ELASTIC ALGORITHMS FOR REGION-OF-INTEREST VIDEO COMPRESSION, WITH APPLICATION TO MOBILE TELEHEALTH

A Dissertation
Presented to
the Academic Faculty

by

Sira P Rao

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2007

ELASTIC ALGORITHMS FOR REGION-OF-INTEREST VIDEO COMPRESSION, WITH APPLICATION TO MOBILE TELEHEALTH

Approved by:

Professor Nikil Jayant
Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Vijay Madisetti
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Russell Mersereau
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Anthony Joseph Yezzi
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Gregory Abowd
College of Computing
Georgia Institute of Technology

Date Approved: 13 August 2007

ACKNOWLEDGEMENTS

My sincere thanks to my advisor Prof Nikil Jayant for his faith, support, and guidance during my PhD years. He has been a source of constant inspiration, and I have learnt from him a spectrum of things - the art of research, teamwork, meeting deadlines, document writing, presentation, and much more. I am also grateful to my parents and my brother for their affection. Their love and prayers have been selfless and unconditional, and their patience unparalleled.

I am also happy to have interacted and worked closely with the doctors at the Medical College of Georgia – Dr Max Stachura, Dr Elena Khasanshina, and Dr Tony Pearson-Shaver. Many thanks to them for their continued assistance with video databases for my research, the publications we worked together on, and more importantly, their time and feedback.

I thank my committee members – Prof Vijay Madisetti, Prof Russell Mersereau, Prof Anthony Yezzi, Prof Krishna Palem, and Prof Gregory Abowd for their time and feedback. Thanks to John Piefer, co-founder of InteleHealth, Jeff Wilson and Scott Robertson from the Interactive Media Technology Center, and Roberto Casas from the Georgia Tech Venture Lab; it has been a real pleasure working with them.

Thanks to the administrative staff in Dr Jayant's office – Barbara Satterfield, Tina Clonts, Rex Smith, and Joanna Shorter, for being ever so helpful and cheerful.

I am fortunate to have worked with two esteemed researchers – Prof David Malah of Technion, Israel, and Dr Zhongkang Lu of I^2R, Singapore. Working with them gave me a broader perspective of research areas in relation to my specific topic.

I would like to make a special mention of my aunts and uncles – Viji Mausi, Lopa Mausi, Mr. Guruprasad, and Jayaram Mama, and also my little sister Anjana, for their love and affection.

Finally, I would like to express my gratitude to my friends - Nitin Suresh, Yogesh Sankarasubramaniam, Souvik Dihidar, Babak Firoozbakhsh, Jeannie Lee, Rajesh Narasimha, Arumugam Kannan, Mayur Chhabra, Navin Viswanath, Gaurav Arora, Bhagat Kota, Karthik Naig, Hemant Sharma, Shirish Srinivas, Vasudev Narayan, Ashvin Lakshmikantha, Sanjeev Kanekal, and Prasad Subraveti, for significantly shaping my personal as well as professional life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Video is the most demanding modality from the viewpoints of bandwidth, computational complexity, and resolution. Thus, there has been limited progress in the field of mobile video technology. In the research, the focus is on elastic wireless video technology, and its adaptation to diagnostic application requirements in real-time clinical assessment. It is important and timely to apply wireless video technology to real-time remote diagnosis of emergent medical events. This premise comes from initial successes in telehealth based on wired networks. The enablement of mobility (for the physician and/or the patient) by wireless communication will be a next major step, but this advance will depend on definitive and compelling demonstrations of reliability. Thus, an important goal of the research is to develop a complete methodology that will be embraced by physicians. Acute pediatric asthma has been identified as a domain where this new capability will be highly welcome.

The research uses flexible and interactive algorithms for Region-of-Interest (ROI) processing. ROI processing is a useful approach to achieve the optimal balance in the quality-bandwidth tradeoff characteristic of visual communication services. The notion of ROI has been traditionally used mostly for foreground-background separation in scene rendering and manipulation, and only more recently for variably quality compression. Even when the latter goal is considered, quality criteria have been ad-hoc and at best useful for video conferencing, given that the medical domain has its own fidelity criteria. The research thus focuses on the design of an elastic ROI-based compression paradigm with medical diagnosis as a central criterion.

The research describes the methodology to achieve elasticity through rate control algorithms at the encoder. An elastic non-parametric approach is proposed that uses a priori user-specified video quality information, quantifies this information, and incorporates this into the encoder in the form of region-quality mappings. This method is compared to a parametric bit allocation approach that is based on region-features and a set of tuning weights. A number of videos of actual patients were filmed and used as the video database for the developed algorithms. In testing the elastic non-parametric and parametric algorithms, both objective measures – in the form of Peak Signal to Noise Ratio (PSNR), and subjective evaluations were used. Thus, in this work, the focus is on domain relevance of the algorithms developed, as opposed to

network related issues such as packet losses. This is justified in that these may have broader value with other applications, and continuation of this work will include realistic network conditions.

To summarize, the research shows the usefulness of ROI processing as a means of achieving a gain (in a bits per pixel sense) over uniform compression at the same bitrate. It also shows how quantifying a notion of functionally lossless video quality – diagnostically lossless video quality in a video-based telehealth system, in a bits per pixel sense is useful from an applications and bitrate perspective. Finally, the research shows how a combination of these two concepts to realize diagnostically lossless ROI video quality, is a viable enabler of mobile medical assessment achievable over bitrate limited wireless channels. The result of the research is to be regarded as an important proof-of-concept in a challenging interdisciplinary area. This thesis lays the scientific foundation for additional validation through prototyped technology, field testing, and clinical trials.

CHAPTER I

INTRODUCTION

The focus of this research is to design an elastic video compression system that is robust to limitations and variations in bandwidth over large classes of wireless networks, and provides for flexibility and adaptivity to video quality requirements. Video is the most demanding modality from the viewpoints of bandwidth, computational complexity, and resolution. Thus, there has been limited progress in the field of mobile video technology. In the research, the focus is on elastic wireless video technology, and its adaptation to diagnostic application requirements in real-time clinical assessment.

It is important and timely to apply wireless video technology to real-time remote diagnosis of emergent medical events. This premise comes from initial successes in telehealth based on wired networks. The enablement of mobility (for the physician and/or the patient) by wireless communication will be a next major step, but this advance will depend on definitive and compelling demonstrations of reliability. Thus, an important goal of the research is to develop a complete methodology that will be embraced by physicians. Acute pediatric asthma has been identified as a domain where this new capability will be highly welcome. The designs resulting from this work, while directly impacting that domain, will also extend to other applications such as stroke assessment, ultrasound analysis and mammogram interpretation.

The research uses flexible and interactive algorithms for Region-of-Interest (ROI) processing. ROI processing is a useful approach to achieve the optimal balance in the quality-bandwidth tradeoff characteristic of visual communication services. The notion of ROI has been traditionally used mostly for foreground-background separation in scene rendering and manipulation, and only more recently for variably quality compression. Even when the latter goal is considered, quality criteria have been ad-hoc and at best useful for video conferencing, given that the medical domain has its own fidelity criteria. The research thus focuses on the design of an elastic ROI-based compression paradigm with medical diagnosis as a central criterion.

As mentioned, elasticity refers to non-uniform coding of different scene segments (for a given total bitrate (TBR)) as well as adaptivity to different TBR budgets (as directed by wireless network conditions). Although much of the research will focus on a single ROI, the prescribed designs will extend

to the case of multiple ROI, as well as to the notion of an Extended ROI (EROI), which is an intermediate region between the ROI and the background (BKGRND). The ROI-EROI-BKGRND framework allows for graceful quality management, as opposed to an abrupt quality degradation obtained with a conventional ROI-BKGRND framework.

The research describes the methodology to achieve elasticity through rate control algorithms at the encoder. An elastic non-parametric approach is proposed that uses a priori user-specified video quality information, quantifies this information, and incorporates this into the encoder in the form of region-quality mappings. This method is compared to a parametric bit allocation approach that is based on region-features and a set of tuning weights. Both schemes will be applied to the medical application to test for two plausible models for clinical acceptability.

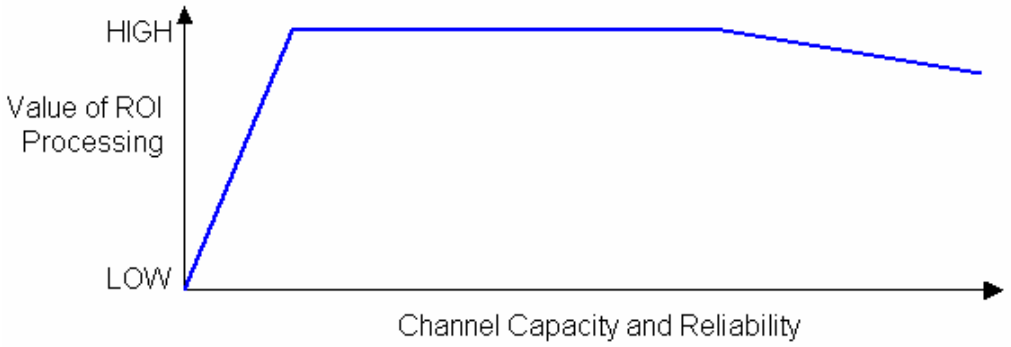


Figure 1. Impressionistic view of the expected value of the proposed research, as a function of wireless channel capacity and reliability: When the channel capacity is very small, the value of ROI processing is small as well, because there is not sufficient capacity in the channel to leverage the benefits of ROI processing. However, the value increases as indicated by the large slope in the initial part of the graph. There is a broad "sweet spot" for intermediate channel capacities where the proposed research promises high value. At very high channel capacities, the value of ROI processing starts to reduce, albeit slowly, because the channel capacity is large enough to support high quality communications by allocating high quality to all pixels in the scene, both ROI and background.

Figure 1 depicts a range of situations where the research offers the greatest value. When the wireless channel is narrowband (the left end of the abscissa), video communications quality will be low, even with the use of ROI processing. When the wireless channel is broadband (the right end of the abscissa), video communications quality will be good enough not to need ROI processing. In the intermediate range (100 to 1000kbps, packet loss rates on the order of 1%), ROI processing is expected to result in visual communications at a quality that will enable services like mobile assessment.

To summarize, the research shows the usefulness of ROI processing as a means of achieving a gain (in a bits per pixel (bpp) sense) over uniform compression at the same bitrate. For example, if the ROI is coded at 0.4 bpp and occupies ¼ of the frame area, and the BKGND is coded at 0.1 bpp and occupies ¾ of the frame area, the average bpp for the complete frame is 0.175bpp. In other words, in a bpp sense, there is a gain of 2.3 relative to uniform compression at the same bitrate. It also shows how quantifying a notion of functionally lossless video quality – diagnostically lossless video quality in a video-based telehealth system, in a bits per pixel sense is useful from an applications and bitrate perspective. This is because uncompressed color video requires 12 bpp (with 8 bpp for luminance and subsampled color components), mathematically lossless color video requires about 4-6 bpp, and coding video at the lowest level of fidelity typically requires 0.1 bpp. Thus, there is a wide intermediate range that corresponds to bpp requirements for various applications, but these have not been quantified. In other words, it is unclear whether the requirement is closer to the lowest level of fidelity, or to the mathematically lossless level. This work intends to determine the answer to this question for the specific application of mobile telehealth. Finally, the research shows how a combination of these two concepts to realize diagnostically lossless ROI video quality, is a viable enabler of mobile medical assessment achievable over bitrate limited wireless channels. The result of the research is to be regarded as an important proof-of-concept in a challenging interdisciplinary area. This thesis lays the scientific foundation for additional validation through prototyped technology, field testing, and clinical trials.

The rest of the document is organized as follows. Chapter 2 develops the background on video compression, video transmission over a broad class of networks, the usefulness of video in remote medical assessment, the motivation for ROI based video compression, and lists some key works in each of these areas. Chapter 3 provides the fundamentals of video rate control, specifically with respect to the MPEG-2 standard. Chapter 4 describes the proposed elastic non-parametric bit allocation algorithms. Chapter 5 describes parametric bit allocation algorithms. Chapter 6 describes the performance results by comparing the elastic non-parametric and parametric bit allocation approaches. Chapter 7 summarizes the work and concludes the thesis by providing insight into possible future work.

CHAPTER II

BACKGROUND

2.1 Wired and Wireless Broadband Communications

The term broadband in the context of data communications refers to high rate data transmissions. There are a variety of broadband services today, but they may be broadly divided into four main categories – wired, wireless, satellite, and fiber. Table 1 summarizes the important technologies within each of these categories, their offered bit rates, their advantages, and their limitations.

Table 1. Summary of broadband communications technologies – bit rates, advantages, and limitations.

	Broadband Communications			
	Technologies	**Bit rates**	**Advantages**	**Limitations**
Wired	DSL	768kbps, 3Mbps	High bit rates, Independent of number of users	Infrastructure costs, No mobility
	Cable	3-30Mbps	High bit rates	Infrastructure costs, Depends on number of users, No mobility
Wireless	WLAN, WiFi	1-54Mbps	Decreased deployment costs, mobility	Range limited to tens of meters Depends on number of users Poor bit rates, Packet loss issues
	3G Mobile	384kbps	Decreased deployment costs, True mobility	Limited range
	3G Stationary	1-2Mbps	Decreased deployment costs, mobility	
Satellite		1-24Mbps	High bit rates	Infrastructure and operational costs, Latency
Fiber	FTTH	10/100/1000Mbps	Very high bit rates	Infrastructure costs, No mobility
	FTTN	25-30Mbps	High bit rates	Infrastructure costs, No mobility

Wired, satellite, and fiber based broadband systems offer very high bit rates at the price of increased infrastructure costs. Thus, the major push towards wireless communications is because of the mobility advantage and decreased deployment costs compared to broadband systems. However, the drawback is that offered bit rates are much lower.

2.2 Video Communications and Region-of-Interest (ROI) Processing

A video signal is spatio-temporal, and its native richness is defined by temporal resolution (number of frames per second) and spatial resolution (number of pixels per frame). For digitization, each pixel (in each of typically three color axes) is typically quantized to an accuracy of 256 intensity levels or $\log_2(256) = 8$ bits per pixel (bpp) (along each color axis), or 24 bpp (for full color video). In the so-called Common Intermediate Format (CIF), the spatial resolution is 352×288 = 101376 pixels per frame. Thus, with 30 frames per second (fps) and approximately 100000 pixels per frame, the video bit rate is 30×100000× (8×3) = 72 Mbps. If a compression algorithm can compress the video by a factor of 72, the bit rate (for transmission and storage) would be 1 Mbps. The bit rate per pixel would be 3×8/72 = 0.33 bpp. This is indeed the state of the art for video coders used for videoconferencing. Standard definition (SD) and high definition (HD) formats for entertainment video use higher spatiotemporal resolutions than CIF. Table 2 summarizes the essential features of some of the state-of-the-art video compression standards – supported bit rates, intended applications, and compression levels assuming a CIF video format at 30fps.

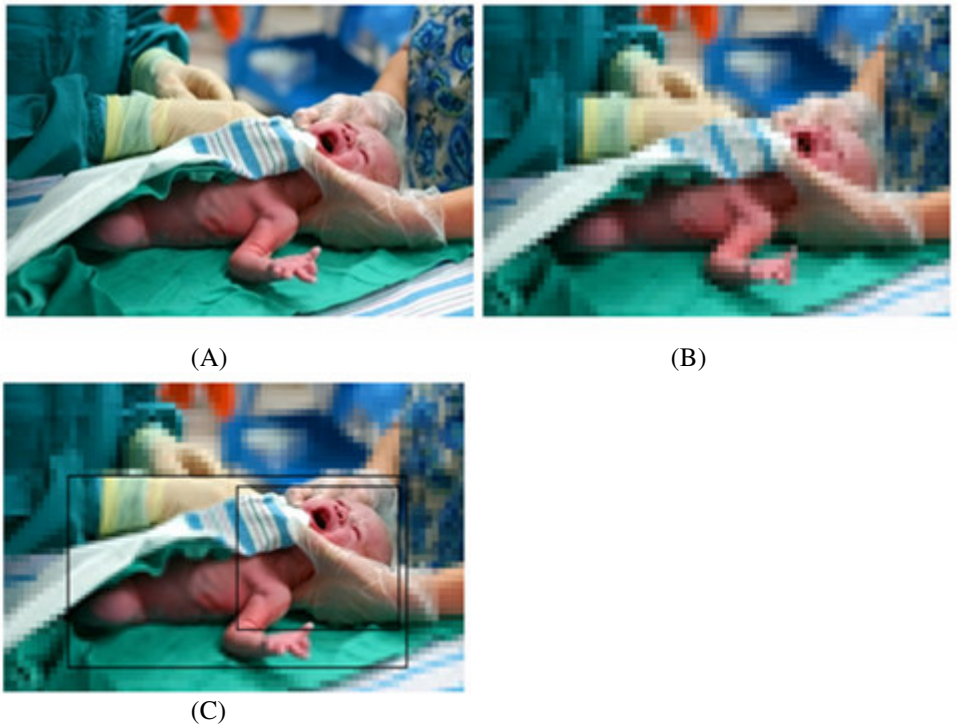
Table 2. Summary of digital video compression standards – bit rates, applications, and compression factors.

Video Compression Standards	Supported Bit Rates	Applications	Compression Factors*
H.261	40kbps-2Mbps	Videoconferencing	36-1800
H.263,H.263+	30kbps-0.5Mbps	Videoconferencing	2400
MPEG-1	1.4Mbps	Video CD	50
MPEG-2	384kbps-30Mbps	Broadcast, Storage	2-180
MPEG-4	64kbps-240Mbps	Broadcast, Storage, Videoconferencing, Video streaming etc.	1-1800
H.264	64kbps-240Mbps	Broadcast, Storage, Videoconferencing, Video streaming etc.	1-1800

* For CIF at 30fps

Region of Interest Processing: Pixels within a region of interest are allocated a disproportionate value of bit rate while allocating a lower bit rate (in bpp) for pixels not in the ROI. The proposed ROI visual communications algorithm does *not* alter the native resolution of the video scene. Figure 2 depicts the ROI principle in a qualitative pediatric scenario. Part A shows perfect video coding with unconstrained bandwidth. Image quality is identical to that of the uncompressed original. Part B shows the consequences of constrained communication bandwidth using the standard procedure of equal allocation of available coding resources (bits) across the entire image. The image is uniformly distorted. In Part C, ROI coding

(the smaller rectangle) enhances the quality of the infant's face at the cost of sacrificing the quality of the background. It also includes an extended ROI (the larger rectangle) which is of better quality than the background but not lossless. The total bit rate is the same in B and C.



(A) (B)

(C)

Figure 2. Illustration of quality management through ROI methods. (A) Standard encoding at unconstrained channel bandwidth; (B) Standard encoding at constrained channel bandwidth; (C) ROI encoding (ROI is smaller rectangle) with high (possibly lossless) quality and Extended ROI (represented by the larger rectangle) with high quality, at the expense of background quality.

2.3 Notion of Functional Quality

In spite of the existence of several objective metrics to measure video quality, there is no standardized metric to measure video quality. Metrics like Peak Signal-to-Noise Ratio (PSNR), Just Noticeable Distortion (JND), Structural Similarity (SSIM) Index etc. are frequently used to compare the performance of different video algorithms. Some of these metrics are even take into account Human Visual System (HVS) properties such as luminance and texture masking i.e. they are perceptually tuned. However, neither of these objective metrics correlates perfectly with a viewer's perception of video quality. More importantly, video quality requirements are very diverse depending on the particular application. For example, a medical imaging application may require very specific detail in the image to be rendered with absolute lossless quality. On the other hand, a remote medical assessment application may be tolerable to

some compression in the video. Likewise, video surveillance may pose its own quality requirements. Thus, it is appropriate to speak of functionally lossless video quality i.e. video quality that is perfect for the particular application. In this work, it is our goal to perform ROI coding to achieve this level of video quality. Specifically, since our focus is on remote medical assessment, the goal is to obtain diagnostically lossless (DL) i.e clinically acceptable video quality.

2.4 Telehealth Systems based on Broadband Video

Clinicians commonly find themselves simultaneously committed to one location but needed in another. This situation can arise [a] within a hospital – in the Intensive Care Unit attending a critically ill patient and called to the Emergency Room to evaluate a patient or assist a trainee, [b] within a community – at work in one hospital and needed urgently to evaluate a patient at another hospital, [c] regionally – possessed of unique expertise and practicing in an urban, possibly academic, medical center and called by a colleague in a rural community to make an urgent assessment and recommendation, [d] practicing in a rural or urban multi-office setting – practicing at one office location when a patient presents unexpectedly at another office location, and [e] at home, either off-duty or on-call when called upon to make an assessment and give recommendations for initial, interim care until necessary expertise can arrive on-site. These situations are extremely common, are far from trivial, and are encountered daily by most clinicians, whether generalists or specialists. Under the best of circumstances they represent merely an inconvenience for the patient and clinician. Usually, however, these situations delay diagnosis and management, can lead to unnecessary consequences, and can even result in inappropriate initial care. Taken as a whole, the result is inefficiency and increased cost. When individual cases are examined, however, the outcome can be unintended, but nevertheless unmistakable, poor care. The only viable solution to this dilemma is to network resources so that access to needed expertise is available wherever the need is encountered and wherever the expertise happens to be located at the time. Such a system is known as a Telehealth system.

Figure 3 illustrates the concept of a telehealth system based on broadband video. A patient (a baby, in the figure) in a hospital is being attended to by local doctors, whereas a medical expert is in a different physical location. In order to carry out real-time remote medical assessment, video of the baby is captured by an acquisition device (e.g. a video camera), and compressed (the camera may have an inbuilt video encoder to perform this function). Then, the video is transmitted over a broadband network, which

could be wired, wireless, fiber, or satellite. At the client side, there is a video decoder that performs video decompression and a display device that displays the output video for the medical expert.

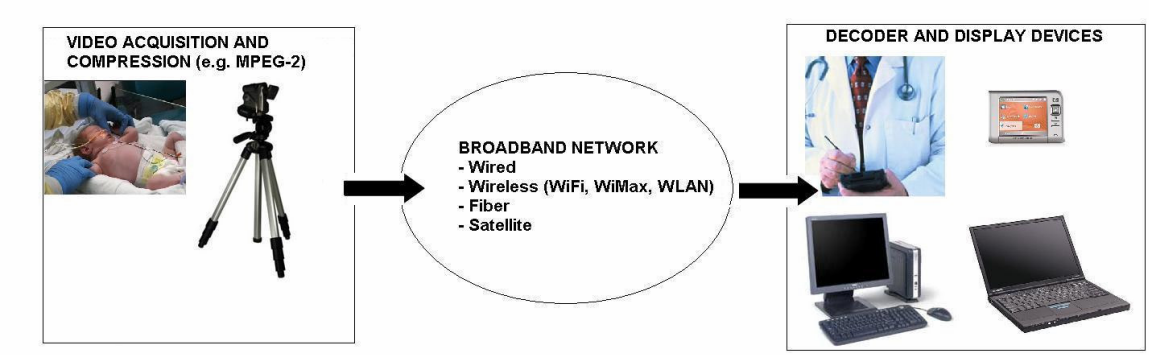


Figure 3. Telehealth systems based on broadband video communications.

Telemedicine systems operating over broadband networks: Stamford, et al [1] address high-speed, high-quality video transmitted from a rural emergency department (ED) to a major medical center ED. The results document improved diagnosis and treatment, as well as improved confidence levels of doctors. Kofos et al [2] studied telemedicine in pediatrics using a broadcast quality real-time audiovisual system and concluded that such a system may have dramatic implications for providing pediatric specialty and subspecialty care in underserved areas.

An inter-hospital system using a wide area network (WAN): Yoo [3] reported the design of an MPEG-2 video system running at 30 frames per second (fps) and requiring 1.5-6 megabits per second (Mbps) to deliver a spatial resolution of 640×480 pixels. Qiao et al [4] addressed the design of a critical care telemedicine system based on video over broadband IP networks. Wang et al [5] described a web-based videoconferencing system (REACH) that allows specialists at an academic medical center to evaluate stroke patients at a rural facility.

While these systems achieve high quality video, sufficient for remote medical assessment, they impose constraints on patient and specialist mobility. Further, rural hospitals frequently have neither the equipment nor personnel infrastructure necessary to support high bandwidth broadband networks. The infrastructure costs and mobility issues of conventional broadband systems motivate telehealth systems based on wireless video.

2.5 Telehealth Systems based on Wireless Video

In today's wireless networks, we are witnessing a convergence of modes that were traditionally disparate: an outdoor mode characterized by high mobility, large coverage area and range and low transmission rates, and an indoor mode characterized by lower mobility, lower coverage area and range, and higher transmission rates.

The so-called 3G cellular radio systems are aiming for a 144 kbps rate while driving and a 2 Mbps rate indoors. Cellular transmission rates that are available pervasively are still in the 100 kbps range. The so-called WiMax and WiFi systems that compete with cellular radio are aiming for ever-greater ranges (several miles or tens of miles in the case of WiMax (fixed wireless) and up to 500 feet in the case of WiFi (mobile or portable wireless), each at rates of several Mbps.) Systems that are currently pervasive are limited to WiFi bubbles in homes, offices and selected urban settings.

Classification wise, wireless was also grouped under the class of broadband networks. Here, however, we treat them separately from the standpoint of mobility and deployment costs. In addition, bit rates for wireless systems are far lower compared to wired, satellite, and fiber based systems. Figure 4 illustrates the concept of a telehealth system based on wireless video. The scenario is exactly the same as in Figure 3 except that the communication network is now a wireless network.

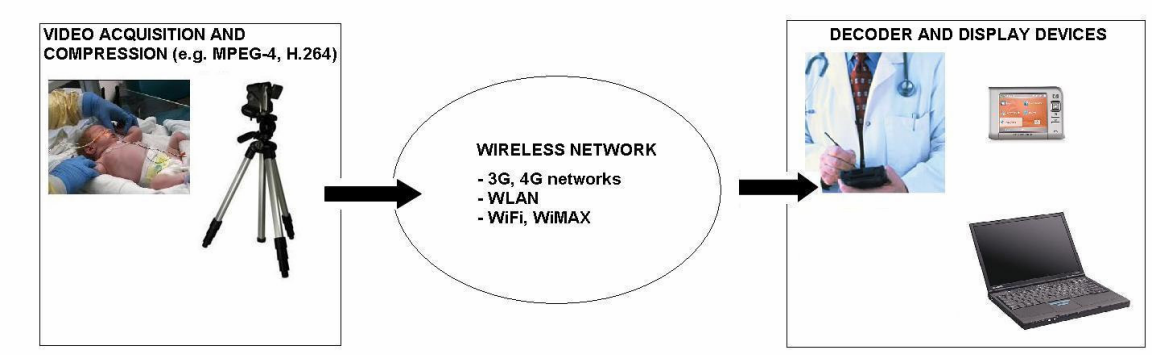


Figure 4. Telehealth systems based on wireless video communications.

Wireless networks: A number of researchers have explored telemedicine systems over wireless networks. Kugean et al [6] discuss a telemedicine system design using a wireless local area network (WLAN). H.261 encoded video is transmitted with FCIF/QCIF spatial resolution at 30fps, requiring about 384kbps. This bandwidth falls well within the WLAN's capacity, but because bandwidth is shared with

other network users, the bandwidth per user varies both with user number and time, with video quality often dropping to levels that create problems for physicians analyzing critical real-time patient features. Banitsas et al's [7] design faces similar issues. Chu et al [8] describe a mobile teletrauma system based on 3G networks where M-JPEG video at 2-25fps with 320×240 resolution is used. The CDMA providers advertised a 153kbs bandwidth, but in fact the average available bit rate was only 50-60kbs. Thus, in these types of systems, limited and varying bandwidth availability are critical issues that are not compatible with clinical uses where the bandwidth must be consistently available and unvarying throughout the encounter.

2.6 Telehealth Systems based on ROI Video

As mentioned, on the one hand, the available transmission bandwidth may place a constraint on the overall compression ratio of the video. On the other hand, for diagnostic purposes i.e. for clinical acceptability, it is essential that the compression process cause no tangible loss of detail and introduces no noticeable artifacts which could otherwise be misinterpreted as being pathological in nature. Region-of-Interest (ROI) based video processing is a useful approach to achieve the optimal balance in the quality-bandwidth tradeoff. The pixels within the ROI are allocated a disproportionately large value of bit rate (in bpp) while allocating a lower bit rate (in bpp) for pixels in the background (BKGRND), for a given constraint on total bit rate (in bits per second: bps, kbps or Mbps). This leads to an ROI with higher quality than the BKGRND. Figure 5 illustrates the concept of a telehealth system based on ROI video over a wireless communications system. The system is identical to Figure 4 except that the encoder is now an ROI encoder, capable of performing the necessary differential bit allocation between the ROI and BKGRND.

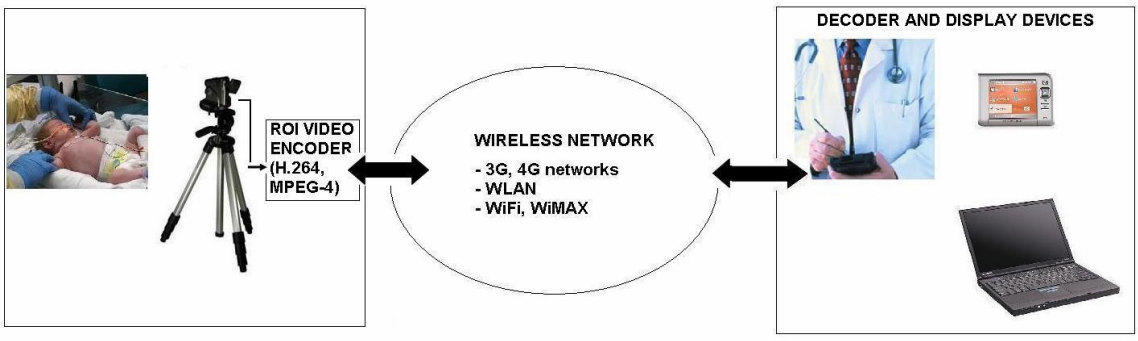


Figure 5. Telehealth systems based on ROI video over a wireless communications system.

Descriptions of several ROI based designs, and also those that have telemedicine systems based upon them have been published. Wong et al [9] proposed an ROI-based channel-adaptive source-coding scheme for wireless channels transmitting H.263 encoded video. Based on the channel state information, the channel bandwidth is computed, and both ROI and background (BKGRND) areas are treated with varying compression ratios in each region. The BKGRND is dropped altogether when necessary.

Chai et al [10] proposed and implemented two ROI coding strategies with the H.261 standard. The maximum bit transfer (MBT) strategy assigns the highest compression level to the BKGRND, and the lowest possible compression level to the ROI that does not result in exceeding the overall available bitrate. The joint bit allocation method allocates bits to the ROI and BKGRND based on the size, motion, and priority characteristics of each region. Chen et al [11] proposed a face detection algorithm integrated into a H.263+ encoder. ROI coding is performed by increasing the distortion weights of ROI macro blocks (MB). Similar to [10], Sun et al [12] propose using size, variance, and weights to perform bit allocation, where the weights are updated based on PSNR difference between ROI and BKGRND. In [13], Lai et al use region-weighted rate-distortion (RD) models. Region distortion is modeled as being directly proportional to a region's weight, the residue variance, and as having an inverse exponential relationship with the bits allocated to the region. Then, Lagrangian RD optimization results in a closed form expression for the bits to be allocated to each region based on the above parameters. In [14], a user specifies the percentage of total bits to be allocated to the ROI. In [15], optimal bit allocation for multiple ROIs is achieved using a weighted Lagrange multiplier technique. In [16], Lin et al consider a videoconferencing application with H.263 video encoded using the TMN-8 rate control scheme, and make decisions on which macroblocks may be skipped, and describe how the bits saved are reallocated to non-skipped macroblocks. In [17], Liang et al perform Lagrangian RDO to arrive at expressions for optimal bits to be allocated to macroblocks. It is a function of the macroblock weight and variance. In addition, a skipping condition is described, and saved bits are reallocated. In [18], Sengupta et al adopt a different strategy wherein they minimize the rate subject to a distortion constraint as opposed to minimizing the distortion subject to a rate constraint. Based on this new Lagrangian optimization, bits are allocated to regions to meet the required distortion constraints.

Gokturk et al [19] proposed a hybrid compression scheme for 3D medical images where ROI is coded in a *lossless* manner, and the BKGRND is coded in a *lossy* manner. Lossless compression is done using first-frame lossless coding, followed by lossless coding of successive motion-compensated frames, and unlike traditional video coders, no DCT is used. The method was tested on CT images of the human colon with the ROI being the diagnostically important colon wall. Gibson et al [20] integrated ROI detection and bit allocation integrated into a 3D wavelet compression scheme. Angiogram video sequences were used to test the proposed technique, with regions containing the coronary arteries of key diagnostic importance. The techniques in [19, 20] are proprietary.

References [21-25] address the quality of medical images or video in telemedicine systems. Martini et al [22] address the design of a quality driven video transmission system for medical applications based on joint source and channel coding, as well as metrics such as PSNR and structural distortion, again, without user interaction. Gibson et al [23] demonstrate diagnostically lossless medical images by using mathematically lossless coding within the ROI, but this cannot be extended to wireless video because of bandwidth limitations. Ashraf et al [24] obtain diagnostically lossless angiogram videos through 3D wavelet based ROI compression, but the diagnostically lossless property is checked only subjectively and is not pursued as part of the algorithm. Wu et al [25] obtain perceptually lossless medical images using a visual pruning function embedded into an advanced human visual system (HVS) model, but this also is not easily extensible to video.

In general, the aforementioned systems employ arbitrary choices for ROI and BKGRND compression ratios and thus are not human-centric. There is no guarantee of video quality, even within the ROI, and there is no opportunity for user interaction or input.

2.7 ROI Segmentation and Tracking

A crucial issue in ROI based systems pertains to segmentation and tracking of the ROI. Our framework is relatively simple as far as ROI segmentation and tracking is concerned. This is because the video camera is fixed, and the background stationary. In this case, tracking may be performed with change detection and background registration techniques. The stationary background assumption is not valid if, in addition to the patient (the ROI), there is a local doctor in the field of view who moves around, and/or nurses entering and

leaving the scene. However, the plus side is that the narrowband properties of skin-color distributions may be used to detect ROI, and tracking may be performed by relatively simple techniques such as object projection or boundary projection. Thus, ROI segmentation and tracking is not the major issue in our ROI system design. For the sake of completeness, the state-of-the-art methods for segmentation and tracking are listed and briefly described below.

Video segmentation methods may be classified with respect to the degree of human intervention involved. Automatic algorithms do not depend on any user interaction, but their implementation presents problems in recognizing and grouping semantically coherent regions into the ROI just as the human eyes do. Thus, they tend to be complex, requiring delicate fine-tuning of parameters and often constitute ad hoc approaches to specific problems. In semi-automatic methods, the user selects the ROI in the first frame using an interaction tool – mouse drawing, polygon method etc., and this is tracked by the algorithm in the following frames. In [26,27], a semi-automatic algorithm for ROI tracking is presented, where the ROI boundary is projected from frame-to-frame using motion information. Uncertain areas are obtained on the boundary based on the regions in conflict by comparing the above results with the results from color information. Finally, the boundary is refined through region growing algorithms to get a precise ROI boundary. In [28], tracking is performed by taking inter-frame differences to obtain regions with motion, performing edge extraction, and finally extracting the ROI in the current frame. In [29], a semi-automatic object segmentation algorithm is proposed that aims at minimizing user assistance by requiring it at the final step of the segmentation process, as opposed to the initial step. This makes it easier to identify objects. For tracking, a single displacement vector for the object is obtained from frame to frame using horizontal and vertical histograms of the object in the previous frame and several candidates in the current frame. In [30], inter-frame differences are used to generate a change detection mask, which supplements a skin detection mask generated using a bivariate normal distribution for skin color. The two masks are fused to generate a face and hands segmentation mask. In [31], ROI segmentation and tracking is performed with neurosurgical video, where the ROI is a field within which surgical procedures takes place, not a single object. Thus, it involves both surgical instruments and a surrounding area, consisting of objects such as biological tissue and fluids. The edge of a selected instrument in the video is computed and utilized as an

input to a histogram-based tracking algorithm that provides a crucial location, such as the tip of the instrument. An ROI is defined around this location and tracked.

CHAPTER III

VIDEO RATE CONTROL

3.1 Video Encoder Operation

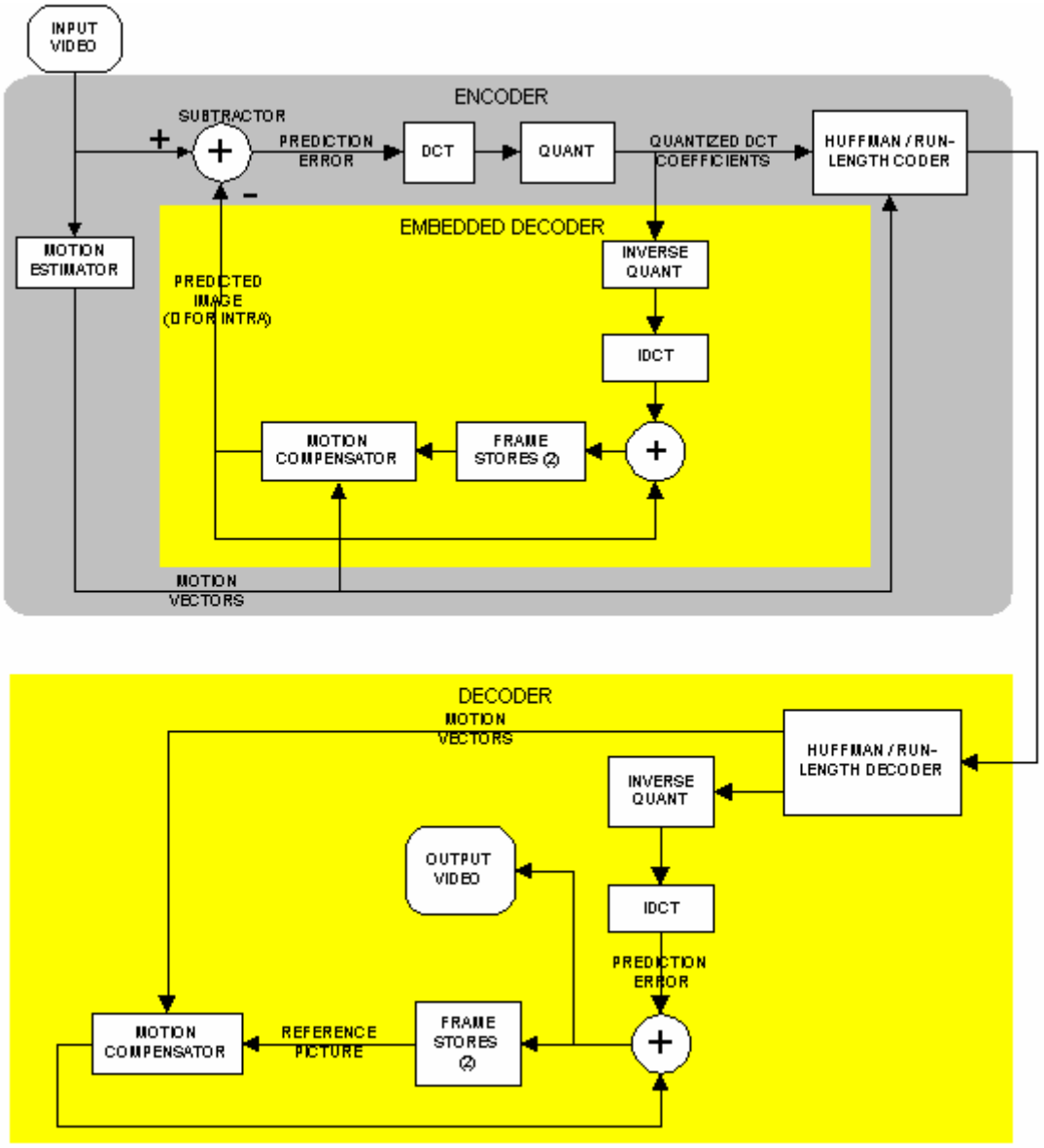


Figure 6. Block diagram of a typical MPEG-2 video encoder and decoder.
*http://www.zenith.com/sub_hdtv/mpeg_tutorial/codecdia1.HTM

Figure 6 shows the block diagram of a typical MPEG-2 video encoder and decoder. The encoder is essentially a DPCM like system with an embedded decoder. This is so that both the encoder and decoder

use the same reference frames in generating the prediction. In the encoder, the motion estimation block receives a macroblock in the current frame and the reference frame(s) candidate macroblocks as its inputs. It generates the optimum motion vectors corresponding to that source macroblock. These motion vectors and the reference frame(s) candidate macroblocks are the inputs to the motion compensation unit, which generates the prediction for the source macroblock. If the source macroblock is from an I (intra) frame, there is no prediction i.e. a zero prediction. If the source macroblock is from a P frame, the prediction is generated by merely selecting the macroblock in the reference frame using the motion vectors. If the macroblock is from a B frame, then the prediction could either correspond to an average of predictions from two reference frames or just one of the two reference frames, depending on which results in the lowest prediction error.

The prediction is subtracted from the source macroblock, resulting in the prediction error. This is then passed through a 2-D DCT, which results in further decorrelation, and also energy compaction. A quantization unit reduces the amplitude resolution of the DCT coefficients. The output from this stage feeds the embedded decoder consisting of an inverse quantization unit, a 2-D IDCT, and an adder. This generates the reference frames for prediction of future frames. The output from the 2-D DCT is also reordered using a zigzag scan (or an alternate scan for field pictures in interlaced video), and then converted to (run, level) pairs which are encoded to bits using a variable length coder (VLC), such as a Huffman coder. Similarly, the motion vectors are also Huffman coded.

3.2 Introduction to Video Rate Control

The video encoder output above is a bitstream. Clearly, with different types of frames – I, P, and B, and with different types of video content, the output bitstream will be at a variable bitrate. This is illustrated in Figure 7, where an uncompressed source results in a compressed video bitstream at a variable bitrate. The quantization parameter (QP) is shown explicitly in this Figure because it is a variable that can control then number of bits generated.

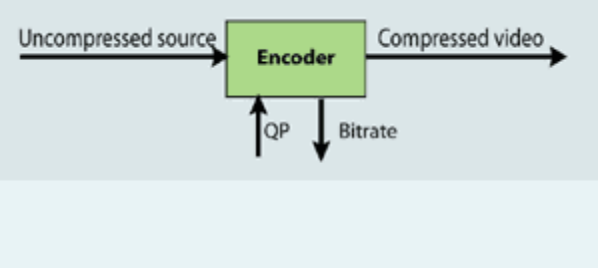


Figure 7. Illustration of a video compression system without rate control.

In practice, an encoder has to meet the channel constraints on the available total bit rate for transmission. Rate control [32, 33, 34] helps achieve this equalization between an inherently variable rate video encoder and a channel bitrate constraint using two key features: a) quantization parameter control in the quantization unit, b) an encoder buffer to buffer encoded video data before channel transmission. This is shown in Figure 8, where a rate controller operates in a negative feedback mode with respect to the channel's demanded bitrate and the actual generated bitrate, thereby controlling QP under a given source complexity estimate.

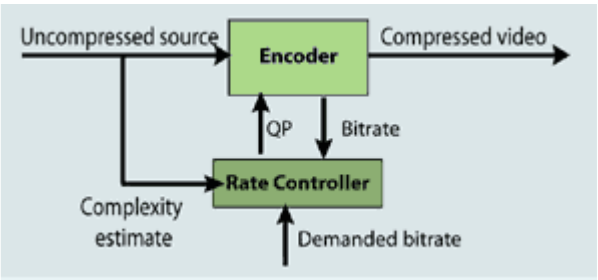


Figure 8. Illustration of a video compression system with rate control.

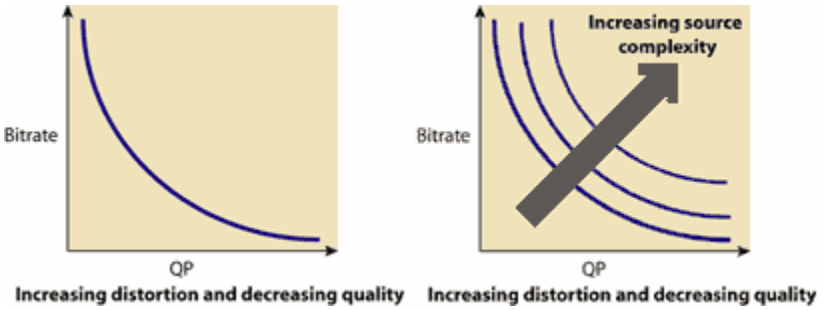


Figure 9. (a) Illustration of relationship between QP and average video bitrate. (b) QP-bitrate curves for different video source complexities.

Figure 9a shows a basic relationship between QP and *average* video bitrate. As expected, with increasing QP, the average bitrate decreases. Figure 9b shows the role played by video source complexity. The more complex the video content, the higher the average bitrate for a given QP.

3.3 MPEG-2 TM5 Rate Control

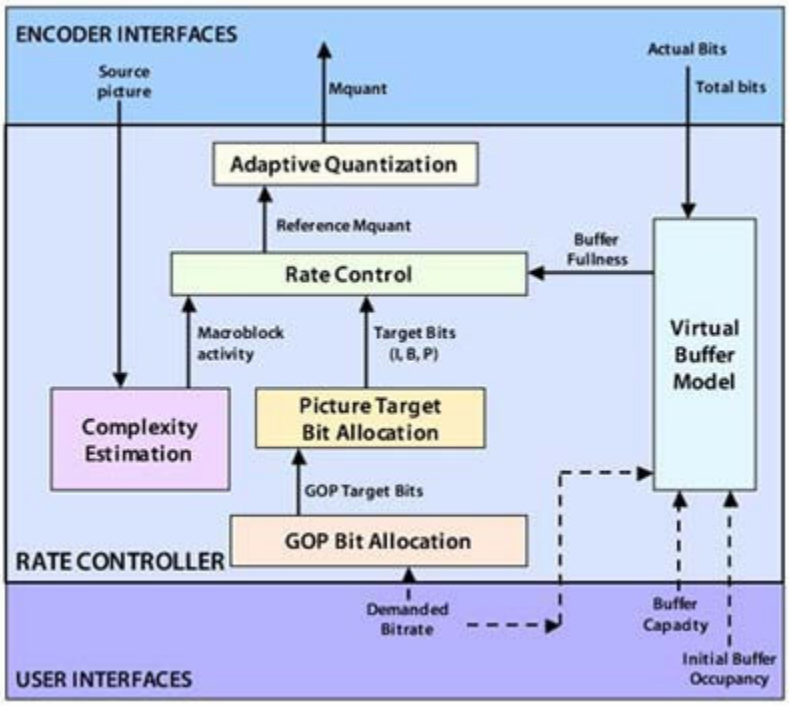


Figure 10. Block diagram representation of the TM5 rate control methodology.
* Source: www.pixeltools.com/figure6.jpg

Figure 10 shows the block diagram representation of the TM5 rate control methodology [35, 36, 37, 38] popularly used with the MPEG-2 standard. For ease of understanding, it is represented as a rate controller with two interfaces – an encoder interface, and a user interface.

The rate controller operates as follows. The Group of Pictures (GOP) bit allocation unit receives the demanded or target bitrate from the user interface. A GOP is a set of frames that consists of exactly one I frame, followed by P and, possibly, B frames. A typical GOP is illustrated in Figure 11.

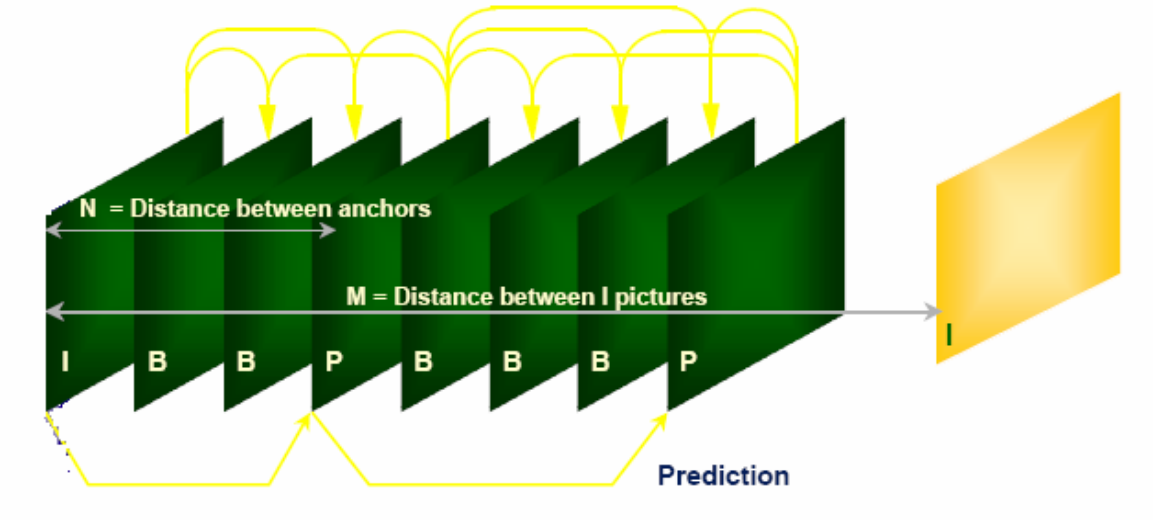


Figure 11. Illustration of a typical GOP.

The GOP allocation unit generates the target bits for the GOP, denoted by R, as follows:

$$R = R + \frac{M \times TBR}{FR\_RATE}$$

where *M* is the number of frames in a GOP, *TBR* is the total bitrate available for encoding in bits per second, and *FR_RATE* is the frame rate in frames per second. The *R* on the right hand side of the above equation is the number of bits remaining after encoding the previous GOP. Thus, this equation represents the average number of bits that will be available for encoding a GOP.

R is the input to the picture target bit allocation unit. This module generates the target bits for each picture (frame). Different picture types – I, P, and B pictures are allocated different number of target bits as follows:

$$T_I = \frac{R}{(1 + \frac{N_P X_P}{X_I K_P} + \frac{N_B X_B}{X_I K_B})}$$

$$T_P = \frac{R}{(N_P + \frac{N_B X_B K_P}{X_P K_B})}$$

$$T_B = \frac{R}{(N_B + \frac{N_P X_P K_B}{X_B K_P})}$$

where R represents the remaining number of bits in the GOP. T_I, T_P, and T_B represent the target bits allocated to I, P, and B frames, respectively. N_P and N_B represent the number of P and B frames in a GOP, respectively. There is one I frame per GOP. X_I, X_P, and X_B represent the average encoding complexities of the I, P, and B frames, respectively. K_I, K_P, and K_B are design parameters that can be used to control the allocation. Note that at the end of encoding each frame, R is updated by subtracting from it the actual number of bits S, used to encode the frame.

The logic behind the above allocation is as follows. Bits in a GOP must be divided between I, P, and B frames based on the number of such frames, their complexities, and design parameters controlling the allocation. For example, in allocating bits to the I frame, the 1 in the denominator represents the number of I frames in a GOP, the second term represents the number of P frames in a GOP and the relative complexities of the I and P frames, and the third term represents the number of B frames in a GOP and the relative complexities of the I and B frames. After allocating bits to the I frame, there remain only P and B frames in the GOP, so the denominators in the target bit formulas for P and B frames contain only two terms, corresponding to the other picture type.

The average frame complexities are obtained as follows:

$$X_I = S_I \times Q_I$$

$$X_P = S_P \times Q_P$$

$$X_B = S_B \times Q_B$$

where S_I, S_P, and S_B represent the *actual* number of bits required to encode the *previous* I, P, and B frames, respectively. Q_I, Q_P, and Q_B represent the QP, averaged over a complete frame, for the *previous* I, P, and B frames, respectively.

Next, a normalized macroblock activity measure *N_act* is computed as follows:

$$N\_act = \frac{2 \times act + avg\_act}{act + 2 \times avg\_act}$$

where *act* is the variance of the current macroblock. *Avg_act* is the average activity of the previous picture, obtained by averaging the variances of all macroblocks. Finally, the QP for the k^{th} macroblock is generated as follows:

$$QP_k = \frac{N\_act_k \times 31 \times d_k}{r}$$

$$r = \frac{TBR}{2 \times FR\_RATE}$$

$$d_k = d_0 + \sum_{j<k-1} A_j - \frac{T \times (k-1)}{N\_MB}$$

where *r*, the reaction parameter, is half the ratio of the bitrate to the frame rate. d_k is the fullness of the virtual buffer, based on a uniform model.

The encoder uses a video buffer verifier (VBV) model that represents its assumptions of the decoder buffer. This is done because the encoder has no information about the decoder's buffer, yet it should take care to avoid overflow or underflow in the decoder buffer. This is achieved by using a VBV model – choosing a VBV size B_V, ensuring that this buffer does not overflow or underflow, and then transmitting the details about this buffer along with the encoded bitstream to the decoder. Thus, in the above VBV fullness equation, d_0 is the initial buffer fullness, A_j is the actual number of bits used by the j^{th} macroblock, *T* is the target bits for the current frame, and *N_MB* is the total number of macroblocks in a frame. The VBV fullness when the k^{th} macroblock is being encoded is obtained by using a uniform model i.e. assuming that bits are equally allocated among all the macroblocks in the picture. Thus, the difference between the actual bits and the target bits according to the uniform model represents the error in the model. The encoder tries to correct or compensate for this error by modulating the normalized macroblock activity with the VBV fullness in the QP equation. So, if there is a positive error, representing a deficit in bits, the VBV fullness increases, and QP also increases. On the other hand, if there is a negative error corresponding

to a surplus in bits, the VBV fullness decreases, and QP also decreases. Figure 12 also illustrates how the VBV fullness changes as a function of the macroblock number.

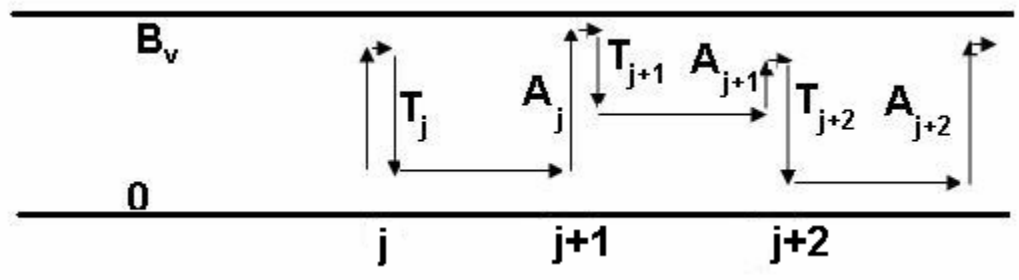


Figure 12. Illustration of VBV buffer fullness as a function of macroblock number.

CHAPTER IV

ELASTIC NON-PARAMETRIC ROI BIT ALLOCATION ALGORITHMS

Based on the motivation developed in Chapter 2, the objectives of the presented work [39, 40, 41, 42, 43] are twofold: (1) Allow for user interaction with the design of the video coding system in order to achieve user-based adaptivity and user-defined quality. (2) Make the video coding system robust for each available total bitrate (TBR), so as to result in optimal performance at each TBR. This *system elasticity* creates advantages over the aforementioned telemedicine systems in Chapter 1. It achieves this via the following steps. First, it uses pilot tests on a set of training medical videos to quantify video quality information. Second, it maps these quality levels to the ROI and BKGRND regions. Finally, based on certain in-built quality criteria, it appropriately modifies these mappings. Each of these steps is described in detail below.

4.1 Quantification of Video Quality

Four hierarchical levels of quality are postulated:

- Mathematical losslessness (ML) - there is no quantization based compression
- Perceptual losslessness (PL) - there are no perceivable artifacts in the video
- Diagnostic losslessness (DL) - there may be visually perceivable artifacts, but they do not compromise visual medical assessment
- Best effort (BE) - the video quality is not distracting or annoying

Based on the current definition of PL, it is always over-designed with respect to DL. In other words, PL always guarantees DL, but not vice versa. To quantify these levels of video quality, a set of training videos are compressed at a variety of TBR and uniform spatial quality. Physician experts are asked to identify maximum levels of compression that can be applied while retaining the aforementioned levels of quality. The training videos are presented in random order to the physicians. For completeness, one of these videos is the original uncompressed video. Clinicians complete an evaluation whose format is shown in Table 3, which in this example is specific to pediatric respiratory distress.

Table 3. Sample evaluation template for quantifying the relationship between compression and video quality levels.

Random Video Sample	Feature Sets	DL? (1-4)	PL? (Yes/No)	Comments
1-n	RR OC WB ….			

In Table 3, each sample number represents a video at a particular TBR. The physician expert lists whichever features he/she can identify in the video at that particular bitrate. The remaining three columns represent the quantification of the video quality information. In the DL? column, the four options relate to the quality of the video for clinical assessment of each feature set: 1 – No, 2 – Maybe Not, 3 – Likely, 4 – Yes. The lowest TBR at which the evaluation is 4 represents the threshold DL for the particular *feature*. Similarly, the lowest TBR at which PL? is 'Yes' represents the PL threshold for the particular *video*. In the 'Comments' column, the expert identifies any annoying or distracting artifacts in the video. This information can be used to obtain the threshold for BE for the particular *video*. Note that the above table represents one video evaluated by one particular expert. The complete test set consists of several videos evaluated by several physician experts. It is noted that ML is not required to be determined from the functional hierarchy assessment, because mathematical losslessness (ML) is a mathematically deterministic video quality level (unlike PL, DL, and BE), obtained when no quantization based compression is applied).

If TBR_DLi represents the TBR required for DL for feature i averaged over several physician experts and TBR_PL and TBR_BE respectively represent the TBR required for PL and BE, also averaged over several experts, then the corresponding bits per pixel (bpp) values denoted by bpp_DLi, bpp_PL, and bpp_BE, are:

$$bpp\_DLi = \frac{TBR\_DLi}{FR\_RATE \times FR\_SIZE}$$

$$bpp\_PL = \frac{TBR\_PL}{FR\_RATE \times FR\_SIZE}$$

$$bpp\_BE = \frac{TBR\_BE}{FR\_RATE \times FR\_SIZE}$$

FR_RATE represents the frame rate (in fps) and FR_SIZE represents the spatial resolution of the training videos (assuming one fixed value). To obtain one value bpp_DL encompassing all the features j necessary for clinical assessment, we choose the maximum value, as follows:

$$bpp\_DL = \max(bpp\_DLj)$$

Figure 13 summarizes all the steps involved in the quantification of video quality.

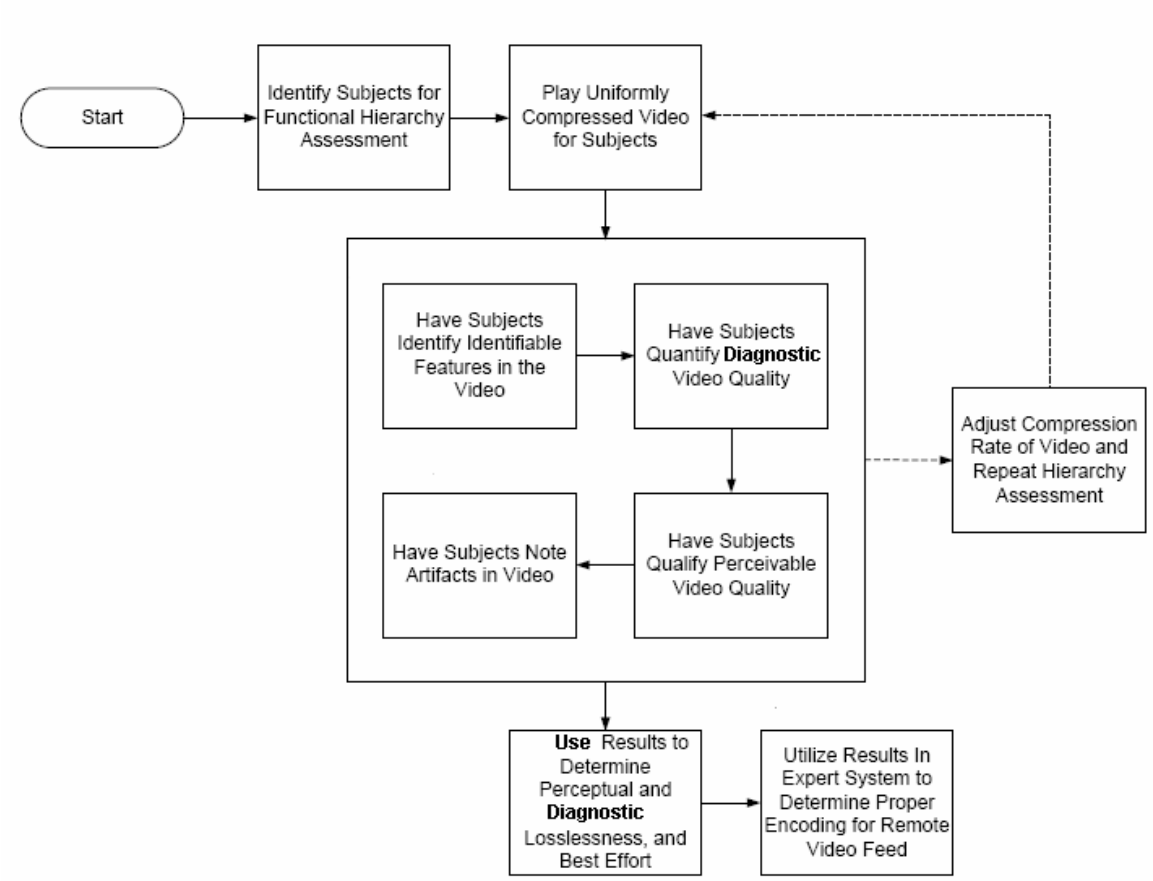


Figure 13. Summary of steps involved in the quantification of video quality.

4.2 Quality Mappings using the Encoder State Machine

The encoder must use the information on quantification of quality levels and map it to the ROI and BKGRND regions in the video. This is done using the notion of an encoder state which is defined by a pair of quality levels, one corresponding to the ROI and the other corresponding to the BKGRND. The complete state table for the encoder is shown in Table 4.

Table 4. Encoder state table showing state numbers, and expected quality levels for ROI and BKGRND.

ROI↓ **BG→**	ML	PL	DL	BE
ML	1	2	3	4
PL	nla	5	6	7
DL	nla	nla	8	9
BE	nla	nla	nla	10

The states are numbered 1 to 10 according to decreasing priority, based on the premise that the ROI quality should be at least the same as or better than the BKGRND quality. For example, state 2 represents ML quality ROI and PL quality BKGRND, and state 8 represents DL quality ROI and DL quality BKGRND. The entries denoted nla represent pairs of quality levels that are not permitted, because they violate the basic premise of the state table. It must be noted that all states in a given row represent the same quality level for the ROI.

The encoder's state of operation is determined through a GOP-level bit allocation algorithm, based on the bpp requirement for ML, and the nominal bpp values for PL, DL, and BE quality levels obtained using training videos. State determination is done at the GOP level in order to average over I, P, and B frames. For the video currently being encoded, the nominal number of bits required per GOP for DL in the ROI denoted by R_ROI_DL, for example, is:

$$R\_ROI\_DL = bpp\_DL \times ROI\_SIZE \times N\_GOP$$

ROI_SIZE is the average size of the ROI for the current GOP, and may be a window of standard shape and size around the feature(s). N_GOP denotes the number of frames in GOP. Similar expressions hold for the nominal number of bits required for other quality levels for the ROI and BKGRND.

At the beginning of each new GOP, the encoder tries to occupy the highest priority state given the current TBR. If R denotes the target number of bits in a GOP, the encoder checks if it can occupy state 1 (highest priority) , governed by the condition: R > R_ROI_ML + R_BKGRND_ML? If so, it occupies state 1, otherwise it checks if it can occupy state 2, governed by the condition: R > R_ROI_ML + R_BKGRND_PL? This procedure continues until it occupies one of the ten possible states, and is summarized in the *state occupancy flow chart* of Figure 14**.**

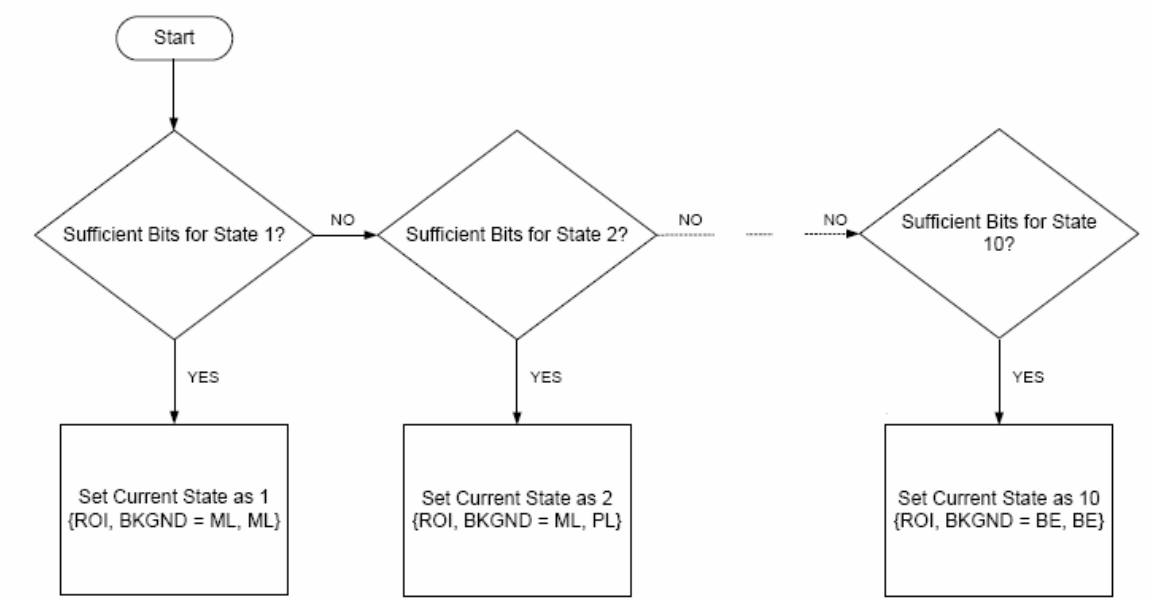


Figure 14. State occupancy flow chart illustrating conditions for occupancy of encoder states.

In each state, the BKGRND is allocated bits first, and then the remaining bits are allocated to the ROI. This is done to ensure that the ROI gets bits in excess of the nominal value. For example, in state 9, where the ROI has DL quality, and BKGRND has BE quality, the allocation is done as:

$$R\_BKGRND = R\_BKGRND\_BE$$

$$R\_ROI = R - R\_BKGRND$$

R_ROI and R_BKGRND denote the target number of bits in a GOP for the ROI and BKGRND, respectively. If the encoder determines that state 9 is the current GOP state, the BKGRND gets assigned R_BKGRND_BE bits, and the ROI gets assigned *at least* R_BKGRND_DL bits. The only exception is state 10, where all the bits are allocated to the ROI. Table 5 summarizes the state based bit allocation algorithm.

Table 5. Summary of state bit allocation algorithm.

State	Quality levels		Requirements	Bit Allocation
	ROI	BKGRND		
1	ML	ML	R > R_ROI_ML + R_BKGRND_ML	R_BKGRND=R_BKGRND_ML R_ROI=R-R_BKGRND
2	ML	PL	R > R_ROI_ML + R_BKGRND_PL	R_BKGRND=R_BKGRND_PL R_ROI=R-R_BKGRND
3	ML	DL	R > R_ROI_ML + R_BKGRND_DL	R_BKGRND=R_BKGRND_DL R_ROI=R-R_BKGRND
4	ML	BE	R > R_ROI_ML + R_BKGRND_BE	R_BKGRND=R_BKGRND_BE R_ROI=R-R_BKGRND
5	PL	PL	R > R_ROI_PL + R_BKGRND_PL	R_BKGRND=R_BKGRND_PL R_ROI=R-R_BKGRND
6	PL	DL	R > R_ROI_PL + R_BKGRND_DL	R_BKGRND=R_BKGRND_DL R_ROI=R-R_BKGRND
7	PL	BE	R > R_ROI_PL + R_BKGRND_BE	R_BKGRND=R_BKGRND_BE R_ROI=R-R_BKGRND
8	DL	DL	R > R_ROI_DL + R_BKGRND_DL	R_BKGRND=R_BKGRND_DL R_ROI=R-R_BKGRND
9	DL	BE	R > R_ROI_DL + R_BKGRND_BE	R_BKGRND=R_BKGRND_BE R_ROI=R-R_BKGRND
10	BE	BE	None	R_BKGRND=0 R_ROI=R

Figure 15 summarizes the procedure involved in elastic non-parametric video encoding.

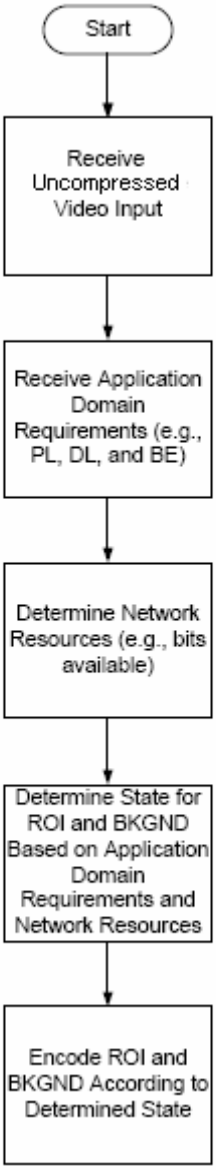


Figure 15. Summary of procedure involved in elastic non-parametric video encoding.

It is useful to carefully consider the properties of the encoder state paradigm.

1) At a given TBR, the encoder occupies a particular state, which is the highest priority under the given conditions.
2) When the TBR changes, as in a VBR channel, the encoder will likely transition to a new state at the beginning of a new GOP. Depending on the change in TBR, the encoder may move to a higher priority state or a lower priority state.

3) Occupancy of encoder states depends on the bpp values corresponding to quality levels, the ROI size, and the BKGRND size. In general, therefore, encoder states are not equidistant in terms of bits required for occupancy. This does not represent a flaw in the design because the algorithm allocates a nominal number of bits to the BKGRND, and the excess to the ROI.

4) If the TBR increases, the number of bits to the ROI increases as long as the encoder remains in the same state. If the encoder can occupy a higher priority state in the same row of the state table, then the number of bits assigned to the BKGRND increases, and so the number of bits assigned to the ROI may decrease. This is, however, not a contradiction because the ROI still gets atleast the nominal number of bits. The design of the elastic non-parametric system is based on optimizing an encoder's state, consisting of both the ROI and the BKGRND, as opposed to merely optimizing the ROI.

5) Based on the above property, it is clear that, at a given TBR, a lower priority state in a row of the state table will allocate a larger number of bits to the ROI than a higher priority state in the same row. This property is pivotal to the next section, where methods of increasing ROI quality, where necessary, are described.

4.3 Quality Update Techniques

The encoder operation is based on bit allocations to ROI and BKGRND based on nominal values obtained from the quantification of video quality step. For example, for DL, the average bpp value over the features is chosen as the nominal value. The encoder is said to have DL quality in the ROI (or BKGRND) if its current encoder state corresponds to a DL quality for the ROI (or BKGRND). However, in some cases, the ROI may not be of DL quality even if its current state indicates that it should be so. This is because DL is a subjective notion, and while it may be quantified to a fairly accurate degree using training videos, there may be imperfections. In other words, DL quality for the ROI may not indeed be achieved as far as the current viewer of the video is concerned. Similarly, PL and BE are subjective notions of quality, and may vary from viewer to viewer. There is no uncertainty involved with ML because it is mathematically deterministic. Under these circumstances, property 4 of the encoder state paradigm is modified i.e. the priority is now the ROI as opposed to the state.

ROI↓ BG→	ML	PL	DL	BE
ML	1	2	3	4
PL		5 →	6	7
DL			8	9
BE				10

Figure 16. Illustration of state transitions.

The system allocates more bits to the ROI in order to alleviate this problem. As motivated by property 5 of the encoder state paradigm, this is done by a state transition from the current state to the next lower priority state in the same row, as shown in Figure 16. In the table, the encoder transitions from state 5 to state 6. The bit allocations in these two states are as follows:

$$STATE\ 5: R\_BKGRND = R\_BKGRND\_PL$$
$$R\_ROI = R - R\_BKGRND$$

$$STATE\ 6: R\_BKGRND = R\_BKGRND\_DL$$
$$R\_ROI = R - R\_BKGRND$$

R_ROI is higher in state 6 than in state 5 because *R_BKGRND_DL* is smaller than *R_BKGRND_PL*. Transitions are not allowed to states in lower rows because the ROI quality level changes across rows. Instead, in states 4, 7, and 9, where the encoder is at the far end of a row (i.e. with BE quality BKGRND), R_ROI is increased by a parameter update i.e. by simply reducing R_BKGRND to a new, lower value R_BKGRND_BE_MIN, compared to R_BKGRND_BE. This is illustrated in Figure 17, where the new state has the same number but is denoted by a " ' ". Thus, state 7 transitions to state 7'.

ROI↓ BG→	ML	PL	DL	BE
ML	1	2	3	4
PL		5	6	7 → 7'
DL			8	9
BE				10

Figure 17. Illustration of state transitions for states at the end of rows.

Now that the issue about tackling inadequate ROI quality has been addressed, attention needs to be focused on detection of such a condition. A manual trigger may be used to indicate that video quality within the ROI is inadequate. This option is straightforward, but cumbersome. Thus, automatic methods based on objective quality metrics such as peak signal to noise ratio (PSNR), video quality metric (VQM), structural similarity metric (SSIM) etc. are necessary to measure quality and detect low ROI quality situations. In this work, two methods based on PSNR are explored. *However, it must be noted that when the*

overall system is being tested for performance, video quality will be measured using subjective tests. PSNR results will also be obtained, but only as a way of showing trends, not for qualifying video quality.

1) ROI-BKGRND PSNR difference - A PSNR difference method is used to automatically trigger a state transition flag, as follows:

$$\begin{aligned} &If\ \ ROI\_PSNR - BKGRND\_PSNR < PSNR\_DIFF\_TH \\ &STATE\_TRANSITION\_FLAG = 1 \\ &else \\ &STATE\_TRANSITION\_FLAG = 0 \end{aligned}$$

ROI_PSNR and *BKGRND_PSNR* respectively denote the average ROI and BKGRND PSNRs for the previous GOP, *PSNR_DIFF_TH* represents a difference threshold, and *STATE_TRANSITION_FLAG* is an update flag that is valid for the current GOP. This flag indicates that there must be a state transition in the current GOP in order to take care of the ROI quality problem.

The detection is based on the expectation of a certain minimal difference in quality between the ROI and BKGRND. Note, however, that *PSNR_DIFF_TH* should be state dependent. This is because a higher difference in ROI and BKGRND PSNRs should be expected in states where the ROI and BKGRND have different quality abstractions. For example, in state 7 (PL ROI and BE BKGRND), the difference between ROI and BKGRND PSNR is expected to be greater than in state 5 (PL ROI and PL BKGRND). Even within a state where the ROI and PSNR have the same quality abstraction, the PSNR difference should be expected to increase if the TBR increases i.e. when the ROI gets allocated more bits. This is also true of states where the ROI and BKGRND have different quality abstractions. The issue with this approach is that in general, the ROI and BKGRND represent different video contents. Hence, it may not be accurate to compare PSNRs of unrelated video material and draw conclusions about video quality.

2) ROI PSNR control – A given row in the state table represents a certain level of ROI quality, and this must be reflected in the ROI PSNR. Thus, a state transition flag is triggered whenever the ROI PSNR falls below a threshold *PSNR_TH*, as follows:

$$\begin{aligned} &If\ \ ROI\_PSNR < PSNR\_TH \\ &STATE\_TRANSITION\_FLAG = 1 \\ &else \\ &STATE\_TRANSITION\_FLAG = 0 \end{aligned}$$

PSNR_TH should be dependent on ROI quality levels. Thus, PSNR_TH values corresponding to PL and DL can be approximately obtained by averaging over ROI PSNRs obtained during uniform compression of the video database at TBR_PL and TBR_DL, respectively. This is justified because content-wise, the ROI is very similar in all videos, whereas the BKGRND may vary significantly.

The above procedures are depicted in Figure 18.

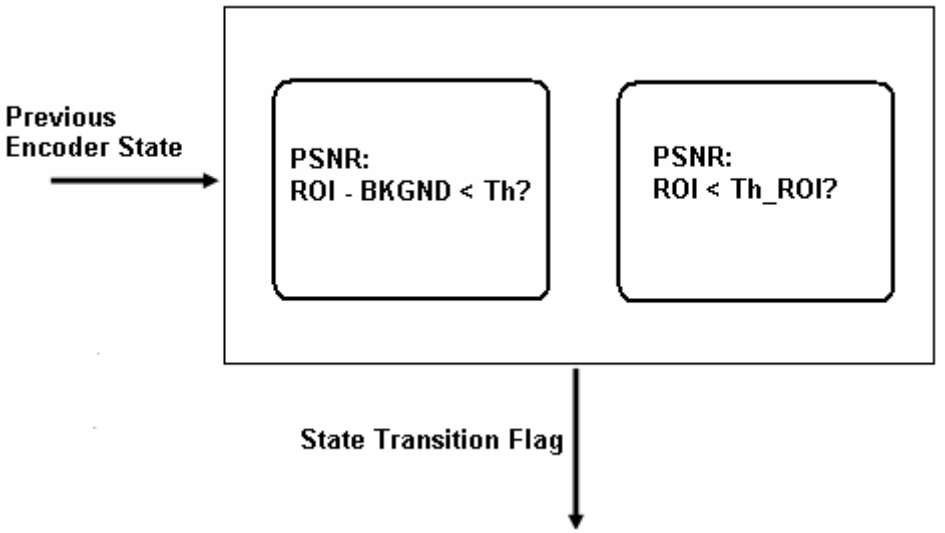


Figure 18. Detection criteria for state transitions.

4.4 Modified Encoder State Machine

The state transition flag as determined by the quality update techniques modifies the state occupancy flowchart of Figure 14. In other words, at the beginning of a new GOP, an encoder has to check the available bits with the bit thresholds of each state, and also check the state transition flag. This results in a modified methodology for determining the new encoder state shown in Figure 19.

First, the encoder uses the state occupancy flow chart of Figure 14 to determine a tentative new state, *state_n*. If the state transition flag corresponding to the previous GOP is 0, then this represents this final new state. If *state_n* and the previous state *state_n-1* are in different rows (due to TBR fluctuations), then the state transition flag is discarded, because the state transition flag is only relevant to states in the same row. The new state is *state_n*.

If, on the other hand, *state_n* is of lower priority than *state_n-1* (due to TBR fluctuations), then again the state transition flag is discarded, because a transition to a lower state has already occurred. The new state in this case also is *state_n*.

However, if *state_n* is either the same as *state_n-1*, or of higher priority than *state_n-1*, then the state transition flag is used i.e. the new state is (*state_n-1* + 1), and *state_n* is discarded. The exception to this rule is if *state_n* is the same as *state_n-1*, and they are at the end of a row. In this case, the new state is the primed state, *state_n'*, or equivalently, *state_n-1'*.

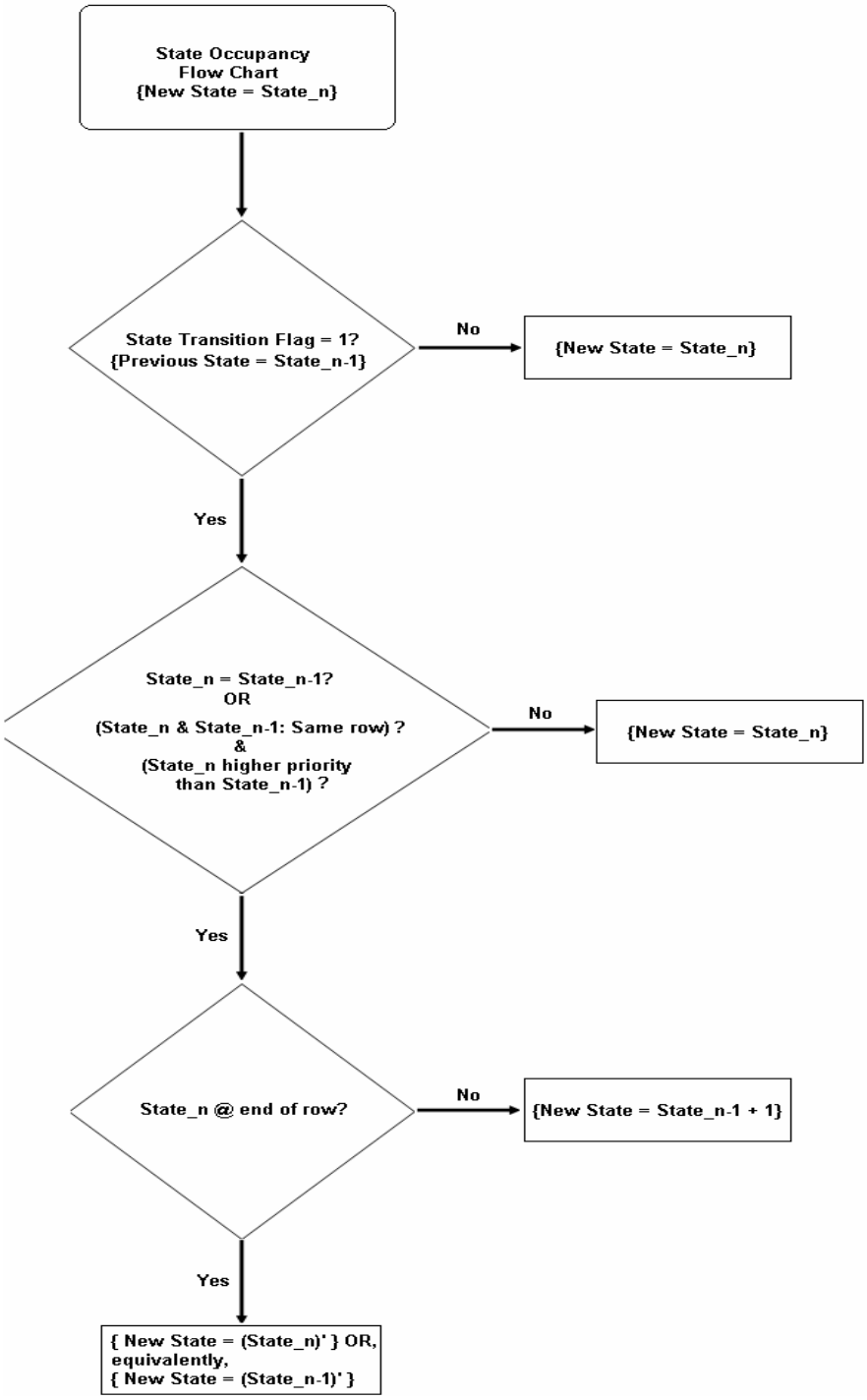


Figure 19. Modified methodology for determining encoder state.

CHAPTER V

PARAMETRIC BIT ALLOCATION

In this approach, bit allocation is shifted from the frame level to individual regions within the frame. A number of criteria are used to determine the number of bits allocated to each region. This differs from the elastic non-parametric bit allocation approach in that ROI bit allocation is done based on well-defined features and even user-defined region weights, but neither ties down as closely with user-quality requirements as the elastic non-parametric ROI bit allocation approach does. Nevertheless, it is a useful approach and merits study and performance comparison with the proposed elastic non-parametric approach.

5.1 Criteria for Regional Bit Allocation

In this approach, a frame level bit budget is derived using state-of-the-art rate control techniques. This budget is then split into region level budgets – denoted by $B^f_{T,ROI}$ for the ROI, and $B^f_{T,BKGRND}$ for the BKGRND. Since macroblock level quantization parameters (QP) depend directly on the number of bits assigned to the region in question, selecting $B^f_{T,ROI}$ and $B^f_{T,BKGRND}$ is crucial to the system design. Thus, as in [10_MontrealStachura], size and motion information are incorporated into the regional bit allocation process. It must be noted, however, that [10_MontrealStachura] uses a fixed QP for each MB in a region. In addition, a user tunable parameter may be used, specified as the region's weight. Thus:

$$B^f_{T,ROI} = (\alpha_S S_{ROI} + \alpha_M M_{ROI} + \alpha_W W_{ROI}) B^f_T$$

$$B^f_{T,BKGRND} = (\alpha_S S_{BKGRND} + \alpha_M M_{BKGRND} + \alpha_W W_{BKGRND}) B^f_T$$

Where:

$$W_x = \frac{\sum_i \omega_{i,x} / s_x}{\sum_y \sum_i \omega_{i,y} / s_y} \qquad x = \{ROI, BKGRND\}, y = \{ROI, BKGRND\}$$

$$S_{ROI} = \frac{N_{ROI}}{N}, S_{BKGRND} = \frac{N_{BKGRND}}{N}$$

$$M_{ROI} = \frac{\sum_{ROI} |MV|}{\sum_{FRAME} |MV|}, M_{BKGRND} = \frac{\sum_{BKGRND} |MV|}{\sum_{FRAME} |MV|}$$

Thus, bits are allocated to individual regions as a fraction of the target bits for the current frame, based on size, motion, and weight information. The α parameters are the priorities of the size, motion, and weight features, and their sum is equal to 1 in order for the regions to meet the required budget for the frame. The S parameters are the normalized size parameters, and the M parameters are the normalized activity parameters – obtained from the motion vectors (MV). W_x represents the normalized weight of region *x*. Note that the weights are size normalized before being normalized over regions. Figure 20 summarizes the parametric bit allocation procedure.

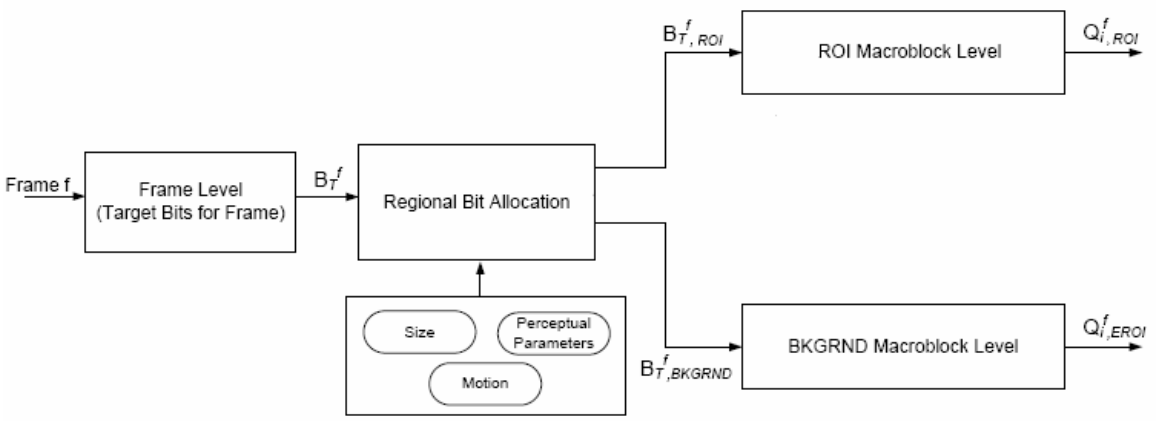


Figure 20. Parametric bit allocation procedure: Bit allocation is first done on a frame level, as in a conventional encoder, and then size, motion, and region weight information are used to determine regional bit allocation. This information is used to finally determine QP values for MBs within the ROI and BKGRND separately.

Two approaches may be used to select the priority weights [a] manual selection, [b] Just Noticeable Distortion (JND) based weights which is motivated out of human visual system (HVS) considerations. The JND method is based on [54], where a Nonlinear Additivity Model for Masking (NAMM) is used to compute the JND as follows:

$$JND(x,y) = T_l(x,y) + T_t(x,y) - C_{l,t}.\min\{T_l(x,y), T_t(x,y)\}$$

In any frame, $T_l(x,y)$ and $T_t(x,y)$ are the visibility thresholds for the two primary masking factors - background luminance masking and texture masking, respectively. (x,y) represents the pixel coordinates. $C_{l,t}$ ($0< C_{l,t} <1$) accounts for the overlapping effect in masking. Luminance masking is modeled by a root

equation for low luminance (below 127) and the other part (above 127) is approximated by a linear function. Texture masking can be determined by local spatial activities (e.g. gradients around the pixel). The expressions for these are detailed in [54] and are implemented in the encoder, but for the sake of simplicity, they are not detailed here.

Since JND is a measure of perceptual masking, it has an inverse relationship with bit allocation. In other words, smaller the JND of a MB, smaller the masking measure for that MB, and so greater should be the number of bits allocated to that MB. Thus, in the implementation, the JND is computed for every pixel in a MB, and the minimum value is chosen as a measure of the masking capacity of the MB. This is done for all MBs in the ROI, and averaged to get a JND value for the ROI. Likewise, the JND value for the BKGRND is obtained. The reciprocal of these (because of the inverse relationship) represent the JND-based region weights for the current frame.

CHAPTER VI

PERFORMANCE RESULTS

In this work we focus upon the problem of acute childhood respiratory distress because [a] it is increasingly common and demands attention, [b] the video that must be transmitted to the clinician is critical to decision making, [c] it occurs with sufficient frequency, [d] initial testing of the technology can occur in settings that are not clinically "live", and [e] "live" clinical testing can occur in a controlled hospital environment in which the usefulness of the technology can be evaluated without putting the patient at risk. With parental informed consent (MCG Human Assurance Committee **#XXXXX**), we collected video databases of pediatric patients in respiratory distress and used clinical experts to test our algorithms' efficacy.

The experimental results consist of three parts. First, the quantification of video quality levels through tests on training videos. Second, the results from encoding test videos, different from training videos, using the proposed elastic non-parametric bit allocation algorithms implemented within an MPEG-2 encoder. Third, the same scenario above, but using the parametric bit allocation algorithms instead.

The experiments use a set of 11 medical videos obtained from MCG. These are videos of patients in respiratory distress, and each includes symptomatic features that are useful to physician experts for visual assessment. Each video has a spatial resolution of 360×240 pixels at 30fps. Figure 21 shows a representative frame from each video used. Table 6 lists the names of the videos used and the number of frames in each video.

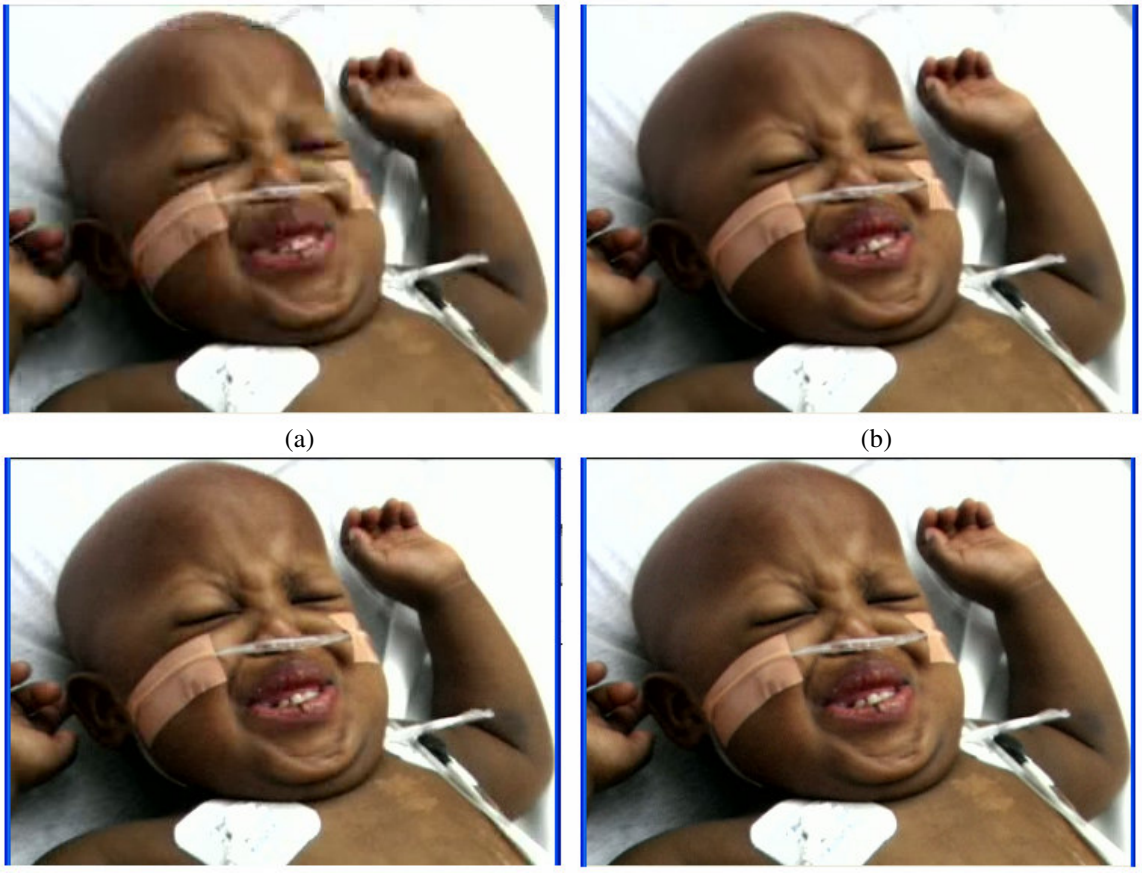
Figure 21. Representative frames from the medical video database.

Table 6. Medical video database names and corresponding frame lengths.

Name	Number of Frames
Er08	1972
Er10	435
Er12	1154
Er13	823
Er14	2506
Er15	1934
Er16	794
Er17	2350
Er19	5111
Er20	1517
Er21	2658

6.1 Quantification of Video Quality Levels

11 videos (all the videos in Table 6) were encoded at uniform spatial quality using a standard MPEG-2 reference implementation (TM5) at 3 bitrates – 500kbps, 1000kbps, and 1500kbps. Figures 22, 23, 24, and 25 show frames representative of overall video quality of er08, er15, er17, and er19, respectively. For comparison, the corresponding frame from the original uncompressed video is also displayed. Similar pictures for other videos are available in the appendix.



(a) (b)

(c) (d)

Figure 22. Uniform compression of video er08. (a) 500kbps. (b) 1000kbps. (c) 1500kbps. (d) Uncompressed video.

Figure 23. Uniform compression of video er15. (a) 500kbps. (b) 1000kbps. (c) 1500kbps. (d) Uncompressed video.

Figure 24. Uniform compression of video er17. (a) 500kbps. (b) 1000kbps. (c) 1500kbps. (d) Uncompressed video.

Figure 25. Uniform compression of video er19. (a) 500kbps. (b) 1000kbps. (c) 1500kbps. (d) Uncompressed video.

They were then evaluated by 2 medical experts from MCG. Tables 7, 8, 9, and 10 (identical to the template in Table 3) list the evaluations of er08, er15, er17, and er19, respectively, by one of the medical experts. They are representative of the overall results, and evaluations other videos at all the TBRs by each expert are available in the appendix. For simplicity of presentation, the video sample names have been converted to the form NAME_TBR. However, as stated before, these videos were presented with random sample names and in random order.

Table 7. Evaluation of er08 compressed at four different compression levels.

Random Video Sample	Feature Sets	DL? (1-4)	PL? (Yes/No)	Comments
Er08_500	WB	2	No	Unclear
	RR	2		
	T	2		
	MS	3		
	HB	2		
	NF	2		
Er08_1000	WB	3	Yes	
	RR	3		
	T	3		
	MS	4		
	HB	3		
	NF	3		
Er08_1500	WB	3	Yes	
	RR	3		
	T	3		
	MS	4		
	HB	4		
	NF	4		
Er08_orig	WB	4	Yes	
	RR	4		
	T	4		
	MS	4		
	HB	4		
	NF	4		

Table 8. Evaluation of er15 at four different compression levels. (Key: RE – Respiratory excursion, LA – Level of activity, STM – Skin tone motting, WB – Work of breathing, RR – Respiratory rate, T-Tachypnea, R – Retractions, MS – Mental status)

Random Video Sample	Feature Sets	DL? (1-4)	PL? (Yes/No)	Comments
Er15_500	RE	2	No	Unclear
	LA	2		
	STM	1		
	WB	1		
	RR	1		
	T	1		
	R	1		
	MS	1		
Er15_1000	RE	3	Yes	
	LA	3		
	STM	3		
	WB	3		
	RR	3		
	T	3		
	R	3		
	MS	3		
Er15_1500	RE	4	Yes	
	LA	4		
	STM	4		
	WB	4		
	RR	4		
	T	4		
	R	4		
	MS	4		
Er15_orig	RE	4	Yes	
	LA	4		
	STM	4		
	WB	4		
	RR	4		
	T	4		
	R	4		
	MS	4		

Table 9. Evaluation of Er17 at four different compression levels. (Key: WB – Work of breathing, RE – Respiratory excursion, A – Activity, R – Retractions, RR – Respiratory rate, T – Tachypnea, MS – Mental status)

Video Sample	Feature Sets	DL? (1-4)	PL? (Yes/No)	Comments
Er17_500	WB	2	No	Unclear
	RE	2		
	A	1		
	R	1		
	RR	1		
	T	1		
	MS	1		
Er17_1000	WB	3	No	-
	RE	3		
	A	3		
	R	3		
	RR	3		
	T	3		
	MS	3		
Er17_1500	WB	4	Yes	-
	RE	4		
	A	4		
	R	4		
	RR	4		
	T	4		
	MS	4		
Er17_orig	WB	4	Yes	-
	RE	4		
	A	4		
	R	4		
	RR	4		
	T	4		
	MS	4		

Table 10. Evaluation of Er19 at four different compression levels. (Key: R – Retractions, LA – Level of activity, WB – Work of breathing, MS – Mental status, RR – Respiratory rate, T – Tachypnea, HB – Head bobbing)

Video Sample	Feature Sets	DL? (1-4)	PL? (Yes/No)	Comments
Er19_500	R	2	No	"Fuzzy"
	LA	3		
	WB	2		
	MS	2		
	RR	3		
	T	3		
	HB	3		
Er19_1000	R	4	No	Clear
	LA	4		
	WB	3		
	MS	4		
	RR	4		
	T	4		
	HB	4		
Er19_1500	R	4	Yes	-
	LA	4		
	WB	4		
	MS	4		
	RR	4		
	T	4		
	HB	4		
Er19_orig	R	4	Yes	-
	LA	4		
	WB	4		
	MS	4		
	RR	4		
	T	4		
	HB	4		

These tables list the particular feature sets that were visible to the specialist in the particular video, and the specialist's opinion of their quality in both a DL and a PL sense. For example, in Table 7, for video er08, the features visible were WB, RR, T, MS, HB, and NF.

In using the evaluations to quantify video quality, only about half of the database was used. This was done so as to leave the remaining videos for testing purposes. The videos used for video quality quantification are – er10, er12, er14, er16, er19, and er21. After averaging over these videos, and over all the experts, the required TBR values for DL for individual features were obtained, and are tabulated in Table 11. The expansions of the feature names are given in Appendix B. Averaging over experts' feedback is justified because they were clustered close together, as opposed to being significantly different. Also, it is

interesting to observe the clustering of TBR values for DL for different medical features, irrespective of the race of the patient etc. Lighting conditions in all the videos were approximately the same.

Table 11. Required TBR for DL of several medical features.

Feature	A	CE	G	LA	MR	MS	NF	RR	R
TBR(kbps)	750	1000	1000	1000	1000	1000	750	1077	786
Feature	T	WB	HB	RE	STM	SUB	SUPRA	INTER	
TBR(kbps)	923	750	1000	1000	1000	750	500	500	

Thus, choosing the maximum TBR value results in DL over all features, and this value is 1077kbps. Similarly, the required TBR for PL, obtained after averaging was found to be 1400kbps. In practical encoders, a certain minimum bpp is required for coding. This minimum is set as the value for the required bpp for BE. Table 12 summarizes the required bpp for PL, DL and BE. These are the results for the quantification of video quality levels – 0.54bpp for PL, 0.42bpp for DL, and 0.10bpp for BE.

Table 12. Summary of bpp values for PL, DL, and BE.

Quality Level	PL	DL	BE
bpp	0.54	0.42	0.10

Table 13 summarizes ROI PSNR values of video sequences when compressed uniformly at bitrates corresponding to PL and DL – 1400kbps and 1077kbps, respectively. It is interesting to note how closely these values are clustered together for the various video sequences because of the similarity in their content. By averaging over *only the training* videos, the PSNR_TH values corresponding to PL and DL are obtained as 35.10dB and 34.01dB, respectively.

Table 13. PSNR values of video sequences at bitrates corresponding to DL and PL.

Video Sample	PSNR DL (dB)	PSNR PL (dB)
Er08	34.08	35.06
Er10	33.87	34.76
Er12	33.60	35.93
Er13	34.48	35.70
Er14	34.89	35.73
Er15	33.85	34.60
Er16	34.50	35.68
Er17	34.18	34.95
Er19	33.67	34.11
Er20	34.21	35.45
Er21	33.57.	34.40.

6.2 ROI Encoding – Elastic Non-Parametric and Parametric Bit Allocation

ROI encoding of all videos was done in four different ways. The first two methodologies were based on the elastic non-parametric ROI bit allocation approach. The bpp values for PL, DL, and BE were used to do ROI encoding of all the videos at three different TBR – 500kbps, 750kbps, and 1000kbps. This was done in two ways: a) without any quality update techniques, b) with quality update techniques i.e. state transitions. In the third and fourth methodology, bit allocation was based on the parametric approach and the encoding was done at the same TBRs mentioned above. Specifically, in the third methodology, only a region's weight was used for bit allocation, and the weights of the ROI and BKGRND were both equal to 0.5. In the fourth methodology, size, motion, and region weights were all given equal priority i.e. the priority weights were 1/3. The ROI and BKGRND weights were still equal to 0.5. In all of the above cases, the ROI was manually selected based on a priori knowledge of the ROI based on experts' feedback. Typically, the size of the ROI varied from 25-50% of the total spatial resolution.

In the following, the above four methodologies are compared in an objective sense using PSNR and bits allocated as measures. More importantly, they are compared in a subjective sense using expert evaluation. Since elastic non-parametric bit allocation uses expert information on video quality levels, videos used in the quantification (training videos) procedure are grouped into one class, and the remaining test videos are grouped into another class.

Objective Comparison

Videos er10, er12, er14, er16, er19, and er21 were used in the quantification of quality levels, but they are also ROI encoded. Videos er08, er13, er15, er17, and er20 are the test videos that are ROI encoded but were not used in the quantification of quality levels procedure.

The following figures show the results after encoding video er08 at 500kbps, 750kbps, and 1000kbps. Figure 26 shows ROI PSNR as a function of GOP number at 500kbps. ROI PSNR per frame was averaged over a GOP to obtain the PSNR for a GOP. The performance of the elastic non-parametric approach with quality updates is the best, followed by elastic non-parametric bit allocation without quality updates, then parametric bit allocation with equal priorities for size, motion, and region weights, and finally parametric bit allocation with only region weights.

The elastic non-parametric bit allocation curves are green and red in color, and where they exceed the PSNR threshold for the encoder's state in the previous GOP, they overlap. At 500kbps, it turns out the encoder is in state 9 i.e. {ROI, BKGRND} = {DL, BE}. The corresponding threshold is 34dB. Thus, it can be seen that the green and red plots overlap when ROI PSNR exceeds 34dB. When it does not, the plots do not overlap because the encoder increases ROI bit allocation by transitioning to state 9'.

The parametric bit allocation curves are blue and black, and in general they fall below the elastic non-parametric bit allocation curves because of the empirical choices for the feature parameters. In the two strategies chosen, one with equal priorities for different features, represented by the blue plot, and one with only region weights, represented by the black plot, the ROI and BKGRND were assigned equal weights of 0.5. Thus, with the black plot, 50% of the bits per frame are allocated to the ROI, irrespective of ROI size, motion etc. With the blue plot, the allocation is more flexible, with some dependency ($1/3^{rd}$) on size and motion as well. However, it turns out that in neither of these cases, the bit allocation to the ROI matches that of the elastic non-parametric scheme.

Figures 27 and 28 show similar ROI PSNR plots at 750kbps and 1000kbps, respectively. At 750kbps, the encoder is in state 9 i.e. {ROI, BKGRND} = {DL, BE}. The corresponding threshold is still 34dB. The green and red plots overlap when ROI PSNR exceeds 34dB. When it does not, the plots do not overlap because the encoder increases ROI bit allocation by transitioning to state 9'. At 1000kbps, the encoder is in state 6 i.e. {ROI, BKGRND} = {PL, DL}. The corresponding threshold is 35.1dB. The green and red plots overlap when ROI PSNR exceeds 35dB. When it does not, the plots do not overlap because the encoder increases ROI bit allocation by transitioning to state 7. The ROI PSNR performance order is the same at each TBR i.e. 500kbps, 750kbps, and 1000kbps.

Figures 29, 30, and 31 show BKGRND PSNR versus GOP number at 500kbps, 750kbps, and 1000kbps**,** respectively. As expected, the order is reversed i.e. the performance of the elastic non-parametric approach with quality updates is the lowest, followed by elastic non-parametric bit allocation without quality updates, then parametric bit allocation with equal priorities for size, motion, and region weights, and finally parametric bit allocation with only region weights has the highest BKGRND PSNR.

It can be noted that the elastic non-parametric bit allocation curves are clustered *somewhat* close to each other, and the parametric bit allocation curves are clustered close to each other. However, there is a

significant PSNR difference between these two sets of curves. The green and red elastic non-parametric bit allocation curves overlap when ROI PSNR exceeds the threshold corresponding to the encoder's state. However, when it does not, ROI bit allocation is increased in the following GOP, and BKGRND bit allocation is decreased. At 500kbps and 750kbps, the encoder transitions from state 9 to 9', thus there is only a small decrease in BKGRND PSNR. At 1000kbps, the encoder transitions from state 6 to 7 i.e. BKGRND quality changes from DL to BE, therefore there is a significant decrease in BKGRND PSNR. BKGRND bit allocation in the two parametric cases is not very different, as will be seen in the following bit allocation plots. Furthermore, the bit allocation percentage is significant, compared to the elastic non-parametric bit allocation case, which explains why the difference in BKGRND PSNR is small (the relatively flatter portions of a generic PSNR versus bitrate characteristic occur at higher bitrates where PSNR changes are relatively small).

Figures 32, 33, and 34 plot bits allocated to the ROI versus GOP number at 500kbps, 750kbps, and 1000kbps, respectively. Likewise, Figures 35, 36, and 37 plot bits allocated to the BKGRND versus GOP number at 500kbps, 750kbps, and 1000kbps, respectively. Consistent with their corresponding PSNR plots, ROI bit allocation increases from parametric bit allocation to elastic non-parametric bit allocation, irrespective of TBR, and BKGRND bit allocation decreases from parametric bit allocation to elastic non-parametric bit allocation. Like the PSNR plots, the green and red curves overlap whenever ROI PSNR exceeds the threshold corresponding to the encoder's state. One important observation is that ROI bit allocation is significantly greater in the elastic non-parametric case compared to the parametric case. However, the ROI PSNR plots do not reflect a corresponding significant difference. This, again, can be attributed to being in the flatter portions of a generic PSNR versus bitrate characteristic i.e corresponding to higher bitrates. On the other hand, BKGRND bit allocation is significantly lower in the elastic non-parametric case compared to the parametric case. The BKGRND PSNR plots do reflect a corresponding significant difference. This can be attributed to being in the steeper portions of a generic PSNR versus bitrate characteristic.

It must also be noted that with increasing bitrate, bits allocated to BKGRND increases with parametric bit allocation whereas it remains relatively constant with elastic non-parametric bit allocation. This emphasizes the important difference between these two bit allocation methods – with parametric bit

allocation, the bit allocation methodology is identical irrespective of bitrate, whereas with elastic non-parametric bit allocation it is adaptive.

The general nature of the above plots for the other videos, whether training videos or test videos, is similar to the ones above. This is not a surprising conclusion with parametric bit allocation, because the classification of a video as a training or test video is only with respect to elastic non-parametric bit allocation. Thus, it is a significant observation that the PSNR and bit allocation plots for ROI and BKGRND at 500kbps, 750kbps, and 1000kbps are behaviorally similar to the plots for the video er08. In other words, the results are similar for both training and test videos.

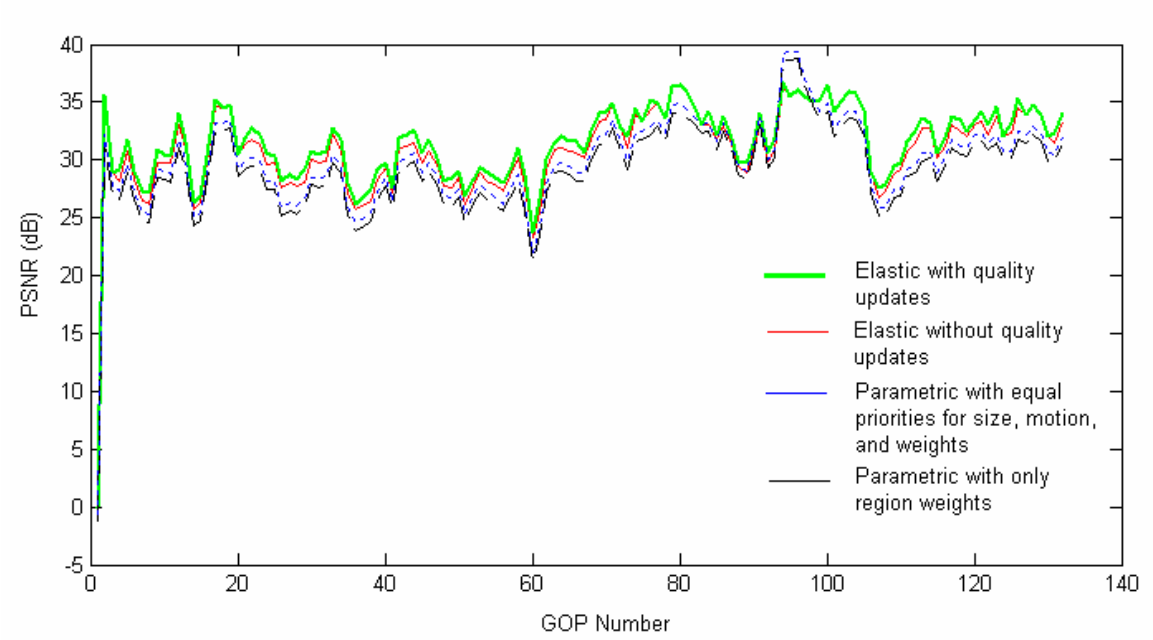


Figure 26. ROI PSNR versus GOP number for er08 at 500kbps.

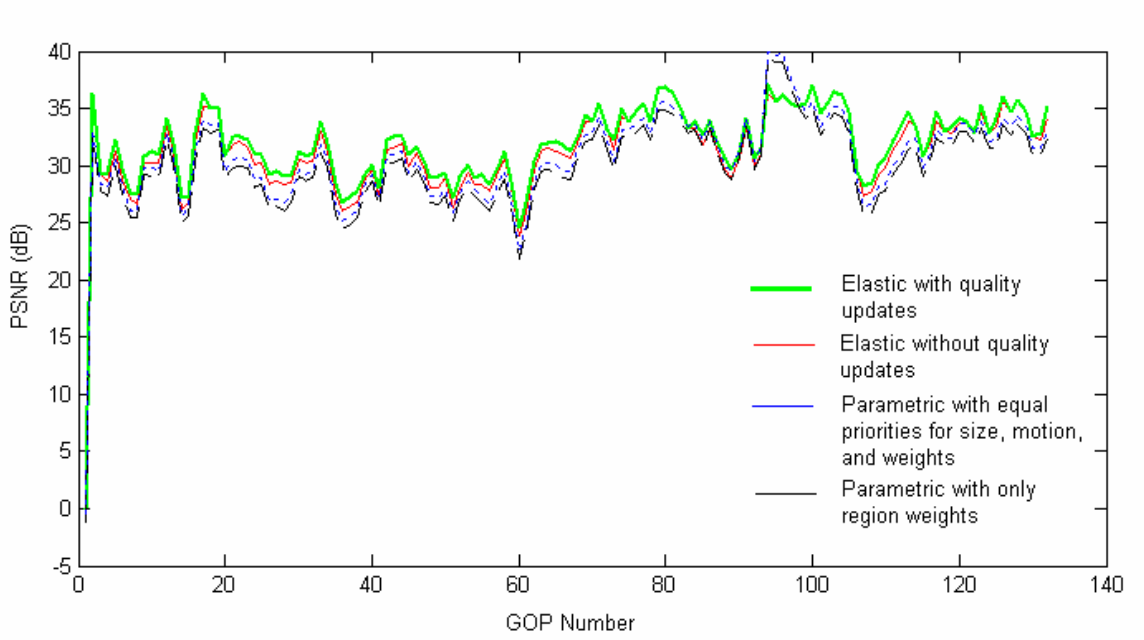


Figure 27. ROI PSNR versus GOP number for er08 at 750kbps.

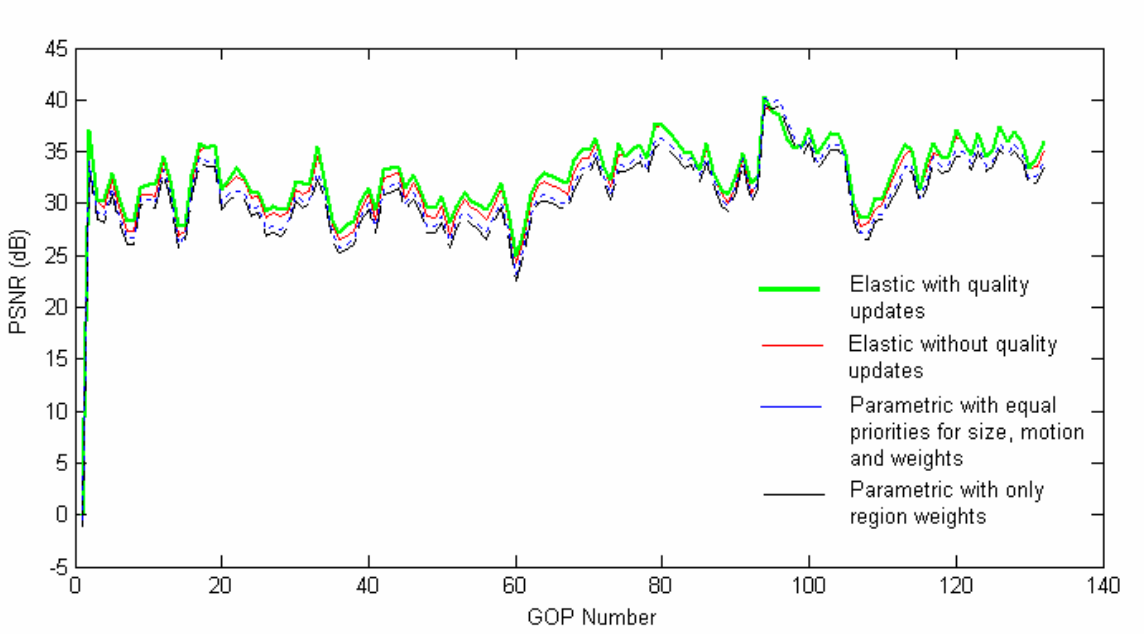


Figure 28. ROI PSNR versus GOP number for er08 at 1000kbps.

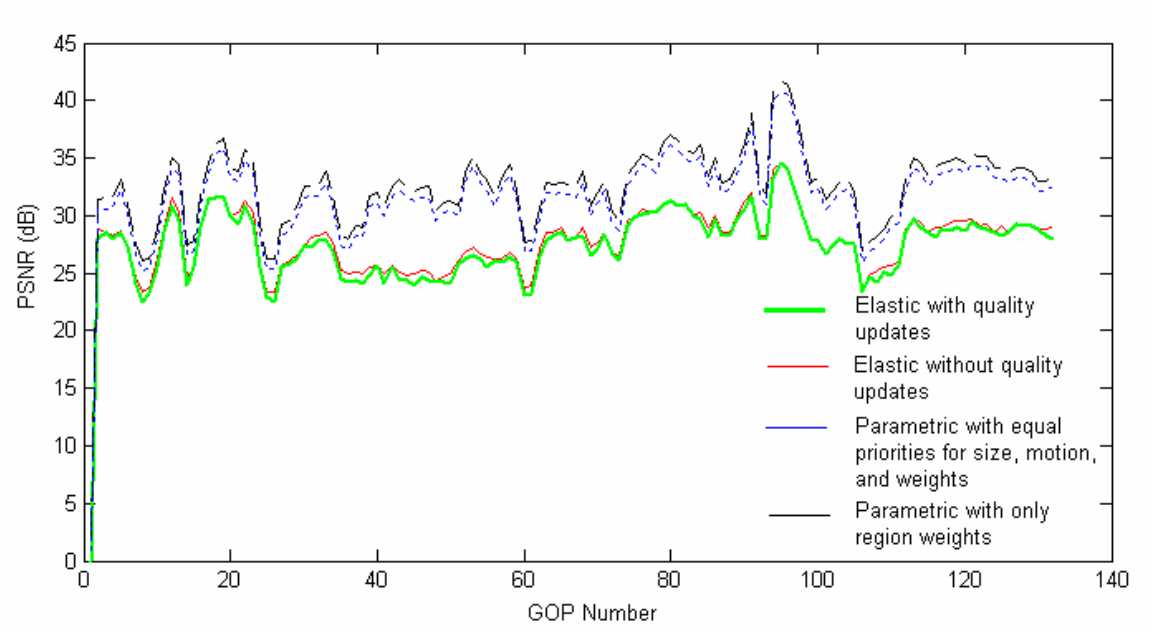


Figure 29. BKGRND PSNR versus GOP number for er08 at 500kbps.

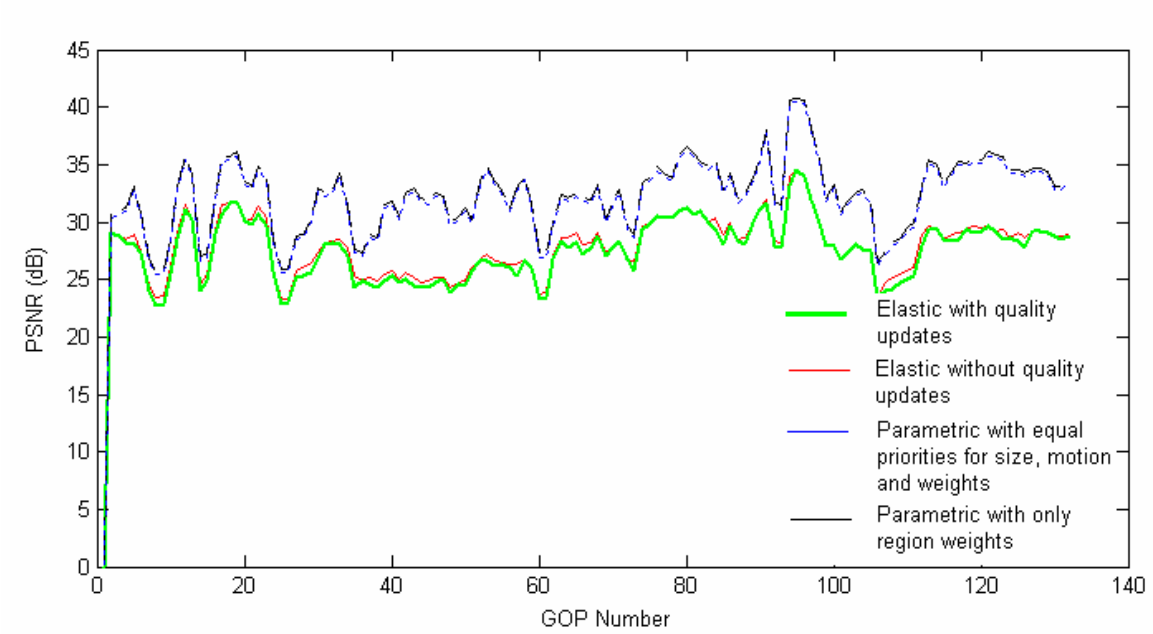


Figure 30. BKGRND PSNR versus GOP number for er08 at 750kbps.

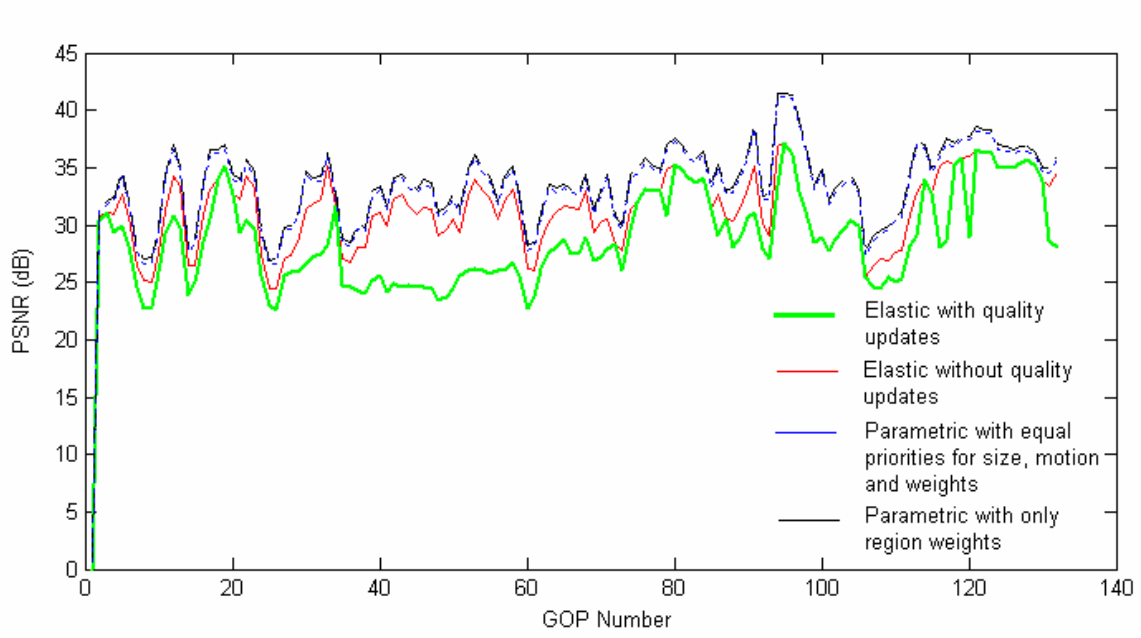


Figure 31. BKGRND PSNR versus GOP number for er08 at 1000kbps.

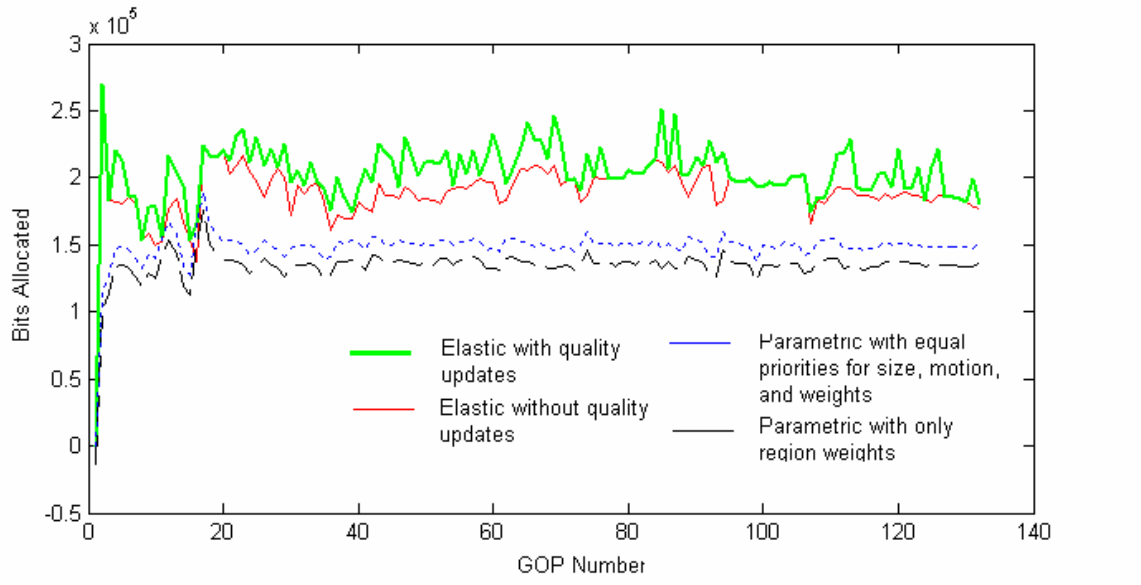


Figure 32. ROI bit allocation versus GOP number for er08 at 500kbps.

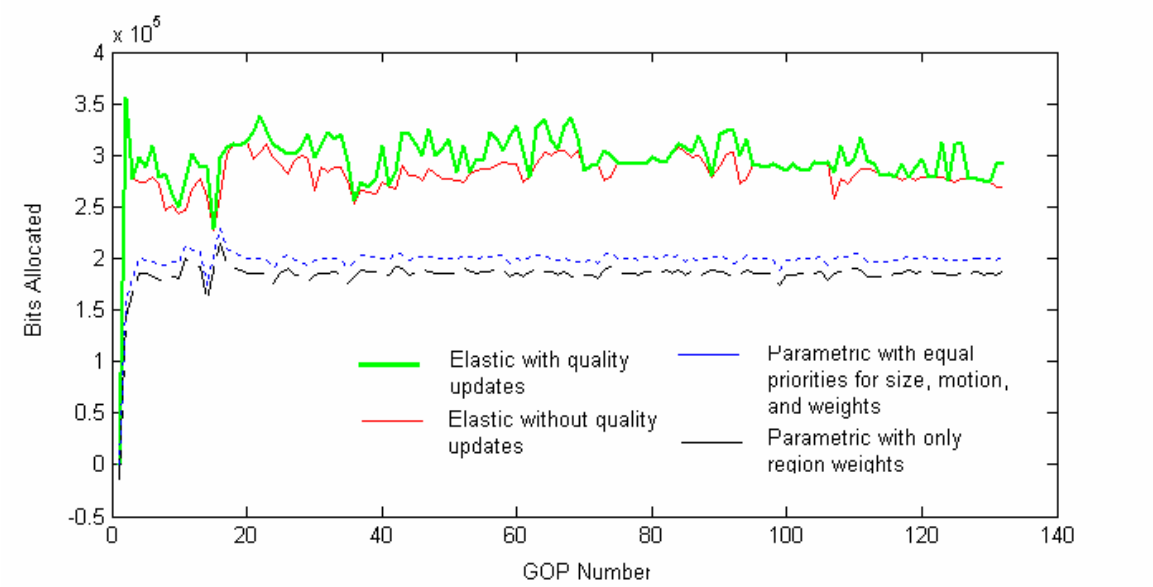


Figure 33. ROI bit allocation versus GOP number for er08 at 750kbps.

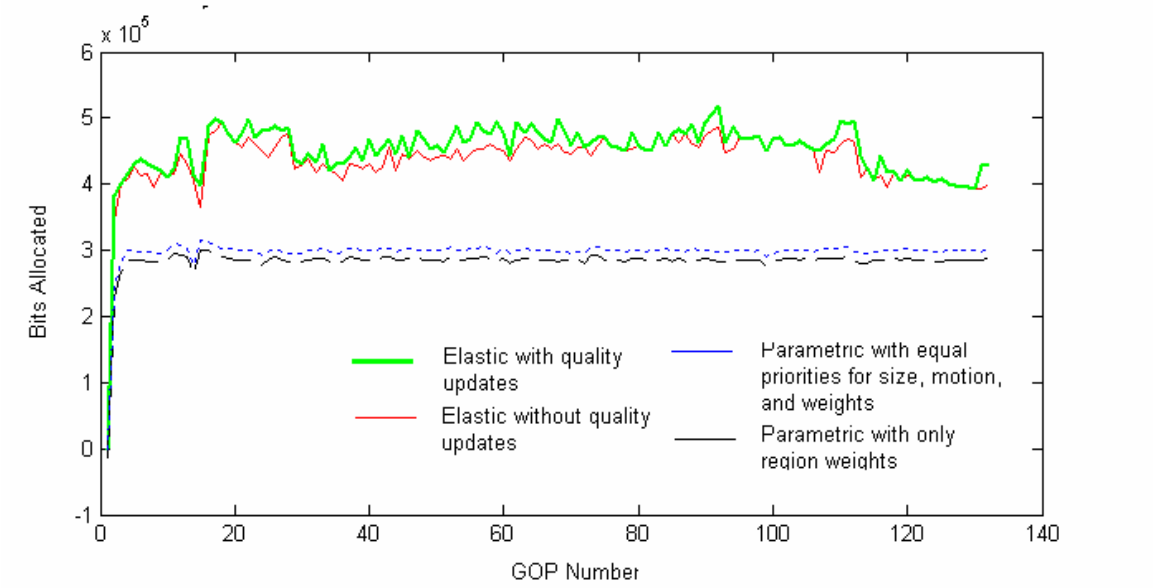


Figure 34. ROI bit allocation versus GOP number for er08 at 1000kbps.

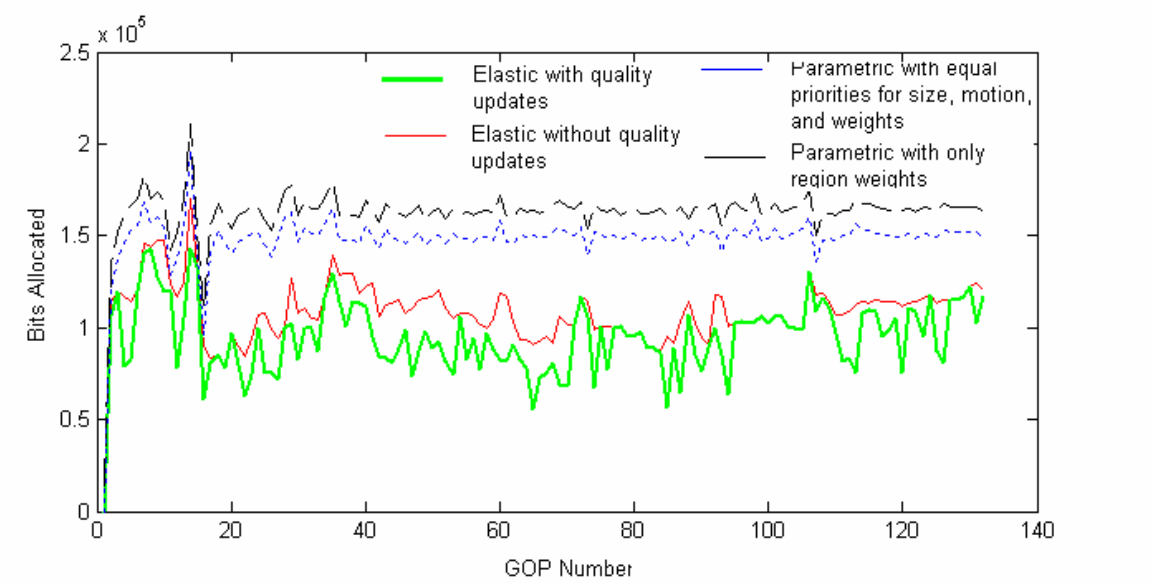


Figure 35. BKGRND bit allocation versus GOP number for er08 at 500kbps.

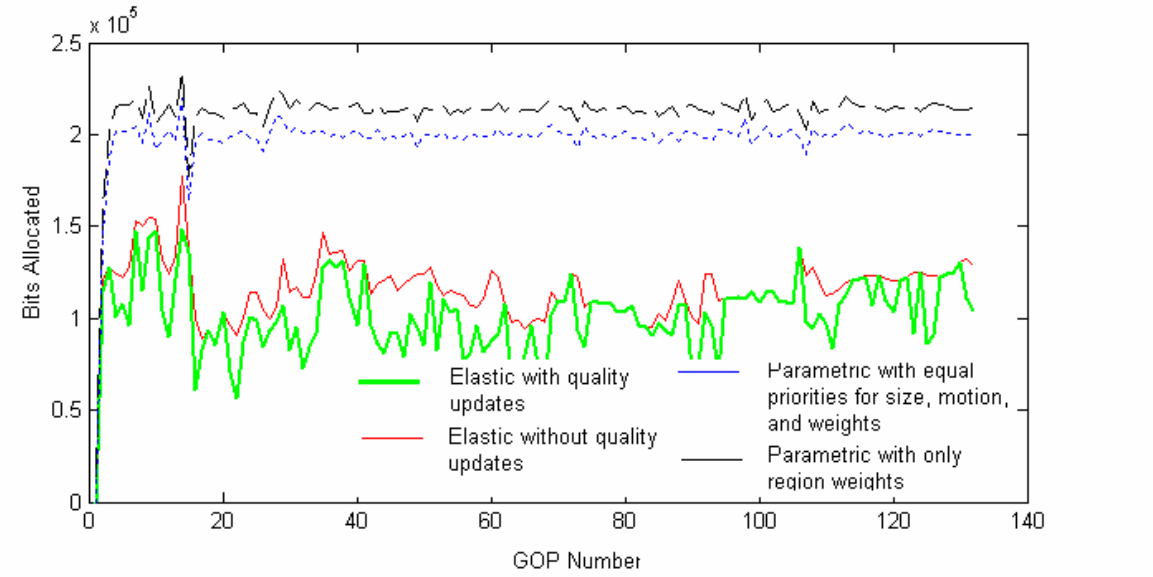


Figure 36. BKGRND bit allocation versus GOP number for er08 at 750kbps.

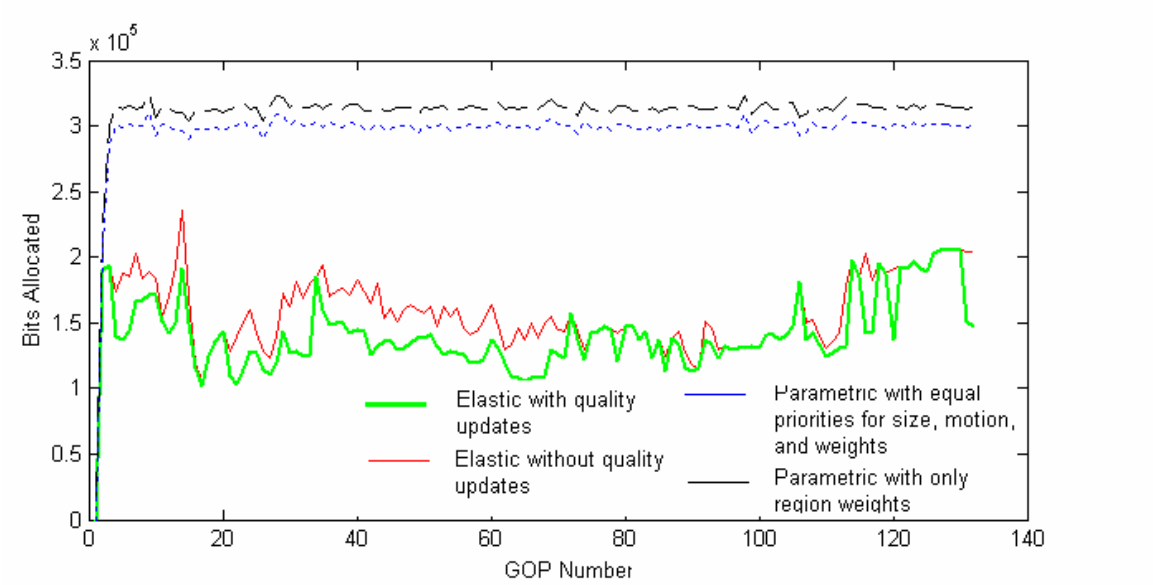


Figure 37. BKGRND bit allocation versus GOP number for er08 at 1000kbps.

Subjective Comparison

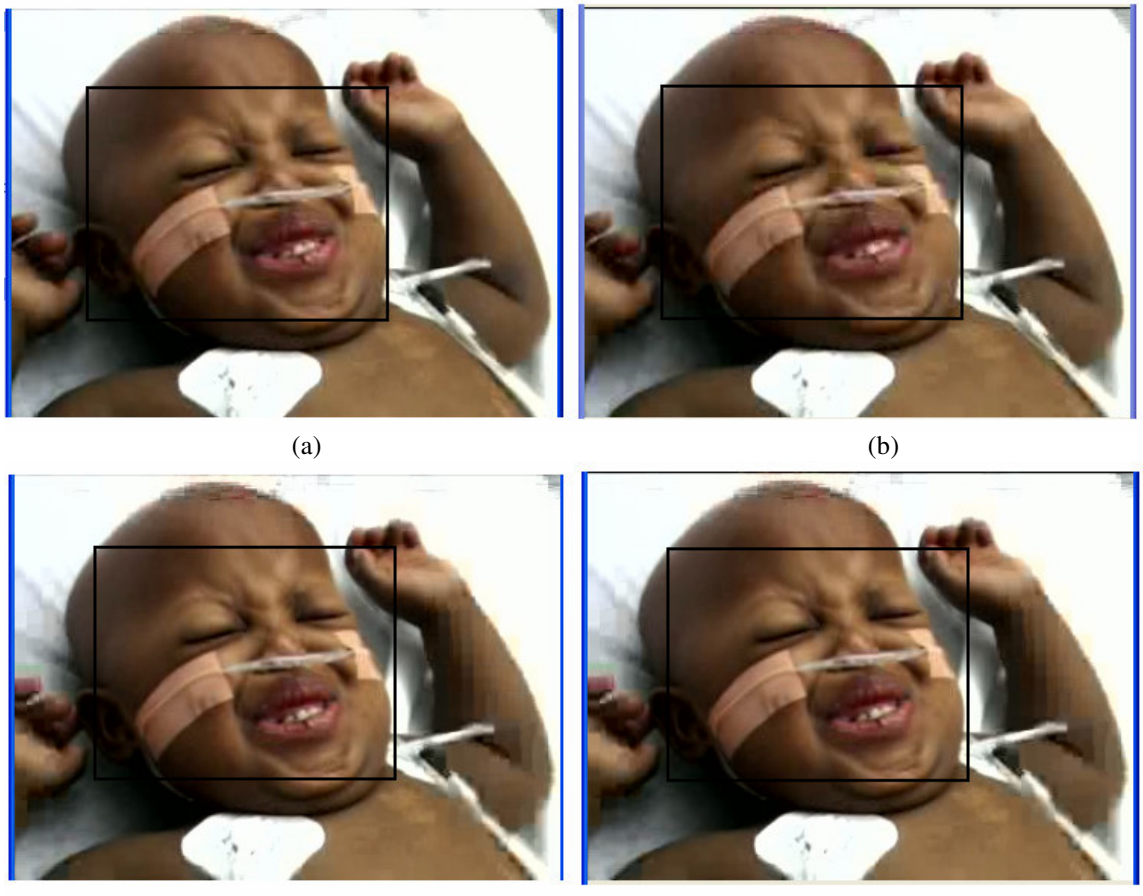
In order to evaluate the usefulness of elastic non-parametric bit allocation and compare it with parametric bit allocation, all ROI based videos encoded using both elastic non-parametric bit allocation methods and both parametric bit allocation methods at 500kbps, 750kbps, and 1000kbps were evaluated by the medical experts who originally provided feedback on uniformly compressed videos. Table 14 shows the evaluation template provided to the experts. It is identical to the template for uniformly compressed videos (Table 3), except that in this case, the assessment of PL and DL apply to the ROI alone.

Table 14. Evaluation template for ROI encoded videos.

Random Video Sample	Features	ROI DL? (1-4)	ROI PL? (Yes/No)	Comments
1-n	RR OC WB ….			

For visual comparison, Figure 38 shows a frame representative of overall video quality of er08 when encoded at 500kbps using both elastic non-parametric bit allocation methods and both parametric bit allocation methods. Likewise, Figures 39 and 40 show the same frames at 750kbps and 1000kbps, respectively. The ROI is indicated by a rectangular bounding box. At 500kbps and 750kbps, with parametric bit allocation, the ROI quality is poorer than with elastic non-parametric bit allocation. At

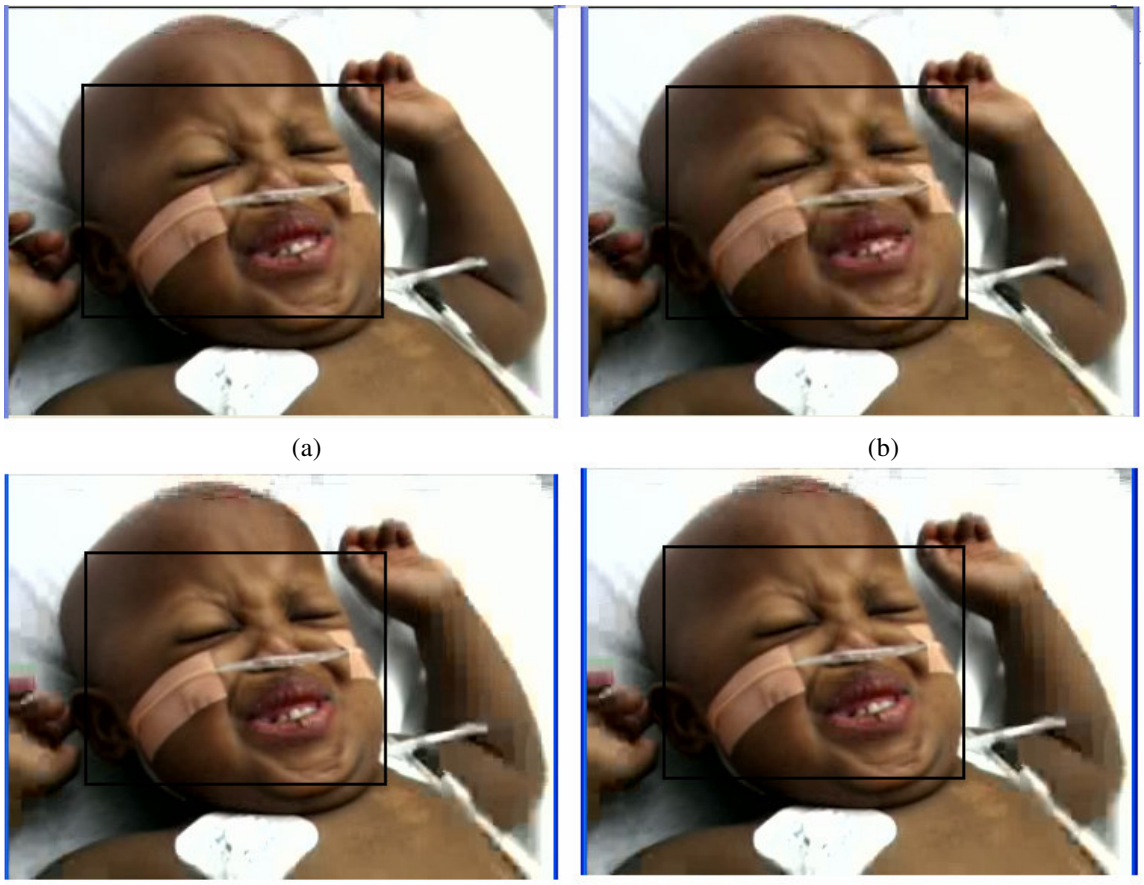
1000kbps, this difference is less perceivable. To illustrate the usefulness of ROI coding per se, it is useful to compare these pictures with pictures of the same frame obtained with uniform coding from Figure 22.



(a) (b)

(c) (d)

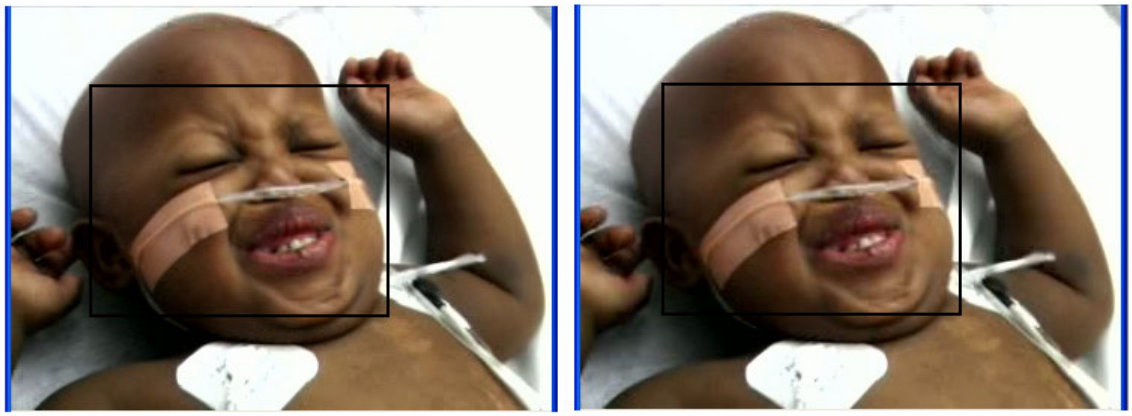
Figure 38. Representative ROI encoded frame of er08 at 500kbps. (a) Parametric bit allocation – only region weights. (b) Parametric bit allocation – size, motion, and region weights. (c) Elastic non-parametric bit allocation without updates. (d) Elastic non-parametric bit allocation with updates.



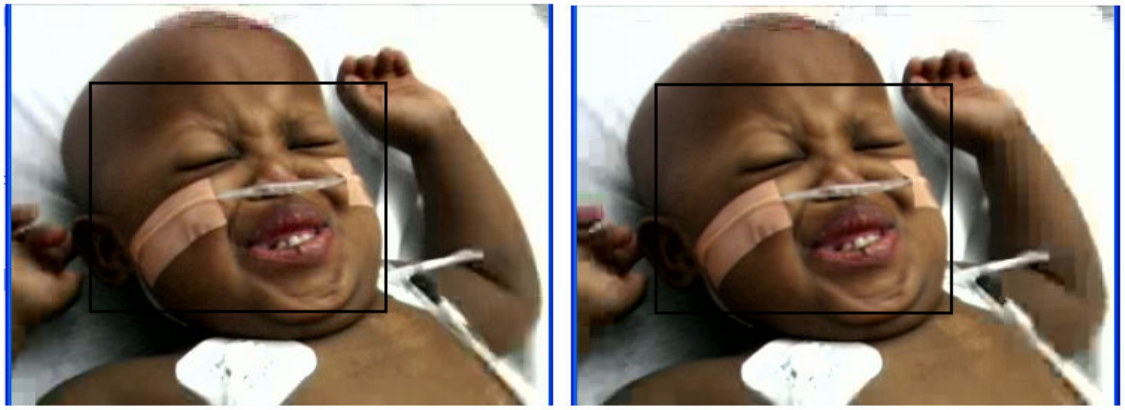
(a) (b)

(c) (d)

Figure 39. Representative ROI encoded frame of er08 at 750kbps. (a) Parametric bit allocation – only region weights. (b) Parametric bit allocation – size, motion, and region weights. (c) Elastic non-parametric bit allocation without updates. (d) Elastic non-parametric bit allocation with updates.



(a) (b)



(c) (d)

Figure 40. Representative ROI encoded frame of er08 at 1000kbps. (a) Parametric bit allocation – only region weights. (b) Parametric bit allocation – size, motion, and region weights. (c) Elastic non-parametric bit allocation without updates. (d) Elastic non-parametric bit allocation with updates.

Table 15 shows the completed evaluation template by an expert of video er08 at 500kbps, 750kbps, and 1000kbps with each of the four adopted bit allocation methods. From the table, it can be noted that any TBR, there is a general decrease in values for *ROI DL?* from elastic non-parametric bit allocation approaches to parametric bit allocation approaches. A comparison with Table 7 also reveals an improvement in performance over the uniform compression case.

Table 15. Completed evaluation template of er08 at 500kbps, 750kbps, and 1000kbps when ROI encoded with both parametric bit allocation methods and both elastic non-parametric bit allocation methods.

Video_TBR	Features	Elastic non-parametric with updates		Elastic non-parametric without updates		Parametric – Size, Motion, Weights		Parametric – Only weights	
		ROI PL?	ROI DL?	ROI PL?	ROI DL?	ROI PL?	ROI DL?	ROI PL?	ROI DL?
Er08_500	WB	Yes	3	Yes	3	Yes	3	No	2
	RR		3		3		3		3
	T		3		3		3		2
	MS		4		3		3		3
	HB		4		4		3		3
	NF		3		3		3		2
Er08_750	WB	Yes	4	Yes	4	Yes	3	No	3
	RR		4		3		3		3
	T		4		4		3		3
	MS		4		3		4		3
	HB		4		3		3		3
	NF		4		4		4		3
Er08_1000	WB	Yes	4	Yes	3	Yes	3	Yes	3
	RR		4		3		3		3
	T		4		4		3		3
	MS		4		4		4		3
	HB		4		4		3		4
	NF		4		4		4		4

To gain an understanding of the general performance of the above four adopted bit allocation methods, the above feedback was considered for all the videos. These were then averaged to obtain a mean (μ) and standard deviation (σ) score for *ROI DL?* at 500kbps, 750kbps, and 1000kbps. Table 16 shows these values. With the elastic non-parametric bit allocation methods, at 500kbps and 750kbps, the ROI is expected to be of DL quality, and at 1000kbps, of PL quality. Since DL is expected irrespective of the particular medical feature, the features column does not appear in the table below. Since *ROI PL?* is answered with a Yes or No, it is displayed as a percentage in the table, representing the percentage score for an answer Yes to the question ROI PL?

From the table, it can be noted that any TBR, there is a general decrease in mean values for *ROI DL?* from elastic non-parametric bit allocation approaches to parametric bit allocation approaches. With the elastic non-parametric approach without updates, the results are consistent with the encoder state at each TBR. For example, at 500kbps and 750kbps, the ROI is in state 9, corresponding to DL quality for the ROI. These are reflected in their corresponding mean *ROI DL?* values – 3.55 and 3.64, respectively. At 1000kbps, the ROI is in state 6, corresponding to PL quality for the ROI. Since PL falls above DL in the quality hierarchy, the ROI is also expected to be of DL quality. This is reflected in its corresponding mean

ROI DL? value – 3.70. A performance improvement is seen with elastic non-parametric bit allocation with updates. At 500kbps, 750kbps, and 1000kbps, the mean *ROI DL?* values increase to 3.67, 3.78, and 3.79, respectively. On the other hand, with parametric bit allocation, the performance deteriorates. With parametric bit allocation using all feature information, the mean *ROI DL?* values at 500kbps, 750kbps, and 1000kbps are reduced to 2.92, 3.08, and 3.25, respectively. With parametric bit allocation using only region weights, these are further reduced to 2.85, 3.05, and 3.17, respectively. The standard deviation values are indicative of the variability about the mean values. To illustrate the advantages of ROI coding per se, it is useful to compare the above values with those obtained with uniform compression at the same TBR (as was done in the quantification of quality levels step). It turns out that at 500kbps, the mean *ROI DL?* value is 2.29, with a standard deviation σ of 0.57. At 1000kbps, mean *ROI DL?* is 2.94, and σ is 0.64. At 1500kbps, mean *ROI DL?* is 3.54, and σ is 0.61.

ROI PL? values are indicated as percentages, and it is interesting to note its values at 500kbps as 82% and 91%, with elastic non-parametric bit allocation without and with updates, respectively. At 750kbps, the values are identical. These results are surprising because at both 500kbps and 750kbps, the ROI is expected to be of only DL quality, and not PL quality. A likely explanation for the above is the expert's perception of the difference in quality between the ROI (as good) and the BKGRND (as bad). However, these did not happen with parametric bit allocation, where the values ranged between 45-65%. Again, to illustrate the advantages of ROI coding per se, it is useful to compared the above percentages with those obtained with uniform compression at the same TBR (as was done in the quantification of quality levels step). At 500kbps, the response to *ROI PL?* was Yes in only 18% of the cases. At 1000kbps and 1500kbps, these values were 45%, and 72%, respectively.

Table 16. Averaged results for *ROI PL?* and *ROI DL?* at 500kbps, 750kbps, and 1000kbps with both parametric bit allocation methods and both elastic non-parametric bit allocation methods.

TBR	Elastic non-parametric with updates		Elastic non-parametric without updates		Parametric – Size, Motion, Weights		Parametric – Only weights	
	ROI PL? (%)	ROI DL? (μ, σ)	ROI PL? (%)	ROI DL? (μ, σ)	ROI PL? (%)	ROI DL? (μ, σ)	ROI PL? (%)	ROI DL? (μ, σ)
500	91	3.67, 0.71	82	3.55, 0.57	45	2.92, 0.68	54	2.85, 0.22
750	91	3.78,0.53	82	3.64, 0.77	54	3.08, 0.55	45	3.05, 0.41
1000	91	3.79,0.44	91	3.70, 0.59	64	3.25, 0.39	54	3.17, 0.44

6.3 Summary of Results

To summarize, the above results show that the elastic non-parametric bit allocation algorithm performs better than the parametric bit allocation algorithm. The bpp required for DL, using the elastic non-parametric bit allocation algorithm, was found to be 0.42bpp. This was found by medical expert feedback on uniformly compressed videos and validated by ROI encoding on ROI compressed videos. Thus, at 500kbps, ROI encoded videos were DL, as compared to 1077kbps for uniformly encoded videos. The ROI in the above videos was user-specified and typically 25-50% of the total spatial resolution. This shows that ROI encoding with our proposed algorithms results in functional losslessness in video quality over bitrates corresponding to the wireless sweet spot i.e. the bitrate range corresponding to state-of-the-art mobile technologies.

CHAPTER VII

CONCLUDING REMARKS

7.1 Summary

In this work, we developed an elastic non-parametric methodology for ROI bit allocation for video compression and transmission over VBR channels. This achieves two important goals – flexibility and adaptivity to user quality requirements, and maintenance of this efficiency even under hostile channel conditions such as limited bandwidth as well as bandwidth fluctuations.

Although the particular application considered in this work is pediatrics based remote medical assessment, the algorithms developed are in general application independent, and even where they are application dependent, they may be easily extensible to other applications. The methods explored in this work may be broadly classified into two categories – parametric bit allocation, and elastic non-parametric bit allocation. Parametric bit allocation operates at the frame layer, and assigns bits to regions based on non-medical feature information such as size, motion, and region priority. Elastic non-parametric bit allocation is based on statistical information on allowable compression levels of medical features for perceptual clarity (PL) as well as diagnostic acceptability (DL). Based on the thresholds for this quality hierarchy, and the available bitrate estimate for the channel, the encoder occupies an optimal state. A state assignment assigns quality levels for the ROI and BKGRND based on the defined quality hierarchy, subject to first maximizing ROI quality, and then achieving the best possible BKGRND quality. In determining the encoder's state, the encoder also determines the bits to be allocated to the ROI and BKGRND. An additional update algorithm to improve ROI quality based on the PSNR metric is also developed.

The above methods were tested on medical videos filmed at the Medical College of Georgia, Augusta. From both an objective as well as subjective standpoint, elastic non-parametric bit allocation performed better than parametric bit allocation. In essence, elastic non-parametric bit allocation determines the required number of bpp for the ROI to achieve the desired quality level, coupled with an automatic update algorithm to improve ROI quality. On the other hand, parametric bit allocation makes the bpp assignment empirically, and is therefore likely to be inadequate compared to elastic non-parametric bit allocation.

To summarize, the research shows the usefulness of ROI processing as a means of achieving a gain (in a bits per pixel (bpp) sense) over uniform compression at the same bitrate. For example, if the ROI is coded at 0.4 bpp and occupies ¼ of the frame area, and the BKGND is coded at 0.1 bpp and occupies ¾ of the frame area, the average bpp for the complete frame is 0.175bpp. In other words, in a bpp sense, there is a gain of 2.3 relative to uniform compression at the same bitrate. It also shows how quantifying a notion of functionally lossless video quality – diagnostically lossless video quality in a video-based telehealth system, in a bits per pixel sense is useful from an applications and bitrate perspective. This is because uncompressed color video requires 12 bpp (with 8 bpp for luminance and subsampled color components), mathematically lossless color video requires about 4-6 bpp, and coding video at the lowest level of fidelity typically requires 0.1 bpp. Thus, there is a wide intermediate range that corresponds to bpp requirements for various applications, but these have not been quantified. In other words, it is unclear whether the requirement is closer to the lowest level of fidelity, or to the mathematically lossless level. This work has determined the answer to this question for the specific application of mobile telehealth as 0.4 bpp. Finally, the research shows how a combination of these two concepts to realize diagnostically lossless ROI video quality, is a viable enabler of mobile medical assessment achievable over bitrate limited wireless channels. The result of the research is to be regarded as an important proof-of-concept in a challenging interdisciplinary area. This thesis lays the scientific foundation for additional validation through prototyped technology, field testing, and clinical trials.

7.2 Future Work

More Extensive Testing with MCG

The results presented in this research have not been statistically validated. For this purpose, many more training and test videos of patients are required, and also many more doctors to evaluate them. Also, the algorithms need to be tested with other video resolutions. Then, a prototype of the system may be built that can be tested on an actual wireless testbed. Following this, actual clinical trials of the ROI system may be performed.

Extensions to H.264 (AVC)

It must be noted that the 0.4 bpp value for DL is specific for the class of medical videos considered in this work and also the MPEG-2 digital video compression standard used to encode the videos. A smaller value

will result if a more compression efficient video standard is used. H.264 [44, 45, 46] is a state-of-the-art digital video coding standard that is identical to MPEG-4 Part 10 or Advanced Video Coding (AVC). It is known for its compression efficiency and also robustness to channel errors. It achieves its compression efficiency as a result of not one, but a host of incremental as well as significant improvements over the MPEG-2 standard. For example, it employs intra prediction, several inter prediction modes based on macroblock partitions and sub-partitions, ¼ pel motion estimation as opposed to just ½ pel motion estimation, context adaptive binary arithmetic coding (CABAC) in addition to variable length coding (VLC), an in-loop deblocking filter etc. It achieves its error resilience due to methods such as data partitioning (DP), redundant slices, and flexible macroblock ordering (FMO), reversible variable length coding (RVLC) etc. For these reasons, it is a strong candidate for video transmission over VBR wireless channels.

Rate control methods in MPEG-2 and H.264 have some similarities but several differences. The similarities are the rate control hierarchy – GOP layer, followed by frame layer, followed by slice layer, and finally macroblock layer. The differences arise due to the plurality of prediction mode options available in H.264. This leads to the so-called chicken-and-egg dilemma in H.264, which is briefly described as follows. Mode decision in H.264 depends on the rate i.e. the bits generated as a result of encoding, which clearly depends on the quantization parameter (QP). The quantization parameter model is a quadratic one, which involves the bit budget and distortion measure (typically, the mean absolute distortion (MAD)). However, the MAD depends on the chosen mode. Thus, there is a cyclical relationship which needs to be broken in order to arrive at a solution. Practical H.264 encoders typically adopt one of two approaches to achieve this – quantization parameter prediction e.g. previous frame average QP, or MAD prediction e.g. previous frame MAD. However, the chicken-and-egg dilemma does not affect our algorithms because our work is more concerned with bit allocation to regions at the frame or GOP level, and does not explicitly deal with QP calculation.

Efficient Semi-Automatic ROI Segmentation and Tracking

In our work, we assumed the ROI was known a priori, based on medical experts' requirements on the diagnostically important regions in the video scene. Furthermore, for convenience, the ROI was assumed fixed from frame-to-frame, and whenever necessary, the ROI was altogether changed to a new set

of coordinates. This assumption is justified based on the fact that the camera was held still throughout the filming period, with some zooming. In other words, the field of view captured by the camera was more or less fixed. Furthermore, the motion of the object of interest i.e. a baby patient, was usually limited, and this was consistent with our assumptions.

However, the above assumption can be relaxed to accommodate a scenario where the ROI may move moderately or even significantly from frame-to-frame e.g. a sudden turn of a baby's face. In this case, with a fixed ROI assumption, the actual object of interest may move out of the ROI and become part of the BKGRND, and thus get coded at lower quality than expected. In order to avoid this, ROI tracking needs to be incorporated into the system.

There are several methods in the literature to perform ROI tracking based on color and/or motion information. For example, one class of motion methods based on projection assumes the object to be completely rigid and estimate a pair of translational motion vectors for the ROI from frame to frame. Another class relaxes the rigidity assumption, and estimates a new boundary based on motion vectors of individual macroblocks that comprise the ROI. The disadvantage of this scheme is its heavy computational complexity, although motion vectors generated in the encoding process itself may be used to perform the tracking. A third simpler class only projects the motion vectors of the boundary macroblocks, but this approach fails if there is occlusion.

It is assumed that the selection of the ROI is the user's task, and is performed at the decoder end. In other words, the user selects the ROI whenever he chooses to, or whenever he/she considers ROI tracking to be inefficient. The size and shape of the ROI is transmitted over a feedback channel from the decoder to the encoder. As already mentioned, in our work, we assumed the ROI size and shape as known a priori. In order to reduce delays, predefined ROI templates i.e. fixed sizes and shapes may be explored, so that this information is quickly transmitted to the encoder. From the standpoint of compression, it may be worthwhile to switch to a new user-defined ROI at the beginning of a new GOP i.e. at the start of a new I frame. Otherwise, the new ROI macroblocks will be predicted from their corresponding motion compensated macroblocks in reference frame(s) which may belong to the BKGRND, and thus may be of lower quality. This means that the prediction will be poor, and this will affect compression because larger macroblock residues require greater bits to encode.

Channel Issues – Error Resilience and VBR Issues

In our work, we operated under the assumption of no packet losses. However, a signal received over a wireless channel exhibits considerable fades in the signal strength. Unlike a wire-line channel wherein the signal strength is relatively constant and the errors in reception are mainly due to the additive noise, the errors in a wireless channel are predominantly due to the time varying signal strength caused by the multi-path propagation from local scatters. Thus, the errors in a wireless channel tend to be bursty, with the duration of bursts being a function of the receiver velocity and the nature of the time varying environment.

For two-way video communication over a narrow-band wireless channel, as we have seen, the video encoder applies motion-compensation and variable length coding to reduce the temporal and spatial redundancies. This increases the compression ratio but makes the signal susceptible to transmission errors. Even a single-bit error may cause the error to propagate over many frames because of the dependencies introduced by motion-compensation and variable length coding. The degree of severity depends on the location of the error(s).

To improve video quality under transmission errors, error-resilience schemes can be performed at the source or channel coding stages. Various source coding schemes [47] like reversible variable length coding (RVLC) and multiple description coding (MDC) have been proposed. As already mentioned, H.264 introduces additional error resilience tools such as DP, FMO, and redundant slices. Another approach is to protect the integrity of the bit-stream by using channel coding schemes, such as forward error correction (FEC) codes or automatic retransmission request (ARQ) schemes. Yet another approach concerns the decoder, which assumes the responsibility of remedying as much as possible the effects of errors and removing visually annoying artifacts. In a ROI-BKGRND video coding scheme such as ours, error resilience is more important for the ROI than the BKGRND. However, in a medical application, ROI error fixes using decoder-based schemes such as region-similarity based matching may be unacceptable. Such fixes are acceptable for the BKGRND. Thus, other robust approaches are required that do not compromise the diagnostic value of the ROI. One approach is to provide better protection against errors in the ROI as opposed to the BKGRND [48].

Delay considerations and buffer design are especially crucial for wireless video transmission. Figure 41 [49] shows a generic wired-wireless system for video transmission. A video source is connected to a wireless access point through a high bandwidth, low error rate channel. Therefore, the transmission between the video source and the access point may be assumed to be error-free. The video client is connected to the wireless access point through a wireless channel. The video transcoder is located at the access point and dynamically adapts the video transmission rate for the wireless channel. The encoder buffer is required to smooth out the mismatch between the encoding rate and the wired channel rate. The transcoder buffer is required to smooth out the mismatch between the transcoding rate and the wireless channel rate. The decoder buffer is required to smooth out the differences between the wireless channel rate and the decoding rate.

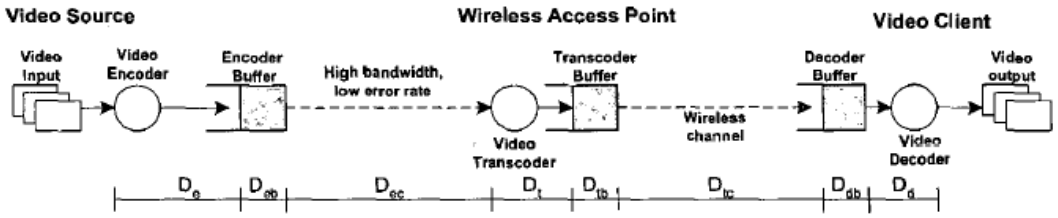


Figure 41. Wireless video communications system with transcoder.

The Ds represent the delay each frame experiences at various stages in the above system. If the overall delay is constant, the system will function normally as long as the decoder buffer does not underflow, which guarantees that the decoder has received the data of a video frame before it is scheduled to be displayed. To achieve this, however, the buffer sizes must be chosen appropriately, and it turns out, as we intuitively expect, that there is dependency between required buffer size and available channel bitrate [47, 49, 50, 51, 52, 53]. The available channel bitrate is thus necessary for this purpose as well as for rate control. Channel models [47, 49, 50, 52, 53] are necessary to estimate expected values of rate for following frames or GOPs.

APPENDIX A

MEDICAL EXPERT EVALUATIONS OF UNIFORMLY COMPRESSED VIDEOS

Table 18 lists the evaluations of 7 videos by expert 1 which were not included in the main document due to space constraints. The key medical features are listed in Table 17.

Table 17. List of key medical features.

A-Activity	RR-Respiratory Rate
CE-Chest excursion	R-Retractions
G – Gasping	T-Tachypnea
LA-Level of activity	WB-Work of breathing
MR – Mild Retraction	HB-Head bobbing
MS-Mental status	RE-Respiratory excursion
NF-Nasal flaring	STM-Skin tone mottling

Table 18. Medical expert evaluations of 7 uniformly compressed videos.

Video Sample Number (NAME_TBR)	Feature Sets	DL? (1-4)	Comments	PL? (Yes/No)
13_1000	WB	3	Difficult to clearly define	No
	RR	2		
	T	3		
13_500	T	3	Not clear image, but good for grow findings	No
	WB	3		
	RR	1		
13_1500	WB	2	Non-diagnostic	No
	RR	3		
	T	3		
13 original	WB	3		Yes
	RR	3		Yes
	T	3		Yes
14_1000	MS	3	Needs better lighting for wide Shate	Yes
	WB	3		Yes
	RR	1		
14_500	WB	3	Very pixilated, few fine details	No
	MS	1		
	RR	1		

Table 18 Continued

14_1500	MS	4	Appears diagnostic	Yes
	WB	4		Yes
	RR	4		Yes
14 original	WB	4	Noted work of breathing	Yes

			by noting movement of the patient's	
	RR	4	gown. Not	No
	MS	4		Yes
16_1000	WB	2		No
	RR	2		
	T	2		
16_500	T	3	Unclear image, poor fine details	No
	WB	1		
	RR	1		
16_1500	WB	2	Not able to see fine features	No
	RR	3	Resp rate Diagnostic	Yes
	T	3		
16 original	WB	4	Patient facing away on NF and MS	Yes
	RR	4		Yes
	T	4		Yes
12_1000	T	3	Remote lighting is poor	No
	R	2		
	WB	3		
12_500	R	3	Not a clear, but an adequate	Yes
	T	3	image	Yes
	WB	3		Yes
12_1500	R	3	Poor light	Good, but not perfect
	T	4		Yes
	WB	3		No
12 original	WB	4		Yes
	R	4		Yes
	T	4		Yes
21_1000	R	4	Good clip	Yes
	A	4		Yes
	MS	4		Yes
	T	4		
21_500	R	3	Adequate detail	Yes
	A	3		Yes
	MS	3		Yes
	T	3		Yes
21_1500	R	4	Clear enough for diagnostic	Yes
	MS	4	purposes	Yes
	A	4		Yes
	T	4		

Table 18 Continued

21 original	R	4		Yes
	A	4		Yes
	MS	4		Yes
	T	4		Yes

10_1000	MS	4	Clear	Yes
	NF	4	Clear	Yes
	A	4		
10_500	MS	3	Fuzzy	No
	A	3		No
	NF	2		No
10_1500	MS	4	Diagnostic	Yes
	NF	4		Yes
	A	4		
10 original	NF	4	Focuses on the face;difficult to get an overall impression	Yes
	MS	4		Yes
	A	4		
20_1000	R	3	Clip is not clear, but it is possible to define details in the image	Yes
	T	3		
	MS	3		
20_500	R	2	Pixilated, not diagnostic	No
	T	2		No
	MS	1		
20_1500	R	4	Diagnostic	Yes
	T	4	Diagnostic	Yes
	MS	4	Gives impression, but not diagnostic for mental status	Good
20 original	R	4	Facing away	Yes
	T	4		Yes
	MS	4		Yes

REFERENCES

[1] P. Stamford et al, The significance of telemedicine in a rural emergency department, *Proc. IEEE Engineering in Medicine and Biology,* Jul/Aug 1999, pp. 45-52.

[2] D. Kofos et al, Telemedicine in pediatric transport: a feasibility study, *Pediatrics 102*, 1998.

[3] S. K. Yoo, Design of multimedia telemedicine system for inter-hospital consultation, *Proc. 26th Annual International Conference of the IEEE EMBS,* 2004, pp. 3109-3111.

[4] R. Y. Qiao et al, A critical care telemedicine system on broadband IP networks, *IEEE Asia-Pacific Conference on Communications,* 2005, pp. 852-856.

[5] S. Wang et al, Remote evaluation of acute ischemic stroke: Reliability of national institutes of health stroke scale via telestroke, *Stroke 2003 34* : 188e, 2003.

[6] C. Kugean et al, Design of a mobile telemedicine system with wireless LAN, *IEEE Asia-Pacific Conference on Circuits and Systems,* 2002, pp. 313-316.

[7] K. A. Banitsas et al, Adjusting DICOM specifications when using wireless LANs: the MEDLAN example, *Proc. 25th Annual International Conference of the IEEE EMBS,* 2003, pp. 3661-3664.

[8] Y. Chu et al, A mobile teletrauma system using 3G networks, *IEEE Transactions on Information Technology in Biomedicine, 8*(4), pp. 456-462, 2004.

[9] A.C.W. Wong et al, On a region-of-interest based approach to robust wireless video transmission, *7th International Symposium on Parallel Architectures, Algorithms, and Networks*. 385-390, 2004.

[10] D. Chai et al, Foreground/background bit allocation for region-of-interest coding, *Proc. 2000 International Conference on Image Processing*, 2000. 2:923-926.

[11] M-J Chen et al, ROI video coding based on H.263+ technique with robust skin-color detection technique, *IEEE Transactions on Consumer Electronics 49*(3): 724-730, 2003.

[12] Sun et al, Region-based rate control and bit allocation for wireless video transmission, *IEEE Transactions on Multimedia*, 2006, *8*(1): pp. 1-10.

[13] Lai et al, A content-based bit allocation model for video streaming, *IEEE Conference on Multimedia and Expo*, 2004, pp. 1315-1318.

[14] Fung et al, Region-based object tracking for multipoint videoconferencing using wavelet transform, *IEEE Conference on Consumer Electronics*, 2001, pp. 268-269.

[15] Yang et al, Rate-distortion optimizations for region and object based wavelet video coding, *IEEE Asilomar Conference on Signals, Systems, and Computers*, 2000, pp. 1363-1368.

[16] Lin et al, Dynamic rate control in multipoint video transcoding, *IEEE International Symposium on Circuits and Systems*, 2000, pp. 17-20.

[17] Liang et al, Design and implementation of content adaptive background skipping for wireless video, *IEEE International Symposium on Circuits and Systems*, 2006, pp. 2865-2868.

[18] Sengupta et al, Perceptually motivated bit allocation for H.264 encoded video sequences, *IEEE International Conference on Image Processing*, 2003, pp. 797-800.

[19] S. B. Gokturk et al, Medical image compression based on region of interest, with application to colon CT images, *23rd Annual EMBS International Conference*. 3:2453-2456, 2001.

[20] D. Gibson et al, A wavelet-based region of interest encoder for the compression of angiogram video sequences, *IEEE Transactions on Information Technology in Biomedicine, 8*(2): 103-113, 2004.

[21] S. N-Jackson et al, Influence of compression and network impairments on the picture quality of video transmissions in tele-medicine, *Proc. 35th Hawaii International Conference on System Sciences,* 2002.

[22] M. G. Martini et al, Quality driven wireless video transmission for medical applications, *Proc. 28th Annual International Conference of the IEEE EMBS,* 2006, pp. 3254-3257.

[23] D. Gibson et al, Diagnostically lossless 3D wavelet compression for digital angiogram video, *Proc. Data Compression Conference,* 2002.

[24] R. Ashraf et al, Diagnostically lossless compression of medical images – 2, *Proc. 8th International Symposium on Signal Processing and its Applications,* 2005, pp. 227-230.

[25] D. Wu et al, Perceptually lossless medical image coding, *IEEE Transactions on Medical Imaging, 25*(3), 2006, pp. 335-344.

[26] Lim, J. and J.B. Ra. 1999. A semantic video object tracking algorithm using three-step boundary refinement. Proceedings, 1999 International Conference on Image Processing. 2:159-163.

[27] Lim, J. and J.B. Ra. 2001. Semi-automatic video segmentation for object tracking. 2001 International Conference on Image Processing. 2:81-84.

[28] Giusto, D.D. et al. 2002. A fast algorithm for video segmentation and object tracking. 14th International Conference on Digital Signal Processing. 2:697-700.

[29] Park, D.W. et al. 2000. Fast object tracking in digital video. IEEE Transactions on Consumer Electronics. 46(3): 785-790.

[30] Habili, N. et al. 2004. Segmentation of the face and hands in sign language video sequences. IEEE Transactions on Circuits and Systems for Video Technology. 14(8): 1086 – 1097.

[31] Liu, B. et al. 2005. Automatic tracking of region of interest in neurosurgical video. Proceedings of the IEEE 31st Annual Northeast Biomedical Conference. 9-10.

[32] Sullivan et al, Rate-distortion optimization for video compression, *IEEE Signal Processing Magazine*, 1998, *15*(6): 74-90.

[33] Wiegand et al, Rate-constrained coder control and comparison of video coding standards, *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, *13*(7): 688-703.

[34] Joch et al, Performance comparison of video coding standards using lagrangian coder control, *IEEE International Conference on Image Processing,* 2002, pp. 501-504.

[35] Zhang et al, Augmenting STM5 bit rate control scheme for improving coding quality of MPEG2 video encoders, *IEEE International Symposium on Circuits and Systems,* 1997, pp. 1297-1300.
[36] Chung et al, Quantization control for improvement of image quality compatible with MPEG-2, *IEEE Transactions on Consumer Electronics,* 1994, *40*(4): 821-826.

[37] Lee et al, Target bit matching for MPEG-2 video rate control, *IEEE International Conference on Global Connectivity in Energy, Computer, Communication, and Control,* 1998, pp. 66-69.

[38] Ding et al, Rate control of MPEG video coding and recording my rate-quantization modeling, *IEEE Transactions on Circuits and Systems for Video Technology,* 1996, *6*(1): 12-20.

[39] Rao, S. and N. Jayant, Optimizing algorithms for region-of-interest video compression with application to mobile telehealth, *IEEE International Conference on Multimedia and Expo*, 2006, pp. 513-516.

[40] Rao, S. and N. Jayant, A flexible approach to spatial quality distribution in wireless medical video, *Third IASTED International Conference on Circuits, Signals, and Systems*, 2005, pp. 194-199.

[41] Rao, S. et al, Region-of-interest medical video over wireless networks using motion compensated temporal filtering, 2005, *Third IASTED International Conference on Circuits, Signals, and Systems*, pp. 212-217.

[42] Stachura ME, Rao SP, Khasanshina EV, Pearson-Shaver A, Robertson SL, Region of interest video compression: Delivering diagnostic quality video over limited throughput mobile telemedicine networks, *Proceedings of the Third IASTED International Conference on Telehealth*, 2007, pp. 176-181.

[43] Jayant N, Rao SP, Khasanshina EV, Stachura ME, A rigorous technical framework for video communications for mobile telehealth, *12th Annual Meeting of the American Telemedicine Association*, 2007 (In Press).

[44] Sullivan et al, Video compression – from concepts to the H.264/AVC standard, *Proceedings of the IEEE,* 2005, *93*(1): 18-31.

[45] Lee et al, Improved lossless intra coding for H.264/MPEG-4 AVC, *IEEE Transactions on Image Processing,* 2006, *15*(9): 2610-2615.

[46] List et al, Adaptive deblocking filter, *IEEE Transactions on Circuits and Systems for Video Technology,* 2003, *13*(7): 614-619.

[47] Aramvith et al, A rate-control scheme for video transport over wireless channels, *IEEE Transactions on Circuits and Systems for Video Technology,* 2001, *11*(5): 569-580.

[48] Jerbi et al, Error-resilient region-of-interest video coding, *IEEE Transactions on Circuits and Systems for Video Technology,* 2005, *15*(9): 1175-1181.

[49] Lei et al, Rate adaptation transcoding for video streaming over wireless channels, *IEEE Conference on Multimedia and Expo*, 2003, pp. 433-436.

[50] Ortega et al, Rate control for video coding over variable bit rate channels with applications to wireless transmission, *IEEE International Conference on Image Processing*, 1995, pp. 388-391.

[51] Lee et al, Optimal rate control for video transmission over VBR channels based on a hybrid MMAX/MMSE criterion, *IEEE Conference on Multimedia and Expo*, 2002, pp. 93-96.

[52] Lee et al, A practical rate control algorithm for VBR MPEG-2 video transmission over ATM networks, *IEEE Transactions on Consumer Electronics,* 2000, *46*(2): 257-264.

[53] Stockhammer et al, Streaming video over variable bit-rate wireless channels, *IEEE Transactions on Multimedia*, 2004, 6(2): 268-277.

[54] Yang, X. et al. 2005. Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology*. *15*(6): 742-751.