P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2008,(2), (1) :73-82

CODING OF VIDEO OVER IP-BASED NETWORKS

Nasser N. Khamiss Alani **Samara A. Elia**

Nahrain University-Baghdad-Iraq

ARTICLE INFO

Received: 1 / 4 /2008
Accepted: 24 / 4 /2008
Available online: 30/4/2008
DOI: 10.37652/juaps.2008.15425

Keywords:
Coding ,
Video ,
IP-Based Networks.

ABSTRACT

This paper is developed to study the compression performance of video by using searching method of motion estimation OTS and SNR scalable coding to enhance the quality of video sample. The work evaluates set of suitable objectives fidelity measures, such as MSE, PSNR and CR. The model of video coding system is designed to treat the video signal as a CBR and implemented for different video samples rates. The two major components inter and intra frames compression is achieved as an optimal compensation for both quality and CBR. In the same time the basic structure of communication networks which is represented by the Transmission Control Protocol/Internet Protocol (TCP/IP) model is taken into consideration within the system BER control. The developed system is implemented using Visual Basic Language (ver 6.0) under Windows Xp operation systems.

Introduction

The increasing demand to incorporate video data into telecommunications services, the corporate environment, the entertainment industry, and even at home had made digital video technology a necessity. However, the problem is that still image and digital video data rates are very large, typically in the range of 150Mbits/sec. Data rates of this magnitude would consume a lot of the bandwidth, storage and computing resources in the typical personal computer. For this reason, video compression is needed to reduce the data to be stored or transmitted through

eliminate picture redundancy [1]. Practical communications channels have a limit to the number of bits that they can transmit per second. In many cases the bit rate is fixed (constant bit rate or CBR, for example IDSN, etc.).

For video the main reduction is achieved by attempting to estimate where areas of the previous frame have moved to in the current frame (motion estimation) and compensating for this movement (motion compensation). The basic H.263 encoder generates a variable number of bits for each encoded frame. If the motion estimation/compensation process works well then there will be few remaining non-zero coefficients to encode. However, if the motion estimation does not work well (for example when the video scene contains complex motion); there will be many non-zero coefficients to encode and so the number of bits will increase. In order to "map" this varying bit rate to a CBR channel, the encoder must carry out rate control. The encoder measures the output bit rate of the encoder. If it is too high, it increases the compression by increasing the quantize scale factor, this leads to more compression and a lower bit rate but also gives poorer image quality at the decoder. If the bit rate drops, the encoder reduces

* Corresponding author at: Nahrain University-Baghdad-Iraq,
E-mail address: nasser_alani@yahoo.com

the compression by decreasing the quantize scale factor, leading to a higher bit rate and a better image quality at the decoder [2]. Fig.1 and Fig.2 show a typical encoder and decoder respectively [3].

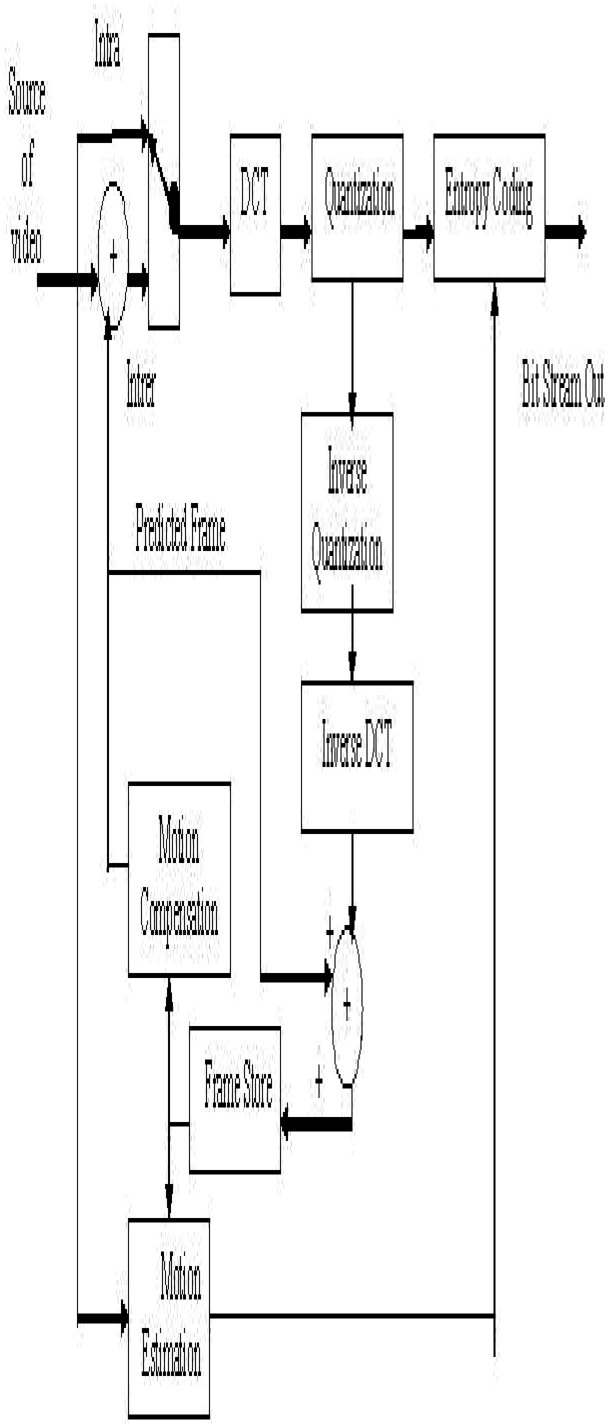

Figure 1 Block diagram of typical video encoder.

In the case of an intra coded macroblock, the encoder performs only the inverse quantization and inverse DCT operations to reconstruct the original macroblock. In case of inter coded macroblock; the decoder performs the inverse quantization and inverse DCT operations on the DCT coefficients corresponding to the prediction residual. The decoder also uses the information in the motion vectors to find the best matching macroblock in the previous reconstructed frame. The latter is then added to the residual to reconstruct the original macroblock. The complete frame is reconstructed then stored for use when decoding the subsequent frame.

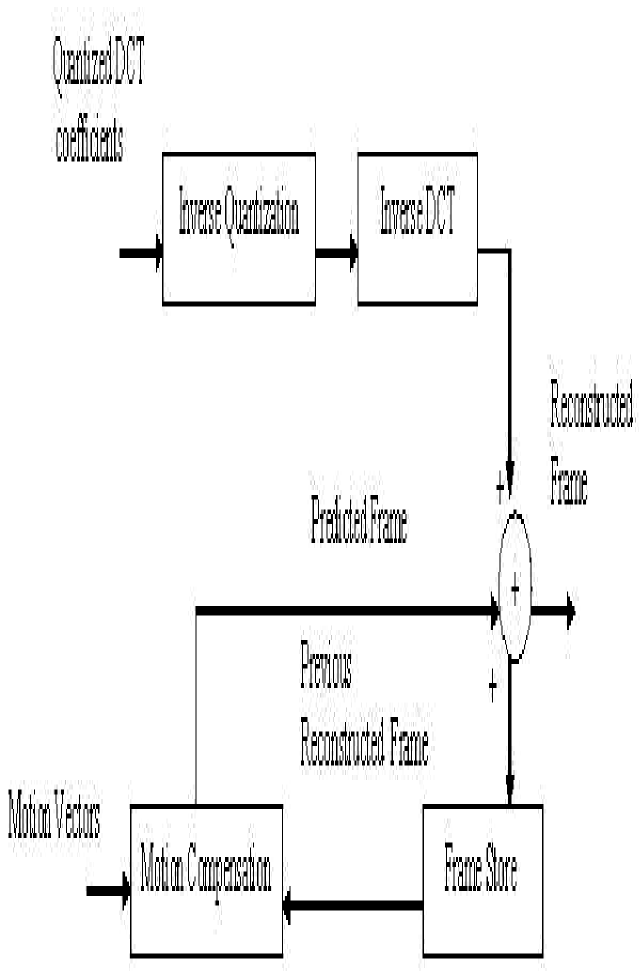

Figure 2 Block diagram of a typical decoder.

Video Quality Scaling

In MPEG and H.263 formats, many scaling techniques of video quality are standardized for scalability, spatial scalability, temporal scalability and peak signal-to-noise ratio PSNR. These scaling techniques are also known QoS parameters that are considered in this work as the objective measurement tool. The main idea behind the objective measures is to calculate Normalized Mean-Square-Error NMSE and Peak Signal-To-Noise Ratio, PSNR which are defined in equations (1) and (2) respectively [4].

$$NMSE = \frac{\sum_{i,j}^{M,N} \left(X_{ij} - \hat{X}_{ij} \right)^2}{\sum_{i,j}^{M,N} X_{ij}^2} \times 100\% \qquad (1)$$

$$PSNR = 10 \log_{10} \frac{\left| X_{P-P} \right|^2}{\frac{1}{MN} \sum_{i=0}^{M} \sum_{j=0}^{N} \left(X_{ij} - \hat{X}_{ij} \right)^2} [dB] \quad (2)$$

Where X_{ij} is the pixel value in the original image and $\hat{X}_{ij}$ is a pixel value in reconstructed image. X^{P-P} is the Peak-to-Peak pixel value of the original image. Usually X^{P-P} equals 255 for an 8-bit representation of the image.

Video Compression Structure

The video coding structure will be build from the original video data that is represented as sequence of frames, the inter frame compression technique will divide the whole frames into two types, the first type consists of the Reference frame (Anchor frames) and the second consists of the Estimated frames. The Reference frames are compressed independently and separately without any considerations to the correlations may exist with the neighboring frames, while the estimated frames are compressed using motion estimation methods, which will encode the estimated frames according to the correlations with the neighboring reference frames as shown in figure 3 [5].

After getting a frames, compress it as a reference frame (RF) by taking one reference frame after each ten frames of video frame then compute the motion estimation (ME) for the next five sequence frames, such that the determined estimation depends on the previous reference frame. Skip five frames and get a new video frame compressing it as a reference frame, where the skipped five frames are compared with this reference frame to compute their blocks motion. The reference frames is compressed using DCT transform (eq.3 and eq.4) and estimated frames are compressed using the OTS technique.

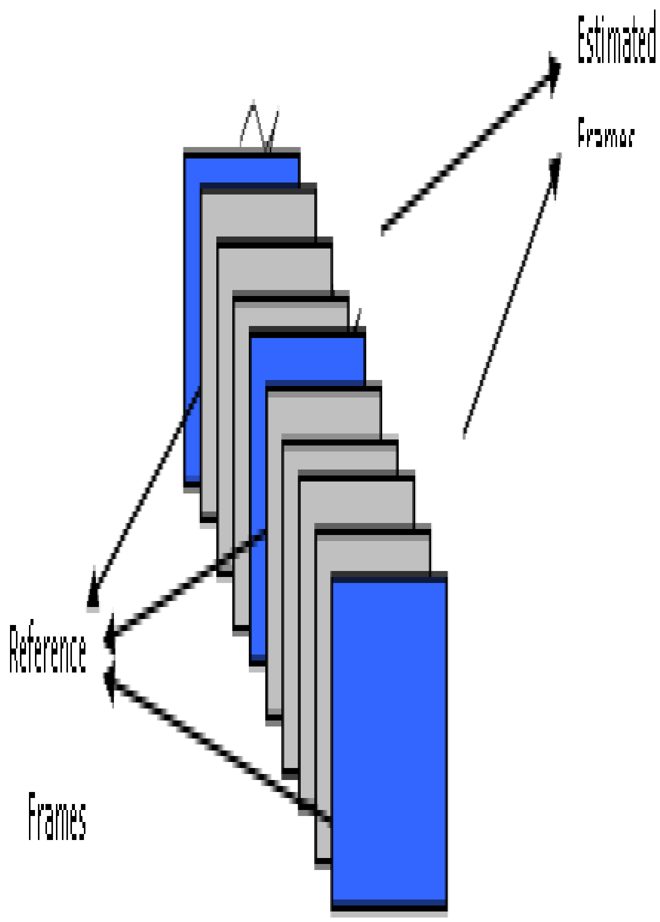

Figure 3 Video Coding Structures.

C(u,v)=

$$\alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (3)$$

For u, v=0, 1, 2,…, N-1 $\alpha(u)$ and $\alpha(v)$ are define in equation (3.6) The inverse transform is defined as

f(x,y)=

$$\sum_{x=0}^{N-1}\sum_{v=0}^{N-1}\alpha(u)\alpha(v)C(u,v)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

(4)

Motion Compensation

It is more efficiently to code the difference between frames, rather than the frames themselves. An estimate for the frame being coded is obtained from the previous frame and the difference between the prediction and the current frame is sent. Most video sequences have moving objects and this motion is one of the reasons for difference between successive frames. If there were no motion of objects between frames, these frames would be very similar. The basic idea behind motion compensation is to estimate this information to build a prediction for successive frames. H.263 supports block based motion estimation and compensation. Motion compensation is done at the MB level, except in the advanced modes, when it is done at the block level. The process involves looking at every MB in the current frame and trying to find a match for it in previous frame. Each MB in the current frame is compared to 16 × 16 areas in a specified search space in the previous frame, and the best matching area is selected. This area is then offset by the displacement between its positions and the position of the current MB. In most cases, it is not possible to find an exact match for the current MB. The prediction area is usually similar to the MB and difference or residue between the two is small. Similarly, predictions for all the MBs in the current frame are obtained and the prediction frame is constructed. This residue frame is then coded using the transform coding procedure. More information about motion compensation can be found in [6]. A major part of the encoding process involves finding these best matches or equivalently the offsets between the position of the best match and the position of the current block. This process of searching for the best matches is called Motion Estimation. These offsets between the position of the best match and the position of the current MB are called motion vectors. H.263 allows these motion vectors to have non-integral values. For instance the motion vector for a MB may have the value (1.5,-4.5), which means that the best match for the current block has pixels that lie at non-integral positions. All pixels in the previous frame are at integer pixel positions. Hence pixels at non-integral positions have values that are computed from the original pixel values using bilinear interpolation. In the last decade, many fast searching algorithms have been developed to reduce the computation and data fetching by reducing the number of comparisons between the blocks, such as the three step search TSS, one at time search OTS and other methods [6].

Once at a Time Search (OTS)

Searching algorithms, first evaluate points within the search window that are on the same row as the center point. If the center point is designated as (m, n), see figure 4, then the three position (m-1, n), (m, n), and (m+1, n) are evaluated. If the best Sum Absolute Difference (SAD) value corresponds to the central point, then the horizontal evaluation ends. If the best point corresponds to one of the end points, then the end point becomes the center point and three more points are evaluated. This continues until the best point is found at the center of the three points. Now the same procedure is conducted in the vertical direction, i.e., at the points (m, n-1), (m, n) and (m, n+1), beginning at the best point found in the horizontal search. Finally the method will continue search for the best SAD but in a diagonal way toward the original center (first center). The result of the diagonal search is the designated best vector location [7].

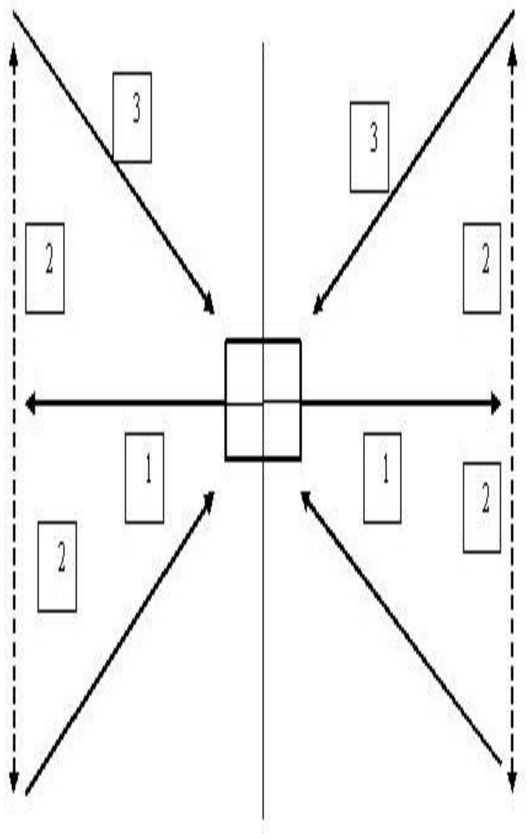

Figure 4 OTS Motion Estimation

Encoder Reference Frame Using DCT

The video frame of RGB components is 2-D matrix. The compression steps scheme could be summarized as follows, see Fig.5:

1. In the first step the system will transform the color components R, G , B to its equivalent components Y , U ,V sequentially by using equations (5) ,(6) and (7).

$$Y=0.299R+0.587G+0.114B$$
$$U=-0.147R+0.289G+0.436B$$
$$V=0.615R-0.515G-0.11B$$

The reason of using the YUV system is that the Human Visual System (HVS) is less sensitive to chrominance or color information then in the luminance information, where Y contains more than 90% of the information of data. The chrominance signals can therefore be represented with a lower resolution than the luminance without significantly affecting the visual quality. This implies that the information in the chrominance component can be reduced without any noticeable degradation in the picture quality. Each of the Y, C^B and C^R components is represented by eight bits. After the sub-sampling, each pixel can be represented by twelve bits, eight bits for the luminance component and two bits for each chrominance component [4].

2. Down sample for each U and V component, the goal of this method is to reduce the image size to quarter (25%) of its original size by replacing each four pixels by one pixel, whose value is the average of the four pixels. This is due to the fact that the components U and V hold only (10%) of the whole color information of image matrix, so this produces distortions in the colors.
3. The DCT transform will take place to the results Ud and Vd in addition to Y band that produce from down sample method.
4. Apply the quantization of the result of DCT transform to reduce the number of bits that needed to represent DCT coefficient.
5. The quantizer scale is changed in range of 1 to 8 to get best quality of video and optimum compression ratio.

6. Apply the Zig-Zag ordering on the quantized of DCT coefficient.
7. Encode the result of Zig-Zag process (ZQY, ZQU, and ZQV) using Run Length encoder.
8. Finally, encode the output of run length encoder using S-shift encoder, then save the result in the output buffer.

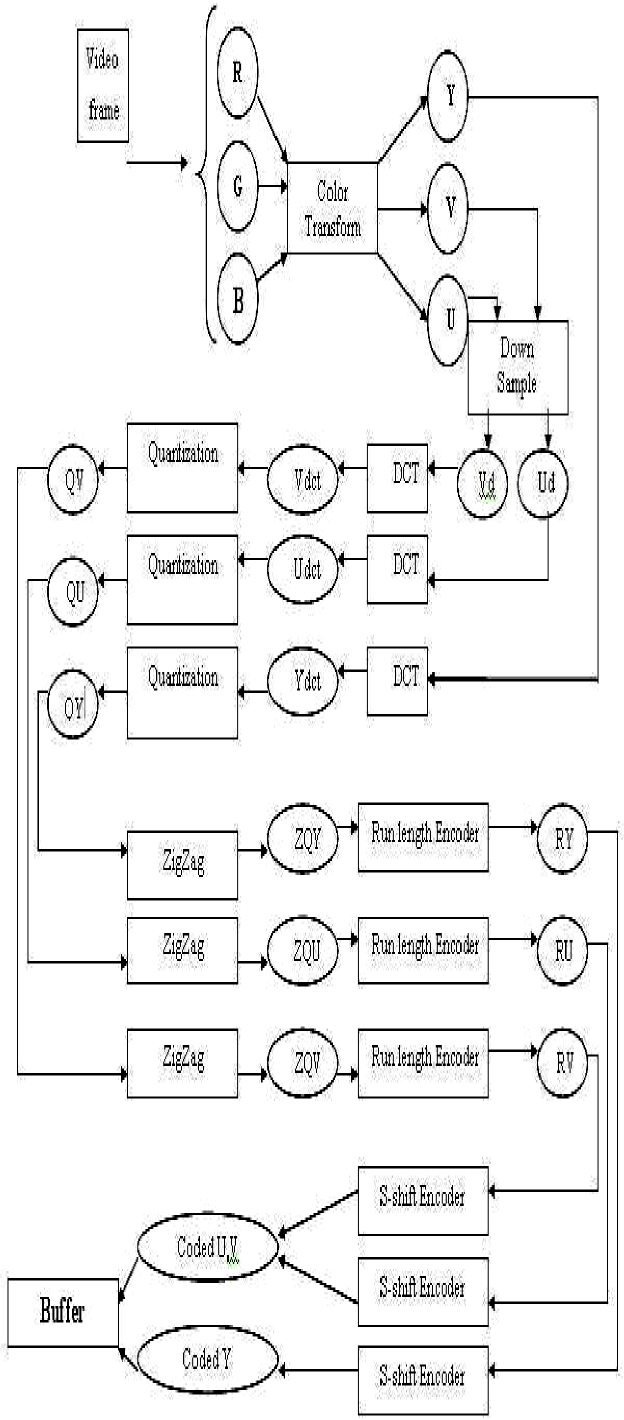

Figure 5 The Compression Scheme for Reference Frame Using DCT

Decoder Reference Frame using IDCT: From the testing results of all samples of the models the results show the following, see Fig.6:

1. The testing result of the tables shows that the PSNR of the References frames and estimated frame increase when Qs=5 and get better quality of video compared with other values (Qs=2,3,4,6,7,8) ,so as the value of Qs increased, the PSNR will be decreased , in other words , the video has lower quality.
2. The testing result of the tables of the References frames using Discrete cosine transform DCT, where the measurement of PSNR of the Reference frames in this table show high quality of the decoded frames in the DCT process. So the fidelity measures are acceptable.
3. The PSNR and MSE fidelity are measured for the estimated frames using one time Search OTS technique in PSNR, the results range between (16 to 38.8).
4. All samples have good compression ratio, where the video samples compress approximately (30 %).
5. From the figures, the references frames are chosen each ten estimated frames have position (1, 11, 22, 33, 44, 55,120) in video sample, and having smaller value of MSE comparing with estimated frames ,alternatively, the PSNR of references frames have larger values ranging between (19 to 40) comparing with estimated frames.

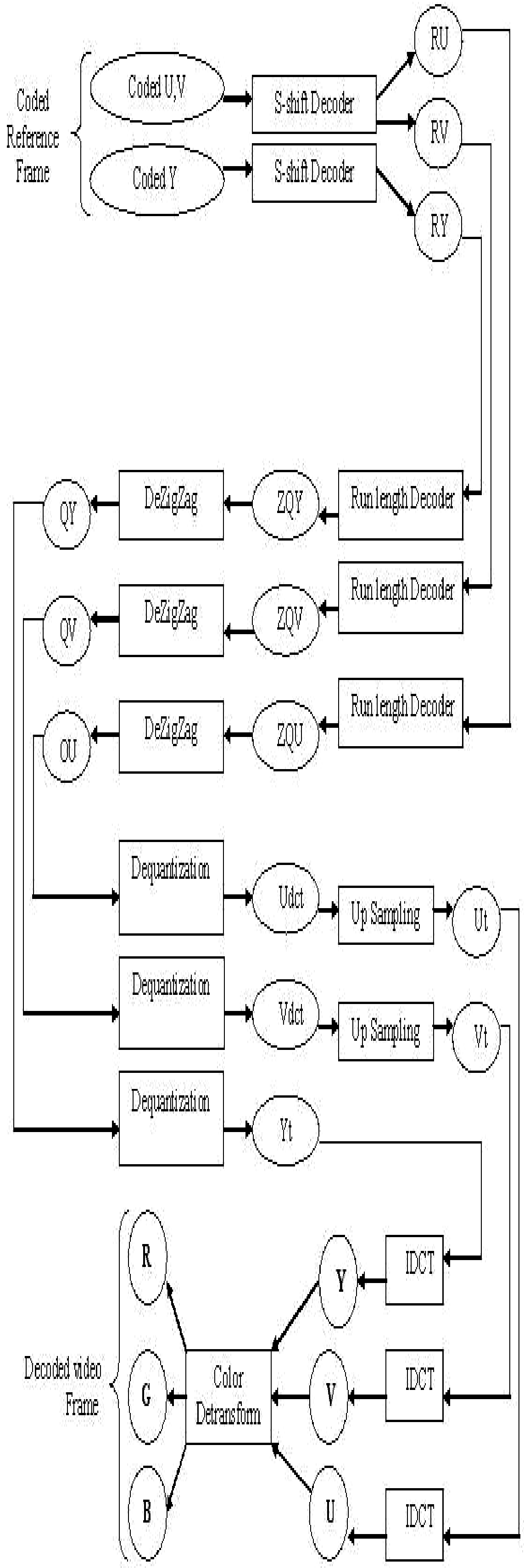

Figure 6 Decoder Reference Frame Using DCT

Test Samples

The testing operation is implemented on different video samples of AVI of size 352×288 (CIF standard picture format) of different number of video frames ranging from 26 to 126. To verify the performance of video the codec algorithm for both subjective and objective is tested and evaluated. the subjective studies rely with human subjects. For computational method (objective), the PSNR measures are used. Fig.7 gives a sample test of one of the video samples used to examine the system. The number of frames for each cycle is 10, which is one of the parameter of compromising between the quality and compression ratio, as a controlling factor of the compressed video. Table.1 shows another test of the system that includes other samples at different frame rate per cycle. The compression ratio for quantization factor of 8 gives a reduction of video rate, as a bit streaming, more than 30 times, which is proper of streaming the video over ISDN. While the overall measurements, as an objective and subjective measurements, are reasonable.

To test the proposed system ov`er communication network, a LAN is used with a rate that is proper for ISDN. To visualize the facility of system establishments and measurements different windows were created, where the IP address for source and destination are named, the quantization factor can be chosen, and all required measurements can be edited for the system work evaluations. Fig.8 shows one of these windows, that gives the objective and subjective system evaluation.

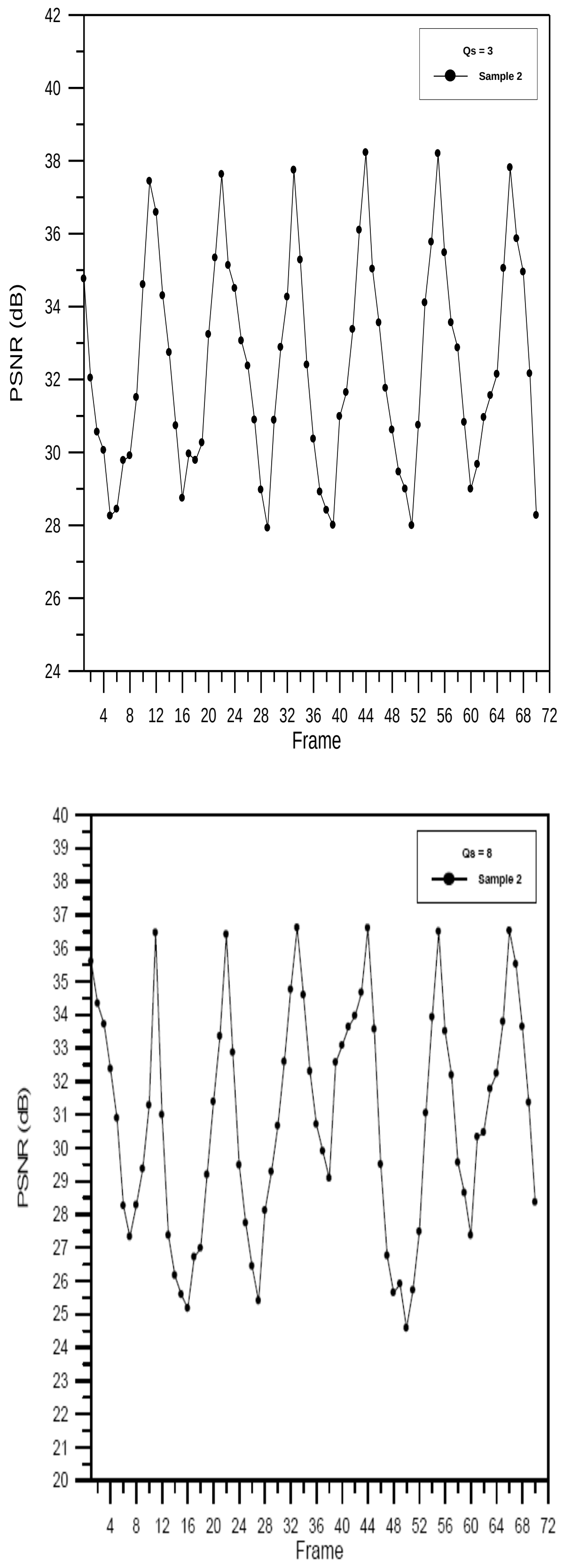

Figure 7 The PSNR of video sample test with quantization step of Qs=3 (left) and Qs=8 (right)

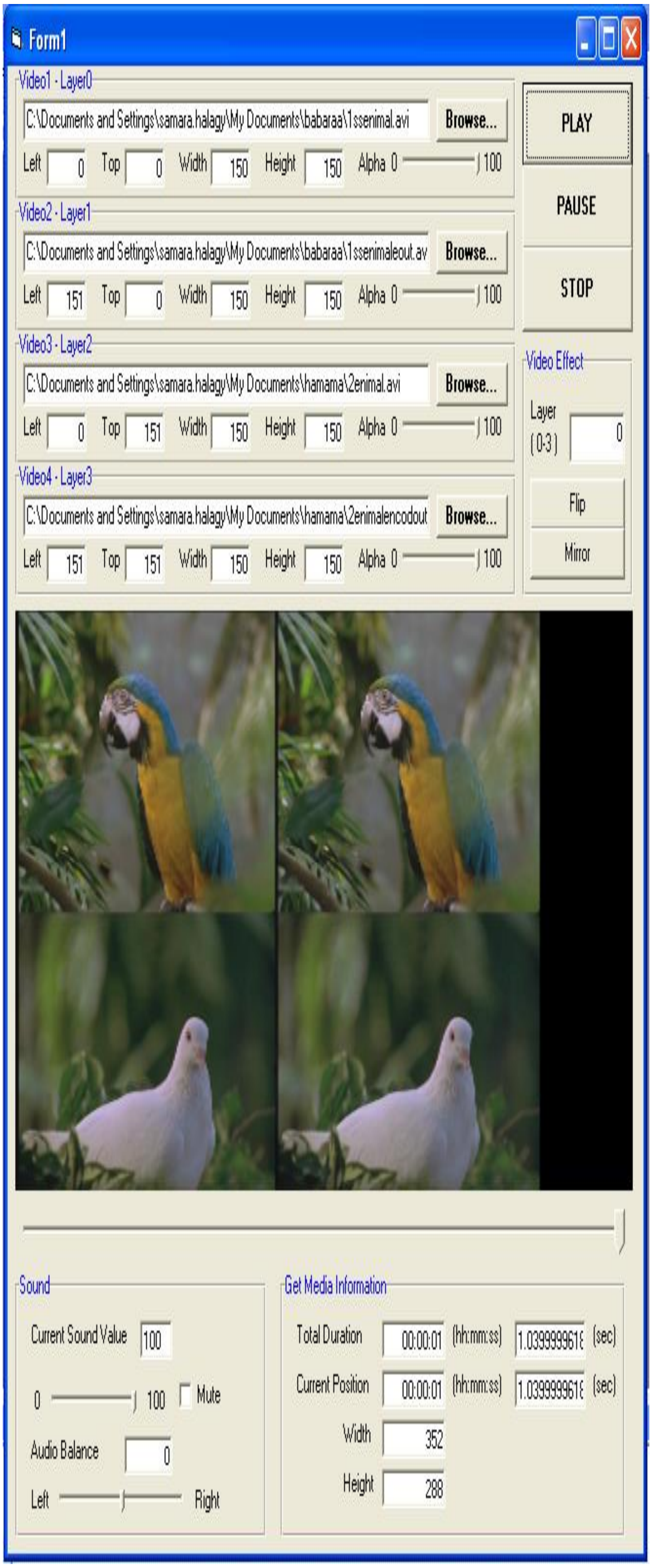

Figure 8 Window for objective and subjective measurements

Table.1 Measurements for sample video at quantization step of Qs=8.

Estimated Frames (EF)					
Samples	Number of frame	Total MSE	Average MSE	Total PSNR(dB)	Average PSNR(dB)
S1	114	5993.994	52.5789	3608.1969	31.65085
S2	114	10233.76	89.76981	3398.949	29.81534
S3	114	14850.244	130.2653	3308.2594	29.01982
S4	114	2887.846	25.33199	3913.338	34.32753
S5	114	5865.366	51.45058	3609.071	31.65852

Conclusions

The incorporation of visualization measurements with the motion detection facilities provides a good tool for increasing the compression ratio and reducing the time needed for encoding the frame sequence by identifying and processing only those blocks that exhibit noticeable changes between successive frames. This tool of system measurements is the good way of building the video bit streaming control system that has the ability of incorporating the network losses in additional to select the proper factors. The proposed system, which is compatible to H.263, gives a compression ration more than 30 times with a high quality of system reception video reconstruction.

References

[1] Array Microsystem, An Introduction to video Compression, Prentince Hall PTR, 1997.

[2] ITU-T Recommendation H.263, "Video coding for low bit rate communication"**, 2000.**

[3] Vancouver, "H.263 Standard Overview and Implementation", 2002 http://dspvillage.ti.com/pdfs/spra018. Fredrik Ydrenius, "Scalable Coding of video over IP-based Networks", M.Sc. Thesis , University of Linkoping, 2000. www.bk.isy.liu.se/publications/en/LiTH-ISY-EX-2066.

[4] Anders Laurin , "Networked Video Surveillance and Compression Technology", June 2002. www.axis.com/documentation/whitepaper/video/video_compression_article

[5] Deepak Turaga and Tsuhan Chen "ITU-T Video Coding Standards", 1998.

[6] Karin R. Gastreich, "OTS/Duk Semester Abroad Program", 2004.

ترميز الفيديو المرسل على شبكات الانترنيت

د .نصر نافع خميس وسمارة إيليا

E- mail: nasser_alani@yahoo.com

الخلاصة:

لتحسين نوعية نموذج العرض الفيديوي، تم تقديم دراسة متطورة لاليات ضغط الصورة الفيديوية بأستخدام نظرية البحث لمرة واحدة (OTS) مع تقييم مستمر لنسبة الأشارة الى الضوضاء SNR كأداة لتحقيق أفضل موائمة بين آليات الضغط وسعة القناة الناقلة.حيث اعتمد التقيس الآلي لتقيم نموذج العرض ، من خلال حساب MSE و PSNR و CR. في الوقت نفسه تم معالجة مشكلة المعدل المتغير للأشارة الفيديوية VBR و التعامل معها كأشارة ثابتة CBR. وكنتيجة كلية لأداء المنظومة تم اعتماد نظام السيطرة لتقييم الخطأ الكلي المسموح به متضمنا الخطأ الناتج عن قناة الأتصال بأستخدام بروتوكولات النقل TCP/ IP. ادخلت لغة البرمجة VBL تحت نظام العمل XP كأداة مساعدة لبناء الواجهات لعرض نتائج التقييس المطلوبة.