Contemporary Engineering Sciences, Vol. 7, 2014, no. 32, 1739 - 1745
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2014.411233

Video Transmission Using Multi-Level Content Aware Compression Based on Object Detection

Artem Lenskiy and Soonuk Seol

School of Electrical, Electronics and Communication Engineering
Korea University of Technology and Education (KOREATECH)
Cheonan, Chungnam, Republic of Korea

Abstract

The computational power of network nodes whether it is a mobile device or a miniaturized computer increases year by year, however the capacity of wireless channels in 4G networks is still falling behind for real-time video transmission. The paper proposes utilizing the computational capacities of mobile devices to detect objects of interest and apply less compression to corresponding regions, while the rest of the video frame is compressed with a higher compression ratio. We generalized this approach on multiple compression ratios depending on the importance of detected objects. For testing purposes we implemented a transmission system that takes into account the content that is being transmitted.

Keywords: Content aware compression, Multi-level compression, Video streaming, Real-time Transfer Protocol

1 Introduction

In the past decade, the spread of mobile devices led to a wide adoption of video telephoning that has become a crucial part of mobile communication. Besides telecommunication providers that offer video calls as a one of the basic communication services along with SMS and voice calls, a number of mobile apps support video communications over 3G, 4G and Wi-Fi networks. Most of the apps as far as we concerned are based on H.264 standard [1]. In this paper we implement a content aware compression technique that potentially can be built on top of H.264 following the similar method described in [2], to further compress video stream keeping important parts of the video at a higher quality. The novelty of our video communication system is the generalization of a binary content aware compression levels to a multilevel compression algorithm such that different regions are compressed with different compression ratios. Somewhat similar approach was proposed earlier [3]. The authors proposed to use non-uniform downscale using salient features before applying any compression algorithms. Those regions that are less salient are downscaled more. An important advantage of this approach is the possibility to easily embed it into any compression pipeline as preprocessing step. The disadvantage of this approach consists in a necessity of transmitting additional information that include a non-uniform grid coordinates and a difference image. That itself occupies part of the bandwidth. An interesting approach of detecting region of interest using salient was proposed in [4]. The authors also used face detection algorithm to detect region of interests. It was demonstrated that applying less compression to region of interests led to better subjective image quality. The ROI based compression scheme finds applications in the field of medical image processing and communication. Similar to previously discussed salient based compression schemes were proposed in [5, 6] for transmitting CT scans images.

In this paper, in order to demonstrate the proposed multi-level compression for video transmission we implement a video communication program. Due to the fact that the most common use of video communication in mobile devices is video chatting, we apply an object detection algorithm for faces and facial features detection. Such image objects serve as regions of interest in our implementation.

2 The Proposed System

The proposed system can be divided into three parts; (a) the sender side processing, (b) the receiver side processing and (c) the communication protocol. A video communication application combines both the sender and the receiver. In this section we describe each part of the system.

The sender side processing consists of data preparation and packet assembling phases. Prior to packet transmission, the sender prepares the payload data. Depending on a compression method the data preparation method varies. In our system we divide the video frame into non-overlapping blocks as shown in figure 1. Each block is compressed independently to others. The compression ratio depends on the importance of the region. We apply an object detection algorithm to define the importance of the region. The details of object detection are given in the following sections. The compression parameters are sent as a part of each block’s header and can be different for each block. In addition to the compression parameters the header conveys a sequence number which is used by the receiver to identify the position of the block in the frame. The sequence number is incremented by one for each block.

The client side is responsible for assembling packets into a complete video frame. Compression parameters are extracted from the packet's header and then the payload data is decompressed accordingly. Then, based on the sequence number, extracted from the packet's header, the position of the block is calculated as follows.

$$b_x = mod(n \cdot s, w),$$
$$b_y = s \cdot \left\lfloor \frac{n \cdot s}{w} \right\rfloor,$$

where w is an image width, s is a block size and n is a sequence number.

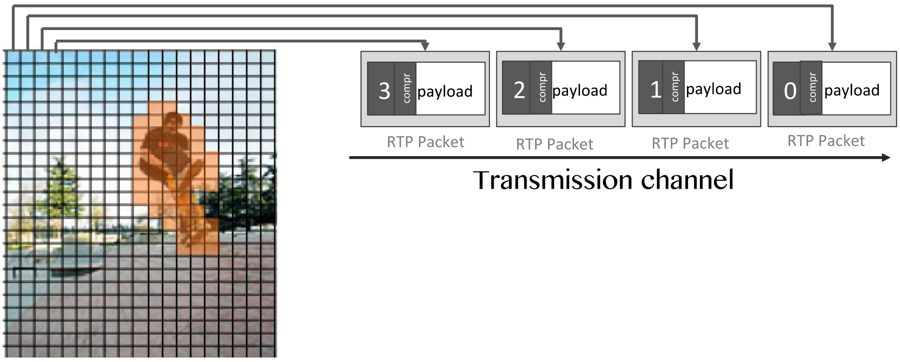


Figure 1. Dividing of a video frame into blocks.

The proposed application layer communication protocol is built on top of RTP and UDP. We apply RTP to be certain that packets are received in the right order as well as to distinguish which packets belong to which frame. The header of our protocol, along with block sequence number, carries block size as a part of compression parameters. Figure 2 shows the packet structure.

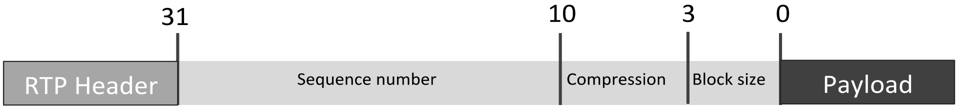


Figure 2. Application layer protocol header format.

We allocate 3 bits to carry a block size, thus eight different block sizes can be represented {1, 2, 4, 8, 16, 32, 64, 128}. The next 8 bits are allocated to store compression parameters. The compression parameters vary depending on the importance of the payload. For the sequence number we allocated 21 bits, which accommodates 2,097,152 blocks.

3 Experiments

We test the proposed system in two different scenarios. In the first scenario, the sender subscribes to an online TV broadcasting server. The incoming video stream is then processed and resent to the receiver. During the processing step important regions are detected. The compression is only applied to the less important regions. For object detection we applied a cascade of boosted classifiers working with Haar-like features [7]. This object detection algorithm is capable of detecting objects in real-time. In our experiment for the region of interest we chose facial region. On the left side of figure 3, we demonstrate frames taken from the original video and the frames on the right are those received by the receiver. As it could be seen the quality of the background has degraded, especially it is noticeable around subtitles and TV channel logo. However, the quality of the faces has not changed. This makes the compression less noticeable due to the fact that humans focus more on faces than anything else. It should be noted that we did not apply low pass filtering at the receiver side to make compression artifacts more visible. Note also that one may choose subtitles or TV channel logos as regions of interest.

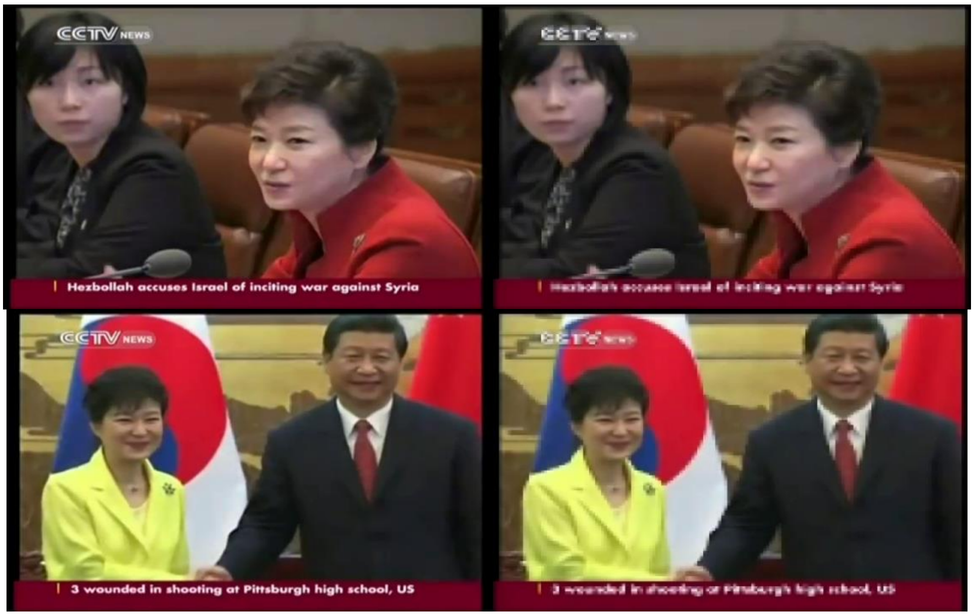

CCTV NEWS
Hezbollah accuses Israel of inciting war against Syria
CCTV NEWS
3 wounded in shooting at Pittsburgh high school, US

Figure 3. The left column depicts original video frames and the right side shows compressed frames using our content-aware compressing algorithm.

In the second scenario, we incorporate the following three levels of compression: no compression, low and high compression. We apply high compression for the background, low compression for the face and no compression for eyes. Snapshots of the original and the received video frames are shown in figure 4. Figure 5 shows zoomed and cropped regions. Notice that the eye regions are not degraded in quality. The facial region has minor drop in quality and the background is significantly distorted due to high compression ratio. We used a trivial down sampling and up sampling for compression and decompression correspondingly. We did not apply Gaussian blurring.

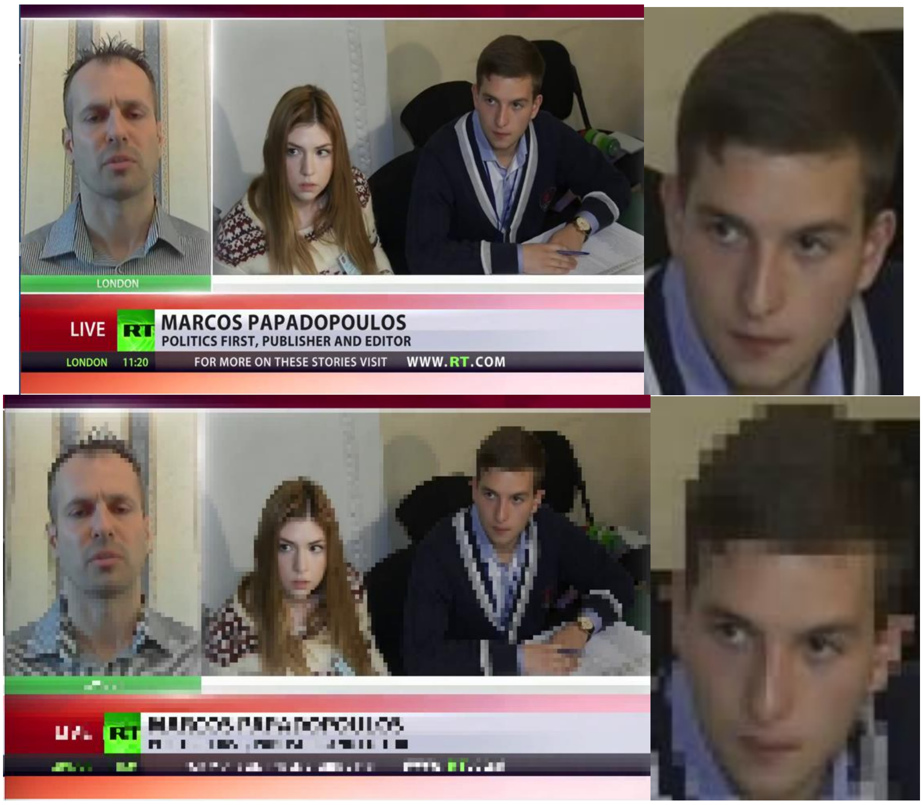
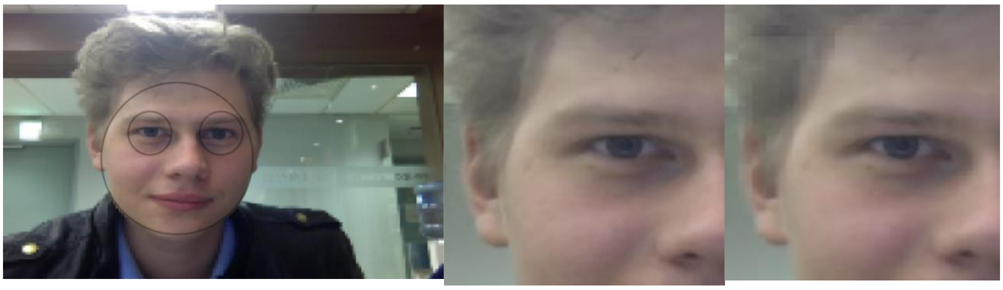

LONDON
LIVE
RT
MARCOS PAPADOPOULOS
POLITICS FIRST, PUBLISHER AND EDITOR
LONDON 11:20
FOR MORE ON THESE STORIES VISIT WWW.RT.COM

Figure 4. The top images show an original video frame and cropped out and zoomed-in ROI, and the bottom are compressed using multi-level compression.

The last experiment is conducted with a video stream captured by a web camera. We apply the three-layer compression for different parts, i.e., background, face, and the eyes as shown in Figure 5. Table 1 compares peak to noise ratio (PSNR) for images compressed with different compression schemes. Images compressed uniformly with one compression level have lower PSNR compared to multi-level compression. However, subjective quality is higher for images with multi-level compression.



(a) Compressed with three levels (b) original (c) compressed

Figure 5. Three-layer compression for the background, face and eyes.

Table 1. PSNR and frame size comparison for different compression conditions

Compression ratio	PSNR	Frame size (bytes)
1	100	1,440,600
16 (one level)	27.54	90,038
8.70 (one level)	31.15	165,540
8.79 (three levels)	28.03	163,870

4 Conclusions

In this paper we implemented a video communication protocol that is based on a content aware compression scheme. We did not focus on the compression algorithm itself rather on the concept of applying a particular compression algorithm for a multi-level compression based on the importance of a region of interest.

In the future we are planning to investigate an adaptive adjustment of compression levels. Depending on whether a region of interest is detected or not, the quality of the background is adjusted. The total size of all ROIs will be taken into account. To maintain the occupied bandwidth at a constant level, the quality of a ROI is improved at the cost of decreasing quality of the background.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H.264/AVC video coding standard, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.13, no.7, (2003), 560 - 576.
http://dx.doi.org/10.1109/tcsvt.2003.815165

[2] F. Zund, Y. Pritch, A. Sorkine-Hornung, S. Mangold, and T. R. Gross, Content-aware compression using saliency-driven image retargeting, *in Proc. ICIP*, (2013), 1845 - 1849. http://dx.doi.org/10.1109/icip.2013.6738380

[3] Yang Liu, Zheng Guo Li, Yeng Chai Soh, Region-of-Interest Based Resource Allocation for Conversational Video Communication of H.264/AVC, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.18, no.1, (2008), 134 - 139. http://dx.doi.org/10.1109/tcsvt.2007.913754

[4] S. Han, N. Vasconcelos, Image Compression using Object-Based Regions of Interest, *IEEE Intern. Conf. on Image Processing*, (2006).
http://dx.doi.org/10.1109/icip.2006.313095

[5] S. B. Gokturk, C. Tomasi, B. Girod, C. Beaulieu, Medical image compression based on region of interest with application to colon CT images, *Engineering in Medicine and Biology Society*, *Intern. Conf of the IEEE*, vol.3, (2001). http://dx.doi.org/10.1109/iembs.2001.1017274

[6] V. K. Bairagi, A. M. Sapkal, Automated region-based hybrid compression for digital imaging and communications in medicine magnetic resonance imaging images for telemedicine applications, *Science, Measurement & Technology, IET*, vol.6, no.4, (2012), 247 - 253. http://dx.doi.org/10.1049/iet-smt.2011.0152

[7] P. Viola, M. Jones, Robust Real-time Object Detection, *International Journal of Computer Vision*, (2001).

Received: October 1, 2014; Published: December 2, 2014