

APPLICATION-AWARE OPTIMIZATION OF PACKET SCHEDULING FOR VIDEO COMMUNICATIONS OVER INTERVEHICLE AD HOC NETWORKS

Enrico Masala

Computer and Control Engineering Department — Politecnico di Torino, Italy
Email: masala@polito.it

ABSTRACT

This paper focuses on optimizing real-time intervehicle video communications from an image analysis perspective. Differently from traditional multimedia communication schemes which are typically optimized to maximize perceptual quality as perceived by human users, in this work we propose a new packet importance estimation method designed to capture the contribution of each packet to the performance of the image analysis process. Packet importance values are used to optimize video packet scheduling and retransmission in the context of ad hoc 802.11 intervehicle communications. Simulations using actual intervehicle transmission traces show PSNR gains up to 2.3 dB and 0.8 dB compared with the standard MAC-layer ARQ and a distortion based optimized ARQ technique respectively, as well as up to 10% accuracy increase with the tested image analysis algorithm.

Index Terms— Intervehicle communications, low-delay video, application-aware packet scheduling

1. INTRODUCTION

Potential applications of wireless multimedia communications are numerous, and include telephony, remote surveillance, entertainment, etc. Typically, users the multimedia communications are human beings. However, automatic multimedia processing and analysis systems, enabled by the constant increase of processing power, have already started to appear, e.g. for video surveillance applications. Automotive applications are another promising example. Some cars already include cameras aimed at the road to improve driving safety, e.g. in difficult environmental conditions such as night, and the interests for solutions which improve driving safety is constantly increasing. In such a context, it could be highly beneficial to driving safety if cars could connect to each other and exchange information, e.g. warning and safety messages. Moreover, in the near future they could also establish more complex form of communications, such as real-time video transmissions, to perform, e.g., cooperative visual processing of road information from multiple point of view. This would improve, for instance, obstacle detection, and in the long term, contribute to accident prevention.

Many efforts have been devoted to investigate wireless multimedia communications issues. In particular, it is well known that multimedia streams exhibit non-uniform perceptual importance, which has often been exploited to improve multimedia communication performance [1, 2]. In these works, perceptual importance is typically defined according to the expectations of human users. To the best

The authors are grateful to the NEDO Project which provided the video traces (<http://www.sp.m.is.nagoya-u.ac.jp/NEDO>).

of our knowledge, we are not aware of any effort aimed at optimizing communication in terms of the performance of automatic image analysis and processing algorithms at the receiver. A partially similar approach is image segmentation of natural scenes, coupled with MPEG-4 object-based video coding. However, this approach requires the use of a specific video codec and the segmentation algorithm is often difficult to tune properly.

In this work we propose and investigate a video communication system optimized from an image analysis application perspective. In particular, a new packet importance estimation method is proposed, based on both the importance of the video data from the image analysis perspective and the peculiar dependency characteristics of a typical compressed video stream. The resulting packet importance values are used to optimize video packet scheduling and retransmission in the context of ad hoc 802.11 intervehicle communications. Results are provided in terms of both image quality and performance of the image analysis process.

The paper is organized as follows. Section 2 introduces the typical intervehicle wireless communication issues, while Section 3 describes the proposed technique in details. Section 4 and 5 present the simulation setup and results, respectively, followed by conclusions.

2. INTERVEHICLE WIRELESS COMMUNICATIONS

Wireless communications are challenging due to noise and interference. Moreover, in the case of intervehicle communications, node mobility is high, because vehicles often move at high speed and distance between nodes can quickly vary. Thus wireless signals can be easily attenuated by obstacles and reflections, which are one of the reasons of the strong time variability of the quality of intervehicle channels.

Some studies investigated the performance of intervehicle communications, in particular using the 802.11 protocol, to experimentally assess the characteristics of intervehicle channels, concluding that the effective throughput is highly variable [3, 4]. However, the recently approved (for trial use) Wireless Access for the Vehicular Environment (WAVE) communication standard [5] uses a variant of the 802.11 protocol, namely the 802.11p, as the data communication layer. WAVE provides different channels, some of them available for custom application deployment, e.g. multimedia communications. Video communication between vehicles, for instance, could be deployed to enable videoconferencing between passengers of cars traveling in the same direction as well as cooperative road video processing applications.

However, the possibility of successfully deploying such multimedia communication applications mainly relies on the availability of techniques which can efficiently cope with the strong time variability of intervehicle channels. A sample packet loss pattern, ex-

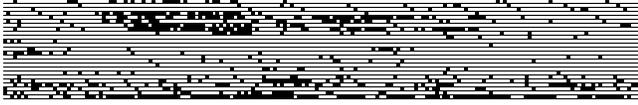


Fig. 1. Sample packet loss pattern (part of Trace 2).

tracted from intervehicle communications experiments [6], is shown in Figure 1. Each square represents the transmission of a single packet. Time increases from left to right, rows show consecutive time segments. White indicates successful packet reception while black means that the packet could not be successfully received, hence black rectangles refer to runs of consecutively lost packets. From the figure it is clear that the wireless channel greatly varies as a function of time, from quite good conditions to instants in which the communication is severely impaired. Therefore, during bad channel conditions it is extremely important to take the maximum advantage of the few available transmission chances by sending the most important information first.

The 802.11 standard uses retransmissions to combat packet losses. However, such a mechanism is data-agnostic, i.e. it simply retransmits packets, one at a time, up to a predefined number of retransmissions, called retry limit, regardless of the data contained in packets. Therefore, the retransmission mechanism could be improved by means of appropriate application-aware packet retransmission and scheduling strategies.

3. APPLICATION-AWARE SCHEDULING OPTIMIZATION

Packet scheduling optimization has often been used to optimize the performance of video communications, and many efforts focused on improving the communication quality as perceived by the user. All these approaches typically include a video quality estimation measure which is used to drive the optimization process, so that it is possible to choose, among each possible transmission policy, which is the one which yields the highest expected video quality. For instance, the most widely used utility measure for video communications is the Peak Signal-to-Noise Ratio (PSNR).

However, when the user of the video stream is an algorithm performing image analysis rather than a human user, different metrics could be more appropriate. In particular, we propose to define a utility measure to reflect the importance of each of the various parts of the image for the purpose of the image analysis process. This is not easy since in general the various parts of the image provide a contribution to the performance of the image analysis process which is difficult to model and which cannot be easily separated from the others.

For the purpose of this work, we focus on a generic road object detection application based on real-time analysis of video captured by a camera positioned in front of the car. Pictures of the same scene captured from different perspectives improve the performance of image analysis algorithms since they allow to perform multi-view processing of the scene. In this work we focus on a scenario in which video data captured by the front camera of a car is compressed and transmitted to a nearby car to allow cooperative visual processing of road information improving, e.g., automatic object detection. Since intervehicle wireless communications might be impaired by a high packet loss rate, it is advisable that, in such conditions, the most important parts of the video stream from the image analysis perspective are sent with higher priority than the rest of the data. Note that we address a *generic* object detection application, thus it is not possi-



Fig. 2. Example of key points determined by the KLT feature tracker.

ble to pre-process images locally at the sender and transmit only the results provided by a given application.

This work proposes a general framework to solve such a problem which relies on the identification of a number of *key points* in each captured video frame, which are considered important from the image analysis perspective. To practically implement this approach we use the Kanade-Lucas-Tomasi (KLT) feature tracker [7, 8] to determine the key points, and on the basis of their position inside the picture, we define the priorities of the video packets which will be used by the scheduling algorithm. However, note that, despite the presented approach makes use of a particular algorithm, i.e. the KLT feature tracker, any image analysis algorithm which is able to provide a set of important points can be used. Moreover, such information can often be obtained with little computational effort as a by-product of image analysis algorithms. Note that, for the purpose of this work, regardless of how key points are obtained, it is sufficient that they are positioned near to the most important parts of the image from the image analysis perspective. Figure 2 shows an example of the results provided by the KLT feature tracker.

We assume that a standard video encoder is running on the sender node. No particular techniques to optimize the encoding process for later image analysis are implemented in the encoder. Moreover, each slice in inserted into one packet hence the maximum slice size is bounded by the network Maximum Transmission Unit (MTU). The encoder produces I and P frames only, due to the low-delay requirements of the real-time visual processing applications. Moreover, resiliency is improved by periodic insertion of intra-coded frames.

In order to determine the best packet transmission policy at a certain time instant, each possible combination of packet loss patterns should be considered to compute the exact effect on the performance of the image analysis algorithm, and the effect should be weighted by the probability that such a combination of events occurs. Clearly this approach is impractical because both probabilities are difficult to estimate (e.g. due to channel burstiness and strong time variability) and performance is a non-linear combination of the received data.

Thus we propose a much simpler heuristic algorithm, based on the expected utility of each packet for the image analysis perspective. For simplicity's sake, we assume that the number of key points computed at the sender is constant for each video frame, however a corrective factor could be introduced to consider a variable number of points for each frame. For each packet i , let k_i the number of

key points belonging the area of the video frame represented by that packet. Let k_i/m_i the utility value per macroblock for that packet, where m_i is the number of macroblocks included in that packet. Then that value is further adjusted by means of two coefficients to obtain the packet priority value. The first coefficient f_i considers the dependency structure implied in video coding. f_i is defined as the number of frames which depends on the frame to which packet i belongs. For instance, its value is equal to the GOP size in frames if the packet belongs to an I frame, while it is equal to one in case of packets belonging to the last P frame of the GOP. Then, a second coefficient t_i is introduced to give packets belonging to I frames higher priority than the rest of the packets since, if correctly received, they stop error propagation. Experimentally it has been found that a good t_i value is the ratio of the average size of I frames over the average size of P frames. t_i is equal to one if the considered packet does not belong to an I frame. In summary, the priority value v_i for the packet i is given by $v_i = k_i \cdot f_i \cdot t_i/m_i$.

The principle underlying the design of the packet priority function is to transmit with higher priority the parts of the image that, in case of loss, greatly reduce the performance of the image analysis algorithms at the receiver side. The function is designed to consider both the position of the key points and the dependency structure which is typical of video coding.

Three packet scheduling and retransmission techniques are considered in this work: the standard 802.11 MAC-layer Automatic Repeat reQuest (ARQ) technique, the Distortion-Based (DB) technique for intervehicle video communications described in [9] and the proposed algorithm. The standard 802.11 technique retransmits packets, in the order given by the queue, up to a maximum number of retransmission attempts (retry limit). The other two techniques select, at each new transmission opportunity, the packet with the highest priority value, regardless of its position in the queue. For the DB technique, the packet priority value is computed according to [10, 11], i.e. it uses the MSE distortion that would be caused by the loss of the packet in the decoded video sequence (including error propagation), while the proposed technique uses the v_i value. Moreover, both techniques remove packets from the transmission queue if their stay in the queue exceeds a given threshold which is determined by the real-time constraints of the application.

4. EXPERIMENTAL SETUP

To assess the performance of the three scheduling techniques, we perform trace-based [6, 12] packet level simulations of 802.11 ad hoc transmissions between two vehicles traveling along the same road. The average packet loss rate ranges from 4.9% to 32.8% depending on the trace. Video resolution is 640x480, 30 fps. The H.264 codec v. JM11 is configured to produce 12-frame GOP with P frames only, with fixed quantization stepsize. Packet size is approximately constant, equal to the MTU size (1500 bytes). Each video segment length is 30 s. The IP/UDP/RTP protocol stack is used, and packetization complies with RFC 3984. Error concealment is performed by substituting missing areas with the same areas in the previous frame.

Since the two devices communicate directly, the MAC-layer ACK/NACK information is used to determine, for each sent packet, the transmission outcome and to eliminate, from the transmission queue, successfully received packets. We configured the feature tracker to provide 100 key points for each video frame. The t_i coefficient is equal to four for I frames. The retry limit for the standard MAC-layer ARQ is set to seven, which is the default on most systems. The maximum communication delay allowed by the

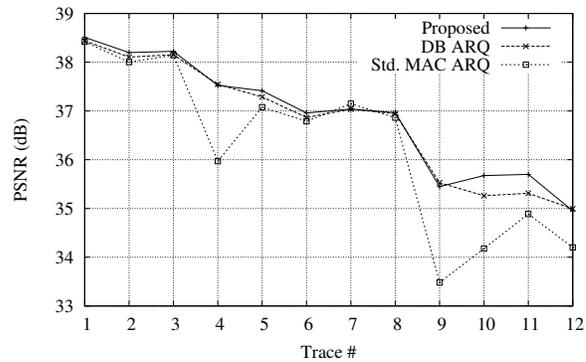


Fig. 3. PSNR performance over the whole image for all techniques.



Fig. 4. Sample of the region considered as the most important for image analysis (brighter areas.)

application is 66 ms.

5. RESULTS

Figure 3 shows a comparison of the PSNR performance computed over the whole image for the three techniques. Results are plotted for various loss traces with increasing packet loss rates. The performance of the proposed technique is higher than the standard MAC-layer ARQ technique and it is similar or slightly higher than the performance achieved by the DB ARQ technique.

We also computed the PSNR value on a region of the image, which is the most representative part from the image analysis perspective. That region includes all macroblocks which contains at least one pixel whose distance from a key point is lower than a given threshold, set to ten pixels. A sample of such a region, which changes for each frame according to the results of the KLT algorithm applied on the original video sequence, is shown in Figure 4. Figure 5 shows the PSNR results computed on that region. The performance gain of the proposed technique over the standard MAC-layer ARQ is up to 2.3 dB, and up to 0.8 dB with respect to the DB ARQ technique. The analysis on confidence intervals shows that with 95% accuracy the average gain is between 0.43 and 1.16 dB in the former case, and between 0.20 and 0.40 dB in the latter. Such results are in accordance with the performance of the image analysis algorithm.

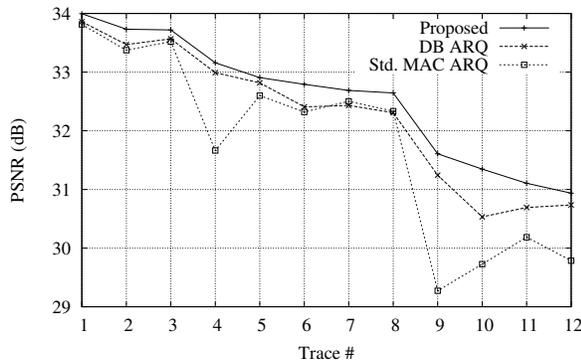


Fig. 5. PSNR performance over the most important region of the image.

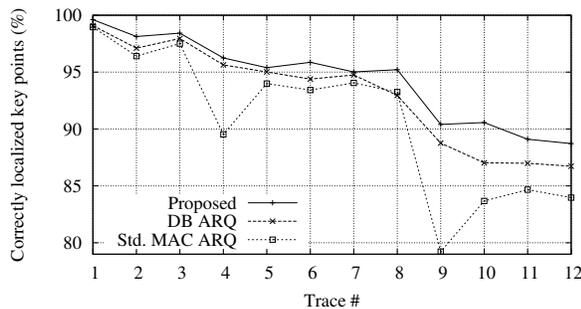


Fig. 6. Percentage of correctly localized key points for all techniques.

Figure 6 shows the percentage of key points in the decoded video whose position is the same as in the original error-free video. The proposed technique allows correct localization of up to 10% more points than the MAC-layer ARQ technique, and 3% more than the DB ARQ technique.

Finally, Figure 7 shows the performance of image analysis for the proposed technique and the best reference technique (i.e. the DB ARQ) as a function of time. The performance of the proposed technique over the whole trace is only 2% higher than the DB ARQ technique, due to averaging with long time periods in which communication is good. However, when transmission is impaired by packet losses, the performance gain is significant, e.g. from 40% to 90% correctly localized key points around 9 s. The performance would be even better if compared with the standard MAC-layer ARQ technique.

6. CONCLUSIONS

This work proposed a technique to optimize video communications from an image analysis perspective. A new packet importance estimation method is introduced, which is based on both the importance of the video data from the image analysis perspective and the peculiar dependency characteristics of a typical compressed video stream. Packet importance values are then used to optimize scheduling and retransmissions in the context of real-time intervehicle communications. Simulation results, evaluated in terms of both image quality and accuracy increase of the image analysis process, show that the proposed technique outperforms the standard 802.11 MAC-layer ARQ mechanism as well as a distortion-based optimized retransmission algorithm.

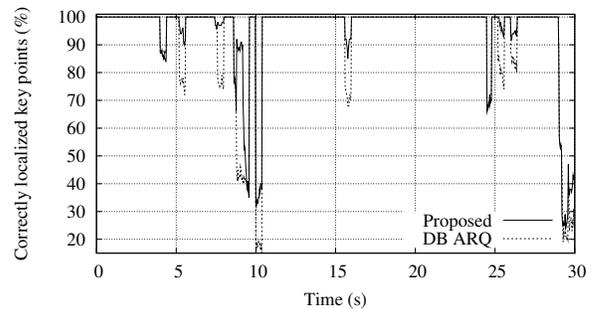


Fig. 7. Percentage of correctly localized key points as a function of time, Trace 8.

7. REFERENCES

- [1] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust internet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 357–372, March 2001.
- [2] P.A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, Apr 2006.
- [3] J.P. Singh, N. Bambos, B. Srinivasan and D. Clawin, "Wireless LAN Performance Under Varied Stress Conditions in Vehicular Traffic Scenarios," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Sept. 2002, vol. 2, pp. 743–747.
- [4] P. Buccioli et al., "Performance evaluation of H.264 video streaming over inter-vehicular 802.11 ad hoc networks," in *Proc. of IEEE Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Germany, Sept. 2005.
- [5] "IEEE trial-use standard for wireless access in vehicular environments (WAVE) — multi-channel operation," *IEEE Std. 1609.4*, November 2006.
- [6] H. Wu et al., "CRAWDAD data set gatech/vehicular," URL: <http://crawdad.cs.dartmouth.edu/gatech/vehicular>.
- [7] C. Tomasi and T. Kanade, "Detection and tracking of point features," *Carnegie Mellon University Technical Report CMU-CS-91-132*, April 1991.
- [8] "KLT: Kanade-Lucas-Tomasi feature tracker," <http://www.ces.clemson.edu/stb/klt/index.html>, 2007.
- [9] E. Masala and J. C. De Martin, "Distortion-optimized retransmission for low-delay robust video communications over 802.11 intervehicle ad hoc networks," in *Proc. of ACM Int. Workshop on Vehicular Ad Hoc Networks (VANET)*, Montreal, Canada, Sept. 2007, pp. 69–70.
- [10] E. Masala and J.C. De Martin, "Analysis-by-synthesis distortion computation for rate-distortion optimized multimedia streaming," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, Baltimore, MD, July 2003, vol. 3, pp. 345–348.
- [11] F. De Vito, D. Quaglia, and J.C. De Martin, "Model-based distortion estimation for perceptual classification of video packets," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, Sept. 2004, vol. 1, pp. 79–82.
- [12] H. Wu et al., "An empirical study of short range communications for vehicles," in *Proc. of ACM Int. Workshop on Vehicular Ad Hoc Networks (VANET)*, Sept. 2005.