Selective Visual Region of Interest To Enhance Medical Video Conferencing

Walt Bonneau Jr.[a], Christopher J. Read, Ph.D. [b], Girish Shirali, MD [c]

[a] Cubic (CTS) Corp., [b] Sony Electronics Inc., [c] Loma Linda University Medical Center

ABSTRACT

The continued economic pressure that is being placed upon the healthcare industry creates both challenge and opportunity to develop cost effective healthcare tools. Tools that provide improvements in the quality of medical care at the same time improve the distribution of efficient care will create product demand. Video Conferencing systems are one of the latest product technologies that are evolving their way into healthcare applications. The systems that provide quality Bi-directional video and imaging at the lowest system and communication cost are creating many possible options for the healthcare industry. A method to use only 128k bits/sec. of ISDN bandwidth while providing quality video images in selected regions will be applied to echocardiograms using a low cost video conferencing system operating within a basic rate ISDN line bandwidth.

Within a given display area (frame) it has been observed that only selected informational areas of the frame are of value when viewing for detail and precision within a image. Much in the same manner that a photograph is cropped. If a method to accomplish Region Of Interest (ROI) was applied to video conferencing using H.320 with H.263 (compression) and H.281 (camera control) international standards, medical image quality could be achieved in a cost-effective manner. For example, the cardiologist could be provided with a selectable three to eight end-point viewable ROI polygon that defines the ROI in the image. This is achieved by the video system calculating the selected regional end-points and creating an alpha mask to signify the importance of the ROI to the compression processor. This region is then applied to the compression algorithm in a manner that the majority of the video conferencing processor cycles are focused on the ROI of the image. An occasional update of the non-ROI area is processed to maintain total image coherence. The user could control the non-ROI area updates. Providing encoder side ROI specification is of value. However, the power of this capability is improved if remote access and selection of the ROI is also provided. Using the H.281 camera standard and proposing an additional option to the standard to allow for remote ROI selection would make this possible. When ROI is applied the ability to reach the equivalent of 384K bits/sec ISDN rates may be achieved or exceeded depending upon the size of the selected ROI using 128K bits/sec. This opens additional opportunity to establish international calling and reduced call rates by up to sixty-six percent making reoccurring communication costs attractive. Rates of twenty to thirty quality ROI updates could be achieved. It is however important to understand that this technique is still under development.

Keywords: videoconferencing, echocardiogram, pediatric, ROI, H.263, H.261, H.320, ISDN

1. INTRODUCTION

One of the more significant opportunities for video conferencing (VC) systems is in the field of medical imaging. There have been several approaches taken by VC manufactures and system integrators to address the needs of the heath care industry. Several problems have surfaced when VC systems designed for home and office with standard video & audio conferencing capabilities were applied to heath care. First issue is whether the VC equipment being used for a heath care application has filed with the FDA and/or received acceptance. Second issue of concern is that the equipment selected is truly performing the medical task required. The third issue of concern is the cost of the equipment and the reoccurring communication line cost.

The subject of heath care equipment that is being addressed in this paper is non-surgical but hospital based. Not to say that at some point in time such systems could not be used in a surgical application. The choice of medical application chosen to validate this research is pediatric echocardiograms. This subject was chosen due to the issues that surround the need to produce high quality images with frame rates equal or greater than 15/sec. This in itself would not be the only reason since there are VC systems today that can achieve 15 to 30 frames/sec. The problem and/or challenge is to make available such systems that offer low to moderate speed telecommunications and low to moderate system cost.

Part of the SPIE Conference on Image Display • San Diego, California • February 1998
SPIE Vol. 3335 • 0277-786X/98/$10.00

ISDN is still the only nationally and internationally available telephone digital line service that offers various choices of bandwidth to both home, small office and large business almost anywhere at an affordable cost for installation and monthly charges. What is known as single BRI (128k bit/sec) ISDN is the only data rate offering that is widely internationally accepted for VC systems. More so the concern however is the cost of a per minute charge. The cost of a BRI line is similar to making two standard voice telephone long distance calls. If a three-BRI or six-BRI version were used, the cost per minute would increase three to six times and the same if not more for installation if at all possible. Also important is that remote rural locations can receive installation of ISDN. In the near future the option of a DSL telephone will further improve upon ISDN capabilities. Unfortunately, this digital line offering has just started in the market place and is extremely difficult to purchase. This being said, the digital phone line service selected for this research is ISDN single BRI.

The Sony 'Mini 1000" was selected as the VC equipment of choice to demonstrate our subject: Region Of Interest (ROI). ROI is a technique to allow a VC system under the control of the technician or doctor to locally or remotely control the region of an image that is of most importance. In selecting an ROI function the technician or doctor has directed the VC processing computer or engine to expend all of the video processing cycles and bandwidth on the most important visual information. This has two net effects; the frame rate will improve two to three times although directly related to the size of the image, and the quality of the image will be improved since there are fewer pixels to process. All of this is accomplished within the telecommunications bandwidth requirement of 128k bps.

2. APPLIED MEDICAL APPLICATION

The challenges involved in providing heath care for patients in remote locations are fundamentally different from those encountered in the care of patients who are already at a tertiary-level center. For example, when a baby is born with a heart defect in a small rural hospital, the step that has the greatest impact on the baby's care is the establishment of whether or not a child has heart disease, and it can be life-saving. If the child is correctly diagnosed as having serious heart disease, treatment can be started immediately, thus preventing damage to vital body organs. If this is not done in a timely fashion, the child may suffer permanent damage, and may even die. Once the diagnosis of heart disease is established, the baby usually needs to be transferred to a tertiary level center for instant access to state of the art testing and highly specialized personnel. If cardiac disease is definitively ruled out, treatment could be provided at the rural hospital, thus sparing the child and family the inconvenience, stress and expense of transfer to another hospital.

When heart disease is suspected in a baby, the best diagnostic test available is an echocardiogram. Due to the relative infrequency of heart defects in babies, most rural hospitals do not have access to trained technicians who can perform adequate studies. Typically, technicians have limited experience with performing these tests on babies. The logistics of interpretation of these tests are also a problem. Only a physician may interpret an echocardiogram; technicians are not permitted to provide official interpretations. These small hospitals usually do not have trained pediatric cardiologists available to interpret these tests. In this context, the availability of videoconferencing technology that would allow rapid transfer of echocardiographic images to an expert at a remote location would be extremely useful to establish the diagnosis and to guide patient management. The two-way audio and video communications capability of such a system would provide for instantaneous feedback from the expert (at the remote site) that could enhance the ability of the technician (at the rural hospital) to obtain useful information from the test.

The requirements for such a system are stringent. With the increasing financial constraints on health care, such a system would need to be affordable. Diagnostic-level image quality would be essential to avoid misinterpretation of diagnostic information. For example, the echocardiographic view presented in Figure 1 is used, in association with other views, to decide whether the artery originating from the left ventricle is the aorta of the pulmonary artery. Specific echocardiographic criteria are used to make this determination. The difference between these two scenarios is the difference between a normal heart, and one with serious heart disease, i.e., transposition of the great arteries (which is eminently treatable by open heart surgery, as long as it diagnosed early). If image quality was compromised, then this determination would not be possible, with potentially lethal consequences. At the same time, high frame rates would be needed, typically in the range of 15 frames per second using de-interlaced video, or 30 frames per second using interlaced video, since babies have rapid heart rates, typically ranging from 120 to 160 beats per minute. In most instances, the useful information displayed on the screen is only 33 to 50% of the surface area of the full screen display. The system would need to be able to provide the user with the flexibility to switch between a full screen display (while accepting lower frame rates or image loss due to compression) and a partial screen display with a user-modifiable Region Of Interest (ROI) that could be used when high frame rates were essential.

Currently available video conferencing systems are either affordable, providing mediocre performance, or , when they do provide excellent image quality, are prohibitively expensive. Transmission of pediatric echocardiograms is truly an acid test of video conferencing technology. If a system were able to satisfy these requirements, then it would, most likely, be perfectly acceptable for most other forms of medical image transmission, almost all of which require lower frame rates.

3. VIDEO CONFERENCING COMPRESSION AND DE-COMPRESSION

At the technical core of video conferencing is compression and de-compression of visual and audio data. The challenge of implementing high quality VC systems involves the complete processing of VC international standards like H.320. As seen in Figure 2, this encompassing H.320 standard contains several other standards. Processing H.320 at a rate of 12 to 15 frames

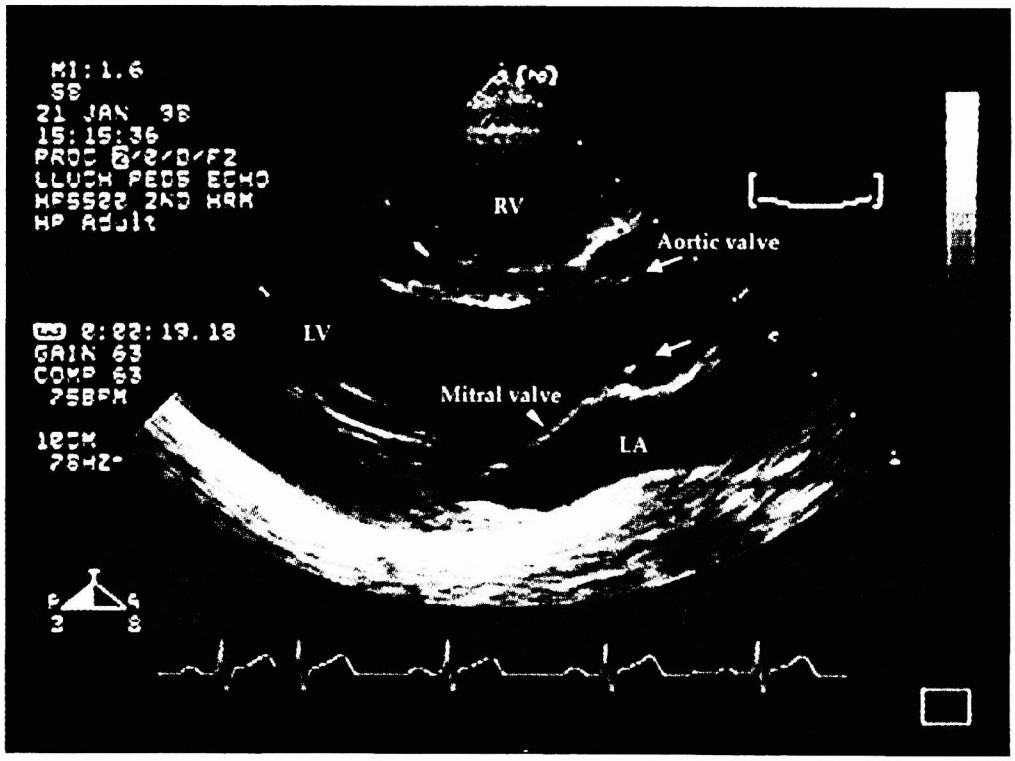


Figure 1. A Sample Echocardiogram

per second presents a significant load on any computer processor. On average this load would consume between 1.2 and 2.5 billion operations per second far exceeding the power of today's PC's. The processor used within the Sony "Mini 1000" is a TMS320C80 digital signal processor that executes five parallel instructions at one time to manage any additional computation of enhance visual algorithms. Any further requirement of processing power by enhancing H.320 must be carefully evaluated so as not to over tax the VC processor thereby, defeating any gain accomplished with execution of ROI. The purpose of this paper is to focus on the video aspects of VC therefore, the H.261/H.263 standards for compression of video data will be our subject of discussion. H.263 is slowly replacing H.261 under H.320 as the method in which video data is compressed and de-compressed.

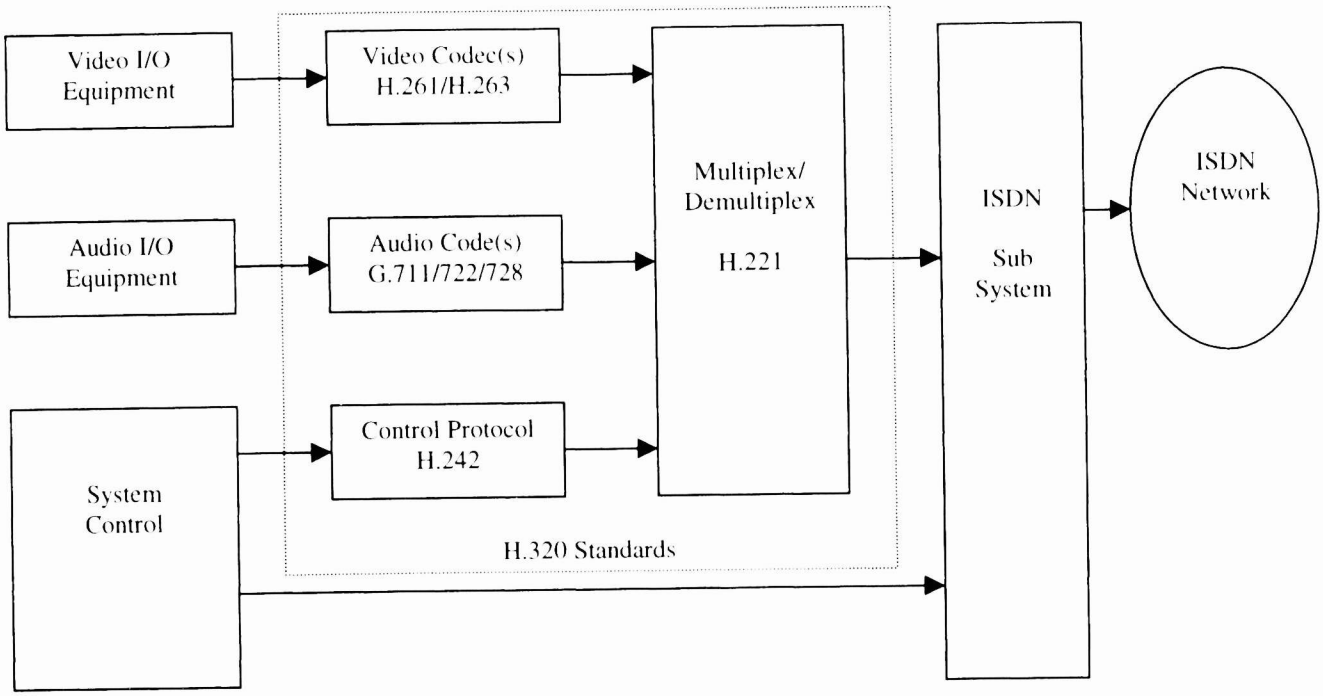


Figure 2. The Standards Components of the H.320 Videoconferencing Standard

4. NORMAL H.261 / H.263 VIDEO COMPRESSION/DECOMPRESSION

H.261 is the international standard for video compression over telecommunication lines of multiples of 64k bits per second. H.263 is a modification of H.261 targeted for lower bit rate applications and better video quality, while requiring more processing power to implement. H.261 and H.263 both operate on 352x288 images and 176x144 images. Some H.263 implementations can also operate on larger and smaller images.

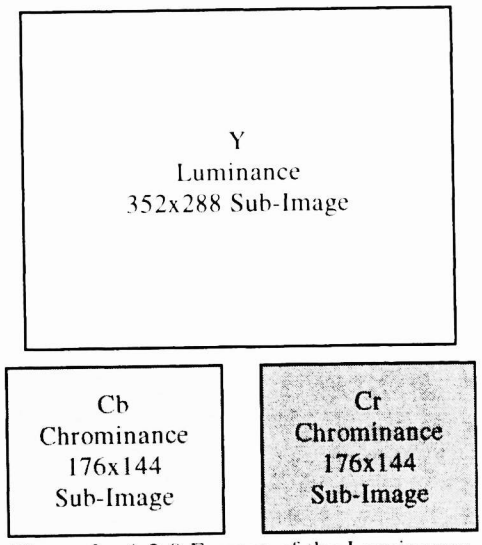


Figure 3. 4:2:0 Format of the Luminance and Chrominance image data

Both H.261 and H.263 are based on the *Discrete Cosine Transform*, or DCT. The DCT is a linear transformation of data samples, that converts pixel data into an array of data points that correspond to frequency measurements. In image compression the DCT is applied in 2 dimensions, yielding 2-D frequencies. One of the strong features of the DCT for image compression is that it concentrates most of the energy in a piece of image data into only a few output values of the DCT. Since only a few output values have much energy, they can be sent, and others not sent, effecting a reduction in the number of bits that must be sent to represent an image.

Both H.261 and H.263 work on color images represented as Y Cb Cr images in 4:2:0 format, as shown if Figure 3. This means that for a 352x288 color image, the intensity or Luminance component of the image (Y) is 352x288, but the Cb and Cr or chrominance components are only 176x144. Each of the three sub-images is broken up into 8x8 blocks. The blocks are then grouped together, 4 Y blocks (2x2), and one each of Cb and Cr that correspond to the Y blocks. This group of six 8x8 blocks is called a *macroblock*. A region of an image that has

been separated into macroblocks is shown in Figure 4.

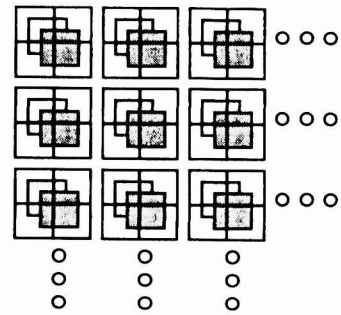


Figure 4. Macroblocks of an Image

DCT Compression

H.261/3 DCT compression applies the DCT to 8x8 blocks of image data. The transformed image data typically has very low amplitude in most of the 8x8 DCT output *bins*. The 64 bins of the DCT data are then quantized, to a fixed step size to reduce their range and minimize the number of non-zero bins. The result is shown in Figure 5. The higher the step size, the greater the image compression ratio. This is the *lossy* step in DCT based compression. The bins are then ordered so as to keep non-zero bins together (a zigzag ordering) and the values are coded using an entropy coder to minimize the data rate.

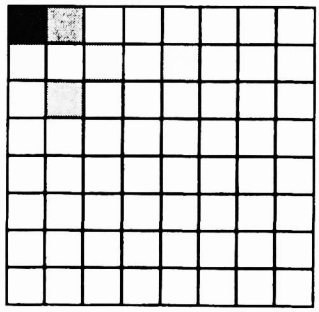


Figure 5. An 8x8 Image Block Transformed by the Discrete Cosine Transform (DCT) and quantized so most bins are zero.

H.261/3 also incorporates frame to frame compression, sending only the differences between frames, to improve compression ratios: an estimate or predicted value for a block is obtained by using the same 8x8 block from the previous frame, and the difference (8x8 block of changes in pixel values) between the block in the new image and the block from the old image is used as the input to the DCT. This greatly reduces the energy that must be transmitted to the other side when differences are small. When the differences are sufficiently small, *nothing* is transmitted for the block, and the predicted value from the previous frame is used without modification.

To this prediction mechanism, H.261/3 adds motion compensation to get a better prediction of the block in the new frame. The encoder can select which area of the old frame best predicts the new block, and thus provide a better estimate of the block in the new image. This further improves compression ratios.

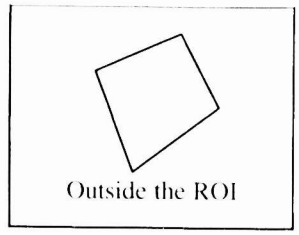
DCT Decompression

The decoder is simply the inverse of the encoder procedure. It decodes the entropy coded values into 8x8 blocks of DCT bins – the exact quantized values that were encoded, scales the bins by the quantization value to restore the proper range, performs and inverse DCT, and adds the result to the predicted value for that block.

This explanation of the encoder and decoder is simplified, as there are various details that have been omitted for brevity, but this sufficiently defines the necessary structure to explain the ROI technique. The ROI technique shown here takes advantage of several of the core features of the H.261/3, and allows an unmodified decompressor to be used.

5. THE ROI TECHNIQUE

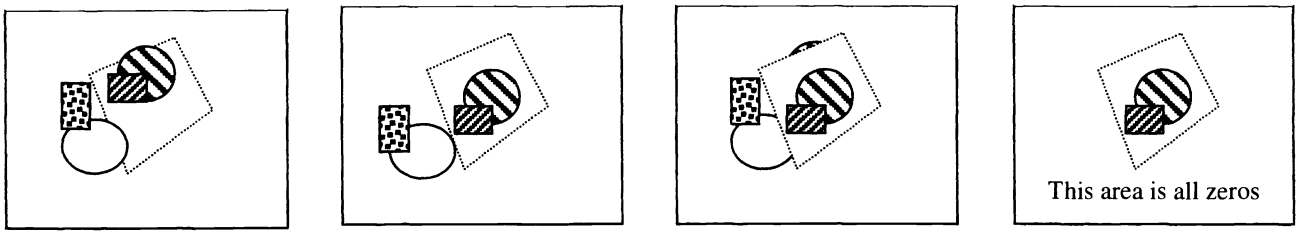
The ROI technique is built directly upon the basic principles of DCT Compression and Decompression. Some simple alterations in the encoder efficiently implement the ROI technique. There are different ways that the ROI technique can be implemented in the encoder. They will each be explained in the following sections. A sample ROI is shown in Figure 6.



The decoder for ROI VC transmissions is simply a normal H.261/3 decoder. The bitstream generated by both techniques is completely standards compliant. No adjustments need to be made to existing H.261/3 decoders.

Pixel by Pixel Merge

The technique that will likely give the most natural and pleasing appearance alters the input image and uses resulting new image as input into the encoder. The rest of the encoding process takes place just as in the normal encoder. The alteration of the input image consists of merging into the input frame the pixels from the encoder reference frame that are outside the ROI, as shown in Figure 7. With this new input image the difference between the previous reference frame and the new frame to be



encoded is zero everywhere outside the ROI. Since the difference is zero, the encoder can ignore those areas macroblocks that are entirely outside the ROI. Only those macroblocks that contain some of the ROI will have any differences between the images and have any bits encoded for them.

The decoder for the ROI version is the same as a normal decoder. The difference images are decoded as normal and added to the old image. In the ROI case, however, areas outside the ROI in the difference image are zero, so in these areas the old image is simply brought forward to the new image, and changes only really occur in the ROI region of the image.

Macroblock Discard

The second technique for implementing an ROI under H.261/3 is much simpler to implement. It selects which macroblocks are going to be in the ROI, and only transmits the entropy coded data for those blocks. This is easy to implement because the encoder already must make a decision as to how to code a macroblock: Not coded / Coded as a difference / Coded with motion compensation / etc. The encoder simply makes blocks that are outside the ROI be classified as "Not Coded", and the balance of the encoder takes care of the rest.

Again, the decoder is a normal H.261/3 decoder, and need know nothing about the ROI technique being used by the remote encoder. The macroblocks that were not coded because they were outside the ROI are simply brought forward from the reference frame like they would be in a normal decoder when a macroblock isn't coded.

6. CONCLUSION

Applying a specialized Region Of Interest (ROI) algorithm to video conferencing (VC) systems provides the possibility of achieving cost effective Tele-medicine solutions. Challenged with a application like two way transmission of echocardiograms provides the first step in validating ROI techniques. The benefits of applying ROI are many; these include higher quality motion images within the technicians or doctors selected areas of visual interest, low cost and internationally available transmission of motion images via a single BRI ISDN telephone line, low cost VC equipment, and real time diagnosis of patients that most likely will save lives, time and money while providing quality first stage heath care to individuals in both urban and rural locations. Additional research and test trial implementations of ROI should prove to be valuable in moving forward the concept of video conferencing applied in the main stream of better heath care.