Adaptive Video Compression Technique Based on Wavelet Transform and NTSS Matching Algorithm

Hameed Abdul-Kareem Younis
Hameedalkinani2004@yahoo.com
Marwa Kamel Hussein
Lava_85K@yahoo.com
Dept. of Computer Science, College of Science, University of Basrah, Basrah, IRAQ

Abstract

Currently, multimedia technology is widely used. Using the video encoding compression technology can save storage space, and also can improve the transmission efficiency of network communications. In video compression methods, the first frame of video is independently compressed as a still image, this is called **intra coded frame**. The remaining successive frames are compressed by estimating the disparity between two adjacent frames, which is called **inter coded frames**. In this paper, intra frame was transformed using Discrete Wavelet Transform (DWT). The disparity between each two frames was estimated by New Three Step Search (NTSS) algorithm. The result of the Motion Vector (MV) was encoded into a bit stream by Huffman encoding while the remaining part is compressed like the compression was used in intra frame. Experimental results showed good results in terms of Peak Signal-to-Noise Ratio (PSNR), Compression Ratio (CR), and processing time.

Keywords: Block Matching Algorithms, Video Compression, Motion Estimation, Motion Compensation, New Three Step Search.

تقنية ضغط الفيديو المعدلة المعتمدة على التحويل المويجي وخوارزمية تطابق الكتلة NTSS

حميد عبد الكريم يونس
Hameedalkinani2004@yahoo.com
مروه كامل حسين
Lava_85K@yahoo.com
قسم علوم الحاسبات، كلية العلوم، جامعة البصرة، البصرة، العراق.

الخلاصة

تستخدم تكنولوجيا الوسائط المتعددة في الوقت الحاضر بشكل واسع، لذلك تساعد عملية ضغط الفيديو كثيرا في التقليل من المساحة الخز نية المطلوبة، وكذلك تساعد في عملية التراسل عبر شبكات الاتصالات.

في هذا البحث، يضغط زوج من إطارات الفيديو بضغط الإطار الأول كما في عملية ضغط الصورة الواحدة ، ثم يضغط الإطار الثاني بتخمين عدم التكافؤ (الاختلاف) بين الإطارين.

لغرض تقدير الحركة (الاختلاف) تم استخدام خوارزمية تطابق الكتلة باستخدام خوارزمية بحث ثلاثي الخطوة الجديدة New Three Step Search (NTSS) تم استخدام ترميز هوفمان لترميز متجه الحركة الناتج أما الجزء المتبقي فتتم عملية ضغطه كما في طريقة الصورة الواحدة. النتائج التجريبية بينت نتائج جيدة عند حساب نسبة قمة الإشارة إلى الضوضاء (PSNR) ونسبة الضغط (CR) ووقت المعالجة.

الكلمات المفتاحيه : خوارزمية تطابق الكتلة، ضغط الفيديو، التحويل المويجي المنفصل، تقدير الحركة، تعويض الحركة، خوارزمية بحث ثلاثي الخطوة الجديدة.

1. Introduction

An images sequence (or video) can be acquired by video or motion picture cameras, or generated by sequentially ordering two-dimension (2D) still images as in computer graphics and animation. As shown in Fig.1.

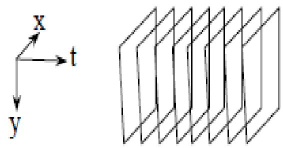

Fig.1: Images sequence (video).

Video processing is special cases of digital processing in which signals are processed are video files or video streams. It is extensively used in television sets, Digital Versatile Disks (DVD), video players, …, etc. Although digital video signals can be transferred over the long distances with a low probability of bit error rate, the raw material of digital video requires high bandwidth for transmission and high storage capacities when compared to its analog equivalent. Therefore, compression basically is necessary to reduce data, a digitized analog video sequence can comprise of up to 165 Mbps of data. To reduce the media overheads for distributing these sequences, the following techniques are commonly employed to achieve desirable reductions in image data [1]:

i . Reduce color nuances within the image.
ii. Reduce the resolution with respect to the prevailing light intensity.
iii . Remove spatial redundancy or correlation between neighboring pixels values.
iv .Compare adjacent images and removes details that are unchanged between adjacent frames in sequence of images.

The first three of above points are image based compression techniques that is called **intra frame**, where only one frame is evaluated and compressed at a time. The last one is called **inter frame,** where different adjacent frames are compared as a way to further reduced image data. All of these techniques are based on the term of motion. Motion is an essential aspect of video sequences. The ability to estimate, analyzes, and compensate for relative motion is a common requirement of many video processing, analysis and compression algorithms and techniques. Fig.2 shows flowchart of video compression.

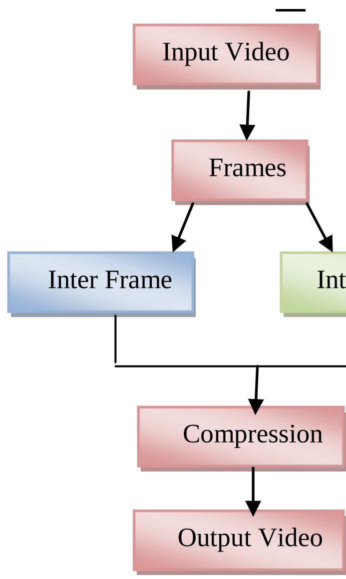

Fig.2: Flowchart of video compression

Bjorn B. in 2005 [2] Developed a video codec. This codec consist of three parts, transformation, quantization and encoding. These three fundamental parts are used for the purpose of compressing the data. Through the transform, the energy in a picture is concentrated to a small region. These regions are then rounded off through quantization to compress the data. Author has also proposed a method to cope the problem of error that can be introduced through transmitted it over channels by sent side information over another channel to describes the information in yet another way. Consequently, a coding scheme called Multiple has been presented.

Hassan B. and Malik K. in 2006 [3] have proposed a method for video compression based on skipping some frames which have little information from the fames sequence. They have designed an algorithm to compare two frames on sub frame level and decides which one is more important for the overall video quality as perceived by eye of the user and skipping the frame which have little information.

The work aims to propose an efficient technique for video compression by wavelet transform by using two systems, intra and inter coded frames. In this present work, video compression system is developed using wavelet transform, and NTSS. The rest of this paper is organized as follows. Some of basic principles has been explained in Section 2. We show our proposed video compression system in Section 3. Section 4 give the experimental results. Finally, the paper has been concluded in Section 5.

2. Basic Principles

2.1 Intra and Inter Coded Frames

The motion estimation and the motion compensation blocks work, only if there is a past frame that is stored. So, question is how do we encode the first frame in a video sequence, for which there is no past frame reference? The answer to this question is fairly straight forward. We treat the first frame of a video sequence like a still image, where only the spatial, i.e., the intra frame redundancy can be exploited [4].

The frames that use only intra frame redundancy for coding are referred to as the **intra coded frame**. The first frame of every video sequence is always an intra-coded frame. From the second frame onwards, both temporal as well as spatial redundancy can be exploited. Since these frames use inter frame redundancy for data compression, these are referred to as **inter coded frame.** However, it is wring to think that only the first frame of a video sequence would be intra-coded and the rest inter-coded. In some of the multimedia standards, intra-coded frames are periodically introduced at regular intervals to prevent accumulation of prediction error over frames. It is obvious that intra-coded frames would require more bits to encode as compared to inter-coded frames since the temporal redundancy is not exploited in the form. As shown in Fig.3 and Fig.4.

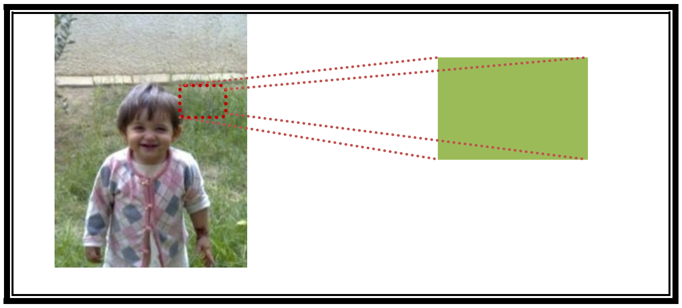

Fig.3: Intra frame.

Fig.4: Inter frame.

2.2 Discrete Wavelet Transform (DWT)

The basic operation in wavelet transform is to filter an image with a low pass filter (L) and a high pass filter (H) and down-sample the output by a factor of 2. Both operations on the x direction, two new images are obtained L and H. They are filtered and down sampled again but this time in the y direction. Four sub bands images are obtained which can be combined to recover the original one. The same amount of information is present, but this new configuration is more suitable for efficient coding [5].

The inverse wavelet transform is performed by enlarging the wavelet transform data to it is original size. Insert zeros between each of four sub images, and sum the results to obtain the original image [6].

2.3 The Embedded Zero Tree Wavelet (EZW) Quantization

An EZW encoder was especially designed by Shapiro to be used with wavelet transform. In fact, EZW coding is a quantization method. The EZW encoder is based on progressive encoding to compress an image into a bit stream with increasing accuracy. This means that when more bits are added to the stream, the decoded image will contain more detail. A zero tree is a tree of which all nodes are equal to or smaller than the root. The tree is coded with a single symbol and reconstructed by the decoder as a tree filled with zeroes [6].

2.4 Arithmetic Encoding

Arithmetic encoding, and its derivative technique, Q-coding, is used to overcome some of the limitations of Huffman codes. It is a non-block code, in that a single codeword is used to represent an entire sequence of input symbols, in contrast to Huffman coding where a source symbol block corresponds to a codeword block. Instead, it uses the real numbers to represent a sequence of symbols by recursively subdividing the interval between 0 and 1 to specify each successive symbol. The limitation of this technique is the precision required in performing the calculations and arriving at the code word which will represent the entire sequence correctly [7].

2.5 Huffman Encoding

A Huffman encoding developed by D.A. Huffman, a Huffman encoder takes a block of input characters with fixed length and produces a block of output bits of variable length. It is a fixed-to-variable length code. Huffman encoding uses a variable length code for each of the elements within the information. This normally involves analyzing the information to determine the probability of elements within the information. The most probable elements are coded with a few bits and the least probable coded with a greater number of bits [8].

2.6 PSNR and CR

Evaluation criteria that usually used in digital image and video compression are in two directions. First direction is to evaluate quality of the reconstructed image. Second direction is Compression Ratio (CR). In terms of quality evaluation, two mathematical metrics are used. First one is Mean Square Error (MSE), which measures the cumulative square error between the original and the reconstructed image. Second meter is Peak Signal-to-Noise Ratio (PSNR). The formula for MSE is giving as [9]:

Peak signal-to-noise ratio (PSNR) is the standard method for quantitatively comparing a compressed image with the original. For an 8-bit grayscale image, the peak signal value is 255. Hence, the PSNR of an M×N 8-bit grayscale image x and its reconstruction $\hat{x}$ is calculated as [9]:

$$PSNR = 10\log_{10}\frac{255^2}{MSE} \qquad \ldots (1)$$

where the Mean Square Error (MSE) is defined as:

$$MSE = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left[x(m,n) - \hat{x}(m,n)\right]^2 \qquad \ldots (2)$$

PSNR is measured in decibels (dB), M: height of the image, N: width of the image.

The second direction of comparing the compressed and the original images is the compression ratio. It is defined as [10]:

$$\text{Compression Ratio} = \frac{Size_c}{Size_u} = \frac{\text{Compressed File Size}}{\text{Uncompressed File Size}} \qquad \ldots (3)$$

In addition to measuring the quality of image, we also measure the compression ratio. Compression ratio is the ratio of the compressed file size to the original file size. In general, the higher the compression ratio, the smaller is the size of the compressed file. Compression speed, on the other hand, is the amount of time required to compress and decompress the image. This value depends on a number of factors, such as the complexity of the algorithm, the efficient of the implementation, and the speed of the processor [10].

2.7 Motion Estimation

Motion Estimation (ME) is the process of analyzing successive frames in a video sequence to identify objects motion. The motion of an object is usually described by a two-dimensional motion vector, which is the placement of the co-ordinate of the best similar block in previous frame for the block in current frame. This placement is represented by the length and direction of motion [11].

2.8 Motion Compensation

Motion Compensation (MC) predication has been used as a main tool to remove temporal redundancy that comes from little change in the content of the image from one video sequence to another. It provides coding system with a high compression ratio. This technique is adopted by all of the existing international video coding standards, such as Picture Expert Group (MPEG) series and H.26x series.

Motion compensation prediction assumes that the current frame can be locally modeled as a translation of the frames in the previous (or reference and next) time. Fig.5 shows motion compensation between two frames [12].

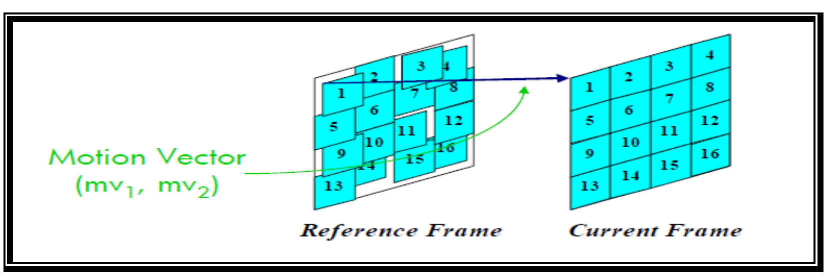

Fig.5: Motion compensation between two frames.

2.9 Motion Estimation Algorithms

These algorithms assume that a frame has been divided into M non-overlapping blocks that together cover the entire frame. Moreover, the motion in each block is assumed to be constant, that is, it is assumed that entire block undergoes a translation that can be encoded in the associated motion vector. The problem of block-based ME algorithms is to find the best MV for each block, as shown in Fig.6 these algorithms are also called **Block Matching Algorithms (BMA)** [13].

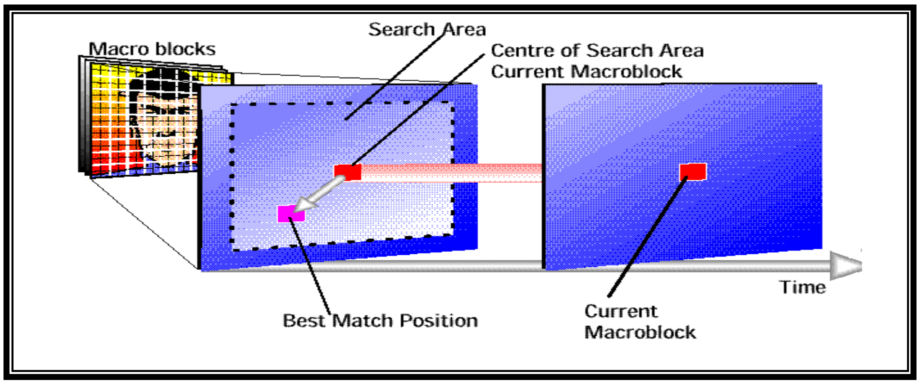

Fig.6: Block Matching Algorithms.

2.9.1 Block Matching Algorithms

Block Matching Algorithm (BMA) is the most popular technique used for motion estimation in which the current luminance frame is divided into non-overlapped macro blocks (MBs) of size NxN that are then compared with corresponding macro block (MB) and its adjacent neighbors in the reference frame to create a vector that stipulates the movement of a macro block from one location to another in the reference frame [14], i.e., finding matching macro block of the same size NxN in the search area in the reference frame.

The position of motion vector has two parts, horizontal and a vertical. These parts can be positive or negative. A positive value means motion was to the right or motion downward while a negative value means motion was to the left or motion upward. This Motion Vector (MV) will be used to predict new frame from the reference which is called **motion compensation**. The matching measurement is usually determined using one of Block Distortion Measure (BDM) like Mean Absolute Difference (MAD) given by Equation1 or MSE given by Equation2. The macro block with the least cost is considered the matching to the current frame macro block [15].

$$MAD = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left|C_{ij} - R_{ij}\right| \qquad \ldots(4)$$

$$MSE = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(C_{ij} - R_{ij}\right)^2 \qquad \ldots(5),$$

Where:

N^2: Block size N x N.
C_{ij}: Pixel in position (i, j) in current block.
R_{ij}: Pixel in position (i, j) in reference block.

2.9.2 New Three Step Search (NTSS)

NTSS improves on Three Step Search (TSS) results by providing a center biased searching scheme and having provisions for half way stop to reduce computational cost. It was one of the first widely accepted fast algorithms and frequently used for implementing earlier standards like MPEG 1 and H.261.The TSS uses a uniformly allocated checking pattern for motion detection and is prone to missing small motions. The NTSS process is illustrated graphically in Fig.7. In the first step 16 points are checked in addition to the search origin for lowest weight using a cost function. Of these additional search locations, 8 are a distance of S = 4 away (similar to TSS) and the other 8 are at S = 1 away from the search origin. If the lowest cost is at the origin then the search is stopped right here and the motion vector is set as (0, 0). If the lowest weight is at any one of the 8 locations at S = 1, then we change the origin of the search to that point and check for weights adjacent to it. Depending on which point it is we might end up checking 5 points or 3 points (Fig.8 (b) &(c)). The location that gives the lowest weight is the closest match and motion vector is set to that location. On the other hand if the lowest weight after the first step was one of the 8 locations at S = 4, then we follow the normal TSS procedure [16].

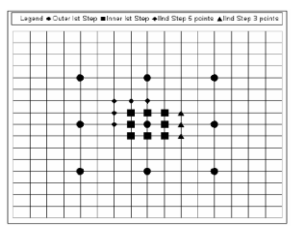

Fig.7: New Three Step Search (NTSS)

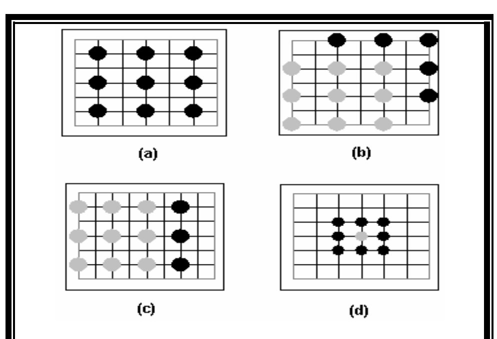

Fig.8: Search patterns of the 4SS. (a) First step (b) Second/Third step (c)Second/Third Step (d) Fourth Step

3. Proposed Compression Video Method

3.1 Intra Coded Frame

The proposed compression system in this method includes three stages, the first stage is the wavelet transform, here we use Haar filter. The output from the transform will be different sub bands with different important information. After that, EZW will quantize these sub bands in efficient manner then the output will be stream of zeros and ones, this stream will be compressed by arithmetic encoding.

Algorithm of Intra Coded Frame.

Input: Digital video clip.
Output: Compressed frame.

Step 1: Partition video into frames.
Step 2: Read first frame in video.
Step 3: Transform first frame using DWT.
Step 4: Quantization of first frame using EZW.
Step 5: Encoding of first frame using arithmetic method.
Step 6: Save the compressed frame.

3.2 Inter Coded Frame

The proposed compression system in this method applies DWT on video frames, then diamond search algorithm is used in order to find MV using forward motion estimation. Motion vector was coded by Huffman encoding. On the other hand, the remaining part (the similar blocks of frames) will be compressed as the compression method that was used in intra frame coded.

Algorithm of Inter Coded Frames.

Input: Digital video clip.
Output: Compressed video.

Step 1: Partition video into frames.
Step 2: Transform these frames by DWT.
Step 3: Divide these frames into 16x16 macro block.
Step 4: Estimate the motion between each two frames using DS algorithm.
Step 5: Encoding the result from Step 4 (MV) by Huffman encoding.
Step 6: Compress the remaining part of frames after find the motion as intra coded frame.
Step 7: Save the compressed video.

4. Experimental Results

This section explains the experiments which have been implemented on two video clips. Clip1 **(fish)** and clip2 **(ran)** as test clips, each one of them is in size of 128*128 and of JPG format. MATLAB version 7.4.0.287 (R2007a) was used as a work environment to carry out these experiments.

4.1 Results of Intra Coded Frame

In this experiment, the first frame of video has been compressed using the first method, namely; intra coded frame. In this method, this frame is compressed as a still image. Two gray scale clips were used in this experiment. Fig.9 and Fig.10 show the result of applying this method on these clips. Table1 and Table2 illustrate the PSNR and CR which are resulted from applying of this method on clip1 and clip2, respectively.

(a) (b)

Fig.9: Reconstructed frame resulted from applying of the first method on clip1.
(a) Original frame. (b) Reconstructed frame.

Table1: PSNR and CR of intra coded method on clip1.

	PSNR (dB)	CR
Frame1	61.697	0.1447

(a) (b)

Fig.10: Reconstructed frame resulted from applying of the first method on clip2.
(a) Original frame. (b) Reconstructed frame.

Table2: PSNR and CR of intra coded method on clip2.

	PSNR(dB)	CR
Frame1	59.389	0.1493

4.2 Results of Inter Coded Frame

In this experiment, frames starting from the second frame have been compressed using inter coded frame method. In this method, these frames are compressed using DWT followed by motion estimation using NTSS algorithm. Two gray scale clips were used in this experiment. Fig.11 and Fig.12 show the result of applying this method on these clips.

Table3 and Table4 illustrate the PSNR, CR, and processing time which are resulted from applying of this method on clip1 and clip2, respectively.

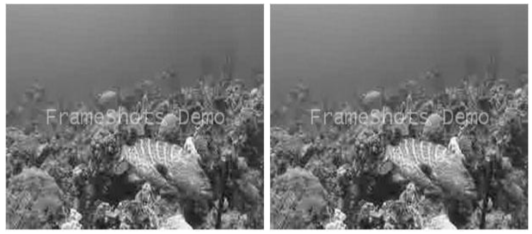

Fig.11: Reconstructed frame resulted from applying of the second method on clip1.
(a) Original frame. (b) Reconstructed frame.

Table3: PSNR, CR, and time of inter coded method on clip1.

	PSNR (dB)	**CR**	**Time (Sec.)**
Without DWT	55.953	0.125	6.1857
With DWT	58.451	0.135	3.2975

(a) (b)

Fig.11: Reconstructed frame resulted from applying of the second method on clip2.
(a) Original frame. (b) Reconstructed frame.

Table4: PSNR, CR, and time of Inter coded method on clip2.

	PSNR (dB)	**CR**	**Time (Sec.)**
Without DWT	58.259	0.135	4.5618
With DWT	60.233	0.145	2.1334

5. Discussion and Conclusions

In this paper, a method for video compression has been proposed. This method based on discrete wavelet transform and New Three Step Search (NTSS) algorithm as a block matching algorithm to find the motion vector which will be used at the stage of motion compensation to finally estimate the current frame depending on reference frame. Use of this method on two clips (one is considered as a standard clip and the another is non standard) has shown good results in terms of PSNR, CR, and processing time. PSNR value in NTSS algorithm is better with DWT proposed approach than this algorithm without PSNR, as shown in Tables 3, and 4 in clip1, and clip 2, respectively. But in CR, the first proposed approach (without DWT) is better than the second proposed approach (with DWT). It is clearly noticed that the use of DWT minimize the processing time, almost, up to 40% - 50%.

6. References

1. Sayood K., "Introduction to Data Compression", Morgan Kaufmann Publishers, 2006.
2. Bjorn B., Sweden, " Image and Video Compression Using Wavelet Transform and Error Robust Transform", M.Sc. Thesis, Stockholm, Sweden KTH Electrical Engineering, September 2005.
3. Hassan B. and Malik K. , "Quality-Aware Frame Skipping for MPEG-2 Video Based on Inter Frame Similarity", The Department of Computer Science and Electronics, Malardalen University, Vasteras, Sweden, 2006.
4. Niehsen W., "Fast Full Search Block Matching", In IEEE Transactions on Circuits and Systems for Video Technology, pp. 241-247, 2001.
5. Keinert F., "Wavelets and Multiwavelets", USA, 2004.
6. Erick S., "Compression of Medical Image Stacks Using Wavelet and Zero-tree Coding ", M.Sc. Thesis, Department of Electrical Engineering, Linkoping University, 'Lith-ISY-Ex-3201', 2002.
7. Saif B., "Wavelet Compression Using Tree and Adaptive Arithmetic Codes", M.Sc Thesis, Baghdad University, College of Science, 2004.
8. Fast Huffman Code Processing", UCI-ICS Technical Report No. 99-43, Department of Information and Computer Science, University of California, Irvine, October 1999.
9. John M., "Compressed Image File Formats", ACM Press, A Division of The Association of Computing Machinery, Inc. (ACM), 1999.
10. Panrong X., "Image Compression by Wavelet Transform", M.Sc. Thesis, East Tennessee State University, Department of Computer and Information Sciences, 2001.
11. Shi Q. and Sun H., "Image and Video Compression for Multimedia Engineering", 2000.
12. Aydin B., " Motion Compensated Three Dimensional Wavelet Transform Based Video Compression and Coding ", M.Sc. Thesis, University of Middle East Technical University, Electrical and Electronics Engineering Science , January 2005.
13. Djordje M., "Video Compression", University of Edinburgh, 2008. http://homepage.inf.ed.ac.uk/rbf/CVonline/LOCALCOPIES/AVo506/s0561282.pdf.
14. Koga T., Iinuma K., Hirano A., Iijima Y. and Ishiguro T., "Motion Compensated Inter Frame Coding for Video- conferencing", Proc. NTC81, Nov. 1981.
15. Jain J. and Jain A., "Displacement Measurement and Its Applications", IEEE Transactions on Communications, Dec. 1981.
16. Zeng B. and Rud L., "A New Three Step Search Algorithm for Block Motion Estimation", IEEE Transactions on Video Processing, 2005.