Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Quality Driven Wireless Video Transmission for Medical Applications

Maria G. Martini and Matteo Mazzotti
CNIT, DEIS - University of Bologna
V.le Risorgimento 2
40136 Bologna, ITALY
(e-mail: mgmartini@deis.unibo.it)

Abstract— **Wireless telemedicine is currently a reality, requiring also the transmission of medical video sequences over often unreliable links. The contrasting requirements of almost lossless compression and low available bandwidth have to be tackled in this case. On one side compression techniques need to be conservative, in order to avoid removing perceptively important information; on the other side error resilience and correction should be provided, with the constraint of a limited bandwidth. An approach based on quality driven, network aware, joint source and channel coding is described in this paper. The approach has been developed in the framework of the IST PHOENIX project (www.ist-phoenix.org), focusing on wireless multimedia transmission over IP networks. After a description of the considered cross-layer approach and of the information to be exchanged among the system component blocks, the techniques considered for this information exchange and the concept of "JSCC/D controllers" are introduced. The implementation of the demonstrator realized is then described.**

I. Introduction

The accessibility of expensive technological resources to any citizen, regardless of where the resources are located, is a clear need of today's medicine. This is the case of medical image resources in hospitals. The existence of this equipment does not imply the presence in hospitals of all specialists involved in the final diagnosis: a solution based on providing remote medical users with access to medical image databases to aid cooperative diagnosis is more rational and profitable.

Furthermore, wireless access to remote video databases for clinical analysis visualization in emergency field is an important application.

Video compression techniques for medical applications have requirements of high fidelity, in order to avoid the loss of information that could help diagnosis. In order to keep diagnostic accuracy, lossless compression techniques are often considered when medical video sequences are involved. Anyway, when transmission is over band limited, error prone channels, a compromise should be made between compression fidelity and protection and resilience to channel errors and packet loss. It has been observed [1] that when lossy compression is limited to ratios from 1:5 to 1:29, compression can be achieved with no loss in diagnostic accuracy. Furthermore, even if the final diagnosis should be done using an image that has been reversibly compressed, irreversible compression still play a critical role when quick access to data stored in a remote location is needed.

For these reasons, lossy compression techniques have been considered for medical images and video compression (see, e.g., [2]).

We describe in the following our approach based on the already known joint source channel coding and decoding (JSCC/D) paradigm, aiming at developing strategies where the source coding (video compression), channel coding (protection from channel errors) and modulation parameters are jointly determined to yield the best end-to-end system performance. The approach has been developed in the framework of the EU IST project "PHOENIX", where wireless telemedicine and wireless access to remote video databases in emergency areas is one of the considered scenarios.

We take into account compression based on recent video coding standards, such as MPEG-4 and H.264, where the available error resilience tools may help keeping an acceptable quality after transmission.

The developed architecture can be then utilized in different practical contexts, whether at the radio link level, with e.g., orthogonal frequency division multiplexing (OFDM) or wide band code division multiple access (WCDMA), with multiple input/multiple output (MIMO) and adaptive antennas or at application level, with e.g. MPEG-4 or H.264/AVC.

This approach is apparently in contrast the so called "separation theorem" derived from Shannon's theory [3], since it was shown that for wireless audio and video transmission separation does not necessarily lead to the less complex solution [4], nor is always applicable, in particular when transmitting data with real-time constraints or operating on sources whose encoded data bit error sensitivity varies significantly. Recently, JSCC/D techniques that include a co-ordination between source and channel encoders were investigated, improving both encoding and decoding processes while keeping the overall complexity at an acceptable level [5] [6] [7].

In this paper, a quality driven approach for wireless video transmission relying on the joint source and channel coding paradigm is thus proposed. We consider several metrics for evaluating video quality, from the "classic" peak signal-to-noise ratio (PSNR) to metrics evaluating structural distortion and metrics that to dot need reference to the original image [9] [10].

The management of the information to be exchanged among the system component blocks is addressed and the logical units responsible for the system optimization, in the following referred to as joint source channel coding and decoding (JSCC/D) controllers, having a key role in the system, are analyzed. The demonstrator of the described system realized in the framework of the IST PHOENIX project is also presented.

II. System Architecture

Figure 1 illustrates the overall system architecture developed in the framework of the PHOENIX project, from the transmitter side in the upper part of the figure to the receiver side in the lower part, and including the signalisation used for transmitting the JSCC/D control information in the system. Beside the traditional tasks performed at application level (source encoding, application processing such as ciphering), at network level (including RTP/UDPLite/IPv6 packetisation, impact of IPv6 wired network and Robust header Compression (RoHC)), medium access (including enhanced mechanisms for WiFi) and radio access (channel encoding, interleaving, modulation), the architecture includes two controller units at physical and application layers. Those controllers are used for supervising the different (de)coders, (de)modulation, (de)compression modules and to adapt said modules parameters to changing conditions, through the sharing of information about the source, network and channel conditions and user requirements. For the controlling purpose, a signalling mechanism has been defined, as detailed in the following.

A. *Side information exchanged in the system and methods for exchanging it*

In particular, the information that are taken into account by the system for optimization are source significance information (SSI), *i.e.* the information on the sensitivity of the source bitstream to channel errors; channel state information (CSI); decoder reliability information (DRI), *i.e.* soft values output by the channel decoder; source a-priori information (SRI), *e.g.* statistical information on the source like importance of data elements; source a-posteriori information (SAI), *i.e.* information only available after source decoding; network state information (NSI), represented *e.g.* by packet loss rate and delay and finally the video quality measure, output from the source decoder and used as feedback information for system optimization. This last measure is critical as the target of the overall system optimisation is the maximisation of the received video quality. In practice, due to the fact this quality measure is regularly sent back to the transmitter side for action, the evaluation should be performed "on the fly" and without reference (or reduced reference) to the transmitted frame.

Naturally, when considering real systems, this control information needs to be transferred through the network and systems layers, in a timely and bandwidth efficient manner. The impact of the network and protocol layers is quite often neglected when studying joint source and channel coding and only minimal effort is made into finding solutions for providing efficient inter-layer signalling mechanisms for JSCC/D. Different mechanisms have been identified by the authors, that could allow information exchange transparently for the network layers (which we call *Network Transparency* concept). Besides the said novel solutions, one should not forget that several transport protocols exist, each of which can carry the payload and also some control information. In particular, UDP, UDPlite and datagram congestion control protocol (DCCP) protocols are considered at transport layer level. For further information on such solutions, please refer to [11].

Finally, it should be noted that additional information is requested by the system for the set-up phase, where information on available options (*e.g.* available channel encoders and available channel coding rates, available modulators, ...) are exchanged, the session is negotiated and default parameters are set (*e.g.* authentication key, modules default settings).

B. *Principle of JSCC/D controllers*

The system controllers are represented by two distinct units, namely the "physical layer (PHY) controller" and the "application layer (APP) controller". The latter collects information from the network (NSI: packet loss rate, delay and delay jitter) and from the source (*e.g.* SSI) and has availability of reduced channel state information and of the quality metric of the previously decoded frame (or group of frames). According to this information, it produces controls for the source encoder block (*e.g.* quantization parameters, frame rate, error resilience tools to activate) and for the network. The physical layer (PHY) controller's task is to provide controls to the physical layer blocks, *i.e.* the channel encoder, modulator and interleaver. A more detailed description of the controllers is made in section III, where exemplifications of their behavior are also provided.

III. The demonstration platform

In order to provide a realistic performance evaluation of the proposed approach, all the involved system layers blocks have been realistically implemented. Namely: application layer controller; source encoder/decoder (three possible codecs: MPEG-4, H.264/AVC and Scalable Video Coding in H.264/AVC Annex F), where soft-input source decoding is also allowed for H.264/AVC; cipher/decipher unit; content level unequal error protection (UEP) block, using rate compatible punctured convolutional (RCPC), alternative to UEP at PHY level, and leading to equal error protection (EEP) at PHY level when activated; real time transport protocol (RTP) header insertion/removal; transport protocol header (*e.g.* UDPLite, UDP, or DCCP) insertion/removal; IPv6 header insertion/removal; IPv6 mobility modelling; IPv6 network simulation; Robust Header Compression (RoHC); DLL header insertion/removal; Radio Link, including physical layer controller, channel encoder/decoder (convolutional, RCPC, low density parity check (LDPC) codes with soft and iterative decoding allowed), interleaver, modulator (OFDM, TCM, TTCM, STTC; soft and iterative demodulation allowed) and channel. A more detailed description is provided in the following for the blocks whose impact is highlighted in the results section, in particular for the controlling units.

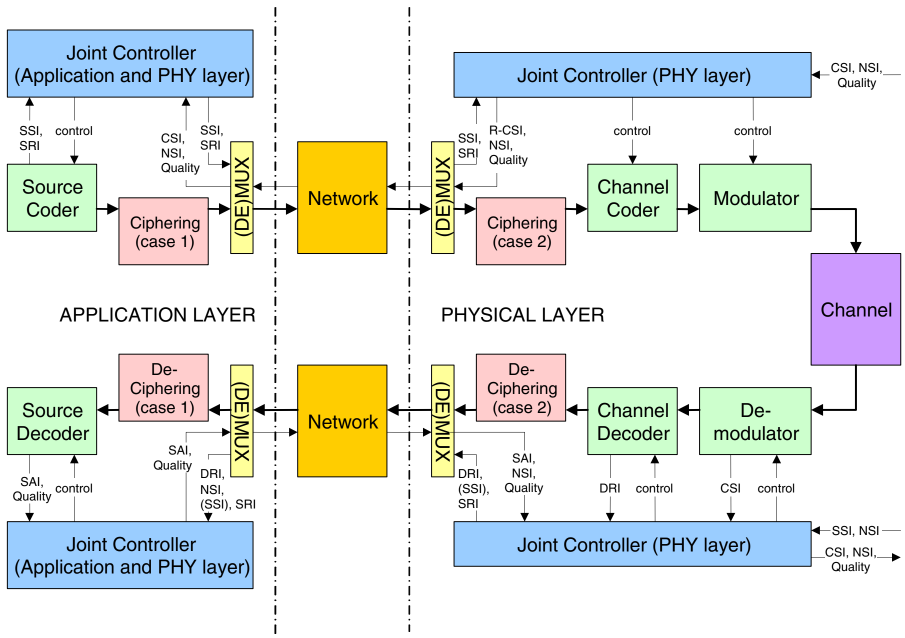

Fig. 1. PHOENIX system architecture for joint optimisation.

A. Application layer (APP) controller

The application controller has been modelled as a finite state machine. At the beginning of each iteration cycle of the duration of one second (corresponding roughly to one or two group of pictures (GOP)), it decides the next operating state, which is defined by a fixed set of configuration parameters for the different blocks of the chain. According to these values, the application layer (APP) controller updates its configuration parameters. The choice of the new state is based on the history and feedback information coming from the blocks at the receiver side, relevant to the previous cycle.

The feedback information are:

- Quality: peak signal-to-noise ratio (PSNR) or other quality metric (e.g. based on structural distortion [9], or without reference to the original sequence [10]). A no-reference or reduced reference metric should be considered in realistic implementation; in particular, attention should be paid on the choice of a metric representing diagnostic accuracy, which is the final goal of medical image and video transmission;
- Reduced CSI: average signal-to-noise ratio (SNR) in one controller step, channel coherence time;
- NSI: number of lost packets, average jitter, average round trip time (RTT).

The main configuration parameters set by the APP JSCC/D and modifiable at each simulation step are the video encoder frame rate, quantization parameters, and GOP size; Code rates $R_{c,i}$, where the index i is related to the ith sensitivity class, for content UEP if applied; average channel code rate, as a consequence of the choice of the source encoding parameters and the knowledge of the available bandwidth.

In order to reduce the dimension of the possible configurations, and to avoid switching continuously from one given set of parameters to another, which is not efficient in terms of compression for any source encoder, the demonstrator takes into account only a limited set of possibilities for these parameters. In particular the choice made considers frame rates of 30, 15 and 7.5 fps; MPEG-4 quantization parameters (frame I, frame P) equal to (6,10) or (10,14); GOP length assuming values of 8, 15 and 30. Furthermore, some constraints on these values must be satisfied, thus the number of the controller states is reduced.

Typically, a low video quality value associated to a negative trend will cause a transition to a state characterized by a higher robustness. Given the bit-rate of the chosen state, the code rate available for signal protection is evaluated considering the total $R_{max} = R_s/R_c$ constraint, where R_s is the average source coding rate and R_c is the target average protection rate. That R_c target information is used either for embedded unequal error protection at the application level or provided directly to the physical layer controller. If physical layer UEP is adopted, given the available total coded bit rate (R_{max}), the average channel coding rate (R_c) is derived by the application JSCC/D controller and proposed to the PHY controller. The knowledge of the bit-rate is of course approximated and based on rate/source parameters models developed by the authors or average values evaluated in previous controller steps.

As an example, for the considered scenario five different

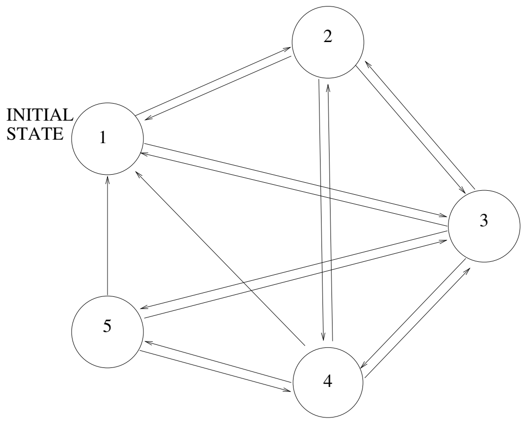

Fig. 2. A graphical representation of the APP-JSCC as a finite state machine with 5 states.

states have been chosen for the APP joint source and channel coding (JSCC) controller, characterized by different set of values for the above mentioned parameters. State 1 corresponds to the lowest source data rate (lowest video quality) and highest robustness, whereas state 5 corresponds to the highest source data rate (highest video quality) and lowest robustness. Thus, increasing the state number means increasing the robustness transmission at cost of a loss in the error free received video quality. Figure 2 depicts the finite state machine describing the APP–JSCC controller with the allowed transitions among states. More precisely, the number of possible parameter sets is seven, since state 3 and state 5 have two different options for the GOP length. The choice of the GOP length is made according to channel conditions: for average channel E_s/N_0 below a prefixed threshold the lower one is chosen, whereas for higher values of E_s/N_0 the higher one is preferred. The adaptive algorithm which has been tested takes into account the trend of the video quality feed-back from the source decoder. When there is a network congestion, indicated by an high value for the packet loss rate (PLR) feed-back in the NSI, the controller sets immediately the state to the first, characterized by the lowest source bit rate, in order to reduce as much as possible the amount of data which have to flow through the IPv6 network.

B. *Physical layer controller*

The Physical layer controller's task is to decide, similarly as in [6], on the channel coding rate for each different sensitivity source layer, with the goal of minimizing the total distortion D_{S+C} with the constraint of the average channel coding rate R_c, provided by the application controller. Furthermore, the controller sets the parameters for bit-loading in multicarrier modulation, interleaver characteristics and performs a trade-off with receiver complexity. Again, the metric chosen for representing distortion should be representative of diagnostic accuracy. In this view, the identification of regions of interest (ROI), e.g. in terms of video objects in the case of MPEG-4 video, allows dedicating a higher protection to the region of interest, allowing an increase in diagnostic accuracy, for a fixed available bandwidth.

IV. Conclusions

A global approach for realistic network-aware joint source and channel system optimization, suited for wireless transmission of medical video sequences, has been outlined in the paper. The information to be exchanged among the system blocks are described, together with the role of the "joint controllers", managing the system according to the received side information. The demonstration platform realized in the framework of the PHOENIX project is also described.

V. Acknowledgement

This work has been partially supported by the European Commission in the framework of the "PHOENIX" IST project under contract FP6-2002-IST-1-001812. The whole "PHOENIX" IST project consortium, contributing to the definition of the system architecture and to the demonstrator implementation, is also gratefully acknowledged.

References

[1] P. C. Cosman et al., "Thoracic CT Images: Effect of Lossy Image compression on Diagnostic Accuracy", Radiology, 190: 517-524, 1994.

[2] H. Yu, Z. Lin, F. Pan, "Applications and Improvement of H.264 in Medical Video Compression", IEEE Transactions on Circuits and Systems I, Special issue on Biomedical Circuits and Systems: A New Wave of Technology, Vol. 52, No. 12, pp. 2707-2716, Dec. 2005.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July-Oct. 1948.

[4] J. L. Massey, "Joint source and channel coding," in *Commun. Systems and Random Process Theory, NATO Advanced Studies Institutes Series E25*, J.K. Skwirzynski editor, pp. 279-293, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.

[5] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proceedings of the IEEE*, vol. 87, no. 10, Oct. 1999.

[6] M.G. Martini, M. Chiani, "Rate-Distortion models for Unequal Error Protection for Wireless Video Transmission," *IEEE VTC 2004 Conference*, Milan, Italy, May 2004.

[7] L. Perros-Meilhac and C. Lamy, "Huffman tree based metric derivation for a low-complexity sequential soft VLC decoding," in *Proceedings of IEEE ICC'02*, New York, USA, vol. 2, pp. 783–787, April-May 2002.

[8] M. G. Martini, M. Chiani, "Proportional Unequal Error Protection for MPEG–4 video transmission", *Proc. IEEE ICC 2001*, Helsinki, June 2001.

[9] Z. Wang, L. Lu, A. C. Bovik, "Video quality assessment based on structural distortion measurement", *Signal processing: Image communication*, vol. 29, no. 1, Jan. 2004.

[10] Zhou Wang, Hamid R. Sheikh and Alan C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images", *Proc. IEEE International Conference on Image Processing*, pp. 477-480, Rochester, New York, September 22-25, 2002.

[11] M. G. Martini, M. Mazzotti, C. Lamy-Bergot, P. Amon, G. Panza, J. Huusko, J. Peltola, G. Jeney, G. Feher, S. X. Ng, "A Demonstration Platform for Network Aware Joint Optimization of Wireless Video Transmission", IST Mobile Summit 2006, Myconos, Greece, June 2006.