Customized Video Quality Metrics for Telemedicine Applications

K.Kardaras, I.Skalkidis, G.Konis, D.Koutsouris

***Abstract*— The paper investigates the qualitative and quantitative characteristics that degrade the perceived video picture quality for a variety of short MPEG-4 video clips. A mixture of subjective and objective picture quality metrics will be tested in the laboratory before a new assessment technique is developed. The new method will be used to assess short MPEG-4 clips, which will be used for telemedicine real time medical video services transmission. The proposed method attempts to degrade the picture quality in order to establish the frame loss that different medical video content can tolerate before picture pauses are encountered. The expected output is a stand-alone technique that process medical video in real time by predicting thresholds before picture pauses occur.**

I. Introduction

The rapid increase in the industry of telemetry and telemedicine applications within the last few years requires bi-directional interactive real time video transmission from the emergency field when this occurs in an outdoors environment that lack broadband communication network infrastructures. Telemetry and telemedicine require guaranteed service delivery i.e anytime and anyplace, from a providers and the End users perspective, as far as the live video transmission is concerned. In this paper we attempt to overcome the severe limitations mobile communications experience by developing new bandwidth efficient video "transcoding" techniques. Initially a set of experiments that took place in the laboratory demonstrates how different parameters affect the perceived picture quality of the video service (QoS). In a latter stage the system that is developed make use of the knowledge gained and decides which video processing parameters will best suit the incoming content. The proposed platform that is a hybrid system that dynamically alters known characteristics, feed back information to the kernel of the system every time an unknown content is inputted and finally decides in real time for the processing parameters that suits best in order to achieve a trade of between the best QoS in the minimum bandwidth. The video coding principles and the vector tracing and refinement techniques are used as a foundation to the system intelligence. The system kernel utilizes all available techniques that are referenced in the ITU-R BT.500 directive for subjective assessment of the perceived picture quality. The scale and the methodology that we follow is described in the Tektronix reports [1] furthermore the video quality metrics versus the quantizer scale is described in the BBC R&D annual report [5]. The proposed system can be used in two discrete operation modes, first in a stand-alone capacity (End-user, provider) and second as a peer-to-peer service between the two ends of the system. The system acts as a prediction tool for the degradation of the picture quality in real time providing and indication of the maximum lossy compression characteristics it can be tolerated before the video becomes unpleasant for viewing and block mosaic effects with picture pauses appear on the screen. The system's intelligence/automaticity allows for interactive data to be feed back and new decisions to be made based on the new data entries. However the system operator can interact in real time and he is capable of altering specific parameters to accomplish the key objective. The reason arisen because MPEG encoders degrade the picture quality so much that the video becomes unviewable. In this case the quality of the picture is significant low and the doctor can't extract the regions of interest. Furthermore there is a worst-case scenario e.g. when picture pauses arisen. The proposed method behaves as countermeasures against unpredicted picture quality degradation in telemedicine video applications that demand guaranteed QoS. Objective and subjective assessment criteria are examined in the laboratory in to establish the behavioral analysis for differentiated medical video content. Certain criteria that directly affect the quality of the video picture are collected in a database in order to create a robust and reliable assessment technique. The parameters will be primarily the frame per second loss with regard to quantization levels and secondary the active window size, quantizer scale code, resolution, buffer size, macroblock size and pixel entropy. All these parameters result in significant picture quality degradation. Exceeding the threshold we set in the laboratory (23 fps) the video becomes unacceptable for medical use. Different content has different characteristics therefore special treatment must be given for individual video sources. The modular implementation allows the addition or extraction of characteristics from one video to the next. The less the number of parameters examined the more universal the model becomes. On the other hand the greater the parameters the more accurate and reliable it becomes. In the second case the model becomes content dependent and it can become customized to meet certain QoS and End user criteria/preferences.

II. PICTURE QUALITY ASSESSMENT TECHNIQUES

A set of test scenes is selected and impaired by imperfect video systems. From the video, we extract a set of candidate objective measurements that seek to quantify the perceptual impact of the video scene impairments. A panel of viewers watches the same set of test scenes and their subjective judgements of the impairments are recorded. The final step in the derivation process is a joint statistical analysis of the objective and subjective data sets. This analysis reveals which objective measurements are meaningful, and how they might be combined to create an objective metric that emulates human perception [6]. If the an assessment system is to work for a wide range of medical scenes, then scenes of sufficient variety must be used in its development. The spatial and temporal information content of the scenes are critical parameters. In order to derive the most general possible system, our library contains a number of scenes with widely varying amounts of spatial and temporal information.

A. Subjective Metrics

MPEG-x compression algorithms make use of the HVS. The human eye in conjunction with the brain's visual cortex is a very precise imaging system. It operates over a wide range of light intensities and detects colour differences. The eyes perceive picture contrast as a function of spatial frequency and light intensity. Picture width, viewer distance and visual acuity define the eye's ability to perceive and separate details within a picture frame [2]. Visual acuity increases subject to electrical signals strength, the brightness, the contrast, the luminance and the chrominance. Picture details, artefacts become visible only if there is significant difference between them and the background. Picture quality evaluation is hard to achieve since quality is subject to the user demands, individual preferences and custom standards. The simplest method is to playback in front of an audience a short video clip before and after the variation of known parameters, i.e the frame rate, the smoothness, the crispness, group of pictures and the video rate. In the end the assessor collects their comments and aggregates their ratings regarding the picture quality. These type of methods can become useful tools but they cannot stand alone to assess the picture quality because they are strictly based on individual custom user criteria. A major conflict arisen due to different users having different preferences, ideas and standards and as a result the opinions very substantially on what it is thought to be acceptable and not acceptable picture quality. Although this type of method is neither accurate nor precise it can be used as a good indication describing the audience opinion/rating. Our method suggests we average all user preferences and finally comment between two distinct levels of picture quality.

- **Acceptable** quality
- **Unacceptable** quality

Both of them refer to the **quality degradation and the number of pauses experienced** during the play back time of a short video clip [3]. The acceptable assessment scale refers to all possible combinations of picture degradation i.e. block mosaic effects, sharp knife-edges and color disorder reproduction. The unacceptable scale refers to all types of picture frames distortion i.e. the frame loss and the picture pauses

B. Objective Metrics

This is the most demanding approach in terms of software and hardware tools. It is the most professional approach because dedicated equipment (quality meter test-bed) is required to measure different properties regarding the quality of the picture. Known parameters that need to be examined is the electrical signal strength, i.e the luminance, the chrominance, the component amplitude etc. However there are other properties that will degrade the picture quality of the encoded video stream [4]. These properties must be examined by special quality meters on a frame-by-frame basis. The differences between the two frames provide a measure of the degradation the reconstructed video stream experiences [7]. The paper proposes a metric to determine the lower threshold, the entropy, beyond which the quality of the video picture becomes unacceptable. The entropy represents the minimum amount of information per picture frame that needs to be preserved in order to remove the uncertainty in the reconstructed video frame before we experience picture

III. DEVELOPMENT METHODOLOGY

The test scene selection is critical for the correct derivation of results, every video stream comprise of different content with different characteristics. Spatial and temporal information content are crucial in determining the amount of video compression that is possible, and consequently, the level of impairment that is suffered when the scene is transmitted over a fixed rate digital channel. The development of the experiment took place in a controlled laboratory environment.

A. Test Scene Selection

Fundamentally the model use the six reference video clips that "Tektronix Corp." supplies for worldwide picture quality assessment [1] and reference. These tests use reference scenes to take advantage of the fact of the human eye excels at making comparisons. A universal model is constructed that can be applied to a wide range of medical and commercial video applications providing an accurate indication regarding the threshold beyond if we exceed picture pauses occur. In order to accommodate different content we need to process hundreds of medical applications.

B. Profile Generation Vs Parameter establishment

Each application is given a unique profile that is composed of two distinct entities. First we categorize the application providing information about the window size, the background, the type of graphics, the spatial and temporal dynamics, editing effects zoom in/out, random camera motion on vertical and horizontal axis, recording angle and the smoothness or crispness from one scene to the next. The second operation is to extract the quantization levels from the source video that is pre-encoded in MPEG-2. At the end the algorithm searches the database for previously encoded content with similar characteristics and take a decision for the bit rate that best suits the application. To avoid being fractionally close to the threshold the algorithm increases the video rate by twenty percent above the minimum value and this is stored in the database. Whenever a new sample is fed to the system that has different quantizing characteristics a new entry is generated to the content profile and updates the system with the most up to date information. If unpredicted mistakes appears the system maintains a record in which encoding mistakes are assessed individually. Mistakes are of prime interest and they contribute significantly to the system intelligence because each mistake enters the database and updates the respective profile. Whenever a mistake enters the database the algorithm updates the "fault tolerance status" to eliminate making the same mistake twice. Each new entry in the profile increases the minimum video rate. The system kernel is capable of all operations without the requirement for external human interactions.

IV. PICTURE QUALITY EVALUATION

In this paper we derived an alternative approach which can be applied as an enhancement tool to the foundation theory. The lower the video rate the worse the picture quality it gets in addition there is a threshold beyond which if you exceed there is no difference to obtain.

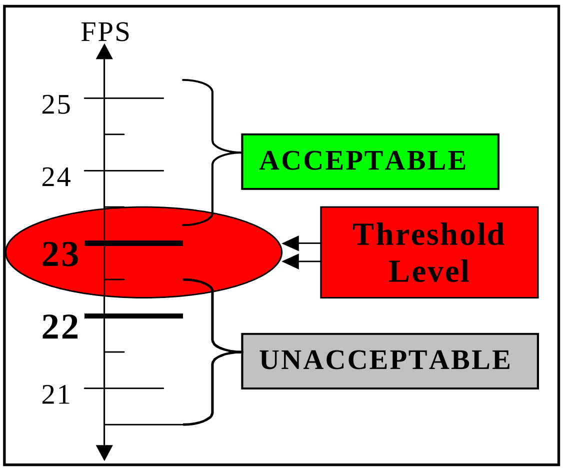

Fig 1: Picture pauses threshold

The experiments indicate that the necessary/minimum requirement for the MPEG-4 video to be played back without picture pauses is 23 frames per second. Anything above 23 fps is thought to be acceptable quality however anything below 22 fps is unacceptable due to picture pauses in figure 1. This is an objective metric for the Perceived picture quality during the evaluation process with respect to differentiated encoding schemes.

V. SOURCE MATERIAL VIDEO ANALYSIS

The source videos are grouped in two main categories [8] and all the individual characteristics are given in detail in the following sections.

A. High Temporal and Spatial Dynamics

Videos that contain scenes with large amounts of unpredictable motion that have little redundancy in their temporal statistics and little temporal compression can be expected.

B. Low Temporal and Spatial Dynamics

Videos that contain scenes with no motion or little motion are highly redundant and can often greatly be compressed with significant low impact in the perceived quality of the picture.

C. Video Specific Characteristics

The source videos have a resolution of 352x288 pixels that equals a CIF window size with standard PAL format and 25 fps. The encoding rate was 2 Mbits/sec while the encoding format was MPEG-2. The reference video clips were transcoded into MPEG-4 at various rates. The reason for choosing MPEG-4 windows media technology (WMT) was the need to deliver Real Time video streaming on demand over IP unicast/multicast to/from handheld portable devices. WMT incorporates all the above features and it is one of the most popular commercial software tools for personal and professional use. The greater the amount of fast moving picture elements within a picture frame (16x16 macroblock) the greater the need for data Bits in the MPEG-4 during video decomposition. Greater spatial and temporal dynamics result to non-linear increase in the number of Quantization Levels (QL). The greater the quantiser scale the more error is introduced. Scanning the video clips we produce a graphical representation of the Quantization scale that is used to digitize the analogue video source. The process takes place throughout the full video playback period. Therefore we produce a dynamic quantizing prediction model for a short period of time. The buffer interval was chosen to be one second. Buffer intervals may differ significantly, from a couple of seconds to minutes. The laboratory has shown that the greater the buffer size the more processing resources are required. In addition when buffer sizes are relatively large there is a greater possibility to loose information during decomposition, hence increase possibility for freezing and picture pauses. The source material, which was examined, covers a broad range of spatial and temporal dynamics. The most demanding is the "Mobile & Calendar" that requires 13.77 linear QL. On the other hand the less demanding video clip is the "Time" which requires 1.5 QL.

Video Clip	Mobl	Tens	Flwr	Cact	Susi	Time
QL	13.77	10.41	10.19	8.44	5.63	1.5

Table 1: Quantization levels for differentiated video content

If we scan through the "Tens/Table Tennis" video clip in terms of motion dynamics this is the second most demanding video clip, its characteristics match the behavior of many athletic events. The video period is 17 seconds, at the beginning one of the players shoots the ball while the camera zooms out and turns to the opposite direction to record the reaction caused by the second player. The first action scenes last about 4 seconds and the game continues for another 10 seconds, at the end the first player shoots the ball once again. As a result we observe a rapid increase in the QL for the first 4 seconds, decrease the QL as the action progressively slows down finally rapid QL increase within the last 3-4 seconds. The graphical illustration of the above scenario is given in figure 2.

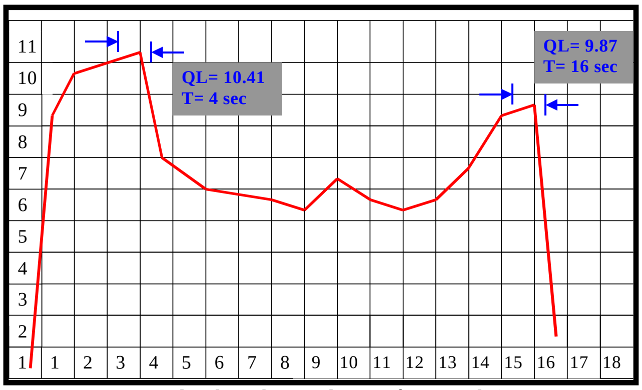

Fig 2: Quantization levels with a resolution of 1 second. (QL Vs Time)

During the video decomposition the two most probable points where picture pauses may occur are the highlighted spots at the peaks in graph. Peak points observed in **T=4 sec → QL=10.41** and **T=16 sec → QL=9.87.** We carry out the same procedure for the complete set of the samples and we derive a database. The results are presented in tabular form in table 1. Repeating the experiment we replace the source videos with medical content. The content is taken from medical surgeries and especially from laparoscope's because they are the most demanding in terms of spatial, temporal dynamics and lighting intensities. This is due to the random camera motion, the rotation and twisting deep inside the organs that goes form the wall of the organ to the inner part. This increases and decreases rapidly the light intensity, changes the texture and the color of the background. The encoder is trying to follow the rapid changes in the electrical parameters and increases the bit quantization levels with regard to the Bit rate. Laparoscope's parameters match the parameters we extracted from the to sports. The vast majority of medical video content uses approximately 11 to 13 quantization levels, that is within the profile we have already produced.

VI. IMPLEMENTATION AND TESTING

When real time medical video streaming is required the scanning process is further limited to the buffer size of the stream encoder. The limitations arisen due to the significant amount of video frames traveling in real time (data on air). In any other case the scanning process may vary subject to the buffer size from a couple of minutes to a couple of hours. For non real time medical video productions post-processing is essential for reliable decomposition. The system compress into MPEG-4 at rates approximately equal to the 40 % of the source encoding. The MPEG-2 source has 2 Mbps therefore the MPEG-4 equivalent will be about 800Kbps with 25 fps. The next step is to play back the video clip to extract if picture pauses occur. If no picture pauses encountered we repeat the procedure reducing the Bit Rate in discrete steps by a predetermined factor (subject to user criteria). The procedure generates a video clip with a Bit rate bellow which picture pauses occur (23 fps). This is the threshold level, bellow, which the system cannot exceed.

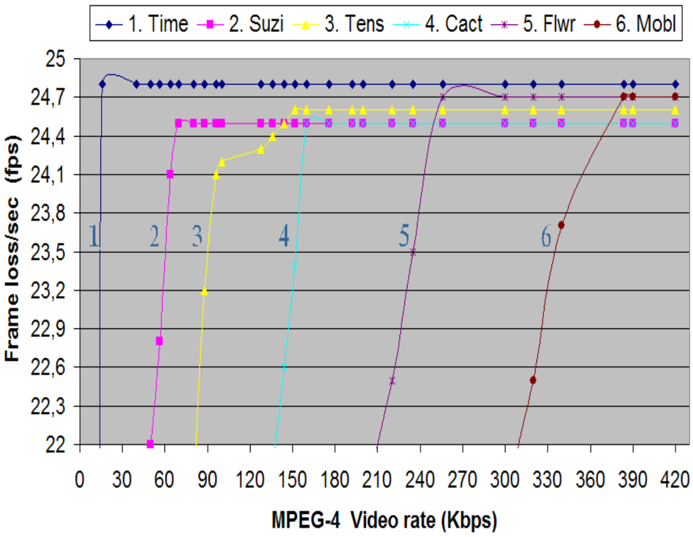

Fig 3: Picture pauses threshold Vs Video rate for differentiated video content

This threshold is a measure of Bits per second (Kbps) with respect to picture frames dropped within a second (fps). The values are stored in a database at the service provider kernel. The sequence of events repeats for each one of the reference video clips.

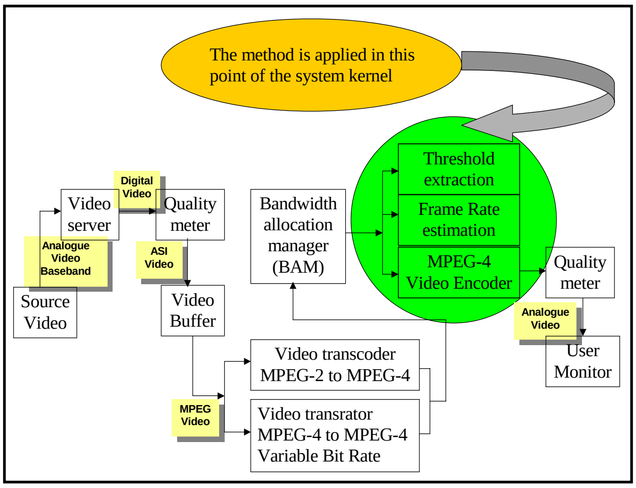

Fig 3: Block diagram of the system kernel

The kernel of the system post-processes the quantization levels versus the bit rate for each of the video clips. A profile is built utilizing the incoming information above. The profile associates the video content with the Bit rate required to succeed in a guaranteed service delivery (real time streaming without picture pauses). The greater the number of fast moving picture elements within the active window (i.e. spatial dynamics) the greater the number of Quantization levels. The system dynamically updates the database in order to generate highly customized profiles. For example a laparoscope surgery is associated with graph N_o6. A doctor giving a speech to his students is associated to graph N_o2 or graph N_o3. The method that is described above (Fig 3) provides the system kernel with artificial intelligence and the QoS is adjusted dynamically with regard to the video content by highlighting the spots where picture pauses might arise.

VII. CONCLUSION

In this paper we propose an alternative approach with regard to the assessment of the quality of a video picture that can be applied in medical applications. The complete proposal of the system provides a rather accurate prediction tool before the time picture pauses and other artefacts degrade the quality of the medical video so much that it's no longer useful. This technique can be used as an enhancement feature for post MPEG-4 processing. Degrading the quality within the limits we specify in the paper result in significant saving in the transmission bandwidth that can be directly applied for mobile video medical treatment. From an End user perspective the system can be deployed as an embedded software tool in the mobile terminal. From a providers perspective the system can operate in the core infrastructure to monitor incoming video and automatically spotting the areas that are most probable to suffer the greater degradation. The system incorporates a mixture of subjective and objective quality assessment criteria that can be used in conjunction to the current metrics. The system can assure the operator that the time intervals with the greater possibility to experience severe degradation i.e pauses/freezing can be spotted out in real time and can be tolerated with a guaranteed level of service.

REFERENCES

[1] Comparing Objective Subjective Picture Measurements
Technical Brief "Tektronix Corp"

[2] Digital Television Fundamentals
Michael Robin & Michael Poulin ISBN: 0-07-053168-4
pp 347, HVS characteristics

[3] Investigation of Wireless provision of Interactive services through the DVB-T
K.Kardaras, MSc project report, Uni KENT
pp 28, September 2002

[4] The development of objective picture quality measures that emulate HVS
Stephen Voran, U.S. Department of commerce
pp 1,Introduction to objective performance measurements

[5] Stephen Wolf,, Margaret H. Pinson, Arthur A. Webster
Institute for Telecommunication Sciences Boulder, CO
pp2 Test method

[6] The development of an objective evaluation video quality assessment syetm that emulates human viewing panels
S.D. Voran, S. Wolf
Institute for Telecommunication Sciences, US Department of Commerce

[7] Objective and Subjective Measures of MPEG Video Quality
Stephen Wolf, Margaret H. Pinson, Arthur A. Webster
Institute for Telecommunication Sciences, US Department of Commerce
pp5 Producing Frame-by-Frame Objective Parameter Values from Features

[8] The development of objective video quality measures that emulate human perception
Stephen Voran
U.S Department of Commerce