DIAGNOSTICALLY RESILIENT ENCODING, WIRELESS TRANSMISSION, AND QUALITY ASSESSMENT OF MEDICAL VIDEO

Andreas Stavrou Panayides

University of Cyprus, 2011

A new framework for effective communication and evaluation of wireless medical video over error-prone channels is proposed. This is motivated by the need to efficiently address unique requirements associated with medical video source encoding, wireless transmission, and quality assessment. The envisioned utilization scenarios target remote diagnosis and care and emergency situations.

A unified framework is developed that: (i) provides a *diagnostically relevant* medical video encoding based on clinical criteria, (ii) enables *diagnostically resilient* medical video encoding for reliable communications over noisy wireless channels, and (iii) introduces objective and subjective criteria for *clinical video quality assessment*.

The approach is based on a spatially varying encoding scheme, where video slice quantization parameters are varied as a function of diagnostic significance. Video slices are automatically set based on a segmentation algorithm. They are then encoded using a modified version of H.264/AVC flexible macroblock ordering (FMO) technique that allows variable quality slice encoding and redundant slices (RS) for resilience over error prone communication channels.

Evaluation of the proposed scheme is performed on a representative collection of ten (10) ultrasound videos, nine of the carotid and one of the femoral arteries, for packet loss rates up to 30%. Extensive simulations incorporating three FMO encoding methods, different quantization levels and display resolutions, and different packet loss scenarios are investigated. Quality assessment is based on a new clinical rating system that provides for independent evaluations of the different parts of the video (subjective). Objective video quality assessment metrics are also employed and their correlation to the clinical quality assessment of plaque type is derived. To this end, some objective quality assessment measures computed over the plaque video slices gave very good correlations to mean opinion scores (MOS). Here, MOS were computed using two medical experts.

Experimental results show that the proposed method achieves enhanced performance in noisy environments, while achieving significant bandwidth demands reductions, providing for transmission over 3G (and beyond) wireless networks. The proposed unified framework can be modified for application to other medical video modalities. This requires the identification of diagnostic ROIs, the adoption of a new clinical diagnostic rating system, and expert validation.

DIAGNOSTICALLY RESILIENT ENCODING, WIRELESS TRANSMISSION, AND QUALITY ASSESSMENT OF MEDICAL VIDEO

Andreas Stavrou Panayides

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Cyprus

Recommended for Acceptance

by the Department of Computer Science

June, 2011

2011

APPROVAL PAGE

Doctor of Philosophy Dissertation

DIAGNOSTICALLY RESILIENT ENCODING, WIRELESS TRANSMISSION, AND QUALITY ASSESSMENT OF MEDICAL VIDEO

Presented by

Andreas Stavrou Panayides

Research Supervisor

_____

Research Supervisor's Name

Committee Member

_____

Committee Member's Name

Committee Member

_____

Committee Member's Name

Committee Member

_____

Committee Member's Name

Committee Member

_____

Committee Member's Name

University of Cyprus

June, 2011

This Ph.D. dissertation is dedicated

to the Memory of my father Stavros,

and to my mother Andriani and my sister Stephania.

ACKNOWLEDGEMENTS

Ultrasound Video ΠΕΝΕΚ/ΕΝΙΣΧ/0308/90 of the Research and Technological Developent 2008-2010.

Last, but certainly not least, I would like to thank all my friends who have generously and unconditionally supported me with their love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

15

TABLE OF ACRONYMS

3G	3rd Generation
3GPP	3rd Generation Partnership Project
4CIF	4xCIF
4G	4th Generation
ASO	Arbitrary Slice Ordering
AVC	Advanced Video Coding
CABAC	Context-based Adaptive Binary Arithmetic Coding
CAVLC	Context-based Adaptive Variable Length Coding
CCA	Common Carotid Artery
CDMA	Code Division Multiple Access
CIF	Common Intermediate Format
DL	Downlink
EAP	Extensible Authentication Protocol
ECG	Electrocardiogram
EDGE	Enhanced Data rates for GSM Evolution
FDD	Frequency Division Duplex
FMO	Flexible Macroblock Ordering
GOP	Group of Pictures
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HEVC	High Efficiency Video Coding
HM	HEVC Reference Software
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
ICA	Inline Carotid Artery
IFC	Information Fidelity Criterion
IMT	Intima Media Thickness
IMT-Advanced	International Mobile Telecommunications-Advanced
IP	Internet Protocol
IQR	Inter Quartile Range
ITU	International Telecommunication Union
JCT-VC	Joint Collaborative Team on Video coding
JM	H.264/AVC Reference Software
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
LAN	Local Area Network
LOS	Line Of Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MB	Macroblock
MBMS	Multimedia Broadcast Multicast Service
MIMO	Multiple Input Multiple Output
M-JPEG	Motion JPEG
MOS	Mean Opinion Score

MOVIE	Motion-based Video Integrity Evaluation
MPEG	Motion Pictures Experts Group
MSE	Mean Square Error
MV	Motion Vector
NAL	Network Abstraction Layer
NLOS	Non-Line of Sight
NQM	Noise Quality Measure
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PHY	Physical
PLR	Packet Loss Rates
PSNR	Peak Signal to Noise Ratio
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QoS	Quality of Service
QP	Quantization Parameter
ROI	Region of Interest
RS	Redundant Slices
RTP	Real-time Transport Protocol
RTSP	Real Time Streaming Protocol
SIP	Session Initiation Protocol
SNR	Signal to Noise Ration
SP/SI	Switching-Predictive/Switching-Intra
SSIM	Structure Similarity Index
SVC	Scalable Video Coding
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UDP	User Datagram Protocol
UE	User Equipment
UEP	Unequal Error Protection
UL	Uplink
UMTS	Universal Mobile Telecommunications System
VCL	Video Coding Layer
VIF	Visual Information Fidelity
VIFP	Pixel-based VIF
VoIP	Voice over IP
VQA	Video Quality Assessment
VQEG	Video Quality Experts Group
VQM	Video Quality Metric
VSNR	Visual Signal to Noise Ration
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Networks
WMAN	Wireless Metropolitan Area Networks
WSNR	Weighted Signal to Noise Ration

Chapter 1

Introduction

1.1 Introduction

The history of telemedicine systems [1] is tightly coupled with the continuous growth of computing technologies and systems in general. Driven by significant advances in computational power, we now have a broad spectrum of new mobile health systems and services that were previously unimagined [2], [3]. Health applications include remote diagnosis and care, home monitoring of patients with chronic diseases and the elderly, and assistive technologies. The underlying technologies range from electronic health records to implantable sensors and wearable devices.

For medical wireless video transmission systems, the two most significant components include the medical video compression technology and the wireless infrastructure that will be used for the transmission. Medical video compression needs to address some of the unique requirements associated with the intended diagnostic use. Efficient video compression systems can be build using the current state-of-the-art video coding standards such as H.264/AVC [4], to provide for both an efficient (size wise) and timely (real time) encoding. On the other hand, increasingly available bitrate through revolutionary wireless transmission channels [5], [6] realize communications previously only available to wired infrastructures [7]. Coverage is extended practically across the globe with the latest mobile cellular and satellite systems. Over the last five years, the emergence of new 3G wireless medical video
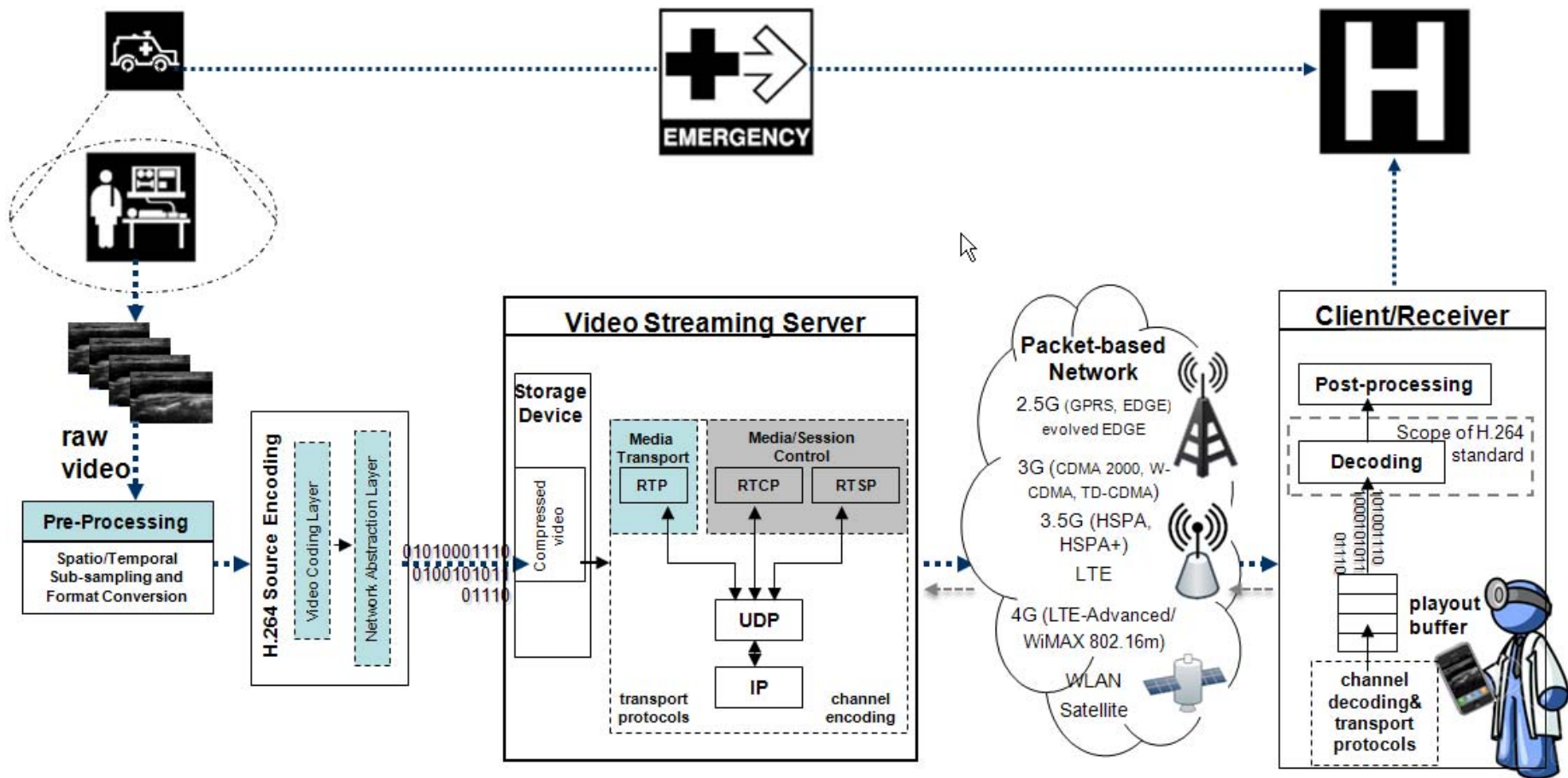
Fig. 1.1. Wireless medical video transmission for emergency telemedicine.

Equipment residing in an ambulance captures and transmits the medical video to a remote medical expert and/or hospital premises for assistance with the diagnosis and to prepare patient admission to the hospital.

transmission systems was facilitated. Significant growth is expected in this area upon wider deployment of 3.5G systems and currently developed 4G technologies.

Fig. 1.1 depicts the basic architecture of an emergency telemedicine system designed for the transmission of wireless medical video. At the incident scene (ambulance, helicopter, ship, airplane), appropriately trained paramedical staff following the established protocol provide the designated patient emergency care. Then, having stabilized the patient's health situation, utilize equipment residing in the ambulance to capture and transmit the vital biosignals, and the medical video to the hospital premises and/or remote medical expert. The reasoning here is twofold. First is the provision of remote diagnosis and care crucial for the patient's health in specialized conditions and second, for better triage and hospital admission related tasks (surgery chamber preparation, etc.). In the event of a trauma incident, the incorporated trauma video will be transmitted. To achieve this, a video camera connected to a portable computer is required. The paramedical staff is responsible for the proper video acquisition and session initialization. Accordingly, for ultrasound video (carotid, femoral, internal organs, cardiac, abdominal aortal aneurism (AAA), etc.), a portable ultrasound device is used [8]. All remaining tasks including video preprocessing, source encoding and transmission through the available wireless transmission medium (3G, HSPA, HSPA+, mobile WIMAX, LTE, emerging 4G) is carried out by a single portable computer which acts both as an encoding and streaming server via an automated procedure. At the receiver's side, the opposite procedure is followed for video reception, decoding, and post-processing.

Despite the rapid growth of telemedicine systems, wireless channels remain error prone, while the continuous bitrate and compression efficiency increase is soon met by the rising expectations on the amount of clinical data to be transmitted. In practice,
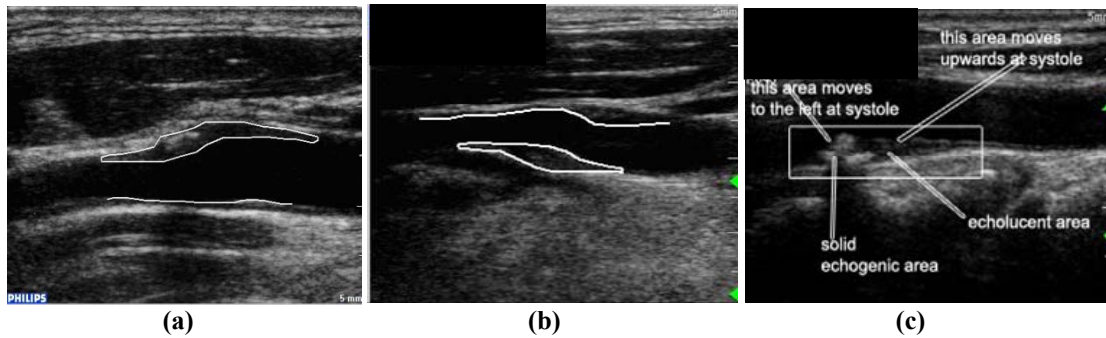
Fig. 1.2. Atherosclerotic plaque video image examples.
The plaque boundaries and nearest walls are outllined by an automatic segmentation algorithm [80].
a) Predominantly echogenic plaque. b) Predominantly echolucent plaque. c) Demonstrative plaque motion at systole and diastole with different motions for the echogenic and echolucent portions.

clinical videos are routinely compressed with a limited understanding on the effects of compression on diagnostic quality. Here, it is interesting to note that the topic of video quality assessment (VQA) is still emerging for general videos [9]. Yet, for medical videos, where crucial clinical information may be deteriorated during compression and transmission, there is relatively very little research being done. Absence of efficient objective and subjective quality assessment metrics for the evaluation of the transmitted medical video contributes to the challenges associated with the streaming of adequate diagnostic quality video at a required bitrate that can be delivered at any time and any location.

The motivation of this study is to develop a framework that: (i) provides a *diagnostically relevant* medical video encoding based on clinical criteria, (ii) enables *diagnostically resilient* medical video encoding for reliable communications over noisy wireless channels, and (iii) introduces objective and subjective criteria for *clinical video quality assessment*. The basic system is demonstrated on the wireless transmission of atherosclerotic plaque ultrasound videos. Here, the envisioned application scenario is to provide a system that allows clinicians to evaluate clinical ultrasound videos for emergency and remote diagnosis and care telemedicine applications.

21

Ultrasound video is widely used in vascular imaging to visualize the arterial lumen, plaque, and wall. Medical experts evaluating carotid artery ultrasound video are mainly interested in identifying plaque presence, the corresponding degree of stenosis, as well as the plaque type. Monitoring of the arterial characteristics like the vessel lumen diameter, the intima media thickness (IMT) of the far wall and the morphology of atherosclerotic plaque are important in order to assess the severity of atherosclerosis and evaluate its progression [10].

The first objective is to ensure that the clinical data in the transmitted video is sufficient to identify the presence of the plaque and its boundary. To assess the degree of stenosis, the boundary of the plaque, its size, as well as the distance to the nearest arterial wall needs to be visualized. In Fig. 1.2(a)-(b), frames of the segmented video plaques with the associated near and far arterial walls is presented. Furthermore, stenosis needs to be visualized throughout the cardiac cycle, over the systolic and diastolic phases, as the plaque moves (see Fig. 1.2(c)). This can be facilitated by the electrocardiogram (ECG) part of the video (see lower right in Fig. 4.4(a)). Visualization of the echolucent and echogenic plaque regions, as well as their corresponding motions throughout the cardiac cycle is of vital importance in assessing plaque stability (see Fig. 1.2(c)). The remaining part of the video carries little diagnostic information.

This study provides a unifying framework for:

- *Mapping clinical criteria to diagnostic video encoding*

Clinical criteria are first used for determining the regions of diagnostic interest (see Table 4.1). The regions are then used to specify video slices with independent coding control. A spatially-varying quality map is used for efficient video slice encoding.

- *Video encoding for mobile communications through noisy channels*

Wireless video transmission requires that decoding performance needs to be evaluated as a function of packet loss rates (PLR), available bitrates, and the mobile device's supported resolutions and frame rates. A unifying framework is proposed that provides error-resilient encoding that allows for reliable performance even at large PLR.

- *Video quality assessment based on clinical criteria*

Both objective and subjective evaluations are used for measuring the quality of the decoded video slices. To establish the validity of the approach, correlation between the medical experts' mean opinion scores (MOS) and a number of objective measurements is used.

- *Coarse to fine parameter optimization based on video quality assessment*

Here, the goal is to determine video encoding parameters that can provide acceptable video quality and the degree of error resilience for reliable medical video communications. The approach allows determining the minimum bitrates needed for transmission, as well as the maximum PLR for which diagnostic quality is preserved.

The aforementioned proposed methodology targets an encoding setting that will allow the transmission of adequate diagnostic quality video over 3G (and beyond) mobile telecommunications networks. Continuous medical expert feedback and objective video quality evaluation guide the process.

1.2 Original Aspects of the Work

The proposed unifying framework for reliable medical video delivery over wireless channels integrates novel concepts that enable enhanced diagnostic

performance over bitrate-limited and error-prone channels. These concepts are summarized below:

- *Association of clinical criteria with certain video portions based on medical expert's feedback enables diagnostically relevant encoding.* In diagnostically relevant encoding, quality levels are varied as a function of the diagnostic significance of the video. The proposed scheme provides for diagnostically lossless encoding at a *significantly reduced bitrate,* while providing for efficient assessment of the system's diagnostic performance.

- *Error-resilient encoding for consistent diagnostic performance* over unstable wireless networks and severe packet losses. Flexible macroblock ordering (FMO) and redundant slices (RS) are two new error resilience techniques defined in H.264/AVC standard that are employed following necessary modifications (to allow for diagnostically relevant encoding) and fine-tuning to recover from high PLR.

- *Coarse to fine-parameter optimization for determining minimum bandwidth requirements for diagnostically robust medical video.* Experimentation using a data sample of ten ultrasound videos with sufficient diagnostic ROI sizes range, and exhaustive investigation of quality levels, packet loss scenarios, and frame rate and display size. Overall, a total of 12600 video instances are considered.

- *Video quality assessment (VQA) based on extensive use of objective and subjective evaluations.* Subjective evaluation is based on MOS provided by two medical experts for preset clinical criteria. Objective evaluation incorporates eight different VQA algorithms. *Correlation between MOS and objective VQA measurements* for different clinical criteria is also investigated.

Different clinical criteria correspond to different video portions. VQA ratings over the specific video portions are correlated with the MOS provided by the medical experts. Ultimately, for high correlations, computerized VQA evaluations may be used to predict medical video's diagnostic yield.

This work has generated 1 journal publication, 2 magazine papers, 2 chapters, and 10 conference papers as documented in APPENDIX 1.

1.3 Guide to Thesis Contents

Chapter 2 includes background information on incorporated technologies used for the successful implementation of the system's objectives. H.264/AVC is discussed, while a detailed portrayal of the error resilience features defined in the standard along with examples that relate to medical video streaming are illustrated. Wireless networks advances is one of the key components in the successful deployment of mobile-healthcare systems and services. The entirety of the candidate heterogeneous networks for medical video transmission is presented. Bandwidth availability with respect to medical video bandwidth demands is depicted. Communication protocols necessary for establishing a connection between the transmitting/receiving parties and responsible for conveying clinical video data are also highlighted. Video quality assessment considerations for both objective and subjective evaluations, is documented. Unique requirements associated with clinical video quality assessment with respect to different medical video modalities and underlying technologies are discussed.

Chapter 3 contains the literature review. The most relevant and recent approaches in the research area of medical video streaming telemedicine systems are analytically presented. The undertaken approaches and incorporated technologies are illustrated.

Efficient source encoding, QoS monitoring for adaptation to network status, and video quality assessment are some of the techniques that summarize the current trends. Limitations of existing systems are highlighted. Future directions that address these limitations and also provide for integration of future technologies are discussed.

Chapter 4 provides the proposed system's methodology. Individual system component methodologies are analytically presented in a step by step fashion employing system diagrams. Emphasis is given to the aspects addressed by each component, along with justification for the undertaken approaches. More specifically, source encoding aspects including compression and error resiliency, candidate wireless networks requirements and associated QoS parameters, and video quality assessment methods for both objective and subjective evaluations are considered. Correlation investigation between objective and subjective ratings is also addressed. Seamless and efficient integration of all system components into a unifying framework for the wireless transmission of robust medical video is depicted.

Chapter 5 presents a comprehensive evaluation of the proposed methods and approaches. Analytical objective video quality assessment is performed to depict the efficiency of the proposed system. Compression efficiency, error resiliency, bandwidth demands for different wireless channels, and system's parameters optimization are included. Subjective video quality assessment provided by two medical experts is used to validate the clinical significance of the transmitted medical video. The degree of correlation between the assessment of different clinical criteria and associated objective ratings is also presented.

Chapter 6 provides a thorough discussion on the performance of the proposed system. Original aspects of the work and the contribution of the proposed system to the research area of medical video transmission telemedicine systems are documented.

A comparative evaluation of the achieved results and those of already published studies is performed. Advantages over existing approaches are depicted. Limitations of the current study are also portrayed.

Chapter 7 includes the conclusions and future work. A report of the system's objectives and degree of fulfilment are presented. Main achievements are summarized both in terms of technical and clinical views. The present challenges are identified and directions as to the enhancement of the proposed system with new techniques and technologies, which appear as the natural continuation of this thesis, are listed. Advances in wireless technologies, namely LTE-advanced and WirelessMAN-advanced, the 4G of mobile communication networks, are expected to allow transmission of higher resolution and frame rates for medical video applications and is a matter of future investigation. Scalable video coding, diagnostically relevant rate control, and new H.264/AVC and H.265 based error resilience encoding techniques for an even increased efficiency are currently planned. Development of new objective video quality assessment techniques with respect to underlying technologies and different medical video modalities will be also considered for future investigation.

Chapter 2

Background on H.264/AVC, Wireless Technologies, and Video Quality Assessment

2.1 H.264/AVC

H.264/AVC is the current state of the art video coding standard [4]. It was jointly developed by the ISO/IEC motion pictures experts group (MPEG) and ITU-T video quality experts group (VCEG) who formed the Joint Video Team (JVT). H.264/AVC met the growing demand of multimedia and video services by providing enhanced compression efficiency significantly outperforming all prior standards (MPEG-x and H.26x, see Fig. 2.1).

H.264/AVC can provide for bitrate reductions of up to 50% for equivalent perceptual quality compared to its predecessors [11]. Its design enables transportation over heterogeneous networks to be carried out in a friendly-manner. To attain the abovementioned, H.264/AVC defines a video coding layer (VCL) and a network abstraction 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 [12]. Its enrichment and refinement resulted in the provided compression efficiency. Fig. 2.2 depicts the basic encoding structure of H.264/AVC. On the other hand, NAL is a 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/AVC coding and transmission network-independent. An example of VCL and NAL functionality is

illustrated in Fig. 2.3. 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/AVC offers a range of error resilience techniques for a wide variety of applications. To this end, H.264/AVC defines different profiles and levels. Each profile and level specify restrictions on bitstreams, hence limits on the capabilities needed to decode this bitstreams [4]. Baseline, main, extended and high profiles assume different processing devices tailored for different applications and offer incremental level capabilities (and therefore complexity), that is alleviating constraints on bitstreams. Fig. 2.4 demonstrates unique features to each profile, while Table 2.1 summarizes some of the different capabilities of each level, including resolution, frame rate, maximum allowed bandwidth, and maximum coded picture buffer (CPB) and decoded picture buffer (DPB).

Telemedicine systems target end user devices such as mobile smart phones. In the context of this study, error resilience methods found in the baseline profile, specifically designed for video streaming to mobile devices, are considered. Error resilience techniques can be further distinguished as to where error control actually takes place. Error resilience at the encoder, error concealment at the decoder, as well as interactive approaches based on feedback communicated from the receiver. In Table 2.2, the entirety of the existing error resilient techniques, both in H.264/AVC, as well as in earlier digital video compression standards are summarized. Exploitation of the aforementioned approaches is application specific and all incorporated aspects should be considered before deploying an error resilient telemedicine system. The focus here is to discuss and demonstrate some of the most significant error resilient features of H.264/AVC. A thorough overview of the standard, performance and

complexity analysis, error resilience features and discussion exploiting H.264/AVC in the context of IP based networks [13] can be found in [11], [12], [14]-[18].
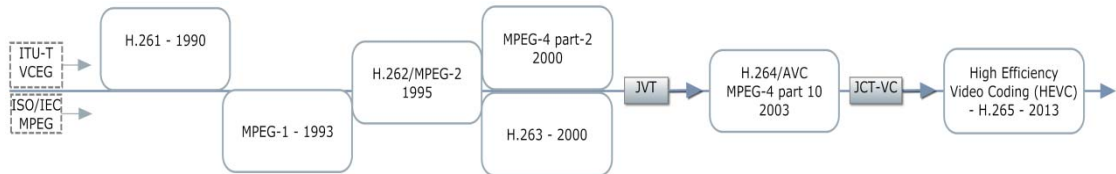
Fig. 2.1. Timeline of video coding standard's development.
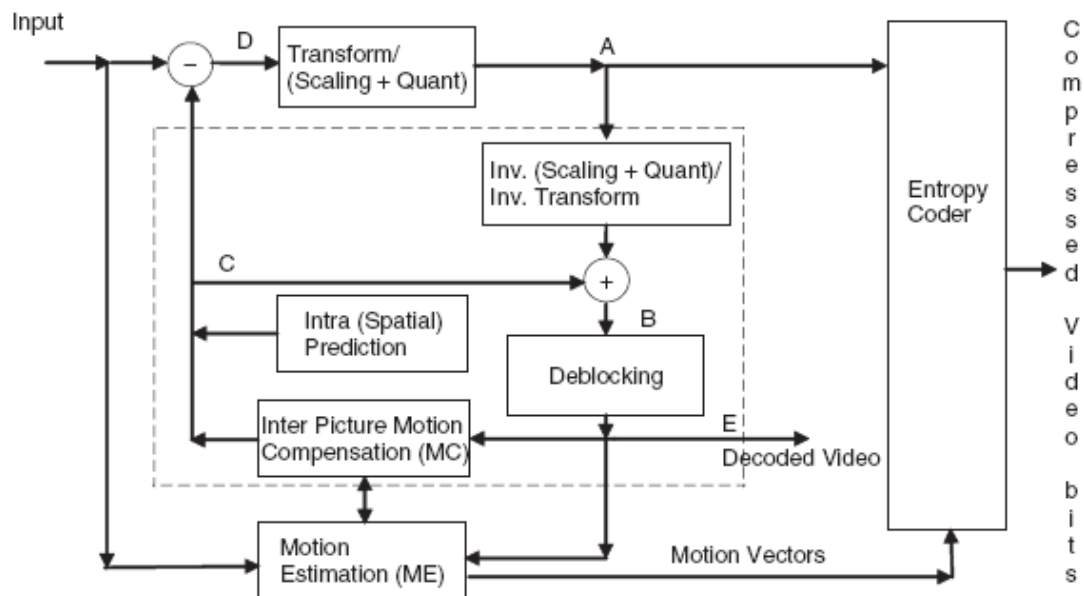
Fig. 2.2. H.264/AVC coding structure (example based on [7])
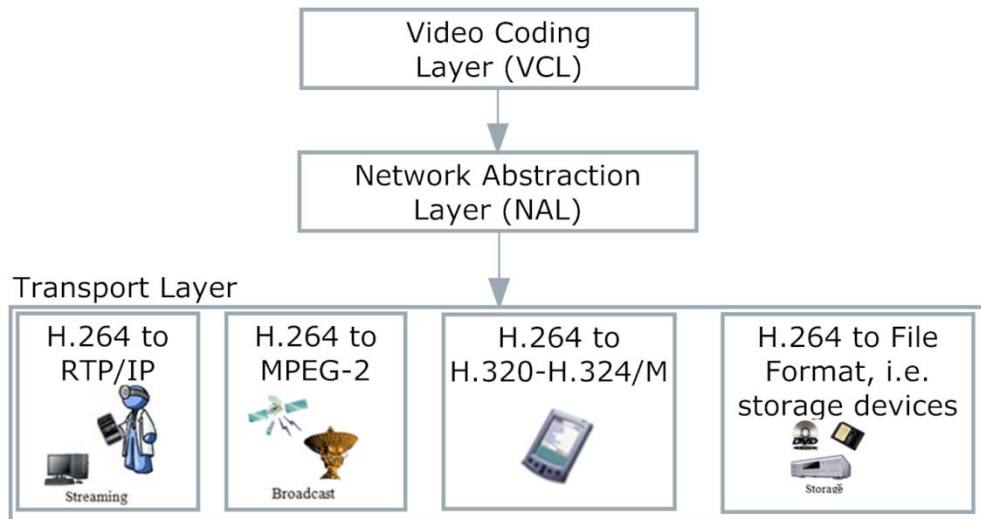
Fig. 2.3. H.264/AVC network abstraction layer (NAL) and video coding layer (VCL).
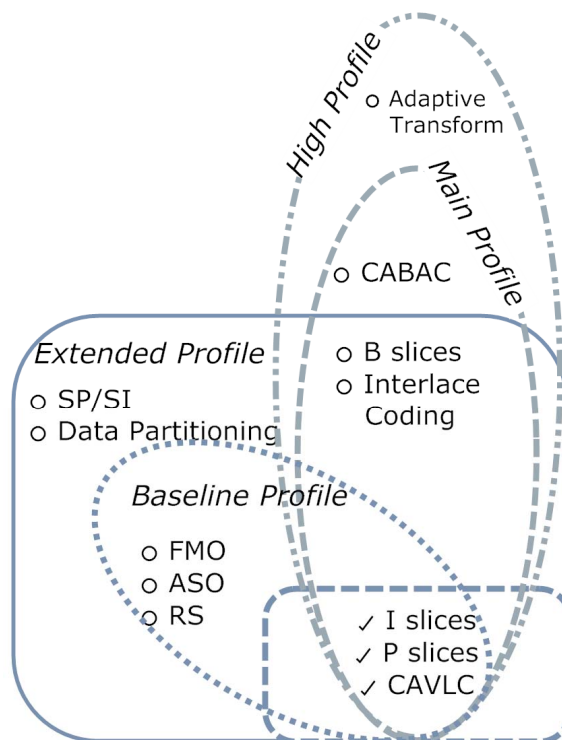
Fig. 2.4. H.264/AVC baseline, main, extended, and high profiles features.
Context-adaptive variable length coding (CAVLC), flexible macroblock ordering (FMO), arbitrary slice ordering (ASO), redundant slices (RS), switching-predictive/switching-intra (SP/SI), context-adaptive binary arithmetic coding (CABAC).

Table 2.1. H.264/AVC Level maximum supported parameters.

Level id.	Max. Video Bitrate	Max Frame Size (MB)	Max MB per second	Max CPB[a] Size (MB)	Max Delay at Max Bitrate	Max DPB[b] size (bytes)	Resolution, Frame Rate, Max Buffer Picture
1	64	99	1485	175	2.7	152 064	**SQCIF** 128× 96 @ 30fps (8) **QCIF** 176×144 @ 15fps (4)
1b	128	99	1485	350	2.7	152 064	**QCIF** 176×144 @ 15fps (4)
1.1	192	396	3000	500	2.6	345 600	**QCIF @** 30fps (9)
1.2	384	396	6000	1	2.6	912 384	**CIF** 352×288 @ 15fps (6)
1.3	768	396	11880	2	2.6	912 384	**CIF ×** 30fps (6)
2	2	396	11880	2	1	912 384	**CIF ×** 30fps (6)
2.1	4	792	19800	4	1	1 824 768	**525 HHR** 352×480 @ 30fps (7) **625 HHR** 352×576 @ 25fps (6)
2.2	4	1620	20250	4	1	3 110 400	**525 SD** 720×480 @ 15fps (6) **625 SD** 720×576@12.5fps(5)
3	10	1620	40500	10	1	3 110 400	**525 SD ×** 30fps (6) **625 SD ×** 25fps (5) **VGA** 640×480@30fps (6)
3.1	14	3600	108000	14	1	6 912 000	**720p HD** 1280×720@30fps(5)
3.2	20	5120	216000	20	1	7 864 320	**720p HD ×** 60fps (5)
4	20	8192	245760	25	1.25	12 582 912	**720p HD ×** 60fps (9) **1080 HD** (1920×1088) × 30fps (4) **2k × 1k** (2048 × 1024) × 30fps (4)
4.1	50	8192	245760	62.5	1.25	12 582 912	as above
4.2	50	8704	522240	62.5	1.25	13 369 344	**1080 HD ×** 60fps (4) **2k × 1k ×** 60fps (4)
5	135	22080	589824	135	1	42 393 600	**2k × 1k ×** 72fps (13)
5.1	240	36864	983040	240	1	70 778 880	**2k×1k @**120fps (16) **4k × 2k** 4096×2048@30fps (5)

[a]Coded picture buffer (CPB), [b]Decoded picture buffer (DPB).

Table 2.2. Error resilience techniques in video coding standards.

Technique		Video Coding Standards	Channel Adaptive Technique
Robust Entropy Coding	Resync Markers	MPEG-1/ H.261	NO
	RVLC	MPEG-4/ H.263	NO
	Data Partitioning	MPEG-2/ H.263	NO
	FMO	H.264/AVC	NO*
	ASO	H.264/AVC	NO*
	Redundant Slices	H.263	NO*
	SP/SI	H.264/AVC	YES*
Intra Updating	Periodic I-MB	MPEG-4/ H.263	NO*
	Preemptive I-coding	MPEG-4/ H.263	NO*
	Random I-coding	MPEG-4/ H.263	NO*
	Intra block refreshing by RD	H.264/AVC	NO*
	Multiple Reference	H.263	NO*
	UEP & LC	MPEG-4/ H.263	YES*
	MDC	MPEG-4/ H.263	YES*

*Error resilience techniques can be used both in a non channel adaptive and a channel adaptive environment. In this Table we record the earliest video coding standard to adopt the listed error resilience techniques. No classification is made between versions of these standards. Thus, these techniques or enhanced versions of them are included in forward standards. Some of these techniques may also be compatible with prior standards.

2.1.1 Encoding Modes and Frame types

Frame encoding modes can have a significant impact on both error propagation and video compression performance. A summary of the different modes is provided next:

- Intra-mode: Intra-mode is the procedure where intra-prediction is used for coding a video frame (I-frame). Here, all the information used for encoding is restricted within the frame. No prediction from previous or future frames is allowed. As a result, intra-mode encoded frames require higher bitrates than inter-mode encoded frames (discussed next). On the other hand, the use of intra-mode coding significantly limits error-propagation in wireless video transmission networks.

- 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. Clearly, decoding errors in the reference picture will be propagated to the predicted frame. Prediction reduces the bandwidth requirements at the expense of error-resilience.

  - 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. Again, errors from previously decoded pictures propagate to the predicted frame. On the other hand, B-mode pictures require less bandwidth than both P-mode and I-mode pictures.

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 around. Intra coding is mostly employed as an error resilience feature for periodic updates (i.e. one I-mode picture in every Group of Pictures (GOP)) (also discussed in the next subsection). The ratio between P-frames and B-frames used is also application specific. A ratio of 2:1, or IBBPBBP coding structure has proved to be a good balance between single-directional and bi-directional prediction, widely used for internet video streaming,

2.1.2 Intra Updating

The insertion of an Intra coded frame in a sequence of every GOP frames is essential in broadcasting applications, given that is utilized for random access to the transmitted bitstream such as joining an ongoing session. However, the transmission of videos using completely intra coded frames in non pre-encoded (real-time) applications is rather limited. This is due to the fact that intra coding requires increased encoding time and involves considerable bandwidth (size) compared to predictive coded frames, usually unacceptable in limited bandwidth, strict time delay applications. Pre-defined, as well as random intra-macroblock refresh is used instead to battle error propagation in error prone wireless environments [14], [19]. Intra-macroblock refreshing can prove particularly efficient in the presence of a low end-to-end delay feedback channel. A feedback channel can provide information as to which part of the picture is affected by losses and needs to be intra-coded in order to limit error propagation.
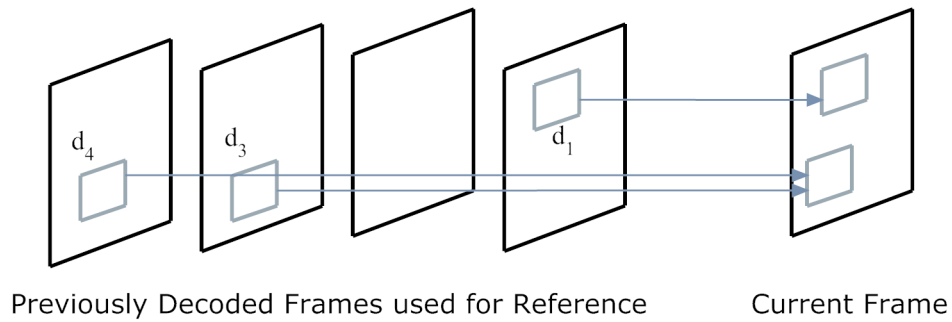
Fig. 2.5. Multiple reference frames prediction. Example based on [11].

2.1.3 Multiple Reference

H.264/AVC allows the utilization of up to 16 reference frames for prediction during encoding. Employing a certain number of previous or future frames which all contribute to the encoding of the current frame, rather than just a single reference frame (past or future) enhances predictive coding accuracy. Given the computational intensity of using multiple reference frames though and increased memory usage at the decoder, the number of reference frames should be selected wisely, especially for real time applications. Studies have shown multiple reference frames to work better in the presence of a feedback channel, notifying the encoder to avoid erroneously received frames at the decoder for motion estimation purposes [19].

2.1.4 Flexible Macroblock Ordering, Redundant Slices and Arbitrary Slice Ordering

An innovative error resilient feature introduced by H.264/AVC is flexible macroblock ordering [20]. 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. A slice contains a number of MBs, the basic block coding unit of H.264/AVC. 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.

36
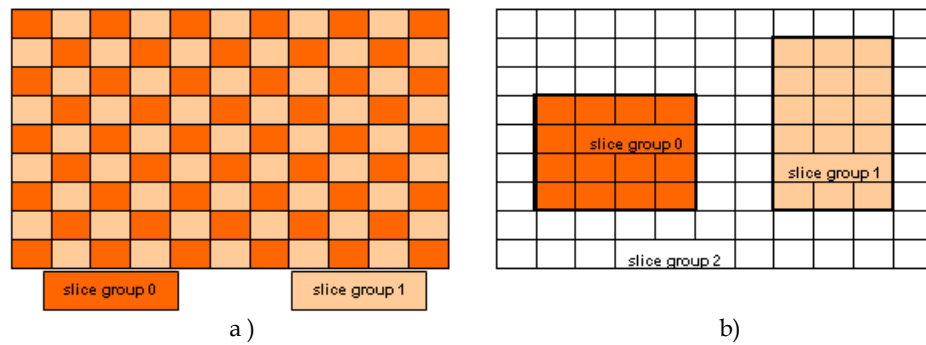
Fig. 2.6. Flexible macroblock ordering. a) Scattered slices and b) Foreground(s) (ROIs) and leftover.

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.

The most interesting case is FMO type 2, designed for defining rectangular slices as foreground(s) and background. These slices can be used to define regions of interest for encoding and transmission. Slices may overlap. However, the MBs can only belong to one slice. In the event of a packet carrying a whole slice gets dropped, H.264/AVC allows the transmission of *redundant slices (RS)*. An RS can be encoded both differently as well as with the same encoding setting as the corresponding primary slice. The decoder is responsible for replacing a corrupted primary slice with its equivalent redundant representation. The latter error resilience technique is highly efficient for communication in noisy environments in the absence of a back channel. More details on FMO can be found in [20]-[21].

Arbitrary Slice Ordering (ASO) [11] enables slices to be essentially transmitted independently of their order within a picture. As a result, they can be also decoded out of sequence, thus reducing the decoding delay at the decoder. ASO is particularly effective in environments where out-of-order delivery of a packet is possible such as the internet or wireless networks, or packet based networks in general.

37

2.1.5 Data Partitioning, SP/SI Slices

The basic idea in data partitioning lies in the observation that not all bits in a bitstream carry equal information. On the contrary, data bits can be categorized according to their importance, with certain bits being more important than others. Data partitioning in H.264/AVC allows the partitioning of a normal slice in up to three parts. Each part can be paired accordingly with unequal error protection (UEP) during transmission. Data partition (DP) A contains the most important slice information such as MB types and motion vectors (MVs), and possible loss or corruption of DP A, constitutes the remaining two partitions of no use. Second in importance comes DP B, which consists of intra-coded block patterns (CBPs) and I-block transform coefficients, while DP C incorporates inter CBPs and P-block coefficients. More detailed description can be found in [11], [16] along with recommended actions when partition loss is detected [11].

Switching-predictive/switching-intra (SP/SI) [22] are two new picture types introduced in the H.264/AVC design that allow the decoder to switch between two or more pre-encoded bitstreams. These bitstreams are constructed from the same source sequence, but are of different bitrate and quality. Besides the obvious benefit of channel adaptation, this dual nature feature proves particularly efficient in terms of error resilience, especially in the presence of a feedback channel which enables the decoder to trigger the encoder to perform a bitstream switch, regaining in that way lost synchronization resulting from data losses or errors. Nevertheless, valuable bandwidth is preserved, since recovering from an error does not incorporate the transmission of an I-frame. The SP/SI scheme can be further used for operations such as fast-forward, reverse, etc. SP/ SI slices and data partitioning are not supported by the baseline and main profiles of H.264/AVC.
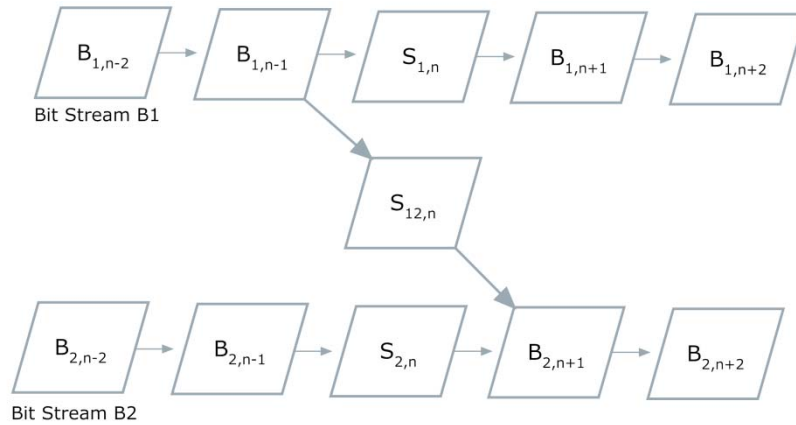
Fig. 2.7. SP/SI Slices.
A switch is triggered to a less aggravating for the network's resources state. Example based on [16].

2.2 Wireless Transmission Technologies

2.2.1 Wireless Transmission Technologies

In terms of wireless infrastructure, the Global System for Mobile communications (GSM) [23] is the most popular standard. GSM signified the transition from analog 1st generation (1G) to digital 2nd generation (2G) technology, and despite originally designed for voice communication, is also capable of data transfer at rates of up to 9.6 kbps. At such low rates, GSM can only be used for still images. It cannot be used for the transmission of medical video. Next, a brief summary of mobile cellular networks' and WiMAX technologies evolution is provided, highlighting theoretical and typical data rates, and associated medical video transfer capabilities. Table 2.3 also summarizes access technologies, operating frequency bands, as well as typical delays.

The evolution of mobile telecommunication systems from 2G to 2.5G (iDEN, GPRS, EDGE) and subsequently to 3G (W-CDMA, CDMA2000, TD-CDMA), 3.5G (HSDPA [24] and HSUPA [25], HSPA+ [26]), mobile WiMAX [27], and LTE [28], [29] systems facilitates both an always-on model (as compared with the circuit-switched mode of GSM), as well as the provision of higher data transfer rates and

lower delays, thus enabling the development of more responsive telemedicine systems [3]. Evolving wireless communications networks' theoretical upload data rates range from 50 kbps - 86 Mbps (GPRS: 50 kbps, EDGE: 236.8 kbps, UMTS: 384 kbps, evolved EDGE: 947 kbps, EV-DO Rev A: 1.8 Mbps, mobile WiMAX Rel.1: 4Mbps, EV-DO Rev B: 5.4 Mbps, HSPA+ Rel. 8: 11.5Mbps, HSPA+ Rel. 9: 23 Mbps, LTE Rel. 8: 86 Mbps [6]). In practice, typical upload data rates are significantly lower. More specifically typical upload data rates range from (i) GPRS: 40-50 kbps, (ii) EDGE: 70-135 kbps, (iii) evolved EDGE: 150-300 kbps (expected), (iv) UMTS: 200-384 kbps, (v) HSPA: 500 kbps - 2 Mbps, (vi) HSPA+: 1-4 Mbps [6].

In terms of bandwidth, both 2.5G and 3G provide sufficient rates for medical image and biosignals transmission. For medical video transmission 3G rates are sufficient for QCIF (176x144) resolution medical video transmission, as well as specific regions of interest (ROIs). Nevertheless, CIF (352x288) resolution video may be transmitted if diagnostic ROI-based encoding is employed, thus lowering bandwidth demands. High speed packet access (HSPA) and HSPA+ 3.5G technologies enable the transmission of high quality CIF resolution video, as well as up to 4CIF (704x576) resolution video and beyond. The clinical benefit of transmitting high-resolution medical video is an open area of research.

WiMAX release 2.0 and LTE-advanced networks [30], [31] conforming to the IMT-advanced requirements [32], [33] will constitute the next generation family of technologies, namely 4G. Low latency, high mobility, high bandwidths (targeting 100 Mbps for high mobility and 1 Gbps for low mobility, in the downlink), and Quality of Service (QoS) provisions, are expected to significantly boost the development of mobile-healthcare systems and services.

Satellite systems provide a variety of data transfer rates starting from limited to high-speed data rates of up to n x 64 kbps and beyond. Satellite links utilization in healthcare benefit from world-wide coverage [34], but require line of sight and comparably higher power for similar bit rates. We refer to [3] for a description of healthcare systems that demonstrate wireless transmission over satellite links.

WLAN is a flexible data communications system implemented as an extension to or as an alternative to a wired LAN. WLANs transmit and receive data over the air, minimizing the need for wired connections. Thus, WLANs combine data connectivity at tens of Mbps with, however limited coverage (in the region of tens of meters at the level of an access point or typically few km within an enterprise) and hence user mobility. To extend coverage over larger distances, wireless mesh networks are also being considered. These networks are peer-to-peer multi-hop wireless networks, in which stationary nodes take on the routing functionality thus forming the network's backbone. Basically, they act as a gateway to high-speed wired networks for mobile nodes (clients) which communicate in a peer manner. Clearly, when tens of Mbps are available, we have sufficient bandwidth for transmitting multiple video bit-streams. The problem of coverage still remains.

2.2.2   4G networks conforming to IMT-advanced requirements

Mobile WiMAX and LTE are today's state-of-the-art deployed networks, incorporating a plethora of sophisticated technologies. However, while they do meet some of the IMT-advanced requirements specified by ITU-R, they both fail to uniquely address all listed requirements. This led to the development of even more efficient techniques and concepts, in order to facilitate conformance with aforementioned specifications. Resulting WirelessMAN-Advanced and LTE-advanced technologies based on IEEE 802.16m and 3GPP Release 10 specifications

respectively, participated in IMT-advanced evaluation process, following ITU-R call for candidate technologies [35]. Evaluation process [36] concluded that both candidate technologies met IMT-advanced requirements and are now officially considered as 4G technologies.

Being backwards compatible, this family of technologies targets improved uplink and downlink rates of 100 Mbps and 1 Gbps respectively, increased coverage and throughput, enhanced mobility support (up to 350 km/h), reduced latencies less than five milliseconds, enhanced QoS provision [37], efficient spectrum usability and bandwidth scalability, and security, with simple architectures, in favour of the end user.

2.2.2.1 Worldwide Interoperability for Microwave Access (WiMAX)

Worldwide Interoperability for Microwave Access (WiMAX) was firstly standardized for fixed wireless applications in 2004 by the IEEE 802.16-2004 [38] and then for mobile applications in 2005 by the IEEE 802.16e [27] standards. After an initial hype of the WiMAX, lately there has been scepticism as to its successful wide deployment, in favour of the LTE. However, current standardization 802.16m [39], also termed as IEEE WirelessMAN-Advanced met the ITU-R IMT-advanced requirements and succeeded as a 4G technology.

WiMAX targets a plethora of applications ranging from common internet access to internet protocol television (IPTV) and voice over IP (VoIP) services, as well as other demanding services that require broadband wireless access (BWA), mobility and QoS support. WiMAX deployment is destined for wireless metropolitan area networks (WMANs) given the high supported throughput and increased coverage. In general, WiMAX can serve as an alternative to DSL/T1, cable and optical lines, as well as mobile cellular systems. Aforementioned technologies are not mutually exclusive, but

42

they rather increase the capacity of the end user for BWA. Data, voice, and video, are some the key candidate applications for WiMAX networks. These services benefit from the capability of WiMAX networks to serve them simultaneously, while at the same time allowing QoS prioritization.

WiMAX originally specified the air interface in frequency ranges between 10-66 GHz (802.16), however later amendments (802.16a) defined frequencies bands below 11GHz (2-11 GHz) [40]. Today's licensed deployment is typically in the range of 2.3, 2.5-2.7, 3.5, and 5.8 GHz, while 4G frequency bands will facilitate deployment between 450-3600 MHz [39]. Channel bandwidth allows great flexibility in the sense that it allows WiMAX operators to consider channel bandwidths between 1.25, 2.5, 5, 10, and 20 MHz (802.16e). In 802.16m scalable bandwidth between 5-40 MHz for a single RF carrier is considered, extended to 100 MHz with carrier aggregation to meet IMT-advanced requirements. WiMAX employs a set of high and low level technologies to provide for robust performance in both line-of-sight (LOS) and non-line-of-site (NLOS) conditions. A thorough overview of WiMAX standardization process and evolving concepts and technologies up to IEEE 802.16e standards appears in [40], while recent advances are described in detail in [7], [41], [42]. Performance evaluation of IEEE 802.16m is documented in [36]. Key features of physical (PHY) and medium access control (MAC) layers are discussed next.

Physical Layer Features

As already mentioned above, WiMAX standards define the air interface and more specifically MAC and PHY layers. PHY layer's central features include *adaptive modulation and coding (QPSK, 16-QAM, 64-QAM)*, *hybrid automatic repeat request (HARQ)*, and *fast channel feedback*. Key technology in the success of WiMAX

Table 2.3. Wireless technologies evolution and associated data transfer rates and delays.

Wireless Technology	Frequency band	Theoretical Data Rates	Typical Data Rates	Delay
2G-GSM (TDMA)	850/900/ 1800/1900 MHz	9.6 – 115 Kbps.	10 Kbps	N/A
2.5G-GPRS (TDMA)	as above	DL:UL: 9.6 - 171.2 Kbps	DL:UL: 30-50 Kbps	< 700ms
2.5G-EDGE (TDMA)	as above	DL:UL: 9.6-384 Kbps	DL:UL: 75-135 Kbps	<600ms (Rel.99) <350 (Rel. 4)
Evolved EDGE (TDMA)	as above	DL: 1.89Mbps UL: 947Kbps	DL: 175-350Kbps UL:150-350Kbps	<200ms
3G-UMTS (FDD, W-CDMA)	800/850/1500/1700/ 1800/1900/2100 MHz	DL:UL: 144 kbps - 2 Mbps.	DL:UL: 220-384 Kbps	<250ms
3G-UMTS (TDD, TD/CDMA)	1900–1920/2010–2025/ 1850–1910/1930–1990/ 1910–1930/2570–2620 MHz	as above	as above	as above
3.5G-HSPA (HSDPA Rel. 5) (HSUPA Rel. 6)	as above	DL: 14Mbps UL: 5.8 Mbps	DL : 1-4 Mbps UL : 500Kbps -2Mbps	<150ms
3.5G-HSPA+ Rel. 8 (Rel. 9)	as above	DL: 42(84) Mbps UL: 11.5(23) Mbps	DL : 1.9-8.8 Mbps UL: 1-4 Mbps	<100ms
3.5G-LTE (OFDMA)	as above + 700/800/1800 MHz 2.6 GHz	DL: 326 Mbps UL: 86 Mbps	DL: 5.9-21.5 Mbps UL: TBD	<70ms
Mobile WiMAX (OFDM) (IEEE 802.16e)	2.3, 2.5-2.7, 3.5, 5.8 GHz (licensed)	DL: 46 Mbps UL: 5.6 Mbps	DL:UL: TBD	<70ms
4G-LTE-Advanced (OFDMA)	As in LTE <6GHz (IMT-Advanced)	DL: 1Gbps UL: 100 Mbps	TBD	TBD (target<5ms)
4G- WiMAX (OFDM) (IEEE 802.16m)	450-470/698-960/1710-025/ 2110-2200/2300-2400/2500-690/3400-3600 MHz	DL: 1Gbps UL: 100 Mbps	TBD	TBD (target<5ms)

N/A: not available, TBD: to be determined, DL: downlink, UL: uplink.

systems in general is OFDM employed in the PHY layer. OFDM, and more specifically scalable *orthogonal frequency division multiple access (SOFDMA)*, allows dividing transmission bandwidth into multiple subcarriers. The number of subcarriers starts from 128 for 1.25 MHz channel bandwidth and extends up to 2048 for 20 MHz channels. In this manner, dynamic QoS tailored to individual application's requirements can be succeeded. In addition, orthogonality among subcarriers allows overlapping leading to flat fading. In other words, multipath interference is addressed by employing OFDM while at the same time available bandwidth can be split and assigned to several requested parallel applications for improved system's efficiency. The latter is true for both downlink (DL) and uplink (UL). *Multiple input multiple output (MIMO)* antenna system allows transmitting and receiving multiple signals over the same frequency. Two types of gain are possible, namely spatial diversity and spatial multiplexing. For spatial diversity, unique configuration enables enhanced link quality by combining independent faded signals resulting from simultaneously transmitted duplications of the same information. For spatial multiplexing, increased throughput is achieved via the parallel spatial channels transmission of multiple streams.

Medium Access Control Layer Features

In the MAC layer, the most important supported features can be summarized in QoS provision through different prioritization classes, direct scheduling for both DL and UL, efficient mobility management, as well as security. The 5 QoS categories are depicted in Table 2.4. According to individual application requirements the appropriate QoS class is considered and the corresponding UL burst will be scheduled and data rate assigned. For real-time video streaming as in the case of emergency telemedicine scenarios, rtPS QoS class best suits the applications requirements.

Table 2.4. QoS classes in WiMAX technology.

QoS Category	Applications	QoS Specifications
UGS – Unsolicited Grant Service	VoIP	• Maximum Sustained Rate • Latency Tolerance • Jitter Tolerance • Grant Interval
rtPS- Real-Time Polling Service	Streaming Audio or Video	• Minimum Reserved Rate • Maximum Sustained Rate • Latency Tolerance • Traffic Priority
ertPS – Extended Real-Time Polling Service	VoIP with Voice Activity Detection/Silence Suppression	• Minimum Reserved Rate • Maximum Sustained Rate • Latency Tolerance • Jitter Tolerance • Traffic Priority
nrtPS – Non-Real-Time Polling Service	File Transfer Protocol (FTP)	• Minimum Reserved Rate • Maximum Sustained Rate • Traffic Priority
BE – Best Effort	Data Transfer, Web Browsing	• Maximum Sustained Rate • Traffic Priority

Example based on [13].

Mobility management is also well addressed in 802.16e and current 802.16m standards, which was an issue in 802.16d primary standard for fixed connections. With a theoretical support of serving users at 120 km/h in 802.16e, established connections provide adequate performance for vehicles moving with speeds between 50-100 km/h. In 802.16m, mobility support is extended for mobile speeds up to 350 km/h as depicted in the evaluation of IMT-advanced requirements.

Enhanced security, especially when compared to competing technologies (like WLANs) is one of the key features in WiMAX networks, shielding the end-user from a variety of threats. Improved security is based on extensible authentication protocol (EAP) for authentication, while advanced encryption system (AES) is employed for encryption. BS and SS are authenticated via the privacy key management (PKM) algorithm.

2.2.2.2 Long Term Evolution (LTE)

Long term evolution (LTE) mobile communication networks have been standardised through the 3rd generation partnership project (3GPP) Release 8. LTE facilitates significant improvements with respect to 3G and HSPA systems. It provides increased data rates (see Table 2.3), improved spectral efficiency and bandwidth flexibility ranging between 1.4 -20 MHz, and reduced latency (less than 5ms user-plane latency, seamless handover). While being backwards compatible enabling seamless deployment on existing infrastructure, LTE shares a set of new cutting edge technologies and simple architecture.

In the physical layer, *multiple-carrier multiplexing OFDMA* is adopted for the downlink, while *single-carrier frequency-division multiple-access (SC-FDMA)* is the access scheme used in the uplink. SC-FDMA utilizes single carrier modulation, DFT-spread orthogonal frequency multiplexing, and frequency domain equalization, and has similar performance to OFDMA. *FDD and TDD* are jointly supported in a single radio carrier. LTE allows *multi-antenna applications for single and multi users through MIMO* technology (up to 4-layers in the downlink and 2-layers in the uplink). QPSK, 16-QAM, and 64-QAM modulation schemes are supported. RLC and MAC implement *automatic repeat request (ARQ)* and *hybrid-ARQ (HARQ)* for increased robustness in data transmission.

Enhanced mobility support (up to 350 km/h), efficient multimedia broadcast multicast service (MBMS), QoS provision, security, and cell capacity up to 200 active users summarize the key features provided by LTE systems.

LTE-Advanced

LTE-advanced bridges the gap between LTE Release 8 and ITU-T IMT-Advanced requirements. LTE-advanced is standardized in 3GPP Release 10. Novel technologies

found in LTE Release 8 and enhancements in Release 9 are incorporated in the design of LTE-advanced systems. Compared to LTE, further improvements in peak data rate, spectrum efficiency, throughput and coverage, as well as latency reductions are facilitated.

Data rates in the order of 1Gbps for low mobility and 100 Mbps for high mobility are achieved via the adoption of new technologies such as *carrier aggregation*. Carrier aggregation enables wider bandwidth transmission up to 100 MHz utilizing a combination of frequency blocks, thus increasing system's peak data rates. An example utilization scenario of 100 MHz system bandwidth utilizes 5 LTE 20 MHz blocks, for uplink and downlink, at 3.5 GHz and FDD. *Enhanced MIMO techniques* in LTE-advanced systems include 8-layer transmission in the downlink and 4-layer transmission in the uplink. *Coordinated multipoint transmission and reception (CoMP)* is another novel technique defined in the standard which provides for increased throughput on the cell edge. The key idea is that multiple e-Node Bs cooperate to coordinate transmission relevant aspects that provide for reduced interference and increased throughput for UEs located near the cell edge. *Relaying techniques* for efficient cell deployment and coverage and parallel processing for even greater reductions in latency targeting less than 50ms and 5ms in control and user planes respectively are also exploited.

A comprehensive review of LTE and LTE-advanced technologies, comparative analysis, utilization scenarios combining different technology components to demonstrate conformance to IMT-advanced requirements, and also IMT-advanced evaluation results can be found in [28]-[31], [36], [43]-[46].

2.3 Communication 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) [47] or alternative Session Initiation Protocol (SIP) [48] 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. The transmission control protocol (TCP) [49] uses retransmission and traffic monitoring to secure packet delivery to destination. This property made TCP highly efficient for HTTP [50] applications. However, when streaming video, retransmission time may result in long delays and even 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) [51] 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.

For real-time audio and video transmission, RTP [52] provides end-to-end delivery services. 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 or 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, see for example [53]. 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 already mentioned, RTP packets are usually transferred over UDP.

The resulting packets use the internet protocol (IP) for delivery through the network. As a result, the packets include RTP/UDP/IP headers.

2.4 Diagnostic Validation

Diagnostic validation is the most significant requirement for emerging medical video transmission systems. Diagnostic validation requires an accurate assessment of the diagnostic capacity of the transmitted medical video.

2.4.1 Objective Video Quality Assessment

Objective VQA largely differs from image quality assessment which has seen significant growth over the last five years. However, today's most established VQA metrics are extensions of algorithms originally designed for image quality assessment. Such examples are the Peak Signal-to-Noise Ratio (PSNR) and the average structure SIMilarity Index (SSIM) [54]. PSNR utilizes the Mean Square Error (MSE) between the original and transmitted videos on a frame basis. PSNR is one of the most established metrics for both image and video quality assessment and is often used as

the benchmark metric for measuring the objective performance of transmitted video. Given a reference, uncompressed video I and a distorted (i.e. following transmission) video K, with mxn spatial dimensions, the PSNR is computed as follows:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (2.1)$$

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \quad (2.2)$$

Structural SIMilarity (SSIM) is another popular image quality assessment metric that has been extended to address video quality assessment [55]. Similarly to PSNR and all image quality assessment algorithms used for VQA, the index is computed separately for each frame and the average yields the video quality assessment index. The following equations are abstracted from [55].

$$S(x,y) = l(x,y).c(x,y).s(x,y) = \frac{4\mu_\chi\mu_y\sigma_{\chi y}}{(\mu_\chi^2+\mu_y^2)(\sigma_\chi^2+\sigma_y^2)} \quad (2.3)$$

$$\mu_\chi = \bar{\chi} = \frac{1}{N}\sum_{i=1}^{N}\chi_i, \qquad \mu_y = \bar{y} = \frac{1}{N}\sum_{i=1}^{N}y_i \quad (2.4)$$

$$\sigma_\chi^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2, \qquad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (2.5)$$

$$\sigma_{\chi y} = \frac{1}{N-1}\sum_{i-1}^{N}(\chi_i - \bar{\chi})(y_i - \bar{y}) \quad (2.6)$$

To overcome the computation stability problem observed when $\left(\mu_\chi^2 + \mu_y^2\right)$ or $\left(\sigma_\chi^2 + \sigma_y^2\right)$ is close to 0, especially for flat image regions, equation (2.3) is modified so that it now reads:

$$SSIM(\chi,y) = \frac{(2\mu_\chi\mu_y+C_1)(2\sigma_{\chi y}+C_2)}{(\mu_\chi^2+\mu_y^2+C_1)(\sigma_\chi^2+\sigma_y^2+C_1)} \quad (2.7)$$

$$C_1 = (K_1 L)^2 \text{ and } \quad C_2 = (K_2 L)^2 \quad (2.8)$$

For the two new constants added, $L = 255$ for 8 bits/pixel gray scale images and K_1 and K_2 are set to $K_1 = 0.01$, and $K_2 = 0.03$ respectively in [55], values which are also adopted here.

Visual Signal to Noise Ratio (VSNR) [56] is computed based on:

$$VSNR = 10log_{10}\left(\frac{C^2(I)}{VD^2}\right) = 20log_{10}\left(\frac{C(I)}{\alpha d_{pc}+(1-\alpha)\frac{d_{qp}}{\sqrt{2}}}\right) (2.9)$$

where *VD* corresponds to the Visual Distortion, $C(I)$ denotes the RMS contrast of the original image *I*, d_{pc} the perceived contrast of the distortion, and d_{qp} the disruption of the global precedence. For detailed algorithmic concepts we refer to [56].

Information Fidelity Criterion (IFC) and extension Visual Information Fidelity (VIF) differentiate from traditional full reference quality assessment algorithms by integrating natural scene statistics (NSS) modeling together with image/video degradation and human visual system (HVS) models in the their algorithmic design. Natural images are modeled in the wavelet domain using Gaussian scale mixtures (GSM) [57], [58]. The corresponding quality indexes are given by:

$$VIF = \frac{\sum_{j\in subbands}I\left(\vec{C}^{N,j};\vec{F}^{N,j}|\vec{s}^{N,j}\right)}{\sum_{j\in subbands}I\left(\vec{C}^{N,j};\vec{E}^{N,j}|\vec{s}^{N,j}\right)} (2.10)$$

where the summation of subbands of interest is computed, and $\vec{C}^{N,j}$ stands for *N* elements of the random field (RF) C_j that correspond to the coefficients of subband *j* and so forth. *E* and *F* similarly denote the visual signal at the output of the HVS model from the reference and test images of the corresponding subbands, respectively. A step by step reasoning is documented in [57].

$$IFC = \sum_{k\in subbands}I\left(C^{N_k,k};D^{N_k,k}|s^{N_k,k}\right) (2.11)$$

As above, $C^{N_k,k}$ denote N_k coefficients from RF C^k of the k*th* subband [58]. $D^{N_k,k}$ and $s^{N_k,k}$ reflect distorted image and an RF of positive scalars subbands respectively.

The Noise Quality Metric (NQM), based on Peli's constrast pyramid accounts for the following parameters: (1) variation in contrast sensitivity with distance, image dimensions, and spatial frequency; (2) variation in the local luminance mean; (3) contrast interaction between spatial frequencies; and (4) contrast masking effects [59]. Quality assessment index is computed as:

$$NQM = 10log_{10}\left(\frac{\sum_x \sum_y O_s^2(x,y)}{\sum_x \sum_y (O_s(x,y) - I_s(x,y))^2}\right) (2.12)$$

Where $O_s(x,y)$ and $I_s(x,y)$ correspond to the simulated instances of modeled restored image and restored images respectively. Model aspects are presented in detail in [59].

Weighted Signal to Noise Ratio (WSNR) was first proposed in [60] and reflects a weighting SNR computation to the human visual system. The latter is based on a contrast sensitivity function (CSF) given by:

$$H(f_r) = 2.6(0.0192 + 0.0114f_r)exp(-(0.0114f_r)^{1.1}) (2.13)$$

where f_r is the radial frequency in cycles per degree that maps spatial frequency to visual frequency and is computed given the spatial frequencies and the viewer's distance input parameters. The implementation found in [62], also uses the angular dependency introduced in [61] for normalizing the radial frequency in CSF computation.

We refer to [62] for algorithmic details and MATLAB implementations for each of the afore-described image to video quality assessment metrics.

In general, there is a growing demand for the development of video quality metrics that are significantly different from image quality metrics. As documented in [9], [63] there are significant problems associated with the use of PSNR for video quality evaluation purposes, as it fails to adequately correlate with perceived video quality. Aspects such as perceived motion, QoS, and Quality of Experience (QoE) [64], [65] need to be considered during the design of new video evaluation techniques. Of particular interest is the new motion-based video integrity evaluation (MOVIE) metric described in [66], which claims to outperform all VQA algorithms to date [9]. Towards this direction, the video quality metric (VQM) [67] developed by the national telecommunication and information administration (NTIA) is also capable of

addressing some of the issues. A thorough study of objective and subjective VQA methods is found in [9]. Having said this, there is also a strong need for new, clinically-driven video quality metrics.

2.4.2 Clinical Quality Assessment

There are unique challenges associated with both objective and subjective clinical VQA. In addition to the motion and QoS aspects of conventional VQA, unique clinical criteria, often different for each medical modality, need to be properly assessed. These clinical criteria often correspond to specific video portions that are of diagnostic interest. These regions of diagnostic interest are much more sensitive to compression and error impairments given their significant clinical contribution to the diagnostic yield of the particular medical video. On the other hand, diagnostic ROI encoding may lead to diagnostically lossless medical videos. Clearly, both schemes are not adequately assessed by current objective VQA algorithms.

For subjective clinical evaluation, while we do expect that the basic features of subjective quality assessment criteria described in [68] will play a role in emerging medical video quality assessment standards, unique clinical criteria will also need to be adequately modeled. Diagnostic yield of transmitted medical video is restricted by a number of factors including resolution, frame rate, and end user equipment. Diagnostic capacity of a QCIF resolution medical video at 15 fps displayed on a PDA largely differs from a 4CIF resolution at 30 fps displayed on a laptop. Hence, appropriate clinical rating schemes should also address the aforementioned implications.

Chapter 3

Literature Review on Wireless Medical Video Transmission Systems

3.1 Case Study: Wireless Medical Video Transmission Systems using 3G

A summary of the most relevant and recent wireless medical video transmission studies with emphasis given on 3G channels appears in Table 3.1 and Table 3.2. These systems are further categorized as ROI and non-ROI based systems. A detailed analysis of the considered parameters in each individual study is provided in Table 3.3. An overview of earlier m-health systems and services classified by wireless transmission technologies appears in [3].

3.1.1   Diagnostic Region of Interest (ROI) based Systems

In this section we present studies that require the identification of a region of interest that is of diagnostic interest. The systems are summarized in Table 3.1.

A study associated with atherosclerotic plaque ultrasound video transmission appears in [69]. Automatic plaque segmentation is used for specifying the most significant ROI. The ROI is extended based on a visual attention model of what human readers would find interesting. Non-diagnostic regions are then blurred using a smoothing filter. In this manner, the smoothed area will require significantly less amount of bandwidth, given the high-compression involved when encoding smoothed frames using existing encoding standards.

Table 3.1. Diagnostic ROI based medical video transmission systems over 3G (and beyond) wireless networks.

Author	Year	Resolution., Frame Rate, BitRate[4]	Encoding Standard	Medical Video Modality	Comments
Tsapatsoulis et al. [69][1]	07	352x288 @ 10fps 10 videos average: 507.2 Kbps	MPEG-2/ MPEG-4	Common Carotid Artery Ultrasound video	A saliency-based visual attention ROI coding for low bit-rate medical video transmission.
Doukas et al. [70][1]	08	Resolution: N/A @ 25 fps 128-256 Kbps	H.264/ AVC	Skin Legion and MRI images, Trauma video snapshots	Adaptive transmission based in context awareness (patient status and network state).
Rao et al. [71][1,3]	09	360x240 @ 30 fps 500Kbps	MPEG-2	Pediatric respiratory distress related videos	ROI coding which incorporates different quantization levels for ROI and non-ROI, targeting diagnostically lossless encoding utilizing physician expert feedback.
Martini et al. [72]	10	480x256@15fps 300 Kbps	H.264	Cardiac Ultrasonography	Context –aware FMO encoding using unequal error protection (UEP).

[1]Simulation, [2]Real-time, [3]Clinical evaluation by medical experts (presented video parameters achieve diagnostically lossless encoding for preset clinical criteria), [4]For real-time transmission, both encoded video bitrates and available channel bitrates are provided.

In [70], the authors propose a context-aware medical image and video transmission system. Here, context is defined based on patient status. It is either normal or urgent status. The urgent state is determined based on the ECG signal, blood pressure, pulse rate, heart rate, and oxygen level. Region of interest encoding is used to encode trauma in optical images or skin lesions for transmission in low bandwidth systems. In the urgent state, data prioritization is achieved through scalable video encoding using a base layer and enhancement layers.

In [71], the authors assign different quantization levels to the foreground and background ROIs (given in bits per pixel). The authors vary bit allocations in the two ROIs to match different bitrates. Six different encoder states are considered. According to bandwidth availability, a switch is triggered to the most relevant state. A state switch is considered every group of pictures (GOP). The system is validated for diagnostically lossless performance based on a variety of clinical criteria for pediatric respiratory distress related videos.

The authors in [72], develop an ROI-based system for cardiac ultrasonography video. Context-awareness is introduced via the identification of diagnostically

important ROIs. These correspond to the ultrasound's most important area denoted by the clinician, the fan-shaped sector produced by the ultrasound scanner, and the background, also containing ECG and patient data. Different quantization parameters are assigned to different ROIs which are also associated with unequal channel error protection to overcome transmission errors.

A new technique for echocardiogram compression is proposed in [73]. The key observation is that echocardiogram operation modes can be divided into two groups, the sweep modes and the 2D modes, each mode relating to particular characteristics in ultrasound video display. Based on individual mode characteristics, intelligent encoding to match mode properties is considered. In the sweep modes, only a slice is compressed for each frame, which corresponds to the new information added with respect to the previous frame. In the 2D modes different algorithms are used to efficiently compress the greyscale (3D SPIHT) and colour (RLE) information. For all modes of operation this new technique outperforms conventional encoding approaches that do not distinguish between encoding modes and mode properties.

3.1.2   Medical video transmission systems without using regions of interest

The use of regions of interest provides for very efficient systems that target specific diseases. While the ROIs provide for a direct way to address clinical criteria, defining ROIs can be a very difficult task and may not be always possible. In this case, other criteria need to be explored.

A portable tele-trauma system that provides for simultaneous transmission of trauma video, medical images, and ECG signals is presented in [77]. Media transformation, data prioritization, and application-level congestion control provide adaptation to network conditions. Data transformation tackles image and video compression using JPEG and M-JPEG standards, respectively. For image

transmission, following manual ROI segmentation of the first transmitted image, only the ROI part is sent in subsequent image transmissions. Priority classes are set to high for ECG transmission, medium for image transmission, and low for video transmission. During simultaneous transmission of the aforementioned priority data classes, when congestion occurs, a congestion-control algorithm which adjusts the trauma video's transmitting frame rate is triggered to alleviate congestion effects leading to packet losses. System's performance demonstration includes effective resolution and frame rate investigation for which transmission of simultaneous streams is possible.

Scalable video coding (SVC) employing spatiotemporal scalability for a number of ultrasound videos is found in [74]. The effectiveness of transmitting different resolution and frame rates over different wireless channels is investigated based on wireless transmission medium parameters and physician's feedback. Data rate, packet losses, delay, and jitter are measured for different scalability layers streamed over wireless local area networks (WLANs) and 3G channels. Based on the examined parameters and also in agreement with diagnostic quality provided by medical experts, higher SVC layer supported by each wireless channel is depicted. Clinical assessment includes evaluation of different clinical criteria.

A performance analysis of an end-to-end mObile Tele-Echography using an ultra-Light rObot (OTELO) system is presented in [78]. Low-bitrate echography ultrasound medical video transmission over 3G channels is investigated and functional bounds are provided. Quality (in terms of PSNR), resolution, frame rate, and maximum delay that provide for acceptable diagnostic performance over the tested 3G link are illustrated. The latter recommendation follows clinical evaluation performed by the

Table 3.2. Non-ROI based medical video transmission systems over 3G (and beyond) wireless networks.

Author	Year	Resolution., Frame Rate, BitRate[4]	Encoding Standard	Medical Video Modality	Comments
Chu et al.[77][2]	04	{320x240 and 160x120} <5fps Channel Bitrate: 50-80 Kbps	M-JPEG	Trauma video	Real time trauma video transmission. Network adaptation enabled through media transformation, data prioritization, and application-level congestion control.
Garawi et al. [78][2,3]	06	176x144 @ 5fps 18.5-60 Kbps Channel BitRate: 64 Kbps	H.263	Echocardiogram	A performance analysis of an end-to-end mObile Tele-Echography using an ultra-Light rObot (OTELO)
Pedersen et al.[74][2,3]	09	320x240 @ 10fps 349 Kbps Channel BitRate: 380 Kbps	H.264/ AVC (Scalable)	Echocardiogram	Spatiotemporal scalability over different wireless networks. Diagnostic quality and how it's affected by network parameters.
Istepanian et al. [75][2,3]	09	176x144 @ 8-10fps 50-130 Kbps Channel BitRate: 360 Kbps	H.264/ AVC	Abdomen	QoS Ultrasound Steaming Rate Control (Q-USR) algorithm based on reinforcement learning that satisfies a medical QoS criterion.

[1]Simulation, [2]Real-time, [3]Clinical evaluation by medical experts (presented video parameters achieve diagnostically lossless encoding for preset clinical criteria), [4]For real-time transmission, both encoded video bitrates and available channel bitrates are provided.

expert physician, also responsible for remotely monitoring patients using the tele-operated robot.

In [75], the authors develop an QoS Ultrasound Streaming Rate Control (Q-USR) algorithm. Based on the concept of reinforcement learning, the frame rate is varied as a function of the state of the network, while at the same time conforming to a pre-defined medical QoS criterion. That is, the varied frame rate and resulting quality indices are monitored not to fall below of what is diagnostically acceptable with respect to the bounds described in [78]. Robotic tele-ultrasonography system OTELO is also used in this study for abdomen ultrasound medical video. Simulations and real time experiments over 3.5G wireless networks validate the efficiency of the proposed mechanism with respect to the default rate–control algorithm of the JM H.264/AVC reference software. This is also verified by clinical video quality assessment.

Ultrasound video acquired using the OTELO system is also used in [79], where multilayer control is employed to optimally tune source and channel encoding parameters for enhanced streaming of medical video. Frame rate, quantization step,

intra-refreshing period and average code-rate channel protection are the key parameters that are varied in these experiments to construct different encoder states. A trigger between states is considered on a per second basis, based on medical video's quality index, channel indications such as average SNR, and network parameters including PLR, delay, and jitter. Experimentations are initially performed over WLANs.

3.2 A Synopsis of Current Status and Associated Parameters

The basic source encoding features that directly impact a video's clinical capacity are quantization levels, resolution, and frame rates. High quality corresponds to increased diagnostic capacity, but also accounts for increased bandwidth demands, a limiting factor in bandwidth limited wireless networks. Resolution and frame rates are also bounded by the wireless channels' upload data transfer speeds as well as the end-user devices. Higher resolutions close to medical video's acquired resolution allows addressing more clinical criteria. Moreover, motion assessment benefits from higher frame rates. Hence, according to the underlying transmission medium data rates and end user device's capabilities, minimum bandwidths, resolutions and frame rates that provide for diagnostically lossless systems for specific clinical criteria should be sought.

Wireless video transmission QoS parameters include packet loss rate, delay, and jitter. Signal attenuation due to mobility, fading channel, and distance from the base station, congestion, background traffic, modulation and coding schemes, and packet sizes are some of the contributing parameters to QoS deterioration. While some packet losses are tolerable, clinical quality cannot be compromised leading to misdiagnosis. Moreover, increased delays usually result in additional dropped packets.

Table 3.3. Medical video transmission systems varied parameters.

Study	QP/Bitrate	Source Encoding			Wireless Transmission			Clinical VQA
		Resolution	Frame Rate	Error Resilience	PLR	Delay	Jitter	
Tsapatsoulis et al. [69][1]	ROI-based	√	√	-	-	-	-	√
Doukas et al. [70][1]	√	-	-	-	°	°	°	-
Martini et al. [72]	ROI-based	-	-	Channel Protection	√	√	-	+
Rao et al. [71][1,3]	ROI-based	-	-	-	-	-	-	√
Cavero et al. [73]	√	√	√	-	-	-	-	-
Chu et al. [77][2]	√	√	√	-	-	°	°	-
Pedersen et al.[74][2,3]	Scalable	√	√	-	°	°	°	√
Girawi et al. [78]	√	√	√	-	-	√	√	√
Istepanian et al. [75][2,3]	√	-	√	-	°	°	°	+
Martini et al. [79]	Encoder States	-	√	Channel Protection	√	√	√	-

√ denotes that the incorporated encoding parameter was investigated during the experiments. For network parameters it signifies that it was taken under consideration to modify encoding state. For clinical VQA it designates a scoring scale for different clinical criteria.
° denotes that measurements are depicted for the specific network parameter.
+ denotes that the transmitted medical video's quality conforms to a preset clinical threshold or verified by a medical expert but not listed in the study.

To address these issues, error resilient encoding, adaptation to varying network conditions by modifying encoding setting, as well as combination of the latter two schemes can be employed. Error resilient encoding for enhanced diagnostic performance in noisy wireless channels is not addressed in the literature. Clearly, error resilient encoding needs to be further exploited in the design of new medical video telemedicine systems given the error prone nature of wireless channels. However, the trade-off between coding efficiency and error resiliency should be thoroughly investigated with respect to available bitrates. Redundant bits used for this purpose occupy bandwidth that can be used for conveying clinical information.

Network monitoring for adaptation to network status by triggering a switch to a different encoder state is used in [71], [75], [79]. This typically involves a combination of quality, resolution, and frame rate downplay which accounts for reduced bandwidth demands and hence channel load. Efficient utilization of the

network's resources can be also achieved through diagnostically relevant ROI encoding [71], [72], where quality levels are varied as a function of the diagnostic significance of the video regions. SVC encoding [70], [74] is an enabling technology for different encoder states. It provides the capability of defining multiple enhancement layers accounting for different quality, resolution, and frame rate that match different bitrates. Network adaptation is also achieved via data prioritization as illustrated in [70], [77].

While the aforementioned techniques efficiently address packet losses due to channel load, they do not account for corrupted packets or burst errors, typical in wireless channels. Given that packet losses are inevitable in wireless transmission, medical video transmission systems should be equipped with appropriate error resilient and concealment mechanisms. Next generation 4G channels promise QoS provision through data prioritization classes for low delay-fixed bandwidth allocation. This will enable the development of telemedicine systems that will exploit a priori adaptation to channel state and provide for a diagnostically robust system at a required bitrate, irrespective of channels conditions.

A summary of the key varying parameters in each of the presented studies in the literature appears in Table 3.3. As rationalized above, these are classified as source encoding and wireless transmission parameters, while clinical VQA is central to the success of medical video transmission telemedicine systems. A brief examination of the table aids in deducting some useful observations regarding implementation aspects of these systems. The key observation is that error resilience encoding is not addressed in the relative literature. While channel encoding is undertaken by two studies, error resilience techniques for robustness in error prone wireless channels are not employed by corresponding studies in the literature.

For source encoding parameters, the frame rate is the most commonly varied parameter for network adaptation, followed by resolution. The latter two parameters are usually adjusted to provide for different bitrates requirements that typically correspond to different encoder states. This can always be facilitated by compression levels and rate control algorithms. Increased PLR and delays resulting in quality degradation are often used to trigger an encoder state switch. Delay and jitter are also measured for system's performance evaluation. Another significant observation is that clinical VQA is employed by more than half of the presented studies. On the other hand, only three studies incorporate clinical assessment of different clinical criteria based on a scoring scale.

Chapter 4

Methodological Framework

4.1 Overall System Description

The overall system diagram, depicting a step-by-step analysis of the incorporated components and associated input parameters at each step, is summarized in Fig. 4.1. An iterative procedure for determining the medical video's minimum bitrate, resolution, frame rate, and amount of error resiliency that achieve acceptable diagnostic performance utilizing medical expert's feedback in noisy environments is depicted. The proposed system's diagram includes two distinct modes of operation. In the first mode, used to determine a diagnostically acceptable encoding setting prior to transmission, the blocks contained in the rectangle annotated by the dotted line are not considered. These blocks correspond to wireless video transmission component and the involved parameters relate to error resiliency and incorporated bitrate. The latter schemes are considered during the second mode of operation. In the first mode, the objective is to determine the minimum threshold values for a diagnostically lossless encoding setting. This translates to the minimum quality, resolution, and frame rate that satisfy the clinical criteria set by the medical expert. The aforementioned parameters are varied through an iterative procedure for a number of medical videos and the medical expert evaluates their diagnostic capacity. The relationship between input parameters that describe the medical video encoding process and the corresponding clinically acceptable parameters is discussed in more detail in subsections 4.1.1 and 4.1.2 (see also Fig. 4.2). Having derived threshold values for the

above-described parameters based on clinical evaluation, the system is now trained to proceed with the actual streaming of the medical video. In the second mode of operation including all blocks, a coarse to fine parameter optimization is used to derive the minimum bitrates (by varying the quality levels) and redundant slices rate that provide for diagnostically lossless video quality over error prone wireless channels. These parameters are varied as a function of the packet loss rates experienced by the wireless network. The trade off between error resiliency and encoding efficiency for reliable performance in noisy environments is investigated. The aim is to (i) maximize medical video's diagnostic robustness by recovering from large PLR while at the same time (ii) conforming to the established diagnostically lossless threshold values and (iii) balancing encoding efficiency. Next, we discuss the input parameters for each system diagram's block.

4.1.1   System diagram's input parameters

According to the underlying medical video modality, the corresponding clinical criteria are determined. The assessed clinical criteria are specified by the corresponding medical expert. These criteria are then used to define the video encoding scheme. The video encoding block is thoroughly discussed in Fig. 4.2 in the following subsection. Channel data rate is used for describing the channel, while the targeted mobile device resolution and frame rate playback are also input to the video encoding block. It is important to note that the channel may not support playback at full resolution and maximum frame rates (due to limited upload data rates). On the other hand, there is little reason to transmit at resolutions and frame rates that will exceed the mobile device playback abilities, which will result in unnecessary power consumption and channel load. In addition, display resolution and frame rate directly relate to medical video's diagnostic yield. For example, diagnostic capacity of a 4CIF

resolution medical video at 30 fps displayed on a portable computer largely differs from a QCIF resolution medical video at 10 fps displayed on a PDA, while there is also a significant bitrate demand variation. Therefore, channel data rate and mobile device specifications *a priori* knowledge also aid in the specification of the clinical criteria by the medical expert. Mobile device's decoding capabilities in terms of supported profiles and levels are also considered. As described in subsection 2.1, different levels account for different supported resolution, frame rate, and buffer size. For the wireless video transmission block, packet loss rate parameter is used to describe the transmission errors over error prone wireless channels. Delayed packets are also treated as lost (hence discarded) if the arrival time exceeds the rendering time, constituting the contained information not effective. Clinical criteria are also input to the video quality assessment block, where the aforementioned implications are taken into consideration during the evaluation of the diagnostic capacity of the transmitted medical video. The video quality assessment procedure is depicted in Fig. 4.3 and is discussed in more detail in subsection 4.1.3.

4.1.2 Medical video encoding process

The basic structure of the encoder is depicted in Fig. 4.2. Video resolution and frame rate adjustments are performed given the available channel data rate, the mobile device specifications, and the corresponding clinical criteria assessed, as already described. The clinical criteria are then used to determine the diagnostic ROIs. The key concept here, which is also applicable to other medical video modalities (see also chapter 3), is that certain video regions contain the clinical information that is needed to assess certain clinical criteria. Classification of these video regions according to the assessed clinical criteria allows the specification of video slices for independent encoding. The video quality of each slice is controlled by setting the value of the

corresponding QP, with respect to its clinical significance. As a result, diagnostically relevant encoding is achieved, providing for efficient utilization of the network's resources. The video slices are then combined to reconstruct the video frames. The RS parameter is used to control error-resilient encoding. The approach here is to use a larger number of RS for recovering from larger packet loss rates. Quantization parameter(s) and redundant slices values are varied through the iterative procedure described in Fig. 4.1. A coarse to fine parameter optimization determines minimum bitrates and also addresses the trade-off between error resiliency and encoding efficiency, for reliable performance in noisy environments.

4.1.3   Video quality assessment

The quality of the decoded video is assessed using both objective and subjective quality criteria (see Fig. 4.3). Subjective evaluation is performed by two medical experts for the preset clinical criteria, and the mean opinion scores (MOS) is documented. Objective evaluation is based on a number of the most established VQA metrics described in subsection 2.3. The correlation between the objective and subjective ratings is also evaluated. Ultimately, for high correlations, the objective measurements may be used to predict subjective quality. Having said this, it is often the case that high diagnostic quality can be established for high values of the objective evaluation. For example, video-slice PSNR values over a certain threshold can be used to establish that the video is of acceptable diagnostic quality. Whereas subjective clinical quality assessment evaluations are performed over the entire video, for objective video quality assessment, the video quality metrics are evaluated over the diagnostic video slice regions that correspond to individual clinical criteria (see subsection 4.4 and Table 4.1).
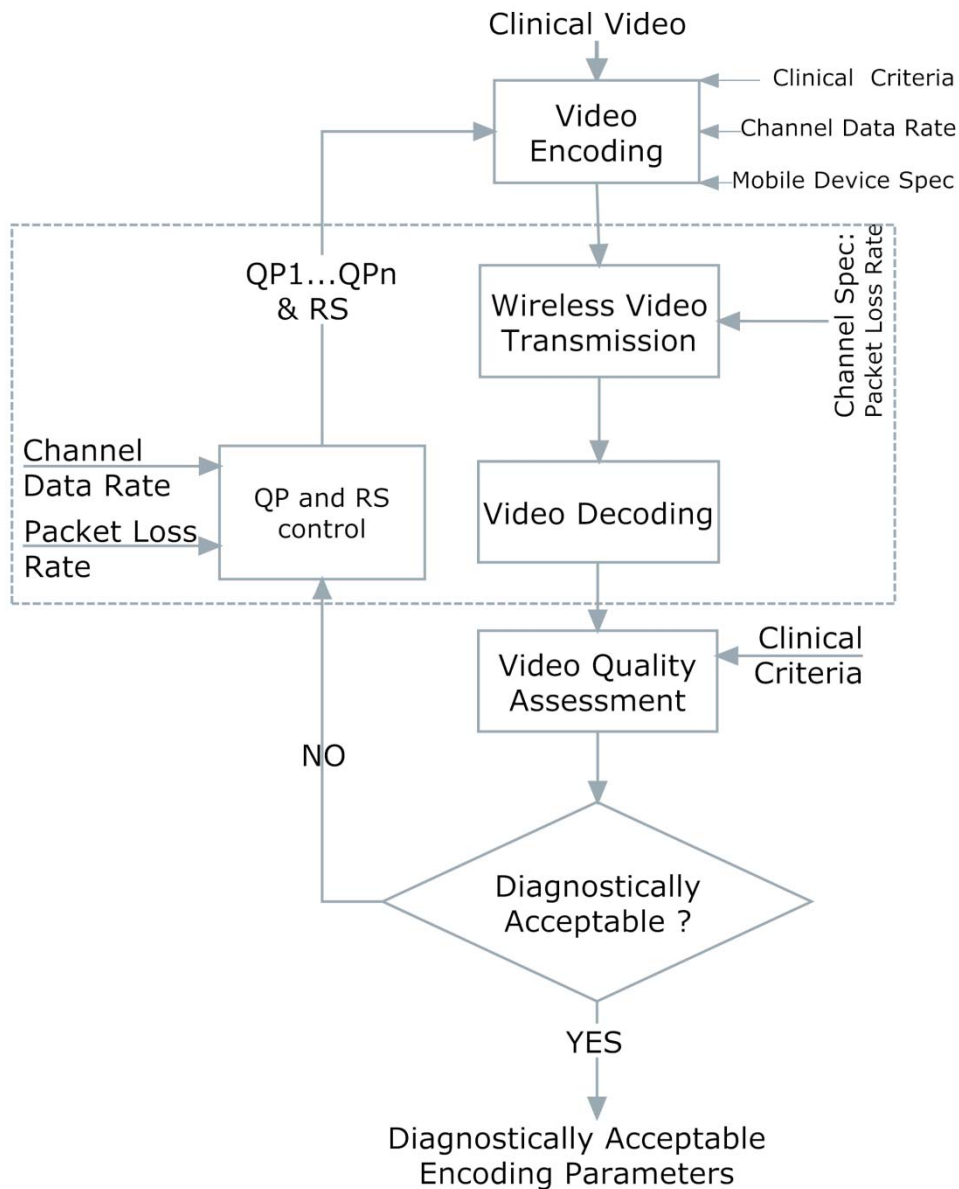
Fig. 4.1. System diagram.

System diagram incorporating all steps in the design of a reliable end to end medical video transmission system. The blocks forming the depicted system diagram correspond to the wireless medical video transmission architecture for emergency telemedicine illustrated in Fig. 1.1. The encoding block is responsible for pre-processing and H.264/AVC source encoding. Wireless video transmission block facilitates the video streaming server and the packet-based network cloud listing candidate wireless transmission technologies. Video decoding, video quality assessment, and resulting diagnostically acceptable setting decision correspond to the client/receiver block of Fig. 1.1, including the hospital premises and corresponding medical expert. Varying parameters through the iterative procedure depicted here are accommodated via the back-channel in Fig. 1.1, annotated by the grey dotted line.

Clinical Video

Channel
Data Rate
Mobile Device
Spec

Video Resolution
& Frame Rate
Conversion

Clinical
Criteria

Video Slice
Specification

ROI Spec

QP1
QP2
⋰ QPn

Video Slice
Encoding

Video Slice
Encoding

········

Video Slice
Encoding

+

RS rate

Frame Error
Resilient Encoding
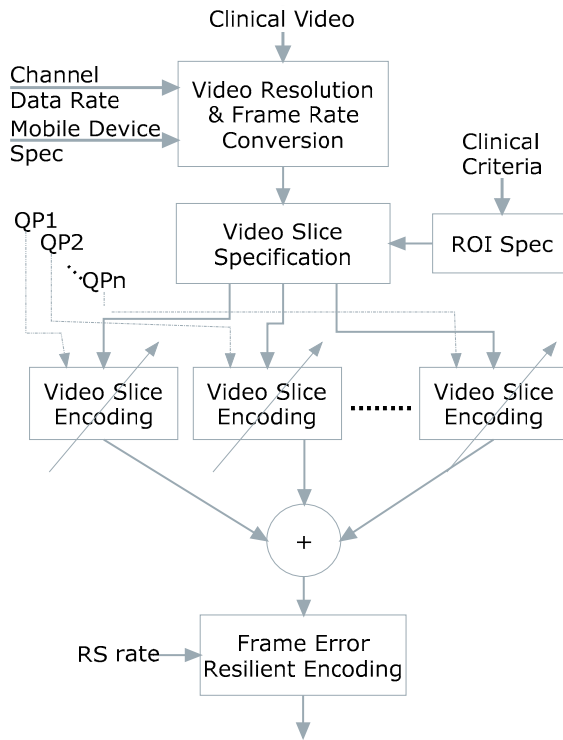
Fig. 4.2. Video encoding process including diagnostically relevant and error resilient encoding blocks. The proposed diagnostically driven variable quality slice encoding scheme using error resilient techniques. Channel knowledge and end user equipment are considered during pre-processing.

Decoded
Video

Decode Video
into Slices

Clinical
Criteria

Clinical
Criteria

Subjective
Evaluation

Objective
Evaluation

Quality
Measurements

Quality
Measurements
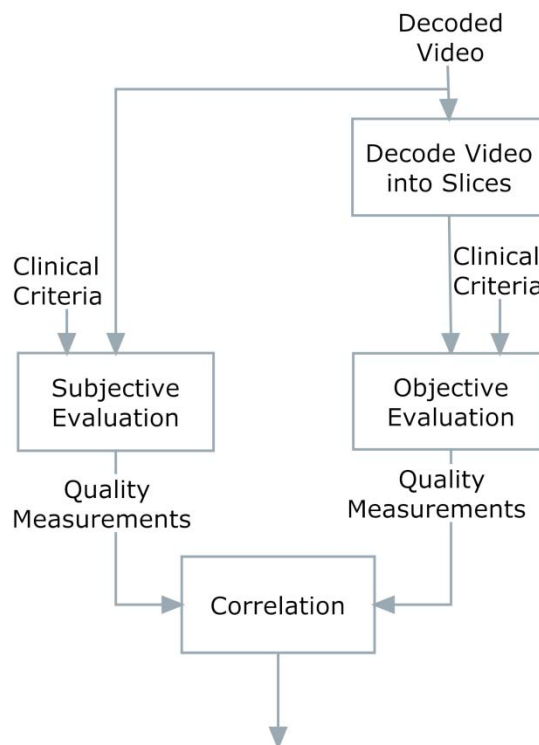
Correlation

Fig. 4.3. Objective and subjective video quality assessment, and correlation investigation, based on clinical criteria.

Fig. 4.4. Video slice variable quality encoding and decoding.
a) Video slice specification. b) The corresponding quantization parameter allocation map (QPAmap). c) The decoded video using variable quality slice encoding. Here with QPs 38/30/28 for background/ wall and ECG ROIs/ plaque ROI and quarter common intermediate format (QCIF-176x144 pixels, 11x9 macroblocks (MB)).

A detailed description of the proposed system's methodology is provided next. An in-depth analysis of the approaches and methods of each of the components parting the proposed system and how they integrate in an efficient system design is depicted. More specifically: (i) the processes of determining diagnostic video slices that correspond to clinical criteria assessment, (ii) variable quality slice encoding according to the slice's diagnostic importance, (iii) wireless transmission simulation, (iv) decoded video's quality assessment (objective and subjective), and (v) incorporated data set (material) and experimental setup are addressed.

4.2 Diagnostic Video Slice Specification

Video slice specification requires the identification of rectangular regions over the diagnostically relevant medical video portions (see Fig. 4.4). A number of techniques can be employed for this purpose which are medical video modality and deployed system specific. In the presence of a back channel, the medical expert can specify the diagnostic ROIs using two mouse clicks for any opposing corners for each ROI. This technique can be applied for other medical video modalities as well. Alternatively, in the absence of a back channel, segmentation algorithms, specific for each modality

might be employed. These rectangular regions of interest will be later used as input for FMO type 2 encoding.

For the purpose of this thesis, atherosclerotic plaque ultrasound video is considered. 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 the atherosclerotic plaque (causing the stenosis) features, including the type of the plaque characterized by its texture composition. Tracking of these features in time can aid in the prediction of the severity of abnormality. Intima media thickness (IMT) of the near and far walls also aid in this direction. The remaining regions of the video carry little diagnostic importance. Subsection 4.3 describes variable quality slice encoding according to the region's diagnostic importance.

To specify the plaque and wall video slice, the single-frame pixel-based segmentation algorithm introduced in [80] is used. To account for both plaque and wall motions, the maximum motions are estimated, by finding the minimum and maximum displacements over a number of frames (considering one every 5 frames for the first two cardiac cycles, and one every 10 frames for the remaining cardiac cycles). This avoids the need for tracking the plaque and wall throughout the streaming process. Here, the goal is not to provide accurate segmentation of the plaque or the different anatomical structures of the video. For video encoding purposes, the segmented plaque is extended out to fixed size MB (16x16 pixels) boundaries (see Fig. 4.4(a)-(b)), significantly larger than the pixel-level accuracy required for video segmentation purposes. Here, note that an MB is the basic block unit that each coded frame is partitioned in all video encoding standards since H.261 [11] and the unit used by FMO type 2. Furthermore, as noted earlier, the pixel-level accuracy of the segmentation method (given by 0.82 (+-) 0.95 mean +/- std pixels) has a limited

71

impact on the final video slice specification. More details regarding the segmentation method appear in the APPENDIX 2.

The wall ROI is similarly defined using the lower plaque boundary and the nearest wall. The success of this approach follows from the fact that these exams follow a clinically-established protocol for visualization of the plaque type, boundary, and stenosis. Pixel-level segmentation for the whole data-set appears in Fig. 4.5 in subsection 4.5. When the ECG video slice is to be transmitted, the lower-right MBs for ECG is simply allocated (see Fig. 4.4(a)). Detecting the ECG is straight-forward since it is the only part of the video that appears in the green colour channel. Recall that the relationship between the selected video slices and the clinical criteria is given in Table 4.1.

4.3 Variable Quality Encoding of Diagnostic Video Slices

Driven by the fact that clinical evaluation is based on the assessment of certain video portions, a spatially varying encoding scheme where quality levels are varied as a function of the diagnostic significance of the incorporated video regions is proposed. Encoding resources (i.e. quantization parameters accountable for video quality) are assigned according to video region's diagnostic importance. Diagnostically important regions are assigned lower QPs (i.e. better quality, more bits), whereas non-diagnostically important regions higher ones (lower quality, less bits). Using the iterative procedure for coarse to fine parameter optimization depicted in Fig. 4.1, a diagnostically lossless video quality is sought. The latter scheme serves the dual purpose of preserving network's resources (by reducing bandwidth demands) without compromising clinical capacity.

For independently encoding the diagnostic video slices, FMO is employed. In particular, FMO type 2 which allows the definition of rectangular slices as foreground(s) and background is used. To enable variable quality slice encoding, the encoder is modified to support a QP Allocation Map (QPAmap), which stores the QP of each MB (see Fig. 4.4(b)). The concept is similar to MB Allocation Map (MBAmap), which stores the corresponding slice number that each MB belongs to, an already implemented scheme in H.264/AVC. The QP of each video slice is parsed via the default configuration file used by FMO type 2 to define the boundaries of the rectangular ROIs, which is accordingly modified. Employing these minor adjustments at the encoder achieves variable quality FMO slice encoding. No change is made at the decoder and hence the resulting bitstream is H.264/AVC compliant (recall that slices are self-contained, hence they can be decoded independently). Moreover, the afore described modifications do not introduce any overhead during encoding, in terms of encoding and motion estimation time.

4.4 Transmission over Wireless Channels Simulation

Transmission over wireless channels presents a number of peculiarities, often unique for each application. Video streaming and in particular medical video telemedicine systems are very sensitive to varying network's conditions. Given that transmission errors are inevitable in today's error prone wireless channels, further video quality degradation due to varying state of the network greatly impacts diagnostic performance of the transmitted medical video. The effort of the current study is directed towards investigating efficient encoding mechanisms that allow medical video streaming over noisy channels. The latter typically involves recovering from large packet loss rates. For this purpose, comprehensive experimentation for

different packet loss rates using a packet loss simulator is performed. Higher packet loss rate values than the typical experienced losses when streaming over wireless channels are simulated. In this manner, errors that do not correspond to transmission errors, like discarded delayed packets, or errors due to channel load, congestion, signal attenuation, service interruption, and other fading channel conditions are addressed.

A modified version of the pseudo-random packet loss simulator included in JM H.264/AVC reference software [81] was employed. The simulator was significantly enhanced by adding an implementation of the random number generator described in [82] to provide drastically improved random performance. To this end, a variety of packet loss distributions were also integrated in the simulator. A uniform packet loss distribution was used throughout the experiments and all results were obtained by averaging 10 consecutive runs (see also Table 5.4). Burst errors were simulated by dropping a maximum of four consecutive packets.

4.4.1   OPNET Modeler Network Simulator

Medical video transmission using OPNET modeler network simulator is also considered. OPNET modeler 16.0 is used to simulate transmission over mobile WiMAX infrastructure. The approach in the proposed framework is to provide a robust encoding setting that recovers from high packet loss rates in extreme channel conditions. High packet losses are simulated using the random packet loss simulator described above. Here, the methodology is to simulate different network parameters such as path loss propagation models, and evaluate resulting QoS measurements with respect to the diagnostically acceptable parameter values derived through the coarse to fine parameter optimization described in Fig. 4.1. The objective is to demonstrate the ability of the proposed framework to adapt to different wireless transmission channels.

Furthermore, to both assess and demonstrate enhanced features associated with 3.5G mobile WiMAX technology.

4.5 Video Quality Assessment

For the assessment of the transmitted medical video quality, both subjective ratings provided by two medical experts, as well as objective ratings based on today's most established VQA algorithms are considered. Correlation investigation between the clinical assessment of plaque type and objective ratings computed over the atherosclerotic plaque region is depicted.

4.5.1 Clinical (Subjective) Video Quality Assessment

Independent grading of each of the three clinical diagnosis criteria listed in Table 4.1 is considered. These criteria correspond to the most important clinical findings for providing an accurate diagnosis and assessment of the patient's status. Having identified the formation of a plaque, the medical expert then proceeds to estimate the degree of stenosis during both systole and diastole, over the cardiac cycle. Assessment of plaque type enables the medical expert to assess the likelihood of a plaque rupture, leading to a stroke incident. Towards this goal, evaluation of the plaque morphology and the plaque components associated motions is of vital importance. Follow up exams evaluate the stenosis progress and plaque's composition components, investigation of which is usually based on texture quantitative analysis [84]. From Table 4.1, one can also see how each clinical quality criterion relates to the encoding of specific ROIs.

During clinical evaluation, a representative sample of one hundred (100) transmitted video instances was evaluated by two medical experts. This sample included different videos, quantization parameters, resolution and frame rates, and

packet loss rates, for all three competing schemes (see subsection 5.2). The videos were played back on a laptop at their original pixel size dimensions. The medical expert was able to view the original, uncompressed video first (at the transmitting resolution and frame rate), before assessing the transmitted video instances. According to Table 4.1, each diagnostic region received an independent evaluation score. The rating scale considered was between one (1) and five (5) (see Table 4.2). A rating of 5 was the highest possible. It signified that the clinical capacity of the decoded video was of the same quality as the original (uncompressed) video. A rating of 4 indicated that there was an acceptable loss of minor details. At the lowest scale, a rating of 1 would signify that the decoded video was of un-acceptably low quality.

Table 4.1. Diagnostic Region of Interest Contribution to each Clinical Rating.

	Plaque ROI	Wall ROI	ECG	Background
Plaque boundary detection	✓	✓		✓
Stenosis	✓	✓	✓	
Plaque Type	✓			

Table 4.2. Clinical Evaluation Rating System.

	Plaque Detection	Stenosis	Plaque Type
5	plaque(s) presence in transmitted video identifiable as in original	degree of stenosis in transmitted video determined as in original	plaque type classification in transmitted video as in original
4	plaque(s) presence easily diagnosed	enough clinical data to determine degree of stenosis	enough clinical data for plaque type classification
3	plaque(s) presence diagnosed, careful attention needed	clinical data only allow approximation of degree of stenosis	plaque type classification is case dependant
2	plaque(s) presence may be diagnosed after freeze of a clean frame	very limited ability to estimate degree of stenosis	not classified
1	not detectable	not determinable	not classified

4.5.2 Objective Video Quality Assessment

For objective VQA, a set of the most well known VQA algorithms found in the literature was employed. More specifically, PSNR, SSIM, VSNR, VIF, VIFP, IFC, NQM, and WSNR metrics summarized in subsection 2.4 were computed. The Metrix_mux software maintained by Cornel University provides a set of wrapper programs for MATLAB implementations of aforementioned algorithms [62]. The software was extended to tackle video quality evaluation since the current software version only supports image quality assessment.

Given that the atherosclerotic plaque ROI is the primary focus point of the medical expert, contributing to the clinical evaluation of all clinical criteria, objective VQA measurements in this study are computed over the plaque ROI slice. Diagnostic capacity of the transmitted medical ultrasound video of the carotid artery is directly associated with the atherosclerotic plaque's ROI quality. High quality visualization of the plaque ROI provides for accurate assessment of plaque presence, degree of stenosis, and plaque type, as it alleviates complications related with erroneous reconstructions. By restricting the objective evaluation to what is diagnostically relevant, the efficiency of the incorporated VQA algorithms is also increased. It is worth noting here that the quality levels of the plaque ROI in each of the investigated methods are set to be equal as described in the next chapter (see subsection 5.2), providing for a fair comparison. Following transmission, the plaque ROI is extracted from the original and transmitted videos and individually processed for objective evaluation. Recall that clinical evaluation is performed by assessing the whole video.

4.5.3 Correlation Investigation between Objective and Subjective evaluations

To evaluate the correlation between objective VQA algorithms and MOS of the clinical ratings, the method described in [9] by VQEG to derive spearman rank order

correlation coefficient (SROCC [85]) and pearson linear correlation coefficient (LCC [85]) is used. SROCC evaluates the monotonicity of the computed VQA metrics with respect to the MOS clinical ratings provided by the two medical experts. LCC corresponds to the prediction accuracy. For the latter, VQA measurements denoted as O_j, $j = \{1,2,...,100\}$ are fitted beforehand to the clinical ratings provided by the medical experts denoted as MOS_j, $\{1,2,...,100\}$. To attain this, a nonlinear regression function as described in [85] is used:

$$O'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(O_j - \beta_3 \|\beta_4\|\right)}} \quad (4.1)$$

By using non linear least square optimization, optimum values for β that minimize the least square error between the clinical ratings MOS_j and the fitted objective scores O_j are derived. The initial parameter estimates are defined as follows [85]:

$$\beta_1 = max\{O_j\}$$

$$\beta_2 = min\{O_j\}$$

$$\beta_3 = \bar{O}_j$$

$$\beta_4 = 1$$

4.6 Material and Experimental Parameters

A total of ten ultrasound videos, nine of the carotid artery (three of the common carotid artery (CCA) and six of the internal carotid artery (ICA)), and one of the femoral artery compose our data set. The videos are depicted in Fig. 4.5 along with associated plaque and wall pixel-level segmentations. Each video consists of 6.5 seconds, sufficient for capturing several cardiac cycles. The videos were collected using the standardization protocol described in [80]. This ensures uniform visualization of the plaque morphology. To evaluate the visualization quality of the plaque type and the degree of stenosis, examples with a large diversity in the sizes of

the ROIs is sought. Of particular interest are the size of the plaque ROI and the size of the wall ROI (see columns 2 and 3 in Table 4.3, respectively). As depicted in Table 4.3 for CIF resolution encoding, strong variation for both the plaque ROIs (27-85 MBs) and wall ROIs (132-204 MBs) is obtained. Diagnostic ROIs range between 42-72% of the total picture size (column 3 in Table 4.3). Overall, larger bandwidth requirements from the larger ROIs are expected.

H.264/AVC source encoding involves a plethora of parameters. As described in subsection 2.1, different profiles and levels target different applications and transmission environments. According to the underlying application and available resources the appropriate encoding setting is selected. For the purpose of this study the baseline profile is considered given that the targeted telemedicine application involves medical video streaming to mobile devices. The following encoding setting, listing the most important parameters common to all simulations is incorporated: Predictive encoding using IPPP frame type encoding structure is considered. Bidirectional prediction is not supported in the baseline profile. The GOP parameter is set to 15. The latter is selected to match the transmitting frame rate of 15 fps as illustrated next in the results chapter. Intra-refreshing period is also considered at the beginning of each GOP. Given that the insertion of an I-frame completely stops error propagation, the key concept is to provide for close to error-free cardiac cycles. A total of 100 frames per video are encoded using UVLC encoding mode. Encoded medical video is encapsulated in RTP packet format. Packets are restricted to contain a maximum of 22(11) MB for CIF (QCIF) resolution. Simple frame copy error concealment method is applied at the decoder to reconstruct corrupted packets. Frame copy algorithm replaces current lost slice with previously displayed slice. One frame is used for

reference for increased encoding efficiency. The aforementioned parameters are
summarized in Table 4.4.

Table 4.3. Diagnostic regions of interest dimensions and overall picture portion (%).

No.	Plaque ROI (pixels,MB)	Wall ROI + opt ECG (MB)	Total ROIs (MB)	ROIs Video Size Percentage (%)
1. CCA #1	176x64 (44)	132+20=152	196	49.5
2. CCA #2	144x48 (27)	127+20=147	174	43.9
3. CCA #3	176x48 (33)	121+20=141	174	43.9
4. Femoral	192x64 (48)	194	242	61.1
5. ICA #1	160x48 (30)	212	242	61.1
6. ICA #2	160x64 (40)	202	242	61.1
7. ICA #3	192x64 (48)	150	198	50.3
8. ICA #4	272x80 (85)	135	220	55.6
9. ICA #5	240x64 (60)	138	198	50
10. ICA #6[a]	192x80 (60)	204+20=224	284	71.7

[a]Video ICA #6 is an outlier, given that the particular video is a close up on the atherosclerotic plaque region, and as a result diagnostic ROIs cover almost 72% of the whole video.

Table 4.4. Experimental setup encoding parameters.

Parameters	Value	Parameters	Value
Profile	Baseline	Level	3
FramesToBeEncoded	100	Resolution	CIF/QCIF
FrameRate	15 (5-30)[a]	Coding Structure	IPPP
IntraPeriod	15	SymbolMode	UVLC
Group of Pictures	15	OutFileMode	RTP
No. Reference Frames	1	No. of MB per Picture	396/99
FMO type used	2	No. of MB per Packet	22/11[d]
No. Slices per Picture	4(3[b])	SearchRange	32
Redundant Slices Rate	1 every 4 (1,2,4,8)[c]	Error Concealment	Frame Copy

[a]Frame rate is varied between 5-30 fps during experimentation
[b]When the ECG is not present the number of diagnostic ROIs and hence slices is set to three.
[c]Redundant Slices rate is considered between 1, 2, 4, and 8 during experimentation.
[d]22(CIF)/11(QCIF)

CCA #1 Predominantly Echogenic
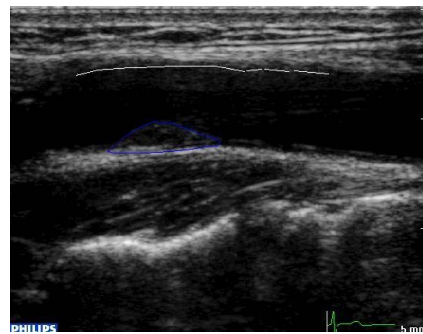
CCA #2 Predominantly Echolucent
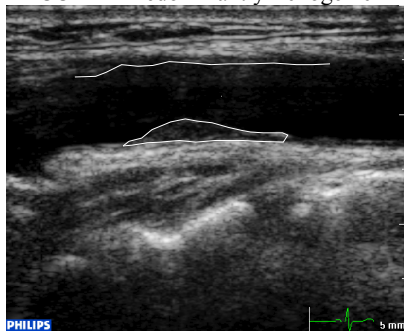
CCA #3 Echogenic

Femoral #1 Uniformly Echogenic

ICA #1 Predominantly Echolucent

ICA #2 Type 5
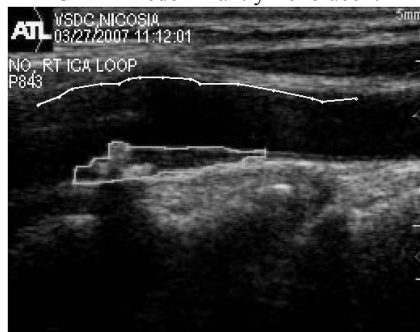
ICA #3 Predominantly Echolucent
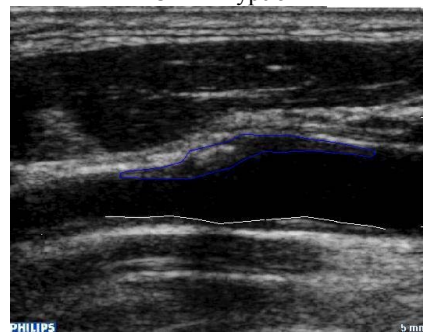
ICA #4 Mixed Plaque
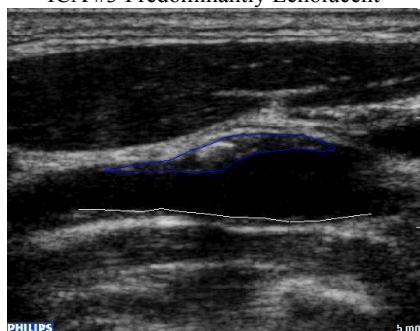
ICA #5 Mixed Plaque
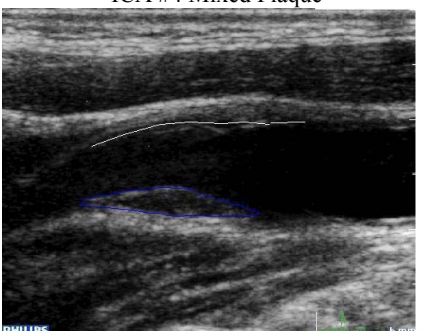
ICA #6 Predominantly echolucent

Fig. 4.5. Medical video data composed of 9 carotid artery and 1 femoral artery ultrasound videos and associated plaque and wall pixel-level segmentation.

Chapter 5

Results

This chapter presents a performance evaluation analysis of the proposed methodological framework. Individual approaches and methods are assessed and their corresponding efficiency is demonstrated through the use of extensive experimentation. Initial testing investigating the effect of bidirectional prediction encoding in noisy environments is depicted first. Bidirectional prediction is not supported in H.264/AVC baseline profile and therefore it was not considered further. Efficient intra-update utilization schemes are considered next so as to select the most suitable approach for limiting error propagation. A comprehensive evaluation of the proposed framework for medical video telemedicine systems follows.

A top-down summary of how minimum diagnostically acceptable parameters including resolution, frame rate, and redundant slices rate were obtained is presented. The experimental setup is then introduced based on these threshold values demonstrating efficiency near the diagnostic limit. This is followed by a thorough comparative evaluation of the proposed methods and the default H.264/AVC encoding. For clinical quality assessment both subjective (provided by medical experts) and objective evaluations are performed. The correlation between medical experts' mean opinion score and the number of employed objective metrics is depicted. Minimum bitrate requirements for diagnostically robust performance are then provided. In the last subsection, results of the ongoing work addressing medical video transmission over 3.5G mobile WiMAX networks are depicted.

5.1 Bidirectional prediction encoding and Intra-update schemes in noisy environments

5.1.1 Bidirectional prediction encoding

Using the JM 15.1 Reference Software, three different encoding schemes were evaluated. More specifically single-directional IPPP, and bi-directional IBPBP and IBBPBBP coding structures were considered. 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. Target bitrate and initial quantization parameter (QP) are the key input parameters for rate control encoding. 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 [86]:

$$QP = \begin{cases} 40 & bpp \leq 0.15 \\ 30 & 0.15 < bpp \leq 0.45 \\ 20 & 0.45 < bpp \leq 0.9 \\ 10 & bpp > 0.9 \quad R \end{cases} \tag{5.1}$$

where

$$bpp = \frac{R}{f x N_{pixels}} \tag{5.2}$$

Here, bpp denotes the target bits per pixel, N_{pixels} is the number of pixels in the Nth frame, R is the target bitrate and f the frame rate of encoded video sequence. Initial quantization parameters for QCIF and CIF resolutions are summarized in Table 5.1. The main encoding parameters are summarized in Table 5.2. Intra MB line update allows a fully intra coded frame every 11 and 18 frames for QCIF and CIF resolutions respectively.

Table 5.1. Target bitrate and initial quantization parameters.

Target Bitrate (QCIF)	Initial QP	Target Bitrate (CIF)	Initial QP
32 Kbps	40	96 Kbps	40
64 Kbps	40	256 Kbps	40
128 Kbps	30	512 Kbps	30
256 Kbps	30	800 Kbps	30
400 Kbps	20	1.2 Mbps	20
512 Kbps	20	1.5 Mbps	20

Table 5.2. Encoding 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[a]

[a]QCIF/CIF

Fig. 5.1 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.

Fig. 5.2 and Fig. 5.3 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 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.

The tested coding structures' performance was also evaluated by a medical expert so as to provide the level of diagnostic quality for plaque presence and degree of stenosis clinical criteria only. Plaque type assessment was not considered during this series of experiments. For bitrates of 128 kbps for QCIF resolution and 512 kbps 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 (5.1) and (5.2), see also Table 5.1) 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.

Above described simulations present initial investigations on the effect of bidirectional encoding. Given that the baseline profile only supports single-directional prediction, bidirectional prediction was not considered further in this thesis.

5.1.2 Intra-Updating encoding

An example of three different intra-update schemes is provided in Fig. 5.4. To measure error-resilience, the decoded video quality versus the packet loss rates (PLR) is plotted. In pre-defined and random intra refresh, 22 macro-blocks per frame are intra coded, giving a completely intra-coded frame every 18 frames (CIF resolution video). In intra-frame, an I-frame is encoded every 18 frames. The latter technique outperforms the former ones since error propagation is completely stopped at each I-frame, whereas with intra macroblock techniques error propagation is limited (but not stopped). On the other hand, note that I-frames require additional bandwidth, as demonstrated by the significant peaks in Fig. 5.4(b). When comparing video encoding at fixed bandwidths, the introduction of I-frames will translate into additional latency to be handled by the rate control algorithm.

For the purposes of this study, I-frame insertion is adopted. The refreshing interval is set to match the transmitting frame rate and encoding GOP. The aim is to completely stop error propagation approximately every cardiac cycle, to maximize the possibility of subsequent diagnostically lossless cardiac cycles. The clinical importance of a diagnostically-lossless cardiac cycle is discussed in subsection 5.3.

5.2 Coarse to Fine Parameter Optimization for Optimum Encoding Setting

For coarse QP optimization, QP values were varied between 20 and 40 for the atherosclerotic plaque region. Similarly, the sequence frame rate was varied between 5 and 30 frames per second (fps). In addition, the number of inserted RS and the corresponding trade off between bitrate, transmission time, and error resilience was investigated. The resolution was considered between quarter common intermediate format (QCIF-176x144 pixels, 11x9 MB) and common intermediate format (CIF-352x288 pixels, 22x18 MB). To reach optimum values, one investigated parameter

was varied while remaining parameters were kept constant. The medical experts then evaluated the diagnostic capacity of the corresponding compressed video. This procedure is depicted in Fig. 4.1.

The medical experts noted that for a QP of 28, clinical quality is preserved in the compressed video, carrying almost as much clinical data as the original. This led to the fine QP parameter optimization around this value. Thus, a selection of QPs of 28 and lower were found to qualify for clinical practice (see Table 5.5). Fifteen frames per second provided acceptable visualization of clinical motion, while clinical quality deteriorated significantly by less than 10 fps. The CIF resolution was selected since it provided quality visualization of the plaque morphology, as opposed to issues observed at the QCIF resolution. The diagnostic capacity associated with different resolutions and frame rates is summarized in Table 5.3. It is important to note here that CIF resolution video transmission was made possible due to the proposed variable quality slice encoding approach, where achieved bitrate reductions allow the transmission over 3G channels without compromising diagnostic quality. RS utilization was set to one RS every four coded slices. The latter was found to be a good balance of the trade-off between increased error resilience and decreased coding efficiency.

Having found an appropriate range of values for the plaque region (being the primary focus point of the clinical evaluation), slightly higher values (more quantization) were considered for the wall region, and a significantly higher value for the background. A low quantization value for the plaque region allows better visualization of the echogenic and echolucent areas that are needed for determining the plaque type. A slightly higher quantization is all that is needed for identifying the nearest wall boundary, for visualizing the stenosis. Most of the bandwidth savings

come from quantizing the background region. Here, it is essential to emphasize that the background regions still need to be visualized to depict the boundaries and relative motion differences between the plaque and the plaque components and other arterial regions. As depicted in Fig. 1.2, plaque components may have different motion patterns, assessment of which is crucial for predicting plaque rupture. Furthermore, quantization differences between the plaque and the wall video slices have helped direct clinical attention towards the plaque and its motion relative to the walls (hence stenosis).

Having concluded a diagnostically acceptable encoding setting including resolution, frame rate, quality indices for variable quality slice encoding, and rate of RS, the proposed setting was comprehensively evaluated including worst case transmission scenarios for extreme error values. The aim was to depict the validity of the undertaken methods, illustrate the corresponding efficiency, and investigate error resiliency in noisy channel conditions. To evaluate the proposed approach the following FMO encoding schemes were considered:

1) FMO type 2 where quality levels are equal for all involved diagnostic ROIs (uniform QP encoding throughout a frame). This encoding setting corresponds to the default H.264/AVC FMO type 2 utilization scheme and serves as the benchmark scheme for comparison purposes. Here, the objective is to show that the incorporated bitrate is a preventing factor for transmission over bitrate-limited 3G wireless channels. 3G is one of the key candidate technologies for medical video transmission upon deployment of the proposed system. Moreover, to depict that despite incorporating higher bitrates and uniformly encoded ROIs, the corresponding diagnostic performance is not superior to any of the proposed approaches.

2) FMO type 2 where quality levels are varied as a function of the diagnostic significance of the video regions. Quality indices are derived using the proposed system diagram depicted in Fig. 4.1, utilizing physician expert feedback. Given that the atherosclerotic plaque ROI corresponds to the most demanding clinical assessment of plaque type, the QP of the plaque ROI is the same for all considered approaches. The latter allows a fair comparison of the demonstrated schemes during both subjective and objective evaluations. The aim is to depict that similar diagnostic performance is attained when compared to the default H.264/AVC encoding scheme described in 1), at a significantly reduced bitrate. An example of QPAMap employing variable quality slice encoding appears in Fig. 4.4(b).

3) Similar to 2) but with the insertion of one redundant frame every four encoded frames. In this manner robust diagnostic performance for communications in noisy environments is sought, at a fixed transmission rate, by slightly increasing bitrate demands and transmission time.

For the rest of the thesis, the following acronyms are adopted: 1) FMO denotes the 1st considered approach corresponding to the default FMO type 2 encoding scheme, 2) FMO ROI corresponds to the 2nd considered approach incorporating variable quality slice encoding according to the ROIs diagnostic importance, and finally 3) FMO ROI RS is the proposed scheme also incorporating redundant slices.

For each method, three sets of quantization levels for the video slices were incorporated. For comparison, we set QP=24, 28, and 32 for the plaque video slice regions for all three cases. Then, for the constant QP case, all video slices are fixed to the plaque QP value. For variable space encoding, (a) low-bandwidth encoding using: QPs=40/34/32 for the background, wall and ECG, and plaque video slices

respectively, (b) medium bandwidth encoding with QPs=38/30/28, and (c) relatively high bandwidth encoding with QPs=36/26/24 is considered. In addition to the no packet loss case, at each quantization level, seven packet loss rates: 5%, 8%, 10%, 15%, 20%, 25%, and 30% were investigated.

Ten packet loss scenarios for each loss rate were simulated and averages were obtained. CIF and QCIF resolutions were considered. We thus had 3 methods x 2 resolutions x 3 quantization levels x 7 loss rates x 10 simulations per loss rate, for a total of 1260 video samples for each of the ten original videos (total=12,600 videos, see Table 5.4).

In the following subsections, all tables and figures describe CIF resolution encodings, which provide for clinical assessment of plaque type and is the proposed encoding resolution. QCIF experimentation is only depicted in Fig. 5.6 and Fig. 5.8 that demonstrate error resilient performance and bitrate demands reductions, respectively.

5.3 Clinical Video Quality Assessment

During the clinical evaluation, the videos were played back on a laptop at their original pixel size dimensions. According to Table 4.1, each diagnostic region received an independent evaluation score. Rating values were between 1 and 5 (see subsection 4.4.1 and Table 4.2 for a description of the scoring scale). As expected, in most cases, higher-quality diagnostic ROI encoding resulted in better clinical scores at lower bitrates (compared to the default FMO encoding). Inter-observer variability between the two medical experts was not extended to more than one point in the above-described clinical rating system. Here, Table 5.5 and Table 5.6, and Fig. 5.5 depict results obtained for the video depicted in Fig. 1.2 (a).

Table 5.3. Clinical evaluation criteria and associated encoding parameters.

	Clinical Significance	Clinical Differentiation for:	
		Display Resolution	Frame Rate
Plaque Boundary	Diagnose plaque(s) presence and plaque boundary	QCIF (176x144), CIF (352x288)	$\geq$5 fps
Stenosis	Estimate the degree of stenosis	QCIF (176x144), CIF (352x288)	$\geq$5 fps Recommended $\geq$10 fps
Plaque Type	Assess plaque morphology and plaque components and determine plaque type	CIF (352x288)	$\geq$10 fps Recommended 15 fps

Table 5.4. Total number of processed videos in this study.

		Instances	Cases
Method		FMO, FMO ROI, FMO ROI RS	3
Resolution		QCIF, CIF	x2
QP	FMO	32/32/32, 28/28/28, 24/24/24	x3
	FMO ROI and FMO ROI RS	40/34/32, 38/30/38, 36/26/24	
Packet Loss Rates (PLR)		5%, 8%, 10%, 15%, 20%, 25%, 30%	x7
Packet loss simulator		Uniform Distribution, averaging 10 runs per PLR	x10
Data Set		10 videos	x10
Total number of processed videos in this study			12600

Table 5.5 records the mean opinion score of two medical experts' ratings on the corresponding compressed video instances of the selected QPs range. The process of determining quality levels that do not compromise clinical information when compared to the original medical video is depicted. The medical experts provide individual ratings for each of the preset clinical criteria based on the available clinical data in the compressed video. The objective is to reach a quantization factor that enables the assessment of the same amount of clinical information in the compressed video as in the original video. Thus, a rating of 5 for all assessed clinical criteria for all videos parting the data set is sought. It is important to note here that while a rating of 4 corresponds to diagnostically lossless medical video, there still exists a loss of minor details. The latter is acceptable when evaluating transmitted medical videos. However, prior to transmission, any loss of clinical data is unacceptable, as it will lead to rapid video quality deterioration following transmission. For the higher quantization level (lower quality) of 32 for the plaque ROI, this is not attainable as depicted by the clinical ratings of 4.5 for the stenosis and plaque type clinical criteria. At such qualities, only plaque presence can be determined as in the original video.

Overall, a selection of ROI QPs of 28 and lower were found to qualify for clinical practice as evident by the clinical ratings of 5. Higher QPs may be selected for urgent clinical practice with respect to bandwidth availability. Having said this, caution must be exercised against the use of higher quantization parameter values in ordinary practice. As an example of what can go wrong, it is noted that in one of the cases (Fig. 1.2(c)), an ulcer on the plaque that was still visible for ROI QP of 32, was not visible for ROI QP of 36. The same observation was true for the video in QCIF resolution.

In Fig. 5.5, the efficiency of the proposed method of spatially varying QPs selection relating to video portion's diagnostic capacity and associated bandwidth

demands reduction is depicted. Results are provided for the objective (luma) Y-PSNR evaluation measurements over the plaque ROI. PSNR was selected for results illustration purposes as despite its documented inefficiencies, it is still the most widely used metric for video quality evaluation, and therefore provides the common ground for comparison with other studies in the literature. As evident in Table 5.5, diagnostic yield of the proposed FMO ROI and FMO ROI RS schemes is not compromised by the variable quality space encoding. In the contrary, efficient quality resources allocation, enables considerable bandwidth savings. Recall that plaque ROI-slices are encoded with equal quality levels in all competing approaches. For this particular video, bitrate demands reductions of the proposed FMO ROI RS scheme are 46%, 48%, and 46% for ROI QPs of 32, 28, and 24 respectively, when compared to the default FMO encoding scheme. Clinical evaluations for PLR up to 15% for plaque ROI QP of 28 are given in Table 5.6. Table 5.6 demonstrates the error resilience of the scheme incorporating RS, even if channel conditions introduce 15% error on the transmitted stream. Furthermore, it depicts the similar behaviour as to video quality degradation of the compared approaches that do not utilize RS. For all investigated PLR scenarios, FMO ROI RS scheme attains MOS ratings greater than or equal to 4, which correspond to a diagnostically lossless performance. Hence, effective utilization of RS error resilience technique allows clinical diagnosis in noisy environments. On the other hand, FMO and FMO ROI ratings fall below of what is diagnostically acceptable. A rating of 3.5 for 5% and 10% PLR for FMO, and for 8% and 10% PLR for FMO ROI is observed, for the plaque type criterion assessment. As already documented, plaque type assessment is much more demanding than identifying plaque presence and estimating the degree of stenosis estimation.

It's worth noting here an initially surprising finding. For PLR of 15%, the scheme incorporating RS attains higher ratings than when compared to lower PLR of 10%. The same observation is also true for FMO and FMO ROI approaches where videos experiencing higher PLR are rated with higher evaluation scores. This detail reveals an important aspect associated with the clinical evaluation of medical videos. Consecutive diagnostically lossless cardiac cycles may prove sufficient for the physician to reach a diagnosis even when the objective VQA ratings suggest the opposite. This is of course medical video modality specific. However, it is an important aspect that needs to be taken into consideration as it dictates the need of designing new, diagnostically driven objective VQA algorithms.

An objective evaluation shown in Fig. 5.6 (QCIF resolution) and Fig. 5.7 (CIF resolution), verifies the aforementioned clinical evaluation observations. Boxplots demonstrating PSNR VQA measurements for PLR up to 30% for the whole data set are provided. In each box, the central mark represents the median, the edges represent the 25th and 75th percentiles, and the whiskers extend to the most extreme values. Beyond outliers, extreme points are plotted with hollow circles. The dotted lines correspond to the median values of each block. The error resiliency of the proposed scheme employing RS for plaque ROI QP of 28 is depicted. FMO ROI RS achieves graceful degradation of the video quality and provides for clinical practice even when transmitting channel introduces 15% of packet losses. FMO and FMO ROI approaches suffer severe quality degradation. At 5% PLR, FMO and FMO ROI schemes receive lower ratings than the proposed FMO ROI RS scheme at 15%.

An important aspect, evident by observing Fig. 5.6 and Fig. 5.7 for the no packet losses scenario, is the lower quality levels associated with the quantization factor of 28 for the QCIF resolution. As a result, PSNR values for corresponding PLR scenarios

are also lower when compared to the proposed CIF resolution. From a clinical perspective, QCIF resolution allows identifying plaque presence and estimating the degree of stenosis. Plaque type cannot be efficiently assessed in all cases as noted by the medical experts in subsection 5.2 (see Table 5.3). Thus, given that plaque type assessment is not considered, these values provide for accurate diagnosis of the plaque presence and degree of stenosis. In every case, higher qualities of QP values of 36/26/24 may be selected. Bitrate demands associated with the aforementioned quality levels in QCIF resolution is not a preventing factor as illustrated in Fig. 5.8.

5.4 Objective VQA and Correlation to Clinical Evaluations

For all cases, Table 5.7 summarizes the correlations between the plaque ROI-based VQA algorithms and the MOS from two clinical experts, for clinical ratings of plaque type. SROCC and LCC were deducted by fitting the VQA ratings to a representative sample of 100 video instances. The best results were obtained by ROI-WSNR with an LCC of 0.690 and an SROCC of 0.715. ROI PSNR, SSIM, and VIF algorithms attained scores higher than 0.5.

The goal here is to provide means of evaluating objective VQA performance with respect to medical video's actual diagnostic capacity. Ultimately, for consistently high correlations between the physicians' evaluations and the computerized measurements, an automated procedure for predicting subjective scores from objective ratings is envisioned. While still at the early stages of investigation, associating diagnostic capacity of encoded video to specific regions of clinical importance, and hence restricting correlation investigation to what is diagnostically relevant, serves as a starting point. For example, high quality reconstruction of the plaque ROI slice is likely to provide for high MOS ratings for the assessment of plaque type. This follows

from the fact that the specific video portion contains all the clinical data assessed by the medical expert when determining plaque type. As already documented, the same allegation is true for plaque presence and degree of stenosis assessment, as plaque type is the primary focus point of the clinical evaluation (see also Table 4.1). Similarly, high correlation between MOS and objective VQA for plaque type criterion is likely to account for high MOS for plaque presence and the degree of stenosis criteria. On the other hand, high quality rendering of the background regions would not contribute to enhanced diagnostic performance. Analogous approaches should apply to different medical modalities.

5.5 Minimum Bitrate Requirements

In Table 5.8, bitrate savings achieved by the proposed FMO ROI RS scheme when compared to the default H.264/AVC FMO scheme are presented, for CIF resolution encoding at 15 fps. Right columns of Table 5.8 incorporate the bitrate gains for each of the three quantization levels sets for all videos in the data set. Fig. 5.8 and Fig. 5.9 depict the required bitrates of all three investigated encoding methods for QCIF and CIR resolutions, respectively.

Bitrate reductions are computed by comparing FMO QPs of 32/32/32, 28/28/28, and 24/24/24 with FMO ROI RS QPs of 40/34/32, 38/30/28, and 36/26/24 respectively. As documented in Table 5.8, significant bitrate demand deductions ranging between 15%-60% (average 42%) is achieved, for all investigated methods and videos, without compromising diagnostic quality. It is important to note that the lower bitrate savings (15%-27%) are observed for the femoral ultrasound video. This is not medical video modality specific, rather it is associated with the slightly different acquisition parameters incorporated for the particular ultrasound video (see Fig. 4.5).

Bandwidth savings are a function of the incorporated plaque sizes and diagnostic ROIs. Hence, the variety of the depicted bandwidth savings is due to the diversity of the considered plaque sizes and diagnostic ROIs. The latter are given in Table 4.3. Diagnostic ROI sizes (total) range from 44%- 72%, while atherosclerotic plaque ROI sizes are considered between 7%-21%.

Fig. 5.8 and Fig. 5.9 also depict the slightly increased bitrate associated with the use of RS. This is observed by comparing the variable quality slice encoding schemes for the same quantization factors (Fig. 5.9: QP4-QP6). The introduction of packet losses produces significant drops in video quality. The drop in video quality more than justifies the overhead of introducing redundant slices. To see this, the example in Fig. 5.7 is re-examined. At 15% PLR, the use of redundant slices keeps the clinical video quality at an acceptable level (>35dB), while all other methods drop below of what is acceptable. In fact, there is a 5 dB drop in quality for FMO and FMO ROI schemes that do not use redundant slices. From Fig. 5.5, it is clear that both FMO and FMO ROI methods cannot match this performance without a huge increase in bandwidth (which would be off the charts of Fig. 5.5).

Next, the candidate wireless transmission channels matching the proposed framework's achieved bitrates are highlighted. To see this, typical upload data rates of today's wireless networks summarized in Table 2.3 are examined. In the absence of a higher-bandwidth channel, QCIF resolution may provide for initial diagnosis purposes. It can also provide for medical video streaming to mobile devices with limited displaying capabilities. Moreover, it can be used by the medical expert to define the diagnostic ROIs before switching to CIF resolution for clinical assessment of plaque type. Overall, as demonstrated in Fig. 5.8, quantization levels of 40/34/32, 38/30/28, and 36/26/24 provide for very low GPRS bandwidths (2.5G), today's typical

EDGE upload date rates (2.5G-2.75G), and are well within the typical 3G rates, respectively. For the proposed CIF resolution encoding depicted in Fig. 5.9, quantization levels of 40/34/32, 38/30/28, and 36/26/24 match typical available upload data rates of 2.5G (EDGE), 3G, and 3.5G of mobile telecommunication networks respectively. The associated bitrate reductions of the proposed diagnostically relevant encoding enable CIF resolution transmission over typical 3G upload rates, otherwise not feasible with conventional uniform encoding. Currently, network operators in Cyprus currently offer theoretical upload data rates at 384 kbps and 2 Mbps for two different 3.5G modems. If the channel data rate input parameter is limited (see Fig. 4.1), a further quantization of the background may be required for seamless streaming.

5.6 Proposed Framework Evaluation over 3.5G mobile WiMAX Networks

Medical video transmission over emerging 3.5G wireless networks is depicted next. Using OPNET modeller 16.0 network simulator [87], the proposed framework's recommended encoding setting is evaluated over 3.5G mobile WiMAX networks. The objective here is twofold. The first objective is to depict that the incorporated approaches and methods described in the proposed unified framework are applicable to different wireless networks. The second objective is to demonstrate the benefits associated with the adoption of next generation wireless networks in the design of new m-health systems and services, especially for investigating the diagnostic capacity associated with higher resolution encodings.

The scenario illustrated in Fig. 5.10 represents a typical scenario where medical video is transmitted from the ambulance to the hospital premises. The basic scenario parameters are summarized in Table 5.10. Video traces for the video ICA#6 depicted in Fig. 4.5 encoded with the proposed system setting summarized in Table 5.9 is used

for a more realistic video content transmission. The video is looped over until the end of the simulation. The ambulance travels with speeds ranging from 30-50 km/h and traverses through 6 base stations (BS) until it reaches to the hospital premises. Different path loss models and more specifically vehicular [88], suburban [89], and free space environments are considered for transmission over 3.5G wireless networks using mobile WiMAX infrastructure. For the vehicular scenario, multipath fading is also considered. The latter scenario corresponds to the more realistic setting towards real-time implementations. The obtained results are averaged over 10 simulations runs for each scenario.

Fig. 5.11 demonstrates instantaneous packet loss rates and end-to-end delay for 4CIF resolution medical video. As evident by observing Fig. 5.11, low delays for all investigated path loss models are achieved. More specifically, the observed delays are in the order of 60 ms for all considered scenarios, dropping to less than 20 ms when the transmitting mobile station (ambulance) enters the same BS serving the receiving station (hospital). This is due to the 50 ms delay manually introduced as internet delay, not related to the mobile WiMAX infrastructure. The spikes observed are caused when the mobile stations leaves the effective coverage zone of one BS and enters the next serving BS. Handover related delays and associated possible service interruption(s) are outside the scope of this work. It is important to note however that for the free space signal propagation model scenario, the increased spike delay is due to the late handover execution time compared to vehicular and suburban scenarios. In any case, consistently delayed packets will be considered as dropped.

Table 5.11 summarizes average QoS measurements for all investigated resolutions and path loss propagation models. Depicted end-to-end delay is well within the range of less than 300 ms required (preferably less than 100ms) for seamless transmission of

medical video [37], [64], [90], and within the range of typical mobile WiMAX delays of less than 70 ms as documented in Table 2.3. Given that 50 ms is the manually introduced internet delay, the actual latency is less than 20 ms for all investigated scenarios (besides the QCIF resolution scenario using free space path loss propagation model, which is in the order of 30 ms). Delay jitter is very limited and is in the order of 1 to 2 ms (recall that dropped packets are not considered during delay jitter computation and that the specific scenario has no background traffic).

Average packet loss rates are within the attainable diagnostically lossless threshold of 15% PLR of the proposed framework using error resilient encoding. Average PLR are less than 10% for all investigated scenarios. However, packet losses are not uniformly distributed and therefore conclusions cannot be deducted before clinical evaluation. Still, the limited PLR experienced when the ambulance is accommodated within the effective coverage zone of the mobile WiMAX BS, ensures reliable medical video delivery. Moreover, effective transmission of higher, 4CIF resolution encoding is demonstrated. The benefits in terms of diagnostic capacity associated with 4CIF resolution transmission, are expected to significantly advance remote diagnosis and care. During 4CIF resolution transmission, 1024 subcarriers are used, hence doubling the network's capacity when compared to the 512 subcarriers considered for QCIF and CIF resolutions (see Table 5.10). Future experimentation involves more complicated scenarios including heavy loaded cells, with increased cell radius, and higher mobility [91].
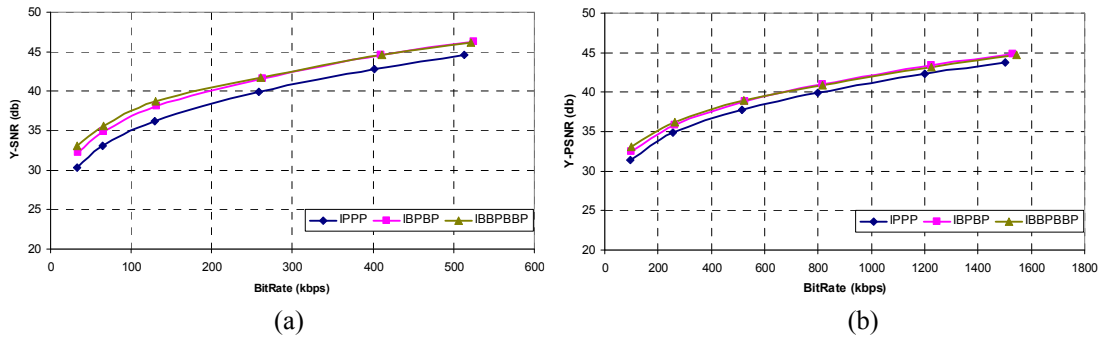
Fig. 5.1. Rate-distortion curves for tested frame encoding schemes. (a) QCIF and (b) CIF.
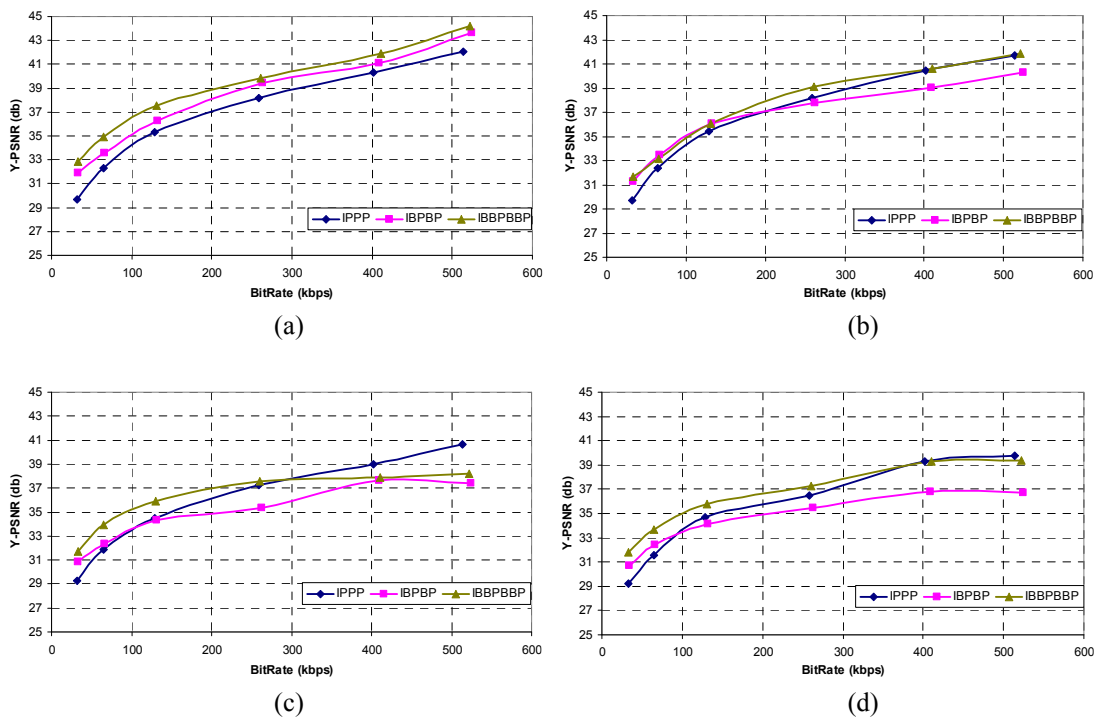
Fig. 5.2. Rate-distortion curves for tested frame encoding schemes, QCIF resolution.
(a) 2%, (b) 5%, (c) 8% and (d) 10% loss rates. IBBPBBP encoding scheme attains higher PSNR ratings in most cases, especially in low-noise (up to 5%) scenarios.
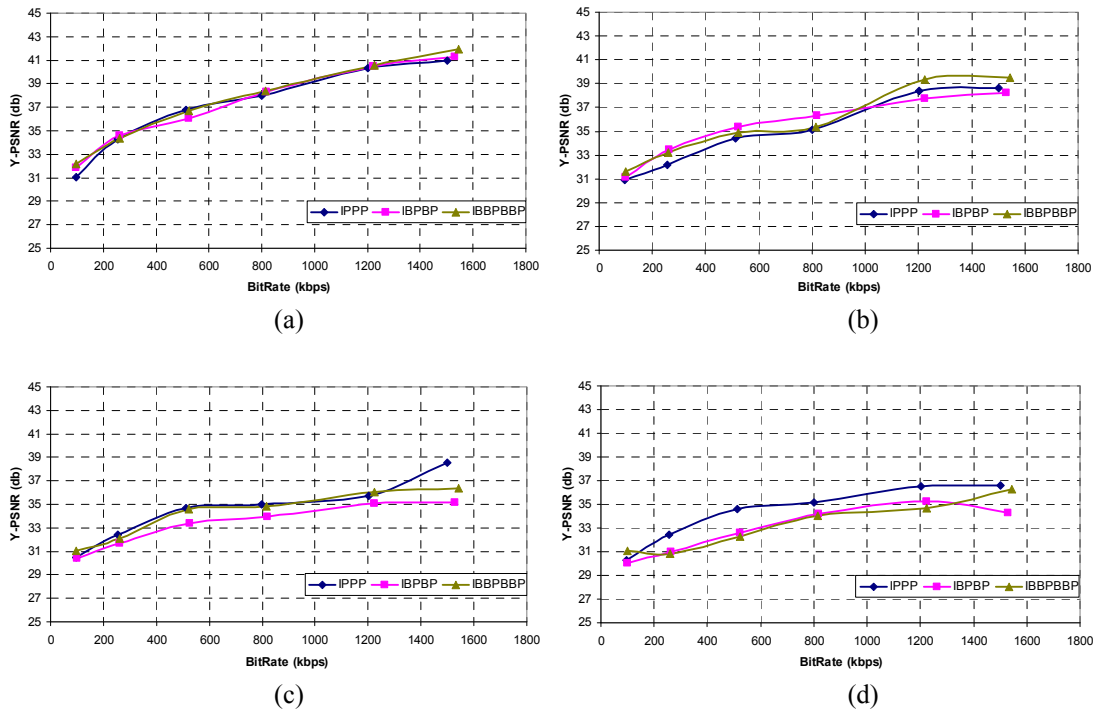
Fig. 5.3. Rate-distortion curves for tested frame encoding schemes, CIF resolution. (a) 2%, (b) 5%, (c) 8% and (d) 10% loss rates. Bi-directional prediction (IBPBP and IBBPBBP) achieves better results up to 5% loss rates (low-noise), whereas as the noise level increases, single – directional (IPPP) provides for better error recovery.

Fig. 5.4. Video Compression example using intra updating.
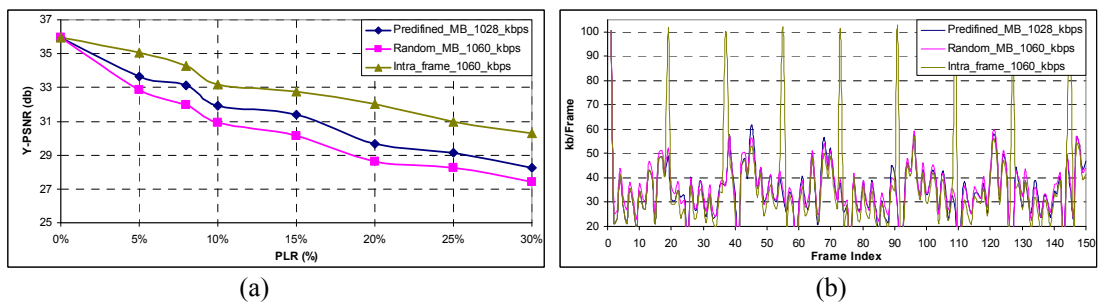
Video Compression example using intra updating. Three intra-update schemes are demonstrated: predefined macro block (MB) intra update, random intra updated, and intra frame update vs packet loss rate (PLR). (a) PSNR quality versus PLR. (b) Single-frame encoding requirements for each method. Rate control is used to smooth variable frame rate requirements (not discussed here). Here, carotid ultrasound video is used.

Table 5.5. Clinical evaluation mean opinion score for determining diagnostically lossless QP.
The results are for the video depicted in Fig. 1.2 (a) and CIF resolution encoding at 15 fps.

FMO	*QP*	32/32/**32**	28/28/**28**	24/24/**24**
	BitRate (kbps)	**416**	**788**	**1295**
Plaque Boundary		5	5	5
Stenosis		4.5	5	5
Plaque Type		4.5	5	5
FMO ROI	*QP*	40/34/**32**	38/30/**28**	36/26/**24**
	BitRate (kbps)	**191**	**355**	**625**
Plaque Boundary		5	5	5
Stenosis		4.5	5	5
Plaque Type		4.5	5	5
FMO ROI RS	*QP*	40/34/**32**	38/30/**28**	36/26/**24**
	BitRate (kbps)	**225**	**411**	**700**
Plaque Boundary		5	5	5
Stenosis		4.5	5	5
Plaque Type		4.5	5	5

1: Lowest Quality, 5: Highest Quality
QPs are given in the order of background/wall ROI/plaque ROI.
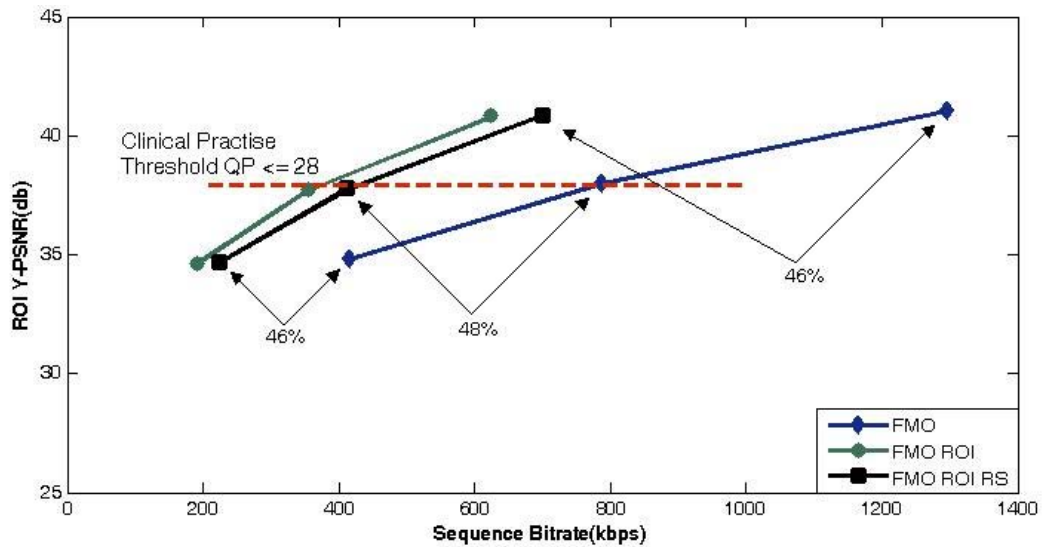
Fig. 5.5. Rate-distortion curves demonstrating compression efficiency near the diagnostic limit.
All three methods are shown for the video depicted in Fig. 1.2 (a) and CIF resolution encoding at 15 fps. Here, FMO stands for uniform QP encoding, FMO ROI denotes the use of variable QP slice encoding, and FMO ROI RS also uses redundant slices. The distortion is measured in terms of the PSNR for the plaque ROI. The key point is the significantly reduced sequence bitrate without compromising clinical quality (verified by Table 5.5). Indicatively, for this particular video, FMO ROI RS requires 46%, 48% and 46% less bitrates than conventional FMO for QPs of 32, 28, and 24 respectively (see text for variable QP parameters). Note that the clinical practice threshold of 35dB or QP $\leq$ 28 is independent of the video.

Table 5.6. Clinical evaluation for the proposed plaque ROI QP of 28 in noisy channels.
The results are for the video depicted in Fig. 1.2 (a) and CIF resolution encoding at 15 fps.

	FMO	FMO ROI	FMO ROI RS
QP	28/28/**28** [a]	38/30/**28** [b]	38/30/**28** [b]
BitRate(kbps)	**788**	**355**	**411**
Loss Rates %	**5/ 8/ 10/ 15**	**5/ 8/ 10/ 15**	**5/ 8/ 10/ 15**
Plaque Boundary	4/4.5/4.5/4.5	4.5/4.5/4.5/4	5/ 5/ 4.5/ 5
Stenosis	4/4.5/4.5/4	4.5/4.5/4.5/4	5/ 5/ 4.5/ 5
Plaque Type	3.5/4/3.5/ 4	4/ 3.5/ 3.5/ 4	4.5/ 4.5/ 4/ 4.5

1: Lowest Score, 5: Highest Score

FMO, FMO ROI and FMO ROI RS is used for constant QP FMO encoding, variable QP FMO encoding and variable QP FMO with RS respectively.

[a] *28:background/ 28:wall ROI/ 28:plaque ROI.*

[b] *38:background/ 30:wall ROI/ 28:plaque ROI.*

Table 5.7. Comparison of the performance of the VQA algorithms for pearson and spearman correlations.

ROI/Correlation	PSNR	SSIM	VSNR	VIF	VIFP	IFC	NQM	WSNR
LCC	0.674	0.630	0.469	0.483	0.605	0.386	0.486	0.690
SROCC	0.667	0.646	0.465	0.468	0.589	0.363	0.480	0.715
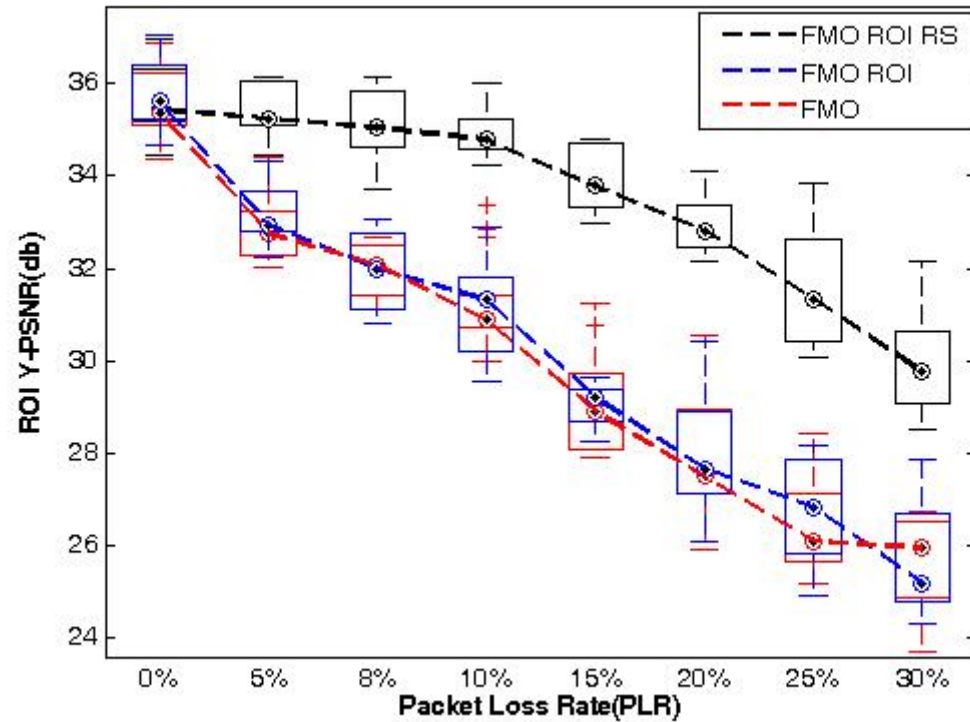
Fig. 5.6. Quality Evaluation for Error Prone Channels (QCIF resolution).

Here, the PSNR vs Packet Loss Rate curve for the plaque ROI QP of 28 is evaluated, by providing box plots for the whole data set (QCIF resolution). FMO ROI RS achieves graceful degradation of video quality in the presence of severe loss rates, qualifying for clinical practice even at 15% loss rate. For QCIF resolution, clinical diagnosis involves identifying plaque presence and the degree of stenosis. Plaque type assessment is not always possible as discussed in subsection 5.2. FMO ROI and FMO suffer severe degradation, as evident by the low PSNR values. In each box, the central mark represents the median, the edges represent the 25th and 75th percentiles, and the whiskers extend to the most extreme values. Beyond outliers, extreme points are plotted with hollow circles.
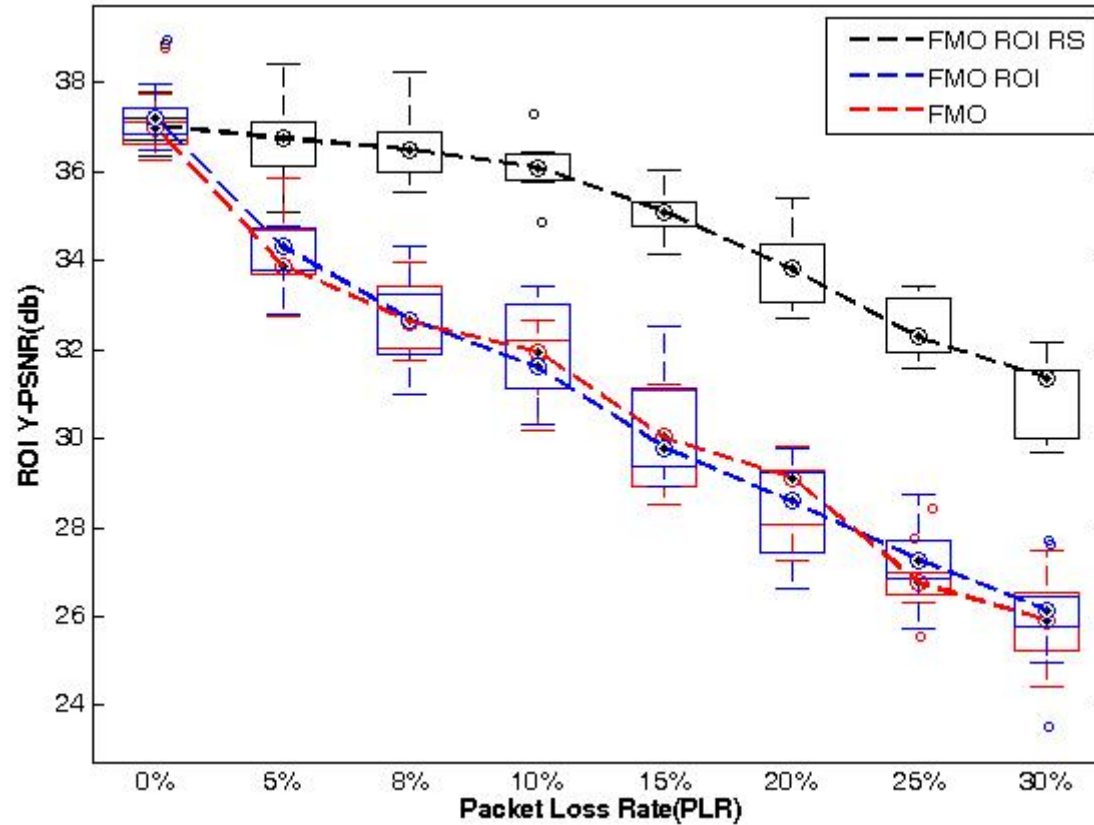
Fig. 5.7. Quality evaluation for error-prone channels (CIF resolution).

Here, the PSNR vs Packet Loss Rate curve for the plaque ROI QP of 28 is evaluated, by providing box plots for the whole data set (CIF resolution). FMO ROI RS achieves graceful degradation of video quality in the presence of severe loss rates, qualifying for clinical practice even at 15% loss rate. FMO ROI and FMO suffer severe degradation, as evident by the low PSNR values. Bandwidth requirements reductions are presented in Table 5.8. In each box, the central mark represents the median, the edges represent the 25th and 75th percentiles, and the whiskers extend to the most extreme values. Beyond outliers, extreme points are plotted with hollow circles.

Table 5.8. Diagnostic regions of interest dimensions and corresponding bitrate savings (%) for CIF encoding at 15 fps.

| No.	ROIs Video Size Percentage (%)	% BitRate Savings for FMO ROI RS vs FMO		
		40/34/32 vs 32/32/32	38/30/28 vs 28/28/28	36/26/24 vs 24/24/24
1. CCA #1	49.5	30	38	43
2. CCA #2	43.9	45	55	58
3. CCA #3	43.9	55	60	58
5. Femoral	61.1	15	21	27
4. ICA #1	61.1	34	40	41
6. ICA #2	61.1	58	59	55
7. ICA #3	50.3	30	36	37
8. ICA #4	55.6	46	48	46
9. ICA #5	50	47	51	49
10. ICA #6[a]	71.7	27	31	30

[a]Video ICA #6 is an outlier, given that the particular video is a close up on the atherosclerotic plaque region, and as a result diagnostic ROIs cover almost 72% of the whole video.
** BitRate savings deducted by comparing FMO QPs of 32/32/32, 28/28/28, and 24/24/24 vs FMO ROI RS QPs of 40/34/32, 38/30/28, and 36/26/24.
*** See Fig. 4.5 for medical video data set and associated plaque and wall ROIs segmentations.

Table 5.9. Minimum proposed settings for atherosclerotic plaque ultrasound video wireless transmission in noisy[a] 3G channels.

Parameter	Value
Encoding Standard	H.264/AVC
Profile	Baseline
Error Resilience	FMO and RS
Resolution	CIF
Frame Rate	15 fps ($\geq$10 fps)
RS	1 every 4 slices
FMO ROI RS QP	38/30/28
Plaque ROI PSNR	$\geq$ 35db
BitRate	Typical 3G and beyond rates (see Fig. 5.9)

[a]Packet loss rates up to 15%
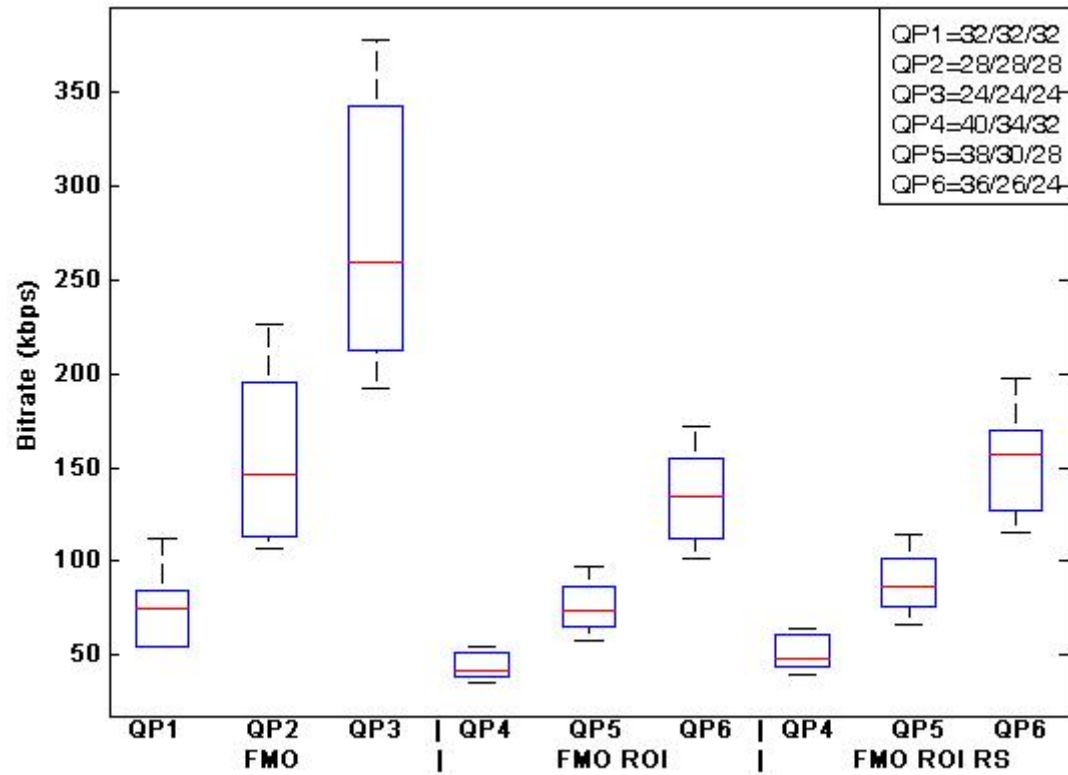
Fig. 5.8. Box plots demonstrating bitrate requirements of the compared schemes for the 9 regular videos of the data set (QCIF resolution).
Here, the last case, in which the video represents a close-up on the plaque is considered an outlier (see Table 5.8 for details). Overall, quantization levels of 40/34/32, 38/30/28, and 36/26/24 provide for very low GPRS bandwidths (2.5G), today's typical EDGE upload date rates (2.5G-2.75G), and are well within the typical 3G rates, respectively. In each plot we display the median, lower, and upper quartiles and confidence interval around the median. Straight lines connect the nearest observations within 1.5 of the IQR of the lower and upper quartiles. The '+' sign indicates possible outliers with values beyond the ends of the 1.5 x IQR.
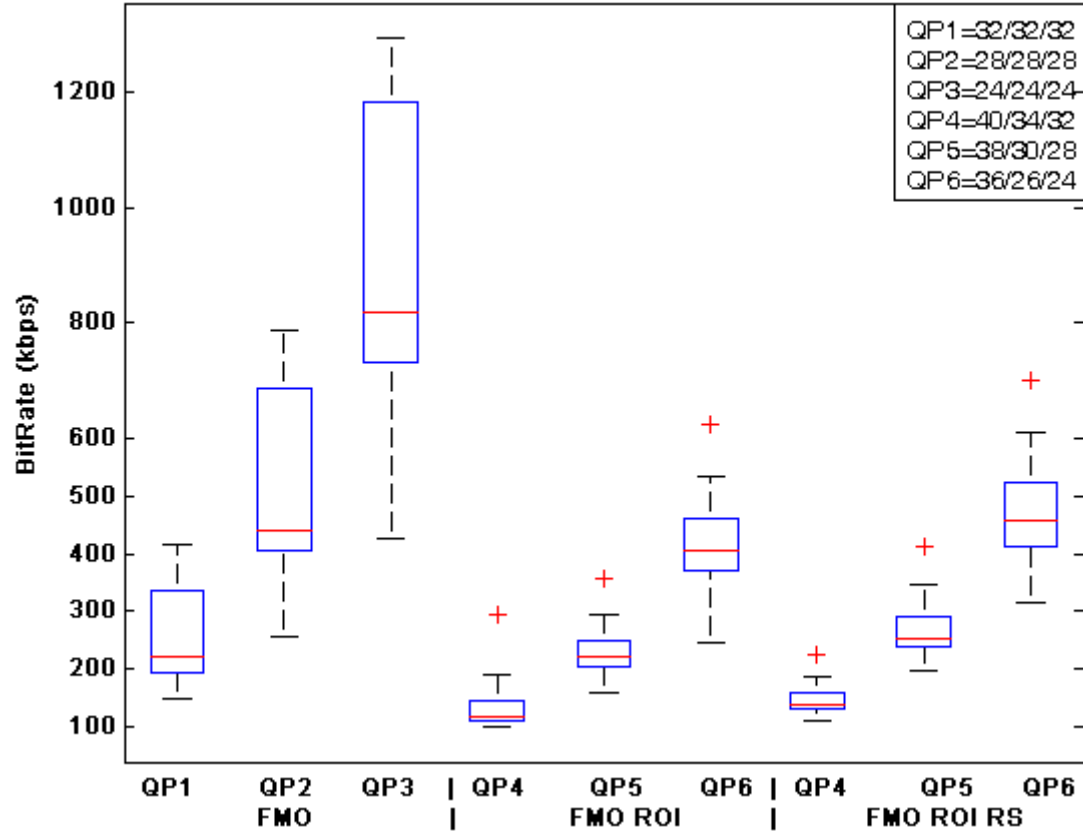
Fig. 5.9. Box plots demonstrating bitrate requirements of the compared schemes for the 9 regular videos of the data set (CIF resolution).
Here, the last case ICA #6, in which the video represents a close-up on the plaque is considered an outlier (see Table 5.8 for details). We observe that lower quality 40/34/32 (QP4) may be transmitted over 2.5G of mobile communication networks, the recommended case of 38/30/28 (QP5) is well within the typical 3G data rates, while the highest quality of 36/26/24 (QP6) is appropriate for 3.5G networks.. In each plot we display the median, lower, and upper quartiles and confidence interval around the median. Straight lines connect the nearest observations within 1.5 of the IQR of the lower and upper quartiles. The '+' sign indicates possible outliers with values beyond the ends of the 1.5 x IQR.

Table 5.10. Mobile WiMAX Configuration Setting.

WiMAX Configuration Parameter	Value
Access Technology	OFDMA
Base Frequency	5.8 GHz
Number of Subcarriers	512(QCIF, CIF) / 1024(4CIF)
Subcarrier Frequency Spacing	10.9375 KHz
Frame Duration	5ms
Symbol Duration	100.8
Duplexing Technique	TDD
Total Capacity DL/UL	2.88 / 0.576 Msps (512 subcarriers)
(mega symbols per seconds)	5.875 /1.267 Msps (1024 subcarriers)
Efficiency Mode	Mobility and Ranging Enabled
Internet Delay	50 ms
Base Station Configuration Parameter	Value
Number Transmit/Receive Antennas	1
Antenna Gain	15 dBi
Maximum Transmission Power	2 W
Effective Cell Coverage	500 m
Subscriber Station Configuration Parameter	Value
Number Transmit/Receive Antennas	1
Maximum Transmission Power	0.5 W
Pathloss Parameter	Vehicular/Suburban/Free Space
MAC Layer QoS Class	Real time polling service (rtps)
Minimum Sustained	128 kbps (QCIF), 500 kbps (CIF),
Data Rate	1.5 Mbps (4CIF)
ARQ/hARQ	Disabled
Modulation and Coding	64-QAM ¾
Medical Video Bitrate	100 kbps (QCIF), 400 kbps (CIF), 1.4 Mbps (4CIF)
Mobility	30-50 Km/h

Table 5.11. Average packet loss rates, end-to-end delay, and delay jitter for different Mobile WiMAX signal propagation scenarios and display resolutions.

Resolution	Path loss Model	Average PLR (%)	PLR Standard Deviation	Average Delay (ms)	Average Delay Jitter (ms)
QCIF	Free Space	5.75	0.28	82	<2
	Suburban	2.84	0.35	66	<1
	Vehicular	4.92	5.57	67	<1
CIF	Free Space	5.03	0.68	68	1
	Suburban	4.23	4.63	54	<1
	Vehicular	8.65	8.16	56	<1
4CIF	Free Space	5.34	0.34	59	<1
	Suburban	4.67	2.76	52	<1
	Vehicular	3.31	0.59	54	<1
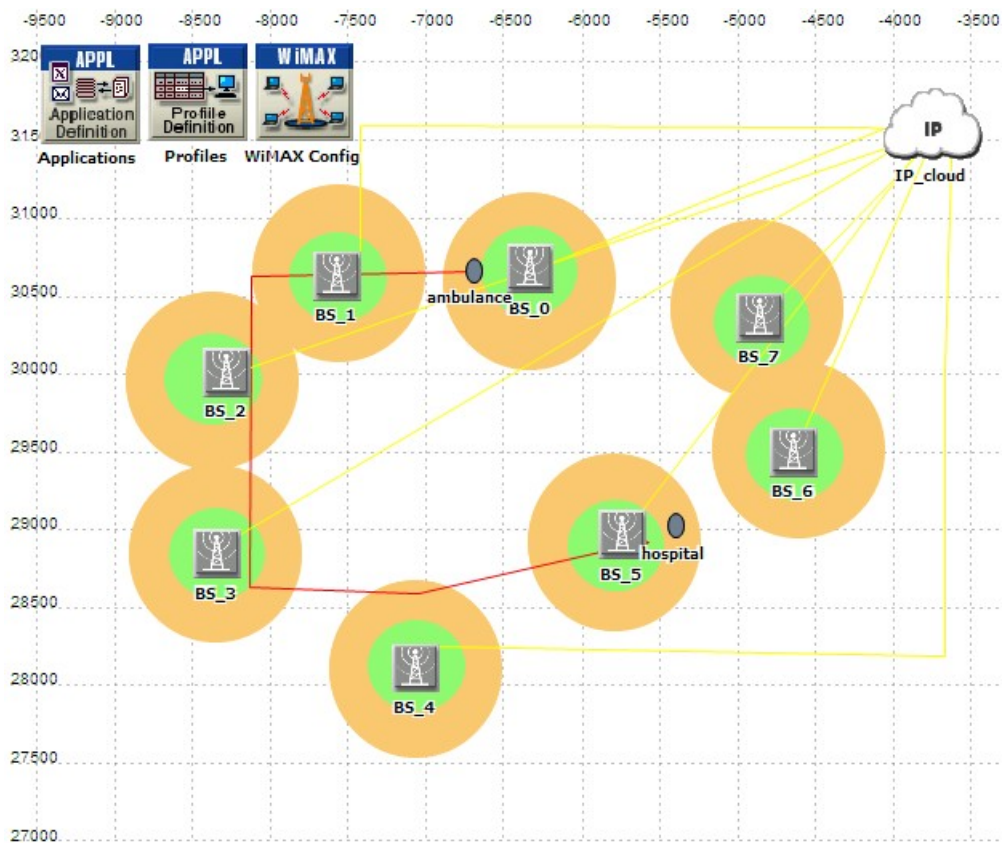
Fig. 5.10. Example topology for medical video transmission over 3.5G Mobile WiMAX using OPNET modeller.

A typical example topology for simulating medical video transmission over Mobile WiMAX. The ambulance travels with speeds ranging from 30-50 km/h and traverses through 6 base stations (BS) until it reaches to the hospital premises. Different path loss models are examined. Video traces for the video ICA#6 depicted in Fig. 4.5 is used for a more realistic video content transmission. Configuration parameters are depicted in Table 5.10.

(a)

(b)

(c)

(d)

(e)

(f)
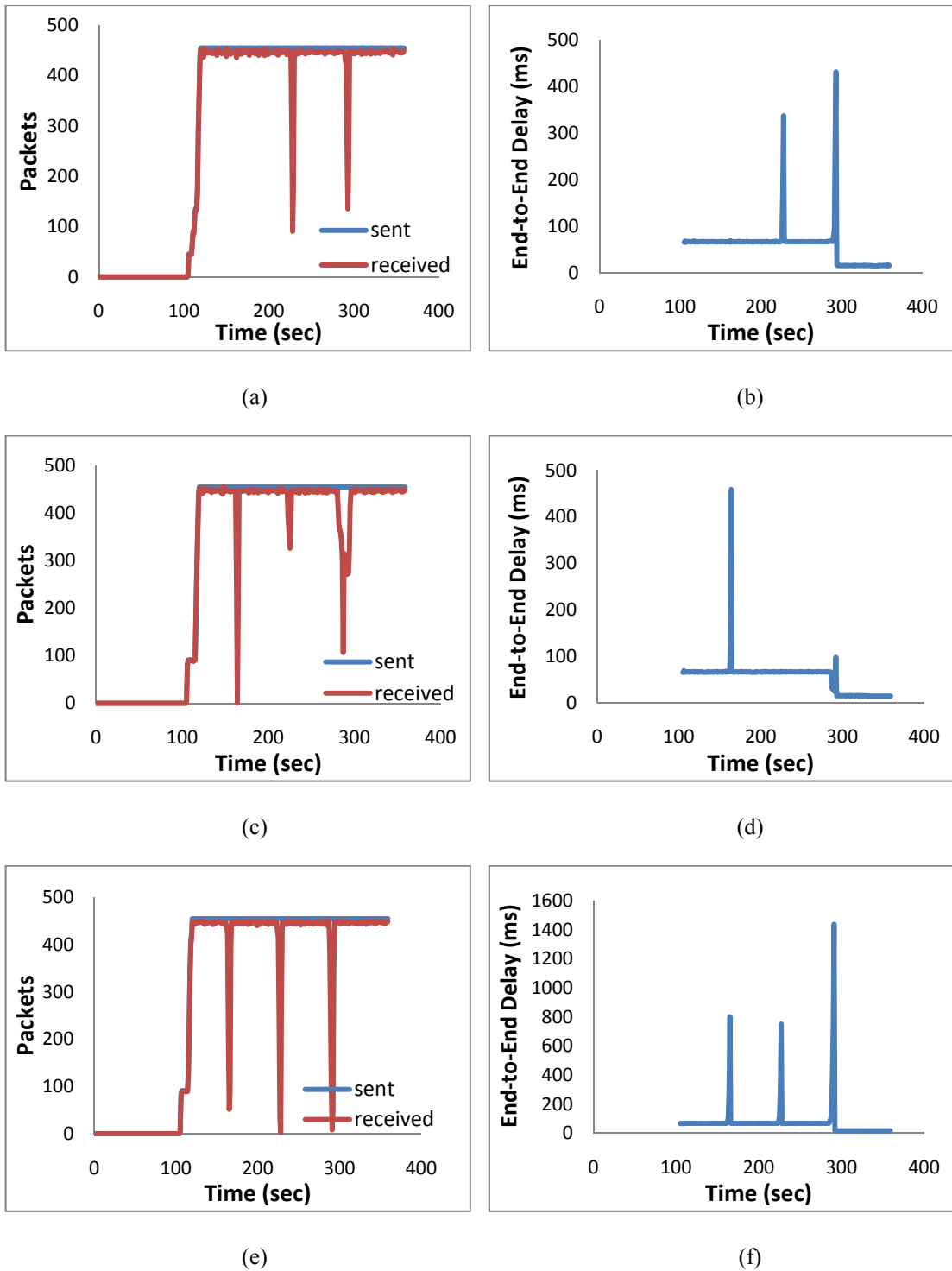
Fig. 5.11. 4CIF resolution medical video transmission over 3.5G Mobile WiMAX wireless infrastructure.

(a) and (b): Packet loss rates and end-to-end delay for vehicular path loss propagation and multipath fading according to ITU vehicular A specification [88].

(c) and (d): Packet loss rates and end-to-end delay for suburban path loss propagation according to [89].

(e) and (f): Packet loss rates and end-to-end delay for free space path loss propagation.

Chapter 6

Discussion

There are significant new factors which differentiate the approaches described in the current study with respect to the existing literature. In what follows, a summary of the key components incorporated in the proposed framework for the design of new and efficient wireless medical video telemedicine systems is discussed. Next, the contribution of the study through the attained results of the proposed methodologies with respect to the existing literature is provided.

Current state-of-the-art H.264/AVC standard is used for efficient source encoding. Baseline profile of H.264/AVC allows low complexity implementation targeting streaming to mobile devices, a scheme which is adopted in the proposed framework. As depicted in Table 6.1, H.264/AVC is now employed by latest studies [70], [72]-[75] as compared to the use of previous standards such as MPEG-2 [69], [71], MPEG-4 part 2 [69], [76], motion-JPEG (M-JPEG) [77], and H.263 [78] in earlier studies. Baseline profile defines new, particularly efficient error resilience techniques. These techniques enable robust medical video transmission, and are particularly useful in the design of reliable telemedicine systems, in the absence of a back channel. For the purpose of this study, FMO and RS error resilience tools, new in the H.264/AVC standard are employed. Efficient Intra-updating interval is also considered aiming to match the frame rate and consequently provide for diagnostically lossless cardiac cycles. While Intra-update interval is discussed by relevant studies in the literature [79], error resilient tools are not adequately addressed in the presented case studies. In fact, besides channel-protection incorporated in [72], [79], error resilience

mechanisms are not integrated in any of the considered studies in the literature. This is partially due to the fact that most studies chose to address possible transmission errors by switching to less aggravating for the network's resources transmission bitrate. While afore-described scheme efficiently addresses errors due to varying network state (limited bandwidth, channel load, congestion, etc.), it does not provide for dealing with transmission errors, inevitable in today's wireless medical video transmission. To address error prone bitrate limited channels and preserve transmission bandwidth, the proposed framework utilizes an efficient variable quality slice encoding scheme based on clinical significance criteria. A similar approach is undertaken by [69], [71]. The key difference is that the proposed scheme provides for multiple ROIs encoding as compared to single-ROI studies in [69], [71], while associated modifications based on FMO type 2 enable for a speedy, H.264/AVC compliant implementation. The benefits associated with the latter implementation, led to the adoption of the proposed approach for cardiac ultrasound videos recently [72]. To evaluate the validity of the proposed methods and demonstrate the resiliency of the proposed system, rigorous testing for PLR of up to 30% is performed. As already mentioned above, most studies trigger a switch to a different encoding state when network state changes (for example, increase in error rate). In [74], the experienced error rate is listed and the impact on video quality degradation is depicted. Comprehensive subjective and objective VQA, with correlation investigation is incorporated in the proposed framework. Correlation between clinical ratings and subjective scores is not present in the literature of medical video transmission systems, to the author's knowledge. Diagnostic validation is incorporated by [69], [71], [72], [74], [75], [78].

Table 6.1. The proposed medical video transmission system and the examined literature's studies targeting transmission over 3G (and beyond) wireless networks.

	Author	Year	Resolution., Frame Rate, BitRate[4]	Encoding Standard	Medical Video Modality	Comments
ROI-based Systems	Tsapatsoullis et al. [69][1]	07	352x288 @ 10fps 10 videos average: 507.2 Kbps	MPEG-2/ MPEG-4	Carotid Artery Ultrasound video	A saliency-based visual attention ROI coding for low bit-rate medical video transmission.
	Doukas et al. [70][1]	08	Resolution: N/A @ 25 fps 128-256 Kbps	H.264/ AVC (Scalable)	Skin Legion and MRI images, Trauma video snapshots	Scalable ROI encoding. Adaptive transmission based in context awareness (patient status and network state).
	Rao et al. [71][1,3]	09	360x240 @ 30 fps 500Kbps	MPEG-2	Pediatric respiratory distress related videos	ROI coding which incorporates different quantization levels for ROI and non-ROI, targeting diagnostically lossless encoding utilizing physician expert feedback.
	Martini et al. [72]	10	480x256@15fps 300 Kbps	H.264	Cardiac Ultrasonography	Context –aware FMO encoding using unequal error protection (UEP).
	Panayides[1,3]	**11**	**352x288 @ 15fps** **10 videos: 197-421 Kbps** **(median: 253 Kbps)**	**H.264/ AVC**	**Carotid Artery Ultrasound video**	**Diagnostically-driven ROI quantization (FMO) and redundant slices for error-resilient encoding. Clinical validation of diagnostic quality.**
Non-ROI based Systems	Chu et al. [77][2]	04	{320x240 and 160x120} <5fps Channel Bitrate: 50-80 Kbps	M-JPEG	Trauma video	Real time trauma video transmission. Network adaptation enabled through media transformation, data prioritization, and application-level congestion control.
	Garawi et al. [78][2,3]	06	176x144 @ 5fps 18.5-60 Kbps Channel BitRate: 64 Kbps	H.263	Echocardiogram	A performance analysis of an end-to-end mObile Tele-Echography using an ultra-Light rObot (OTELO)
	Pedersen et al. [74][2,3]	09	320x240 @ 10fps 349 Kbps Channel BitRate: 380 Kbps	H.264/ AVC (Scalable)	Echocardiogram	Spatiotemporal scalability over different wireless networks. Diagnostic quality and how it's affected by network parameters.
	Istepanian et al. [75][2,3]	09	176x144 @ 8-10fps 50-130 Kbps Channel BitRate: 360 Kbps	H.264/ AVC	Abdomen	QoS Ultrasound Steaming Rate Control (Q-USR) algorithm based on reinforcement learning that satisfies a medical QoS criterion.

[1]Simulation, [2]Real-time, [3]Clinical evaluation by medical experts (presented video parameters achieve diagnostically lossless encoding for preset clinical criteria),
[4]For real-time transmission, both encoded video bitrates and available channel bitrates are provided.

115

Objective VQA utilizes PSNR and SSIM metrics in most of the considered studies. In this study, 8 different VQA metrics are computed and their correlation to the medical expert's MOS is depicted. Experimentation is based on a data set composed of ten (10) videos with multiple video instances (1260), as compared to limited data sets incorporated by corresponding studies in the literature with a smaller amount of video instances. Eleven videos are used in [71] (5 for training the demonstrated system and six for evaluation), ten videos in [69], seven videos in [73], while four (4) videos part the data set in [74]. Remaining studies depict results obtained for one video with multiple video instances.

6.1 Diagnostic Region of Interest based Systems

The commonly accepted observation that medical video modalities incorporate video portions that directly relate to specific clinical assessment criteria led to the development of medical video transmission telemedicine systems that exploit this property. These systems can be characterized as diagnostic ROI-based systems. While this has been a known concept for medical imaging since JPEG-2000 as illustrated in [7], [12], [70], [71], there has been a rather limited adoption by medical video transmission systems. In [71], diagnostic ROI for pediatric respiratory distress related videos is utilized. The ROI is defined by the medical expert, and is considered to be in the centre of the video, occupying between 25%-50% of the video's spatial resolution. In [69], the diagnostic ROI is based on a visual attention model for what the reader (medical expert) would find interesting, for ultrasound videos of the carotid artery. Both studies employ MPEG-2 for encoding, while [69] also demonstrates MPEG-4 part 2 utilization.

While the aforementioned approaches address single diagnostic ROI encoding, the proposed framework utilizing H.264/AVC flexible macroblock ordering, allows for multiple ROI encoding with varying QPs, according to the region's diagnostic importance. The diagnostic ROIs are specified using a pixel-level segmentation algorithm which is accordingly extended to macroblock-level for encoding. The diagnostic ROIs can be also specified utilizing physician expert feedback. The ROIs spatial size range between 44%-72% of the video's size for the considered data set as depicted in Table 4.3, for CIF resolution. The proposed approach has been also adopted by [72] for echocardiogram ultrasound videos.

Overall, automated ROI computation and medical expert feedback are utilized for specifying the diagnostic ROIs. The sizes of the ROIs vary according to the regions of diagnostic interest and the incorporated medical video modality. The key concept in the success of diagnostic ROI-based systems is the bitrate demands savings without compromising diagnostic quality. The former is achieved by allocating bitrate resources to what is diagnostically relevant and the latter by diagnostic validation performed by medical experts. Bitrate demands reductions are depicted in subsection 6.4 while diagnostic validation is discussed in subsection 6.3.

6.2 Error Resilience Tools and Network Adaptation

H.264/AVC defines a plethora of new and powerful error resilience tools. Error resilience is an area that attracted considerable attention of the research community over the last decade. Given the fact that transmission errors are inevitable in video streaming over error prone wireless channels, the effort is directed toward minimising video degradation. Significant drop in quality leads to poor diagnostic performance in medical video transmission systems.

For the purposes of this thesis, FMO, RS, and efficient intra-update interval has been integrated in the proposed framework. While FMO also serves the dual purpose of defining diagnostic ROIs and is used for implementing variable quality slice encoding, RS is the key technology that enables error resilient performance in error prone environments. As documented in the results chapter, the proposed FMO ROI RS scheme for CIF resolution video transmission at 15 fps provides for diagnostically lossless performance for PLR up to 15% for atherosclerotic plaque ROI QP of 28. The proposed framework's system setting is summarized in Table 5.9. As illustrated in Fig. 4.1, this is attained by utilizing channel and end-user device knowledge prior to transmission, followed by a coarse to fine parameter optimization for efficient adaptation to the channel's conditions.

Error resilient implementation for a robust diagnostic performance is not efficiently addressed in the current literature as discussed in chapter 3 and also depicted in Table 6.2. The absence of error resilience mechanisms that will allow for error recovery is a significant factor that should be integrated in future telemedicine systems for remote diagnosis and care, especially in emergency situations. Most presented studies make use of network adaptation schemes utilizing feedback messages conveyed from the receiver end to the transmitting party. Triggering a switch to a different encoding state to lower the bandwidth requirements of the transmitted medical video efficiently addresses errors occurring from channel load and congestion. However, transmission errors, and especially burst errors are not considered. In general, fading channel conditions may trigger a switch, but the absence of error resilience schemes will not prevent medical video degradation and diagnostic capacity deterioration. Nevertheless, medical's video diagnostic capacity immediately prior and after a switch is not sufficiently discussed in presented studies.

More specifically, errors triggering a transmission compromise diagnostic quality, while reduced bandwidth typically involves resolution, frame rate, and quality relegation.

The proposed system demonstrates an *a priori* adaptation to the expected network conditions. This feature is particularly efficient in 3.5G networks and beyond, where a requested minimum data rate is sustained for the duration of a session, as illustrated in subsection 5.6. Efficient incorporated methods enable significantly reduced bandwidth demands while they are specifically designed for transmission in noisy wireless channels. Diagnostic performance robustness even if channel conditions introduce 15% PLR sufficiently tackles both transmission as well as congestion and channel load errors. For real-time adaptation, the proposed scheme can provide for increased RS rate for increased error resilience, while bandwidth demands can be further reduced following additional background compression without compromising diagnostic quality.

SVC is incorporated in [74], where different networks capabilities are evaluated. With respect to available bandwidth, different video enhancement layers of varying QoS parameters including resolution and frame rates are streamed through WLAN and 3G networks. The transmitted videos are clinically evaluated and the corresponding acceptable values are depicted. These are discussed in more detail in the following subsection. For the purposes of this thesis, SVC encoding was not considered due the relatively slow SVC software penetration to widely used media players, limiting potential usage. SVC will be addressed in the future work (see also subsection 7.2.3).

6.3 Clinical Video Quality Assessment

Successful implementation of medical video streaming systems assumes adequate diagnostic capacity of the transmitted video. To evaluate the diagnostic yield of the streamed medical video both objective and subjective approaches exist. The term subjective addresses the subjective assessment of perceived video quality of a human subject for conventional videos. For medical video quality assessment this is very different as illustrated in subsection 2.4. Clinical quality assessment by experienced technicians essentially validates the deployed system's performance. Therefore clinical quality assessment should be always considered in the design of medical video telemedicine systems.

For the purposes of this study, a new clinical rating system for the evaluation of different parts of atherosclerotic plaque ultrasound video that correspond to clinical assessment of different clinical criteria is proposed. According to Table 4.1 and Table 4.2, each criterion received independent evaluation scores by two medical experts. Following a coarse to fine parameter optimization, diagnostically lossless thresholds for QP values, resolution, and frame rate were deducted. As demonstrated, medical video's clinical capacity is directly related to the video's size, frame rate playback, and amount of compression. While different clinical criteria are assumed for different medical video modalities, the incorporated methods of the proposed framework can be applied with minor modifications to other medical video modalities as well.

Diagnostic validation has been incorporated in studies [69], [71], [72], [74], [75], [78]. In [71] a similar approach for deriving diagnostically lossless quality threshold for the incorporated clinical criteria in paediatric respiratory distress videos is depicted. In [76], a comprehensive evaluation platform for compressed echocardiogram video is presented. While this study does not correspond to a medical

video streaming system, it consists of a relevant clinical evaluation framework. Based on a set of clinical criteria, minimum bitrates are derived for different modes parting the echocardiogram video. In [75], [78] PSNR threshold value above which abdominal and echocardiogram ultrasound videos are deemed as diagnostically lossless is provided. Recommendations for resolution and frame rate are also given. Clinical ratings that describe diagnostically acceptable channel's parameters are presented in [74]. A rating scale between 1 and 4 is considered in [71], [74], while the modelling score ranges between 1 and 5 in [76].

Objective VQA both for medical, as well as conventional videos is largely based on PSNR and SSIM measurements. PSNR is a well established metric, however as documented in [9], it fails to efficiently correlate with perceived quality. SSIM has gain wide acceptance over the past five years. Both metrics were initially designed for image quality assessment. The latter metrics are the ones employed in the literature.

To overcome aforementioned limitations in the proposed framework, a series of the most established VQA algorithms have been employed during objective evaluation. Then, the computed measurements were fitted to the clinical MOS of a representative sample of 100 video instances and the correlation between objective and subjective scores was explored. The correlation was considered for the clinical assessment of plaque type and objective measurements were computed over the atherosclerotic plaque region. As depicted in Table 4.1, the diagnostic ROI describing the atherosclerotic plaque is accountable for determining the plaque type. The aim was twofold. First, to depict that new, diagnostically driven objective VQA algorithms should be developed, where diagnostically relevant assessment of the clinically important regions should be considered. Second, that ultimately for high correlations between computerized and clinical ratings, objective VQA could be used to determine

the diagnostic capacity of the transmitted medical video. Correlation investigation is not undertaken by any study in the presented literature.

6.4 Bandwidth Requirements and Clinically Acceptable Parameters

To demonstrate the bandwidth requirements and associated QoS parameters considered in the literature for medical video transmission over 3G wireless channels, we refer to Table 6.1 and Table 6.2. 3G channels are today's dominant technology for wireless medical video transmission facilitating worldwide deployment, extended coverage and mobility support. Limited upload date rates (typically between 220–384 kbps) however impose a bound on the amount of transmitted clinical data.

The following source encoding parameters and resulting bandwidths are considered in the literature targeting medical video transmission over 3G networks (see Table 6.1):

1. Resolution: 160x120 [77], 176x144 (QCIF) [75], [78], 320x240 (QVGA) [74], [77], 352x288 (CIF) [69], 360x240 [71], and 480x256 [72].

2. Frame Rate: 5 [77], [78], 10 [69], [74], [75], 15 [72], 25 [70], [73], 30 [71].

3. Bandwidth: 507 kbps [71], 500 kbps [69], 128-256 kbps [70], 50-80 kbps [77], 349 kbps [74], 50-130 kbps [75], 300 kbps [72].

The proposed framework transmits CIF resolution videos at 15 fps with resulting bitrates ranging between 197-420 kbps (median: 253 kbps) for the considered data set of ten ultrasound videos. Considerable bitrate demands reductions starting from 15% and extending up to 60% are achieved (see Table 5.8), with a median value of 44% for the proposed encoding setting. The variety of the bitrate savings is associated with the diversity of the ROI sizes ranging between 44%-75% of the video's display size. The ROI-based system described in [69], incorporates a similar data set composed of ten

ultrasound videos of the carotid artery. The proposed methodology for CIF resolution video transmission at 10 fps produces a bitrate average of 507 kbps, providing bitrate savings equivalent to 21.3% using MPEG-4 for encoding. Compared to the estimated bitrate of diagnostically lossless uniform encoding, 53.6% less bitrate is required for a fixed ROI-size of 25% of the video's size in [71]. The considered resolution is 364x240 at 30 fps.

Compared to the aforementioned ROI-based approaches, the proposed system's corresponding bitrates provide for transmission over 3G wireless channels. Bitrate requirements in the order of 500 kbps [69], [71] match typical 3.5G data rates. Multiple diagnostic ROI encoding is considered with respect to single-ROI schemes, while the current state-of-the-art H.264/AVC standard is employed. A higher frame rate of 15 fps is incorporated with respect to 10 fps in [69] for the same resolution, with higher bitrate savings. The incorporated ROI-sizes more than double the ROI-size considered in [71] in some cases, while slightly increased resolution is transmitted by the proposed framework (352x288 vs 360x240). On the other hand, the considered frame rate in [71] doubles the frame rate of the proposed system. Bitrate savings of 53.6% is in the range of depicted savings of 15-60%. The latter however being a function of the ROI-sizes, incorporates higher bits per pixels (bpp) compared to [71], both for diagnostic ROI and background video portions. None of the aforementioned studies addressed medical video quality deterioration after transmission.

Next, the proposed framework is weighed against non-ROI based systems evaluated over real wireless channels. The successfully streamed SVC layer in [74] provides a good indication of 3G networks' capabilities. Resolution and frame rate are

Table 6.2. Medical video transmission systems varied parameters.

Study	QP/ Bitrate	Source Encoding			Wireless Transmission			Clinical VQA
		Resolution	Frame Rate	Error Resilience	PLR	Delay	Jitter	
Tsapatsoullis et al. [69][1]	ROI-based	√	√	-	-	-	-	√
Doukas et al. [70][1]	Scalable	-	-	-	º	º	º	-
Martini et al. [72]	ROI-based	-	-	Channel Protection	√	√	-	+
Rao et al. [71][1,3]	ROI-based	-	-	-	-	-	-	√
Cavero et al. [73]	√	√	√	-	-	-	-	-
Chu et al. [77][2]	√	√	√	-	-	º	º	-
Pedersen et al.[74] [2,3]	Scalable	√	√	-	º	º	º	√
Girawi et al. [78]	√	√	√	-	-	√	√	√
Istepanian et al. [75][2,3]	√	-	√	-	º	º	º	+
Martini et al. [79]	Encoder States	-	√	Channel Protection	√	√	√	-
Panayides	**ROI-based**	√	√	√	√	**OW**	**OW**	√

√ denotes that the incorporated encoding parameter was investigated during the experiments. For network parameters it signifies that it was taken under consideration to modify encoding state. For clinical VQA it designates a scoring scale for different clinical criteria.

º denotes that measurements are depicted for the specific network parameter.

+ denotes that the transmitted medical video's quality conforms to a preset clinical threshold or verified by a medical expert but not listed in the study.

- denotes that the feature is not supported.

OW denotes ongoing work.

set to 320x240 and 10 fps respectively, with a matching bitrate of 349 kbps. The specific setting achieves clinically acceptable physicians' evaluation scores after transmission over a 3G wireless network with upload data rates of 380 kbps and experienced packet losses of 2%. The proposed diagnostically lossless encoding setting incorporates higher resolution (CIF) and frame rate (15) at comparable bitrates, while it provides for clinical practise for up to 15% of PLR. The key features enabling the aforementioned increases are diagnostically relevant ROI encoding and associated reduction in resulting bandwidth demands, and error resilient implementation. For the three videos of the considered data set exceeding available data rates, a further quantization of the background video regions can be employed for seamless streaming, without compromising medical video's diagnostic capacity.

Low bandwidth encoding and transmission is undertaken by studies [75], [77]. In [77], resolutions of 320x240 and 160x120 are transmitted at frame rates less than 5 fps. At such frame rates, clinical motion assessment is not always possible. However, the primary goal of this study is ECG and medical image transmission. For that reason, data prioritization algorithms favour ECG and medical image clinical data transmission at the expense of reduced frame rates. Moreover, limited bandwidth availability of the incorporated early 3G network with typical bandwidths ranging between 50-80 kbps (theoretical 153 kbps), prohibit matching a higher frame rate.

The rate control algorithm described in [75], allows varying the frame rate based on reinforcement learning. QCIF resolution video transmission is considered and the frame rate is varied between 8 and 10 frames. Given that 5 fps represent the diagnostically acceptable frame rate for the considered criteria in [78], bandwidth savings resulting from frame rate reductions enable increasing quality levels by an average of 2.5 dB with respect to the default JM rate control algorithm at 10 fps. The

latter is demonstrated for bitrates ranging from 50 kbps to 130 kbps. While using a 3.5G modem, the available upload data rates of 384 kbps are not uncommon in 3G networks. At such rates, higher resolutions and frame rates can be considered as already described above. For the considered resolution and bitrates, the proposed variable quality slice encoding scheme for QPs 38/30/28 achieves slightly lower ROI PSNR ratings, at 15 fps. At 10 fps, quantization set of 36/26/24 allows better quality ratings for comparable bitrates.

Chapter 7

Concluding Remarks and Future Work

7.1 Concluding Remarks

7.1.1   Overall framework's description

This thesis presents a new H.264/AVC based framework approach for effective communication and evaluation of wireless medical video over error-prone channels. Motivated by the need to efficiently address unique requirements associated with medical video source encoding, wireless transmission, and quality assessment, a unified framework which addresses individual requirements is proposed. The envisioned utilization scenarios target remote diagnosis and care and emergency situations. Resulting system's performance is demonstrated for the transmission of atherosclerotic plaque ultrasound video over 3G wireless networks. While particular features are medical video modality specific, the proposed framework can be efficiently applied with minor modifications to other medical video modalities as well. This is verified by initial experimentations on femoral and cardiac ultrasound videos [72]. Adaptive nature of the incorporated components provide for seamless integration of emerging and newly developed wireless technologies such as 3.5G and 4G wireless systems. Using diagnostically relevant encoding, the proposed framework achieves significant bandwidth demands deductions, while error resilient encoding provides for diagnostically lossless performance in noisy environments. Overall, coarse to fine parameter optimization and exhaustive experimentation allows providing recommendations regarding diagnostically lossless encoding setting for transmission over noisy 3G (and beyond) channels. Performance evaluation is based on a new

clinical rating system using the mean opinion score (MOS) provided by two medical experts. Objective assessment includes a series of the most established VQA algorithms. Correlation investigation between medical experts' ratings and objective measurements restricted to what is diagnostically relevant, highlights the need for designing new, diagnostically driven VQA metrics.

7.1.2 Incorporated methodologies

Clinical quality assessment of atherosclerotic plaque ultrasound videos requires the visualization of the plaque, the plaque boundary, and the associated near and far walls. This enables the medical expert to provide a diagnosis regarding plaque presence, associated degree of stenosis, and determine plaque type. Automated segmentation is used to identify diagnostic ROIs. These ROIs are then mapped to video slices and encoded utilizing H.264/AVC error resilience feature, FMO type 2. The FMO type 2 concept is modified to support variable quality slice encoding according to the slices' diagnostic importance. Diagnostically relevant encoding enables efficient utilization of the network's resources. By inserting redundant representations within the transmitted sequence, the encoded video becomes resilient to the presence of extensive PLR. To demonstrate the efficiency of the proposed framework a data set of ten ultrasound videos is used. The selected videos include a sufficient diversity of diagnostic ROIs sizes. This ensures the consistency of the depicted performance gains. A plethora of scenarios including different quality levels, resolutions, frame rates, packet loss rates, and amount of error resilience (redundant slices) are simulated. Overall, coarse to fine parameter optimization investigating 1260 scenarios that demonstrate efficiency near the diagnostic limit and under significant packet drops are simulated. To assess the proposed framework's performance, both subjective (clinical) and objective evaluations are performed.

Clinical evaluation is based on a new clinical rating that provides for independent ratings of the different clinical criteria, while objective video quality assessment includes today's most widely used metrics. The correlation between medical experts MOS and objective measurements for clinical assessment of plaque type is also investigated.

7.1.3 Accomplished results

Comprehensive evaluation designated that enhanced diagnostic performance is attained in noisy environments at significantly reduced bitrates. More specifically, bitrate demands reductions ranging between 15%-60% with a median of 44% for the proposed QP levels (function of diagnostic ROIs sizes) are achieved, while efficient error resilient encoding enables diagnostically lossless performance even at 15% packet drops. The proposed system setting for diagnostically lossless performance for the aforementioned clinical criteria at 3G data transfer rates includes CIF resolution encoding at 15 fps, with one RS inserted every four coded slices. Proposed diagnostically relevant quality levels are 38 (background)/ 30 (Walls and ECG) / 28 (plaque) or above 35db for the plaque ROI. The proposed system setting is summarized in Table 5.9. For determining the plaque type, both the (ROI) PSNR and the WSNR gave very good correlations to the MOS provided by two medical experts (LCC: 0.674 and 0.69, and SROCC: 0.667 and 0.715, respectively), highlighting the need of designing new, diagnostically driven VQA algorithms.

7.2 Future work

Future work includes investigating higher quality encodings over emerging high-bandwidth 3.5G and 4G wireless networks. Higher resolution and frame rate encodings close to acquired medical video's resolution and frame rate is envisioned.

4CIF resolution transmission (and higher) is currently planned and associated diagnostic capacity is to be investigated. Towards this end, new MOVIE and VQM algorithms specifically designed for video quality evaluation purposes will be exploited. A higher correlation with the medical experts' MOS is expected. In addition, threshold values above which diagnostically lossless quality is attained for the remaining VQA algorithms considered in this study will be considered.

A rate control algorithm that will allow for diagnostically relevant encoding according to a predefined data rate will be implemented. For an even improved network's resources utilization, the proposed framework will be enhanced to accommodate a mode of operation where only diagnostic ROIs redundant representations will be included in the transmitted bitstream. For error resilience, effective integration of H.264/AVC FMO type 1 and type 2 will be considered (see subsection 2.1).

Scalable video coding for generating fast adaptive encoding states over candidate heterogeneous networks will be also employed. The latter will follow the implementation of the proposed diagnostically relevant encoding to SVC software, as well as determining diagnostic capacity of higher resolution encodings. It is important to note here that SVC was not considered within the framework of this thesis due to the relatively slow adaptation of commercially available SVC players. It is expected that this will be solved in the near future.

The joint collaborative team on video coding (JCT-VC) standard development was established in 2010 [92] to develop a video coding standard more advanced than the current H.264/AVC. While still at the early stages, high efficiency video coding (HEVC) test model (HM) reference software 1.0 has been already released [93]. A final recommendation is expected in early 2013. A close monitoring of the developing

procedure will allow for a timely integration of the increased efficiency features (targeting 50% reduction in bandwidth for equivalent subjective quality) in the future.

7.2.1 Emerging 4G wireless networks and associated bandwidth increases linked with transmitted medical video's diagnostic capacity

Standardization of WirelessMan-Advanced and LTE-advanced 4G technologies conforming to IMT-advanced requirements, refinement of already developed 3.5G systems and worldwide deployed 3G channels include a range of cutting edge technology features. As discussed in [94], these technologies will be fully developed within the next years and meet a peak acceptance within the next two decades. At the same time, ongoing initiation regarding the successor of the highly successful current state-of-the-art H.264/AVC video coding standard dictate immediate actions regarding the development of new m-health systems and services. Such systems must benefit by the foreseen new horizons formed before them and especially in the area of remote diagnosis and care. Systems incorporating the wireless transmission of medical video must integrate these technologies and assure to fully take advantage of the new promising perspectives.

The proposed framework described in this thesis is capable of integrating the abovementioned technologies' features given its scalable and adaptive design. Based on medical expert's feedback, channel knowledge, and end-user device capabilities, higher bandwidth availability through 3.5G and 4G networks will be investigated with respect to the impact on medical video's diagnostic capacity. Higher quality, resolution, and frame rate encodings, close to medical video's acquisition setting will be exploited. As described in the methodology chapter and verified by clinical assessment in the results chapter, increased quality, resolution, and frame rate accounts for increased diagnostic capacity. Given that 4CIF resolution transmission is

now possible via 3.5G (and beyond) channels, the effect on diagnostic quality remains to be investigated. Overall, a scalable recommendation setting associating different resolutions (QCIF, CIF, and 4CIF) and existing 3G and 3.5G, and emerging 4G wireless networks with attainable diagnostic capacity is envisioned. Ultimately, the objective is transmitting medical video that is of the same diagnostic quality as the acquired medical video from the ultrasound device.

7.2.2 Diagnostically relevant rate control algorithm and proposed framework system enhancements.

A rate control algorithm for diagnostically relevant encoding according to the region's diagnostic importance will be developed. The obvious benefits of controlling the transmitted data rate while also monitoring individual ROIs quality for diagnostically lossless performance is envisioned.

Further enhancement of the proposed framework will also be considered for increased error resilience and even more efficient utilization of the network's resources. The proposed framework will be enhanced to provide a mode of operation where redundant slices will be only transmitted for the most important diagnostic ROI slices, also paired with unequal error protection. Furthermore, a combination of FMO type 1 (scattered slices) and currently employed FMO type 2 error resilience features will be examined. In this way, a more robust setting for the plaque ROI components is sought.

7.2.3 Scalable Video Coding for Medical Video Applications

Scalable video coding [33] can be used to allow for a single medical video bitstream to be displayed on different devices with different requirements on spatial resolution and frame rates. The basic idea is to first encode video into different layers. The base layer is to be decoded by every device. Then, additional (enhancement)

layers need to be decoded in order to provide for higher spatial resolutions and frame rates.

A typical example of spatiotempral scalability using atherosclerotic plaque ultrasound video is presented in Table 7.1. In Table 7.1, we can see that the base layer required the lowest bandwidth at the lowest possible spatial resolution and frame rates. The first of enhancement layers provided higher frame rates at increased bandwidth requirements. Then, the highest enhancement layers provide higher spatial resolutions at the even higher bandwidth requirements. The highest layer provided the highest spatial resolution at the highest possible bitrate. It is clear that scalable video coding can also be used to control bandwidth requirements. It can thus be used to adapt the video compression bitstream to available bandwidth.

7.2.4 New video quality assessment algorithms and diagnostically lossless threshold values.

Fine QP parameter optimization using medical expert's feedback showed that for ROI QP of 28, clinical quality is preserved in the compressed video. Subsequent experimentations revealed that PSNR ratings over 35db provide for diagnostically lossless medical videos, attaining medical expert's ratings greater or equal to 4, following transmission. The latter are summarized in the proposed framework's system setting in Table 5.9. Given that 8 VQA metrics were used for quality evaluation in this thesis, remaining quality metrics thresholds are also to be investigated. Further exploitation of objective VQA algorithms for deriving the threshold value above which diagnostic quality is preserved for each metric is also planned.

To this end, MOVIE and VQM VQA metrics will be also used for evaluation purposes. These two metrics were not considered during quality evaluation of the

current study, given the small plaque ROI sizes associated with CIF resolution. As documented in chapter 2, both algorithms do not work very well with small spatial dimensions. Higher bandwidths available through emerging wireless networks enabling 4CIF resolution transmission will provide for larger plaque ROI sizes, and therefore also allowing evaluating quality using these current state of the art VQA algorithms. Correlation with medical expert's MOS is expected to be even higher than the depicted correlations in this thesis, given the fact that both algorithms were specifically developed to address video aspects, such as motion evaluation, and were not originally designed for image assessment.

Table 7.1 – An Atherosclerotic Plaque Ultrasound Example of Spatiotemporal Scalability.

Layer	Resolution	Frame Rate	Bitrate (kbps)
0	176x144	1.875	108.2
1	176x144	3.75	150.6
2	176x144	7.5	194.3
3	176x144	15	227
4	176x144	30	243.5
5	352x288	1.875	320.1
6	352x288	3.75	469.7
7	352x288	7.5	641.6
8	352x288	15	798.6
9	352x288	30	886.4
10	704x576	1.875	465.3
11	704x576	3.75	661.5
12	704x576	7.5	887.8
13	704x576	15	1103
14	704x576	30	1248

Input Videos: {4CIF: 704x576, CIF: 352x288, QCIF: 176x144}@30 fps, QP: 28, Software version: JSVM 9.19.9.

Bibliography

[1] C. S. Pattichis, E. Kyriacou, S. Voskarides, M. S. Pattichis, R. Istepanian, "Wireless telemedicine Systems: An Overview," IEEE Antennas and Propagation Magazine, vol 44, no. 2, pp.143-153, 2002.

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

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

[4] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), "Advanced Video Coding for Generic Audiovisual Services," v1, May, 2003; v2, Jan. 2004; v3 (with FRExt), Sept. 2004; v4, July 2005; v5, June 2006; v6, June 2006; v7, April 2007; v8 (with SVC), Nov. 2007; v9, Jan. 2009; v10 (with CPB), March 2009; v11(with MVC), March 2009; v12, March 2010; v13, March 2010.

[5] H. S. Ng, M. L. Sim, C. M. Tan, and C. C. Wong, "Wireless technologies for telemedicine," BT Technology Journal, Springer Netherlands, vol. 24, no. 2, pp. 130–137, 2006.

[6] Rysavy Research, LLC, "Transition to 4G, 3GPP Broadband Evolution to IMT-Advanced," 2010. Available: http://www.4gamericas.org/.

[7] H. Wang, L. Kondi, A. Luthra, and S. Ci, 4G Wireless Video Communications. New York: John Wiley & Sons, 2009.

[8] Portable Ultrasound Device. Available: http://www.technologyreview.com/biomedicine/.

[9] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video," *IEEE Transactions on Image Processing*, vol. 19, no.6, pp.1427-1441, June 2010.

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

[11] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst.Video Technol.*, vol. 13, pp. 560–576, July 2003.

[12] A. C. Bovik, The Essential Guide to Video Processing (2nd ed.). Academic Press, 2009.

[13] M. Grayson, K. Shatzkamer, and S. Wainner, IP Design for Mobile Networks. Networking Technology Series, Cisco Press, 2009.

[14] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Proc. Mag.*, vol. 17, pp. 61-82, July 2000.

[15] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer and T. Wedi, "Video coding with H.264/AVC: tools, performance and complexity," *IEEE Cir. Syst. for Video Technol. Mag.*, Vol. 4(1), pp.7-28, 2004.

[16] S. Kumar, L. Xu, M. K. Mandal, and S. Panchanathan, "Error Resiliency Schemes in H.264/AVC Standard," *Elsevier J. of Visual Communication & Image Representation* (Special issue on Emerging H.264/AVC Video Coding Standard), Vol. 17(2), April 2006.

[17] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Cir. Syst. Video Technol.*, vol. 13, pp. 645–656, 2003.

[18] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, pp. 657–673, July 2003.

[19] W. Zia, T. Afzal,W.Xu, T.Stockhammer, and G.Liebl, "Interactive Error Control for Mobile Video Telephony," *in Proc. ICC 2007*, Glasgow, UK, Jun. 2007.

[20] S. Wenger, "FMO: Flexible Macroblock Ordering," ITU-T JVT-C089, May 2002.

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

[22] M. Karczewicz and R. Kurçeren, "The SP and SI Frames Design for H.264/AVC", *IEEE Transactions on Circuits and Systems*, Vol. 13, No. 7, Jul. 2003.

[23] GSM World, Available: http://www.gsmworld.com/.

[24] 3GPP TS 25.308 V5.4.0 (2002-10) High Speed Downlink Packet Access (HSDPA) Stage 2 - Release 5.

[25] 3GPP, "Overview of 3GPP Release 6," V0.1.1, 2010, Available: http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/.

[26] 3GPP TR 25.913, "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," Release 7, v7.3.0, March 2006.

[27] IEEE std. 802.16e-2005: IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, and IEEE Std.802.16-2004/Cor1-2005, Corrigendum 1, Dec. 2005.

[28] 3GPP, "Overview of 3GPP Release 8," V0.2.2, 2011, Available: http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/.

[29] 3GPP, "Overview of 3GPP Release 9," V0.2.1, 2011, Available: http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/.

[30] 3GPP TR 36.913, "Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access (EUTRA)," v. 8.0.1, Mar. 2009; ftp://ftp.3gpp.org.

[31] 3GPP, "Overview of 3GPP Release 10," V0.1.0, 2011, Available: http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/.

[32] ITU-R Rep. M.1645, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," Jan. 2003.

[33] ITU-R Rep. M.2134, "Requirements Related to Technical Performance for IMT-Advanced Radio Interface(s)," Nov. 2008.

[34] P. Reinaldo, Wireless Communications Design Handbook, Volume I: Space (Interference: Aspects of Noise, Interference, and Environmental Concerns). San Diego: Academic Press, 1998.

[35] ITU-R SG5, "Invitation for submission of proposals for candidate radio interface technologies for the terrestrial components of the radio interface(s) for IMT-Advanced and invitation to participate in their subsequent evaluation", Circular Letter 5/LCCE/2, March 2008.

[36] ITU-R, Doc. IMT-ADV/10-23 - Evaluation of IMT-Advanced candidate technology submissions in Documents IMT-ADV/4-5-7 (IEEE) IMT-ADV/6-8-9 (3GPPP). Available: http://www.itu.int/ITU-R/index.asp?category=study-groups&rlink=rsg5-imt-advanced&lang=en.

[37] M. Alasti, B. Neekzad, C. Jie Hui, and R. Vannithamby, "Quality of service in WiMAX and LTE networks [Topics in Wireless Communications]," *Communications Magazine, IEEE*, vol.48, no.5, pp.104-111, May 2010.

[38] IEEE Std. 802.16-2004: IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Jun. 2004.

[39] The Draft IEEE 802.16m System Description Document (SDD), IEEE 802.16 Broadband Wireless Access Working Group, Jul. 2008.

[40] A. Ghosh, D. R. Wolter, J. G. Andrews, R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *Communications Magazine, IEEE*, vol.43, no.2, pp.129-136, Feb. 2005.

[41] S. Ahmadi, "An overview of next-generation mobile WiMAX technology," *Communications Magazine, IEEE*, vol.47, no.6, pp.84-98, June 2009.

[42] I. Papapanagiotou, D. Toumpakaris, L. Jungwon, M. Devetsikiotis, "A survey on next generation mobile WiMAX networks: objectives, features and technical challenges," *Communications Surveys & Tutorials, IEEE*, vol.11, no.4, pp.3-18, Fourth Quarter 2009.

[43] T. Nakamura, "Proposal for Candidate Radio Interface Technologies for IMT-Advanced Based on LTE Release 10 and Beyond (LTE-Advanced)," *ITU-R WP 5D 3rd Workshop on IMT-Advanced*, Dresden, Germany, 15 Oct., 2009.

[44] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, and K. Zangi, "LTE-Advanced - Evolving LTE towards IMT-Advanced," *Vehicular Technology Conference, 2008,VTC 2008-Fall, IEEE 68th*, pp.1-5, 21-24 Sept. 2008.

[45] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *Wireless Communications, IEEE*, vol.17, no.3, pp.10-22, June 2010.

[46] S. Parkvall, A. Furuskär, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *Communications Magazine, IEEE*, vol.49, no.2, pp.84-91, Feb. 2011.

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

[48] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, SIP: Session Initiation Protocol, Internet Engineering Task Force, RFC 2543, Mar. 1999.

[49] J. Postel, "Transmission Control Protocol," RFC 793, 1981.

[50] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hypertext Transfer Protocol-HTTP/1.1," RFC 2616, 1999.

[51] J. Postel, "User Datagram Protocol," RFC 768, 1980.

[52] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, RTP: A Transport Protocol for Real-Time Applications, Internet Engineering Task Force, RFC 1889, Jan. 1996.

[53] P. Antoniou, V. Vassiliou, A. Pitsillides, "ADIVIS: A Novel Adaptive Algorithm for Video Streaming over the Internet," *in Proc of PIMRC'07*, Athens, Greece, Sep. 2007.

[54] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error measurement to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[55] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process.: Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[56] D. M.Chandler and S. S.Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images," *IEEE Transactions on Image Processing*, vol.16, no.9, pp.2284-2298, Sept. 2007.

[57] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006. (VIF)

[58] H.R. Sheikh, A.C. Bovik and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol.14, no.12pp. 2117- 2128, Dec. 2005.

[59] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol.9, no.4, pp.636-650, Apr. 2000. (NQM)

[60] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images", *IEEE Trans. Inf. Theory*, IT-20(4), pp. 525-535, July 1974. WSNR

[61] T. Mitsa and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms", *ICASSP '93-V*, pp. 301-304. WSNR

[62] Metrix_mux objective video quality assessment software, Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/.

[63] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*, A. B. Watson, Ed.,1993, pp. 207–220.

[64] ITU-R Recommendation G.1010, "End-user multimedia QoS Classes," Nov. 2001.

[65] ITU-R Recommendation G.1011, "Reference guide to quality of experience assessment methodologies." Jun. 2010.

[66] K. Seshadrinathan and A. C. Bovik, "Motion Tuned Spatio-temporal Quality Assessment of Natural Videos", vol. 19, no. 2, pp. 335-350, IEEE Transactions on Image Processing, Feb. 2010.

[67] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[68] ITU-R Recommendation BT.500-11: "Methodology for the subjective assessment of the quality of television pictures." International Telecommuncation Union, Geneva, Switzerland, 2002.

[69] N. Tsapatsoulis, C. Loizou, and C. Pattichis, "Region of Interest Video Coding for Low bit-rate Transmission of Carotid Ultrasound Videos over 3G Wireless Networks," in Proc. of *IEEE EMBC'07*, Aug. 23-26, 2007, Lyon, France.

[70] C. Doukas and I. Maglogiannis, "Adaptive Transmission of Medical Image and Video Using Scalable Coding and Context-Aware Wireless Medical Networks," *EURASIP JWCN*, vol. 2008, ID 428397, 12 pages, 2008.

[71] S. P. Rao, N. S. Jayant, M. E. Stachura, E. Astapova, and A. Pearson-Shaver, "Delivering Diagnostic Quality Video over Mobile Wireless Networks for Telemedicine," *International Journal of Telemedicine and Applications*, vol. 2009, Article ID 406753, 9 pages, 2009.

[72] M. G. Martini and C. T. E. R. Hewage, "Flexible Macroblock Ordering for Context-Aware Ultrasound Video Transmission over Mobile WiMAX," *International Journal of Telemedicine and Applications*, vol. 2010, Article ID 127519, 14 pages, 2010. doi:10.1155/2010/127519.

[73] E. Cavero, A. Alesanco, and J. Garcia, "A new approach for echocardiogram compression based on display modes," *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, pp.1-4, 3-5 Nov. 2010.

[74] P. C. Pedersen, B. W. Dickson, and J. Chakareski, "Telemedicine applications of mobile ultrasound," in *Proc. of MMSP '09*, pp.1-6, 5-7 Oct. 2009.

[75] R. S. H. Istepanian, N. Y. Philip, M.G. Martini, "Medical QoS provision based on reinforcement learning in ultrasound streaming over 3.5G wireless systems," *Selected Areas in Communications, IEEE Journal on* , vol. 27, no. 4, pp. 566-574, May 2009.

[76] A. Alesanco, C. Hernandez, A. Portoles, L Ramos, C. Aured, M. Garcıa, P. Serrano, and J. Garcıa1, "A clinical distortion index for compressed echocardiogram evaluation: recommendations for Xvid codec," *Physiological Measurement*, vol. 30, no. 5, pp. 429-440, 2009.

[77] Y. Chu and A. Ganz, "A mobile teletrauma system using 3G networks," *Information Technology in Biomedicine, IEEE Transactions on* , vol. 8, no. 4, pp. 456-462, Dec. 2004.

[78] S. A. Garawi, R. S. H. Istepanian, and M. A. Abu-Rgheff, "3G wireless communication for mobile robotic tele-ultrasonography systems," *IEEE Comms. Mag.*, vol. 44, no. 4, pp. 91–96, April 2006.

[79] M. G. Martini, R. S. H. Istepanian, M..Mazzotti, and N.Philip, "Robust multi-layer control for enhanced wireless tele-medical video streaming," *IEEE Trans. on Mobile Computing*, vol. 9, no. 1, pp. 5-16, Jan. 2010.

[80] C. P. Loizou, C.S. Pattichis, M. Pantziaris, and A. Nicolaides, "An integrated system for the segmentation of atherosclerotic carotid plaque," *IEEE Trans. on Inform. Techn. in Biomedicine*, vol. 11, no. 5, pp. 661-667, Nov. 2007.

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

[82] S. Park and K. Miller, "Random Number Generators: Good Ones Are Hard To Find," ACM Commun., vol. 39, no. 10, pp. 1192-1201, Oct. 1988.

[83] D. Williams, M. Shah, "A Fast Algorithm for Active Contour and Curvature Estimation", *GVCIP: Imag. Und.*, vol. 55, no. 1, pp. 14-26, 1992.

[84] C. P. Loizou, C. S. Pattichis, C.I. Christodoulou, R.S.H. Istepanian, M. Pantziaris, and A. Nicolaides, "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery," *IEEE Trans. Ultrasonics Ferroelectrics and Frequency Control*, vol. 52, no. 10, pp. 1653-1669, 2005.

[85] Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment, 2000 [Online]. Available: fttp://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI.

[86] 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, Mar. 2003.

[87] OPNET University Program: Available: http://www.opnet.com/services/ university/.

[88] ITU-R Recommendation M.1225, "Guidelines for evaluation of radio transmission technologies for IMT–2000," 1997.

[89] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communication*, vol.17, no.7, pp. 1205-1222, July 1999.

[90] D. Niyato, E. Hossain, and J. Diamond, "IEEE 802.16/WiMAX-based broadband wireless access and its application for telemedicine/e-health services," *IEEE Wireless Communications,* vol.14, no.1, pp.72-83, Feb. 2007.

[91] E. Reetz and R. Tonjes, "Deliverable D-7.4, MORYNE IST-FP6 Project," March 2008, http://www.fp6-moryne.org/doc/MORYNE-D7.4-V1.1.pdf.

[92] ITU-T, "Terms of Reference of the Joint Collaborative Team on Video Coding Standard Development," Jan. 2010.

[93] H.265. Available: http://www.h265.net/.

[94] 4G Americas. Available: http://www.4gamericas.org/.

APPENDIX 1

Publications

IEEE Journal Papers

1) A. Panayides, M.S. Pattichis, C.S. Pattichis, C. P. Loizou, M. Pantziaris, and A. Pitsillides, *Atherosclerotic Plaque Ultrasound Video Encoding, Wireless Transmission, and Quality Assessment Using H.264*, **IEEE Transactions in Information Technology in Biomedicine**, vol.15, no.3, pp.387-397, May 2011. doi: 10.1109/TITB.2011.2105882

IEEE Magazine Papers

2) A. Panayides, M.S. Pattichis, C.S. Pattichis, and A. Pitsillides, *A Tutorial for Emerging Wireless Medical Video Transmission Systems,* **IEEE Antennas & Propagation Magazine,** to be published May 2011.

3) E. Kyriacou, M.S. Pattichis, C.S. Pattichis, A. Panayides, A. Pitsillides, *e-Emergency m-Health Systems: Current Status and Future Directions,* **IEEE Antennas & Propagation Magazine,** Vol. 49, No. 1, Feb. 2007, pp. 216-231.

Book Chapters

4) E. Kyriacou, P. Constantinides, C.S. Pattichis, M.S. Pattichis, and A. Panayides , *eEmergency Healthcare Informatics*, Ed. by J. Bronzino, **Handbook of Biomedical Engineering**, CRC Press, to be published in 2011.

5) A. Panayides, M. S. Pattichis, C. S. Pattichis, C. P. Loizou, M. Pantziaris, and A. Pitsillides, *Towards Diagnostically Robust Medical Ultrasound Video Streaming using H.264*, in **Biomedical Engineering**, Ed. by Carlos Alexandre Barros De Mello, IN-TECH, Vienna, Austria, pp. 219-237, 2009.

Conference Papers

6) E. Kyriacou, P. Constantinides, C. Pattichis, M. Pattichis, and A. Panayides, "eEmergency Health care Information Systems," invited paper, submitted to **33rd Annual International IEEE EMBS Conference of the IEEE Engineering in Medicine and Biology Society**, IEEE EMBC'11, Aug. 30 – Sep. 3, 2011, Boston, MA, USA.

7) A. Panayides, M. S. Pattichis, C. S. Pattichis, C. N. Schizas, A. Spanias, and E. Kyriacou, "An Overview of Recent End-to-End Wireless Medical Video Telemedicine Systems using 3G," **in Proc. of 32nd Annual Conference of the IEEE Engineering in Medicine and Biology Society**, IEEE EMBC'10, Aug. 31-Sep. 4, 2010, Buenos Aires, Argentina.

8) A. Panayides, M. S. Pattichis, C. S. Pattichis, C. P. Loizou, and M. Pantziaris, "Wireless Ultrasound Video Transmission for Stroke Risk Assessment," **International Workshop on Video Processing and Quality Metrics for Consumer Electronics, (VPQM 2010)**, Scottsdale, Arizona, Jan. 13-15, 2010.

9) A. Panayides, M. S. Pattichis, C. S. Pattichis, C. P. Loizou, M. Pantziaris, and A. Pitsillides, "Robust and Efficient Ultrasound Video Coding in Noisy Channels Using H.264," **in Proc. of 31ˢᵗ Annual Conference of the IEEE Engineering in Medicine and Biology Society**, IEEE EMBC'09, Sep. 2-6, 2009, Minnesota, U.S.A.

10) A. Panayides, M. S. Pattichis, and C. S. Pattichis, "Wireless Medical Ultrasound Video Transmission Through Noisy Channels," **in Proc. of 30ᵗʰ Annual Conference of the IEEE Engineering in Medicine and Biology Society**, IEEE EMBC'08, Aug. 28-30, 2008, Vancouver, Canada.

11) A. Panayides, C. Christophorou, J. Antoniou, A. Pitsillides, V. Vassiliou, "Power Counting for Optimized Capacity in MBMS Enabled UTRAN", **in Proc. of IEEE Symposium on Computers and Communications 2008 (ISCC'08)**, July 6 - 9, 2008, Marrakech, Morocco.

12) R. Hockmann, C. A. Jotten, C. Sgraja, C. Christophorou , A. Panayides , A. Pitsillides, R. Chiang , E. Reetz, R. Tonjes, "Evaluation of RAN Concepts for enhancing MBMS in the framework of C-MOBILE"**, in Proc. of ICT Mobile Summit 2008,** Stockholm, Sweeden, June 2008.

13) C.S. Pattichis, E. C. Kyriacou, M.S. Pattichis, , A. Panayides, S. Mougiakakou, A. Pitsillides, C. Schizas, "A brief overview of m-Health e-Emergency Systems," Invited paper**,** *Proceedings of the 2007* **IEEE International Conference on Information Technology Applications in Biomedicine**, *ITAB'07*, Tokyo, Japan, Nov. 8-11, 2007.

14) C.S. Pattichis, E. Kyriacou, M.S. Pattichis, A. Panayides, and A. Pitsillides, A review of m-Health e-Emergency Systems, in **Proc. of ITAB 2006, International Special Topic Conference on Information Technology and Biomedicine**, Ioannina - Epirus, Greece, October 28-28, 2006.

15) A. Panayides, M.S. Pattichis, C.S. Pattichis, and A. Pitsillides, *A Review of Error Resilience Techniques in Video Streaming,* in **Proc. of ISYC 2006, International Conference On Intelligent Systems and Computing: Theory and Applications**, Ayia Napa, Cyprus, July 6-7, 2006, pp.39-48.

Workshops

16) A. Panayides, M. S. Pattichis, C. S. Pattichis , A. Spanias, C. P. Loizou, and E. Kyriacou, "Ultrasound Video Transmission in M-Health Systems," **in Proc. of 3ʳᵈ Cyprus Workshop on Signal Processing and Informatics**, Nicosia, Cyprus, July 15ᵗʰ , 2010.

17) A. Panayides, M. S. Pattichis, C. S. Pattichis , C. P. Loizou, and A. Pitsillides, "Diagnostically Resilient Medical Ultrasound Video Streaming using H.264," **in Proc. of 2ⁿᵈ Cyprus Workshop on Signal Processing and Informatics**, Nicosia, Cyprus, July 14ᵗʰ , 2009.

18) M. S. Pattichis, S. Murillo, and A. Panayides, "Diagnostically Driven Image Processing Systems," **in Proc. of 2ⁿᵈ Cyprus Workshop on Signal Processing and Informatics**, Nicosia, Cyprus, July 14ᵗʰ , 2009.

19) A. Panayides, M. S. Pattichis, C. P. Loizou, and C. S. Pattichis, "Error Resilience Wireless Transmission of Medical Ultrasound Video Through Noisy Channels Using

Segmentation," **in Proc. of 1st Cyprus Workshop on Signal Processing and Informatics**, Nicosia, Cyprus, July 8, 2008.

APPENDIX 2

Williams &Shah Snake

A snake contour may be represented parametrically by $v(s) = [x(s), y(s)]$, where $(x, y) \in \Re^2$ denotes the spatial coordinates of an image, and $s \in [0,1]$ represents the parametric domain. The snake adapts itself by a dynamic process that minimizes an energy function defined as [83].

$$
\begin{aligned}
E_{snake}(v(s)) &= E_{int}(v(s)) + E_{image}(v(s)) + E_{external}(v(s)) = \\
&\int_s (\alpha(s)E_{cont} + \beta(s)E_{curv} + \gamma(s)E_{image} + E_{external})ds.
\end{aligned} \tag{1}
$$

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 $E_{int}(v)$, 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 [83]. 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 [83].

Figure 1 (a) shows the first frame from an ultrasound video of the CCA, whereas Fig. 1(b) shows the blood flow image. After cross-correlating the original image with the blood flow image, the initial blood flow contour is extracted (see Fig, 1(c)). The user then selects an area of interest, which will be used an initial contour. This will be used as initialisation for the snakes segmentation algorithm, which then deforms and converges as shown in Fig. 1(d). Figure 1(e) shows the manual segmentation results made by an expert radiologist. Additioanlly in order to help the snake converge better the lsmv despeckle filter [84] was applied on the original image. 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 [80], 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.
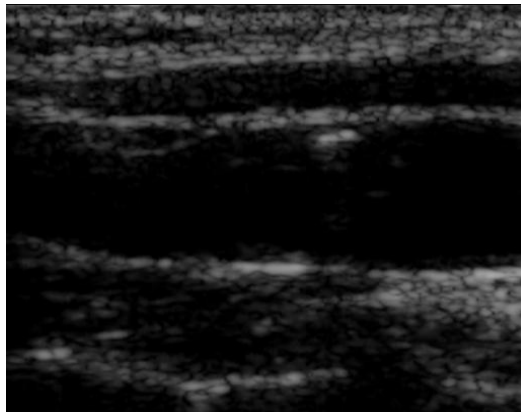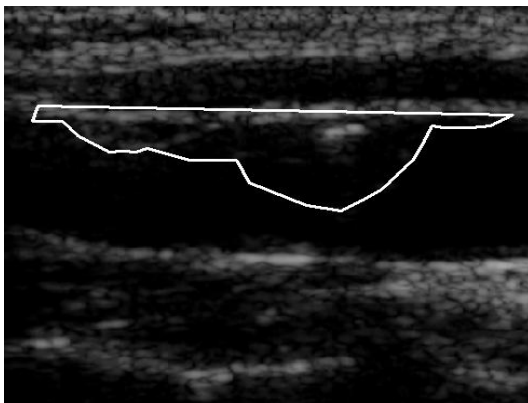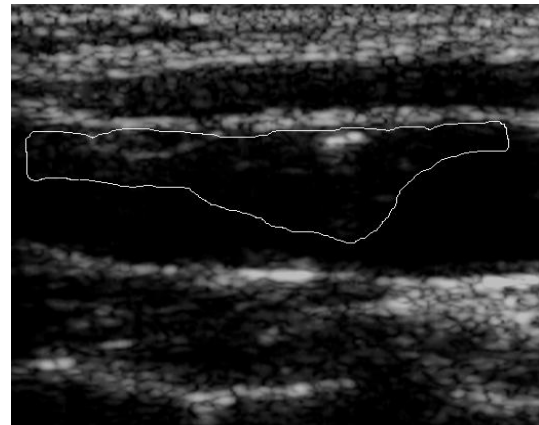
(a) Original B-mode image.
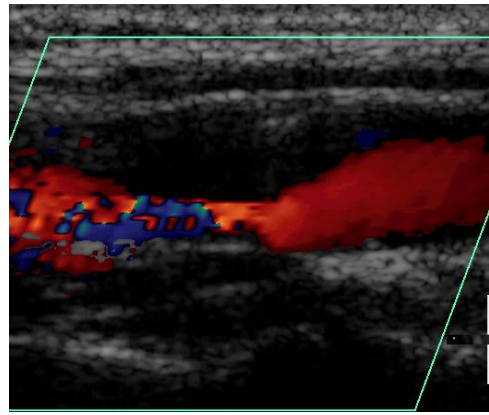
(b) Blood flow image.
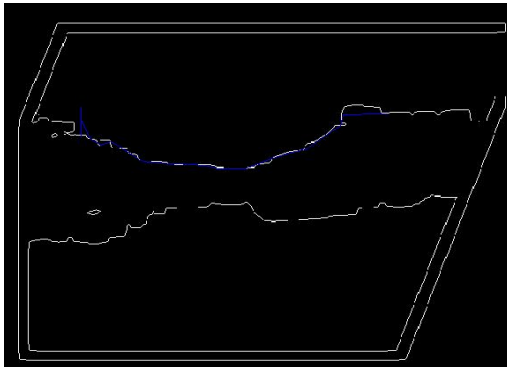
(c) Initial blood flow edge contour.

(d) Williams & Shah snakes segmentation results.

(e) Manual segmentation results.

Fig. 1. Plaque initialization using the blood flow image: (a) Original ultrasound B-mode image of a carotid artery with plaque at the far wall, (b) blood flow of the image 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.