TEM Journal. Volume 5, Issue 3, Pages 277-283, ISSN 2217-8309, DOI: 10.18421/TEM53-04, August 2016.

Research of Block-Based Motion Estimation Methods for Video Compression

Tropchenko Andrey [1], Tropchenko Alexandr [1], Nguyen Van Truong [1]

[1] *ITMO University, Department of Software Engineering and Computer Systems, Chair of Computation Technologies, Kronverkskiy pr. 49, S-Petersburg, Russia*

Abstract **– This work is a review of the block-based algorithms used for motion estimation in video compression. It researches different types of block-based algorithms that range from the simplest named Full Search to the fast adaptive algorithms like Hierarchical Search. The algorithms evaluated in this paper are widely accepted by the video compressing community and have been used in implementing various standards, such as MPEG-4 Visual and H.264. The work also presents a very brief introduction to the entire flow of video compression.**

Keywords **– Video compression, spatial redundancy, temporal redundancy, motion estimation, motion compensation.**

1. Introduction

In the early eighties the processing sequences of digital images became a subject of research of many scientific communities. This is not surprising because group of images has much more information about the object than its individual image. Today transfer systems, processing system, storage systems of the video became integral part of ordinary people's life who have not special skills. The ever-growing computational complexity of processing algorithms and high cost of storing video data are becoming more visible even with the ever-increasing computing power.

Video compression algorithms use, firstly, the features of the original data, such as information redundancy, smoothness of its changes, and secondly, the characteristics of human perception, i.e. weak sensitivity of the eye to a slight distortion when restoring. Algorithms for lossy compression (data compression in which some of the information is lost and the quality is damaged) actively uses this feature.

Speaking about the peculiarities of human perception, in the color planes of image there is some redundancy which is called redundancy of color space. Indeed, image brightness is the most important for the perception. As an application of this knowledge, the standard RGB color representation scheme must be replaced by YUV scheme with decimation of the corresponding components.

Redundancy of video data is divided into spatial and temporal. Spatial redundancy or the similarity values of neighboring pixels/smoothness color transitions in the frame means the predominance of low frequency signal representation over it's high frequency representation. Its elimination is used in the algorithms based on different types of discrete transformations. Removal of temporal redundancy uses the assumption that in a short period of time corresponding to several frames the objects presented in video scene changes insignificantly. In this connection per-pixel difference between two successive frames will be close to zero. Although compression of only differences between adjacent frames instead of themselves frames imposes some restrictions on the processes of compression and decompression. This approach is used by almost all algorithms for video compression.

2. The main stages of video compression

Consider a typical procedure of the video sequence compressor (Figure 1).

There are two main functional units: a temporary model and a spatial model. Time model seeks to reduce temporal redundancy; spatial model also uses the similarity or likeness of neighboring samples of the frame, reducing the spatial redundancy.

DOI: 10.18421/TEM53-04
https://dx.doi.org/10.18421/TEM53-04

Corresponding author: Tropchenko Andrey
ITMO University, Department of Software Engineering and Computer Systems, Chair of Computation Technologies, Kronverkskiy pr. 49, S-Petersburg, Russia
Email: zayka_98rus@mail.ru

In the first step of the encoding, each frame is converted from RGB representation to YUV. Thereafter the video sequence is preprocessed through a series of filters, from which the minor parts and the jitter (unwanted rapid camera motion) are removed that increase the compression ratio. This eliminated the high frequency components. Then the color decimation (gamma correction) is executed. Due to the fact that the human eye responds to the luminance variation is nonlinear, luminance of pixels is scaled using a power function.

Further in the simplest case of the independent frame compression it falls on input transducers from spatial representation of the signal to spatial frequency, otherwise it is the reference frame and is activated the motion estimation and compensation scheme and only then frame is converted. The target of this step is reflection of digital data of frames into another coordinate space (transform domain).

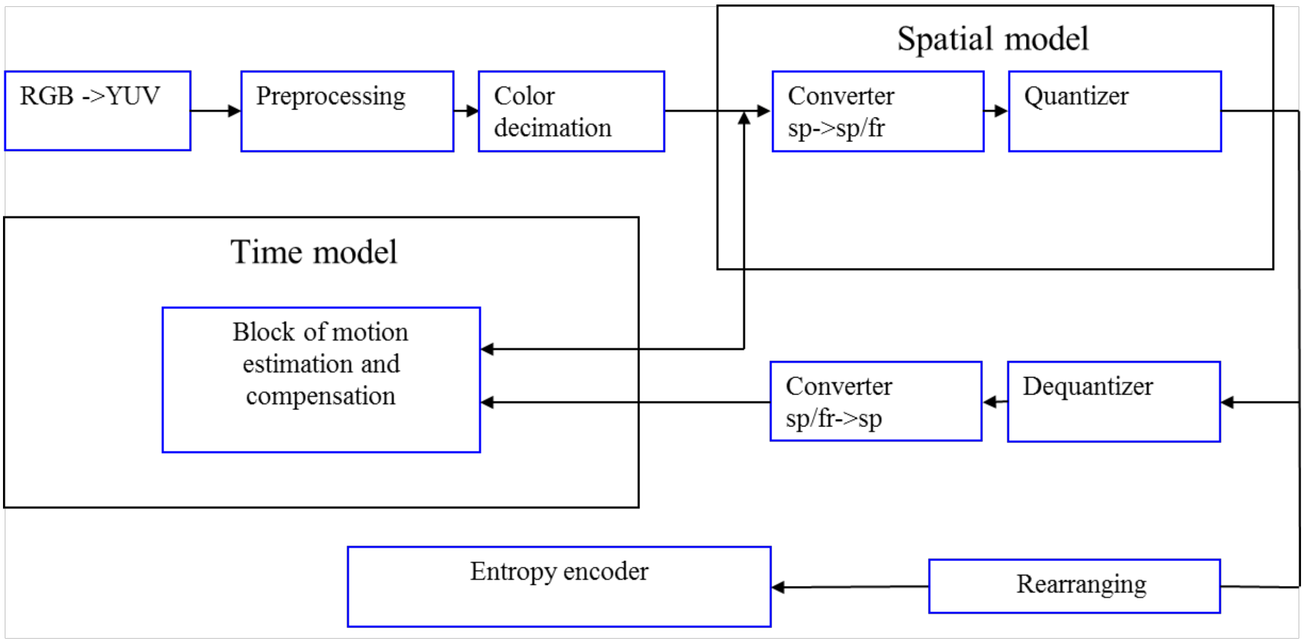

Figure 1: Typical procedure of the compressor of a video sequence

The input stream of converter is counts of the original signal, at the output expansion coefficients are obtained in the basis which consists from localized functions by frequencies and in the space. One can allocate the conversion based on the block and the conversion based on the image (discrete wavelet-transform).

The first type of transformation operates with square image blocks, elements of which are samples (usually samples of the image or values of the differences from a prognosis of image), and after a series of operations it generates an equilateral coefficient block. Any block image can be restored by a linear combination of N x N basic patterns, where the basic patterns are multiplied by the appropriate weight coefficients (conversion coefficients).

Block transformations have low memory requirements and better suited for the compressing image retention (after the passage of time model), but as a result there are a lot of arising at the interface blocks artifacts.

In the second type to the original signal as a sequence of discrete values is used the pair of filters which splits it into two components: a low-frequency (L) and high-frequency (H) subbands. Each subband is decimated through one element and in each sequence of frequencies remains N/2 samples. This operation is correct with proper selection of filters.

Wavelet-transform is applied by rows and columns for each level of decomposition. The advantages of this method is that it does not result in correct form artifacts (blocking – effect) and has good scalability. For example, in order to reduce the resolution of already compressed image in four times it is possible to leave only one factor of four.

However, this solution has a number of drawbacks such as the effects of blurring edges, loss of detail and artifacts near the boundaries of objects. Also for their work we required a large amount of memory.

A converted video sequence supplies to a quantizer. Each sample in scalar or group of samples with vector quantization of the original signal is placed in correspond to the number.

In the first case the signal samples are split in accordance with the quantization step. In the quantization process the value of each count is replaced by index interval in which it falls. When decoding the index is replaced by the centroid - the average value of count of the signal of a particular interval. This type of quantization can be used to lower the accuracy of the image after applying the encoding conversion, for example, DCT or wavelet - transformation.

In the second case the original signal is divided into rectangular areas, which is grouped by similarity to some criterion. For this group is calculated the weighted average area which enters in the dictionary tables. During decoding process these fields are replaced by the corresponding indices of the dictionary, the relevant portions of which are similar to the original. At the equal level of signal distortion, this scheme promotes greater reduction of its correlation compared with the scalar. However it is inherent in a serious drawback - the high computational complexity of the dictionary constructing that prevents the widespread use of this method. However in connection with feature of video compression one can use the dictionary of the previous image to build a new, greatly simplifying the process

After quantization, the coefficients must be reordered into a group of non-zero elements. Optimal scan order depends on the distribution of non-zero coefficients (after scanning they should be located before zero). For a typical image block, the best procedure is the zigzag scanning starting at the upper left corner.

Efficient motion estimation and compensation reduces the correlation of samples, allowing efficient compression of image sequences compared to the original video frames. This model usually uses algorithms for the prediction encoding. In this case, the encoder makes a prediction for the current field based on previous or future frames and subtracts this area - the forecast from the current scope. If the prediction was done correctly, the resulting residue can be represented by fewer bits.

Parameters of time model (motion vectors) and of spatial model (conversion factors) are supplied to the input of the entropy encoder. The motion vectors are usually presented in the form of coordinates of movement vector of all the motion-compensated blocks with integral or fractional pixel resolution. The input data can also be markers (codes for the synchronization point in the video sequence) and headers (headers of macroblocks, images, sequences and other objects).

During the passage of the video sequence through the entropy encoder, it is compressed element by element. This is achieved by the use of information about the probability of occurrence of each symbol sequence. The following types of coding: the modified variable-length Huffman codes and the arithmetic coding.

Decoding scheme works in the opposite direction.

3. Block-based methods

As noted earlier, the purpose of the temporary model is removal of redundancy between transmitted frames. The main purpose of this scheme is the compilation of frame - of forecast. The faster record subjects move, the more different blocks of the current frame from the corresponding previous frame and the greater prediction error and the volume of transmitted data, which reduces the compression efficiency. The better the prediction, the less energy is contained in the residual block (the difference between the current frame and the prediction). Decoder, receiving the frame, reproduce frame-forecast by reference frames and adds it to the residual frame. The result is as much as possible approached frame to the original frame.

Consider in more detail the nature of the differences between adjacent frames of a video sequence. They are caused by moving objects, cameras, objects overlap each other, changes in illumination.

Depending on the intermediate data, all approaches to the analysis of motion are divided into two categories: determination of optical flow or continuous approach and correlation characteristics or discrete approach.

In the first case the difference between adjacent frames is considered as a number of moving pixels on the frame. Thus, it is possible to trace the motion trajectory of the pixel between sequential frames. In this case, the optical stream is called generated field of trajectories. Determining the field of the stream one can construct an accurate prediction for the majority of the pixels of the frame. Each pixel in this case would correspond to its optical flow vector for the prediction. This approach appeared to compensate for motion of the first one. The scheme to be considered to account for the movement takes into account only the linear shifts.

Thus, it is assumed that the pixel value is obtained using a linear function of its position in the frame. However, this hike is justified only to a small neighborhood of the point, which significantly reduces the scope of this algorithm.

The situation may be somewhat corrected, estimating the difference of the shift vector with some vector prediction. This approach reduces the amount of transmitted corrections that is the decision of the discharge translated true / false in the category enough/sufficient accuracy. Thus, iterative algorithm of motion vector search is obtained. At each step achieving some precision of construction is checked, and if it is not achieved, the vector is specified. Despite the simplicity of formulation this method is not characterized as practical and it has serious deficiencies.

Firstly the calculation process of the optical flow of resource is intensive due to multiple iterations for each pixel. Secondly, the method significantly increases the volume of traffic information transmitted due to the need to forward all the vectors of the optical flux to the decoder.

Therefore at present time this algorithm is rarely used and investigated.

The straight-line method for the analysis of motion instead of defining two-dimensional motion parameters, involves the assessment of three-dimensional motion without the use of a direct solution for intermediate values. Thus we can create a system of equations that is characterized as the displacement vector in the two-dimensional plane of the image, and motion parameters three-dimensional space, the solution of which will determine the parameters of motion in three-dimensional space. This method can be used for reconstruction of the object surface. Of course, this method has a number of limitations imposed by features of the geometry of the object.

In the case of class methods of correlating features only exceptional two-dimensional image features are analyzed under consideration. It should be noted that under this approach, there are two tasks. Firstly it is the task of feature extraction, and secondly it is the task of determining the correlation. In this case, the separation and overlapping of objects of a video sequence with each other contributes to the appearance and disappearance of the characteristic features, which greatly complicates the second task. This approach is in direct use is highly labor intensive, therefore it is very rarely used in the process of video compression in this form.

As a method of correlating features it is possible to allocate an object and a segmental approach to the analysis of motion.

Speaking about the first of them we can see that today there is no generally accepted method of selection of objects from the source images, but it is assumed that natural video scenes are presented by information about the shape of the objects in addition to the usual brightness and color components. Data on the shape are usually presented as a binary segmentation mask or alpha - plane with a gray scale to represent objects with multiple overlapping. Alpha mask determines whether the pixel belongs to the object. Mask with a gray scale provides accurately determined transparency of each pixel.

In this approach, a number of difficulties associated with the need of accurate and reliable description of the boundaries of the objects, segmentation and contour coding object boundaries for the decoder, encoding residue after motion compensation and so on.

Class of methods based on the segmental approach eliminates much of the disadvantages of the per pixel algorithm. Rectangular block is a reimbursable essence in this method. Motion is described by the two-dimensional displacement vector of the block. In this approach is used the assumption that within the framework of two adjacent frames the location and the shape of objects are changed slightly. Then this change can be compensated by a parallel translation of the segment by some vector. This assumption works for the vast majority of frames of a video sequence, except for the sections complete change of frame when switching stage (Figure 2) [1-4].

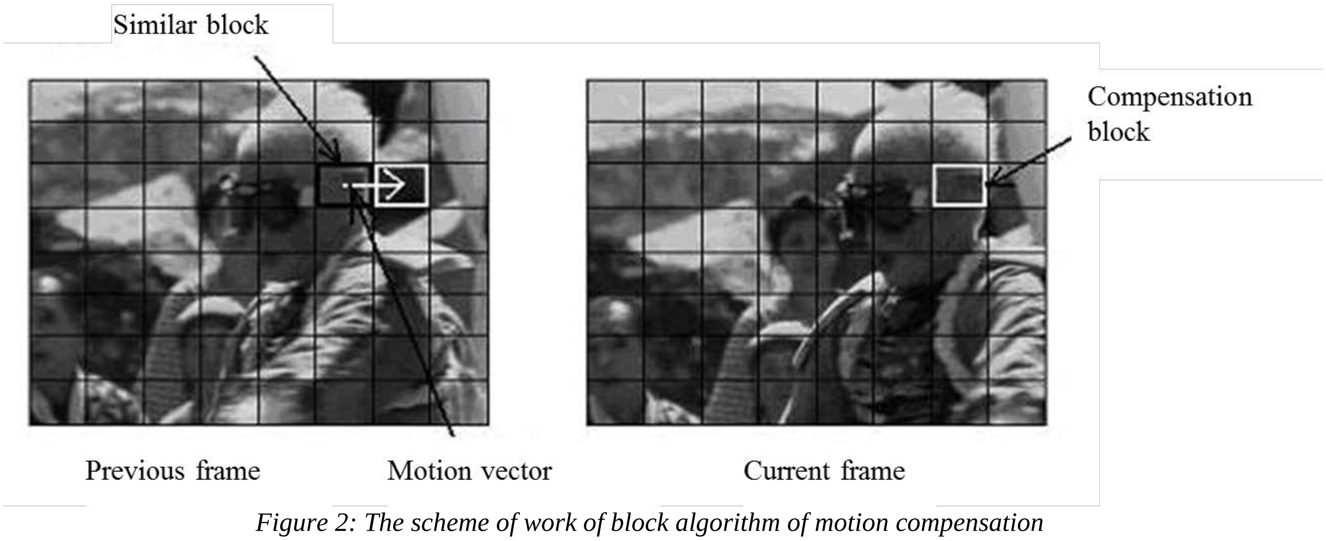

Figure 2: The scheme of work of block algorithm of motion compensation

In operation of the algorithm the frame is partitioned into disjoint blocks of one size. Then, for each block in a neighborhood is searched for a block of the previous frame corresponding to the minimum criterion. The thus obtained vector is the displacement vector for that block.

In its turn, the set of algorithms of the segmented approach is differed by method of selecting blocks for comparison.

Algorithms are the basic algorithms of this class [5-11]: Full Search (FS), One at a Time Algorithm (OTA), Orthogonal Search Algorithm (OSA), a Three-Step Search (TSS), Two-Dimensional Logarithmic Search (TDL), Four-Step Search (FSS), a Hierarchical Search – the method of averaged pyramid (MP).

- **Algorithm FS** (algorithm of full search). The algorithm assumes search of all possible variants of the forecast for the block. This scheme has high computational complexity in predicting the maximum quality and can be used as a reference for comparison.

- **Algorithm OTA** (One at a Time Algorithm). The algorithm is simple to implement, but is effective in finding the optimum position of the block. During the initial stage of the algorithm it is searched the block with minimal deviation in horizontal direction, and then search started in the vertical direction with a minimum deviation block. Obligatory condition expanding the search area is to reduce the amount of deviation; otherwise one of the stages of the algorithm ends. The scheme of work of algorithm is shown in Figure 3, where block A - minimum on the horizontal step, and B - seeking compensation block.

The disadvantage of this algorithm is that it is impossible to predict the number of processing blocks. It is also necessary to take into account the fact that the function of error compensation is almost never monotonous, often a set of its local extrema, complicates the search for a global extremum. Therefore, it seems appropriate to use the algorithms with different templates.

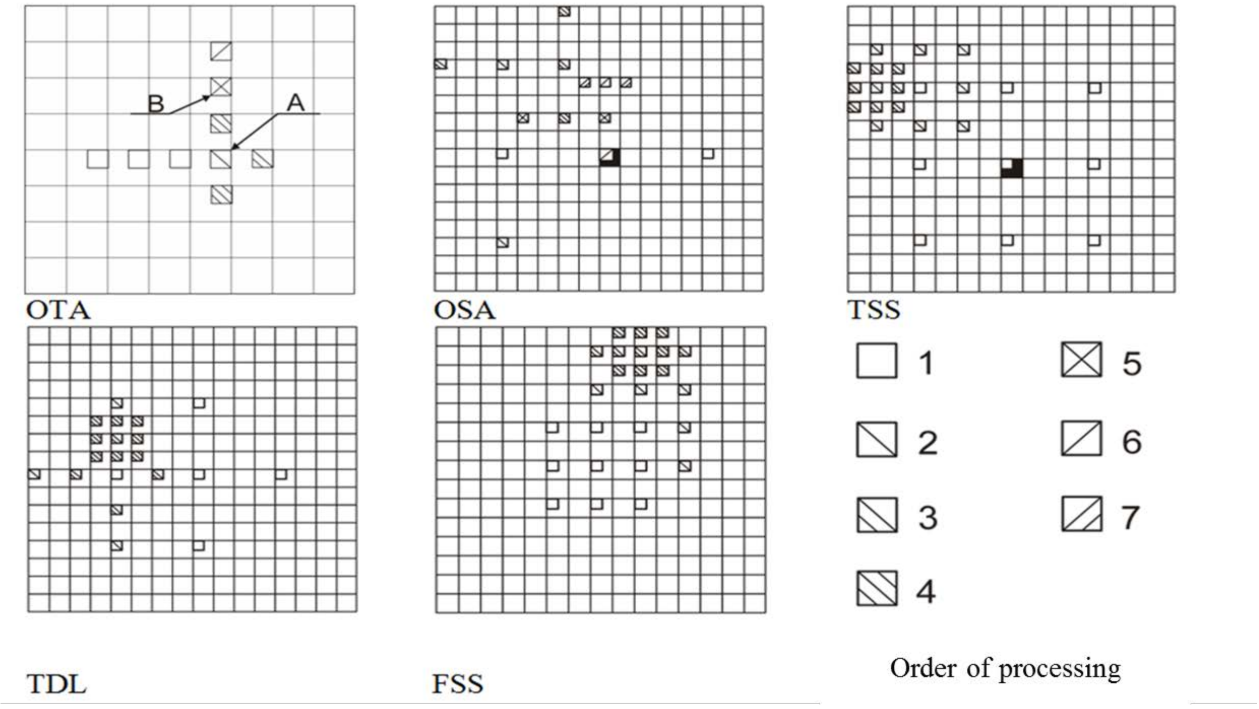

Figure 3: The scheme of work of the various algorithms

- **Algorithm OSA** (orthogonal search algorithm). To reduce the probability of finding of a local minimum instead of the global algorithm using orthogonal search it solves the problem of finding a block of a finite number of steps. It uses two alternating templates in the process of moving the center of the field in the block position with a smaller deviation with the cyclic reduction of the shoulder (Figure 3). A series of experiments showed that the use of this template also does not eliminate the possibility of accidental coincidences. This leads to the need for algorithms of multi-point templates.

- Algorithms of multipoint templates TSS, TDL, FSS.

- **Algorithm TSS** (three-step search) was developed in 1981 and is still popular because of its simplicity, reliability and high performance (Figure 3). The main problem of the algorithm is to remove points of uniformly distributed pattern, which makes it inefficient for small areas of motion.

- **Algorithm TDL** (two-dimensional logarithmic search) requires several more steps, but may be more accurate, especially in cases of large windows. In contrast to TSS the current center search area is included in the set of candidates. Condition of reduction step of the algorithm is the determination of the center of the field by block of the smallest differences. In the last step of the algorithm, the pattern is changed (Figure 3). There are many variations of this algorithm, which is characterized by condition changing step pattern moreover, the reduction step twice is not always the best solution.

- **Algorithm FSS** (four-step search) is based on this property of most video sequences, that they are orientated to the center of the frame. The first and the last stage of the algorithm, a nine-point pattern is used in two other depending on location block with the minimum criterion function to select one of the six patterns (Figure 3). As a rule, this algorithm shows higher reliability with preservation of efficiency for complex variants of motion and scaling operations. It makes FSS attractive strategy for selecting the blocks in a motion compensation.

- **Algorithm MP** (hierarchical search – method of averaged pyramid). To reduce the complexity of the motion search algorithms, course-to-fine hierarchical searching schemes have been suggested. This reduction is achieved by compensating on low-resolution frame. At the beginning, to eliminate the effect of noise the low-resolution image is obtained by low-pass filter. In the future, for a multi-level hierarchy of low-resolution images, a simple averaging of image pixels of the previous level as shown in Figure 4. Thus, when using three hierarchical levels, one pixel of the level 2 corresponds to a block 4×4 of level 0 and block 2×2 – 1, respectively. At the same time, the block size 16×16 level 0 will correspond to the block (16/2L) × (16/2L) level L.

After building of the averaged pyramids to count mean absolute deviation (MAD) and the choice of vector having the smallest MAD as a rough motion vector 2 level. The found vector moves to level 1 and made his clarification. The same process is repeated for level 0, on which the desired vector is turned. To increase the accuracy of the algorithm, this procedure may be performed for several vectors of level 2, having similar MAD values using windows (search areas) of reduced size (Figure 4).

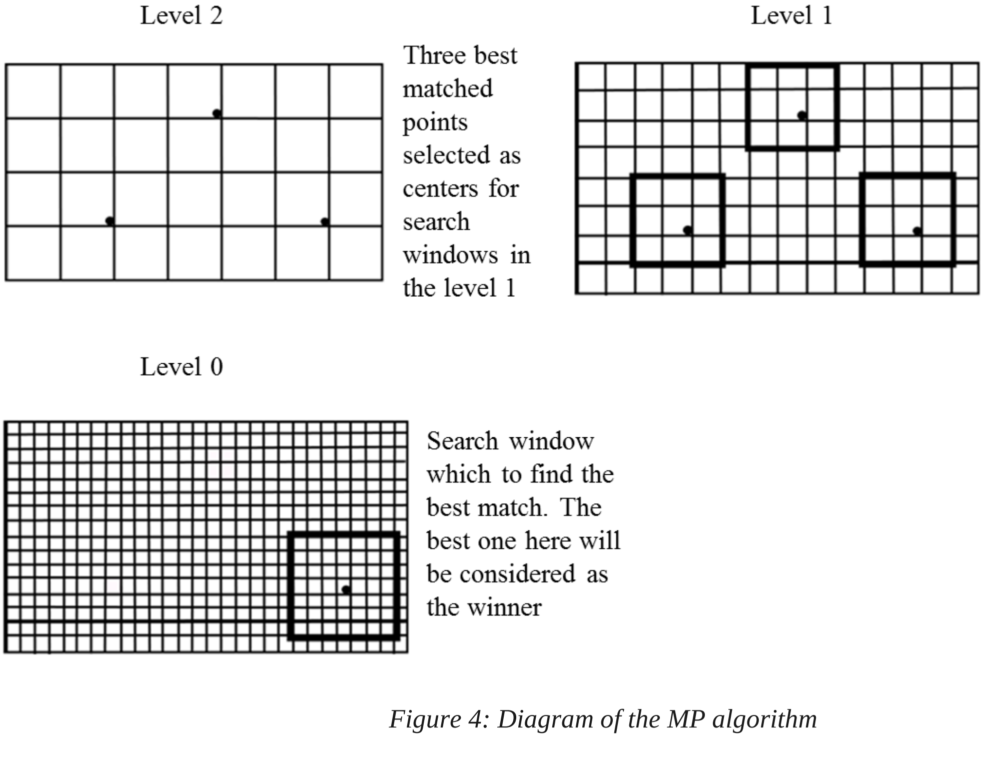
Level 2
Three best matched points selected as centers for search windows in the level 1
Level 1
Search windows in where to find the best match in the second level
Level 0
Search window which to find the best match. The best one here will be considered as the winner

Figure 4: Diagram of the MP algorithm

When we describe the basic algorithms of selection units for comparison within the segment approach to motion estimation and compensation, we move on to the overall analysis of the disadvantages of modern video compression algorithms.

4. Conclusion

To date is developed a number of compression algorithms of video sequences described in the video compression standards, where the vast majority of encoding/decoding schemes is based. Among them, a special place is occupied by the MPEG-4 Visual and

H.264 which were developed by experts from around the world. Having common sources, these standards went by their separate ways and are specialized for specific subject areas.

Developers of H.264 standard set a goal to raise the level of reliability, the compression efficiency, while maintaining compatibility with existing technologies such as streaming multimedia, video storage, high-definition television.

The compression algorithm uses a half pixel motion compensation cyclic filter, foreign motion vector, context- oriented arithmetic coding, an improved prediction mode, as well as all types of frames, four motion vectors per macroblock, motion compensation blocks with overlapping, as well as alternative quantizer.

The use of H.264 allows to lower the considerably load on the data network and reduce the cost of storage device to store the video. Unfortunately, today there are factors that limit the ubiquitous standard. Is in its infancy, the standard requires the use of high-performance cameras and requires a relatively large computing power. Adhering to segmentation approach of motion compensation, the developers of the standard overlook information about community groups of blocks within the frame.

Standard MPEG-4 Visual has already gained popularity among a wide range of users. Aimed at variability, he came to the object of arbitrary form, is flexible and adaptive, providing transparent transmission stream, allowing to maintain a decent level of video compression.

This implies the use of a subset of functions to support coding of specific actions such as basic video coding processing interlaced video encoded form description of the object - element of rectangular frames of a video sequence, regions of arbitrary form or a still image for encoding by using one or more tools.

MPEG-4 Visual provides a high level of interaction with the structure of the object, which allows the transmission over networks with low bandwidth. Today, however, only a small part of the possibilities of the standard is used. Also, there is relatively less opportunity to reduce the average density of the data stream. The presence of a noticeable blocking effect at high degrees of video compression is undoubtedly a serious drawback of the standard and requires no additional hardware. In this case, today has special complexity the process of determining the object's shape prediction in natural video scenes.

Based on the described features of the existing algorithms, it is possible to put the research problem as follows. It is necessary to develop a method of compression of video sequences, combines the advantages of the segmentation and objective approach to maximize the quality of the reconstructed video sequence and minimize their weaknesses by reducing the computational costs and reduce the amount of information transmitted on the motion.

References

[1]. Moschetti, F., Kunt, M., & Debes, E. (2003). A statistical adaptive block-matching motion estimation. *IEEE transactions on circuits and systems for video technology*, *13*(5), 417-431.

[2]. Babu, D. V., Subramanian, P., & Karthikeyan, C. (2006, December). Performance analysis of block matching algorithms for highly scalable video compression. In *Ad Hoc and Ubiquitous Computing, 2006. ISAUHC'06. International Symposium on* (pp. 179-182). IEEE.

[3]. Barjatya, A. (2004). Block matching algorithms for motion estimation. *IEEE Transactions Evolution Computation*, *8*(3), 225-239.

[4]. Cuevas, E., Zaldívar, D., Pérez-Cisneros, M., & Oliva, D. (2013). Block-matching algorithm based on differential evolution for motion estimation.*Engineering Applications of Artificial Intelligence*, *26*(1), 488-498.

[5]. Tauraga, D., & Alkanhal, M. (1998). Search algorithms for Block Matching Estimation. *Mid-term Project, spring*.

[6]. Toivonen, T., Heikkilä, J., & Silvén, O. (2002, November). A new algorithm for fast full search block motion estimation based on number theoretic transforms. In *Proceedings of the 9th International Workshop on Systems, Signals, and Image Processing* (pp. 90-94).

[7]. Chen, M. J., Chen, L. G., & Chiueh, T. D. (1994). One-dimensional full search motion estimation algorithm for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *4*(5), 504-509.

[8]. Li, R., Zeng, B., & Liou, M. L. (1994). A new three-step search algorithm for block motion estimation. *IEEE transactions on circuits and systems for video technology*, *4*(4), 438-442.

[9]. Chau, L. P., & Jing, X. (2003, April). Efficient three-step search algorithm for block motion estimation in video coding. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on* (Vol. 3, pp. III-421). IEEE.

[10]. Po, L. M., & Ma, W. C. (1996). A novel four-step search algorithm for fast block motion estimation. *IEEE transactions on circuits and systems for video technology*, *6*(3), 313-317.

[11].Nam, K. M., Kim, J. S., Park, R. H., & Shim, Y. S. (1995). A fast hierarchical motion vector estimation algorithm using mean pyramid. *IEEE Transactions on Circuits and Systems for Video technology*, *5*(4), 344-351.