Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering

Basılı/Printed ISSN: 1302-9304. Elektronik/Online ISSN: 2547-958X

Bi-Directional Motion Estimation For Motion Compensation In The Wavelet Domain

Dalgacık Alanında Hareket Dengeleme İçin Çift Yönlü Hareket Kestirimi

Vildan Atalay Aydın

Tekirdağ Namık Kemal Üniversitesi Çorlu Mühendislik Fakültesi, Tekirdağ, TÜRKİYE
Sorumlu Yazar / Corresponding Author: vatalay@nku.edu.tr

Geliş Tarihi / Received: 24.09.2019 *Araştırma Makalesi/Research Article*
Kabul Tarihi / Accepted: 08.05.2020 *DOI:10.21205/deufmd.2020226627*
Atıf şekli/ How to cite: ATALAY AYDIN, V., Bi-Directional Motion Estimation For Motion Compensation In The Wavelet Domain. DEUFMD 22(66), 941-948.

Abstract

One of the main goals of video coding is to reduce the required bitrate to encode a video file to be transmitted or stored. Central to the video coding research is the concept of motion compensation, which has a required pre-processing step of motion estimation. Motion estimation methods used in video coding can be categorized into two groups employing either one (i.e. uni-directional) or two-directional (i.e. bi-directional) search algorithms. Although the uni-directional algorithms are faster than the bi-directional ones, the latter methods have the advantage of recovering objects affected by occlusion which in turn helps to increase the quality of reconstructed video sequences in the decoder side and to reduce the required bitrate to encode a video file. Therefore, our focus in this paper is to analyze the impact of a bi-directional wavelet-based motion estimation algorithm for motion compensation. The proposed method is performed in the wavelet domain using the relationship between wavelet subbands which decrease the computational cost of the search technique by reducing the size of the video frames with the wavelet decomposition. Experimental results demonstrate that the proposed technique outperforms both uni- and bi-directional motion estimation methods in reducing the required bits to be encoded both in spatial and wavelet domains.

Keywords: *Motion estimation, Motion Compensation, Video Coding, Wavelet Transform*

Öz

Video kodlamanın ana hedeflerinden biri, iletilecek ya da depolanacak bir video dosyasını kodlamak için gerekli bit oranını azaltmaktır. Video kodlama araştırmalarının merkezinde, hareket kestiriminin gerekli bir ön-işlem adımı olduğu hareket dengeleme kavramı bulunur. Video kodlamada kullanılan hareket kestirimi yöntemleri, bir (tek yönlü) veya iki yönlü (çift yönlü) arama algoritmaları kullanan iki gruba ayrılabilir. Tek yönlü algoritmalar çift yönlü olanlardan daha hızlı olmasına rağmen, ikinci gruptaki yöntemler, kod çözücü tarafında yeniden oluşturulan video dizinlerinin kalitesini artırma ve bir video dosyasını kodlamak için gereken bit oranını azaltmaya yardımcı olan, örtmeden etkilenmiş nesneleri kurtarma avantajına sahiptir. Bu nedenle, bu makalenin odağı, hareket dengeleme için iki yönlü dalgacık-tabanlı bir hareket kestirimi algoritmasının etkilerini analiz etmektir. Önerilen yöntem, dalgacık ayrıştırma ile video çerçevelerinin boyutunu küçülterek arama tekniğinin hesaplama maliyetini düşüren dalgacık alt bantları arasındaki ilişkiyi kullanarak dalgacık alanında gerçekleştirilir. Deneysel sonuçlar, önerilen tekniğin, kodlanması gereken bitlerin azaltılmasında hem tek yönlü hem de çift yönlü hareket kestirimi yöntemlerinden uzaysal ve dalgacık alanlarında daha iyi performansı olduğunu göstermektedir.

Anahtar Kelimeler: *Hareket kestirimi, Hareket dengeleme, Video Kodlama, Dalgacık Dönüşümü*

1. Introduction

Motion compensation is a method used in several different research areas such as video frame interpolation [1], video compression [2], action recognition [3], medical imaging [4], and remote sensing [5]. Our focus in this paper will be on motion compensation techniques for video compression. The goal of motion compensation in video coding is to reduce the required bitrate to encode a video sequence taking advantage of the fact that consecutive frames consist of similar content. The most employed solution is to encode a reference frame and the difference between a reference and a target frame. Motion compensation decreases the energy in this difference image even further. In the decoder side, the target frame is reconstructed using the difference image and the motion parameters. Since the difference image is motion-compensated, a motion estimation process is required as a pre-processing step.

Motion estimation is the process of finding the displacement between two frames in a video sequence. The most used methods are called pel-recursive and block matching. While pel-recursive methods recover the displacements per pixel, block matching algorithms assume that pixels in a block, in general, move in the same direction. Since it is easier to realize block matching algorithms than pel-recursive ones in hardware, former methods are employed more in general. In order to estimate motion between two blocks, the first step is to select a reference and a target frame. The second step is to divide the target frame into blocks, and search for the blocks of the target frame in a search area in the reference frame. The search area is determined by a window size defined beforehand. After motion parameters are discovered for all blocks, the reference image blocks are mapped onto the target ones, and the difference frame is generated subtracting the mapped reference image from the target frame, which is called the motion compensation procedure.

Motion estimation can be performed using either uni- or bi-directional search which is determined by the type of frame compression. In video coding, frames are encoded in three different ways, which are called I (Intra-coded, encoded in full resolution), P (Predicted, uni-directional motion estimation from a previous frame), and B (Bi-directional predicted, bi-directional motion estimation from both a previous and a future frame) frames. While uni-directional estimation is faster due to the fact that bi-directional counterpart needs the future frames in the video sequence to be encoded/decoded; bi-directional estimation can, in fact, recover the objects in a frame that could be affected by occlusion. This has the advantages of both recovering higher quality videos in the decoder side and decreasing the required bitrate to be encoded in the encoder side. Therefore, in this paper, we propose a bi-directional motion estimation method which is an improvement on the method proposed by Aydin and Foroosh [2]. While Aydin and Foroosh [2] demonstrate encoding P frames in a video sequence, we, on the other hand, expose a method to encode B frames to prove a reduction in bitrate. The proposed method is later employed in a motion compensation step in order to demonstrate that the bi-directional method decreases the motion-compensated difference between frames. Experimental results demonstrate that the proposed method has higher quality results compared to other techniques both for uni- and bi-directional motion estimation in spatial and wavelet domains.

The rest of the paper is organized as follows. In Section 2, we provide a literature review, then in Section 3, we explain the proposed method in detail. In Section 4, we analyze the experimental results, and finally, in Section 5, we conclude the paper as well as point out future research directions.

2. Literature Review

Algorithms proposed to decrease the computational cost of block matching based motion estimation algorithms are summarized as follows. Zhu and Ma [6] propose a diamond search algorithm (DSA) that performs motion estimation in two steps where the first step searches nine points and the second one searches five points. Lam et al. [7] improve the DSA method by changing the pattern to a cross-shaped search. Moreover, Luo and Konofagou [8] utilize pre-calculated sum tables for fast normalized cross-correlation calculations for motion estimation in ultrasound applications; and, Werlberger et al. [9] employ non-local total variation regularization with a low-level segmentation process. Cuevas et al. [10] reduce the number of search locations in block matching motion estimation by a method based on Artificial Bee Colony optimization; while, in order to decrease the encoding complexity of H.265/HEVC encoder, Pan et al. [11] partition the prediction units into two groups as parent and children.

Summary of the works on the use of motion estimation methods in motion compensation is given in the following way. Nakaya and Harashima [12] propose using affine, bilinear, and perspective transformation between images instead of only translational displacements and introduce a matching-based motion estimation method. Furthermore, Fan et al. [13] use the single view with a depth map format for 3D video and propose a compression method for the 3D depth map based on motion estimation/compensation.

Motion estimation/compensation is used in video frame rate-up problem, as well. As an example, Wang et al. [14] propose a technique called irregular grid expanded block weighted motion compensation to decrease blocking problems and a method called block-wise directional hole interpolation to fill the holes caused by motion compensation step.

3. Bi-Directional Wavelet-Based Motion Estimation for Motion Compensation

In this section, we describe the proposed method in detail.

The flowchart of the proposed algorithm is illustrated in Figure 1. As can be seen from the figure, the first step of the algorithm is to select three consecutive frames from a video sequence. Uni-directional methods use a reference and a target image; while, bi-directional ones employ two reference images, a previous and a future frame with respect to the target one. After selecting the images, 1-level Discrete Wavelet Transform (DWT) is performed on each frame.

A pictorial explanation of the uni-directional motion estimation step in the wavelet domain is shown in Figure 2. Here, A, H, V, and D stand for approximation, horizontal, vertical, and diagonal detail coefficients of the DWT of images. Subscripts R and T demonstrate reference and target frames, respectively.

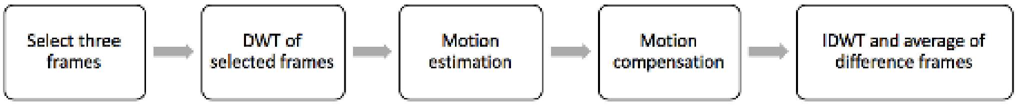


Figure 1. Flowchart of the proposed algorithm

After decomposing the reference and target frames into wavelet subbands (i.e. Step 2 in Figure 1), target frame subbands are divided into non-overlapping blocks as in the spatial domain block matching algorithms where the original frame is divided into such blocks. One block for each subband is displayed as an example in Figure 2 with dark-squared areas. Later, these subband blocks are searched in the reference frame subbands looking for where they are coming from. The reference frame search blocks are overlapping and the search windows which are shown in dark squares in Figure 2 are determined beforehand. Exhaustive search is performed for all blocks. Block matching algorithms, in general, employ interpolation techniques in order to achieve sub-pixel precision in motion estimation. We, on the other hand, utilize the method proposed by Aydin and Foroosh [2] where the relationship between

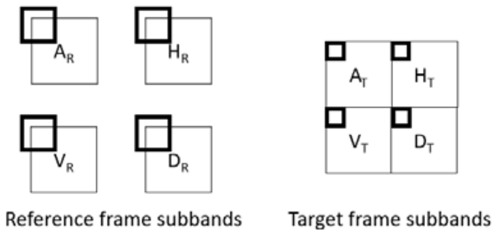


Figure 2. Block matching-based motion estimation in the wavelet domain

reference and target frame wavelet subbands is used to estimate motion with sub-pixel accuracy. For each block in the search window of the reference frame which is the size of blocks in the target frame, the formula from [2] to calculate blocks in the target frame is employed. In other words, blocks shown in dark squares for the target image in Figure 2 for all subbands (i.e. A_T, H_T, V_T, D_T) are searched in the reference frame subbands, using Equation (1). In order to make the paper self-contained, we will include the formula in [2] here, which is used to define the relationship between the subbands.

$$\begin{aligned} A_T = {} & F_y A_R F_x + F_y H_R K_x \\ & + K_y V_R F_x + K_y D_R K_x \\ H_T = {} & -F_y A_R K_x + F_y H_R L_x \\ & - K_y V_R K_x + K_y D_R L_x \\ V_T = {} & -K_y A_R F_x - K_y H_R K_x \\ & + L_y V_R F_x + L_y D_R K_x \\ D_T = {} & K_y A_R K_x - K_y H_R L_x \\ & - L_y V_R K_x + L_y D_R L_x \end{aligned} \tag{1}$$

Here, F, K, and L are called the coefficient matrices where x and y subscripts stand for x and y directions; the matrices are derived and explained in detail by Aydin and Foroosh [2]. The block matching algorithm is performed by calculating target frame subband blocks using the reference frame blocks from the search windows. Later, the block from the reference image search window with the sum of lowest mean square error for all subbands is chosen as the optimal solution. The formula to select the block with minimum error based on mean squared error is given as follows;

$$\min_b \sum_s \left(\frac{1}{n \times n} \sum_i \left(S_T - \hat{S}_T \right)^2 \right) \tag{2}$$

where S_T stands for all target frame wavelet subbands, $\hat{S}_T$ demonstrates related estimated subbands, i shows each pixel in the subbands, and $n \times n$ is the size of subbands. Finally, s and b demonstrate the sum over all subbands and minimum value in all searched blocks, respectively.

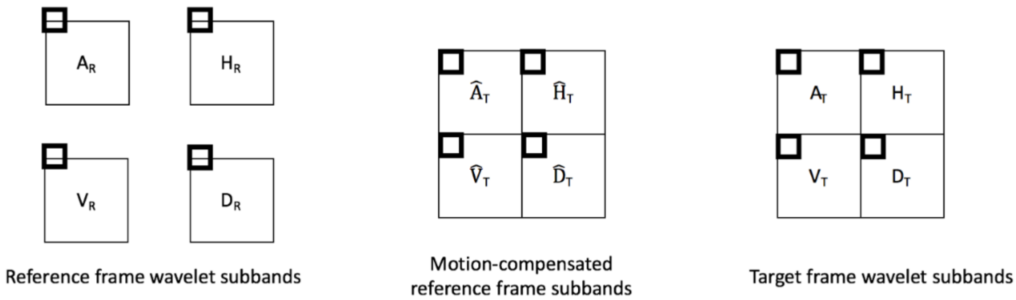


Figure 3. Motion compensation in the wavelet domain

Figure 3 shows a pictorial explanation of the motion compensation step in the wavelet domain, assuming that, as an example, the blocks with the minimum error are selected as the dark-squared areas in reference frame subbands.

When motion estimation is performed and motion parameters for each block in the target image are acquired, the reference frame blocks are shifted for the amount of motion vectors using Equation (1) to find motion compensated reference frame subbands (i.e. $\hat{A}_T, \hat{H}_T, \hat{V}_T, \hat{D}_T$). Later, the difference between all subbands of motion-compensated reference and target frames is found, and Inverse Discrete Wavelet Transform (IDWT) is performed on the difference subbands to find the difference image.

In the encoder side of a video coding framework, the obtained difference image and motion vectors for each block are encoded together with the reference frame, instead of encoding both reference and target frames in full resolution, in order to decrease the required bits to be encoded.

We update the method in [2] to reduce the bits even further, by employing bi-directional motion estimation. The pictorial explanation of the proposed method is shown in Figure 4. One can observe from the figure that two reference frames are chosen for the bi-directional motion estimation step. Reference frames 1 and 2 are selected as coming before and after the target frame in the video sequence, respectively. As in the uni-directional counterpart, target frame blocks are searched in the reference image search areas for both reference frames, looking for where the blocks are coming from.

directional search has the advantage of recovering objects that could be affected by occlusion, which in turn reduces the bits to be encoded for the video coding process that will be presented in the next section.

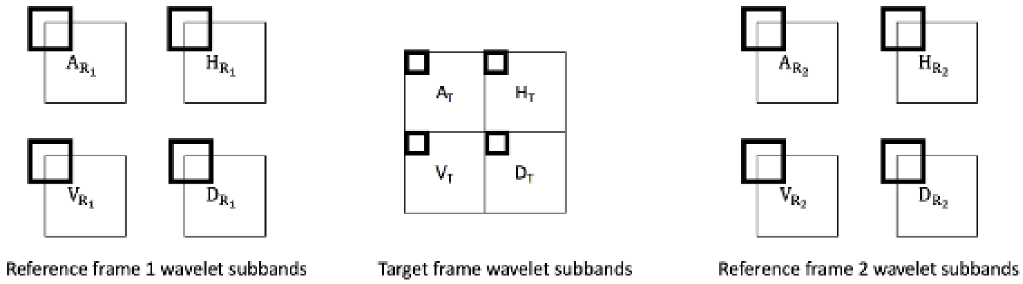


Figure 4. Bi-directional motion estimation in the wavelet domain

Figure 5. Example CIF video frames used in experiments (Bridge, Bus, and Container respectively Images courtesy of [15])

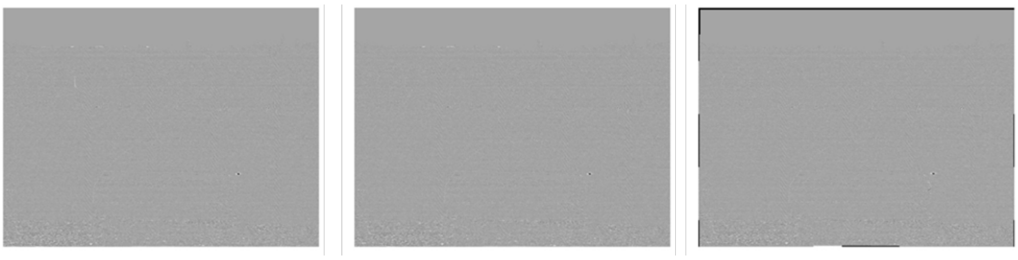
Table 1. *Comparison of bi-directional methods based on MSE and MAD.*

Video sequence	Bi-directional Spatial (Bilinear)		Bi-directional Spatial (Bicubic)		Bi-directional Wavelet (Proposed)	
	MSE	MAD	MSE	MAD	MSE	MAD
Bridge	10.15	2.12	10.05	2.1	10.01	2.12
Bus	104.84	5.25	104.63	5.24	92.9	4.62
Container	2.46	1.04	2.45	1.04	2.4	1.04

After motion vectors from each reference frame to the target one are obtained, motion compensation step is performed for the two sets of frames. Finally, after the IDWT step, the average of difference frames is calculated in the bi-directional version. Employing the bi-

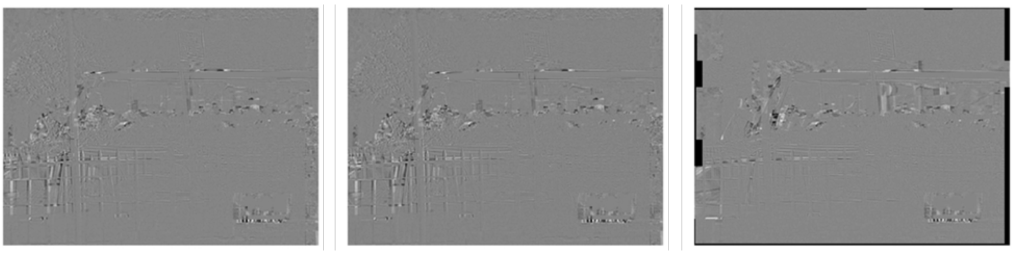
3. Experimental Results

This section draws the experimental results together obtained using the proposed method compared to other techniques. The results are based on CIF video sequences with resolution 352 × 288. Experiments are performed for block size of 16 × 16, subpixel accuracy is set to ¼ for the proposed and all compared methods, and comparisons based on mean squared error (MSE) and mean absolute differences (MAD) are provided for the acquired difference images. Compared methods include bi-directional spatial domain and uni-directional wavelet domain techniques. Bi-directional spatial domain methods are divided into two categories by the interpolation technique employed, namely for bilinear and bicubic interpolation. Average computational time taken by each method is also provided in Figure 5. Table 1 compares the proposed technique to the bi-directional spatial domain methods. Bi-directional spatial domain (bilinear) indicates that the motion estimation method employs bilinear interpolation in order to achieve subpixel accuracy; whereas, bi-directional spatial domain (bicubic) implies bicubic interpolation is utilized. For the proposed and all compared methods, same motion compensation technique is used to let comparisons be fair, employing bicubic interpolation. MSE and MAD results are obtained



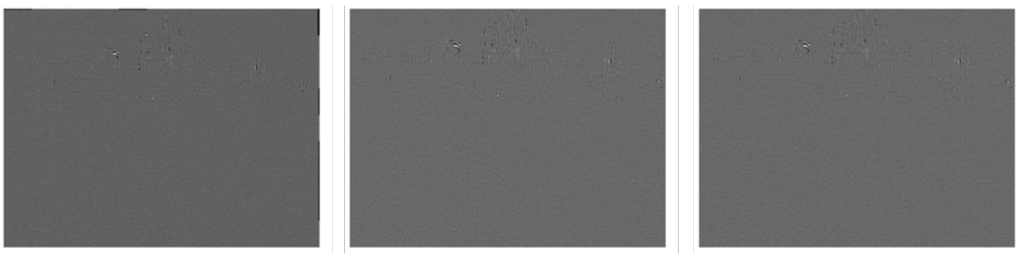
(a) Bilinear (b) Bicubic (c) Proposed

Figure 6. Difference images for Bridge sequence for bi-directional methods



(a) Bilinear (b) Bicubic (c) Proposed

Figure 7. Difference images for Bus sequence for bi-directional methods



(a) Bilinear (b) Bicubic (c) Proposed

Figure 8. Difference images for Container sequence for bi-directional methods

analyzed. All experiments are evaluated in MATLAB 2015b on a PC with 2.7 GHz Intel Core i5 CPU and 8 GM RAM. Example CIF video sequences used in the experiments are obtained from [15] and first frames of the sequences are for difference images after motion estimation and motion compensation steps are performed. Looking at Table 1, it is apparent that results demonstrate that the proposed bi-directional wavelet domain method has better results compared to spatial domain techniques. The average time taken in seconds by bi-directional spatial domain (bilinear) method is 1036.2, bi-directional spatial domain (bicubic) method is 1580.2, and the proposed technique is 359.82.

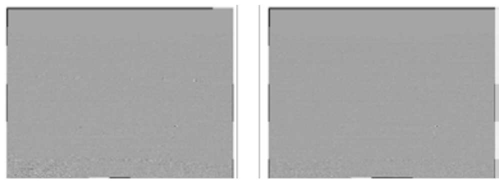
Table 2. *Comparison of wavelet domain uni-directional method with proposed technique.*

Video sequence	Uni-directional Wavelet [2]		Bi-directional Wavelet (Proposed)	
	MSE	MAD	MSE	MAD
Bridge	10.15	2.12	10.01	2.12
Bus	104.84	5.25	92.9	4.62
Container	2.46	1.04	2.4	1.04

Figures 6-8 demonstrate the difference images obtained for Bridge, Bus, and Container video sequences, respectively. As can be seen in the given figures, the proposed technique suffers from boundary problems. Therefore, MSE and MAD evaluations for all methods are carried out after removing the borders of the difference images. One can see in the provided figures (especially for the Bus sequence) that the difference images acquired after the motion compensation step for the proposed technique, indeed, have lower energy compared to the other techniques.

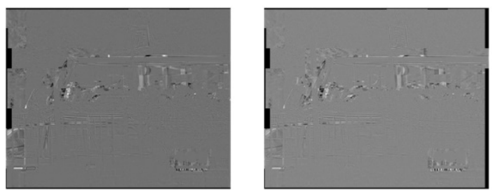
Table 2 shows the comparison based on MSE and MAD values between uni-directional wavelet domain motion estimation technique [2] and the proposed bi-directional wavelet domain technique. Again, the table indicates the that the proposed method outperforms the compared ones. The average time taken in seconds by uni-directional wavelet domain method [2] is evaluated as 178.12; whereas, the proposed technique is determined as 359.82. Since the motion estimation step is performed for both forward and backward directions in the bi-directional method, the average time of the proposed technique is expected to be higher. The results in Table 2 proves that even though the proposed technique takes longer time to run, it compensates this disadvantage by lowering the MSE and MAD results.

Figures 9-11 display the difference images achieved by uni-directional [2] and bi-directional (proposed) wavelet domain methods. Again, as shown in the figures, the difference images have lower energy compared to the other techniques, which indicates lower bitrate is required to encode these difference images in a video coding scheme.



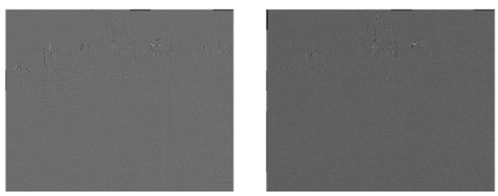
(a) Uni-directional [2] (b) Proposed

Figure 9. Difference images for Bridge sequence with wavelet domain methods



(a) Uni-directional [2] (b) Proposed

Figure 10. Difference images for Bus sequence with wavelet domain methods



(a) Uni-directional [2] (b) Proposed

Figure 11. Difference images for Container sequence with wavelet domain methods

4. Discussion and Conclusion

In this paper, we propose a wavelet-domain bi-directional motion estimation algorithm based on block matching which is employed in motion compensation. The method is performed in the wavelet domain taking advantage of the relationship between wavelet subbands. Experimental results confirm that the proposed bi-directional method reduces the energy in the difference images to be encoded. The proposed method can be employed in a scalable video coding framework in order to demonstrate the effective usage of wavelet-based motion estimation/compensation in reducing the bitrate while keeping the quality of the reconstructed video sequences. Another future research area in which the proposed method can be employed is to select block sizes automatically where the computational cost can further be decreased, and bitrate can also be reduced to lower amounts.

References

[1]. Bao, W., Lai, W.S., Zhang, X., Gao, Z., Yang, M.H. 2018. MEMC-Net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. arXiv preprint arXiv:1810.08768.

[2]. Aydin, V.A., Foroosh, F. 2017. Motion compensation using critically sampled dwt subbands for low-bitrate video coding. IEEE International Conference on Image Processing (ICIP), pp. 21-25. DOI: 10.1109/ICIP.2017.8296235

[3]. Jain, M., Jegou H., Bouthemy, P. Better exploiting motion for better action recognition. 2013. IEEE conference on computer vision and pattern recognition, pp. 2555-2562. DOI: 10.1109/CVPR.2013.330

[4]. Kesner, S. B., Howe, R.D. Position control of motion compensation cardiac catheters. 2011. IEEE Transactions on Robotics 27, no. 6: 1045-1055. DOI: 10.1109/TRO.2011.2160467

[5]. Chen, Y.C., Li G., Zhang Q., Zhang Q.J., Xia, X.G. Motion compensation for airborne SAR via parametric sparse representation. 2016. IEEE Transactions on Geoscience and Remote Sensing 55, no. 1: 551-562. DOI: 10.1109/TGRS.2016.2611522

[6]. Zhu, S., Ma, K.K. A new diamond search algorithm for fast block-matching motion estimation. 2000. IEEE transactions on Image Processing 9, no. 2: 287-290. DOI: 10.1109/83.821744

[7]. Lam, C.W., Po, L.M., Cheung, C.H. A new cross-diamond search algorithm for fast block matching motion estimation. 2003. IEEE International Conference on Neural Networks and Signal Processing. vol. 2, pp. 1262-1265. DOI: 10.1109/ICNNSP.2003.1281100

[8]. Luo, J., Konofagou, E.E. A fast normalized cross-correlation calculation method for motion estimation. 2010. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 57, no. 6: 1347-1357. DOI: 10.1109/TUFFC.2010.1554

[9]. Werlberger, M., Pock, T., Bischof, H.. Motion estimation with non-local total variation regularization. 2010. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2464-2471. DOI: 10.1109/CVPR.2010.5539945

[10]. Cuevas, E., Zaldívar, D, Pérez-Cisneros, M., Sossa, H., Osuna, V. Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC). 2013. Applied Soft Computing 13, no. 6: 3047-3059. DOI: 10.1016/j.asoc.2012.09.020

[11]. Pan, Z., Lei, J., Zhang, Y., Sun,X., Kwong, S. Fast motion estimation based on content property for low-complexity H. 265/HEVC encoder. 2016. IEEE Transactions on Broadcasting 62, no. 3: 675-684. DOI: 10.1109/TBC.2016.2580920

[12]. Nakaya, Y., Harashima, H. Motion compensation based on spatial transformations. 1994. IEEE Transactions on circuits and systems for video technology 4, no. 3: 339-356. DOI: 10.1109/76.305878

[13]. Fan, Y.C., Wu, S.F., Lin, B.L. Three-dimensional depth map motion estimation and compensation for 3D video compression. 2011. IEEE Transactions on Magnetics 47, no. 3: 691-695. DOI: 10.1109/TMAG.2011.2112641

[14]. Wang, D., Vincent, A., Blanchfield, P., Klepko, R. Motion-compensated frame rate up-conversion—Part II: New algorithms for frame interpolation. 2010. IEEE Transactions on Broadcasting 56, no. 2: 142-149. DOI:10.1109/TBC.2010.2043895

[15]. http://trace.kom.aau.dk/yuv/