High Quality Region-of-Interest Coding for Video Conferencing based Remote General Practitioner Training

Manzur Murshed, Md. Atiur Rahman Siddique, Saikat Islam
Mortuza Ali, Guojun Lu, Elmer Villanueva†
Gippsland School of Information Technology
†Gippsland Medical School
Monash University,Churchill, Victoria 3842, Australia
Emails: {manzur.murshed,md.siddique,saikat.islam, mortuza.ali,guojun.lu,elmer.villanueva}@monash.edu

James Brown
Southern GP Training Ltd
Churchill, Victoria 3842, Australia
Email: james.brown@sgpt.com.au

***Abstract*—In a video conferencing based remote teaching system, visual quality is of critical importance, specially when it is used for medical training. However, transmission of high quality video data requires substantial amount of bandwidth. Unfortunately, communication systems in remote areas suffer from low transmission rate which demands significant compression of videos at the expense of visual quality. To strike a balance between the requirements of high visual quality and high compression ratio, in this paper, we propose to achieve higher visual quality only in the area of critical importance of a video known as *Region of Interest* (ROI). In the proposed scheme, the increase in the bit rate due to the improved visual quality in ROI is compensated with a degradation in the visual quality of non-ROI so that the effective bit rate meets the available bandwidth constraint. One of the salient features of the proposed method is that it operates outside the rate-distortion optimization (RDO) process of standard video codecs and thus ensures easy integration of the scheme into existing video coding standards and devices. Experimental results demonstrated that the proposed scheme significantly improves the visual quality of ROI without much degradation of the overall visual quality.**

***Keywords*-Region-of-interest; video coding; video conferencing.**

I. INTRODUCTION

With the ubiquitous availability of broadband Internet, video conferencing systems are replacing traditional face-to-face meeting and teaching methods. Video conferencing, which relieves the need for traveling across distance, is not only cost-effective but also energy efficient in terms of carbon footprint. Besides, video conferencing allows participants in rural areas to access quality learning where traveling to a distance is not an option. Indeed, our research motivation comes from the need of effective supervised training of newly appointed *general practitioners* (GPs) in regional Australia. Under the Australian General Practice Training (AGPT) program, every newly recruited GP goes through a mandatory supervised period. During this period, an experienced supervisor assesses the GP's consultation performance, offers on-demand consultation in complex cases, and conduct workshops along with a group of GPs. Ideally, the GP and the supervisor should be co-located at the same facility which is often infeasible as a supervisor has to supervise several GPs practicing at different remote sites. Clearly, in this scenario, remote supervision using video conferencing is an effective alternative.

While in remote teaching, based on video conferencing, the visual clarity is not a ruling factor in general, it is often a critical requirement in medical training. For instance, when a GP wants the supervisor to look at a particular skin rash or the tone of an infection of a patient, a proper investigation is impossible if the video quality is not clear. However, high quality video transmission requires high bandwidth which is often not available in rural areas. In a brute approach, video data need to be compressed significantly, given the low available bandwidth in remote sites, albeit at the expense of visual quality. However, in practice, the participants in a video conference are mostly interested in a small region of the video frames, termed as *region of interest* (ROI), which needs detailed inspection. Therefore, an effective approach is to allocate the scarce bandwidth unevenly between the ROI and non-ROI of a video. Instead of transmitting the whole frame at the same visual quality, the ROI area needs to be transmitted at a high quality which must be compensated by a degradation of the visual quality of non-ROI area.

ROI based video coding has been proposed in a number of contemporary research works [1]–[4]. Since video coding standards such as H.264 and MPEG are characterized by high-complexity encoding, which is often considerable for real time video communications, these works mainly focused on reducing the encoding complexity. Liu *et al.* [1] used skin tone and frame difference to detect ROI and proposed region based computational power and bit allocation by adjusting encoding parameters adaptively. Considering the fact that lower frequency coefficients of an image is less detectable to Human Visual System (HVS) than higher frequency components, Zheng *et al.* [2] proposed adaptively suppressing the low frequency coefficients of non-ROI blocks which, in

 ISBN: 978-1-61208-252-3

turn, reduces the overall computational complexity. Wang *et al.* [3] proposed using the texture and motion features of a video to determine its ROI and non-ROI. The authors then proposed a dynamic parameter allocation scheme to reduce the computational complexity to attain the low power requirement of portable devices. Considering the importance of rate control in ROI based video coding, Yang *et al.* proposed a rate control mechanism in [4]. The scheme proposed in [4] determines the *quantization parameter* (QP) for the ROI, based on the user defined interest level, and adaptively allocates bits between ROI and non-ROI regions.

All above schemes essentially focused on reducing the computational complexity of encoding. However, with the advancement of VLSI techniques, high computational power is now available even in portable devices like Apple iPAD or Sumsung Galaxy III. Moreover, in recent days, neither the computational power nor the battery life is a serious concern for PC and laptop users. More importantly, these schemes require custom rate-distortion algorithms which are difficult to accommodate in the framework of existing video coding standards. The *rate-distortion optimization* (RDO) algorithms implemented in the standard codecs are well studied and are widely being used in real life applications. Therefore, one of the objectives of our research is to design an effective variable bit allocation scheme that can easily be integrated with the standard video codecs. In this paper, we propose a scheme to transmit the ROI at a high quality where the additional bits used to transmit the ROI is compensated by transmitting non-ROI regions at a lower (than suggested by the standard codec) quality. In effect, we attempt to maintain the same level of bandwidth usage while transmitting the ROI at a higher quality.

The organization of the rest of the paper is as follows. We briefly review the architecture of current video coding standard H.264 and its rate-distortion control mechanism in Section II. In Section III, we then propose our scheme to improve the visual quality of ROI without much degradation of the overall quality of the video while meeting the bandwidth constraint. Extensive experimental results are then presented in Section IV to validate the efficacy of the proposed method. In Section V, finally we conclude our paper.

II. Visual Quality Control in Standard Codecs

Video is a sequence of image frames that are played back at a specific frame rate. The feasibility of video compression stems from exploiting its *interframe redundancy and intraframe redundancy*. The similarity among the successive frames of a video is referred to as interframe redundancy. On the other hand, intraframe redundancy refers to the correlation that exists between a pixel and its neighbors in the same frame. Effective video compression relies on efficient exploitation of all these redundancies.

The techniques of exploiting the intraframe and interframe redundancies are referred to as intra coding and inter coding, respectively. In both of the techniques, a frame is partitioned into non-overlapped, fixed sized, rectangular blocks called *macroblock*. For each of the macroblocks in the current frame, a prediction is made based on the previously encoded data. In intra coding, the prediction is made from previously encoded macroblocks of the same frame, while in inter coding a set of recently coded frames called a Group of Pictures (GOP) are used for prediction. Then the residual, i.e., the difference between the actual block and the predicted block, is transformed. After transformation, such as discrete cosine transform (DCT), most of the energy of the residual is concentrated into a few low frequency coefficients. Since HVS is less sensitive to the distortions at high frequency components, significant compression can be achieved by discarding these high frequency coefficients, by using a coarse quantizer, without much degradation of the visual quality of the reconstruction.

Indeed, in the video coding standards, it is the quantization step that controls the trade off between bit rate and visual quality. In the H.264 video coding standard, the rate-distortion trade off is controlled by the parameter QP which can take a value between 0 and 51 and refers to a matrix of Qsteps, i.e,. the quantization step size. In general, a higher Qstep achieves more compression, however, results in higher distortion and vice versa. The quantization matrices are such that most of the high frequency coefficients become zero after quantization. For more details on H.264 see [5].

In real-time applications, given the available bandwidth, the value of QP is a function of network usage (determined from the emptiness of Coded Pixel Buffer (CPB)) and the estimate of bit allocations for the current frame and its GOP. More specifically, in H.264 video coding standard a RDO algorithm allocates a number of bits to the GOP depending on the emptiness of the CPB. The GOP includes a number frames that can be any of I, P, or B frames. A GOP always begins with an I frame which is followed by a sequence of P and B frames. Depending on the type of current frame and the remaining bit allocation for the current GOP, a frame level bit allocation is decided in the RDO process.

Since QP essentially controls the trade off between the visual quality and the compression ratio, it is an effective tool to achieve varying image quality and enforcement of bit allocation for individual macroblocks. In the following section, we explain how we can adjust QP for individual macroblocks to offer better image quality in ROI.

III. Rate Adjustment for High Quality ROI Encoding

To increase the visual quality of the ROI, we need to decrease the QP values of the macroblocks within the ROI so that the quantization distortion during encoding is less for these macroblocks. This allows the decoder at receiver to reconstruct a better representation of the original frame. However, using a lower QP to offer lower quantization

 ISBN: 978-1-61208-252-3

distortion at a ROI also increases the data size. If the increase in the bit-rate is not compensated in the non-ROI then it will result in buffer or network overflow. To overcome this problem, the non-ROI needs to be coded using higher QP values to compensate for the increase in data size due to improved encoding of ROI. The compensation mechanism needs to ensure that, while satisfying the bandwidth constraint, the maximum possible information is retained.

Adjusting the visual quality of non-ROI to maximize the overall visual quality, subject to the available bandwidth, is challenging. If a lower value of QP is used for non-ROI, although the visual quality would be better, the resulting bit rate will then be higher than the available bandwidth which will lead to call termination. On the other hand, if the QP for non-ROI is unnecessarily high, some of the available bandwidth will be left unused which could have been used to improve the overall visual quality. The primary obstacle in selecting the appropriate QPs for ROI and non-ROI macroblocks is that although bandwidth and quality control parameters are closely related, their relationships can not be expressed with simple functions. In the following, we explain our proposed method of improving the visual quality of ROI while maximizing the overall visual quality subject to available bandwidth. Although, the proposed scheme assumes the x264 codec as an efficient implementation of H.264, it can easily be adopted to other implementations of H.264 as well.

In x264 codec, a award winning implementation of H.264, after allocating a bit budget to a frame, QP for each of the macroblocks is computed adaptively based on the complexity and compressibility of the macroblock. Therefore, we assume that while adjusting the QP values after the RDO of x264 encoder, the variations in QPs among the macroblocks of ROI and among the macroblocks of non-ROI should be maintained. More specifically, the QP of each of the macroblocks in ROI should be decreased by a constant factor C as specified by the user. Similarly, a fixed factor T must be added to the QP values of each of the macroblocks in non-ROI. Now we need to determine the value of T such that the resulting bit rate after the adjustment of QPs matches the available bandwidth. After extensive experimentation we have observed that to ensure that the effective bit rate is close to the target bit rate, the mean of the QPs of a frame after adjustment should be same as the mean of the QPs computed by the RDO process of x264 encoder. Therefore, QP values of non-ROI should be increased by $T = nC/(N-n)$, where N is the total number of macroblocks in a frame and n is the number of macroblocks in ROI. Clearly, this method operates outside the RDO process and thus can be easily incorporated into the standard codecs.

IV. Experimental Results

In image and video compression, *peak signal to noise ratio* (PSNR) is used as an approximation to human perception of reconstruction quality. A higher PSNR value indicates lower loss during compression and thus is preferable. Therefore, in our experiments we used PSNR as a metric for visual quality.

We implemented the proposed scheme in x264 codec [6]. To implement a real time video communication system with ROI based coding, we extended the linphone [7] software, which is a mature softphone application for Linux, Windows, and Mac, to use in the modified x264 codec. For comparison purpose, we used this modified x264 application on a number of YUV test video sequences. The QCIF and CIF YUV videos were collected from [8] while the higher resolution (4CIF and 720p) video sequences were downloaded from [9] and [10], respectively. Under different target bit rate requirements, we determined the effective bit rate and PSNR of ROI, non-ROI, and overall frame for these video sequences having different resolution and image complexity.

Fig. 1 demonstrates the impact of ROI based encoding on the visual quality of four standard test video sequences: 'Silent,' 'News,' 'Mother & Daughter,' and 'Foreman'. In each of the video sequences, the ROI contained the face which is detected using the Face Detection module of OpenCV [11]. The figure shows that the perceptual visual quality of ROI is improved without much degradation of overall frame quality.

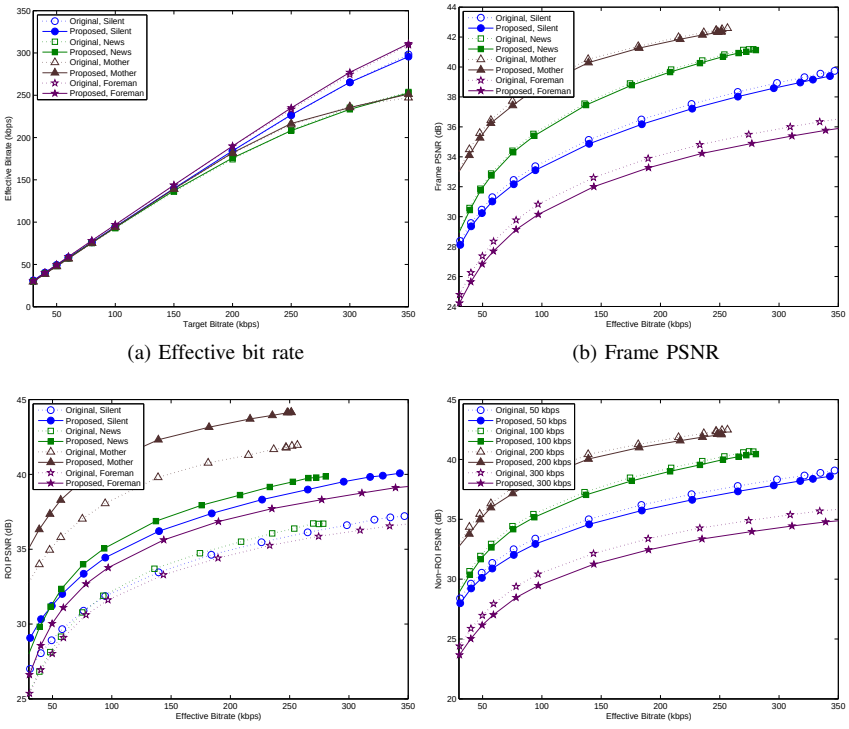
The performance of the proposed technique for a QP-shift of $C = 5$ is illustrated in Fig. 2a-2d in terms of effective bit rate, overall frame PSNR, ROI PSNR, and non-ROI PSNR, respectively. It follows from Fig. 2a that the bit rate resulting from the proposed method is considerably close (within ± 1.5 kbps on average) to the one obtained by original x264 codec. This validates our equi-mean QP hypothesis of readjusting QP values of ROI and non-ROI.

Fig. 2b shows that the overall frame PSNR obtained with the proposed method degraded slightly than that with original x264 codec. The average decrease in overall frame PSNR was found to be 0.3 dB, 0.13 dB, 0.19 dB, and 0.61 dB for 'Silent', 'News', 'Mother & Daughter', and 'Foreman' video sequences, respectively.

It is worth mentioning that the visual quality of ROI improved in every test video sequences. The ROI PSNR was found to be 2.62 dB, 3.09 dB, 2.39 dB, and 2.2 dB higher than that obtained with original x264 codec in 'Silent', 'News', 'Mother & Daughter', and 'Foreman' test video sequences, respectively. This increase in visual quality is due to lowering of the ROI QPs which resulted in additional bit allocation for ROI. This increase in bit rate for ROI was compensated by increasing the QPs of non-ROI which lead to a decrease in visual quality of the non-ROI. The decrease in non-ROI PSNR using the proposed method was found to be 0.45 dB, 0.27 dB, 0.34 dB, and 0.9 dB for 'Silent', 'News', 'Mother & Daughter', and 'Foreman' video sequences, respectively.

 ISBN: 978-1-61208-252-3

Figure 1: Impact of high quality ROI encoding on 'Silent,' 'News,' 'Mother & Daughter,' and 'Foreman' video sequences. The figures demonstrate that the proposed method improves the visual quality of ROIs without much degradation of the visual quality of non-ROIs.



(a) Effective bit rate (b) Frame PSNR

(c) ROI PSNR (d) Non-ROI PSNR

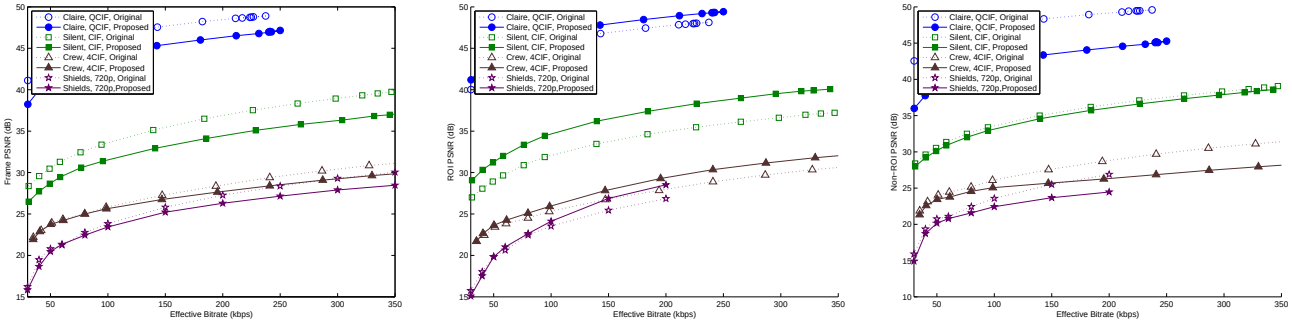
Figure 2: The proposed ROI based encoding with a QP-shift of $C = 5$ improves the visual quality of ROI at the cost of moderate decrease in non-ROI and overall frame PSNR.

The ROI area in the above videos contained only face which is about 5% of the total frame. Intuitively, the performance of the proposed scheme will vary with the ROI area and frame resolution. In the following, we present and analyze our experimental results to understand the impact of these parameters on the proposed ROI based coding scheme.

A. *Impact of Video Resolution*

To determine the impact of image resolution on the performance of the proposed method, a number of tests were

 ISBN: 978-1-61208-252-3



(a) Frame PSNR (b) ROI PSNR (c) Non-ROI PSNR

Figure 4: Impact of varying video resolution on the performance (PSNR) of ROI based video encoding.

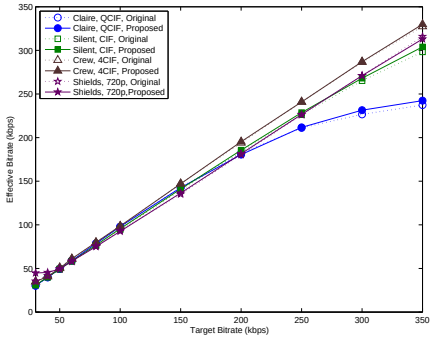


Figure 3: Impact of varying video resolution on effective bit rate.

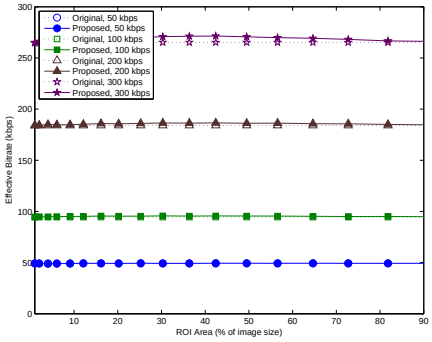


Figure 5: Impact of ROI size on effective bit rate.

conducted on several QCIF, CIF, 4CIF, and 720p videos using a QP shift of $C = 5$. In each of the video sequences, (left) half of the frame was used as ROI and the other (right) half was used as non-ROI. For space constraint, results for only one video per resolution are reported in this paper. However, similar results were observed for other video sequences as well.

It follows from Fig. 3 that the proposed method resulted in good agreement between effective bit rate and target bit rate. The effective bit rate achieved with the proposed method was found to be within 5.2 kbps, 1.03 kbps, 0.48 kbps, and 1.75 kbps of the target bit rate for 'Claire' (QCIF), 'Silent' (CIF), 'Crew' (4CIF), and 'Shields' (720p) video sequences, respectively.

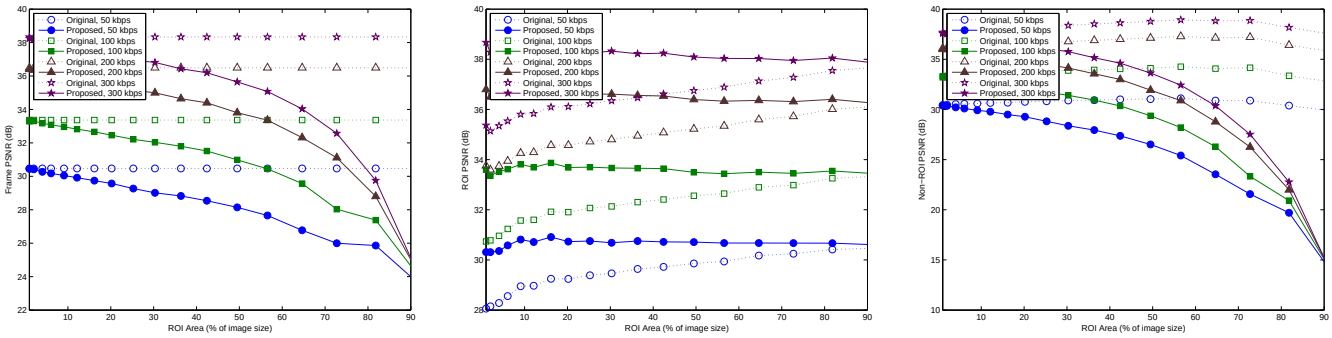
Fig. 4a illustrates that when the proposed method was used the overall frame PSNR degraded slightly for CIF, 4CIF, and 720p video sequences. The CIF, 4CIF, and 720p video sequences showed an increase in ROI PSNR by 2.6 dB, 0.54 dB, and 0.38 dB (see Fig. 4b) at the cost of slight degradation in non-ROI PSNR by 0.45 dB, 1.95 dB, and 1.1 dB (see Fig. 4c). The adverse result for the QCIF video was primarily due to its low resolution for which the QP shift of $C = 5$ was considerably large. The lower QP used for ROI was compensated by using a higher QP in non-ROI which resulted in significantly lower PSNR for non-ROI and for the whole frame.

B. *Impact of ROI Size*

To demonstrate the impact of varying ROI size on the performance of the proposed method, we use the CIF 'Silent' video and change the ROI size from 1% to 90%. The center of ROI was at the center of the video since the important visual objects are usually located at the centre of the video. With varying ROI size, the effective bit rates were found to be considerably close to that obtained by the original x264 codec (See Fig. 5). At target bit rates as high as 200 kbps and 300 kbps, the bandwidth usage in the proposed method differed by 1.49 kbps and 3.47 kbps, respectively. This observation demonstrates the robustness of our equi-mean QP hypothesis with the variation of ROI size, relative to the frame size.

It follows from Fig. 6a that at each bit rate, the frame PSNR decreased in the proposed method when the size of ROI was increased. This pattern can be explained from the resulting ROI PSNR (Fig. 6b) and non-ROI PSNR (Fig. 6c). A much higher ROI PSNR (compared to original x264 codec) was obtained with the proposed technique when the ROI was small. As the ROI area increased, the ROI PSNR degraded and approached the ROI PSNR obtained with the original x264 codec. For instance, at 50 kbps, the increase in ROI PSNR by the proposed method from that obtained by the original x264 codec was 2.25 dB and 0.15 dB when the ROI area was 1% and 90% of the total frame, respectively.

 ISBN: 978-1-61208-252-3



(a) Frame PSNR (b) ROI PSNR (c) Non-ROI PSNR

Figure 6: Impact of varying ROI size on the performance (PSNR) of the proposed scheme on 'Silent' video sequence.

While determining the effective QP shift, we first looked for the minimum original QP within the ROI and then its difference from the QP shift limit was subtracted from the QP of each ROI macroblock. Therefore, as the area of ROI increases, the probability of finding a lower QP within ROI also increases. For instance, when the ROI size was 9.09%, 18.18%, and 27.27% of the whole image, the minimum original QP-adjustment within ROI was found to be around 4, 7, and 8, respectively.

An inverse impact on the image quality of non-ROI was also observed. When the ROI was small, the decrease in ROI QP was compensated with slightly higher QP for non-ROI. Therefore, the non-ROI image quality did not degrade considerably and the non-ROI PSNR using the proposed method was nearly same as that obtained with original x264 codec. However, when the ROI became larger, with the same QP shift for ROI, we needed a much higher QP for the non-ROI to ensure that the mean QP of the whole frame remained same. Therefore, the non-ROI image quality degraded more when the ROI area was larger. For instance, at 50 kbps, the decrease (compared to original x264 codec) in non-ROI PSNR was 0.06 dB and 15.61 dB when the ROI area was 1% and 90% of the total frame, respectively.

In effect, when the ROI size is small, the proposed method gives a considerably higher ROI PSNR and slightly lower non-ROI PSNR compared to the original. However, when the ROI is increased, the increase in ROI PSNR diminishes and the decrease in non-ROI PSNR intensifies. As a result, the overall frame PSNR degrades with increasing ROI size.

V. Conclusion

In this work, we presented a novel approach to transmit ROI at a higher quality than the non-ROI counterpart, subject to the bandwidth constraint. The bandwidth usage is maintained by the use of original H.264 rate control algorithms which are proven to be reliable. The desired result is achieved by decreasing QP for ROI and compensating it in non-ROI without affecting the overall bit allocation per frame. Thus this method can be easily integrated with commercially available standard codecs to selectively transmit the ROI at a higher quality for effective real-time remote medical training.

Acknowledgment

This research was supported by the Education Integration Program (EIP) grant from General Practice Education and Training (GPET) Ltd., Australia.

References

[1] 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. Circuits Syst. Video Technol.*, pp. 134–139, 2008.

[2] Y. Y. Zheng, X. Tian, and Y. W. Chen, "Adaptive frequency coefficient suppression for ROI-based H. 264/AVC video coding," in *Proc. IEEE ICNSC*, 2008, pp. 714–718.

[3] M. Wang, T. Zhang, C. Liu, and S. Goto, "Region-of-interest based dynamical parameter allocation for H. 264/AVC encoder," in *Proc. PCS*, 2009, pp. 1–4.

[4] L. Yang, L. Zhang, S. Ma, and D. Zhao, "A ROI quality adjustable rate control scheme for low bitrate video coding," in *Proc. PCS*, 2009, pp. 1–4.

[5] I. Richardson. (2009, Apr.) H.264 quantization parameter. [Online]. Available: http://vcodex.blogspot.com.au/2009/04/h264-quantization-parameter.html

[6] (2012, Jan.) VideoLAN Organization. [Online]. Available: http://www.videolan.org/developers/x264.html

[7] (2012, Jan.) Linphone: Open source video SIP phone for desktop & mobile. [Online]. Available: http://www.linphone.org/

[8] (2012, Jan.) Video Trace Library. [Online]. Available: http://trace.eas.asu.edu/yuv

[9] (2012, Feb.) Network Systems Lab (NSL) at Simon Fraser University (SFU). [Online]. Available: http://nsl.cs.sfu.ca/video/library/YUV/4CIF/

[10] (2012, Feb.) Xiph.org Video Test Media (derf's collection). [Online]. Available: http://media.xiph.org/video/derf/

[11] (2012, Jan.) Open Source Computer Vision (OpenCV). [Online]. Available: http://opencv.willowgarage.com

 ISBN: 978-1-61208-252-3