12

Towards Diagnostically Robust Medical Ultrasound Video Streaming using H.264

A. Panayides[1], M.S. Pattichis[2], C. S. Pattichis[1], C. P. Loizou[3], M. Pantziaris[4], and A. Pitsillides[1]
[1]*Department of Computer Science, University of Cyprus, Nicosia, Cyprus*
[2]*Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico, USA*
[3]*Department of Computer Science, Intercollege, Limassol, Cyprus*
[4]*The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus*

1. Introduction

M-Health systems and services have seen significant growth over the past decade. Increasingly available bitrate through new wireless technologies linked with compression advances provided by current state-of-the-art H.264 advanced video coding standard, have brought rapid growth to the development of mobile health (m-Health) healthcare systems and services. Advances in nanotechnologies, compact biosensors, wearable devices and clothing, pervasive and ubiquitous computing systems have broadened the applicability areas of such systems and services. A recent overview of the current status, highlighting future directions, while also covering incorporated wireless transmission technologies, can be found in (Kyriacou et al., 2007; Istepanian et al., 2006).

Despite this momentum towards m-Health systems and especially e-Emergency systems, wireless channels remain error prone, while the absence of objective quality metrics (Wang & Bovik, 2009) limits the ability of providing medical video of adequate diagnostic quality at a required bitrate. Recently, new image and video quality metrics have been proposed, for evaluating the original and transmitted (or despeckled) (Loizou & Pattichis, 2008) video over a communication path, but these have not yet been applied to wireless transmission. The development of effective medical video streaming systems requires the implementation of error-resilient methods that are also diagnostically relevant. Although a plethora of studies has been published in video transmission, very few have focused on the difficulties associated with the wireless transmission of medical ultrasound video. In (Doukas & Maglogiannis, 2008), scalable coding is used for adaptive transmission of medical images and video snapshots over simulated wireless networks, while in (Tsapatsoulis et al., 2007), a saliency-based visual attention ROI for low bit-rate medical video transmission is proposed. Mathematically lossless (M-lossless) coding of the ROI after video denoising using motion compensated temporal filtering (MCTF) is practiced in (Rao & Jayant, 2005). A new rate

control algorithm which takes into account a new motion complexity measure together with perceptual bit allocation is proposed in (Hongtao et al., 2005).
Here, we investigate different encoding schemes and packet loss rates in video transmission and provide recommendations regarding efficiency and the trade-off between bitrate and diagnostic quality. We propose the use of different frame type ordering schemes and provide methods for quantifying video quality for different packet loss rates. We also provide lower bounds for acceptable video quality and discuss how these bounds can be achieved by different frame type encoding schemes. A preliminary study appears in (Panayides et al., 2008).
Then, we are interested in exploiting new error resilience techniques and how they relate to diagnostic error resilient encoding. Loss tolerance is subject to the amount of clinical data recovered and whether this amount is suitable for providing acceptable diagnosis. Failure to do so may result in imprecise diagnosis. Consequently, the effort is directed towards maximizing error protection and recovery while at the same time conforming to channel limitations (bandwidth) and (varying) conditions (introducing delay, jitter, packet loss, and data corruption). We propose the development of effective wireless video transmission systems by extending the current state-of-the-art H.264/AVC standard to provide for video encoding that is driven by its intended diagnostic use.
Medical experts evaluating carotid ultrasound video are mainly interested in identifying possible stenosis of the carotid artery. Having diagnosed a stenosis, they aim at extracting atherosclerotic plaque (causing the stenosis) features, tracking of which in time can aid in the prediction of the severity of this abnormality. Intima media thickness (IMT) of the near and far wall also aid in this direction. The remaining regions of the video carry little, yet some diagnostic significance.
Bearing in mind the aforementioned, segmentation algorithms identifying these region(s) of diagnostic interest (ROIs) were developed by our group. The key concept is that once diagnostic ROIs have been defined, the remaining part of the video can be effectively compressed without significantly altering diagnosis. We investigate the use of a spatially-varying encoding scheme, where quantization levels are spatially varying as a function of the diagnostic significance of the video. To implement this approach, we use flexible macroblock ordering (FMO). Having derived the diagnostic ROIs of carotid ultrasound medical video, we use them as input for FMO type 2 slice encoding. We extend the FMO concept by enabling variable quality slice encoding (i.e. different quantization parameter for each ROI), tightly coupled by each region's diagnostic importance, targeting high diagnostic quality at a reduced bitrate. Subjective (clinical evaluation) as well as objective (PSNR) quality assessment shows considerable bitrate reductions while at the same not compromising diagnostic quality.
The rest of the chapter is organized as follows. Section 2 introduces the fundamental concepts of video streaming architecture and protocols. Section 3 briefly highlights H.264, describes frame types and encoding modes, as well as Flexible Macrblock Ordering. Section 4 illustrates the methodology while section 5 presents an analysis of the conducted experiments. Finally Section 6 provides some concluding remarks.

2. Video Streaming Architecture and Protocols

2.1 Architecture

Video streaming is the delivery of a video sequence over a network in real time, where video decoding starts before the entire video has been transmitted. That is, the client plays the incoming video stream in real time as the data is received. However, we should distinguish between live and on demand video streaming. Essentially, we have two different delivery methods: live video streaming where the live source (video) capture and encoding is done in real time, and on-demand video streaming, which tackles archived pre-encoded video.

A typical video streaming architecture is illustrated in Figure 1. The raw video is captured by a camera or ultrasound device, before it is fed to a source encoding software (i.e. H.264/AVC encoder). The resulting stream is then forwarded to the storage device of a server (encoding software and server can reside on the same computer, for example a laptop in an ambulance). The server is responsible for channel encoding and delivery of the stream to the client(s) through the network. The reverse procedure is followed at the client's side. After channel decoding the stream enters a playout buffer before it's forwarded to the decoder for source decoding and rendering of the transmitted video.

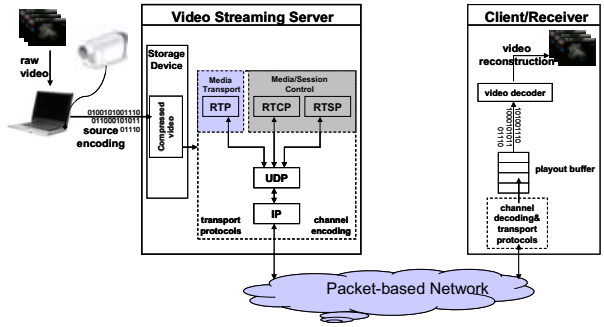


Fig. 1. A typical video streaming architecture.

2.2 Protocols

Unique requirements associated with video streaming impose strict time-delay constraints on video transmission. Video streaming protocols can be classified as follows: *session control* protocols, *transport* protocols and *network* protocols.

Session control protocols such as the Real Time Streaming Protocol (RTSP) (Schulzrinne et al., 1998), or alternative Session Initiation Protocol (SIP) (Hanfley et al., 1999) are responsible for session initialization between client and server.

Transport protocols are further distinguished in upper and lower layer, Real-Time Transport Protocol (RTP), and UDP/TCP respectively. Transmission control protocol (TCP) (Postel, 1981) uses retransmission and traffic monitoring to secure packet delivery to destination. This property constituted TCP highly efficient for HTTP (Fielding, 1999) applications. However, when streaming video, retransmission time may result in alternations of temporal relations between audio and video and is in most cases unacceptable. Given the fact that a limited number of packet losses are tolerable in video streaming and error resilience techniques are employed both at encoder and decoder, TCP's

no loss tolerance simply introduces additional jitter and skew. The user datagram protocol (UDP) (Postel, 1980) on the other hand, does not provide any error handling or congestion control mechanisms, allowing therefore packets to drop out. Given the aforementioned, UDP is primarily established as the lower layer transport protocol.
The design of a new internet protocol that would enhance the existing protocols while being suitable for real time data delivery was more than essential. RTP (Schulzrinne, 1996) provides end-to-end delivery services for data with real time characteristics such as interactive video and audio. Despite being able to provide real time data delivery, RTP itself does not contain any mechanisms to ensure on time delivery. In the contrary it relies on UDP, TCP for doing so. It does provide however the appropriate functionality for carrying real time content such as time-stamping and control mechanisms that enable synchronization of different streams with timing properties. 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 distinguishes data delivery and control mechanisms and consists of basically two parts: the RTP part which carries the real time data and the Real Time Control Protocol (RTCP) part which is responsible for Quality of Service (QoS) monitoring and extracting information regarding the participants in an RTP session. This information can be later used to improve QoS as it can be supplied as feedback for the encoder to adapt to varying network conditions. As we have already mentioned, RTP packets are usually transferred over UDP.
The resulting packets are encapsulated using the internet protocol (IP), responsible for delivering a packet from source to destination through the network, hence RTP/UDP/IP headers.

3. H.264/AVC and related Error Resilience techniques

H.264/AVC is the current state of the art video coding standard (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), and was jointly developed by the ISO/IEC MPEG and ITU-T VCEG experts, who formed the Joint Video Team (JVT). H.264 met the growing demand of multimedia and video services by providing enhanced compression efficiency significantly outperforming all prior standards (MPEG-1, 2 and 4, and H.262x). All features used together in a fashionably manner can provide for bitrate reductions of up to 50% for equivalent perceptual quality compared to its predecessors (Wiegand et al., 2003). Its design enables transportation over heterogeneous networks to be carried out in a friendly-manner. To attain the abovementioned, H.264 defines a video coding layer (VCL) and a network adaptation layer (NAL). VCL, as its name suggests, is responsible for video coding and is a unit already known from prior standards, maintaining its block-oriented coding functionality. Its enrichment and refinement resulted in the provided compression efficiency. On the other hand, NAL is novel concept aiming at a network-friendly adaptation of VCL content to candidate heterogeneous networks (or storage devices). NAL functionality is a substantial improvement constituting H.264 coding and transmission network-independent. As always, the scope of the standard is centered on the decoder. That is, only the decoder is standardized, allowing great flexibility to the encoder. H.264 offers a plethora of error resilience techniques with a wide range of applications. A thorough overview of the standard, error resilience features and discussion exploiting H.264 in the context of IP based networks can be found in (Wiegand et al., 2003; Wenger, 2003).

3.1 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, originates from the picture itself where 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 the utilization of forthcoming ones.

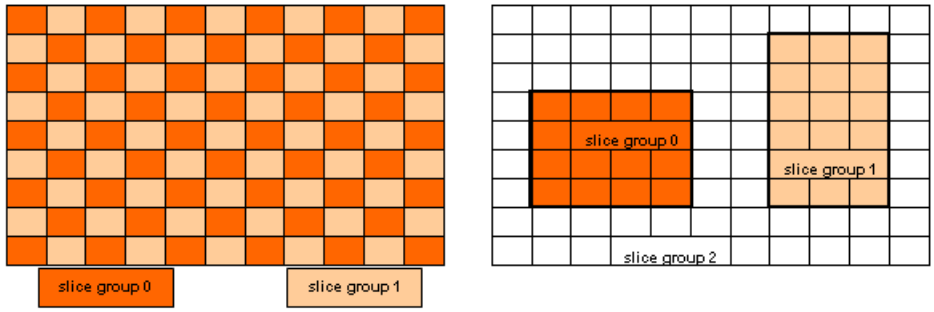
The extensive use of predictive coding (P-frames, B-frames) or not (I-frames), is application specific. Depending on time and quality constraints imposed, one mode may be preferred over the other and the other way round. Recent video coding standards have largely adopted predictive coding and it is considered one of the key components in the video streaming success. Intra coding on the other hand is mostly employed as en error resilience feature for periodic updates (i.e. inserting an I-frame every Group of Picture-GOP). The predictive coding question which arises next is the relationship (ratio) between P-frames and B-frames used, and is a matter of investigation in this particular study. P-frames employ unidirectional prediction as mentioned above whereas B-frames use 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.

3.2 Flexible Macroblock Ordering

An innovative error resilient feature introduced by H.264/AVC is flexible macroblock ordering (Wenger, 2002; Wenger & Horowitz, 2002). FMO is essentially a slice structuring approach, where a frame is partitioned into independently transmitted and decoded slices. Each frame may be partitioned in up to eight different slices and a frame may still be decoded even if not all slices are present at the decoder. In this manner and in conjunction with proper utilization of the spatial relationships between error free slices and macroblocks (MBs) therein, concealment of errors becomes much more efficient. Seven different types of FMO are defined (i.e. patterns for MB to slice allocation). A macroblock allocation map (MBAmap) is used to keep track of macroblocks assigned to slices.

Figure 2 depicts FMO utilization of FMO type 1 (Figure 2a)) and FMO type 2 (Figure 2b)). FMO type 1 aims to evenly distribute errors throughout a frame in such a way that for every lost slice, the neighboring MBs of the lost slice's MBs residing in a different slice may be used for concealment, providing for a very efficient error recovery design. FMO type 2 is

designed for defining rectangular slices as foreground(s) and leftover. A well suited scheme for region of interest coding and transmission. Slices may overlap but a MB may only belong to one slice, the first it was assigned. In the event that a packet carrying a whole slice is dropped, H.264/AVC allows the transmission of redundant slices (RS) (a redundant slice being a slice describing the same MBs in a bitstream). RS can be coded in a different manner with respect to the primary slices (i.e. different coding parameters) and are utilized in the absence of a clear primary slice. A thorough overview of FMO, performance evaluation and overhead analysis, can be found in (Wenger, 2003; Lambert et al., 2006).



a) b)

Fig. 2. Flexible Macroblock Ordering. a) Scattered Slices and b) Foreground(s) (ROIs) and leftover.

4. Methodology

4.1 Material and Video Acquisition

Cardiovascular disease accounts for about half of the deaths in the western world, with stroke being the third leading cause of death and disability worldwide. The predictive ability to identify which patients will have a stroke is poor, where the current practice of assessing the risk of stroke relies on measuring the thickness of the intima media (Intima-Media-Thickness, IMT) of the carotid artery wall or the artery lumen stenosis by identifying the plaque borders in the carotid artery. However, the IMT and the degree of stenosis do not provide adequate information of an individual stroke risk, and treatment of asymptomatic patients remains controversial. Thus the need exists for the development of new tools and techniques for the assessment of the risk of stroke.

Video analysis of the common carotid artery (CCA) can be used to analyze plaque motion patterns. Plaque motion analysis patterns can then be used to differentiate between asymptomatic and symptomatic atherosclerotic plaques. Diagnostically, the clinicians rely on visual evaluation of the CCA video motion to differentiate between asymptomatic and symptomatic plaques.

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.

CCA videos from B-mode longitudinal ultrasound segments were digitally recorded by the Philips ATL HDI-5000 ultrasound scanner (Advanced Technology Laboratories, Seattle, USA) with 256 gray levels, 768x576 resolution, having frame rate of 100 frames per second. For detailed technical characteristics of the ultrasound scanner (multi element ultrasound scan head, operating frequency, acoustic aperture, and transmission focal range), we refer to (Loizou et al., 2007).

4.2 Segmentation Procedure

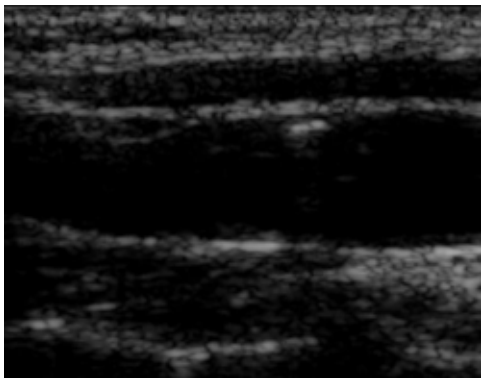
For the purposes of this study, we use snake segmentation of the plaque region and use it to define a region of interest over the video (see section 4.3.2, Fig. 4). We begin with the basic definitions.

We delineate a segmentation region using a snake contour. A snake contour is represented parametrically by $v(s) = [x(s), y(s)]$, where $(x, y) \in \Re^2$ denotes the spatial coordinates of an image, and $s \in [0,1]$. The snake adapts itself by a dynamic process that minimizes an energy function defined as in (Williams & Shah, 1992):

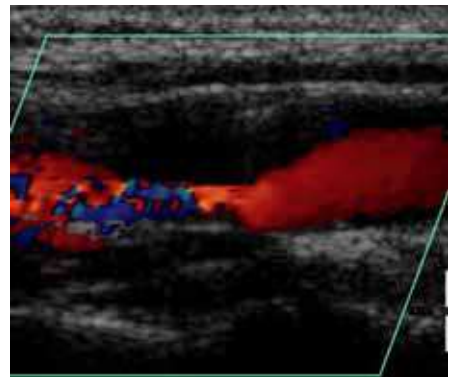
$$E_{snake}(v(s)) = E_{\text{int}}(v(s)) + E_{image}(v(s)) + E_{external}(v(s)) = \int_s (\alpha(s)E_{cont} + \beta(s)E_{curv} + \gamma(s)E_{image} + E_{external})ds. \quad (1)$$

At each iteration step, the energy function in (1), is evaluated for the current point in $v(s)$, and for the points in an mxn neighborhood along the arc length s, of the contour. Subsequently the point on $v(s)$ is moved to the new position in the neighborhood that gives the minimum energy. The term $E_{\text{int}}(v)$ in (1) denotes the internal energy derived from the physical characteristics of the snake and is given by the continuity $E_{cont}(v)$, and the curvature term $E_{curv}(v)$. This term controls the natural behaviour of the snake. The internal energy contains a first-order derivative controlled by $\alpha(s)$, which discourages stretching and makes the model behave like an elastic string by introducing tension and a second order term controlled by $\beta(s)$, which discourages bending and makes the model behave like a rigid rod by producing stiffness. The weighting parameters $\alpha(s)$ and $\beta(s)$ can be used to control the strength of the model's tension and stiffness, respectively. Altering the parameters α, β, and γ, affect the convergence of the snake. The second term in (1) E_{image}, represents the image energy due to some relevant features such as the gradient of edges, lines, regions and texture (Williams & Shah, 1992). It attracts the snake to low-level features such as brightness and edge data. Finally the term $E_{external}$, is the external energy of the snake, which is defined by the user and is optional. In our study we used a modification of the greedy algorithm as presented in (Williams & Shah, 1992).

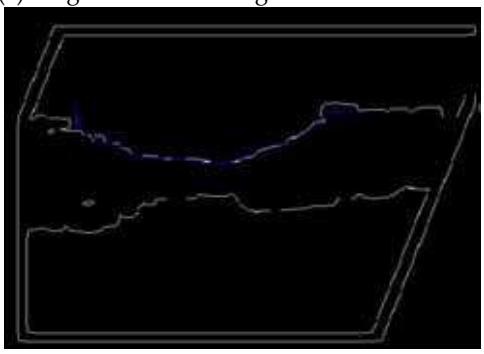
Figure 3 a) shows the first frame from an ultrasound video of the CCA, whereas Fig. 3b) shows the first frame of the blood flow video. After cross-correlating the first frame of the video with the first frame of the blood flow, the initial blood flow contour for the first video frame is extracted (see Fig, 3c)). The user then selects an area of interest defining a contour segment, which will be used as an initial contour for the snake. The snake then deforms and converges as shown in Fig. 3d). Figure 3e) shows the manual segmentation results made by an expert radiologist. Additionally, in order to help the snake to converge better the lsmv (local statistics filter based on mean and variance of the local neighborhood) despeckle filter (Loizou et al., 2005) was applied on the original video frame in Fig. 3a.



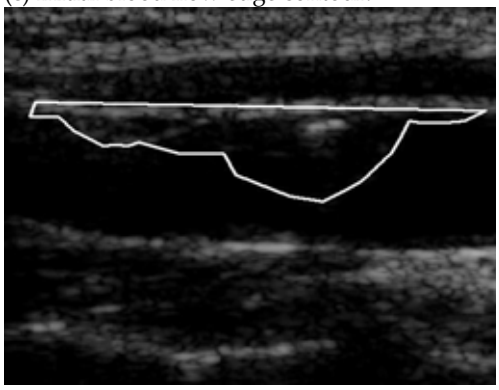
(a) Original B-mode image.



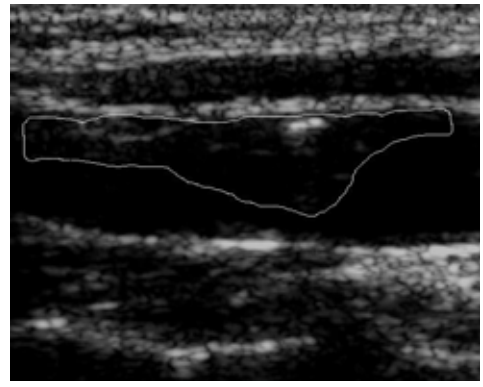
(b) Blood flow image.



(c) Initial blood flow edge contour.



(d) Williams & Shah snakes segmentation results.



(e) Manual segmentation results.

Fig. 3. Plaque initialization using the blood flow image: (a) Original ultrasound B-mode image (first video frame) of a carotid artery with plaque at the far wall, (b) first frame of the blood flow video in a), (c) initial blood flow edge contour with the area for the initial contour selected by the user, (d) Williams & Shah snakes segmentation of plaque, and (e) manual segmentation of plaque.

The automated snakes segmentation system used for the segmentation of the CCA plaque in each video frame, was proposed and evaluated on ultrasound images of the CCA in (Loizou et al., 2007), and is based on the Williams & Shah snake as described above. Initially the plaque on the first video frame was segmented and then the segmentation of the first frame was used as an initialization for the next frame. This procedure was repeated until all video frames were segmented. The snake contour iterations in each frame varied from 10 to 22.

4.3 Encoding/ Decoding Procedure

4.3.1 Frame Encoding

Using the JM 15.1 Reference Software (http://iphome.hhi.de/suehring/tml/), we evaluated three different encoding schemes, namely IPPP, IBPBP and IBBPBBP. A series of four videos encoded at QCIF and CIF resolutions were used. The default JM rate control algorithm was applied to explore the trade-off between video quality and bitrate, performing frame level adaptation. Besides the target bitrate, initial quantization parameter (QP) is another fundamental parameter which impacts on rate control algorithms performance. The first frame is encoded using the provided QP and given the remaining frames and available bitrate, QPs of following frames are adjusted accordingly to try to match the target bitrate. The initial QP was selected according to the following formula (Li et al., 2003):

$$QP = \begin{cases} 40 & bpp \le 0.15 \\ 30 & 0.15 < bpp \le 0.45 \\ 20 & 0.45 < bpp \le 0.9 \\ 10 & bpp > 0.9 \end{cases} \tag{2}$$

where

$$bpp = \frac{R}{f \times N_{pixels}}. \tag{3}$$

Here, *bpp* denotes the target bits per pixel, *Npixels* is the number of pixels in the N-th frame, *R* is the target bitrate and *f* the frame rate of the encoded video sequence. Initial quantization parameters for QCIF and CIF resolutions are summarized in Table 1.

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 modified to provide better random simulation performance by adding an implementation of the random number generator described in (Park & Miller, 1998). The simulator was also enhanced with a number of loss distributions. The uniform distribution was used throughout the experiments and all results were obtained by averaging 10 consecutive runs. The main encoding parameters are summarized in Table 2. Intra MB line update allows a fully intra coded frame every 11 and 18 frames for QCIF and CIF resolutions respectively. Simple frame copy error concealment method is applied at the decoder to reconstruct corrupted packets. Here, the use of this simple error-concealment scheme is allowed due to the fact that stenosis characteristics are only slightly affected by frame copying.

QCIF	Initial QP	CIF	Initial QP
32000	40	96000	40
64000	40	256000	40
128000	30	512000	30
256000	30	800000	30
400000	20	1200000	20
512000	20	1500000	20

Table 1. Target BitRate and Initial Quantization Parameters.

Parameters	Value	Parameters	Value
Profile	Main	MbLineIntraUpdate	1
FramesToBeEncoded	100	NumberBFrames	0/1/2
FrameRate	25	SymbolMode	CABAC
IntraPeriod	0	OutFileMode	RTP
SearchRange	32	RateControlEnable	1
NumberOfReferenceFrames	5	BasicUnit	99/396

Table 2. Encoding Parameters.

4.3.2 Variable QP Encoding

A modified version of the JM 15.1 Reference Software was used to support FMO type 2 variable quality slice encoding. FMO type 2 was selected amongst other FMO types since it's specifically designed for defining foreground(s) (ROIs) with leftover. By defining upper left and lower right corner points (on a MacroBlock basis) we select the rectangular diagnostic ROIs as illustrated in Figure 4. These bounds are defined at the beginning of each encoded sequence and are directly derived from the segmentation procedure. Taking plaque movement into account we select a slightly broader area as ROI, thus avoiding redefinition of the rectangular area further in the sequence (incorporating additional bits for picture parameter sets). Following a similar concept with MB Allocation Map (MBAmap), where each macroblock is allocated a slice id according to the slice group it belongs to, we define a QP Allocation Map (QPAmap), which stores the QP of each macroblock (see Figure 4). The QP of each ROI slice is parsed via the same configuration file used to define the boundaries of the rectangular ROIs. Each slice's QP is carried within its slice header during transmission, hence no additional bits are transmitted through picture or sequence parameter sets and the resulting bitstream is H.264/AVC compliant. Employing these minor adjustments achieves variable quality FMO slice encoding.

A series of four videos encoded at QCIF and CIF resolutions for QPs listed in Table 3 were used to evaluate our proposed approach. The videos were encoded using:

1) Constant QP throughout the sequence, which is the default encoding scheme. Here, we are interested in showing that despite achieving the best overall quality video, the required bitrate is a preventing factor, while at the same time not achieving the best diagnostic performance in the presence of errors.

2) FMO type 2 with variable QP according to the ROIs diagnostic importance (following a low to medium, medium and medium to high QP allocation pattern for non-important regions, IMT (and ECG when available) and atherosclerotic plaque slices respectively, see Table 3). The QP of the constant QP encoding and slice 0 (describing the plaque) are equal so as to be able to deduct conclusions regarding diagnostic quality. Our aim is to depict that the best diagnostic performance is attained at a significantly reduced bitrate.
3) Rate control encoding with target bit rate being the output bitrate of 2). In this manner we are able to compare the diagnostic performance of the latter two schemes for the same bitrate.

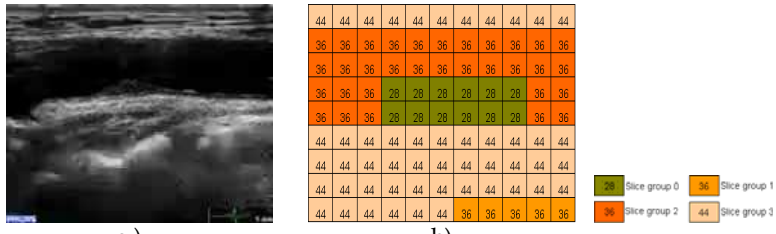
Packet loss simulation and decoding procedure are similar to the ones described in section 4.3.1. Encoding parameters are summarized in Table 4.

Constant QP Encoding	Variable QP FMO slice 3/ 2(1)/ 0	Rate Control
32	48/40/32	Target Bit Rate is the output Bit Rate of Variable QP FMO
28	44/36/28	
24	40/32/24	
20	36/28/20	
16	32/24/16	

Table 3. Quantization Parameters and Target BitRate.

Parameters	Value	Parameters	Value
Profile	Baseline	MbLineIntraUpdate	0/1*
FramesToBeEncoded	100	NumberBFrames	0
FrameRate	25	SymbolMode	UVLC
IntraPeriod	16/0*	OutFileMode	RTP
SearchRange	32	RateControlEnable	0/1*
NumberOfReferenceFrames	1	BasicUnit	99/396

Table 4. Encoding Parameters.
*IntraPeriod is enabled for Constant QP and Variable QP FMO encodings and disabled for RateControl encoding. The opposite stands for MbLineIntraUpdate and RateControlEnable.



a) b)

Fig. 4. a) Frame 1 of compressed QCIF carotid ultrasound video using variable QP FMO. b) The corresponding Quantization Parameter Allocation map (QPAmap). Slice groups: 0: atherosclerotic plaque, 1: ECG, 2: Upper and lower intima media complex, including the lumen diameter, 3: other components.

5. Experimental Results

5.1 Frame Encoding

5.1.1 Technical Evaluation

Figure 5 depicts the trade-off between quality and bit rate for one of the investigated videos in QCIF and CIF resolutions. Naturally, the more bits that are allocated for source encoding using rate control, the higher PSNR quality is attained. IBPBP and IBBPBBP coding structures behave similarly, while IPPP coding structure achieves slightly lower PSNR values. Typically, bidirectional prediction requires fewer bits during encoding than single-directional prediction for the same quantization parameters, however, since we are using rate control, this is translated into increased quality. On the other hand, single-directional prediction is marginally quicker in terms of encoding time due to the increased motion estimation time required for bidirectional encoding.

Figures 6 and 7 demonstrate the performance of the three tested encoding schemes under losses of 2%, 5%, 8% and 10% of transmitted RTP packets for the same video, for QCIF and CIF resolutions respectively. For QCIF resolution, IBBPBBP achieves better PSNR output than IBPBP and IPPP, especially up to 5% loss rate. For 8% and 10%, IPPP coding structure attains higher PSNR ratings in some cases. For CIF resolution, bidirectional prediction (IBPBP and IBBPBBP) achieves slightly better results up to 5% loss rates, but then it is outperformed by single-directional (although for 8% loss rate, ratings are comparable with IBBPBBP). In general, in low-noise environments bidirectional prediction gives the best performance. However, as the noise level increases, the use of single-directional prediction provides for better error recovery and better results.

Another important aspect which was observed by examining the results obtained by averaging 10 consecutive runs for each scheme is 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.

5.1.2 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 PSNR ratings higher than 30.5 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 regarding the presence of an atherosclerotic plaque and the degree of stenosis. This information was also available to videos attaining lower PSNR ratings in some cases. This was made possible when the medical expert freezed a relatively clean frame to use for diagnosis. Moreover, for bitrates of 128kbps for QCIF resolution and 512kbps for CIF resolution, the medical expert could almost identify as much diagnostic information in the compressed video as in the original video sequence. It is worth noting here that for the abovementioned bitrates, the initial QP switches from 40 to 30 (according to (2) and (3)) for the target bitrates chosen for this series of experiments.

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 videos used in this particular study were very clear cases.

5.2 Variable QP Encoding

5.2.1 Technical Evaluation

Diagnostic quality for carotid ultrasound video evaluation can be defined as the PSNR over the (atherosclerotic) plaque, being the primary focus point of the clinical evaluation. In Figures 8 and 9 we provide rate-distortion curves of both entire video (Figures 8a) and 9a)) and atherosclerotic plaque (region of diagnostic importance, Figures 8b) and 9b)) and the impact that can have on bitrate.

As expected, the incorporated R-D curves differentiate only with respect to variable QP FMO encoding, as the latter is the only scheme to employ variable QP encoding. In Figures 8a) and 9a) constant QP encoding achieves the best overall performance with rate control encoding following, since both schemes employ equal QP throughout a frame. When it comes to diagnostic performance however (Figures 8b) and 9b)), variable QP FMO attains similar PSNR ratings with constant QP, the key observation being the drastically lower sequence bitrate it involves. Compared to rate control incorporating the same bitrate, variable QP FMO achieves higher PSNR ratings.

Figure 10 demonstrates the performance of the three tested encoding schemes when extracting the atherosclerotic plaque (diagnostic ROI) of the decoded video under losses of 5% of transmitted RTP packets. We have significant bandwidth requirement reductions without sacrificing diagnostic quality. Furthermore, FMO provides increased error resilience by coding diagnostically relevant regions independent of the rest. Figures 8-10 show the results of QCIF and CIF resolution video with ECG lead while Table 5 records indicative db gain and bitrate reductions for a video with no ECG lead. Results are video specific but the trend is the same for all investigated videos.

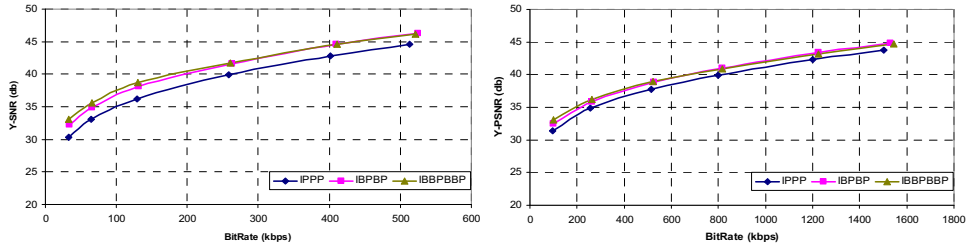
5.2.2 Clinical Evaluation

The evaluation recorded that all CIF resolution variable QP FMO encoding videos were well within the desired diagnostic range. The medical expert could identify the presence of a plaque (stenosis), come to conclusions regarding the degree of stenosis and quality of the plaque, and classify the plaque as echogenic or echolucent. Naturally, as the quality rose, diagnosis became easier and more precise. As the medical expert designated, all video qualities qualified for urgent clinical practice. Some distortion in ECG for 48/40/32 QPs (40 being the QP of the ECG) was observed, not present for lower QPs or affecting diagnostic quality.

Similar conclusions were drawn for constant QP encoding. Rate control encoding also performed quite well, however, for videos attaining PSNR ratings higher than 30.5db, corresponding to output bitrate of quantization parameters 40/32/24 of variable QP FMO and lower.

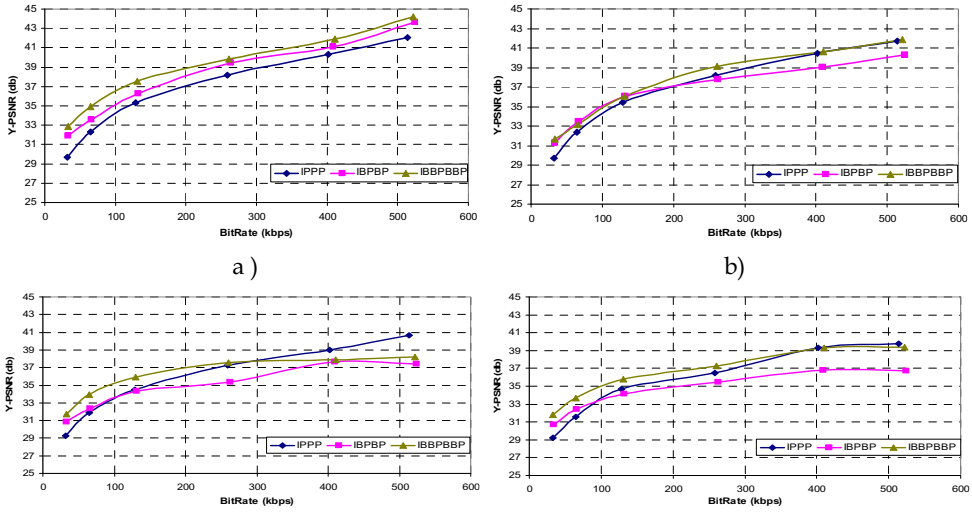
For QCIF resolution, the medical expert underlined that plaque classification could not be derived for all videos, irrespective of video quality, rather its dimensions. This of course is video specific but for the same video in CIF resolution classification could be obtained. Furthermore, in some cases, and under loss rates of 5%, 48/40/32 variable QP FMO corresponding to PSNR slightly lower than 31db, required extra attention to reach

diagnosis. Again, all video qualities qualified for urgent clinical practice, however QPs of 44/36/28 is recommended. The same allegations stand for constant QP encoding, whereas for rate control, similar to CIF resolution, videos attaining PSNR higher than 30.5 db.



a) b)

Fig. 5. Rate-distortion curves for tested frame encoding schemes. a) QCIF and b) CIF.



a) b)

c) d)

Fig. 6. Rate-distortion curves for tested frame encoding schemes, QCIF resolution. a) 2%, b) 5%, c) 8% and d) 10% loss rates. IBBPBBP encoding scheme attains higher PSNR ratings in most cases, especially in low-noise (up to 5%) scenarios.

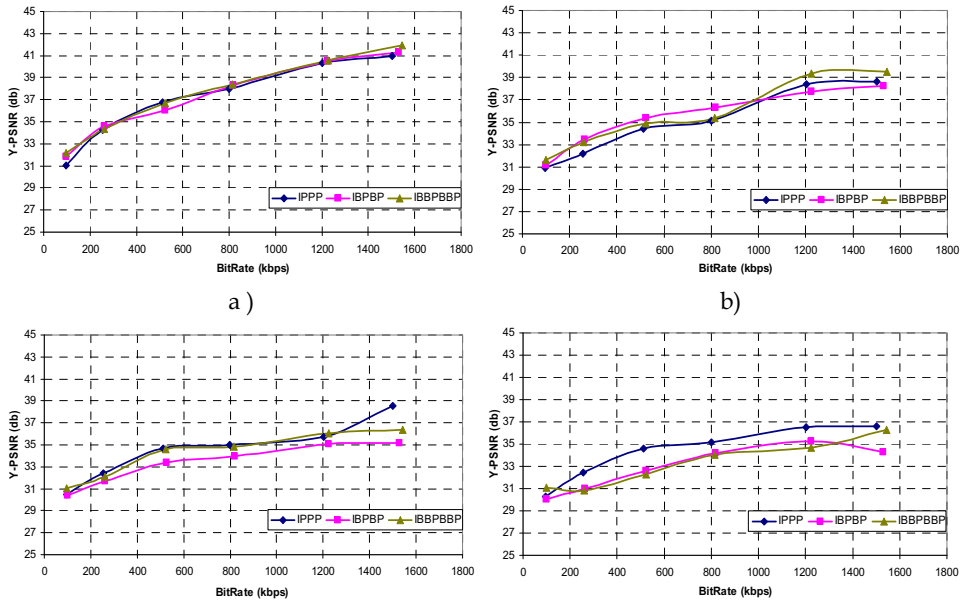


Fig. 7. Rate-distortion curves for tested frame encoding schemes, CIF resolution. a) 2%, b) 5%, c) 8% and d) 10% loss rates. Bi-directional prediction (IBPBP and IBBPBBP) achieves better results up to 5% loss rates (low-noise), whereas as the noise level increases, single – directional (IPPP) provides for better error recovery.

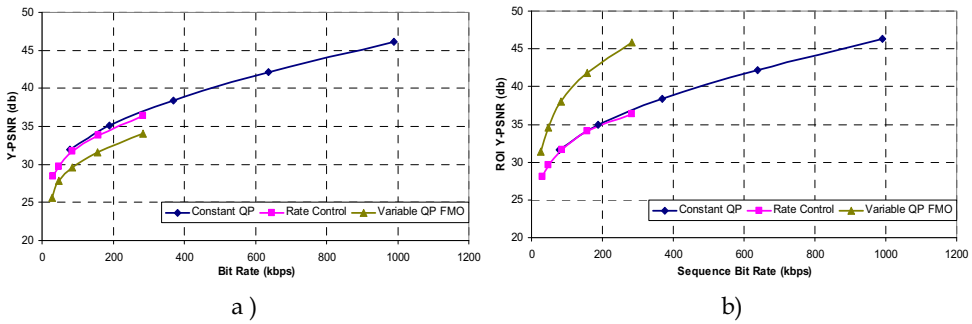
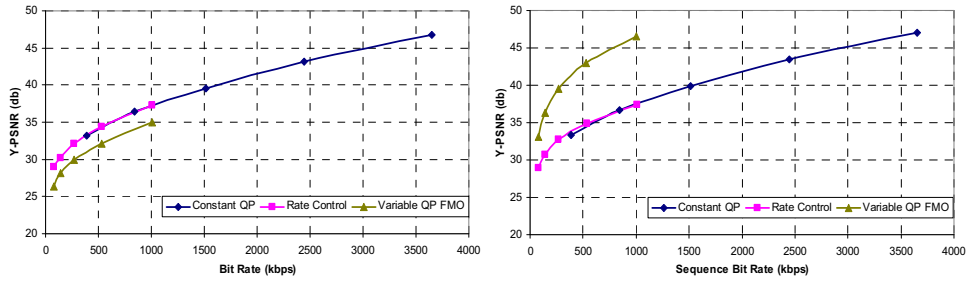
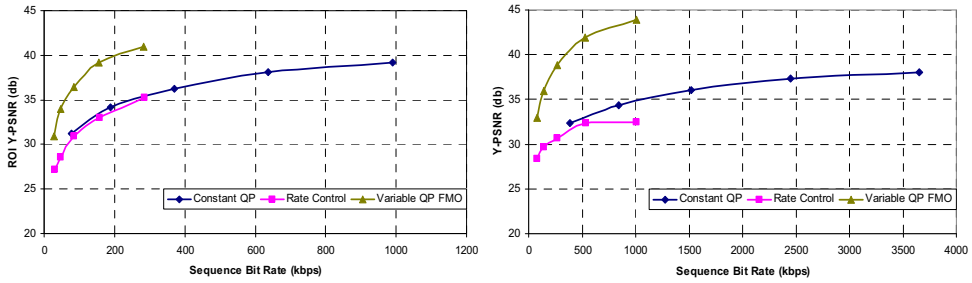


Fig. 8. Rate-distortion curves for a) entire video, QCIF resolution with ECG lead and b) atherosclerotic plaque extracted from QCIF resolution video with ECG lead (diagnostic ROI). Observe that Variable QP FMO encoding attains inferior quality for the whole video, when it comes to diagnostic quality however it outperforms rate control encoding, while it achieves similar PSNR ratings with constant QP encoding, the key observation being the drastically lower bitrate it involves.



a) b)

Fig. 9. Rate-distortion curves for a) entire video, CIF resolution video with ECG lead and b) atherosclerotic plaque extracted from CIF resolution video with ECG lead (diagnostic ROI). Observe that Variable QP FMO encoding attains inferior quality for the whole video, when it comes to diagnostic quality however it outperforms rate control encoding, while it achieves similar PSNR ratings with constant QP encoding, the key observation being the drastically lower bitrate it involves.



a) b)

Fig. 10. Rate-distortion curves for a) atherosclerotic plaque extracted from QCIF resolution video with ECG lead, 5% loss rate and b) atherosclerotic plaque extracted from CIF resolution video with ECG lead, 5% loss rate. Variable QP FMO encoding attains the best diagnostic performance. Better error recovery compared to constant QP encoding is due to the fact that FMO employs slice encoding. Bandwidth requirements reductions as to Figures 8-9.

Constant QP		Rate Control		Variable QP FMO		Constant QP vs Variable QP FMO		Rate Control vs Variable QP FMO	
PSNR	Seq. BitRate	PSNR	Seq. BitRate	PSNR	Seq. BitRate	Db Gain	BitRate Reduction	Db Gain	BitRate Reduction
33.08	235	29.19	82	33.19	82	0.11	153	4	Negligible
34.88	508	30.69	157	36.06	156	1.18	352	5.37	
36.51	960	33.01	302	38.65	301	2.14	659	5.64	
37.47	1642	33.67	562	40.77	561	3.30	1081	7.10	
38.04	2554	35.6	960	42.56	959	4.52	1595	6.96	

Table 5. Atherosclerotic plaque extracted from CIF resolution video, no ECG lead - 5% Loss Rate.

6. Conclusion and Future Work

M-Health systems and services facilitated a revolution in remote diagnosis and care. Driven by advances in networking, video compression and computer technologies, wide deployment of such systems and services is expected in the near future. Before such a scenario becomes a reality however, there are a number of issues that have to be addressed. Video streaming of medical video over error prone wireless channels is one critical issue that needs to be addressed. Remote diagnosis is very sensitive to the amount of clinical data recovered, hence the effort should be directed towards the provision of robust medical video at a required bitrate for the medical expert to provide a confident and accurate diagnosis.

H.264/AVC encompasses powerful video coding and error resilience tools, exploitation of which can significantly improve video quality. We present an evaluation of different frame types and encoding modes of H.264/AVC and how they relate to diagnostic performance. In addition, an efficient, diagnostically relevant approach is proposed for encoding and transmission of medical ultrasound video of the carotid artery. Driven by its diagnostic use, ultrasound video is segmented and encoded using flexible macroblock ordering (FMO). FMO type 2 concept is extended to support variable quality slice encoding. Diagnostic region(s) of interest are encoded in high quality whereas the remaining, non-diagnostic region, is heavily compressed. Both technical and clinical evaluation show that enhanced diagnostic performance is attained in the presence of errors while at the same time achieving significant bandwidth requirements reductions.

Future work includes the insertion of redundant slices (RS) describing diagnostically important region(s) in the resulting bitstream, maximizing medical video's error resilience under severe packet losses (Panayides et al., 2009). We will also explore the application of these technologies to other medical video modalities.

7. Acknowledgement

This work was funded via the project *Real-Time Wireless Transmission of Medical Ultrasound Video* of the Research and Technological Development 2008-2010, of the Research Promotion Foundation of Cyprus.

8. References

Doukas, C. & Maglogiannis, I. (2008). Adaptive Transmission of Medical Image and Video Using Scalable Coding and Context-Aware Wireless Medical Networks, *EURASIP Journal on Wireless Communications and Networking*, Vol. 2008, Article ID 428397, 12 pages. doi:10.1155/2008/428397.

Fielding, R.; Gettys, J.; Mogul, J.; Frystyk, H.; Masinter, L.; Leach, P. & Berners-Lee, T. (1999). Hypertext Transfer Protocol-HTTP/1.1., Internet Engineering Task Force, RFC 2616, 1999.

H.264/AVC JM 15.1 Reference Software, Available: http://iphome.hhi.de/suehring/tml/.

Handley, M.; Schulzrinne, H.; Schooler, E. & Rosenberg, J. (1999). SIP: Session Initiation Protocol, Internet Engineering Task Force, RFC 2543, Mar. 1999.

Hennerici, M. & Neuerburg-Heusler, D. (1998). *Vascular Diagnosis With Ultrasound*, Thieme, 0865776032, 9780865776036, Stutgart - New York.

Istepanian, R.H.; Laxminarayan, S. & Pattichis, C.S. (2006). *M-Health: Emerging Mobile Health Systems*, Springer, 0387265589, 9780387265582, New York.

Joint Video Team of ITU-T and ISO/IEC JTC 1. (2003). 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, Mar. 2003.

Kyriacou, E.; Pattichis, M.S.; Pattichis, C.S.; Panayides, A. & Pitsillides, A. (2007). M-Health e-Emergency Systems: Current Status and Future Directions [Wireless corner], *Antennas and Propagation Magazine, IEEE* , Vol. 49, No. 1, Feb. 2007, pp. 216-231, 1045-9243.

Lambert, P.; De Neve, W.; Dhondt, Y. & Van De Walle, R. (2006). Flexible macroblock ordering in H.264/AVC, *Journal of Visual Communication and Image Representation*, Vol. 17, No. 2, Apr. 2006, pp. 358-375, 10473203.

Li, Z.G.; Pan, F.; Lim, K.P.; Feng, G.N.; Lin X. & Rahardaj, S. (2003). Adaptive basic unit layer rate control for JVT, JVT-G012, 7th meeting, *Pattaya II, Thailand*, 7-14, Mar. 2003.

Loizou, C.P.; Pattichis, C.S.; Christodoulou, C.I.; Istepanian, R.S.H.; Pantziaris, M. & Nicolaides, A. (2005). Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery, *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, Vol. 52, No. 10, Oct. 2005, pp. 1653-1669, 0885-3010.

Loizou, C.P.; Pattichis, C.S.; Pantziaris, M. & Nicolaides, A. (2007). An integrated system for the segmentation of atherosclerotic carotid plaque, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 11, No. 5, Nov. 2007, pp. 661-667, 1089-7771.

Loizou, C.P. & Pattichis C.S. (2008). Despeckle filtering algorithms and Software for Ultrasound Imaging, *Synthesis Lectures on Algorithms and Software for Engineering*, Ed. Morgan & Claypool Publishers, 13: 9781598296204, USA.

Panayides, A.; Pattichis, M. S. & Pattichis, C. S. (2008). Wireless Medical Ultrasound Video Transmission Through Noisy Channels, *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'08)*, pp. 5326-5329, 1557-170X, Aug. 2008, Vancouver, Canada.

Panayides, A.; Pattichis, M. S.; Pattichis, C. S.; Loizou, C. P.; Pantziaris, M. and Pitsillides, A. (2009). Robust and Efficient Ultrasound Video Coding in Noisy Channels Using H.264, to be published in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'09)*, Sep. 2009, Minnesota, U.S.A.

Park S. & Miller, K. (1998). Random Number Generators: Good Ones Are Hard To Find, *Communications of the ACM*, Vol. 31, No. 10, Oct. 1988, pp. 1192 - 1201,0001-0782.

Postel, J. (1980). User Datagram Protocol, Internet Engineering Task Force, RFC 768, 1980.

Postel, J. (1981). Transmission Control Protocol, Internet Engineering Task Force, RFC 793, 1981.

Rao, S. & Jayant, N. (2005). Towards high quality region-of-interest medical video over wireless networks using lossless coding and motion compensated temporal filtering, *Proceedings of the fifth IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'05)*, pp. 618-623, 0-7803-9313-9, Dec. 2005, Athens, Greece.

Schulzrinne, H.; Casner, S.; Frederick, R. & Jacobson, V. (1996). RTP: A Transport Protocol for Real-Time Applications, Internet Engineering Task Force, RFC 1889, Jan. 1996.

Schulzrinne, H.; Rao, A. & Lanphier, R. (1998). Real-Time Session Protocol (RTSP), Internet Engineering Task Force, RFC 2326, Apr. 1998.

Tsapatsoulis N.; Loizou, C. & Pattichis, C. (2007). Region of Interest Video Coding for Low bit-rate Transmission of Carotid Ultrasound Videos over 3G Wireless Networks, *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'07),* pp. 3717-3720, 978-1-4244-0787-3, Aug. 2007, Lyon, France.

Wang Z. & C. Bovik, A. (2009) Mean squared error: love it or leave it? - A new look at signal fidelity measures, *IEEE Signal Processing Magazine*, Vol. 26, No. 1, Jan. 2009, pp. 98-117.

Wenger S. (2002). FMO: Flexible Macroblock Ordering, *ITU-T JVT-C089*, May 2002.

Wenger, S. & Horowitz, M. (2002). Flexible MB Ordering – A New Error Resilience Tool for IP-Based Video, *Proceedings of International Workshop on Digital Communications (IWDC'02)*, Sept. 2002, Capri, Italy.

Wenger, S. (2003). H.264/AVC over IP, *IEEE Transactions on Circuits and Systems for Video Technolology*, Vol. 13, No. 7, Jul. 2003, pp. 645–656, 1051-8215.

Wiegand, T.; Sullivan, G. J.; Bjøntegaard, G. & Luthra, A. (2003). Overview of the H.264/AVC video coding standard, *IEEE Transactions on Circuits and Systems for Video Technolology*, Vol. 13, No. 7, Jul. 2003, pp. 560–576, 1051-8215.

Williams, D. & Shah, M. (1992). A Fast Algorithm for Active Contour and Curvature Estimation, *GVCIP: Imag. Und.*, Vol. 55, No. 1, 1992, pp. 14-26.

Yu, H.; Lin, Z. & Pan, F. (2005). 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, Dec. 2005, pp. 2707-2716, 1549-8328.

Biomedical Engineering

Edited by Carlos Alexandre Barros de Mello

ISBN 978-953-307-013-1
Hard cover, 658 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

Biomedical Engineering can be seen as a mix of Medicine, Engineering and Science. In fact, this is a natural connection, as the most complicated engineering masterpiece is the human body. And it is exactly to help our "body machine" that Biomedical Engineering has its niche. This book brings the state-of-the-art of some of the most important current research related to Biomedical Engineering. I am very honored to be editing such a valuable book, which has contributions of a selected group of researchers describing the best of their work. Through its 36 chapters, the reader will have access to works related to ECG, image processing, sensors, artificial intelligence, and several other exciting fields.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

A. Panayides, M.S. Pattichis, C. S. Pattichis, C. P. Loizou, M. Pantziaris and A. Pitsillides (2009). Towards Diagnostically Robust Medical Ultrasound Video Streaming using H.264, Biomedical Engineering, Carlos Alexandre Barros de Mello (Ed.), ISBN: 978-953-307-013-1, InTech, Available from: http://www.intechopen.com/books/biomedical-engineering/towards-diagnostically-robust-medical-ultrasound-video-streaming-using-h-264

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821