Wireless Medical Ultrasound Video Transmission Through Noisy Channels

A. Panayides[1], M.S. Pattichis[2], and C. S. Pattichis[1]

Abstract—**Recent advances in video compression such as the current state-of-the-art H.264/AVC standard in conjunction with increasingly available bitrate through new technologies like 3G, and WiMax have brought mobile health (m-Health) healthcare systems and services closer to reality. Despite this momentum towards m-Health systems and especially e-Emergency systems, wireless channels remain error prone, while the absence of objective quality metrics limits the ability of providing medical video of adequate diagnostic quality at a required bitrate. In this paper we investigate different encoding schemes and loss rates in medical ultrasound video transmission and come to conclusions involving efficiency, the trade-off between bitrate and quality, while we highlight the relationship linking video quality and the error ratio of corrupted P and B frames. More specifically, we investigate IPPP, IBPBP and IBBPBBP coding structures under packet loss rates of 2%, 5%, 8% and 10% and derive that the latter attains higher SNR ratings in all tested cases. A preliminary clinical evaluation shows that for SNR ratings higher than 30 db, video diagnostic quality may be adequate, while above 30.5 db the diagnostic information available in the reconstructed ultrasound video is close to that of the original.**

***Index Terms*—H.264/AVC, mobile health systems, wireless video streaming, carotid ultrasound video, error resilience.**

I. Introduction

M-Health systems and services gained a tremendous momentum over the past decade. A thorough overview of the current status, highlighting future directions, while also covering incorporated wireless transmission technologies, is given in [1], [2].

Despite numerous studies in video transmission, very few are focused on medical video transmission over wireless environments [3], [4]. There is a clear need of exploiting new features introduced by current state of the art H.264/MPEG-4 (part 10) [5] video compression standard (i.e. error resilience), linked with new wireless transmission technologies such as 3G (UMTS [6]) and WiMax [7], etc.

This paper reports on a preliminary study attempting to tackle the aforementioned implications. Namely, different frame encoding schemes are tested using the JM 13.2 H.264 Reference Software, while RTP packet drops aim to simulate the transmission errors likely to occur when streaming over error prone wireless channels. Choosing the appropriate encoding scheme is essential. Different encoding schemes incorporate different encoding times, result in different compression sizes and yield different resilience to the presence of errors. In other words, the appropriate coding structure must fulfill certain application specific criteria.

The rest of the paper is organized as follows. Section II introduces the fundamental concepts of video streaming architecture. Section III describes the methodology, while Section IV presents an analysis of the preliminary results. Finally Section V provides some concluding remarks.

II. Video Streaming Architecture, Protocols and Encoding Modes

A typical video streaming architecture is illustrated in Fig. 1. For the purposes of this paper, we are interested in describing (i) the protocols and their relationship to the packets drops in a noisy environment and (ii) the frame encoding scheme and how they relate to error resilience.

A. Protocols

Video streaming protocols involved in the procedure can be classified as follows: *session control protocols*, *transport protocols* and *network protocols*. Session control protocols such as the Real Time Streaming Protocol (RTSP) [8] or alternative Session Initiation Protocol (SIP) [9], are responsible for session initialization between client and server. Transport protocols are further distinguished in upper and lower layer, Real-Time Transport Protocol (RTP) [10], and UDP and TCP respectively. RTP payload contains the real time data being transferred while the RTP header contains information characterizing the payload such as timestamp, sequence number, source, size and encoding scheme. RTP packets are usually transferred over UDP, which in turn are encapsulated in the network layer in IP packets, hence RTP/UDP/IP headers. Given that RTP

This work was partially funded via the *Mobile and Robotised Telemedicine for Emergency Vehicle* (MARTE III) project, under the program Cyprus – France (Zenon) cooperation, funded by the Research Promotion Foundation of Cyprus.

[1]A.Panayides and C.S.Pattichis are with the Department of Computer Science, University of Cyrus, Nicosia, Cyprus (tel:+357-22892697, e-mail: {panayides, pattichi}@ucy.ac.cy).

[2]M.S.Pattichis is with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, USA (e-mail: pattichis@ece.unm.edu).

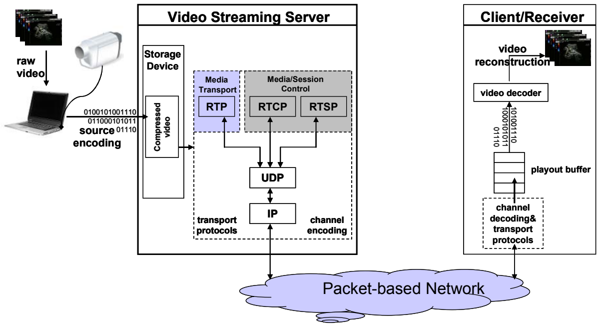


Fig. 1. A typical video streaming architecture.

header and payload contain all video data and information characterizing the transmitted sequence, for the purposes of this paper we simulate transmission errors occurring in the network layer by dropping RTP packets (see Section III).

B. *Encoding Modes and Frame Types*

Frame encoding modes can have a significant impact on both error propagation and video compression performance. We provide a summary of the different modes:

Intra-mode: Intra-mode is the procedure where intra-prediction is used for coding a video frame (I-frame). That is, all the information used for coding originate from the picture itself and block samples are predicted using spatially neighboring samples of previously coded blocks.

Inter-mode: Inter-mode is the procedure where inter-prediction is used for coding a video frame:

P-mode: P-mode uses prediction from previously decoded frames. In inter-mode, the encoder's side provides all the necessary information for accurate motion estimation of the spatial displacement between the decoder's reference picture and the current picture in the sequence at the encoder. This procedure is described as motion compensation.

B-mode: Whereas in P-mode at most one motion compensated signal is employed, B-mode provides the ability to make use of two motion compensated signals for the prediction of a picture. B-mode is also referred to as bi-prediction as not only it allows the utilization of previously decoded pictures but also forthcoming ones.

Again, the extensive use of predictive coding (P-frames, B-frames) or not (I-frames), is application specific. Depending on time and quality constraints imposed, one mode may be preferred over the other and the other way round. Intra coding is mostly employed as en error resilience feature for periodic updates (i.e. every Group of Picture-GOP).

The question which arises next is the ratio between P-frames and B-frames used. P-frames employ unidirectional prediction as mentioned above whereas B-frames bidirectional. For P-frames this is translated into less motion estimation time with more bits per picture. The opposite stands for B-frames where the prediction from both previous and forward frames results in greater motion estimation time during encoding but less bits per picture for transmission.

III. METHODOLOGY

A. *Material*

Ultrasound is widely used in vascular imaging because of its ability to visualize body tissue and vessels in a non invasive and harmless way and to visualize in real time the arterial lumen and wall, something that is not possible with any other imaging technique. B-mode ultrasound imaging can be used in order to visualize arteries longitudinally from the same subject in order to monitor the development of atherosclerosis.

Monitoring of the arterial characteristics like the vessel lumen diameter, the intima media thickness (IMT) of the near and far wall and the morphology of atherosclerotic plaque are important in order to assess the severity of atherosclerosis and evaluate its progression [12]. For this particular study, three carotid ultrasound videos were used, illustrated in Table I. The 1st video sequence records a normal (without plaque) case while the 2nd and 3rd an abnormal one (with plaque).

B. *Encoding/Decoding Procedure*

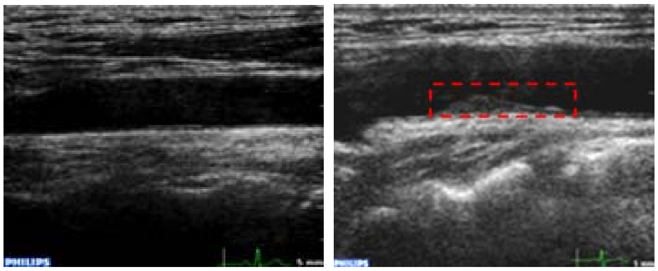
Using the JM 13.2 Reference Software [11], we conducted tests for three different encoding schemes summarized in Table II. The default JM rate control algorithm [13] was applied to explore the trade-off between video quality and bit rate, performing frame level adaptation. Firstly, we need to select the proper initial quantization parameter (QP). The algorithm encodes the first frame using the provided QP and given the remaining frames (in the GOP) and available bitrate, it adapts accordingly to try to match the target bit rate. The initial QP was selected according to the following formula [13]:

$$QP = \begin{cases} 40 & bpp \le 0.15 \\ 30 & 0.15 < bpp \le 0.45 \\ 20 & 0.45 < bpp \le 0.9 \\ 10 & bpp > 0.9 \end{cases} \quad (1),$$

where
$$bpp = \frac{R}{f \times N_{pixels}} \quad (2)$$

and *bpp* stands for the target bits per pixel, N_{pixels} is the number of pixels in the N-th frame, R is the target bit rate and f is the encoded video sequence frame rate.

To evaluate the performance of the aforementioned encoding schemes in error prone wireless environments, the pseudo-random RTP packet loss simulator included in JM was employed [11]. By dropping packets in the generated RTP files, we simulate the transmission errors likely to occur in today's unreliable wireless links. However, making use of the keep_leading_packets option, the initial RTP packets preceding the video data RTP payload packets carrying picture and sequence parameters sets is avoided, loss of which is beyond the scope of this study. One frame



a) b)

Fig. 2. The 4th frame of uncompressed source carotid ultrasound videos with ECG, 176x144, 25fps a) normal, without plaque, and b) abnormal, with atherosclerotic plaque (depicted by the dashed red line).

TABLE I
ENCODED ULTRASOUND VIDEOS OF THE CAROTID ARTERY

Videos	Resolution	Frames per second	No. of Frames	Case
1st	QCIF, 176x144	25	102	normal
2nd	QCIF, 176x144	25	119	abnormal
3rd	CIF, 352x288	25	100	abnormal

TABLE II
ENCODING SCHEMES

Scheme	Encoding/ Decoding Order	No. I-P/B - Frames 1st	2nd	3rd
IPPP…	1 2 3 4 …	1-101/0	1-118/0	1-99/0
IBPBP…	1 3 2 5 4 …	1-50/50	1-59/59	1-49/49
IBBPBBP…	1 4 2 3 7 5 6 …	1-33/66	1-39/78	1-33/66

roughly equals one RTP packet. Frame Copy error concealment method is applied at the decoder to reconstruct corrupted packets.

IV. RESULTS

A. Technical Evaluation

Three typical ultrasound videos were encoded using three different encoding schemes as illustrated in Tables I and II. Figure 3 depicts the trade-off between quality and bit rate for the three videos investigated.

Naturally, the more bits that are allocated for source encoding using rate control, the higher SNR quality is attained. IBPBP and IBBPBBP coding structures behave similarly, while IPPP coding structure achieves slightly lower SNR values. Typically, bidirectional prediction requires less bits during encoding than single-directional prediction. However, single-directional prediction is marginally quicker in terms of encoding time due to the increased motion estimation time required for bidirectional encoding. Due to the increased resolution over QCIF (4x), the CIF sequence (3rd video) requires significantly higher bitrate to achieve comparable SNR values.

An important aspect which is clearly visible by looking at the graph is that the ultrasound video of the normal carotid case attains higher SNR ratings than the abnormal case while at the same time it consumes considerably less bits (QCIF sequences). This is partly due to the fact that these extra bits are needed to encode the plaque and richer artery structures (and their motions) in the abnormal video.

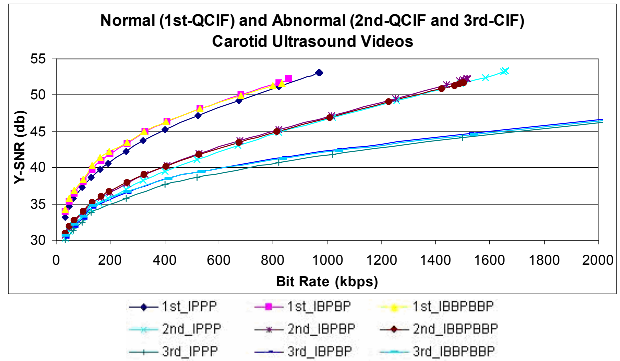


Fig. 3. Trade-off between quality and bit rate for the tested coding structures after applying the default JM H.264 Reference Software rate control algorithm on normal (1st-QCIF) and abnormal (2nd-QCIF and 3rd-CIF) carotid ultrasound videos.

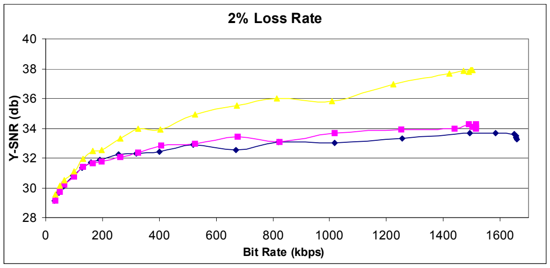
Figure 4 demonstrates the performance of the three tested encoding schemes under losses of 2%, 5%, 8% and 10% of transmitted RTP packets for the 2nd video. One can clearly see that IBBPBBP coding structure has the best SNR output in all aforementioned packet losses but the last one of 10%, where it is comparable with IBPBP. In low-noise environments, bidirectional prediction gives the best performance. However, as the noise level increases, the use of more frames with single-directional prediction provides for better error recovery and better results, still performing better than the exclusive use of single-directional prediction (see Figs 4c, 4d). On the other hand, the IPPP coding structure performs worst in all cases, although for packet loss rates of 2% and 5% its SNR output is similar to IBPBP.

By examining the results obtained by averaging 10 consecutive runs for each scheme with different number of leading packets preserved, we came to the conclusion that quality is directly affected by the loss ratio of P to B frames. High ratio (more P-frames dropped) is translated into poor quality, whereas low ratio (more B-frames dropped) results into better quality. Given these measurements we deduct that IBBPBBP encoding scheme is more suitable for encoding and transmission for this series of experiments as it attained the higher SNR values for the majority of cases.

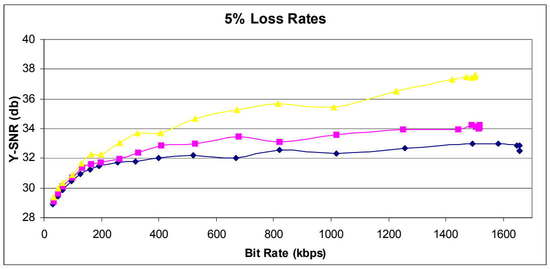
B. Clinical Evaluation

The tested coding structures' performance was also evaluated by a medical expert so as to provide the level of diagnostic quality. The videos were played back on a Laptop at their original pixel size dimensions.

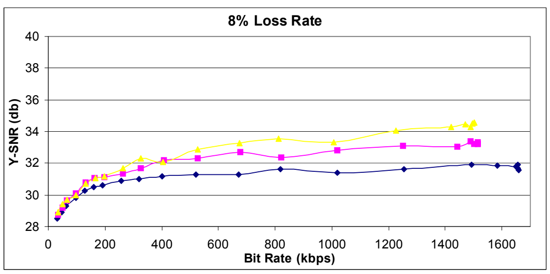
The evaluation recorded that videos achieving SNR ratings higher than 30 db may be suitable for providing diagnosis. That is, there is sufficient information in the ultrasound video that enables the medical expert to make a confident diagnosis. This information was also available to videos attaining lower SNR ratings in some cases. This was made possible when the medical expert freezed a relatively clean frame to use for diagnosis. Moreover, above 30.5 db the medical expert could almost identify as much diagnostic



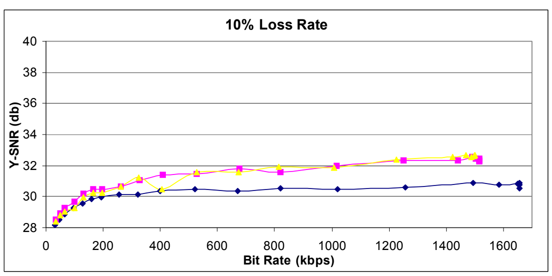
a)



b)



c)



d)

Fig. 4. Tested encoding schemes performance after decoding the 2nd video stream at the presence of a) 2%, b) 5%, c) 8% and d) 10% packet loss rate. Y-SNR(db) vs Bit Rate(kbps).

information in the compressed video as in the original video sequence. It is worth noting here that at 96 kbps, when the intial QP switches from 40 to 30 (according to (1) and (2)), for 2% and 5% loss rates this is translated into SNR ratings higher than 30.5 db (except at 5%, IPPP, 30.44 db), while for 8% higher than 30 db.

We also observed motion delays when using bidirectional prediction (more obvious on IBBPBBP, in the presence of heavy loss rates). However, diagnostic quality is not affected by this observation. The medical expert was emphatic that the carotid ultrasound video used in this particular case was a very clear case.

V. Concluding Remarks

This paper deals with medical video streaming over wireless networks. It attempts to exploit H.264/AVC features towards resilient video streaming over today's error prone wireless networks. More specifically, to find which encoding scheme achieves better quality in the presence of error introduced by the network. For this particular set of experiments we derive that IBBPBBP coding structure has the best performance in terms of SNR quality.

Future work includes testing of more encoding schemes (such as hierarchical coding), exploiting H.264/AVC error resilience techniques for different cases (i.e. different medical videos), while using a network simulator/ emulator to introduce packet loss for a more realistic approach. Scalable video coding applying spatio-temporal and quality scalability is planned.

Acknowledgment

The Authors would like to thank medical experts Prof. A. Nicolaides for providing the carotid ultrasound videos and Dr. M. Pantziaris for the perceptual video evaluation.

References

[1] E. Kyriacou, M.S. Pattichis, C.S. Pattichis, A. Panayides, and A. Pitsillides, "m-Health e-Emergency Systems: Current Status and Future Directions [Wireless corner]," *Antennas and Propagation Magazine, IEEE* , vol.49, no.1, pp.216-231, Feb. 2007.

[2] R.H. Istepanian, S. Laxminarayan, and C.S. Pattichis, Eds, *M-Health: Emerging Mobile Health Systems*. New York: Springer, 2006, ch. 3.

[3] C.N. Doukas, I. Maglogiannis, T. Pliakas, "Advanced medical video services through context-aware medical networks," *Conf Proc IEEE Eng Med Biol Soc*.2007:3074-7.

[4] A. Alesanco, J. Garcla, S. Olmos, and R. S. H. Istepanian, "Resilient ECG Wavelet Coding for Wireless Real-Time Telecardiology Applications," in *M-Health: Emerging Mobile Health Systems*, R. S. H. Istepanian, S. Laxminarayan, and C. S. Pattichis, Eds., New York, Springer 2006, pp. 293-312.

[5] Joint Video Team of ITU-T and ISO/IEC JTC 1, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, March 2003.

[6] www.umtsworld.com

[7] J. G. Andrews, A. Ghosh, and R. Muhamed, "*Fundamentals of WiMAX – Understanding Broadband Wireless Networking*," Prentice Hall PTR, Upper Saddle River, NJ, USA, 2007.

[8] H. Schulzrinne, A. Rao, and R. Lanphier, "Real-Time Session Protocol (RTSP)", RFC 2326, April 1998.

[9] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: Session Initiation Protocol. RFC 2543, March 1999.

[10] RFC 1889, RTP: A Transport Protocol for Real-Time Applications. Jan 1996.

[11] H.264/AVC JM 13.2 Reference Software, Available: http://iphome.hhi.de/suehring/tml/.

[12] M. Hennerici, D. Neuerburg-Heusler, Eds, *Vascular Diagnosis With Ultrasound*. Stuttgart, Germany: Thieme, 1998.

[13] Z.G. Li, F. Pan, K.P. Lim, G.N. Feng, X. Lin and S. Rahardaj, "Adaptive basic unit layer rate control for JVT," JVT-G012, 7th meeting, Pattaya II, Thailand, 7-14, March, 2003.