ENERGY EFFICIENT ADAPTIVE VIDEO COMPRESSION SCHEME FOR WVSNS

Othmane Alaoui Fdili[1,2], *François-Xavier Coudoux*[2], *Youssef Fakhri*[1,3], *Patrick Corlay*[2], *Driss Aboutajdine*[1]

[1]LRIT-CNRST URAC29, Université Mohammed V-Agdal, Rabat, Morocco;
[2] IEMN UMR 8520, Department OAE, UVHC, Valenciennes, France;
[3] LARIT équipe réseaux et Télécommunications, Université IbnTofail, Kénitra, Morocco.

ABSTRACT

This paper presents an energy efficient adaptive video compression scheme dedicated to wireless video sensor networks. This scheme based on the recent H.264/AVC standard operates using two modes, namely the standby mode and the rush mode. In the standby mode, frame rate is limited while image quality is maintained to a high level. In the rush mode, the frame rate is increased and the Flexible Macroblock Ordering coding tool provided by the H.264/AVC video compression standard is used to produce two service-differentiated macroblocks categories: region of interest (ROI) and background (BKGD). In addition, bit rate adaptation based on frequency selectivity is applied to reduce background data amount. Hence, transmission energy consumption is decreased while maintaining a high quality for ROI. Simulations proved the energy efficiency of the proposed scheme that enables to extend network's lifetime up to three times, and achieves high quality ROIs even at low bit rates.

Index Terms— Video Compression, Energy efficiency, Flexible Macroblock Ordering, H.264/AVC, WVSN.

1. INTRODUCTION

Thanks to recent developments in low-cost CMOS cameras, nodes are able to capture and process images and videos. As a result, a new type of networks has emerged called Visual Sensor Networks (VSN) [1]. This category regroups Wireless Video Sensor Networks (WVSNs) and Wireless Image Sensor Networks (WISNs). Consequently, all previous sensor-based applications can be enriched with visual data allowing more efficient wildlife monitoring, surveillance, tracking, face/posture recognition and many others abilities that need visual information [1]. Besides all these novel opportunities, new research challenges have been introduced. First of all, nodes are still in lack of processing resources (CPU and memory) despite the few improvements that have been introduced. Second, in addition to coverage area that already needs to be studied before deploying the network and during its lifetime, camera field of view must be taken into consideration as well for optimal coverage. Finally, data amount generated by each visual node is very large compared to scalar nodes. Thus, energy budget needed for processing and then transmitting data is more important and an efficient management of nodes' resources is of great importance than ever. For this purpose, image and video compression algorithms are used to take advantage of temporal and spatial correlation inherent to video signal.

It has been shown in the literature that different ways for video compression could be adopted in WVSN. In some works, authors rely on the standard itself and try to reduce energy consumption in lower layers of the protocol stack. For example, in [2], the H.264/AVC standard is used since it can achieve higher compression efficiency than any previous standards [3]. The amount of energy consumed during motion estimation is compensated by an energy efficient routing protocol called LEACH [2] in addition to a frame dropping scheme based on a recursive distortion prediction model. In [4], MPEG-4 was also employed for video compression. The authors proposed a differentiated service packets management based on the relevance of each video packet to the MPEG-4 codec. Hence, a queue management scheme with prioritization is adopted to drop packets related to intra-coded frames less frequently than others.

In other works, authors slightly modify standard codecs in order to be more appropriate to WVSN in terms of energy consumption and complexity, and also to introduce differentiated service paradigm that is strongly recommended for WVSN [5]. For instance, the video compression algorithm introduced in [6] called Modified-MPEG (M-MPEG) uses two different types of frames: Main-Frame (M-Frame) and Difference-Frame (D-Frame). Unlike MPEG, M-MPEG uses an adapted period related to background variation to transmit the M-Frame in order to reduce energy consumption. In [7], video compression is conducted by a modified version of JPEG that was first proposed in [8] and called Triangular JPEG (T-JPEG). Actually, the authors in [7] propose to divide the frame into two categories: the so-called Region Of Interest (ROI), and the background (BKGD). Then, they introduce two parameters namely the High Coefficient Count (HCC) and the Low Coefficient Count (LCC) that define the number of retained DCT coefficients for the block belonging to the ROI and the BKGD respectively. The HCC parameter allows ROI to attain a maximal quality higher than T-JPEG because this last offers a maximal coefficients number of only 36 corresponding to the upper diagonal half part of the block.

From this brief bibliographical survey, it is clear that introducing differentiated service in addition to codec adaptation are the most used schemes that have proved their efficiency. In this paper, we propose an energy efficient adaptive video compression scheme designed for WVSNs based on H.264/AVC standard. The proposed scheme exploits the H.264/AVC video standard benefits in terms of image quality and lowering the bit rate. In addition, it applies a low complexity bit rate adaptation that is the frequency selectivity (FS) for additional transmission energy efficiency. Furthermore, this scheme outputs streams containing ROI and BKGD, at two different video quality levels, with only one quantization operation that is known to be energy consuming. Finally, the resulting streams are able to be transmitted with service differentiation. Unlike [7], we prove through simulations, scheme's ability to energy efficiently transmit the video stream under bit rate constraint of 250 Kbps, which is generally the case is WVSN platforms. The rest of this paper is organized as follows. Section 2 describes different tools used in our proposed scheme. Section 3 details our energy efficient adaptive video compression scheme. In section 4, we present experimental results and finally section 5 concludes the paper.

2. H.264/AVC AND BIT RATE ADAPTATION

This section briefly presents the different tools used in our proposed scheme for efficiently compress videos in WVSN environments.

2.1. H.264/AVC standard

The H.264/AVC standard was completed in the form of final draft international standard in 2003, by the Joint Video Team (JVT) formed by ISO Motion Picture Experts Group and ITU-T Video Coding Experts Group (MPEG and VCEG). H.264 achieves up to 50% bit rate gain compared to MPEG-2 for the same video quality. This standard relies on two layers: Video Coding Layer (VCL) and Network Adaptation Layer (NAL). Each frame is processed in (16x16 pixels) macroblock units that are encoded in intra or inter mode. In H.264/AVC standard, the intra-prediction is applied to the whole image following one of the 9 optional intra-prediction modes for 4x4 or 8x8 luminance blocks, or one of the 4 optional modes for 16x16 luminance macroblocks. In Inter mode several improvements are offered by H.264/AVC standard for instance variable block size and multiple references. Afterwards, H.264/AVC offers two transforms that depend on the data to be encoded: Hadamard transform for DCs coefficients of the 16x16 residual luminance and chrominance macroblocks respectively, and 4x4 integer transform for all blocks in the residual data. The resulting coefficients are then quantized, zigzag scanned and entropy coded. H.264/AVC standard offers two entropy coding techniques, namely Context-based Adaptive Binary Arithmetic Coding (CABAC) and Context Adaptive Variable Length Coding (CAVLC) that is slightly less efficient than CABAC but can approach the entropy of the source with reduced complexity. Actually, the H.264 defines a set of three profiles: Baseline, Main and Extended. In this paper we consider the Baseline profile. For more details the reader is referred to [3].

Flexible macroblock ordering is an error resilience mechanism introduced by the H.264/AVC standard. It is actually a frame structuring that proposes the Slice Group as an intermediate level between the frame and its slice(s). Each macroblock is assigned to a slice group according to a MacroBlock-to-slice Allocation map(MBAmap). The term of FMO can only be used when dealing with more than one slice group. In addition, FMO allows the decoder to modify the order in which macroblocks were encoded since each slice is independently encoded and transmitted [9]. Consequently, a frame may still be decoded even if not all slices are presented at the decoder. H.264/AVC proposes seven FMO schemes for different macroblock allocation strategies. Among these, the FMO scheme noted Type 6 allows to fully customize the slice groups in terms of size and position [9].

2.2. Bit rate adaptation based on frequency selectivity

As previously mentioned, in our proposed scheme we use FS for bit rate adaptation of the intra-coded frames. In fact, it is a low complexity technique that permits to remove several transform coefficients based on their frequency position while maintaining the remaining coefficients with complete precision. Actually, this technique was also applied for high-definition streams transrating and proved its efficiency [10]. The number of coefficients to be maintained, in zigzag order, is specified by the Frequency Position (FP) varying from 1 DC coefficient to all of the 16 coefficients. According to the authors, the FS allowed them to produce an efficient transrater that does not need a large buffer and does not have a high computational cost [10]. Consequently, these features make this technique an interesting candidate to be used in a WVSN context.

3. ENERGY EFFICIENT ADAPTIVE VIDEO COMPRESSION SCHEME

In the following, we introduce our proposed scheme for video compression in WVSNs. For modes switching, we assume that the scheme is integrated in a general network architecture providing an Intelligent Motion Detector System (IMDS) deployed in the area of interest with video nodes (see Fig.1). When an event occurs, the system is responsible of reporting this state to video nodes that shift to the so-called rush mode. By the end of the event, the system notifies video nodes to shift back to the standby mode.

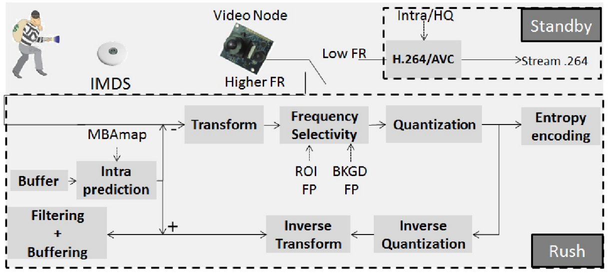

Fig. 1. Block diagram of the proposed scheme (FR: Frame Rate)

3.1. Standby mode

For energy saving, nodes capture the scene following a low frame rate since there is no important event to be reported. Each frame is then intra-coded according to a Quantization Parameter (QP) that defines quantization level of transformed macroblocks. We adopt intra-coding because we trust that it is adequate to a real-time WVSN. Actually, in [11] the authors prove that Intel-imote2 nodes are able to run the H.264 video compression standard. In addition, when comparing intra-coding to inter-coding at 30 dB, they showed that intra-coding is more energy efficient than inter coding [11]. Indeed, motion estimation algorithms are approximately 10 times more complex than intra prediction ones. Hence, even when dealing with static backgrounds, the complexity is still high comparing to intra-coding despite the decrease of the bit rate. Last but not least, intra-coding avoids any temporal propagation of errors due to possible channel distortion or network variation since each frame is independently encoded.

3.2. Rush mode

The first changing in this mode is frame rate adjustment. In fact, an increase of nodes' scene capture frequency is applied because we believe that in a monitoring application, nodes should report the event including accurate motion information. Nevertheless, the desired motion precision of the reported event varies from an application to another. In addition, frame rate increase can be considered as an error resilient mechanism since the importance of each frame is proportionally decreased. Consequently, losing a frame, which is very frequent in such error prone networks (due to congestion or channel error), will be less tragic.

The video signal is then intra-coded using FMO option of H.264 standard. FMO allows reorganizing each frame into several slice groups which are ROI and BKGD in our case. Actually, in the present work the slice groups are manually signalled to the encoder and FMO Type 6 is used as previously mentioned. Hence, for a ROI's macroblock intra prediction, only previous reconstructed macroblocks belonging to the same slice group are used. At the decoder side, as consequence, the spatial address of the next decoded macroblock is not necessary the next one is scan order. This flexibility is important for protecting the ROI from error propagation when an error occurs in the BKGD.

After the 4x4 intra prediction, a 4x4 integer transform is applied to the residual blocks. Subsequently, bit rate adaptation is done for each block of the current macroblock depending on which slice group it belongs, namely the BKGD or the ROI. In the proposed scheme we distinguish two use cases: bit rate adaptation of BKGD only, or bit rate adaptation of both BKGD and ROI. Hence, for the first use case, one low FP is used for BKGD bit rate adaptation while the ROI remains intact. For the second one, we propose to use two different FP values in the same frame: a higher one for the ROI and a lower one for the BKGD. Finally, after CAVLC entropy encoding, slice groups are separately encapsulated in NAL units for further transmission.

3.3. Computational energy overhead discussion

The computational energy of our proposed scheme does not differ significantly from H.264/AVC standard using intra-coding mode. The major additions that we propose are FMO and the bit rate adaptation. The FMO, as stated in [9], does not add a significant complexity at both encoder and decoder sides. Since computational energy consumption is related to computational complexity, we can assume that there is no additional computational energy consumption. In addition, the FS does not add any computational energy since it is only a logical control of the considered coefficients after DCT transform.

4. VIDEO QUALITY AND ENERGY EFFICIENCY EVALUATION

The simulations have been conducted using the QCIF Hall_Monitor sequence composed of 300 frames in YUV 4:2:0 format. We use JM18.4 implementation of the H.264/AVC video coding standard. Frame rates of standby mode and rush mode are set to 5 and 10, respectively. In the following, we consider four study case for comparison:

- noSCQPOnly refers to intra-coded sequence according to a given QP with no frame rate adjustment,
- SCQPOnly refers to the previous one with frame rate adjustment,
- SCROInoBRA refers to the previous one with BKGD bit rate adaptation,
- SCROIBRA refers to the previous one with bit rate adaptation of both BKGD and ROI.

For perceptually satisfying balance between horizontal and vertical frequency components, the FP of BKGD is set to 3 while when bit rate adapting the ROI it is set to 6. We use the well-known Peak signal-to-noise ratio (PSNR) as video quality metric.

4.1. Video quality evaluation

Fig.2 reports the rate distortion trade-off at the output of the decoder for the different investigated use cases and their ROIs (squares). Each point is a realisation at a given QP that varies from 40 to 10. For the sake of visibility we do not show the curve regarding the noSCQPOnly that needs a minimum bit rate of 785.7 Kbps. As can be seen, frame rate adjustment allows achieving for example a global PSNR of 32.59 dB for a bit rate of 241.2 Kbps. In addition BKGD bit rate adaptation permits to achieve more bit rate gain. When bit rate adapting the ROI as well this gain comparing to noSCQPOnly reaches 84%. Furthermore, for SCROInoBRA use case, despite reducing the overall quality, ROI is maintained to a high video quality. We observe the effect of FMO that slightly decreases ROI's (comparing ROInoBRA and SCQPOnly) PSNR values. Nevertheless, the visual quality is identical (see Fig.3). When comparing SCROIBRA and SCROInoBRA we observe that the global PSNR values (circles) are close to each other while their ROIs' PSNR values are very different. Obviously, this drop of the quality is due to the bit rate adaptation. This PSNR values decrease can be explained by the fact that the ROI is logically very rich in term of content (textures and objects) and hence it is characterized as high energy video content. Therefore, removing coefficients that are for sure non zeros values leads to this result. The visual quality was controlled in order to remain acceptable while reducing the bit rate.

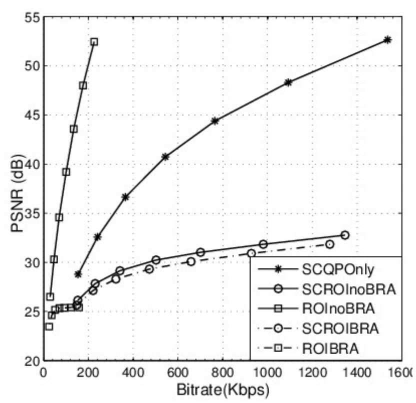

Fig. 2. Rate distortion curves

4.2. Energy efficiency evaluation

Next, we evaluate the energy efficiency of the proposed scheme in different use cases. The first experiment consists in considering a multihop network with no bit rate constraint (maximizing video quality). We study the energy consumed per frame during the transmission toward destination while

Fig. 3. Zoom on ROIs at QP=25, from left to right: SCQPOlny, SCROInoBRA, SCROIBRA

varying the number of hops. In this network, we consider that nodes are located 10 meters away from each other as it would be placed in a real video surveillance application. The energy consumed for transmitting a single bit is given by equation (1)

$$E_{Tx} = E_{elec} + \epsilon_{fs} \times d^2 \quad (1)$$

Where ϵ_{fs} is the energy consumed by the amplifier to transmit at short distance, E_{elec} is the energy dissipated in the electronic circuit to transmit and receive the signal and d is the distance between transmitter and receiver.

The total transmission energy drained for N hops is given by equation (2) [11].

$$E_N = E_{Tx} \times N + E_{Rx} \times (N - 1) \quad (2)$$

Where E_{Tx} and E_{Rx} are the energy consumed by a node for transmitting and receiving the compressed bitstream respectively. We assume that E_{Rx} is equal to E_{elec}.

Fig.4 shows the energy efficiency of the proposed scheme. In fact, one can see the effect of BKGD bit rate adaptation that decreases the amount of energy drained for frame transmission. In addition, bit rate adapting the ROI as well permit to reach a reduction of the energy consumed per frame of 30% but with a controlled visual distortion. In the second experi-

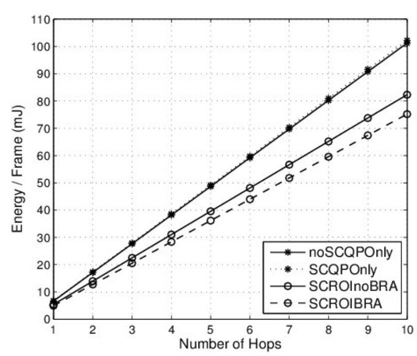

Fig. 4. Cumulative energy consumed per frame for N hops

ment, we consider a network where the source is 4 hops away from destination with bit rate constraint of 250 Kbps as in WVSN platforms.

Fig.5 depicts the maximum number of video sequences that the network was able to deliver until the death of all nodes. Note that the initial energy of nodes is chosen 10 Joules [8]. As reported on this figure, our proposed scheme, whatever the use case, achieves high performances and extends network's lifetime. The average extension of node's lifetime is 2.8 times the one of noSCQPOnly.

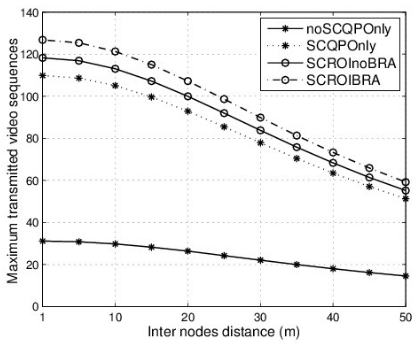

Fig. 5. Maximum number of transmitted video sequences

Fig.6 summarizes the overall average video energy transmission gain for the different use cases with different QP values at one hop of the destination.

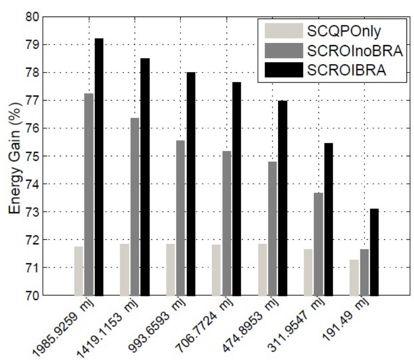

Fig. 6. Video energy transmission gain

5. CONCLUSION

Our motivations behind this work were to propose an energy efficient compression scheme, based on H.264/AVC standard and adapted to WVSN context, that offers to lower layers of protocol stack the possibility to manage differentiated priority video data. We have shown through simulations that SCROInoBRA offers a satisfying video quality of about 30 dB with a bit rate of 230.3 Kbps and video transmission energy efficiency of 73.6%. On the other hand, at same bit rate, SCROIBRA achieves a reduced global quality of 25 dB but with a video transmission energy efficiency of 75.4%. Furthermore, this last permits an average network lifetime extension of 3.06 times.

6. REFERENCES

[1] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, 2009.

[2] I. Politis, M. Tsagkaropoulos, T. Dagiuklas, and S. Kotsopoulos, "Power efficient video multipath transmission over wireless multimedia sensor networks," *Mobile Networks and Applications*, pp. 274–284, 2008.

[3] I. Richardson, "H.264 and MPEG-4 video compression," *Wiley Press*, 2006.

[4] J. Zhang and J. Ding, "Cross-layer optimization for video streaming over wireless multimedia sensor networks," in *International Conference on Computer Application and System Modeling (ICCASM)*, Taiyuan Shanxi, October 2010, pp. 295–298.

[5] I. F. Akyildiz, T. Melodia, and K.R. Chowdury, "Wireless multimedia sensor networks: A survey," *Computer Networks*, vol. 51(4), pp. 921960, 2007.

[6] H.S. Aghdasi, M. Abbaspour, M.E. Moghadam, and Y. Samei, "An energy-efficient and high-quality video transmission architecture in wireless video-based sensor networks," *Sensors*, pp. 4529–4559, 2008.

[7] A.B.S. Bambang, C.M.L. Victor, and N.Panos, "Energy efficient multiple description coding for video sensor network," in *International Symposium on Parallel and Distributed Processing with Applications Workshops (ISPAW)*, Busan, South Korea, May 2011, pp. 153–158.

[8] A. Mammeri, A. Khoumsi, D. Ziou, and B. Hadjou, "Modeling and adapting JPEG to the energy requirements of VSN," in *International Conference on Computer Communications and Networks (ICCCN)*, St. Thomas U.S. Virgin Islands, August 2008, pp. 806–811.

[9] S. Wenger and M. Horowitz, "FMO: Flexible Macroblock Ordering," *JVT-C089*, 2002.

[10] C. Deknudt, P. Corlay, A.S. Bacquet, M. Zwingelstein-Colin, and F.X. Coudoux, "Reduced complexity H.264/AVC transrating based on frequency selectivity for high-definition streams," *IEEE Trans. Consumer Electronics*, pp. 2430–2437, 2010.

[11] S. Ullah, J.J. Ahmad, J. Khalid, and S.A. Khayam, "Energy and distortion analysis of video compression schemes for wireless video sensor networks," in *Military Communications Conference (MILCOM)*, Baltimore, MD, USA, November 2011, pp. 822–82.