video A (Original or Processed) 12s,
- 2) Gray 3s,
- 3) Video B (Processed or Original) 12s,
- 4) Gray 3s

Assessors evaluated the picture quality of both videos using an ITU-R quality scale (Excellent=5, Good=4, Fair=3, Poor=2, Bad=1) [26].

The final subjectively picture quality assessment score was calculated from the mean over all assessors;

$$U = \frac{1}{N} \sum_{i=1}^N u_i \quad (7)$$

where u_i is score which is determined by each assessor, and N is number of assessors.

The measurement of the subjective picture quality was performed by distributing reconstructed videos to 10 assessors without notification of the methods and the focus regions. We got the average score of the subjective picture quality obtained from 10 evaluators as shown in Table 2.

TABLE 2: Results of the subjective picture quality for the concerned regions (bitrate of 0.8Mbps).

Video	Existing method	Proposed method				
		Dif4	Dif6	Dif8	Dif10	Focus
Bus	3.2	3.6	3.4	3.2	3.0	central
Flower	3.4	3.5	3.6	3.5	3.3	lower
Foreman	4.2	4.1	4.2	4.4	4.1	central
Waterfall	4.0	4.1	4.3	4.2	3.7	periphery

The Foreman and Bus videos represented good subjective picture quality on assigning of the quantization step-size with the central focus. For the Foreman video, the desirable difference between the maximum and the minimum values of the quantization parameter between slices is 8. However, for the Bus video, the desirable difference between the maximum and the minimum values of the quantization step-size between slices is 4. Because the peripheral focus shows the reappearance of periphery as the picture is zoomed out as the same as that of the Waterfall video, the periphery is more important than the central region in terms of subjective appearance. Therefore, the assigning small values of quantization step-size to the slices in the periphery brings good subjective picture quality. It can be seen that a difference of 6 between the maximum and minimum quantization parameter between slices will present the most excellent subjective picture quality. The Flower video shows the movement of the object located in the lower region of the video. Therefore, focusing the ROI on the lower region gives more emphasis to those movements rather than that of the whole region of the video. The desirable difference between the maximum and minimum quantization parameter values between slices is 6. For all of the videos, if the difference between the maximum and the minimum values of the quantization parameter is set to more than 10, then the subjective picture quality deteriorates.

V. CONCLUSIONS

This paper classified the videos into one of four types based on region of interest: central focus, peripheral focus, lower focus, and upper focus. Then, this paper allocated the quantization step-size (parameter) to the region of interest differentially without an abrupt change between the ROI and adjacent regions. Also, this paper demonstrated an improvement in subjective picture quality by adapting the step-size, compared to the existing method of allocating the quantization parameter. This paper verified that a reasonable difference between the maximum and the minimum values of the quantization parameter within a video was about 4-8 in the case of a differentially applied quantization parameter. Although this paper focused on the H.264 video coding, it can be directly used to other types of video coding standard. In addition, it can be expected that this paper will be extensively applied with other ROI extraction methods and the macroblock-based control besides the slice-based control.

REFERENCES

- [1] S.-k. Kwon, A. Tamhankar, and K. R. Rao, "Overview of H.264/MPEG-4 part 10," *Journal of Visual Communications and Picture Representation*, Vol. 17, No. 2, pp. 186-216, 2006.
- [2] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced video coding," *Final Committee Draft, Document JVT-F100*, Dec. 2002.

-
- [3] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 3, No. 7, pp. 560-576, 2003.
- [4] T. Shanableh and M.Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formalis," *IEEE Trans. Multimedia*, Vol. 2, No. 2, pp. 101-110, 2000.
- [5] Y. Liu, Z. G. Li, and Y. C. Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technology*, Vol. 17, No. 1, pp.68-78, 2007.
- [6] MPEG-2: ISO/IEC JTC1/SC29/WG11 and ITU-T, ISO/IEC 13818-2, Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: Video, *ISO/IEC and ITU-T International Standard Draft*, 1994.
- [7] L.J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Video Technology*, Vol.8, No. 4, pp. 446-459, 1998.
- [8] T. Chiang and Y.Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Video Technology*, Vol.7, No. 1, pp. 153-180, 1997.
- [9] H. Song and C.-C. J. Kuo, "A region-based H.263+ codec and its rate control for low VBR video," *IEEE Trans. Multimedia*, Vol. 6, No. 3, pp. 489-500, 2004.
- [10] L. Tong and K. R. Rao, "Region of interest based H.263 compatible codec and its rate control for low bit rate video conferencing," *International Symposium on Intelligent Signal Processing and Communication Systems*, pp.249-252, Dec. 2005.
- [11] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 18, No. 1, pp. 134-139, 2008.
- [12] M.-C. Chi, C.-H. Yeh, and M.-J. Chen, "Robust region-of-interest determination based on user attention model through visual rhythm analysis," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 19, No. 7, pp. 1025-1038, 2009.
- [13] Y. Zhang, G. Jiang, M. Yu, Y. Yang, Z. Peng, and K. Chen, "Depth perceptual region-of-interest based multiview video coding," *Journal of Visual Communication and Image Representation*, Vol. 21, No. 5-6, pp. 498-512, 2010.
- [14] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 8, No. 8, pp. 928-934, 1998.
- [15] D. Chai, K. N. Ngan, and A. Bouzerdoum, "Foreground/background bit allocation for region-of-interest coding," *International Conference on Image Processing*, Vol.2, pp. 438- 441, 2000.
- [16] Q. Liu and R.-M. Hu, "Perceptually motivated adaptive quantization algorithm for region-of-interest coding in H.264," *Advances in Multimedia Information Processing, Lecture Notes in Computer Science*, Vol. 5353/2008, pp. 129-137, 2008.
- [17] M.-C. Chi, M.-J. Chen, C.-H. Yeh, and J.-A. Jhu, "Region-of-interest video coding based on rate and distortion variations for H.263+," *Signal Processing: Image Communication*, Vol. 23, No. 2, pp. 127-142, 2008.
- [18] Y.-M. Chen and I. V. Bajic, "Motion vector outlier rejection cascade for global motion estimation," *IEEE Signal Processing Letters*, Vol. 17, No. 2, pp. 197-200, 2010.
- [19] F. Shang, G. Yang, H. Yang, and D. Tian, "Efficient global motion estimation using macroblock pair vectors," *International Conference on Information Technology and Computer Science*, pp. 225-228, July 2009.
- [20] M. Haller, A. Krutz, and T. Sikora, "Evaluation of pixel- and motion vector-based global motion estimation for camera motion characterization," *International workshop on Image Analysis for Multimedia Interactive Services*, pp. 49-52, May 2009.
- [21] M. Hu, S. Ali, and M. Shah, "Detecting global motion patterns in complex videos," *International Conference on Pattern Recognition*, pp. 1-5, Dec. 2008.
- [22] J.-h. Huang and Y.-s. Yang, "Effective approach to camera zoom detection based on visual attention," *International Conference on Signal Processing*, pp. 985 – 988, Oct. 2008.
- [23] T.-Y. Kuo and C.-H. Wang, "Fast local motion estimation and robust global motion decision for digital image stabilization," *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 442-445, Aug. 2008.
- [24] JM Reference Software, <http://iphome.hhi.de/suehring/tml/download>.
- [25] M. Bernas, "Image quality evaluation," *International Symposium on Video/Image Processing and Multimedia Communications*, pp.133-136, June 2002.
-