Video Clip Image Compression Using DCT Technique

Dr. Nabeel Hashem Kaghed Dr. Tawfiq A. Abbas Wafaa Hassan AL-Marsomi

Department of Computer Science , University of Babylon, Iraq

Abstract

This paper presents a system for video compression which can be summarized into four stages the first stage is opening Audio/ Video Interleaved file (AVI) Video file. second stage is to compress sequence of video frames by using hybrid techniques presented in lossy compression in video codec through used of transform technique which is presented by discrete cosine transform (DCT) to reduce spatial redundancy in image data itself and to reduce temporal redundancy by using motion compensation technique which is employed associated with DCT to get error image produced by motion technique based on skew measurement and at last of this stage we used efficient method to code them, the third stage we inverse all hybrid compression techniques to obtain the decompressed sequence image of video file. At the fourth stage which is used to reconstruct video file AVI type.

The proposed system achieves compression ratio close to the best known video compression techniques MPEG between (1:20- 1:40). This paper consist of five sections, in section 1 explain the introduction to video compression , in section 2 introduce to structure of AVI movie file , in section 3 explain the video codec (the proposed system), in section 4 display the experimental results with PSNR measurement, in section 5 display the conclusion .

1. Introduction

Digital video has become very important form of information technology and is now used in many different areas, such as board casting, teleconferencing, mobile telephone, surveillance, and entertainment. People now expect to be able to access video through a wide range of different devices and over various networks. To provide these kinds of services we must know what is video compression for storage and transmission(K.Holtz).Compression of video imagery has become a necessity and very important because of transmission and storage of uncompressed video would be extremely costly and impractical. For instance, a video sequence running for (90) minutes, at (25) frames per second, at standard resolution of (720 *576), and with (24) bit per pixel would require (1343692800000) bit or approximately 156.43 gigabytes". This is not a problem if we only wish to access and deal with video through high - end systems with a lot of storage and network band width. However, since it is desired that video will be accessible from a wide range of devices having different capabilities and connected to different networks, therefore some form compression for digital video is required (B. Tower ,2001). Compression refers to the process of reducing the number of bits required to represent the image and video comes in two forms lossless and lossy (N. Schaab ,2004). The lossless compression is a process to reduce image or video data for storage and transmission while retaining the quality of original image (i.e. the decoded image quality is required to be identical to image quality prior encoding. In lossy compression, on the other hand, some information present in the original image or video is discarded so that the original raw representation of image or video can only be approximately reconstructed from the

compressed representation with high compression ratio. For more compression can be achieved with approximate quality to source image lossy compression is almost usually used than lossless compression to compress digital video (L. Kizewsk ,2004). In generall, in video compression the video sequences contain significant amount of statistical and subjective redundancy within and between frames. The ultimate goal of video source coding is the bit-rate reduction for storage and transmission by exploring both spatial and temporal redundancy and to encode “minimum set” of information using entropy coding techniques .This is usually results in compression of coded video data compared to original source data (T.sikora ,1998). These statistical and subjective redundancy that can be exploited for compression are called spatial and temporal redundancy. The performance of the video compression techniques depends on the amount of spatial and temporal redundancy contained in video data as well as on actual compression techniques used for encoding – with practical code schemes a trade – off between coding performance (high compression with sufficient quality) and implementation complexity is targeted – thus one of the most method which is used for exploiting spatial redundancy is the transform coding and for exploiting temporal redundancy used motion estimation methods (T.sikora ,1998), all of these methods used in video codec which explain in figure (1) (P. Lindmark , 2005).

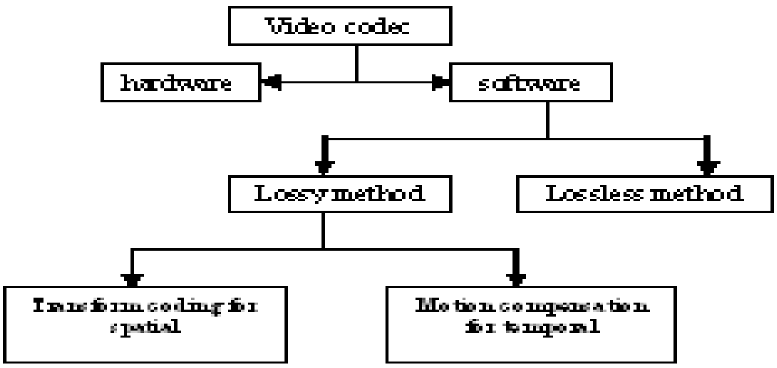


Figure (1): Types of video codec

Video Codec is a codec i.e. computer program that compress/ decompress digital video data according to given file format and can be either software application or part of hardware as explained in figure (1) (Bamboo Web Dictionary). They process the video through complex algorithm that compress and decompress video files. Video codec implemented in hardware is most efficient way to code and decode video file, they are faster and demands much less processor power than a software codec. hardware codec are often use at video conferencing where the equipment at the sender and receiver are configured at the same way. Software codecs are cheaper and there are a lot of codecs on the internet to use for free such as (MPEG-2, MPEG-4, H.261 and H.283). After we are introduced to redundancy which is contained in image sequence of video file we are used DCT to reduce spatial redundancy, and differences between successive frames to reduce temporal redundancy by using skew measurement to measure asymmetry between them by calculating standard deviation and mean which tell us about contrast and brightness respectively to compress video file and stored it in another form . To construct video file again we are needed to store all necessary information about AVI header and keep the same structure of AVI file .

. Amir Said (A. Said -2004) proposed new coding system, in this system the encoder perform coding depend on coding image block classification method that is very simple and low computational–complexity. Ayman Darwish (A. Darwish ,2005) submitted thesis in scalable video coding, she used DWT to reduce Spatial Redundancy at 1, 2, 3, 4-level and use quantization factor at 8,16,32, when increasing result in more quantization & worse quality reconstructed image.

2. Audio/ Video Interleaved File

The Audio Video Interleave file format was originally developed by Microsoft for Intel. The AVI format is now marketed as "Video for Windows". Due to the wide user base of Microsoft Windows, the AVI format has become very popular, extension is AVI which is used the AVI RIFF form. The AVI RIFF form is identified by the four-character code ``AVI '' – which A FOURCC (four-character code) is a 32-bit unsigned integer created by concatenating four ASCII characters. . The AVI file format uses FOURCC codes to identify stream types, data chunks, index entries, and other information-. All AVI files include two mandatory LIST chunks. These chunks define the format of the streams and stream data. AVI files might also include an index chunk. This optional chunk specifies the location of data chunks within the file

3. Design of the Proposed System

The proposed system consist of two parts : the first part is the Coder which contain the stages shown in figure (2) which represent the compression of movie file, and the second part is Decoder which contain the stages shown in figure (3), in the next sections we explain each part in details .

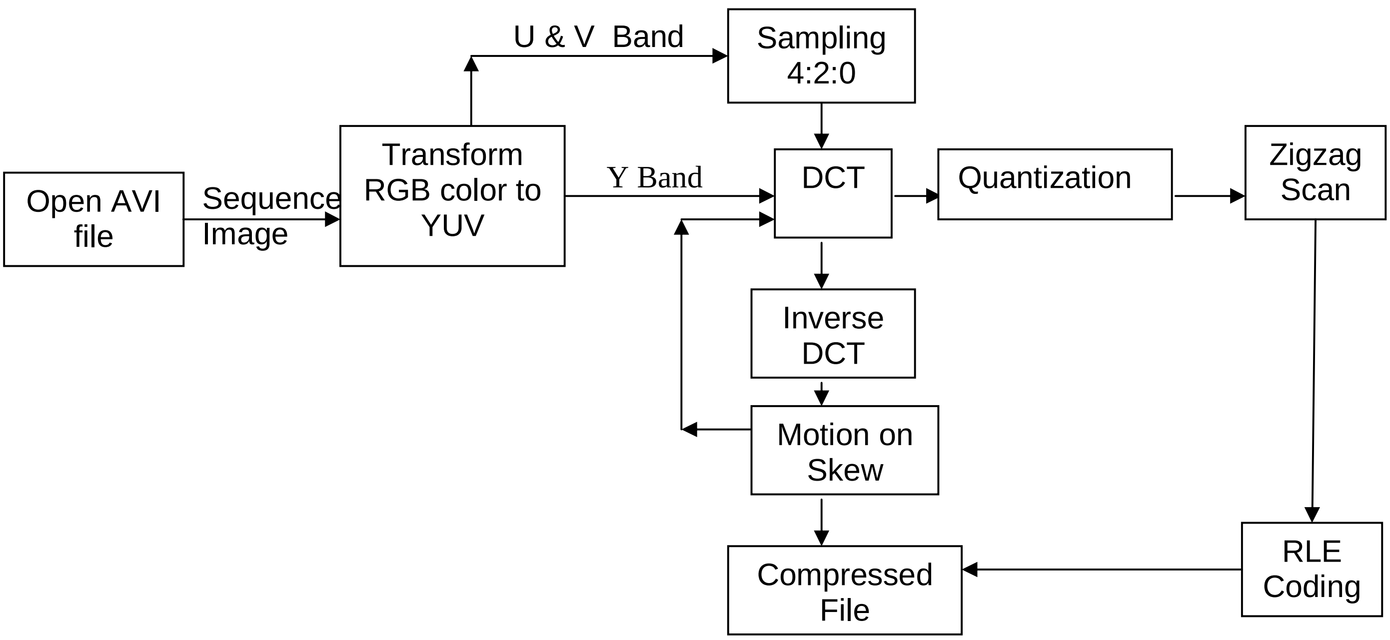


Figure (2): Block Diagram of Coder Part

3.1. Coder Part

This part consists of several algorithms which are needed to compress video data into another form stored in compressed file. These algorithms will be explained as follows:

3.1.1. Open AVI File & Convert to still Images

In this stage we proposed algorithm to open AVI structure and get list movie (i.e. sequence frames / images), then after that take information from AVI header such as total of frame , height , and width of the image, and split it into three bands that save it into three individual matrices called RED,GRN, and BLU which are used in the next stage. Are needed in the next stages to compress it.

3.1.2. Transform Color RGB to YUV

The 'Transform Color YUV' perform algorithm to convert RGB pixels into Luminance / Chrominance components by using YUV European which convert Red into Luminance called Y and the value lies in [0..255] while convert Green and Blue to color information called Chrominance U and V respectively which is a bipolar value (i.e. positive & negative), the range for U is [-112 .. 112] and V is [-157 .. 157], this conversion is efficient in compression. By using the following equation:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0.000 & 1.140 \\ 1 & -0.396 & -0.581 \\ 1 & 2.029 & 0.000 \end{pmatrix} * \begin{pmatrix} Y \\ U \\ V \end{pmatrix}$$

3.1.3. Sampling 4:2:0

The sub- sampling take only U and V components (i.e. chrominance) which carries color information and apply the algorithm which reduce each color component to half by dividing by two in both direction height and width to reduce the size that take in compressed file as shown in figure(3).

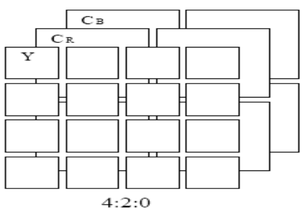


Figure (3) : Sampling type of 4:2:0

3.1.4. Discrete Cosine Transform DCT (E .Scott, 1998)

Apply the DCT transform to the three Y, U, and V images matrices to transform image data from spatial domain to frequency domain by dividing each frame into block of 8*8 pixels and apply DCT transform to each block 8*8 pixel value to produce 64 coefficients. By using the 2D-DCT which is the amongst block transform that partitions an input image into non – overlapped blocks, then maps them into blocks of coefficients so as it can be used in JEPG & MPEG compression successfully at decorrelating and concentrating the energy of pixel data into spatial domain. Given an image of any array that partition into 8*8 blocks is then the DCT applied separately on 8*8 blocks of data so that the equations can be written as:

$$C(u,v) = \frac{\alpha(v)}{2} \cdot \frac{\alpha(u)}{2} \sum_{y=0}^{7} \sum_{x=0}^{7} f(y,x) \cdot Cos\left[\frac{\Pi(2x+1)u}{16}\right] \cdot Cos\left[\frac{\Pi(2y+1)v}{16}\right]$$

For *u, v* = 0 ,1 ,2 ,… *7*

$$f(x,y) = \sum_{y=0}^{7} \frac{\alpha(v)}{2} \sum_{x=0}^{7} \frac{\alpha(u)}{2} \cdot C(v,u) \cdot Cos\left[\frac{\Pi(2x+1)u}{16}\right] Cos\left[\frac{\Pi(2y+1)v}{16}\right]$$

for *x* , *y* = 0 ,1 ,2 ,…*7*

$$\alpha = \left[\begin{array}{ll} 1/\sqrt{2} & \text{if } u = 0 \\ 1 & \text{if } u <> 0 \end{array}\right]$$

The DCT make it possible to compress the image by concentrating most of the image information in the lower spatial domain, so that the 64 coefficients produced by that equation is represent as in figure (4). The coefficient of the top -left- corner is called DC which represents the average of whole image, and it is low frequency, the other 63coefficients is high frequency which is called AC.

36	32	26	19	14	11	10	10		75	39	21	0	0	0	0	0
32	29	22	16	11	9	8	8	DCT	42	23	0	0	0	0	0	0
26	22	17	11	7	5	4	5	⇒	19	0	0	0	0	0	0	0
19	16	11	6	2	1	1	2		0	0	0	0	0	0	0	0
14	11	7	3	0	0	0	1	⇐	0	0	0	0	0	0	0	0
12	9	5	2	0	0	2	3	IDCT	0	0	0	0	0	0	0	0
11	9	5	2	1	2	4	6		0	0	0	0	0	0	0	0
11	9	6	3	3	4	6	8		0	0	0	0	0	0	0	0

Figure(4): The Effect of a Discrete Cosine Transform

Given this interpretation of the DCT, the way to lose the unimportant image information is to reduce the size of 8*8 numbers of coefficient especially the ones of the right bottom. There is a chance that this won't degrade the image quality much. In general these coefficients must be quantified and then compress using one of compression lossless method such as run length coding, Huffman, Arithmetic coding , etc.

After each 8*8 matrix of DCT coefficients *c (u ,v)* is calculated, it is quantized, this step where information is loss (except for some unavoidable loss because of finite precision in other steps) occurs (D. Salomon , 1998). The equalizer divides the DCT coefficient by its corresponding quantum. The quantization matrix is 8*8 matrixes of step sizes called quantum and then rounds to the nearest integer. To increase the number of zero coefficient by increasing R factor which is used in (1+(row + col)*R) to divide DCT coefficients and to avoid negative coefficients by using normalized operation by computing the maximum coefficient and minimum coefficient can be obtained by the DCT transform in block and then difference the minimum value from the maximum to produce positive value of coefficients.

3.1.5. Motion based On Skew

As we illustrated previously video compression is involved with removal of the spatial and temporal redundancy, therefore intra – frame which are used to remove spatial redundancy and inter –frame is used for to remove temporal redundancy, by using this fact, the current frame is often not entirely different from the previous one. Many part of blocks of a frame are identical to corresponding block the previous frame .Coding efficiency is achieved with prediction from the previous reference frame, therefore we take only Y luminance component of sequence frame and calculate the skew for each image (frame), then make absolute difference for successive frames and compare the result with some threshold and the value of threshold will be indicated by the user in the proposed system, the threshold value is 0.01 which is chosen to split sequence image into scenes based on skew and comparison, then in each scene copy the first frame as key frame and then make the difference between key frame and successive frame to reduce pixel value because the successive frame contain many identical information, after that stored the number of key frame of each scene in compressed file, skew is to measure asymmetry about the mean in the gray level distribution . it is defined as

$$SKEW = 1/\sigma g^3 * \sum_{g=0}^{L-1} (g - \hat{g})^3 * p(g)$$

where *σg* is the standard deviation , *ĝ* is the mean , *p(g)* is the probability of histogram Another method to measure the skew uses mean, mode, and standard deviation where the mode is defined as the peak, or highest value :

$$SKEW = \hat{g} - \mathrm{mod}\, e / \sigma g$$

This method of measuring the skew is more computationally efficient, especially considering that, typically, the mean and standard deviation have already been calculated (R. Gonzallez 2002).

3.1.7. Zigzag Scan and RLE

After quantization, it is not unusual for more than half of the DCT coefficients to equal zero, thus we can use zigzag type scan pattern shown in figure (5) which re-order the coefficient so that gathering the zero coefficient one often another ,which is useful in lossless compression methods .

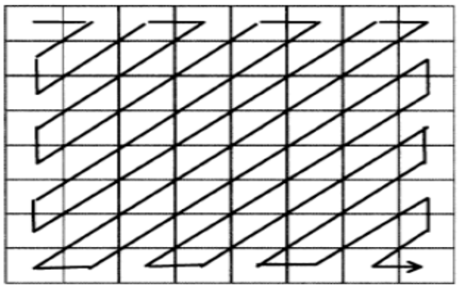
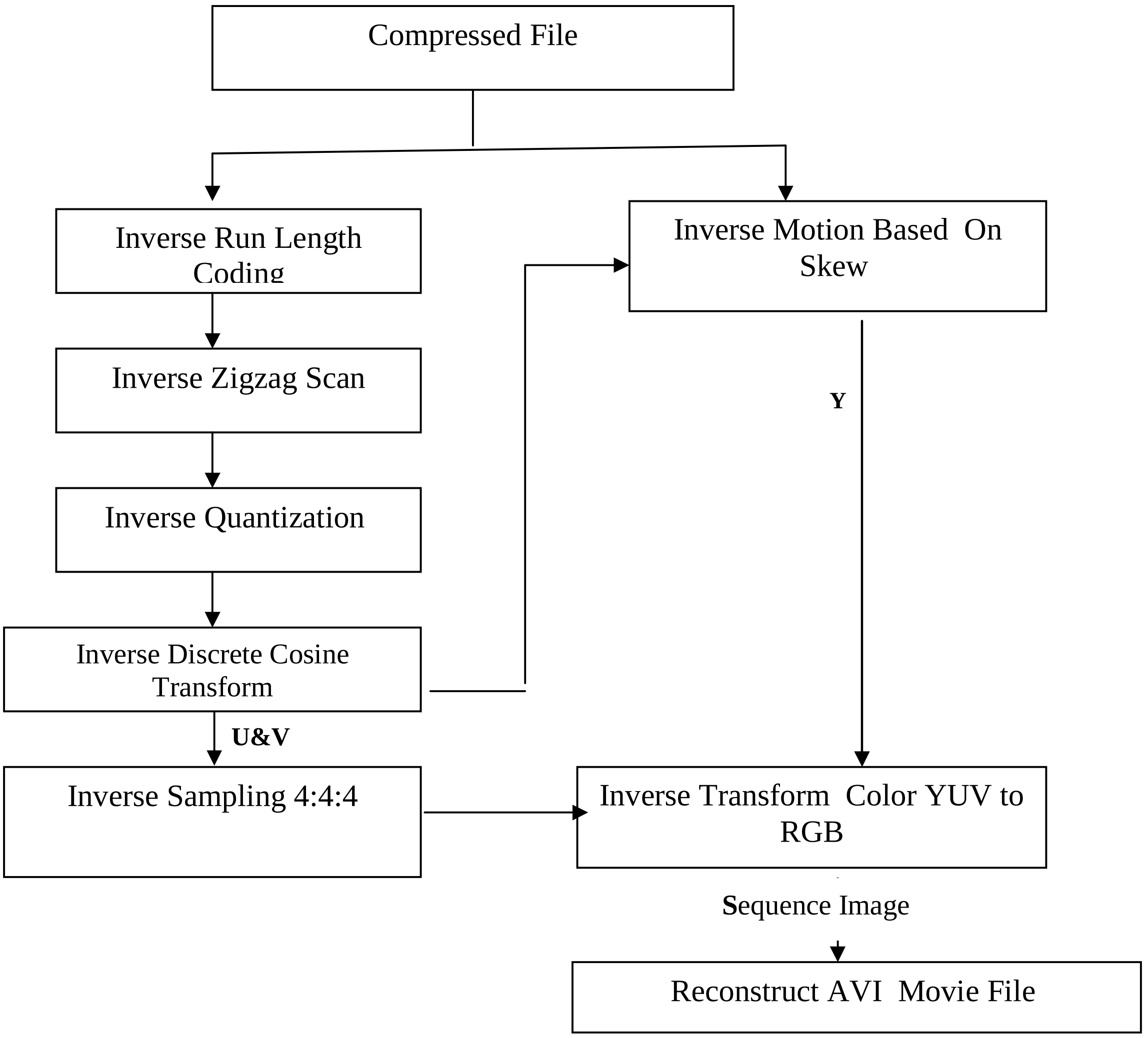


Figure (5) : ZIGZAG –SCAN TYPE

The process of zigzag scan apply algorithm to convert quantized coefficients block 8*8 into one dimensional vector of 64 coefficient scan in zigzag type which is useful to accumulate the coefficients that closely same together. and RLE then used to compressed the coefficients.

3.2 Decoder Part

This Part consist of several algorithms which are needed to decompress video data from compressed file (i.e. inverse compression operation) as in figure (6) which are represent the inverse of each stage in Coder Part .



Figure(6) Decoder Part

3.2.1 Reconstruct AVI Movie File

Write AVI block is the last stage of the Decoder , in this stage we form the structure of AVI movie file to play back again by using sequence images that obtained from previous stages of the Decoder which extract them from compressed file by using two process load AVI header which are used to store the all information needed to write AVI file.

4. THE Experimental Result

A wide variety of techniques were used for video compression, they include lossy and lossless as explained in detail in the previous section. In this section we test the proposed compression system on video clip files which illustrated in Table (1). The results of the test were summarized below.

Table (1) the compression ratio, and average PSNR of RGB of whole video clip

No.	Movie sample	Source file in Byte	Compressed file	Compression ratio	PSNR
1		1333160	33101	40.28:1	37.84
2		7895304	241763	32.66:1	29.55
3		10794856	343517	31.43:1	27.31
4		6243256	204606	30.52:1	35.39
5		5377272	182979	29.39:1	29.87
6		19760276	754744	26.19:1	30.65
7		19188296	784617	24.46:1	29.28

In the proposed system numbers of video file movies was taken which are difference in natural complexity, scenes quality, and size. These movie implemented with proposed system give the results which explain in table (1), and we noticed the compression ratio increased if the files have few complexity with stable scene (i.e. the file contain one scene) such as the movie no.1 which contain 16 frames has high compression ratio that equal to (40:1)while the movie no. 7 has which contain 259 frame distributed over several scenes has low compression ratio (24.46:1) , this mean that when scenes were varying and increased in video file the compression ratio decreased .

In addition the PSNR is in negative with compression ratio for the same movie file, also depend on the complexity of the movie, such as movie no.1 because have simple image without complexity the PSNR is high ,and can also increased for that movie if the compression ratio is decreased while movie no.7 because has natural image with some complexity the PSNR is decreased in spite of the compression ratio is low.

5. Conclusions

In this section we will present some conclusions:

1. **The proposed system can split movie file in the scene depending on the skew measurement with a specific threshold for that purpose.**
2. **Natural of the motion make the sequence frames similar with some differences in some regions, thus the differences between frames due to smaller pixel value. Therefore the coefficients produced through DCT tend to zero which make the compression operation by using Run Length Coding very high.**
3. **The transform of color from RGB to YCC luminance /chrominance components reduced the inter pixels redundancy (i.e. reduce pixel value), subsequently increased similarities between them.**
4. **The quantization process not used only to exclude the real value of coefficients but also to exclude negative value from it, which makes Run Length Coding effective.**
5. **Scan of data is played a vital role to convert 2-D matrix to 1-D matrix, therefore a suitable scan type as zigzag which gives best result should be used.**
6. **The compression ratio based on the number of frames in each scene because of the relation between them is positive while the relation is in negative between the compression ratio and number of scene (i.e. when increase number of frame, increased compression Ratio, but when increase the number of scene the compression ratio is decreased) .**

References

K.Holtz, "Image and Video Compression Techniques" Site: http // www. AutoSophy.com / video com.html

T.sikora,"MPEG-1 and MPEG-2 digital video coding standard ",digital consumer electronic hand book, Mc Grow -Hill Book Company,Ed.R.Jurgens,1998.

B.Tower,"H.263 video compression : the standardized advantage " March Network corporation®, September 2001.

P. Lindmark ,"Optimizing Video Compression and Streaming Media for E-Learning" MS_SC thesis IT university of GÖteborg, Sweden 2005.

Bamboo Web Dictionary, "Video Codec", Site: http // www.Bamboo Web.com/articls/v/i/videocodec.html

W. Niblack, ”An Introduction of Digital Image processing” ,prentice hall international ,1986.

E .Scott ,”Computer Vision and Image Processing : A practical approach using CVIP tools “ Prentice –Hill , NewJersy, 1998

B. Andrew “Image Compression Using Discrete Cosine Transform “, Mathematics Journal vol.4, no.1, pp 81-88, 1994.

R. Gonzalez & R. Woods ,”Digital Image Processing (2nd addition) “ NewJersy , prentice –Hill ,2002.

D. Salomon ,”Data Compression” ,dept of computer science , California University, Verlag NewYourk,1998.

A. Darwish ,”Scalable Video Coding “ , Clyton School of Information Technology, Monash University, November, 2005.

S. Einerman ,”Residual Coding “ ,Department of signal ,Sensors and Systems, Royal Institute of Technology (KTH), December 2001 .

A. Said , “Compression of Compound Images and Video for Enabling Rich Media in Embedded Systems”, Imaging System laboratories HP laboratories Palo Alto, © Hewlett – Packard company, May 2004.

L. Kizewsk ,” Image De-blocking Using Local Segmentation” School of Computer Science and Software Engineering, Monash University, November, 2004.

O. K. AL-Shayk, E. Milo, T. Nomura, R. Neff and A. Zakhor ,” Video Compression Using Matching Pursuits”, Department of Electrical Engineering & Computer Science, California University, IEEE Transaction on circuits & system for video technology , February 1999.

N. Merhav & V. Bhaskaran ,” A Fast Algorithm for DCT – domain Inverse Motion compensation “, Comuter Systems laboratory & Hewlett – Packard laboratories (HP) in Science Center, Technion city , Haifa -3- 2000.

L. shen ,”Image Analysis and Processing “ , Computer and Information Science U Mass Dartmouth, 2005.

N. Schaab, “ Compressed Video In the Mobile Environment “, Virginia Polytechnic Institute and University © by Chris N.Schaab, 2004.

B. Digan , & Koyanagi ,”Streaming Video “ 1997.

C. Nilsson ,& B. Marianne “Introduction Till Digital Video “ 1998 site ://http//www.mosda. .VXU.Se /multimedia/ introdigi.html, 9-3-2005