Manufacturing Agility and Hybrid Automation -I, Proceedings of the International Conference on Human Aspects of Advanced Manufacturing: Agility & Hybrid Automation, HAAMAHA'96, Maui, Hawaii, August 1996, pp183-187.

MULTI-RESOLUTION PERCEPTUAL ENCODING FOR INTERACTIVE IMAGE SHARING IN REMOTE TELE-DIAGNOSTICS

Javed I. Khan and D. Y. Y. Yun

Laboratories of Intelligent and Parallel Systems
Department of Electrical Engineering, University of Hawaii at Manoa
Holmes# 492, 2540 Dole Street, Honolulu, HI-96848, USA
javed@hawaii.edu

This paper presents our research on an interactive image transmission scheme for fast high resolution diagnostic medical video transmission. In this scheme, the viewer (or some other high level analytical or sensor mechanism) can specify either or both the maximum bit rate and the minimum quality constraints on various spatial patches of the transmitted video. The encoder, attached with a control mechanism, dynamically adjusts to the specifications with appropriate redistribution of spatial resolution and bit allocation. The proposed scheme does not require any new encoding/decoding algorithm to be designed rather it works with most existing ones. This novel scheme offers the viewer a new flexibility to efficiently manage the image resolution and the available bandwidth.

1. INTRODUCTION

This paper presents a video transmission scheme that tries to utilize interactive perceptual information to maximize bandwidth utilization and user satisfaction. Although this scheme can be generalized for many other applications involving on-demand video transmission, our particular focus is video based tele-diagnostics. CT-scan, MRI, Ultrasonogram, X-ray are few examples of image types used in medical diagnostics.

Conceptually a transmission scheme involves the (i) source, (ii) the channel, and (iii) the sink. Most current techniques for information compression analyze and take advantage of specific source and channel characteristics [Jain81]. The scheme presented in this research, looks into the very sink characteristics [Wong92]. Human eye probably is not physically capable of processing all the information that a regular video screen wants to pump into it, and only selectively process a fraction of the information what it provided to our eye in terms of digital estimate. Current encoding schemes (such as DCT, JPEG, VQ) does not make distinction between the varied perceptual importance among the various regions in the scene. But, higher compressibility can be achieved by spatially adapted local adjustments [Wong92, Mall89], if the perceptual importance of the image patches can be properly harnessed.

In this particular scheme we demonstrate an integrated video transmission (encoding, networking, and decoding) scheme, which accepts direct input about the perceptual importance of the spatial segments. This feedback can come interactively from viewer. This information is then fed to a dynamic encoder. The dynamic encoder tries to encode the video or image with variable resolution compression scheme and tries to meet the viewer ‘s expectation.

The ability to explicitly accept perceptual importance can significantly reduce bandwidth consumption, without effecting perceptual quality because, many compression “clues” are not computable just by bit-level statistical source analysis. Perceptual visualization is a highly complex cognitive process. This new viewer’s perceptual qualification based scheme can obtain “clues” directly from viewer’s feedback (or from high-level automatic analysis) and can unveil whole new scopes of video compaction.

The proposed scheme involves a perceptual constraint specification formalism and a dynamic control based encoding system. Section 2 and 3 briefly explain these topics. Finally section 3 explains the overall communication and network setup which is based on NASA’s Advanced Communications Technology Satellite (ACTS) technology [Gedn95].

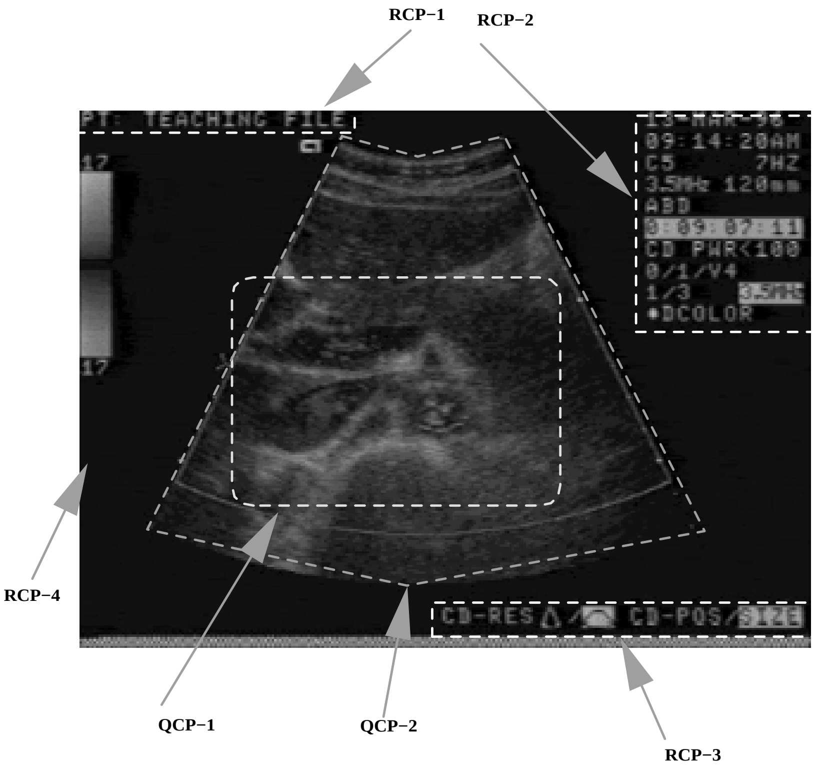

Fig-1 Patch Labels on a Scene

2 PATCH CONSTRAINT SPECIFICATION

Any image compression attempt is a compromise between (i) quality of reconstructed image and (ii) size of the encoded image or bit-rate. Encoding schemes attempt to achieve a balance between these two quantities. Consequently, the following two types of constraints are used on image segments called *time-patch*. A *time-patch* is a 3D sub-array of the 3D video data set (with *x,y* spatial and *t* temporal dimension).

Rate Constrained Patch: A *rate-constrained* time-patch *RCP(R)* is transmitted in such a way that it maintains a specified bit-rate *R* for the time-patch, and tries to maximize the quality for this rate.

Quality Constrained Patch. A *quality-constrained* time-patch *QCP(Q)* is transmitted in such a way that it maintains a quality *Q* for the patch and tries to minimize the bit rate for the this quality.

An element in the video-stream may be part of more than one time-patches, It may have both rate and time constraints. However, in such multi-constraint case a *constraint precedence order* (CPO) specification should also be given. There may be situations, where to satisfy one constraint the other may have to be violated. The CPO is used to resolve such constraint conflicts.

A possible label assignment for various time patches is shown in the example ultra-sonogram of Fig-1. The central part of the video has the actual diagnostics information. The image is also flanked by textual patches displaying various changing measurements. Perceptually, the text regions (CRP-1, CRP-2, CRP-3) need some minimum guaranteed resolution, if not the best. Therefore, these are assigned rate-constraints of values R1,R2, R3, etc. Clearly, most of the bit-rate should be allocated at the central portion of the video. This however is segmented into two quality-constraint patches. The smaller patch (QCP-1) representing the physician's principal focus of attention is assigned the highest possible quality label Q1. Viewer (physician) can dynamically move QCP-1 as needed. For the region common to QCP-1 and QCP-2, a CPO is needed. QCP-1 should take precedence over QCP-2. The entire frame patch (RCP-4) can be rate constraint. Assuming lowest order constraint for RCP-4, the encoding of background can be determined by the bits remaining after encoding all other time patches.

3. ENCODING DECODING SCHEME

Most of current data compression algorithms work in such a way that neither the quality, nor the bit-rate can be predicted exactly apriori. Even if such an algorithm is devisible, there is doubt if it would be efficient enough. In the proposed scheme, therefore, an approach is taken, which can be used in conjunction with most existing algorithms to produce the effect of constant bit-rate or constant quality for any given time patch.

The scheme requires an encoding algorithm $C=E(B,v)$, such that given a data block B, and a control parameter v it produces a code C. We will also need a decoding algorithm $B'=D(C,v)$. Let, us consider that the number of bits required to represent B is b, and to represent C is c. Then the compression ratio is $r=c/b < 1$. And the quality of reconstruction is given by $q=Dist(B',B)$. Where *Dist()* is some way of measuring the distance between the codes. We only need an additional condition that both q and r are monotonic functions of v. Most of current data compression algorithms can be modified to satisfy these requirements. For example, in MPEG-1 [Gal91], the *scale-factor* for macro blocks can serve as the valve v. Increasing Q decreases quality but increases compression. MPEG-2 offers additional features known as “scaleable modes” to encode images with variable resolution. *Vector Quantization* (VQ) based techniques, would require code books to be continuously trained with weights proportional to perceptual importance.

Our, approach is to institute a dynamic feedback control structure for (E,D) pair. Fig-2 explains the scheme. An incoming video-stream into E can be considered as a collection of data blocks $\{B(11,1), B(12,1),...,B(nn,1), B(11,2),......,B(nn,2)....,B(xy,t)...\}$. When fed to the encoder, this stream is converted to a collection of code blocks $\{C(11,1), C(12,1),....,C(nn,1), C(11,2),......,C(nn,2)....,C(xy,t)...\}$.

Once the code blocks are out, an array of measurements is taken about the rate and accuracy of the encoding. The measurements are then stored in a set of buckets arranged in a stack. There is one bucket corresponding to each time-patch constraint. For each passing code block, patch statistics are stored and accumulated in its appropriate bucket. On the other side of the bucket mechanism, the control unit reads in the cumulative statistics from the bucket and estimates the difference between the expected and the cumulative measurements. Based on this difference, it generates the appropriate corrective feedback to the valve *v*.

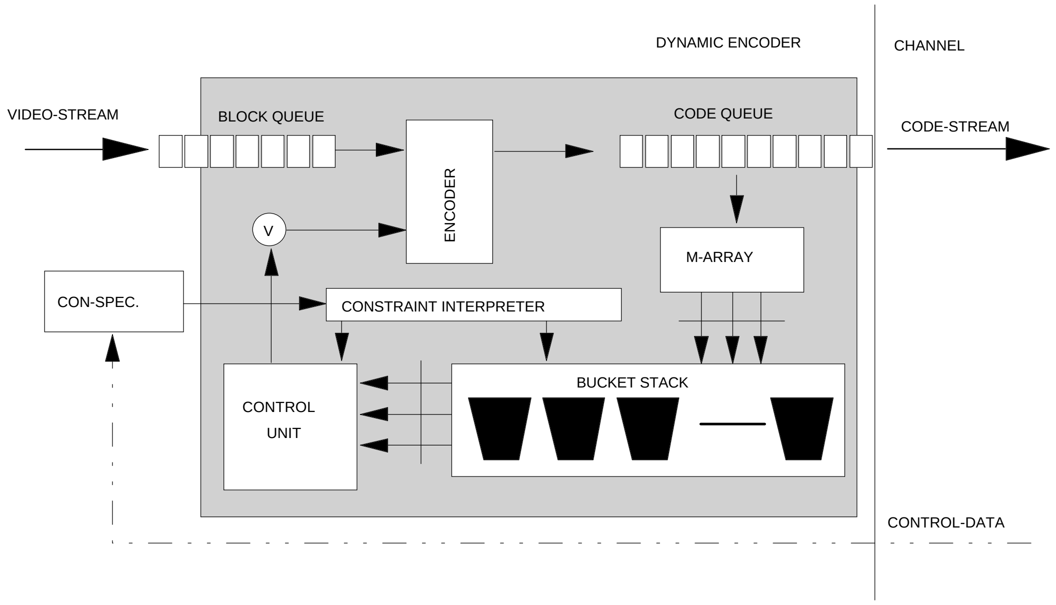

Fig-2 Transmitter Model

For encoding a RCP time-patch, the measurement unit continuously estimates the number of code bits produced as each of the code-block corresponding to this patch passes through the channel. On the other side, as each of the data-block of this patch arrives at the encoder, the control unit adjusts the valve *v* in such a way that the average bit rate follows the specified constraint rate. If a modified MPEG-1 encoder is used, then for the case of higher cumulative average, *v* should be increased. Similar control can be exerted on quality measurement (or , on any other measure on the code-stream).

The above scheme provides a number of advantages. First, the scheme can be used with many existing encoding and decoding algorithms. Secondly, it is dynamic and robust. Thirdly, the control scheme is independent from the principal data flow stream and can be executed concurrently, thus, it will have almost negligible effect on the encoding/decoding time.

4. COMMUNICATION NETWORK

The proposed transmission scheme is being developed as a part of an on-demand interactive tele-diagnostics service for remote and mobile locations based on NASA’s Advanced Communications Technology Satellite (ACTS) technology [Gedn95], ACTS enables on-demand communications between very small aperture low-cost earth stations/terminals (which can be manufactured as “consumer electronics”). Although satellite communication suffers from weaknesses such as transmission latency and medium stability, it has the unique advantage that provides on demand connectivity to any remote or mobile site (such as island nations, ships, or troops). This capability is particularly critical for medical applications.

In this scheme, the encoder at a remote medical sensor site will be connected through a two way on demand band-limited link to a physician's terminal through a combination of ACTS and fiber optics based land network. At the first phase of our implementation, we are using an enhanced MPEG based encoding and decoding scheme. Varying macro-block *scale factor* generally results in some overhead in the MPEG code. But here the *scale-factor*s are determined by control information originating from the receiver side. As a result our modified scheme can strip off this overhead from code-stream.

We expect, that for overloaded or underloaded viewer expectation, the control unit will eventually be able to estimate new upper-limit of the bandwidth requirement for an aggregate bit-rate and quality demand, and will dynamically be able to negotiate with SONET/ATM protocol to raise or lower the allocated channel capacity when necessary.

5. CONCLUSIONS

This new transmission scheme also offers several new challenges regarding the viewer interface. For example, there will always be a delay between viewer's patch specification, and transmission scheme to adjust and conform to any change in the patch specification. How much delay will be acceptable? The proposed scheme with single-hop ACTS is expected to respond within 30-50 frame delay (600-1000ms). A second concern is the artifacts due to abrupt variations in spatial resolution. We are considering various "quality smoothing" techniques to ease such artifacts.

The other principal research challenge in this initiative is the adaptive determination of the perceptual significance of image regions. As a related research we are investigating a scheme where, the spatial resolution distribution is computed from the (i) anatomical and optical geometry of eye and ratina (ii) the direction and distance of ratina from the screen. The second part of the information can be traced from a dynamic eye-glance tracer. In addition to the spatio-geometric characteristics of perceptual focus also is important is the temporal properties of eye movement. Such as the speed to tracking, the fixation time, perceptual response to loop-back time, etc. Also, it may be possible to obtain perceptual significance from high-level analysis of scene (such as, edge detection) in the encoder side.

REFERENCES

[Gall91] Gall, Dilider Le,"MPEG: A Video Compression Standard for Multimedia Applications", *Comm. of the ACM,*Vol.34, 1991, pp46-58.

[Gedn95] Gedney, R. T., "Results from ACTS Development and On-Orbit Operations", *Proceedings of NASA ACTS Results Conference,* Sept 1995, Cleveland.

[Jain81] Jain, A. K. "Image Data Compression: A Review"*, Selected Papers on Image Coding and Compression*, Ed. M. Rabbani, SPIE Vol. MS 48, 1992, pp418-458.

[Mall89] Mallat, S., "A theory of Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, no 7, July, 1989.

[Wong92] Wong, P. W., "A Multiscale Image Coder*", SPIE proc. on Image Processing Algorithms and Techniques*, v.1657, 1992, pp46-57.