

PERCEPTUALLY-OPTIMIZED PACKET SCHEDULING FOR REAL-TIME 802.11 VIDEO COMMUNICATIONS BETWEEN VEHICLES

Enrico Masala, Juan Carlos De Martin

Computer and Control Engineering Dept., Politecnico di Torino, I-10129 Torino, Italy
{masala,demartin}@polito.it

ABSTRACT

The automotive industry is increasingly looking at solutions for intervehicle wireless communications for applications ranging from safety to entertainment. Warning signals are the most immediate applications, but more complex forms of communications, and video in particular, could be used by innovative applications such as multi-vehicle-based visual processing of road information, multi-vehicle radar systems for obstacle avoidance and automatic driving, and more generally swarm communications among cars traveling along the same road. In this paper we present a perceptually-optimized packet scheduling algorithm which can transmit video data captured by an on-board camera to another vehicle in proximity, using standard 802.11 wireless technology. Low-delay transmission of on-board captured H.264 video is simulated using actual intervehicle packet transmission traces. The results show that the proposed perceptually-optimized algorithm achieves a consistently higher quality compared to two reference techniques, with gains up to 2 dB PSNR. Moreover, timely dropping of packets which cannot reach the receiver on time for playback has been shown to greatly improve the performance of the transmission system.

Index Terms— Packet scheduling optimization, low-delay video, intervehicle communications

1. INTRODUCTION

The potential applications of intervehicle wireless communications are numerous, ranging from safety to entertainment. Warning signals are the most immediate applications, but more complex forms of communications, and video in particular, could be used by innovative applications such as multi-vehicle-based visual processing of road information for obstacle avoidance and automatic driving, and more generally swarm communications among cars traveling along the same road.

Numerous issues must be addressed for efficient intervehicle communication. The Wireless Access in Vehicular Environment (WAVE) suite of communications standards, that has recently been approved for trial use, aim

at addressing issues such as security, management of multiple radio channels and system resources. Other standards are under active development, such as the IEEE 802.11p, an extension of the well-known 802.11, which is expected to cover protocol and networking services in WAVE.

Intervehicle and vehicle to infrastructure communications, however, have already been experimented by several research projects based, for instance, using the 802.11 standard [1][2]. Those works mainly focused on throughput and connectivity issues. Others investigated routing [3] and information dissemination issues [4]. Modifications of the existing MAC access protocols have also been proposed [5] to provide important features such as bounded maximum access delay.

However, for more specific applications, such as real-time multimedia as opposed to generic data communication, it is possible to exploit the peculiar characteristics of the data, and in particular their non-uniform importance, to further optimize the communication performance. In this work we present a perceptually-optimized packet scheduling algorithm aimed at providing reliable low-delay intervehicle video communications. The proposed algorithm is tested using actual intervehicle packet transmission traces and video sequences captured from on-board cameras, and results are compared with two reference techniques, that is the standard 802.11 MAC-layer retransmission scheme and a delay-constrained retransmission scheme.

The paper is organized as follows. Section 2 briefly introduces the H.264 standard, explains the distortion estimation technique and discusses the main issues of multimedia transmission over 802.11. The proposed perceptually-optimized scheduling technique as well as the two reference techniques are analyzed in details in Section 3. Section 4 describes the experimental setup and Section 5 discusses the results. Finally, conclusions are drawn in Section 6.

2. H.264 VIDEO COMMUNICATIONS OVER 802.11

2.1. The H.264 Video Coding Standard

In this work we consider video communications based on

the state-of-the-art H.264 video codec [6][7], which is particularly suitable for packet networks communications. In fact, one of the most interesting characteristics of the H.264 standard is the attempt to decouple the coding aspects from the bitstream adaptation needed to transmit it over a particular channel. The part of the standard that deals with the coding aspects is called Video Coding Layer (VCL), while the other is the Network Adaptation Layer (NAL).

As in previous video coding standards, the H.264 VCL groups consecutive macroblocks into *slices*, that are the smallest independently decodable units. Slices are important because they allow to subdivide the coded bitstream into independent packets, so that the loss of a packet does not affect the ability of the receiver to decode the bitstream of others.

Differently from other video coding standards, the H.264 provides a NAL which aims to efficiently support transmission over IP networks. In particular, it relies on the use of the Real-Time Transport Protocol (RTP), which is well suited for real-time wired and wireless multimedia transmissions. This NAL has been used in the simulations.

However, some dependencies exist between the VCL and the NAL. For instance, the packetization process is improved if the VCL is instructed to create slices of about the same size of the packets and the NAL told to put only one slice per packet, thus creating independently decodable packets. The packetization strategy, as the frame subdivision into slices, is not standardized and the encoder has the possibility to vary both of them for each frame. Usually, however, the maximum packet size (hence slice size) is limited and slices cannot be too short due to the resulting overhead that would reduce coding efficiency. In this work we employ one slice per packet with a maximum allowed packet size.

2.2. Analysis-By-Synthesis Distortion Estimation for Video Packets

The quality of multimedia communications over packet networks may be impaired in case of packet loss. The amount of quality degradation strongly vary depending on the importance of the lost data. In order to design efficient loss protection mechanisms, a reliable importance estimation method for multimedia data is needed. Such importance is often defined *a priori*, based on the average importance of the elements of the compressed bitstream, as with the data partitioning approach.

In order to provide a quantitative importance estimation method at a finer level of granularity, we define the importance of a video coding element, such a a macroblock or a packet, as a value proportional to the distortion that would be introduced at the decoder by the loss of that specific element.

The potential distortion of each element, could, therefore, be computed using the analysis-by-synthesis technique [8]. The conceptual scheme is depicted in Figure 1.

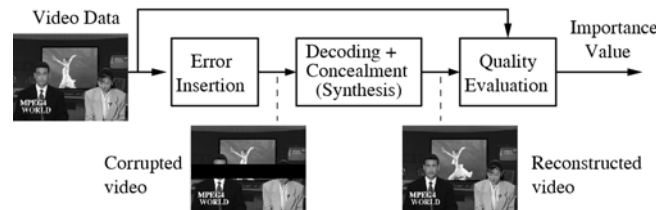


Figure 1: Block diagram of the analysis-by-synthesis technique.

Note that in this work we apply the analysis-by-synthesis technique on a packet basis. The analysis-by-synthesis distortion estimation algorithm performs, for each packet, the following steps:

1. Decoding, including concealment, of the bitstream simulating the loss of the packet being analyzed (synthesis stage);
2. Quality evaluation, that is, computation of the distortion caused by the loss of the packet. The original and the reconstructed picture after concealment are compared using, e.g., MSE;
3. Storage of the obtained value as an indication of the perceptual importance of the analyzed video packet.

The previous operations can be implemented with small modifications of the standard encoding process. The encoder, in fact, usually reconstructs the coded pictures simulating the decoder operations, since this is needed for motion-compensated prediction. If step 1 of the analysis-by-synthesis algorithm exploits the operations of the encoding software, complexity is only due to the simulation of the concealment algorithm. In case of simple temporal concealment techniques, this is trivial and the task is reduced to provide the data to the quality evaluation algorithm.

The analysis-by-synthesis technique, as a principle, can be applied to any video coding standard. In fact, it is based on repeating the same steps that a standard decoder would perform, including error concealment. However, due to the inter-dependencies usually present between data units, the simulation of the loss of an isolated data unit is not completely realistic, particularly for high packet loss rates. Every possible combination of events should ideally be considered, weighted by its probability, and its distortion computed by the analysis-by-synthesis technique, obtaining the expected distortion value. For simplicity, however, we assume that all preceding data units have been correctly received and decoded. Nevertheless, this leads to a useful approximation as demonstrated by some applications of the analysis-by-synthesis approach to MPEG coded video [8].

The results section will show the effectiveness of the proposed video transmission algorithm which relies on these distortion values.

Finally, note that, to reduce complexity, statistical studies on many different video sequences have been conducted and a model-based approach [9] has been developed. According to that model the encoder computes the distortion that would be caused by the loss of the packet into the current frame and then, using a simple formula, it computes an estimation of the total distortion which includes future frames. The reader is referred to the cited work for further details.

2.3. Multimedia Communication Issues over 802.11

This work aim at improving multimedia communications using the 802.11 [10] standard in ad hoc mode for point-to-point transmissions. Note that an extension of the 802.11 standard, namely 802.11p, is currently being developed to provide networking services and protocols in the context of the WAVE communications standards.

The original 802.11 standard implements an immediate per-packet acknowledgment scheme to combat packet losses due to collisions and unreliability of wireless channels. The scheme is based on sending a positive acknowledgment MAC packet for each correctly received one. Packets which are not correctly received are retransmitted by the sending device up to a maximum number of retransmissions, which is referred to as *retry limit*. Thus the exact outcome of the transmission is immediately available at the sender, apart from acknowledgment packet losses. However, traditional MAC implementations do not pass the information to higher layers in the protocol stack, which usually relies on other protocols, such as TCP, to detect losses and retransmit packets if needed.

But end-to-end application-layer retransmission mechanisms are often too slow or would involve sending excessively frequent acknowledgment information to quickly retransmit corrupted or missing packets. Note, in fact, that the maximum end-to-end delay allowed in case of audio and video conference is approximately 150 ms [11].

Hence, using the MAC level acknowledgment information, which is available at no cost, is desirable in case a direct communication is performed. Moreover, it contributes to reduce the risk of a drastic performance decrease of the 802.11 access scheme caused by an excessively high number of packets offered to the network. Cross-layer communication between the MAC and the application layer can be used to signal the outcome of the transmission of each MAC packet. Current 802.11 MAC implementations are designed to automatically retransmit the same packet a number of times up to the retry limit value, but they could be easily modified to communicate with the application

layer to implement different retransmission strategies. Note that these modifications are fully compliant with the 802.11 standard since no changes in the receivers are needed.

Note that an automatic retransmission request mechanism is well suited for the rapidly changing characteristics of intervehicle wireless channels. Differently from alternative solutions such as redundant codes, such retransmission mechanism automatically adapts to different channel conditions, it does not waste bandwidth when channel is good, modifications are standard compliant and it only requires very small modifications of the MAC layer of devices.

In order to improve the performance of multimedia communications, the time sensitiveness of the transmission can also be exploited. In fact, packets which cannot successfully received before their playback time, should not be retransmitted since this would only result in bandwidth waste. Information about the application-level deadline associated to each packet can be communicated to the MAC layer so that it can eliminate from the transmission queue the packets which cannot arrive on time at the decoder, thus reducing the queuing delay of all the remaining packets.

3. THE PERCEPTUALLY-OPTIMIZED PACKET SCHEDULING ALGORITHM

In this work we propose a perceptually-optimized packet scheduling algorithm for video communication over 802.11 networks. The algorithm is compared with two reference transmission algorithms. The proposed algorithm takes advantage of the packet importance estimation technique described in Section 2.2. That technique allows to compute, for each packet, an estimation of the importance of each packet i , expressed as the value of the distortion d_i caused in the decoded video in case the packet is lost. Let L be the set of the packets which have not correctly been received. Under the assumption that the distortion D of a sequence at the receiver can be approximated as the sum of the distortion of the not correctly received packets (as computed by the analysis-by-synthesis technique), i.e.

$$D = \sum_{i \in L} d_i ,$$

the distortion value D of the sequence can be minimized scheduling for transmission, at each transmission opportunity, the packet which presents the highest distortion value. Thus, each time a new packet can be transmitted by the MAC layer, the packet which presents the highest distortion value is selected and sent. This allows to reduce the distortion D of the sequence of the greatest amount, thus the solution is optimal. In case the packet is not correctly received (i.e. MAC acknowledge has not been sent), the same packet is retransmitted until is either correctly

received or another packet with a higher distortion value enters the transmission queue (and it takes priority over the last one). The validity of the approximation is confirmed by the simulation results.

Note that the sender knows which packets have been correctly received and which have not, if the lost acknowledgment case is neglected. Packets which cannot reach the receiving node on time for playback are immediately dropped from the transmission queue not to waste transmission resources. Moreover, there are no a priori limits on the number of times a packet can be retransmitted, differently from the other two reference techniques. The results will show, however, that the average number of retransmissions is similar for all the algorithms. The proposed retransmission algorithm will be referred to as Perceptually-Optimized Scheduling (POS) in the rest of the paper.

Two reference techniques have been implemented for comparison purposes. The first is the standard MAC layer ARQ scheme, which retransmits non-acknowledged packets up to a certain number of times, given by the retry limit value. All 802.11 devices implement this algorithm. Various retry limit values have been used in the comparisons. Note also that the standard MAC layer ARQ scheme discards packets which have been queued at the transmission interface for more than a certain time (called lifetime in the standard, default value is about 500 ms).

A variation of the standard MAC-layer ARQ algorithm, which uses the same retransmission policy, has also been tested. It implements automatic packet dropping from the transmission queue when it determines that the packet cannot reach the receiving node on time for playback. Obviously this requires a form of communication between the application layer, which has the deadline information of each packet, and the MAC layer, which discards packets. Due to this behavior the algorithm is referred to as "Maximum-Delay MAC-layer ARQ" (MD-ARQ) in the rest of the paper.

4. SIMULATION SETUP

In this work an 802.11 transmission between two vehicles traveling along the same road is simulated using actual intervehicle transmission traces, available online at [12]. In the communication experiments packets of 1470 bytes were continuously transmitted between two vehicles using 802.11 devices operating at physical transmission speed of 2 Mbit/s, reaching an approximate packet rate of 150 packet/s. An external antenna was used to improve signal reception. Each trace has been collected trying to keep the distance from the preceding car as constant as possible, but

variations are present due to traffic conditions. See [13] for further details.

This work focuses on an intervehicle communication scenario in which on-board captured video is transmitted to another vehicle, which can use it, for instance, for cooperative visual processing of road information (in case of road video) or for videoconference between the driver and a passenger in the other car (in case the video represents the driver or a passenger).

The data collection effort of the NEDO project [14] includes capturing road and driver's face video sequences, which have been used in this work. Videos have been captured at a resolution of 640x480 pixels, at 30 fps, compressed at high bitrate and stored. They have been later re-encoded for these experiments using the H.264 codec [15], simulating a live encoding with a fixed quantization stepsize, so that video quality is kept approximately constant. The simulations assume the use of the IP/UDP/RTP protocol stack which is well-suited for real-time multimedia transmissions. All simulations use a fixed GOP structure of an I-type frame followed by 11 P-type frames. The length of the video segments used in the experiments is 30 s, and all results are averaged over four simulations using consecutive segments of a given packet loss trace.

The PSNR distortion measure, which is widely accepted in the multimedia communications community, is used to evaluate the performance of the proposed and the reference algorithms. In case of missing packets at the decoder, a simple temporal concealment technique is implemented, that is the missing macroblocks are replaced by the data in the same area in the previous frame. Finally, note that for simulation simplicity, we assume that for all the three techniques acknowledgment MAC packets are always received without errors.

5. RESULTS

This section analyzes the performance of the proposed perceptually-optimized scheduling algorithm (POS), compared with the two techniques described in Section 3, that is the standard MAC-layer ARQ and the MD-ARQ techniques. We employ two packet loss traces, named "A" and "B", which respectively present a packet loss rate of about 27 and 36%. The distance between the two cars is approximately between 300 and 400 m. However, due to the variability of the traffic conditions, each single segment of the traces present values which may significantly differ. Note also that numerous factors may affect channel quality, such as vehicular traffic conditions, presence of obstacles etc., besides the distance between the two cars.

Trace	Standard MAC ARQ	MD-ARQ	POS	Encoding quality
A	38.80	40.19	40.89	44.16
B	37.32	37.82	39.11	

Table 1: PSNR performance of the three techniques. Maximum delay is 150 ms, retry limit equals to seven.

Table 1 shows the PSNR performance of the proposed POS algorithm, compared to the other two reference algorithms. The POS technique outperforms the other two reference techniques, with gains up to 1.3 dB PSNR with respect to the MD-ARQ technique and up to 2.1 dB compared to the standard MAC-layer ARQ technique. These values refer to

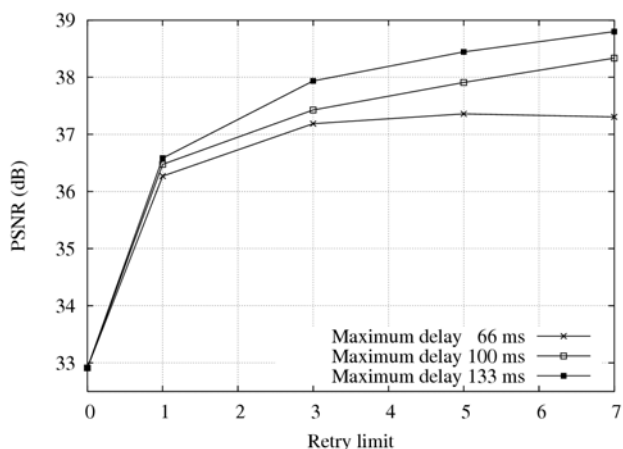


Figure 2: PSNR as a function of the retry limit for different values of the maximum delay. Standard MAC ARQ technique, Trace A.

a videoconference scenario, in which the camera is aimed at the driver's face. Moreover, note that the values for the MD-ARQ and the standard MAC-layer ARQ techniques are the

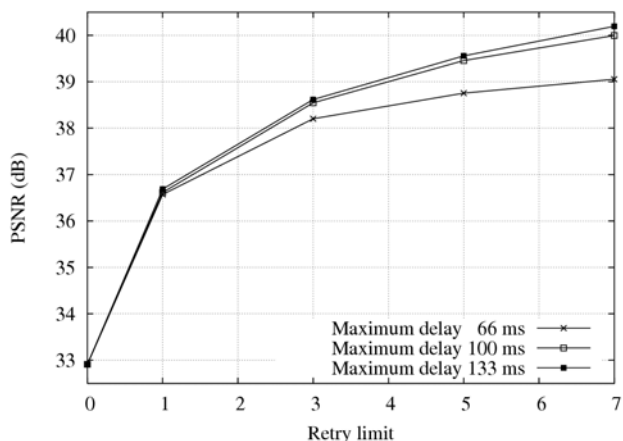


Figure 3: PSNR as a function of the retry limit for different values of the maximum delay. MD-ARQ technique, Trace A.

maximum performance achieved for the given packet loss trace.

Different values of the retry limit parameter, in fact, have been tested for the standard MAC-layer ARQ and the MD-ARQ algorithms. The video quality increases as a function of the retry limit for both the standard MAC-layer ARQ and the MD-ARQ algorithms, as shown in Figure 2 and 3, respectively. They achieve the maximum performance when the retry limit is set to seven. Various curves show the performance for different values of the maximum allowed delay. The MD-ARQ technique presents a significant performance gain if the maximum delay is increased over 66 ms, which corresponds to the time of two frames, then the performance tends to saturate. The standard MAC-layer ARQ, instead, significantly benefits from the increase of the maximum allowed delay, with gains up to 1.5 dB PSNR.

Table 2 shows the average number of transmissions of MAC layer packets, for the same experiments of Table 1. The standard MAC-layer ARQ technique presents the highest average number of retransmissions but many packets are then dropped by the application-layer of the receiver because they are too late for playback. The performance of the MD-ARQ and the proposed POS technique is similar for high retry limit values. However, note that the proposed POS technique presents a slightly lower channel usage in terms of number of transmitted MAC packets, while the PSNR value is higher than the other techniques, as shown in Table 1.

Finally we assessed the performance of the proposed POS technique and the two reference techniques as a function of the maximum allowed delay. Results are shown in Figure 4. The proposed POS technique shows a clear PSNR performance increase (up to 0.7 dB) compared with the MD-ARQ technique when the maximum allowed delay is increased over 100 ms, which is the time interval corresponding to 3 video frames. The performance of the standard MAC-layer ARQ technique is much lower, about 1-1.5 dB less than the MD-ARQ technique, regardless of the maximum allowed delay, because it is greatly influenced by the delays caused by packet accumulation in the queue when channel conditions are bad.

Trace	Standard MAC ARQ	MD-ARQ	POS
A	1.64	1.60	1.59
B	1.75	1.73	1.72

Table 2: Average number of MAC-layer transmissions for all the techniques. Maximum delay is 150 ms.

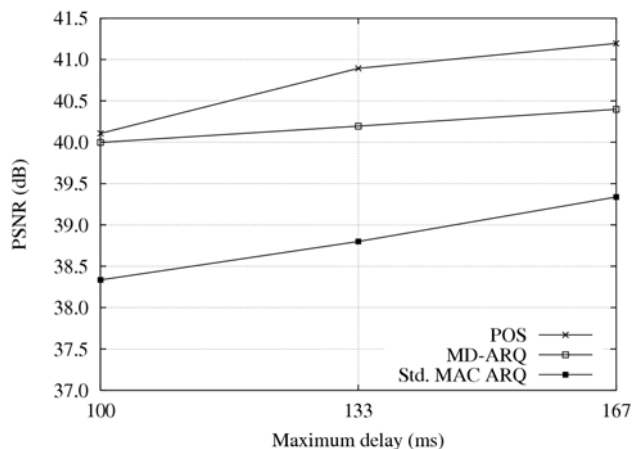


Figure 4: PSNR performance as a function of the maximum delay for all the techniques. Maximum delay is set to 150 ms, Trace A.

6. CONCLUSIONS

In this paper we presented a perceptually-optimized algorithm for robust low-delay intervehicle video transmission using standard 802.11 wireless technology. The algorithm exploits the non-uniform perceptual importance of packets to give higher priority to the most important ones. Transmission of in-car and road video has been simulated using 802.11 intervehicle packet level error traces. The results show that the proposed perceptually-optimized algorithm achieves a consistently higher quality compared to two reference techniques, with gains up to 2 dB PSNR. Various error traces and different parameters settings have been used to verify the consistency of the improvements. Moreover, the results also show that timely dropping packets which cannot reach the receiver on time for playback greatly improves the performance of the transmission systems.

7. ACKNOWLEDGEMENT

The authors would like to thank the NEDO project [14] and in particular prof. Takeda, prof. Abut and all the participants of the project, who collected and provided the video traces used in this work, and the CRAWDAD project for the intervehicle transmission traces.

11. REFERENCES

[1] J. Ott, D. Kutscher, "Drive-thru Internet: IEEE 802.11b for Automobile Users," *Proc. IEEE INFOCOM*, March 2004, vol. 1, pp. 362-373.
 [2] J.P. Singh, N. Bambos, B. Srinivasan, D. Clawin, "Wireless LAN performance under varied stress conditions in vehicular

traffic scenarios," *Proc. IEEE Vehicular Technology Conference (VTC)*, Sept. 2002, vol. 2, pp. 743-747.

[3] S.Y. Wang, C.C. Lin, Y.W. Hwang, K.C. Tao, and C.L. Chou, "A practical routing protocol for vehicle-formed mobile ad hoc networks on the roads," *Proc. of IEEE Intl. Conf. on Intelligent Transportation Systems*, Vienna, Austria, September 2005, pp. 161-166.

[4] L. Wischhof, A. Ebner, H. Rohling, "Information dissemination in self-organizing intervehicle networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 90-101, March 2005.

[5] A. Pal, A. Dogan, F. Ozguner, U. Ozguner, "A MAC layer protocol for real-time inter-vehicle communication," *Proc. of IEEE Intl. Conf. on Intelligent Transportation Systems*, Singapore, September 2002, pp. 353-358.

[6] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," ITU-T, May 2003.

[7] 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, pp. 560-576, July 2003.

[8] E. Masala, J.C. De Martin, "Analysis-by-synthesis distortion computation for rate-distortion optimized multimedia streaming," *Proc. IEEE Int. Conf. on Multimedia & Expo*, Baltimore, MD, July 2003, vol. 3, pp. 345-348.

[9] F. De Vito, D. Quaglia, 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, September 2004, vol. 1, pp. 79-82.

[10] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ISO/IEC 8802-11, ANSI/IEEE Std 802.11, 1999.

[11] ITU-T G.114, "One-way transmission time," ITU-T, May 2000.

[12] CRAWDAD project, "A community resource for archiving wireless data at Dartmouth," URL: <http://crawdad.cs.dartmouth.edu>.

[13] H. Wu, M. Palekar, R. Fujimoto, R. Guensler, M. Hunter, J. Lee, J. Ko, "An empirical study of short range communications for vehicles," *Proc. of the 2nd ACM international workshop on Vehicular ad hoc networks*, Cologne, Germany, September 2005, pp. 83-84.

[14] Project NEDO, "Driving behavior signal processing based on large scale real world database," URL: <http://www.sp.m.is.nagoya-u.ac.jp/NEDO>.

[15] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Joint Model v.11.0," URL: <http://iphome.hhi.de/suehring/tml>, Jan 2007.