Network-Aware H.264/AVC Region-of-Interest Coding for a Multi-Camera Wireless Surveillance Network*

Pierpaolo Baccichet**, Xiaoqing Zhu, and Bernd Girod

Information Systems Laboratory,
Stanford University, Stanford, CA 94305
{bacci,zhuxq,bgirod}@stanford.edu

Abstract. Surveillance video is characterized by short periods of activity in a small region of the image, and long periods of inactivity. When transmitting such video sequences over a wireless network, it is important to dynamically adapt the encoding of each stream according to both the video content and the underlying network conditions. We propose a video surveillance system that exploits H.264/AVC compliant region-of-interest (ROI) coding for more efficient representation of the captured video signal. Information exchange between the application and transport layers is used to guide dynamic selection of reference pictures for temporal prediction, as well as adaptation of encoding quantization parameter. Performance gains resulting from the various proposed schemes are confirmed in simulations of an 802.11 surveillance network comprising 7 cameras.

Index Terms - Video surveillance, H.264/AVC, cross-layer design, Region of Interest.

1 Introduction

The combination of decreasing cost in video cameras and flexible deployment of ad hoc networks finds an appealing application in multi-camera wireless surveillance networks. Such application requires content adaptation to efficiently represent the video signal and cross-layer information exchange to rapidly counteract congestion and transmission errors over the wireless network.

Video surveillance sequences typically contain short periods of activity confined in a small region of the image and long periods of inactivity. These characteristics can be exploited to reduce the traffic over the network, by analyzing the video content and encoding only important frames and regions. In [1], for instance, ROI coding in JPEG2000 is used for video transcoding in a surveillance system.

Multiple simultaneous video streams can easily congest a shared wireless network. It is therefore important to dynamically adapt encoding parameters to observed network conditions. The encoder quantization parameter (QP) can be adjusted to control the source rate, whereas reference frames for temporal prediction can be chosen to prevent error propagation, in case packets are lost due to congestion [2].

In this work, we propose an H.264/AVC [3] standard compliant solution which combines *frame rate adaptation* with *region-of-interest* (ROI) coding, so as to represent the surveillance video content more efficiently. We introduce a simple preprocessor to identify regions of interest and to signal them by means of a standard-compliant *Flexible Macroblock Ordering* (FMO) mapping function. Delay and loss statistics estimated from acknowledgment packets at the transport layer are used at the application layer for QP adaptation and *reference picture selection* (RPS). The system block diagram is presented in Fig. 1.

In the rest of the paper, we describe in Section 2 details of the proposed frame rate adaptation and ROI coding scheme. Network-aware QP adaptation and reference picture selection are explained in Section 3. Experimental results illustrating the performance gains of the proposed schemes are discussed in Section 4.

2 Region-of-Interest Determination and Frame Rate Adaptation

In a conventional camera network, the static background regions in a surveillance scene would still be encoded and transmitted, even if they do not contribute significant information. This waste of resources may penalize the performance of the overall system, especially when multiple active video streams need to be supported over a bandwidth-limited wireless network. Alternatively,

* This work has been supported, in part, by NSF Grant CCR-0325639.

** On leave from I.E.I.I.T. - Consiglio Nazionale delle Ricerche, Italy and with support, in part, by MIUR (Italian Ministry of Education and Research) under Research Project PRIMO - "Reconfigurable platforms for wideband wireless communications".

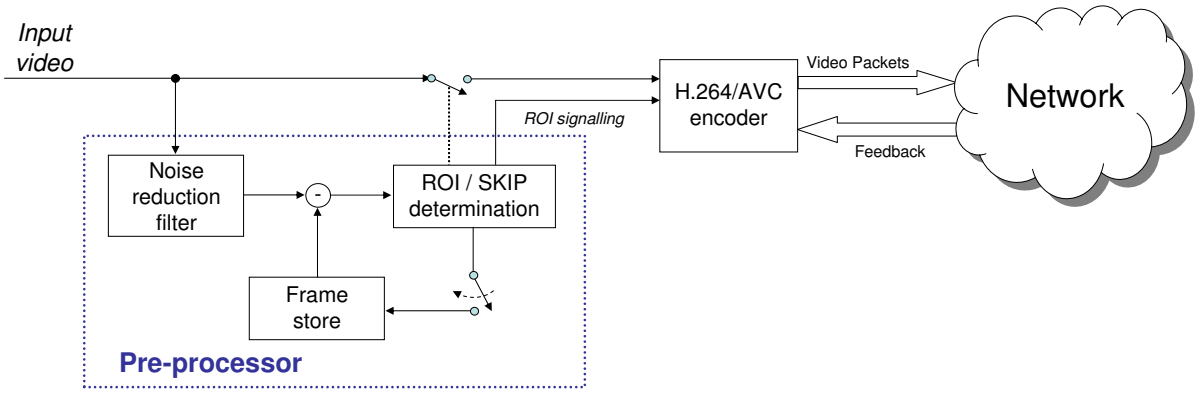

Fig. 1. Blockdiagram of the proposed video-surveillance scheme. A preprocessor determines if the current frame contains significant information to be encoded by the H.264/AVC encoder. Feedback information from the network allows adaptation of reference picture and quantization parameter during encoding, to counteract errors and congestion.

one can try to identify background regions in a video frame, and avoid encoding and transmitting them explicitly. If an entire frame does not contain significant changes with respect to the previous one, it can be *skipped* completely, resulting in reduced frame rate. In the proposed scheme, the automatic determination of the regions of interests and frame rate adaptation are performed in a preprocessing stage, involving only simple operations on the un-encoded input video signal. As illustrated in Fig. 1, the preprocessor comprises of the following procedures:

- *Noise reduction*
  A simple 3x3 spatial median filter is applied to the input video signal, to alleviate the impact of camera noise before the frame is compared with the filtered version of the last transmitted picture in the frame store.
- *ROI determination.*
  An activity map is computed for each input frame to capture the perceptual importance of each macroblock. This is achieved by calculating the *Mean Absolute Difference* (MAD) between the pixels in the filtered version of current and latest encoded frames. We then threshold these MAD values to identify a set of macroblocks corresponding to the region-of-interest. An illustration of the procedure is shown in Fig. 2.
- *Frame rate adaptation*
  If the MAD values of all macroblocks in an image fall below the threshold, the entire frame is skipped, i.e., not passed to the encoder. In effect, the encoding frame rate is reduced.
- *ROI/Frame-Skip signalling.*
  The ROI determination and frame-skip decisions are signalled to the H.264/AVC encoder in a standard-compliant fashion. The timing information in the system layer is used to indicate the skipping of a frame, whereas the regions of interest are signalled as a set of rectangles using *Flexible Macroblock Ordering* (FMO) Type 2, and recorded in a *Picture Parameter Set* (PPS).

3 Network-Aware Adaptation

When acknowledgment(ACK) packets are sent for each received video packet, this feedback information from the network can also be used for encoder adaptation. In the following, we discuss two mechanisms to avoid network congestion and to prevent error propagation at the decoder in case of packet losses:

- *QP adaptation*
  To avoid network congestion, the encoder adjusts the *quantization parameter* (QP) based on the statistics reported in the ACK packets. A coarser QP is chosen when the observed end-to-end delay or the number of unacknowledged packets exceeds a pre-defined threshold. Conversely, the video quality can be improved by selecting a finer QP when acknowledgments are received with a short round trip delay. In order to avoid excessive fluctuations in the visual quality of consecutive images, the step size of QP adaptation is confined to one unit per encoded frame.

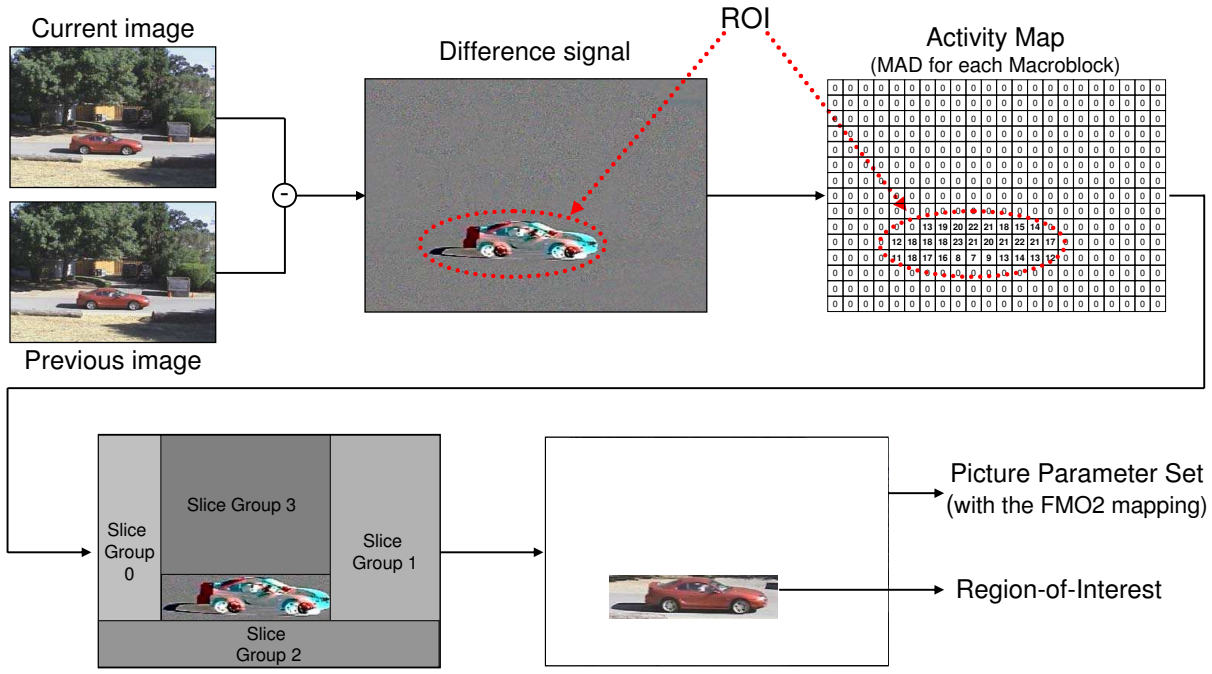

Fig. 2. The ROI/Skipping determination procedure considers the difference between the current and the previously encoded frames to compute an activity map. The ROI is then composed by the macroblocks where the activity exceeds a given threshold. If no ROI is detected, the frame is skipped.

– *Reference picture selection*
For live video coding with network feedback, Reference Picture Selection (RPS) is an effective technique to *recover from* error propagation at the decoder in case of packet losses [4]. For surveillance video, especially when frames containing only the background are skipped during encoding, the typical round trip time is shorter compared to frame intervals. In this case, RPS can be used to *prevent* error propagation altogether, by only using the acknowledged frames as references at the encoder. Compared to inserting a fixed ratio of intra refresh macroblocks, RPS can adapt to packet losses dynamically, and incurs less penalty in coding efficiency when the network is not congested.

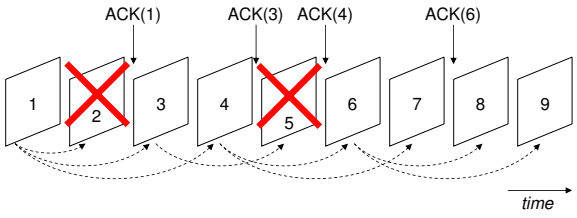

Fig. 3. *Reference Picture Selection (RPS)* avoids the insertion of intra coded pictures to counteract transmission errors.

4 Experimental Results

We evaluate the performance of the proposed schemes using surveillance video sequences captured at 7 different locations around a residential area on the campus of Stanford University. The layout of the area and the camera positions are illustrated in Fig. 4, where Node 0 acts as the central location for video collection. The available transmission links in the network are marked with dashed lines. The recorded video sequences last for 90 seconds, showing a car and a bike riding in opposite directions on the path. The topology of the simulated network is matched with the real camera positions, and the simulated wireless nodes follow the 802.11b protocol in ad hoc mode. For this simple and static wireless network, routes are set manually.

Video sequences are captured with frame size of 352×240 pixels and frame rate of 15 fps. A preprocessor as described in Section 2 is applied to the unencoded video signal for frame rate adaptation and ROI identification. The selected pixel regions are then encoded by the H.264/AVC reference codec [5],

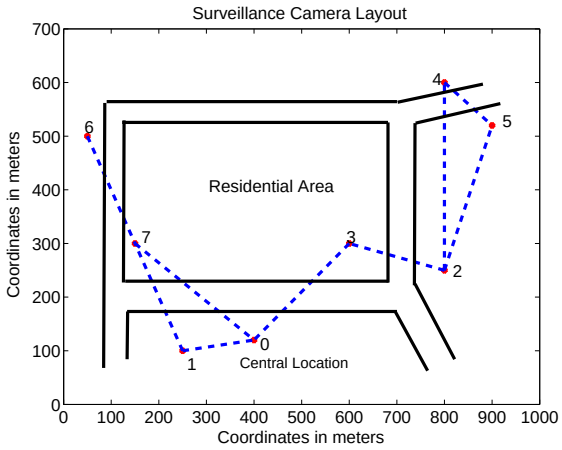

Fig. 4. Camera layout and network topology.

which runs in conjunction with the network simulator `ns-2` [6], mimicking a live coding and streaming scenario. To protect packet from losses, unacknowledged video packets are retransmitted up to 3 times at the transport layer. At the application layer, all frames are predictively encoded in an IPPP... structure. The playout deadline is set to 0.5 second.

4.1 Encoder performance

Figure 5 compares the encoding performance of three schemes: conventional H.264/AVC coding, the scheme with frame rate adaptation (named as *frame-skipping*) and the one combining frame skipping with ROI coding (named as *ROI*). Note that the encoded video quality is measured in terms of *peak-signal-to-noise-ratio* (PSNR) within the ROI identified by the preprocessor, as this better corresponds to the visual perception of received surveillance videos. Compared to conventional H.264/AVC, the *frame-skipping* and *ROI* schemes achieve rate savings in the range of 5-50% and 25-90% respectively, for various quality levels and cameras. This is achieved by omitting all background regions that can be efficiently reconstructed at the decoder by repeating the corresponding blocks from the previous frame.

4.2 Performance over the network

When transmitted over the wireless network, the rate savings from the *frame-skipping* and *ROI* schemes can potentially lead to reduction of network congestion. This is illustrated in the trace of encoded frame sizes in Fig. 6 for Camera 2. The *ROI* scheme consistently yields fewer and smaller packets over time, without sacrificing the encoded video quality.

When the wireless network is congested, certain video packets may be discarded at intermediate nodes due to lack of transmission opportunities, or dropped at the decoder if they arrive later than the playout deadline. Figure 7 shows the loss ratio corresponding to the three schemes, as a function of the fixed quality level. As expected, since the conventional H.264/AVC scheme introduces highest amount of traffic, it leads to highest packet loss ratios for each quality level. The *ROI* scheme, on the other hand, can sustain the highest quality level among the three, without incurring significant losses over the network.

Figure 8 plots the decoded video quality as a function of encoded rate, for the three schemes using fixed QP encoding without RPS. To mitigate the effect of error propagation due to packet losses, one line of macroblocks is intra coded every 4 frames. Nevertheless, at high rates the decoded video quality drops abruptly, due to high packet loss ratio in a congested network. The *frame-skipping* and *ROI* schemes therefore outperform the conventional H.264/AVC scheme by 2 dB in terms of highest sustainable video quality. It can also be noted that the *ROI* scheme achieves the same optimal quality at a lower rate. However, its encoded bitstream is more sensitive to losses, therefore the quality drop also occurs at a lower rate.

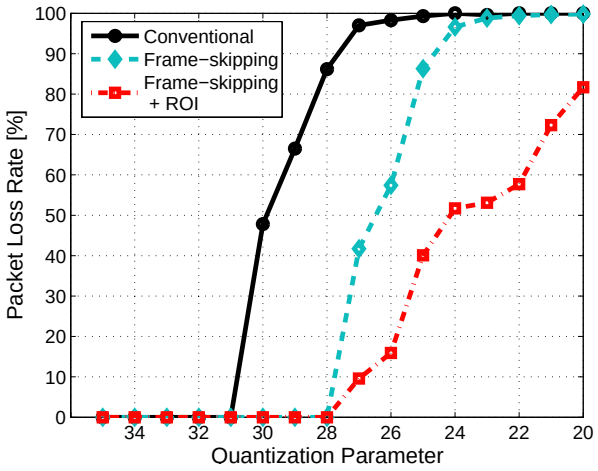

Fig. 7. Packet loss ratio versus quantization parameter (QP) for all three schemes, at Camera 5.

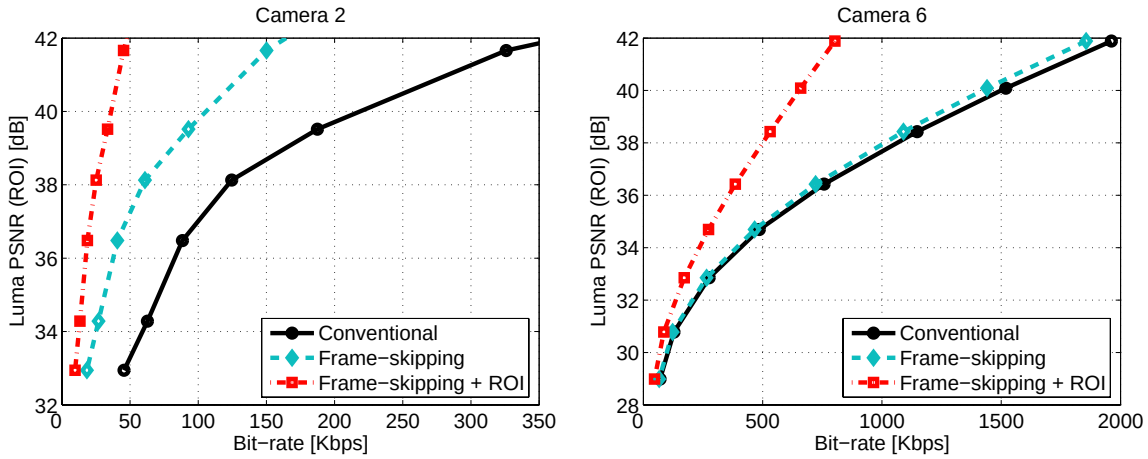

Fig. 5. Encoding performance comparison for the three considered schemes for two cameras.

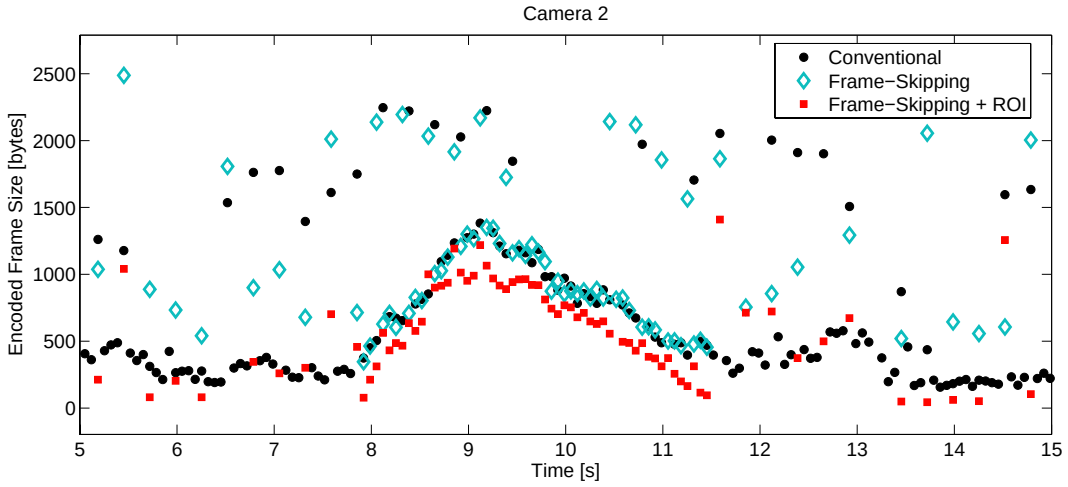

Fig. 6. Encoded frame size over time obtained with a fixed QP of 30 for all three schemes. The proposed schemes avoid encoding many frames during inactive periods. Compared with *frame-skipping*, the *ROI* scheme leads to smaller frame sizes during active periods.

4.3 Benefit of RPS and QP adaptation

In this section, we investigate the additional performance gains introduced by RPS and QP adaptation.

Figure 9 compares the performance of the intra-refresh-based scheme against the proposed scheme using RPS, in terms of decoded video quality versus encoded rate, with fixed-QP encoding and *frame-skipping* or *ROI*. As no macroblocks are forced into intra-coding mode, RPS can further reduce the encoding rate by 60-70%, hence achieving a higher sustainable video quality by 1-2 dB. On the other hand, RPS is more susceptible to increase in network congestion, as coding efficiency degrades quickly when the reference pictures lag too far behind the frames to be encoded. Consequently, optimal video quality for the RPS schemes are achieved at a lower rate in comparison with the intra-refresh scheme.

Finally, the benefit of QP adaptation is shown in Tab. 1, where we compare the decoded video quality resulting from fixed or adaptive QP schemes, both using RPS and ROI coding. QP adaptation improves the average video quality at 5 out of the 7 cameras. It should also be noted that even though all the cameras follow the same QP adaptation procedures as described in Section 3, the network congestion experienced by the active periods of each camera is different. In particular, the active scenes of Cameras 4, 5 and 6 occur at around

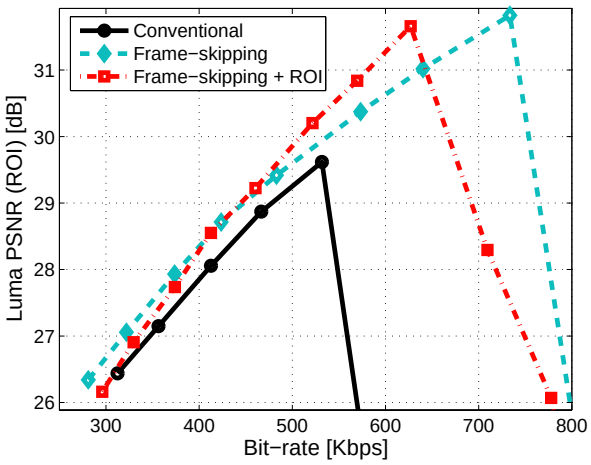

Fig. 8. Decoded video quality in PSNR versus encoded rate at Camera 5, for all three schemes, using fixed-QP encoding and intra refresh. Similar relative performances are observed at other cameras as well, with the performance gain of *frame-skipping* and *ROI* over the conventional H.264/AVC ranging from 1 to 4 dB in terms of highest sustainable video quality.

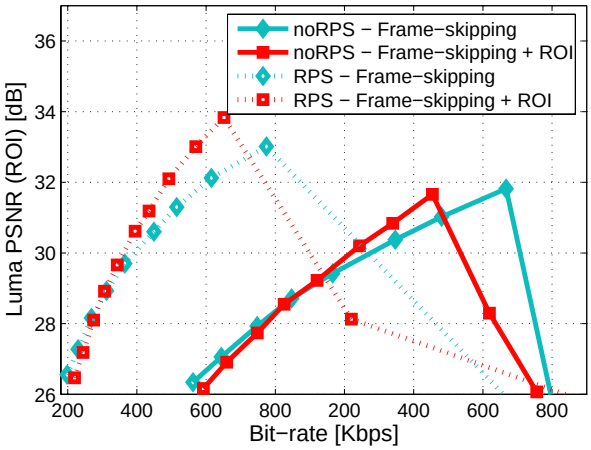

Fig. 9. Advantage of RPS versus a fixed intra refresh policy. At low bit-rates RPS allows to avoid error propagation without introducing an excessive overhead, thus increasing the highest sustainable quality.

the same time, leading to higher network congestion and lower video qualities at these cameras, as a reaction from the congestion-avoidance QP adaptation procedure.

5 Conclusions

In this paper, we propose an H.264/AVC compliant solution for content and network adaptive coding of surveillance video sequences. Automatic frame rate adaptation and ROI determination in the preprocessing stage leads to higher encoding efficiency. Network-aware adaptation of reference pictures and quantization parameter during live encoding further mitigates the impact and possibility of packet losses due to congestion. Possible performance gains are demonstrated in the simulation results of a 7-camera wireless surveillance network. ROI coding and frame rate adaptation outperforms the conventional scheme by 2 dB in terms of highest sustainable video quality over the network. Additional improvement of 1-2 dB is achieved by reference picture selection, and 0.4-8.1 dB over different cameras with QP adaptation .

Camera	1	2	3	4	5	6	7
Fixed QP	30.7	33.8	33.0	33.7	33.8	33.1	32.5
Adaptive QP	38.8	39.3	38.2	32.1	33.1	33.5	39.1
Gain	8.1	5.5	5.2	-1.6	-0.7	0.4	6.6

Table 1. Decoded video quality in PSNR (dB) for different cameras using fixed or adaptive QP, RPS and ROI coding. The fixed QP is chosen at 27, corresponding to the highest sustainable video quality for most of the cameras. The adaptive QP scheme is initialized with a QP of 30, and the thresholds for increasing/decreasing QP are tuned empirically given the playout deadline.

References

1. Hata, T., Kuwahara, N., Nozawa, T., Schewenke, D.L., Vetro, A.: Surveillance system with object-aware video transcoder. Proc. Multimedia Signal Processing, (MMSP-05), Shanghai, China (2005)
2. Liang, Y.J., Girod, B.: Network-Adaptive Low-Latency Video Communication over Best-Effort Networks. IEEE Transactions on Circuits and Systems for Video Technology **16** (2006) 78–81
3. Joint Video Team (JVT) of ISO/IEC MPEG & ITU–T VCEG: Draft ITU–T recommendation and Final Draft International Standard of Joint Video Specification (ITU–T Rec. H.264 — ISO/IEC 14496/10 AVC - JVT–G050r1.doc). ISO/IEC MPEG & ITU–T VCEG (2003)
4. Fukunaga, S., Nakai, T., Inoue, H.: Error Resilient Video Coding by Dynamic Replacing of Reference Pictures. Proc. of Global Telecommunications Conference, GLOBECOM **3** (1996) 1503–1508
5. Suehring, K.: (JVT JM reference software) `http://iphome.hhi.de/suehring/tml/`.
6. (NS-2) `http://www.isi.edu/nsnam/ns/`.