Robot Vision And Video Transmission

A. RITCHIE, J. CONRADI, A. PREVOT, and E. A. YFANTIS
Image Processing Computer Graphics Intelligent Systems Lab. (ICIS)
School of Computer Science
Engineering College
University of Nevada, Las Vegas, 89154-4019
UNITED STATES OF AMERICA
yfantis@cs.unlv.edu

Abstract: In this research paper we examine the problem of distance resolution using four cameras and a noninvasive laser light for robot vision. The cameras are placed in a cross configuration and the noninvasive laser is in the center of the cross. The system constitutes the vision part of a robot used for remote patient visits by physicians. The system enables physicians to perform between visit comparisons in order to assess health patient improvements. During the patient's hospital robot visits made with the help of a nurse, the cameras take the relevant patient images and transmit them wirelessly to a file server housed in the hospital where they are stored and from where they are also being transmitted to the Phycisian's office for real time examination, and comparison with the images of the previous visit. The ROY is compressed lossless and the remaining part of the image is compressed lossy. Each time the robot visits a patient positions itself on a fixed distance from the center of the ROY. Therefore distance resolution and distance tolerance interval estimation is very important. In this research paper we model the distance with a Gamma probability distribution and we compute the tolerance interval of the distance. Our system is part of a sophisticated system that includes personal patient networks, and radiological image – video management, and secure network transmission.

Key-Words: Network programming, De-Interlacing, Compression, RAID Arrays, Client, Server, Real-Time.

1 Introduction

As the population of many industrial countries ages the demand for health care increases while the number of physicians remains about the same or, in some places, has decreased. Telemedicine, in many cases, provides an efficient way for health care professionals to increase the quality of care they provide to their patients. Robots, with the assistance of a nurse and a duplex network transmission, can enable physicians to visit with their hospital or home patients from the comfort of their office or clinic. Thus a physician can make patient visits to a number of different hospitals without having to spend time traveling from her/his office and wasting valuable time on the road, thus avoiding long road trips to remote hospitals, traffic accidents, delays etc. The time saved can be invested in her/his patients care. An important element of a robot assisting visit is the computer vision, which consists of four cameras positioned on a cross-design with the center of the cross having a non-invasive laser light. The purpose of the vision system is to obtain high quality images, and also resolve distances accurately. The robot also downloads information from personal patient network, which might include sensors monitoring the patient's heart or other organs. Furthermore with our system ambulances picking up patients from an accident scene, home, office or elsewhere, are able to transmit multiple camera video and other important information for real time evaluation and qualified medical professionals can provide advice real time. This knowledge could help paramedics treat patients efficiently, and emergency rooms to prepare for incoming patients so that they can provide the best of care. The importance of using quality cameras as opposed to capturing the video ad hoc with such methods like a cell phone is that one can control the video quality better, and also incorporate the video and audio as part of the software-hardware diagnostic intelligent system. Also, with multiple cameras it is easier to use de-interlacing [1], device noise reduction filters, resolve distances, compare and

assess progress, and incorporate the rate of progress into a healing prediction model. The system consists of raid array servers, serving several robots in a hospital with many client computers at physician sites. The system also includes administration software that adds or deletes users, schedules robots, schedules robot visitation times, and adds or deletes robots. All network programming, and all other programming modules work in a parallel multithreaded environment. Although there are high quality cameras using high quality (R, G, B) CCDs producing progressive video at 30 frames per second, for many applications is advantageous to use cameras producing interlaced video, and then deinterlace the video and produce 60 frames per second video. Video is very demanding and requires a great deal of storage. Our system performs lossless and lossy compression on the same frame. Thus on the region of interest (ROY) we perform lossless compression [2], while on the rest of the frame we perform a lossy compression, using an algorithm we have developed, which is an extension of h.264, and conditional probability gradient prediction. Camera calibration is a very important part of computer vision [3-6], and is incorporated in our system. The lossless compression used is an improvement of our previous algorithm [7-8]. Our network programming deals with Layer 3 Network layer and layer 4 transport layer of the Open System Interconnection model (OSI model), while the integration between our wireless and wired connections deals with Layer 1 Physical layer. Our system facilitates both Wi-Fi and WiMAX. Wi-Fi belongs to WLAN devices based on the IEEE 802.11 standards which are the most widespread WLAN class today. IEEE 802.11 devices are installed in personal computers, smartphones, printers, and most of the laptops, or palm-sized computers. Wi-Fi uses single-carrier direct sequence spread spectrum radio technology which is part of the family of spread spectrum systems, and multicarrier orthogonal division multiplexing (OFDM). As certain frequencies became available for public use (deregulation of radio frequencies for unlicensed spread spectrum development) the development of Wi-Fi products was enabled which competes with Bluetooth for short range transmission, and extends transmission to longer distances than Bluetooth. In the US, the Federal Communications Commission (FCC) made unlicensed spread spectrum available on May 9, 1985. Many other countries followed the lead of US and adopted the FCC regulations, making it possible for other countries to use Wi-Fi. Wi-Fi allows direct communications from one computer to another without the involvement of an access point. The Worldwide Interoperability for Microwave Access (WiMAX) is a telecommunications protocol that provides fully mobile internet access up to 40 Mbits per second and is based on the IEEE 802.16 protocol. The bandwidth is expected to increase up to 1 Gbits/sec. The WiMAX is also referred to as Broadband Wireless Access. The name was branded by the WiMAX forum in 2001, and is a standard designed to deliver “last mile” wireless broadband access. In our system the robot transmits wireless using WiMAX or Wi-Fi, to the file server, using cryptographic packages with forward error correction and automatic repeat request using Real Time Protocol (RTP) which is usually used in conjunction with RTP Control Protocol (RTCP). RTP was developed by the Audio Video Transport group and was first published in 1996 as Request For Comments (RFC 1889), it was superseded by RFC 3550 in 2003. We use the RTP in our work to transfer real time multimedia data, which includes video and audio, timestamps for synchronization and database management, sequence numbers for packet loss detection and correct replay, payload format to denote the encoded format of the data. In addition we use the RTCP for Quality of Service (QoS), and synchronization between the media streams. A session using the RTP protocol consists of IP address with a pair of ports one for RTP and the other for RTCP. Each stream has a separate session. Also, when transmitting audio and video, the video has a separate stream than the audio. In the case of streaming audio and video our software provides the option of deselecting the audio. The file servers use Ethernet Passive Optical Network (EPON). Thus we provide integration of Worldwide Interoperability for Microwave Access (WiMAX) and Ethernet Passive Optical Network (EPON). The EPON network provides tremendous bandwidth while the WiMAX network supports mobility. Integration of PON and WiMAX combines the best of both technologies. However,

the integration requires advanced secure mechanisms to overcome vulnerabilities of wireless mobile protocols. Our software provides a security framework that using our own extension of the AES protocol. We use a pre-authentication method for the authorization key (AK) pre-distribution. The paper is organized as follows: Section 2 describes the problem specifications, section 3 gives the theoretical results, which include de-interlacing, camera calibration, distance resolution, and distance tolerance intervals. Section 4 describes the system our robot vision is part of. Section 5 is the conclusions, followed by the references.

2 Robot Vision The Problem

Robot vision is a very complex problem that manifests itself in each application in a way that is unique to the application [7]. The solution to the problem is even more complex than the problem specification. In our case the requirements are that we need to be able to resolve distances of objects located with one meter from the robot with less than a centimetre accuracy. The reason for this is because the application is a medical application, whereby to ensure comparability and repeatability we need to position the centre of the vision component of the robot at the fixed prespecified distance from the centre of the region of interest of the patients operation scar. The reason for that is because our intelligent system needs to segment the extent of the scar (dimensions), the area of the tissue affected by the operation, and compare with previous visits to assess progress.

3 Robot Vision The Solution

Although television has made the transition from analogue to digital with a typical resolution of 1920 pixels across a scan-line and 1080 lines per frame, at 30 frames per second. Television stations still use interlaced cameras and transmit interlaced video which is de-interlaced and converted to progressive by the monitors used to view digital television, since the technology used to view digital television is based on computer monitor technology which is built for progressive video rather than interlaced.

The industry has been producing interlaced video cameras for a long time and that technology is more mature than the relatively new progressive video cameras which are far more expensive than the interlaced cameras. Interlaced cameras produce a frame by first producing the even lines of the frame which constitute the even field, and then $1/60^{th}$ second later the odd lines are produced which constitute the odd field. The sixty field recording is better when fast motion is in effect. It has the drawback that if the frames are not just for viewing but they are to be used for pattern recognition then due to motion within the $1/60^{th}$ of a second (muscle spasm, patient movement due to breathing or other reasons), the two fields due not much and that affects the segmentation and pattern recognition algorithms as well as distance resolution algorithms. This is the reason that before we perform camera calibration, or distance resolution or segmentation, it is desirable to perform de-interlacing first, and after that to perform any other algorithm. There are many de-interlacing algorithms developed over the past 15 years. We have developed one based on the Sinc function has desirable frequency domain properties producing frames visually pleasing and having desirable mathematical properties that facilitate camera calibration, distance resolution and segmentation.

3.1 De-Interlacing

An image with resolution N*M pixels can be thought of as a group of N signals each on having period of (1/M). Thus to interpolate along a column we can use the Sinc function:

$$f(x) = \sum_{n=0}^{N-1} f(n) \frac{\sin(\pi(x-n))}{\pi(x-n)} \quad (1)$$

Thus,

$$f(K) = \sum_{n=0}^{N-1} f(n) \quad [\text{Sin } \pi(K-n) / \pi(K-n)] \quad (2)$$

$$f(K+\tfrac{1}{2}) = \sum_{n=0}^{N-1} f(n) \quad [\text{Sin } \pi(K+\tfrac{1}{2}-n) / \pi(K+\tfrac{1}{2}-n)] \quad (3)$$

In order to reduce the number of operations, if we need to double the resolution of the column, we can use 4 coefficients to the left and 4 to the right

of the mid point to be interpolated. In order to do that the coefficients (-2/7π), (2/5π), (-2/3π), (2/π), (2/π), (-2/3π), (2/5π), (-2/7π) have to be normalized.
The normalized coefficients are,
(-2/7πa), (2/5πa), (-2/3πa), (2/πa), (2/πa), (-2/3πa), (2/5πa), (-2/7πa) (4)

where,

$a = 4 / \pi (1 – 1/3 + 1/5 – 1/7)$

$= (4 / \pi) * (76/105)$ (5)

Thus the normalized coefficients are,
(-105 / 14 *76), (105 / 10 *76), (-105 / 6 *76), (105 / 2 *76), (105 / 2 *76), (-105 / 6 *76), (105 / 10 *76), (-105 / 14 *76)
So the coefficients are,
-0.0986842, 0.138158, -0.230263, 0.690789, 0.690789, -0.230263, 0.138158, -0.0986842

f (K+ ½) = -0.0986842 f(K-3)+ 0.138158 f(K-2)-0.230263 f(K-1)+ 0.690789 f(K) + 0.690789f(K+1)-0.230263f(K+2)+ 0.138158f(K+3)-0.0986842f(K+4) (6)

Floating point operations are computationally expensive, an integer approximation to the above formula is:

f(K+½)=1/100[-10f(k-3)+14f(k-2)-23f(k-1)+69f(k)+69f(k+1)-23f(k+2)+14f(k+3)-10f(k+4)+30] (7)

The frequency response function of the above formula constitutes a bandpass filter which passes both low and high frequencies and produces sharp frames from fields, it increases the video quality and the video frequency from 30 frames per second to 60 frames per second. De-Interlacing is very helpful for camera calibration, noise-motion, separation, motion detection, motion estimation, and motion compensation, segmentation, feature vector formulation, classifier formulation, and pattern recognition.

3.2 Camera Calibration

Using the pinhole camera model [3-6], we can devise a mathematical model for the camera based on the principles of perspective projections. Using a very precise grid designed with accuracy of a mil we calibrate each camera and during the calibration we estimate the Intrinsic Camera Parameters (ICP), we also create the proper linear transformations that take us from the camera coordinate system to a global Euclidean coordinate frame. The affine transformations enabling us to go from the local camera coordinate system to the global coordinate system consist of a rotation and a translation. Let **R** and **t** be the rotation and translation that take the global coordinate system into the camera coordinate system, respectively. These are referred to as the *Extrinsic Camera Parameters*.

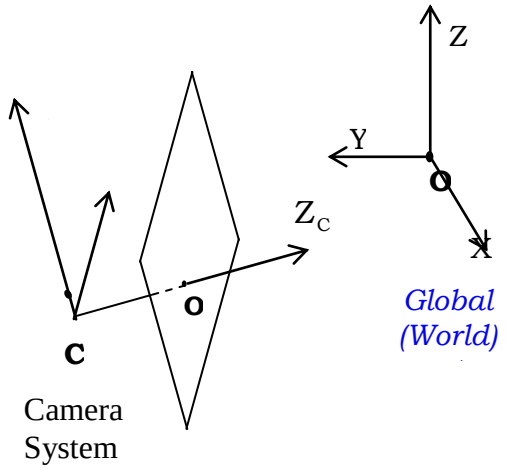

Fig. 1: The Camera Coordinate System with respect to the Global (World) Coordinate System.

In order to estimate the coordinates of a point in the three dimensional space we need at least two cameras focused on the point of interest. Each one of the cameras has to be calibrated at different depths, including the depth (or close to the depth) that the point of interest is located. Part of the calibration is the correction for the lens effect. The spherical lens deforms the image, and the deformation is proportional to the distance from the intersection between the ray and the lens from the center of the lens.

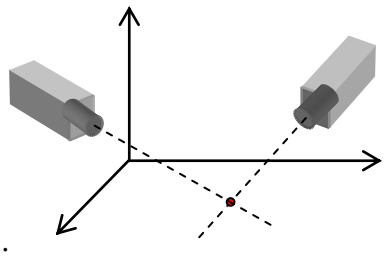

Fig. 2: Two Camera System for Stereo

3.3 Distance Resolution And Accuracy Assessment

In this section not only we develop the formulas that estimate the distance to the target but we also give tolerance intervals of the estimated distance. Due to the noise associated with the estimation process, the estimated distance is a random variable that can be modeled with the Gamma probability density function. The Gamma probability function is extremely flexible and depending on its two parameters could be an exponential, or a Chi-square, or close to a normal distribution. It could be symmetric, or skewed to the left, or skewed to the right.

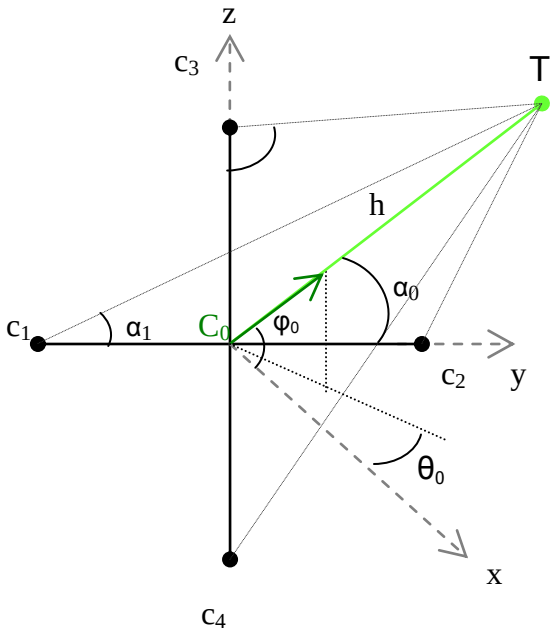

Fig. 3. The system consists of a non-invasive laser light C_0, cameras c_1, c_2, c_3, c_4 in a cross formation.

T is the target, $h = C_0T$ is the distance to the target.

Let θ_i be the angle of the projection of c_iT on the xy plane with the x-axis, and φ_i the angle between c_iT and its projection to the xy-plane. The θ_i , φ_i, , therefore are the spherical coordinates of the unit vector along the direction c_iT, i=0,1,2,3,4. The coordinates of the unit vector along the direction $c_iT = (X_i, Y_i, Z_i)$ are :

$$X_i = \cos\theta_i \cos\varphi_i \quad (8)$$
$$Y_i = \sin\theta_i \cos\varphi_i$$
$$Z_i = \sin\varphi_i$$

The dot product of the above vector with the unit vector Y = (0, 1, 0) along the Y-axis is equal to $\cos\alpha_i$ of the angle between the unit vector

$$c_iT = (X_i, Y_i, Z_i) \quad (9)$$

and the Y-axis. Thus

$$\cos\alpha_i = \sin\theta_i \cos\varphi_i \quad (10)$$

and

$$\alpha_i = \arccos(\sin\theta_i \cos\varphi_i) \quad (11)$$

i=0,1,2,3,4.

The index i=0 corresponds to the non invasive laser light source. Let β_i, i=0,1,2,3,4 be the angle between C_iT and the Z axis then

$$\cos\beta_i = \sin\varphi_i, \quad (12)$$

therefore

$$\beta_i = \pi/2 - \varphi_i, \quad i=0,1,23,4. \quad (13)$$

Let $h=|C_0T|$ be the distance of the light source from the target , then from the triangle $C_1\,C_0$ T we obtain

$$\frac{|C_1C_0|}{\sin(180 - \alpha_1 + \alpha_0)} = \frac{h}{\sin\alpha_1} \quad (15)$$

Or

$$h= \frac{|C_1C_0|sin\alpha_1}{\sin(\alpha_1-\alpha_0)} \quad (16)$$

Similarly from the triangle C_0 TC_2 we have:

$$h=\frac{|C_2C_0|sin\alpha_2}{\sin(\alpha_0-\alpha_2)} \quad (17)$$

Also from the triangle C_0 TC_3 we obtain

$$h=\frac{|C_3C_0|Cos(\varphi_3-\varphi_0)}{\sin\varphi_3} \quad (18)$$

Finally from the triangle C_0 TC_4 we obtain

$$h=\frac{|C_4C_0|Cos\varphi_4}{\sin(\varphi_4-\varphi_0)} \quad (19)$$

The value of h obtained by each camera should theoretically be the same, but in practice the four values obtained are all different. The fluctuation between the four estimates of the distance h depends on the quality of the cameras used. The camera quality is a function of the quality of the image capture chip, which is usually a charge couple device, the quality of the lens used, and the quality of the hardware software design of the camera. The camera calibration is also very important to the accuracy of distance resolution. Once the laser projects a color spot on the target, each camera moves by (θ_i, φ_i), i=1,2,3,4, until the center of the camera image space has the image of the center of the laser spot. Let H be the true distance then:

$$h_i = H + \varepsilon_i, \quad i=1,2,3,4$$

Where ε_i is the error committed by the ith camera, and is a random variable. The error is affected by several factors, included the camera, and many environmental factors. The estimated h_i $therefore$ is a random variable. Based on our experimental work the Gamma probability density function with parameters α and β provides enough flexibility and enables us to adequately model the estimated distance. Thus

$$f(h_i) = \frac{h_i^{\alpha-1}e^{-\frac{h_i}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, \quad h_i > 0, \ \alpha>0, \ \beta>0 \quad (20)$$

I=1,2,3,4. The two parameters of the gamma distribution are the same for each one of the four cameras, provided that the cameras are identical. The moment generating function of the Gamma distribution with parameters α and β, is

$$M(t) = \int_0^{\infty} e^{tx}\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}dx \quad (21)$$

Or

$$M(t) = \frac{1}{(1-\beta t)^{\alpha}} \quad (22)$$

If $W = (h_1 + h_2 + h_3 + h_4)$

Then the moment generating function of W is:

$$M_W(t) = E(e^{Wt})$$

After integration we obtain

$$M_W(t) = \frac{1}{(1-\beta t)^{4\alpha}} \quad (23)$$

Which implies that W is a Gamma distribution with parameters 4α, and β.

The average $\overline{h}$ = W/4, and the moment generating function of $\overline{h}$ is

$$M_{\overline{h}}(t) = E(e^{\frac{Wt}{4}}) \quad (24)$$

Or

$$M_{\overline{h}}(t) = \frac{1}{(1-\frac{\beta}{4}t)^{4\alpha}} \quad (25)$$

Which implies that $\overline{h}$ has a Gamma distribution with parameters 4α and $\frac{\beta}{4}$

Based on this mathematical foundation therefore in order to estimate the parameters α and β, we proceed as follows:

Let $\overline{h} = \frac{\sum h_i}{4}$, then

$\alpha\beta=\bar{h}$, (26)

and if $S^2=\frac{\sum(h_i-\bar{h})^2}{3}$, then

$\alpha\beta^2=S^2$ (27)

The estimate of β therefore is

$$\beta=\frac{S^2}{\bar{h}} \quad (28)$$

and the estimate of a is

$$\alpha=\frac{S^2}{\bar{h}^2} \quad (29)$$

The values of the parameters are introduced to the gamma distribution in order to make an assessment with regard to the accuracy of the distance resolution process. From the above we compute the formula for the probability of $\bar{h}$ which has mean α β and variance $\frac{\alpha\beta^2}{4}$ and then we compute the limits a,b for which

$$\int_a^b f(\bar{h})d\bar{h}=c \quad (30)$$

Or

Given a c value 0<c<1, find the parameters a,b, for which

$$\int_a^b \frac{x^{4\alpha-1}e^{-\frac{4x}{\beta}}}{\Gamma(4\alpha)(\frac{\beta}{4})^{\alpha}}dx=c \quad (31)$$

The values for a and b are computed numerically. The value of c is usually chosen to be 0.99. Thus the true distance h is in the interval [a,b] with probability 99%. Notice that due to the relatively small number of measurements the classical statistical methods which are based on the central limit theorem and the normal distribution are not applicable. This method is based on the estimation of the parameters α, β, from the four cameras. One can test the validity of this method during calibration, using a very accurately measured square grid where the accuracy of the lines and distances could be less than 3 mils (3/1000 of an inch), and the design of the grid is made by a very accurate cad method similar to the one used for circuit design (we used Altium package for circuit design and circuit layout for our experimental verification).

4 The System

Although the problem we solve here is the robot-vision part of the robot, in order for the reader to understand the importance of robot vision it is very important to understand how robot vision integrates with the system and the importance of our solution to the rest of the system.

Due to the HIPA regulations patient information has to be private. To ensure secure communication we use network programming (client-server paradigm) where the robot is the client and the Doctor's office is also a client. The server is in the hospital and consists of a RAID-Array which stores several terabytes of video obtained by the robot time-stamped, patient radiological images and video and patient data obtained from the patient personal network (PPN). Not all patients have PPN, only heart patients, critical condition patients, and patients that are being monitored for a number of various reasons. The four cameras of the robot are being serviced by an electronic card (hardware) which captures the video from every camera, processes it compresses it and transmits it wireless to the server. The wireless transmission is facilitated via Wi-Fi transmission compliant to the IEEE 802.11. Thus for smaller hospitals where the distances from one side of the hospital to another are relatively small, and the demand for the robots services are relatively small, Wi-Fi is sufficient. For larger hospitals WiMAX is used, based on the IEEE 802.16e standard. WiMAX offers larger reach and wider bandwidth. The transmission protocol used is the Real Time Protocol (RTP), along with the Real Time Control Protocol (RTPC) for quality of service (QoS). In case the patient has PPN, all the sensors conclude below the feet of the patient, away from the heart, and if there is a request by the physician to view the sensors, then they are uploaded to the robot using Bluetooth. Every time the robot gets video of the patient the four cameras view the region of interest (ROY) from a fixed distance. The viewer at the physician's computer shows the current view along with the previous viewing to enable the physician to make a visual comparison and understand better if there is progress or regression. Furthermore our software compares the three RGB channels to detect degree of infections and quantify progress. The images are compressed. Specifically, the region of interest is compressed losslessly while the remaining frame is

compressed lossy. In addition to compression we provide Cryptography based on our own extension of AES and also we provide forward error correction to ensure that the images have not been corrupted during their wireless and then possibly wired channel transmission. The Hospital raid array server consists of over 100 terabytes and it is a software controlled server, that hosts video from many patients and a database associated patients with physicians, robots used in each visit, nurses escorting the robots, time of the visit, duration of the visit, and commends from the physician related to this visit. Each physician has access to her/his patient video, radiological images and videos, patient personal network data, and other data and information that are patient related and can download a copy to her/his computer dedicated to this activity. The processing computer which is in the robot draws a defult region of interest based on the artificial intelligence software residing in the robot computer. The four cameras create a 3-D view of this region then they calculate the area of the region. They also identify the infected regions and quantize each one of them assigning a degree of infection based on the volume of each one of the infected areas. Then they compare the infected area of the trauma or operation as calculated in this visit with the infected area calculated by the previous visit. The intelligent system assigns improvement, relative improvement, absolute improvement, and makes a prediction about the healing period. The healing period prediction is upgraded after each visit. The physician is able to control the region of interest remotely with the mouse. Thus if she/he wants to change the default region of interest calculated by the intelligent system, then she/he can use the mouse to redraw the region of interest. The intelligent system adjusts the four cameras and computes the area and degree of infection for the new area. The region of interest is compressed lossless, using linear prediction coding based on the gradient of previeously encoded pixels in the x and y directions to predict the current pixel. The predicted value is subtracted by the current pixel value obtaining the residual. The residual is being predicted using nonlinear prediction based on conditional probability theory, were by we estimate the maximum likelihood estimate of the residual condition on the values of the neighbors to the immediate left, immediately north, left diagonal and right diagonal. The estimated residual value is being subtracted from the actual residual value and the new value is a relatively small positive or negative number. Most of the numbers obtained after the residual estimation are zero, about equal amount of numbers is 1, and -1, And about equal amount of numbers are 2 and -2. There is less amount of numbers being 1 than zero, and less amount of numbers being 2 than 1, and the numbers attenuate rapidly as we move away from zero. The probability distribution function of the numbers remaining after the residual estimation are double exponential with mean zero. After the residual estimation the remaining number have relatively small entropy. In order to avoid dealing with negative numbers first we transform all numbers to non-negatives. Thus we represent the 0 as 0, the -1 as 1, the 1 as 2, the -2 as 3, the 2 as 4, in general the –i as (2i-1) and the i as 2i. After this transformation we apply variable length encoding in the remaining data. The options in variable length encoding ad Huffman encoding or arithmetic, or vlc, or cabac, or a custom made algorithm. We use arithmetic encoding but not in the classical way where each number is encoded separately but we encode a group of four numbers at a time. This multivariate arithmetic encoding provides larger compression ratio than memoryless arithmetic encoding. For the remaing of the image we use lossy compression. The algorithm we use is an extension of the H.264. In this algorithm we divide the frames into I frames P frames and B frames. The Group of pictures (GOP) includes a relatively large number of pictures. The number of frames is a function of the amount of motion or change from one picture to the next. On the average 50 pictures are included in a GOP. Every GOP includes one I frame, every 3rd frame is a P frame, and there 2 B frames following a P frame or an I frame. The I frame is divided into 4x4 squares. The pixels in each one of the 4x4 squares are being predicted from the pixels to the left, directly above, left diagonal and right diagonal of the 4x4 square. The predicted values are being subtracted from the actual values and the obtained residuals are relatively small numbers. The residuals are transformed using an orthogonal transformation in order to decorrelate the values of the residuals. The transformation rearranges the

energy so that the maximum energy is at the left corner of the 4x4 and attenuates as the distance from the left corner increases. A 4x4 quantization matrix is applied to the transformed residuals reducing their strength and reducing the energy of most of the transformed residuals to zero. Subsequently all remaining values except for the value at the left corner are encoded using a lossless variable length encoded algorithm. The values of the corner of the 4x4 square are being encoded separately by using wavelet with the lifting scheme to further reduce the strength of the numbers. After that we use a zero-tree wavelet quantization, and conditional arithmetic encoding. The compression algorithm obtained this way is a function of the quantizer used, and for our application we obtain about 50:1 for the I frames. The P frames are being predicted by the restored I frames using forward prediction. The predicted values are being subtracted from the original values. The obtained residuals are being treated the same way as the I frames. Sections of the P frames that are very close to corresponding sections of the I frame are simply replaced from the I frame. The B frames are not used to predict any frames, but they are predicted by both I frames and P frames from left or right. The predicted values are subtracted from the original values, and the residuals are being compress the same way as the residuals of the P frames.

5 Conclusion

In this research paper we examined the vision, comparison, and transmission aspect of a robot used for remote medical hospital visits. The vision of the robot consists of two cameras or more cameras, calibrated to resolve distances with centimeter accuracy. The robot transmits the multiple images wirelessly to a file server maintained in the hospital. The hospital file server transmits to the physician's office real time, or store and forward, and provides a comparison of the present images with the images of the previous visit. The physician makes a visual comparison and the software of our intelligent system makes a comparison and provides intelligent and quantized progress information. In this research paper we addressed issues related to video capture for multiple cameras, noise, de-interlacing, camera calibration, distance resolution, network programming, secure transmission, and video comparison.

References:

[1] Yfantis, E. A. M. Au, and G. Miel, (1992) Efficient Image Compression Algorithm for Computer Animated Images, Journal of Electronic Imaging, (1)4, pp. 381-388.

[2] Yfantis, E. A., 1993, A New Quadratic and Biquadratic Algorithm for Curve and Surface Estimation, Journal of Computer Aided Geometric Design, 10(1993), pp.509-520.

[3] M. B. Teruel, (2005) Color Detection and 3D Rreconstruction for 3D Hand Pose Estimation as Part of an HCI System, MS ThesThesis, School of Computer Science, UNLV, pp. 1-85.

[4] Trucco and A. Verri, "Introductory to TechTechniques For 3-D Computer Vision", PrenPrentice Hall, pp. 1 – 343, March 1998.

[5] Salvi, X. Armangué and J. Batlle, "A Comparative Review of Camera Calibrating Methods with Accuracy", Patern recognition, pp 1617-1635, Vol. 35, Issue 7, July 2002

[6] D. Faugeras and G. Toscani, "Camera Calibration for 3D Computer Vision", Proceedings of the International Workshop for Industrial Applications of Machine Learning and Machine Intelligence, pp. 240-247, Feb. 1987.

[7] E. A. Yfantis, Method and Apparatus for Determining the Size and Shape of a Foot, Patent no. US 6,735,547 B1, 5-11-2004.

[8] E. A. Yfantis, Visually Lossless Still Image Compression for CMYK, CMY and Postscript Formats, Patent no. US 6,633,679 B1 10-14-2003.

[9] Z. Bojkovic, B. Bakmaz, A Survey On MPEG-4 Standard and Digital Television Deployment WSEAS Transactions on Communications Issue 1, Vol. 9, 2010, pp. 33-42.

[10] B. Paul, M. A. Matin, M. J. Shoakat, Z. Rahman, Optimal Placement of Base Stations in Two Tiere Wireless Sensor Network, WSEAS

Transactions on Communications, Issue 1, Vol. 9, pp. 43-52.

[11] M. A. Matin, A. I. Sayeed , A design Rule for Inset-fed Rectangular Microstrip Patch Antenna, Issue 1, Volume 9, 2010, pp. 63-72.

[12] W. Gu, S. V. Kartalopoulos, P. K. Verma, Fast and Secure Handover Schemes Based on Proposed WiMAX over EPON Network Security Architecture, Issue 2, Vol. 9, 2010, pp. 115-126

[13] C-F. Lin, An Advanced Wireless Multimedia Communications Application: Mobile Telemedicine, Issue 3, Vol. 9, 2010, pp. 207-215.