Content-Adaptive Robust H.264/SVC Video Communications over 802.11e Networks

Dario Gallucci, Attilio Fiandrotti, Enrico Masala and Juan Carlos De Martin
Dipartimento di Automatica e Informatica
Politecnico di Torino, Italy
Email: [dario.gallucci, attilio.fiandrotti, enrico.masala, juancarlos.demartin]@polito.it

Abstract

In this paper we present a low-complexity traffic prioritization strategy for video transmission using the H.264 scalable video coding (SVC) standard over 802.11e wireless networks. The first part of this work focuses on assessing the perceptual impact of data loss in the various enhancement layers using a wide set of H.264/SVC encoded videos. The analysis shows that perceptual impairments are highly correlated with the motion activity in the video sequence. Thus, we propose an adaptive unequal error protection strategy which identifies the most perceptually important parts of the enhancement layers in the video sequence by means of a low complexity macroblock motion analysis process. The algorithm is tested by simulating a realistic 802.11e based home network scenario. Results obtained on a large set of video sequences show that the proposed content-aware traffic prioritization strategy enables PSNR gains up to 2.5 dB as well as noticeable visual quality improvements with respect to a traditional prioritization strategy aiming at minimizing error propagation.

1. Introduction

The increasing availability of cheap devices with wireless communication capabilities is fostering the demand for reliable wireless multimedia communications. However, wireless devices are highly heterogeneous in terms of communication bandwidth and processing capabilities. Therefore efficient mechanisms, such as scalable multimedia coding systems, are needed to handle such heterogeneity of networks and devices. For the case of video communications, advances in the development of video compression technologies, e.g., the recently standardized H.264/SVC [1], provides significant efficiency gains compared to previous standards [2], therefore the H.264/SVC is expected to be widely adopted in the near future. Currently, the H.264/SVC standard offers temporal, spatial and fidelity scalability. Those scalability mechanisms allow to easily trade-off temporal and spatial video resolution as well as image quality for different bandwidth requirements and types of client devices (mobile devices, personal computers, etc.).

However, wireless multimedia communications, and video in particular, remains challenging because of the intrinsic unreliability of wireless communications, which demands robust error protection mechanisms. Protection is usually provided either by means of application-level error control mechanisms or by exploiting the lower-level QoS capabilities offered by some network technologies, for instance 802.11e. Moreover, in the context of multimedia communications, Unequal Error Protection (UEP) schemes are often employed to optimize resource usage by protecting each part of the multimedia stream in proportion to its importance. However, for these schemes to be the most effective, a reliable data importance measure has to be defined. Such importance is often defined a priori, for instance as a function of the encoding strategy. For non-scalable video coding, for instance, I-frames and P-frames are usually more important than B-frames, therefore they are prioritized by transmitting them first as in [3], [4], or by assigning them a high network protection level [5].

Besides the advantages in heterogeneous scenarios, scalable video coding is also best suited for UEP, since it naturally separates the content into layers with different levels of importance: a base layer which contains the most important information and one or more enhancement layers with less important refinement information. In a UEP scheme involving scalable coding, high priority is used for the base layer to allow the reconstruction, even in the worst case, of a limited quality version of the content, while low priority is assigned to enhancement layers for an efficient use of network resources [6]. Moreover, within the same layer, decoding dependency can be used to determine data priority [4], so that error propagation in case of loss is mitigated.

However, if multiple scalability options are employed, as the H.264/SVC allows, it is not easy to decide how to prioritize data among different types of enhancement layers. For instance, the user might prefer to receive either the temporal enhancement layer, thus trading off details for smoother motion, or the spatial enhancement layer, thus having better image sharpness. Moreover, such preference is not static, since visual degradation usually depends on the video content being compressed as, e.g., in sport clips [7]. Generally, for high-motion content it is visually more pleasant to have higher temporal resolution, i.e., a higher frame rate, than an increased spatial resolution, while the opposite consideration holds for static content.

In this paper a preliminary analysis is conducted to assess the perceptual impact of losses on the various enhancement layers of a H.264/SVC scalable video. Then, on the basis of this analysis, a content-adaptive strategy is proposed to optimize the performance of the communication. In particular, the amount of motion in the video content is estimated by a low-complexity algorithm which exploits side information produced by the video encoder. Such information is used to drive the video prioritization in the context of an 802.11e ad hoc video communication. The perceptual performance of the proposed algorithm is then evaluated using the *ns* network simulator [8].

The rest of this paper is organized as follows. In Sec. 2 we briefly review the H.264/SVC video coding standard and the 802.11e network architecture, while Sec. 3 focuses on the statistical analysis of the perceptual importance of different enhancement layers and introduces the principle of the proposed motion adaptive video prioritization algorithm, described in Sec. 4. Sec. 5 introduces the simulation scenario, and Sec. 6 provides simulation results. Conclusions are drawn in Sec. 7.

2. Technical Background

2.1. The H.264/SVC Coding Standard

The H.264/SVC [9] amendment extends the earlier H.264/AVC standard with spatial, temporal and SNR scalability options, allowing to encode a video as an independently decodable AVC-compatible base layer and one or more SVC enhancement layers. H.264/SVC inherited from H.264/AVC the decomposition of the encoder functionalities among a Video Coding Layer and a Network Abstraction Layer [2]. While the former layer encompasses all the encoder core functionalities (e.g.: macroblock coding), the latter is responsible for the encapsulation of encoded data into independently decodable transport units known as NALUs. Each NALU is prefixed by a header whose fields describe the characteristics of the data it contains, such as the *Type* field which specifies whether the NALU contains AVC base layer or SVC enhancement layer data. Enhancement NALU headers provide additional information, such as the temporal scalability level index (TID).

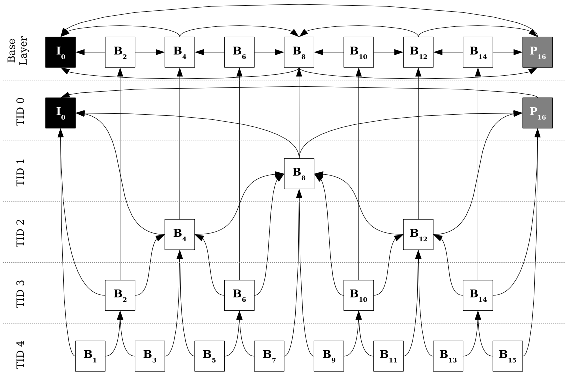

Figure 1. GOP structure of the H.264/SVC encoding scheme used in this work.

As in other video coding standards, frames are considered in groups, called Group Of Pictures (GOP), for the coding and decoding process. Figure 1 depicts the H.264/SVC GOP structure used in this work. The GOP includes the AVC-compatible base layer, one spatial and temporal enhancement layer. Each box represents a NALU, the letter inside the box the corresponding picture type (Intra, Predictively or Bipredictively coded) and the subscript number the display order of the picture. Each GOP is 16 frames long and every 32 frames a picture is Intra-coded, therefore the video encoding scheme is $I_0B\ldots BP_{16}B\ldots BI_{32}$. As it can be surmised by the figure, the base layer provides the decoder with the data needed to reconstruct an half spatial resolution, half frame rate video. NALUs whose TID ranges from zero to three provide the decoder with the information needed to reconstruct a full spatial resolution, half frame rate video (i.e., composed by even frames only). Finally, NALUs with TID equal to four allow to decode the remaining odd frames, which are needed to rebuild the full frame rate video. NALUs with temporal index equal to four are hence referred to as the *temporal enhancement layer*. The arrows, shows the dependencies between the NALUs and indicate that the NALUs are decoded in increasing Temporal Index (TID) order. Therefore, the TID value can be used as a coarse index of NALU importance. During the decoding process, in fact, the loss of a low-TID NALU generates a distortion which propagates to higher TID NALUs only.

2.2. The 802.11e Communication Standard

The recently standardized IEEE 802.11e amendment [10] supersedes the legacy 802.11 DCF specifications for channel access in wireless ad hoc networks with the Enhanced Distributed Channel Access (EDCA) mechanism which provides QoS features. For each host station, four distinct transmission queues known as Access Categories (ACs) are introduced in place of the unique queue of the DCF standard. Each queue is characterized by specific contention window and interframe spacing values, the parameters which control the amount of time the transceiver has to wait before attempting to access the channel for transmission. Because of the different settings of such parameters for each AC, the result is that the queues are characterized by unfair chances of getting access to the channel. As a result, packets in high priority queues are elected for transmission before those in low priority queues, resulting in an effective intra and inter hosts traffic prioritization mechanism.

3. Preliminary Analysis

3.1. Layer-Dependent Distortion Estimation

The 802.11e traffic differentiation capabilities can be exploited to minimize error propagation within a GOP by better protecting pictures used for prediction. With respect to the encoding scheme in Figure 1, the base layer could be mapped to the highest priority traffic class AC3, while the enhancement data could be mapped to lower priority classes on the basis of the TID index [11]. Enhancement NALUs with TID ranging from zero to two could be mapped to AC2, NALUs with temporal index 3 to AC1 and the temporal enhancement layer, that is NALUs with TID equal to four, to AC0. Such TID-based prioritization strategy is used in this work as a reference strategy for comparison purposes. Simulations in various network scenarios showed that in the common case of moderately loaded networks only the low priority AC0 traffic, that is a minimal fraction of video information, is lost [11]. The loss of low priority traffic causes the loss of the temporal enhancement layer: several odd-numbered frames (e.g.: B_1, B_3, etc. in Figure 1) are lost and have to be concealed in the reconstructed video. The loss of odd-numbered frames results in reduced temporal video resolution, which causes a motion jerkiness sensation that is especially noticeable when objects move rapidly, as in the case of sport clips. In order to mitigate such impairments without increasing the amount of high priority bandwidth used for video transport, the temporal enhancement layer needs to receive better protection with respect to other enhancement information. However, this strategy should be applied only if the reduction in distortion achieved by correctly receiving the temporal enhancement layer exceeds the potential distortion introduced by the loss of a lower temporal index enhancement layer, also accounting the distortion propagation to the temporal enhancement layer itself.

	Average bitrate [kb/s]	Encoding PSNR [dB]	Δ PSNR [dB] due to loss of	
			spatial layer	temporal layer
Akiyo	66.50	37.38	-0.17	**-0.26**
Bus	489.00	30.46	-1.07	**-5.40**
Coastguard	337.82	30.28	-1.47	**-1.86**
Container	108.45	33.62	**-0.31**	-0.12
Flower	514.79	29.24	-0.80	**-4.45**
Football	571.57	31.07	-0.66	**-4.80**
Foreman	212.54	33.13	-0.60	**-2.66**
Irene	169.42	34.94	-0.70	**-2.09**
Mobile	501.62	29.52	-0.76	**-2.53**
Mother	80.58	36.68	-0.24	**-0.45**
News	165.16	35.13	**-2.52**	-1.42
Paris	305.61	31.99	**-1.97**	-1.13
Silent	150.87	33.30	-0.24	**-1.14**
Soccer	331.08	32.98	-0.77	**-5.33**
Students	141.74	34.28	-0.26	**-0.34**
Tempete	377.56	30.46	-0.47	**-1.48**

Table 1. Test video characteristics and preliminary analysis of different enhancement layer prioritization strategies. Lower Δ PSNR values (bold) identify the perceptually most important enhancement layer.

A set of test video sequences with widely different characteristics was used to experimentally determine the prioritization strategy which, for each sequence, minimizes the distortion. Table 1 summarizes the characteristics of the videos, which are all CIF size (352×288 pixel), are from 260 to 300 frames long and are encoded at 30 fps according to the scheme depicted in Figure 1. Each enhancement layer of each test sequence was alternatively stripped to simulate its loss during a transmission. The loss of the enhancement layers with temporal index zero, one or two systematically results in either undecodable or extremely distorted video sequences. Vice versa, the loss of the enhancement layers with temporal index three, referred to as the *spatial enhancement layer* in the rest of the paper, and four, that is the *temporal enhancement layer*, results in the quality degradation shown, respectively, in the fourth and fifth columns of Table 1. The loss of the spatial enhancement layer directly affects one frame out of four and, because of error propagation, every one frame out of two. With respect to Figure 1, the loss of the B_2 NALU with TID equal to three would propagate distortion to NALUs B_1 and B_3 with TID equal to four. However, its loss can be effectively concealed by upsampling the corresponding, lower quality, information found in the B_2 NALU in the base layer. On the other hand, the loss of the temporal enhancement layer affects one frame out of two, and no error is propagated. In this case, techniques for error concealment based on frame copy and frame averaging are employed since there is no lower layer information to exploit. The frame averaging technique usually provides better results.

The figures in Table 1 indicate that the reference traffic prioritization strategy provides results that can be improved for about half the set of the considered test sequences. In such cases (e.g.: Bus and Soccer sequences), the loss of the temporal enhancement layer introduces higher distortion (fifth column) than the loss of the spatial layer does (fourth column). Sequences such as Bus or Soccer are characterized by fast movements of objects and the loss of the temporal enhancement layer produces an annoying motion jerkiness effect and a PSNR decrease in excess of 5 dB with respect to the encoded video. On the contrary, the loss of the spatial enhancement layer causes a less perceptible quality degradation (less than 1 dB) and it results in much better video quality than in the previous case. If the temporal enhancement layer is available, the motion jerkiness effect is not present and the artifacts due to error concealment of the lost spatial layer are hardly noticeable. We attribute such effect to the fact that in high-motion scenes the user cannot easily spot image reconstruction artifacts because his attention is captured by fast-moving objects. On the contrary, for sequences such as News or Paris where objects move slowly, no significant motion jerkiness is noticeable even if the temporal enhancement layer is missing. Higher quality degradation is rather observed when the spatial layer is lost, with an average PSNR decrease close to 2.5 dB.

Since for half of the test video sequences it would be better to give higher transmission priority to the temporal enhancement layer rather than to the spatial layer, a content adaptive traffic prioritization strategy is needed to ensure that the optimal traffic prioritization strategy is chosen depending on the characteristics of the sequence. However, analyzing the video content as done in Table 1 requires heavy computation, i.e. multiple decoding and distortion evaluation for each single GOP, which might not be suitable, e.g., in case of multiple concurrent live transmissions. The problem of finding the optimal H.264/SVC prioritization scheme is related with finding the optimal encoding scheme for a given type of video content. The latter issue was recently addressed by means of a fuzzy logic based approach in [7]; however, it specifically focuses on optimizing soccer video and requires a heavy training process. To overcome such limitations, we propose a low complexity traffic prioritization method suitable for generic video sequences.

3.2. Motion Adaptive Video Prioritization

Because the reference traffic prioritization strategy performs poorly if fast-moving objects are present in the video, we propose to use an average macroblock motion measurement to decide, on a GOP basis, whether the temporal or the spatial enhancement layer should receive better protection. If the average macroblock motion level M is below a given threshold, the GOP is considered *static* and the spatial enhancement layer, which, in this case, is more important than the temporal layer, receives better protection. Otherwise, the GOP is classified as *dynamic* and the opposite protection strategy applies.

The motion information M required to implement the proposed traffic prioritization strategy can be easily extracted from the compressed video bitstream. For every picture, the number of macroblocks coded in *Direct* mode N_{DirectMB}, that is macroblocks encoded without using a motion vector, is determined. Such number is then subtracted from the total number of macroblocks of the picture, N_{MB}, thus determining the number of macroblocks with at least one non-zero motion vector. Then, to determine the required motion measure M, such value is divided by N_{MB} and then averaged over the N pictures that constitutes the GOP, as shown in Equation (1).

$$M = \frac{1}{N} \sum_{i=1}^{N} \frac{N_{\mathrm{MB}} - N_{\mathrm{DirectMB}}}{N_{\mathrm{MB}}}. \qquad (1)$$

Figure 2 allows to investigate the relationship between M and the traffic prioritization strategy which minimizes the distortion. For each GOP of every video sequence, the figure shows the M value and the

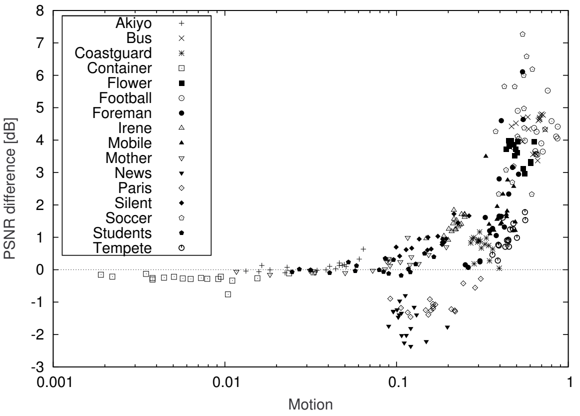

Figure 2. PSNR difference between the case of loss of the temporal enhancement layer and the case of loss of the spatial layer as a function of the motion measured in the video.

difference in PSNR between the case of loss of the temporal enhancement layer and the case of loss of the spatial layer. Points with positive PSNR difference correspond to GOPs whose temporal enhancement layer is perceptually more important that the spatial enhancement layer. In this case, it is better to assign a higher protection level to the temporal enhancement layer than to the spatial enhancement layer. The opposite consideration holds for points with negative PSNR difference.

Figure 2 shows that two main clusters of GOPs exist. The first cluster encompasses GOPs located in the leftmost area of the graph (low M values); such GOPs mainly present a negative PSNR difference. Since they require better protection of the spatial enhancement layer, we refer to them as *static*. The second cluster encompasses GOPs located in the rightmost area of the graph (high M values); almost all GOPs present a positive PSNR difference. Since they require better protection of the temporal enhancement layer, we refer to them as *dynamic*.

From the figure, a reasonable M value to discriminate between static and dynamic clusters of GOPs is in the 0.2 to 0.3 range. Note also that if the M value is lower than about 0.07, giving better protection to either the temporal or the spatial enhancement layer approximately provides the same performance.

4. The Traffic Prioritization Algorithm

In this section an implementation of the motion adaptive traffic prioritization strategy proposed in Section 3 is described. For each GOP, the base layer is encapsulated into RTP packets in compliance with the RFC draft [12] and mapped to the highest priority AC3 class, while enhancement layers with temporal index ranging from 0 to 2 are mapped to AC2. The algorithm then computes the GOP average motion index using Equation (1) and, by comparison against a threshold value, it classifies the GOP either as static or dynamic. Then, NALUs are encapsulated into RTP packets and queued into a temporary list. Depending on GOP classification, the spatial enhancement layer packets are enqueued before the temporal enhancement layer (static case) or vice versa (dynamic case). Therefore, the perceptually more important packets are located at the head of the list regardless of the GOP classification. Finally, the first half of the packets in the list are extracted from the head of the list and mapped to AC1, while the remaining packets are mapped to AC0. The algorithm complexity is linear with the number of macroblocks, as the process only requires to analyze the encoding type of each macroblock.

5. Simulation setup

An ad hoc wireless network is simulated using the NS network simulator [8] and the 802.11e EDCA extension developed at the Berlin Technishe Universität [13].

Figure 3 illustrates the simulated network topology, which represents a typical domestic environment where different types of hosts are located in the various rooms of a building and operate within the same collision domain. Host A is a gateway which provides internet access to all the other hosts in the building. Host G and D are digital TV sets which receive a stream composed of an SVC video and an AAC audio substream. Host F is a videoconferencing device which communicates with a remote host located in the internet by sending and receiving VoIP traffic and low bitrate H.263 video. Hosts B and E are videophones which communicate among themselves, each generating traffic whose characteristics, in type and bandwidth, are similar to the ones of node F. Host C is a PC which exchanges data with a host on the internet via a TCP connection, and it additionally acts as a domestic media server which streams AAC audio and H.264/SVC video to hosts D and H.

In such a scenario data flows with different bandwidth and delay requirements coexist. For example, a VoIP call requires limited bandwidth, although it loads the network with a high number of small packet which have tight maximum delay and jitter requirements. Videoconferencing traffic demands more bandwidth than VoIP and, similarly, requires timely packet de-

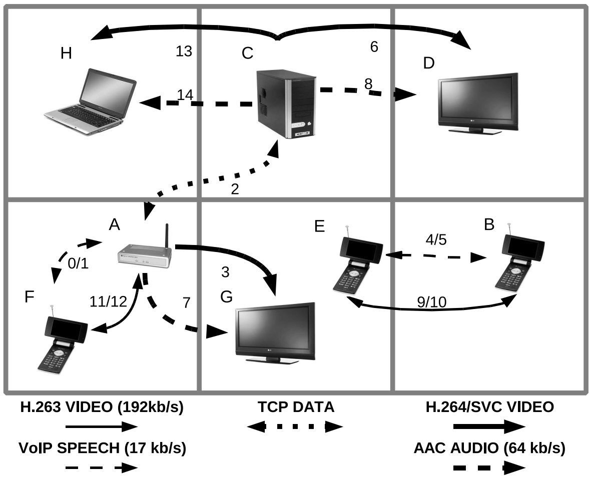

Figure 3. The simulated network topology, showing the analyzed video flow (i.e, flow #3) and all the interfering background traffic.

livery. Maximum delay requirements for streaming of pre-recorded contents are less stringent, albeit a minimum bandwidth is required to ensure a smooth playback. Therefore the four VoIP streams are delivered using the highest importance AC3 class provided by the 802.11e standard. Traffic class AC2 encompasses the four H.263 video streams and the 50% perceptually most important traffic, i.e., approximately all the base layer of the three H.264/SVC flows. AC1 encompasses the three AAC streams and the 25% of the SVC enhancement traffic, i.e., the most important part of the enhancement data. Finally, the background TCP traffic and the remaining H.264/SVC enhancement traffic are assigned to the lowest importance AC0.

Since we aim at investigating the performance of the proposed traffic prioritization algorithm in relation with the QoS capabilities offered by 802.11e, we prefer not to account the effect that the noise on the channel would have on the video transmission, thus no link error model is simulated. Nevertheless, in the considered scenario packet losses are caused by the limited size of the transmission queues (50 packets), together with collisions in channel access happening due to the heavy traffic sustained by the network, whose channel bandwidth is set to 11 Mb/s. Transmission queues also implement, as suggested in the 802.11 standard, a timeout mechanism which drops packets after 0.5s of stay in the queue. The amount of traffic offered to the network is determined by the interfering background traffic, which is kept constant, and the bitrate of each test video, which varies from sequence to sequence as reported in Table 1.

Two simulation sets are performed. Each sequence is first transmitted using the motion adaptive traffic prioritization algorithm described in Section 4 (threshold motion value set to 0.20) and then, for comparison purposes, using the reference temporal index based traffic prioritization. Fairness in the comparison of the two traffic prioritization strategies is ensured, for each sequence, by an identical allocation of video traffic among 802.11e classes.

6. Results

Table 2 shows the simulation results for each combination of test sequence and traffic prioritization strategy for the H.264/SVC video under test, i.e., the stream flowing from node A to G. For each scenario, the global byte loss rate (fraction of bytes lost on sent), the loss rate of spatial and temporal enhancement information and the PSNR of the reconstructed video is reported. The table also reports the GOP classification produced by the algorithm: GOPs of a given sequence may be classified as all static (S), as all dynamic (D) or some as static and some as dynamic (S/D).

The byte loss rate ranges from as few as 4% (Container) up to nearly 20% (Football) and depends on the traffic offered to the network. The traffic offered to the network in turn depends on both the background and the H.264/SVC video traffic, the latter ranging from 80 kb/s (Mother) up to nearly 600 kb/s (Football). Losses mainly affect the low priority AC0 class, but when the network is heavily loaded, e.g., in case of the Soccer sequence, also a few packets in the AC1 class are lost.

As it can be surmised from the table, the motion adaptive traffic prioritization strategy always assigns more protection to the enhancement layer identified as the most perceptually important one, which implies a different byte loss ratio for the temporal and spatial enhancement information. For instance, in the case of the all static Akiyo sequence packet losses only affect the temporal enhancement layer, while in the case of the all dynamic Bus sequence packet losses mainly affect the spatial enhancement layer.

As expected, the motion adaptive traffic prioritization strategy provides the same performance as the reference technique in the case of sequences classified as static, e.g., Akiyo or Container. In the case of sequences made of both static and dynamic GOPs (i.e., Paris and Silent), minor PSNR improvements are achieved. In the case of the Paris sequence, a slight performance degradation is recorded because three GOPs (the rightmost ones of Paris in Figure 2) exceed in motion the 0.25 threshold, thus they are classified as dynamic even though they present a negative PSNR

		Reference strategy				Proposed content-adaptive strategy					
Sequence	Encoding PSNR [dB]	Byte loss rate [%]	Spatial layer loss [%]	Temporal layer loss [%]	PSNR [dB]	Sequence type	Byte loss rate [%]	Spatial layer loss [%]	Temporal layer loss [%]	PSNR [dB]	Performance gain [dB]
Akiyo	37.38	7.15	0.00	74.79	**37.18**	S	7.22	0.00	75.18	**37.18**	**0.00**
Bus	30.46	19.97	0.40	73.52	**23.47**	D	19.23	96.20	12.60	**25.55**	**2.08**
Coastguard	30.28	13.00	0.00	86.96	**25.87**	D	13.65	90.61	0.26	**26.62**	**0.75**
Container	33.62	4.22	0.00	90.65	**33.52**	S	4.25	0.00	90.62	**33.53**	**0.01**
Flower	29.24	19.37	0.17	81.97	**21.81**	D	18.71	96.48	2.04	**24.30**	**2.49**
Football	31.07	20.34	1.40	68.74	**26.44**	D	20.15	93.70	27.87	**28.07**	**1.63**
Foreman	33.13	12.39	0.00	67.79	**31.24**	D	12.39	92.21	15.92	**31.85**	**0.61**
Irene	34.94	12.56	0.11	66.36	**33.52**	S	12.37	0.00	65.15	**33.55**	**0.03**
Mobile	29.52	18.70	0.00	82.83	**22.16**	D	18.38	96.23	2.38	**23.44**	**1.28**
Mother	36.68	9.98	0.00	74.49	**36.36**	S	9.65	0.00	72.26	**36.37**	**0.01**
News	35.13	9.02	0.00	89.38	**33.82**	S	8.76	0.11	87.45	**33.86**	**0.04**
Paris	31.99	10.08	0.00	87.82	**30.22**	S/D	11.16	25.09	63.68	**29.44**	**-0.78**
Silent	33.30	9.88	0.00	67.16	**32.54**	S/D	9.90	6.46	62.94	**32.54**	**0.00**
Soccer	32.98	15.08	0.00	66.79	**28.89**	D	14.85	91.44	13.09	**31.09**	**2.20**
Students	34.28	7.45	0.00	83.58	**34.07**	S	7.37	0.00	82.53	**34.07**	**0.00**
Tempete	30.46	13.39	0.00	70.35	**27.87**	D	13.47	93.19	11.96	**28.79**	**0.92**

Table 2. Simulation results comparing the reference temporal index based strategy with the proposed content-adaptive traffic prioritization strategy.

difference value. If a different motion threshold value (e.g. 0.30) is used, all GOPs of the Paris sequence would be classified as static and hence the PSNR performance would be equivalent to the reference technique. With such a threshold value, however, the Irene and Silent sequences would be classified as all-static, therefore no gain will be achieved with respect to the reference technique.

The motion adaptive strategy greatly improves the quality of the received video in the case of markedly dynamic sequences. Flower and Soccer videos show the most noticeable quality improvement, with PSNR gains up to 2.5 dB. A visual inspection of the reconstructed videos shows almost no error concealment artifacts. The playback of the video is smooth and provides a better visual experience than the reference traffic prioritization technique. Summarizing the previous results, the proposed motion-adaptive technique achieves an average quality improvement in excess of 1 dB for half of the tested sequences, while in all other cases the strategy behaves exactly as the reference, TID-based, strategy.

7. Conclusions

In this paper we propose a content adaptive traffic prioritization strategy for H.264/SVC scalable video communications over 802.11e wireless networks in moderate packet loss conditions. By means of an extensive analysis on a large set of test video sequences, we point out the relationship between the presence of high motion in the video sequence and the relative importance of the temporal enhancement layer over the spatial one. We exploit such relationship to design a content adaptive traffic prioritization algorithm which classifies each part of a video sequence either as static or dynamic on the basis of the motion features extracted from the compressed bitstream and selects the optimal traffic prioritization strategy accordingly. Simulations in an 802.11e wireless ad hoc network scenario show that the proposed strategy achieves significant PSNR gains, up to 2.5 dB, over a reference content-unaware technique, as well as a remarkable increase in visual quality and smoothness. Due to its low complexity, the proposed strategy is suitable for real time video transmissions and can also be adapted to other network architectures which provide traffic differentiation capabilities.

References

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H. 264/AVC Standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, p. 1103, 2007.

[2] T. Wiegand, G. Sullivan, G. Bjntegaard, and A. Luthra, “Overview of the H. 264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.

[3] S. H. Kang and A. Zakhor, "Packet scheduling algorithm for wireless video streaming," in *Proc. Packet Video Workshop*, Pittsburgh, PA, Apr. 2002.

[4] Y. Cho, D.-K. Kwon, C.-C.J. Kuo, R. Huang, C. Lima, J.D. Black, and J.J. Pan, "Video streaming over 3G networks with GOP-based priority scheduling," in *Proc. Intelligent Information Hiding and Multimedia Signal Processing*, Dec. 2006, pp. 201–204.

[5] A. Ksentini, M. Naimi, and A. Guéroui, "Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture," *IEEE Communications Magazine*, vol. January, 2006.

[6] K. Gao, W. Gao, S. He, P. Gao, and Y. Zhang, "Real-time scheduling on scalable media stream delivery," in *Proc. Intl. Symp. on Circuits and Systems*, vol. 2, May 2003, pp. 824–827.

[7] E. Akyol, A. M. Tekalp, and M. R. Civanlar, "Content-aware scalability-type selection for rate adaptation of scalable video," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 10236, 2007.

[8] UCB/LBNL/VINT, "Network simulator - ns - version 2.31," http://www.isi.edu/nsnam/ns, 2007.

[9] *Advanced video coding for generic audiovisual services — Annex G*, ITU-T & ISO/IEC Std. H.264 & 14 496-10, Nov. 2007.

[10] *Wireless LAN Medium Access Control and Physical Layer specifications Amendment 8: Medium Access Control Quality of Service Enhancements*, IEEE Std. 802.11e, 2005.

[11] A. Fiandrotti, D. Gallucci, E. Masala, and E. Magli, "Traffic prioritization of H.264/SVC video over 802.11e ad hoc wireless networks," in *Proc. of the IEEE ICCCN 17th International Conference on Computer Communications and Networks*, 2008.

[12] S. Wenger, T. Schierl, and A. Eleftheriadis, "RTP payload format for SVC video," draft-ietf-avt-rtp-svc-14 (work in progress), Sep. 2008.

[13] S. Wiethölter, M. Emmelmann, C. Hoene, and A. Wolisz, "TKN EDCA model for ns-2," Telecommunication Networks Group, Technische Universität Berlin, Tech. Rep. TKN-06-003, 2006.