# ADAPTING THE STREAMING VIDEO BASED ON THE ESTIMATED MOTION POSITION

*Hussein Muzahim AZIZ*[1], *Marcus FIEDLER*[1], *Hakan GRAHN*[1], *Lars LUNDBERG*[1]

[1]School of Computing, Blekinge Institute of Technology, Blekinge Tekniska Hokskola, Karlskrona, SE 371 79 Sweden

hussein.aziz@bth.se, marcus.fiedler@bth.se, hakan.grahn@bth.se, lars.lundberg@bth.se

***Abstract.*** *In real time video streaming, the frames must meet their timing constraints, typically specified as their deadlines. Wireless networks may suffer from bandwidth limitations. To reduce the data transmission over the wireless networks, we propose an adaption technique in the server side by extracting a part of the video frames that considered as a Region Of Interest (ROI), and drop the part outside the ROI from the frames that are between reference frames. The estimated position of the selection of the ROI is computed by using the Sum of Squared Differences (SSD) between consecutive frames. The reconstruction mechanism to the region outside the ROI is implemented in the mobile side by using linear interpolation between reference frames. We evaluate the proposed approach by using Mean Opinion Score (MOS) measurements. MOS are used to evaluate two scenarios with equivalent encoding size, where the users observe the first scenario with low bit rate for the original videos, while for the second scenario the users observe our proposed approach with high bit rate. The results show that our technique significantly reduces the amounts of data are streamed over wireless networks, while the reconstruction mechanism will provides acceptable video quality.*

## Keywords

***Mean opinion score, region of interest, streaming video, sum of squared differences.***

## 1. Introduction

Nowadays mobile cellular networks are able to support different type of services, such as video streaming that makes a great demand on the wireless networks bandwidth. Bandwidth is the most critical resource in mobile networks [5], therefore it is important to employ adaption mechanisms for efficient use of the available bandwidth. Network adaptation refers to how much network resources (e.g. bandwidth) a video stream should utilize for video content, resulting in designing an adaptive streaming mechanism for video transmission [3]. The main feature of H.264/SVC is to provide bandwidth-optimized transmission of real time video streaming by observing current network conditions [6]. H.264 contains a rate-control algorithm that are dynamically adjusts the encoder parameters to achieve a target bit rate by allocates a budget of bits to the video frames sequence. The main concept of the rate-control algorithm is a quantitative model that describes the relationship between the quantization parameter and the actual bit rate [8].

The quantization parameter (QP) has a great impact on the encoder performance, because it regulates how much spatial details can be saved. As the increases of the QP, some of the details are aggregated so that the bit rate drops with some increases in distortion and some loss of the video quality [12]. The frame size can be reduced to eliminate the artifacts at low bit rate environment. However, the reduction of size does not guarantee a good quality, as the original video contents are in high resolution where the video quality will be poor when the bit rate is low [14]. The limitation of the available bit rate is one of the key technologies required for efficiently allocating bits for the purpose of the video contents for transmitting the Region-Of-Interest (ROI) [13]. The user attention is the ability to detect the interest parts of a given scene that called attention area or ROI [11]. ROI extracted from the streaming video, as the ROI consider the most interesting and important parts of the video frames, while the background (non-ROI) are dropped as it considered less important region.

The major idea for coding the ROI video is to reduce the bit rate by sacrificing the quality of the non-ROI; the other is to allocate more bandwidth to ROIs and enhances the ROI quality by giving priority or important factor for determination the quantization parameters (QP) in the encoder side [13]. In this paper, we present an adapting technique to reduce the amounts of data to be streamed over the wireless networks. The streaming server will set the reference frames and extract the slice region from the frames that are between reference frames. After the mobile device has received the video stream, linear interpolation between reference frames are performed to reconstruct the pixels has been dropped in

 

the server side. Mean Opinion Score (MOS) measurements that are obtained from a panel of human observers are used to evaluate the videos after the dropping pixels (non-ROI) are reconstructed.

The remainder of this paper is organized as follows. Section 2 provides related work to spatial adaptation for slicing the video frames over a wireless network. Section 3 explains the proposed video streaming scenario. Section 4 adaptive the video according to the quantization parameters for the both scenarios, while the analyses of the subjective data from the experiments study are presented in Section 5. Section 6 provides the experimental test results. Finally, we conclude this study in Section 7.

## 2. Related Work

Several techniques have been proposed for spatial adaptation for slicing the video frames. Wang and El-Maleh [15] proposed an adaptive background (non-ROI) skipping approach where every two consecutive frames grouped into a unit. In each unit, the first non-ROI is code, while the second non-ROI is skipped (using predicted macro blocks with zero motion vectors). The ROI is been identified either automatically detected or specified by the end-user, while non-ROIs will be skipped and the numbers of bits will be allocated to ROI, to ensures the best visual quality of the video sequence.

Shuxi et al. [7] proposed a spatial domain adjustable resolution method based on the ROI. The proposed method is to divide the video into ROI and non-ROI. The ROI have more details, as it is the most important region in the video frames, while the details of the non-ROI are ignored. The ROI will perform coding on the resolution of the original frame. The non-ROI will perform coding on the low resolution. Coding the frame from the region of non-interest after down-sampling in order to achieve adjustable resolution feature based-on the region. They claimed that the proposed method would reduce the complexity of the encoding to guarantee the subjective and objective quality of ROI.

Inoue et al. [11] proposed data format based on Multiview Video Coding (MVC) for two types of partial delivery method with and without lower-resolution. The two types of partial delivery method are considered in their work for multi-bit rates and resolution to maximize the partial panoramic video quality under restricted bandwidth. The first type is partial delivery method that used to deliver the frames without lower-resolution; while the second partial delivery method with lower-resolution. In their work, they examined the impact of subject image quality in terms of delivery method and ROI movement. The work examined three delivery methods, 'Deliver-all', 'Partial delivery' with and without lower-resolution. They claimed that the two types of the proposed partial delivery methods could achieve higher subjective video quality than the deliver-all methods when the ROI on the move. The above researchers identify the ROI as the most attractive object or regions to the viewers. Where some researcher considered two different resolutions, a high resolution for the ROI and low resolution for the non-ROI, while others considering skipping the non-ROI or encoded with low resolution to provide a good quality to the ROI in the video sequence that can cope with the bandwidth limitation.

In our study, we define the ROI as the most motion regions in the video frames by calculating the sum of the motion differences between the video frames and drop the part that are less motion, which will be outside the ROI (non-ROI).

## 3. The Proposed Technique

The Sum of Squared Differences (SSD) metric is computed to detect the most motion regions, which we call it, the ROI. The ROI will be extracted from the frames between reference frames in the server side. On the mobile side, the part of the region that is outside the ROI (non-ROI) will be reconstruct by using linear interpolation between reference frames, as shown in Fig. 1.

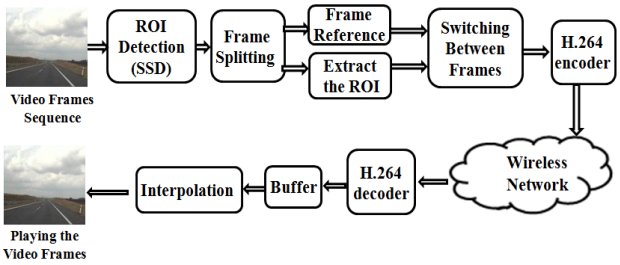


**Fig. 1:** The proposed streaming technique.

### 3.1. Detecting the ROI

The SSD technique is a commonly used technique for motion estimation for video encoding standards like H.264 [2]. Computing the SSD for the consecutive video frames will be similar except for the changes that might be induced by objects moving within the frames.


**Fig. 2:** Scanning the slice region based on SSD *(k)*.

The SSD is computed to detect the estimate position of the ROI within the frame for the consecutive

video frames based on the highest intra-slice differences. The SSD is computed by scanning the consecutive video frames from top to bottom to identify the highest intra-slice differences. The SSD goes high as an indicator that it is the most motion and important part in the video frames that we will considered as the ROI, as shown in Fig. 2 and according to (1):

$$SSD(k) = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\sum_{x=0}^{L-1}\left(F_x(i, j+k) - F_{x-1}(i, j+k)\right)^2 , \qquad (1)$$

where $L$ is the length of the frame sequences, $N \times M$, is the height and width, while $k$ is the fixed region which is.

The test videos used in this work were the samples of video sequences Highway, Akiyo, Foreman, News, and Waterfall, with a resolution of 144 × 176 [17]. The videos have been chosen as they have different characteristics. For Highway video that is shown in Figure 3. The SSD is lower in the top of the frames as indicators that there are less activities, therefore the motions are lesser. The SSD is higher in the bottom of the frames as there are high activities in the video frame, therefore the motion is high.

For Waterfall video that is shown in Fig. 4. The SSD is increasing dramatically from the top of the frames until the bottom of the frames, as the video is zooming out all the times, where the highest SSD is in the bottom of the frames as an indicator that there are the most motion regions. For News video that is shown in Fig. 5. The SSD is the lower in the top and in the bottom of the frames but it is slightly higher in the middle part that is closed to the top as there are more activates, therefore the motion is high. For Foreman video that is shown in Fig. 6. The SSD value is approximately within the same range, because the Foreman video is shaking all the time, therefore, the SSD is relatively high in all regions but it is slightly higher in the bottom of the frame. For Akiyo video that is shown in Fig. 7. The SSD is lower in the top and in the bottom of the frames, while it is higher in the middle of the frames as an indicator that there are high activities in the middle of the frames, therefore the motion is high. The highest value of the SDD is the most important regions in the frame, therefore it considered as the ROI in this work.

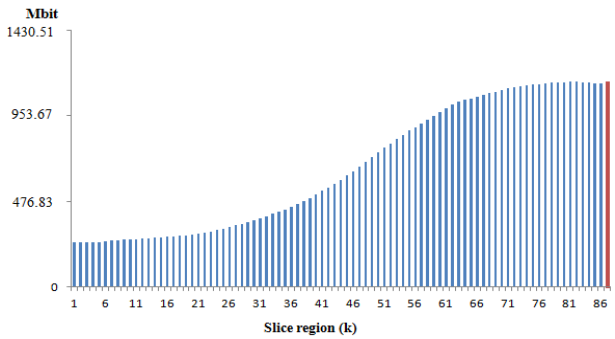


**Fig. 3:** The SSD ($k$) for Highway video.

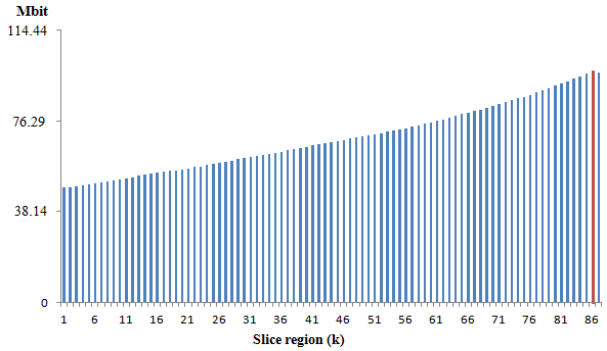


**Fig. 4:** The SSD ($k$) for Waterfall video.

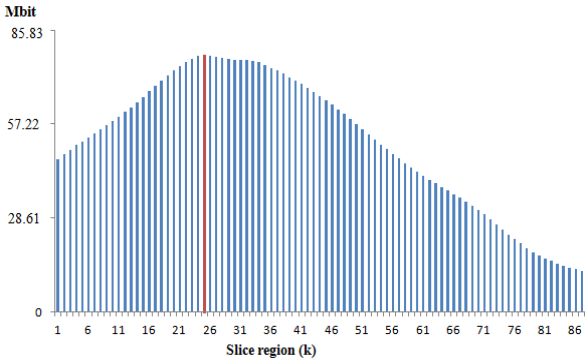


**Fig. 5:** The SSD ($k$) for News video.

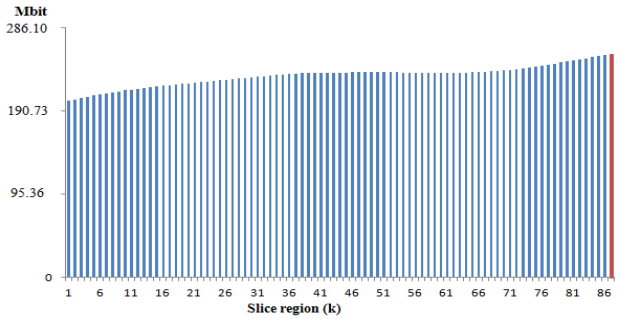


**Fig. 6:** The SSD ($k$) for Foreman video.

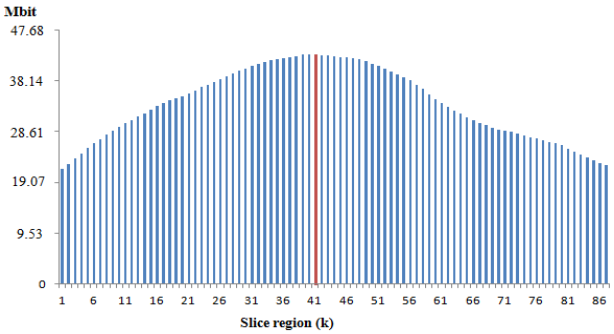


**Fig. 7:** The SSD ($k$) for Akiyo video.

## 3.2. Extracting the ROI

The streaming server will establish the connection according to the mobile request. The server will compute the SSD value to the consecutive video frames, to detect and extract the position of the ROI based on the highest

intra-slice differences and drop the pixels that are outside the ROI.

The sequences of the video frames are split to set the reference frames and to extract the ROI from the frames between reference frames. The sequence position of the reference frames are considered in this study is every fifth frame, as the maximum distance between frames that do not have high affect on the quality of the viewers' perception [1]. The reference frames with the ROI are combined by a switching mechanism for encoding and transmitting in the normal way by using H.264 codec, as shown in Fig. 1.

### 3.3. Reconstructing the Video Frames

After the mobile device start receiving the video frames (reference frames and ROI), it will be held in the buffer to reconstruct the surrounding pixels that are outside the ROI (non-ROI) that been dropped on the server. The method is used to reconstruct the non-ROI is linear interpolation [16]. Linear interpolation is applied between reference frames to reconstruct the non-ROI pixels. After the frames been returned to their original shape, the video will be played on the mobile screen.

## 4. Quantization Parameter Adaptation

The video will encode to obtain the optimum visual quality within the available network bandwidth. The bit allocation for the video should achieve the tradeoff between encoding video quality and bandwidth limitation.

The bit allocation for the video is encoded by using H.264 ffmpeg codec [18]. The videos are encoded to identify the effectiveness of the bit rates and the QP on the encoding size. The encoding size will be different from one video to another as the videos had different characteristics.

**Tab.1:** The encoding videos size for the two scenarios.

| **Test videos** | Highway | Water-fall | News | Foreman | Akiyo |
|---|---|---|---|---|---|
| **Size (KB),QP= 2** | 4564 | 1280 | 1128 | 616 | 350 |
| **Scenario 1: QP** | 14 | 10 | 11 | 12 | 6 |
| **Size (KB)** | 707 | 165 | 167 | 146 | 165 |
| **Coding Efficiency Gain** | 84,50% | 87,10% | 85,19% | 76,29% | 52,85% |
| **Scenario 2 : QP** | 10 | 8 | 7 | 9 | 4 |
| **Size (KB)** | 693 | 175 | 172 | 139 | 172 |
| **Coding Efficiency Gain** | 84,81% | 86,32 | 84,75% | 77,43% | 50,85% |

Two scenarios are proposed to encode the videos, the first scenario; where the original video frames are encoded with default QP for a bit rate of 64 kbps. The second scenario (the proposed scenario) where the videos are encoded with a bit rate of 128 kbps, while the QP is adaptive to get equivalent size to the videos that are in the first scenario, as shown in Tab. 1. Encoding the videos in the second scenario with adaptive QP is to gain equivalent encoding size for the videos that are in the first scenario that can cope with limited bandwidth.

## 5. Subjective Viewing Test

### 5.1. Test methods

It is well known that the peak signal-to-noise ratio (PSNR) does not always rank quality of an image or video sequence in the same way as a human being. There are many other factors considered by the human visual system and the brain [9]. One of the most reliable ways of assessing the quality of a video is a subjective evaluation of the Mean Opinion Score (MOS). MOS is a subjective quality metric obtained from a panel of human observers. It has been regarded for many years as the most reliable form of quality measurement technique [10].

### 5.2. Testing Materials and Environments

The videos are displayed on a 17-inch FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of $1680 \times 1050$ pixels. The videos are displayed with resolution of $144 \times 176$ pixels in the centre of the screen with a black background with duration of 66 seconds for Highway video and 10 seconds for Akiyo, Foreman, News and Waterfall videos.

The MOS measurements are used to evaluate the video quality in this study and based on the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). We use a lab with controlled lighting and set-up according to the ITU-R recommendation. The score grades in these methods range from 0 to 100. These ratings are mapped to a 5-grade discrete category scale labeled with Excellent, Good, Fair, Poor and Bad [4].

The subjective experiment was conducted at Blekinge Institute of Technology in Sweden. The participants of thirty non-expert test subjects were 25 males and 5 females. They were all university students and their ages range from 20 to 35. The users observed two scenarios for displaying the videos; the first scenario is to display the videos for low bit rate for the original videos and the second scenario by implementing our proposed technique with linear Interpolation for high bit rate, where the playing rates for both scenarios are 30 frames per second.

The amount of data gathered from the two subjective experiments groups with respect to the opinion scores that were given by the individual viewers. Concise representations of this data are achieved by calculating conventional statistics such as the mean score and 95 % confidence interval [4].

## 6. Experimental Results

A panel of users are evaluating the two scenarios according to the mean opinion score (MOS) measurements. In the first scenario the original video are decoded with a bit rate of 64 kbps. The second scenario (the proposed scenario), where linear interpolation are used to reconstruct the pixels outside the ROI that are decoded with a bit rate of 128 kbps, as shown in Fig. 8.

For Highway videos the observers evaluate both scenarios within the same score range, as an indicator that the observer had a similar opinion to the quality of the videos. For Waterfall video, the MOS for both scenarios is shows that is larger than 3 and lower than 4, while the first scenario is slightly higher than the second scenario.

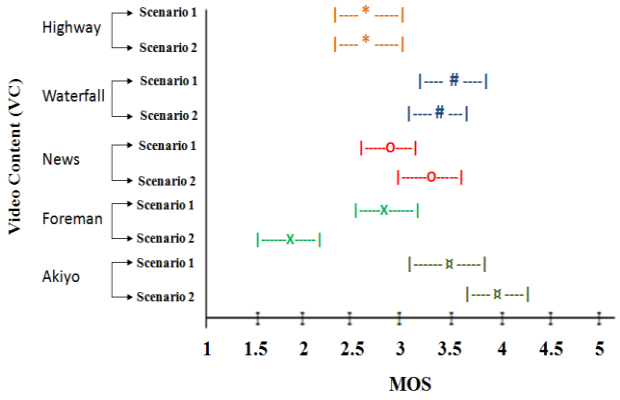


**Fig. 8:** The MOS for different videos content and for different scenarios.

For News and Akiyo videos, the MOS for the second scenario is better score than the first scenario, as the ROI fit in the motion region as shown in Fig. 5 and Fig. 7, respectively. Therefore, the observers did not manage to recognize the effect of interpolation on the video frames.

For Foreman video, the MOS for the second scenario is shows the worst score than the first scenario, as the observers manage to recognize the effect of interpolation, although the first scenario is been encoded with low bit rate.

## 7. Conclusion

In this study, we proposed an adaptive scheme to identify and extract the appropriate slice (ROI) by computing the Sum of Squared Differences (SSD) on the server side and drop the pixels that are outside the ROI. The receiving video on the mobile device will reconstruct the dropping pixels that are outside the ROI by using linear interpolation and from the reference frames.

In general, it seems that the highest SSD results as an indicator to the important region in the video frames. A panel of users observers and evaluates the two scenarios by using the MOS measurements. The user's panel observed the first scenario with a low bit rate for the original videos and the second scenario with a high bit rate (the proposed adaptive scheme for estimate the position of ROI). It is been notice from that, the MOS score is the highest for the videos like Waterfall, Akiyo and News, while for a video like Foreman, the MOS is the lowest score as it is not easily to estimate the ROI as the video frames are shaking all the time.

Even the quality of the videos is degraded; it could still be a satisfactory technique for reducing the encoding size of the streaming video over limited bandwidth.

## Acknowledgements

We would like to thank the students at Blekinge Institute of Technology, Sweden for participating in the subjective experiments. We also would like to thank the Swedish Knowledge Foundation for sponsoring a part of this work through the project QoEMoS (d-nr 2008/0520).

## References

[1] KAUR, A., P. SIRCAR and A. BANERJEE. Interpolation of lost frames of a video stream using object based motion estimation and compensation. In: *Annual IEEE India Conference, 2008. INDICON 2008*. Kanpur: IEEE, 2009, pp. 40-45. ISBN 978-1-4244-3825-9. DOI: 10.1109/INDCON.2008.4768798.

[2] SANCHEZ, G., F. SAMPAIO, R. DORNELLES and L. AGOSTINI. Efficiency evaluation and architecture design of SSD unities for the H.264/AVC standard. In: *VI Southern Programmable Logic Conference (SPL), 2010*. Ipojuca: IEEE, 2010, pp. 171-174. ISBN 978-1-4244-6309-1. DOI: 10.1109/SPL.2010.5483019.

[3] SCHWARZ, H., D. MARPE and T. WIEGAND. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE transactions on circuits and systems for video technology: a publication of the Circuits and Systems Society*. 2007, vol. 17, iss. 9, pp 1103-1120. ISSN 1051-8215. DOI: 10.1109/TCSVT.2007.905532.

[4] ITU-R BT.500-11. *Methodology for the Subjective Assessment of the Quality of Television Pictures.* International Telecommunication Union. 2002. Avaible at: http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-11-200206-S!!PDF-E.pdf.

[5] CHANG, J.-Y. and H.-L. CHEN. Service-oriented bandwidth borrowing scheme for mobile multimedia wireless networks. In: *ePress UTS Publishing* [online]. 2006. Available at: http://epress.lib.uts.edu.au/dspace/handle/2100/131?show=full.

[6] LEE, J.-H. and C. YOO. Scalable ROI Algorithm for H.264/SVC-Based Video Streaming. In: *IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas: IEEE, 2011, pp. 201-202. ISBN 978-1-4244-8711-0ISSN 2158-3994. DOI: 10.1109/ICCE.2011.5722540.

[7] SHUXI, L., S. YONGCHANG and X. YANG. Method of Adjustable Code Based on Resolution Ratio of Spatial Domain in Surveillance Region of Interest. In: *International Conference on Multimedia Technology (ICMT)*. Ningbo: IEEE, 2010, pp. 1-4. ISBN 978-1-4244-7871-2. DOI: 10.1109/ICMULT.2010.5631260.

[8] CZUNI, L., G. CSASZAR and A. LICSAR. Estimating the optimal quantization parameter in H.264. In: *18th International Conference on Pattern Recognition (ICPR)*. Hong Kong: IEEE Computer Society, 2006, pp. 330-333. ISSN 1051-4651. ISBN 0-7695-2521-0. DOI: 10.1109/ICPR.2006.502.

[9] MARTINEZ-RACH, M., O. LOPEZ, P. PINOL, M. P. MALUMBRES, J. OLIVER and C. T. CALAFATE. Quality assessment metrics vs. PSNR under packet loss scenarios in MANET wireless networks. In: *Proceedings of the International Workshop on Mobile Video*. New York: ACM, 2007, pp. 31-36. ISBN 978-1-59593-779-7. DOI: 10.1145/1290050.1290058.

[10] MARTINEZ-RACH, M., O. LOPEZ, P. PINOL, M. P. MALUMBRES, J. OLIVER and C. T. CALAFATE. Behavior of quality assessment metrics under packet losses on wireless networks. In: *XIX Jornadas de Paralelismo* [online]. Castellon, 2008. ISBN 978-84-8021-676-0. Available at: http://www.grc.upv.es/calafate/download/martinez_jornadas08.pdf.

[11] INOUE, M., H. KIMATA, K. FUKUZAWA and N. MATSUURA. Partial delivery method with multi-bitrates and resolutions for interactive panoramic video streaming system. In: *IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas: IEEE, 2011, pp. 891-892. ISSN 2158-3994. ISBN 978-1-4244-8711-0. DOI: 10.1109/ICCE.2011.5722922.

[12] HRARTI, M., H. SAADANE, M. LARABI, A. TAMTAOUI and D. ABOUTAJDINE. Adaptive quantization based on saliency map at frame level of H.264/AVC rate control scheme. In: *3rd European Workshop on Visual Information Processing (EUVIP)*. Paris: IEEE, 2011, pp. 61-66. ISBN 978-1-4577-0072-9. DOI: 10.1109/EuVIP.2011.6045539.

[13] CHI, M.-C., Ch. MEI-JUAN, Y., CHIA-HUNG and J. JYONG-AN. Region-of-Interest Video Coding based on Rate and Distortion Variations for H.263+. *Signal Processing: Image Communication*. 2008, vol. 32, iss. 2, pp. 127-142. ISSN 0923-5965. DOI: 10.1016/j.image.2007.12.001.

[14] LEE, S. H. and N. I. CHO. Low bit rates video coding using hybrid frame resolutions. *IEEE Transactions on Consumer Electronics*. 2010, vol. 56, no. 2, pp. 770-776. ISSN 0098-3063. DOI: 10.1109/TCE.2010.5506000.

[15] WANG, H. and K. EL-MALEH. Joint Adaptive Background Skipping and Weighted Bit Allocation for Wireless Video Telephony. In: *International Conference on Wireless Networks, Communications and Mobile Computing*. Maui, IEEE, 2005, vol. 2, pp. 1243–1248. ISBN 0-7803-9305-8. DOI: 10.1109/WIRLES.2005.1549590.

[16] PENG, Y.-Ch., Ch. HUNG-AN, L. CHIA-KAI, H., CHEN and K. CHANG-JUNG. Integration of image stabilizer with video codec for digital video cameras. In: *IEEE International Symposium on Circuits and Systems (ISCAS'05)*. Kobe: IEEE, 2005, vol. 5, pp. 4871-4874. ISBN 0-7803-8834-8. DOI: 10.1109/ISCAS.2005.1465724.

[17] YUV Video Sequences: YUV Sequences. In: *Arizona State University* [online]. 2008. Available at: http://trace.eas.asu.edu/yuv/index.html.

[18] FFmpeg. *FFmpeg* [online]. 2012. Available at: http://www.ffmpeg.org.

# About Authors

**Hussein Muzahim AZIZ** has obtained his Licentiate degree in Computer System Engineering in 2010, from the School of Computing at Blekinge Institute of Technology (BTH) in Sweden. He is currently a Ph.D. student at the same School and Institute. His research interest is real time video streaming.

**Marcus FIEDLER** has obtained his M.Sc. and Ph.D. degrees in Electrical Engineering with focus on Information and Communication Technology (ICT) from the University of the Saarland, Saarbrücken, Germany, in 1993 and 1998, respectively. Since then, he has been with Blekinge Institute of Technology (BTH) in Sweden. He is a Professor of Teletraffic Systems and leading the Communication and Computer Systems Research Laboratory (CCS) at the School of Computing (COM) at BTH. His research interest focuses on the quality of experience, performance modeling and analysis, and future Internet. He is also the co-chair of the Future Internet Cluster of the European Commission.

**Hakan GRAHN** is a professor of Computer Engineering at Blekinge Institute of Technology (BTH) in Sweden. He received a M.Sc. degree in Computer Science and Engineering in 1990 and a Ph.D. degree in Computer Engineering in 1995, both from Lund University. His main interests are computer architecture, multicore systems, parallel computing, and performance evaluation. He has published more than seventy papers on these subjects. Since January 2011, he is the Dean for research at the Faculty of Engineering. He is a member of the ACM and the IEEE Computer Society.

**Lars LUNDBERG** is a professor of Computer Systems Engineering at Blekinge Institute of Technology (BTH) in Sweden. He received a M.Sc. degree in Computer Science from Linköping University (Sweden), and a Ph.D. from Lund University (Sweden). Professor Lundberg has had a number of tasks at BTH, including being the Dean of the technical faculty for six years, and heading a department with more than 100 people for five years. He is currently the research coordinator at the School of Computing, and he was heading a research group called Communication and Computer Systems Research Laboratory (CCS). He has published more than 100 papers in international journals and conferences. His research interests include real-time systems, high-performance processing, software engineering and cloud computing.