Spectral-Based Semi-automatic Segmentation of Video Object Using Constraint Estimation

RURI SUKO BASUKI
Computer Science
Dian Nuswantoro University
Jl. Imam Bonjol 205 - 207, Semarang, 50131
INDONESIA
ruri.basuki@dsn.dinus.ac.id

Abstract: - Motion vector is acquired by calculating the conclusive dissimilarity of blocks in the current frame and search block on next frames. Block Matching Algorithm (BMA) is employed to obtain the motion vector value. The result is added to pixel coordinates in current frame correlated with user constraints. Next, the object segmentation process is performed by matting techniques, after constraint scribble automatically occupies the next frame. However, matte extraction reveals a high error rate value after evaluation of segmentation results, caused by motion vector calculation which is as the driving of constraint parameter conducted in entire block. As result, position of pixel scribble is extending and far from object expected when motion vector value is applied. To solve the problem, calculation of motion vector performance is only on the block directly correlated to pixels scribble. This research presents an approach estimating constraint on semi-automatic segmentation of video object and the aims is to estimate the constraint in driving position of pixels scribble, where in the object extraction in a single frame is done with image matting, while the temporal domain motion estimation algorithm performed by Exhaustive Search of the BMA, but it is not robust algorithms for motion estimation on the label (scribble). Thus, in this study improved with the ES algorithms are developing and applying adaptive block SAD (Sum of Absolute Difference) to determine the distance vector. At final, the motion vector value is used to move the label from current frame to next frame. The result reveals accuracy improvement of 71.19%.

Key-Words: - object segmentation, constraint estimation, motion vector prediction

1 Introduction

Multimedia technologies for storage media such as CD/DVD and video streaming disseminated by internet become very popular. For this reason, the demand of multimedia technology, particularly to video coding is increasing dramatically. Previously, the video coding standard defined by H.264/MPEG-4 AVC [1] technologies has been widespread on the HD (High Definition) TV signal which is transmitted by fixed line (such as cable or fiber optic), satellite, and terrestrial transmission. In addition, the H.264/MPEG-4 AVC standard also give contribution in multimedia and network application such as video editing, streaming in mobile network, security, etc.

However, the rapid popularity of high quality videos well as the advent of higher resolution technology (4kx2k or 8k x 4k) encourage demand of video coding which is more efficient than the performance of H.264/MPEG-4 AVC standard [2][3]. Increased use of mobile-based applications and tablet PCs as well as transmission needs to services video on demand is a challenge on the networking field. Moreover, the demand for services on the high-resolution video is attended by multi stereo camera. The display is also increasing rapidly, so that the standard H.264/MPEG-4 AVC is insufficient to meet the user needs.

Recently, video coding standards defined by High Efficiency Video Coding (HEVC)offershigh technology of video compression allowing the development of new various types of content-based applications which are focused on two main keys; resolution enhancement and increased use of parallel processing architecture [3][4]. Before, the MPEG-4 AVC / H.264 standard using the macro block (MB) as the unit of fundamental process with the size 16x16. However HEVC can support larger sizes of the basic processing unit, starting from 8x8 up to 64x64 in size. Application of single instruction multiple data (SIMD) [5] on HEVC is proven to improve the time efficiency up to 80%. Whilst the use of wave-front parallel processing (WPP) [6] for HEVC encoder and decoder has reached parallel speeds up to factor 3.

To achieve more function and benefits, it is required video processing based on object. The process is significant in computer vision applications for example feature extraction, object extraction, object annotation etc. Unfortunately, illposed problem [7] becomes the biggest issue in the extracting objects process in video sequences since semantic information is implicitly provided in the video data. Thus, it is only the human vision which is aware the semantic information in the video object. In other cases, manual segmentation in which a semantic object defined manually by human assistance performs the addition of special effects on film production, frequently by considering the video context. This causes ineffective processing procedures so that it cannot be applied in a video having a large data.

In the last decades, video object segmentation researches have been performed. In general, the algorithms applied in the study are divided into two types, semi-automatic segmentation [8][9][10] and automatic segmentation [11][12][13]. In automatic segmentation, specific information such as texture, movement and color used as primary characteristic of the scene [7]. It has also a difficulty for automatic separating object in its method since it has no semantic information. Therefore, the today research has no guarantee of satisfaction for the results of the automated semantic segmentation [10].

From the statement above, it is clear that semi-automatic segmentation methods become the problem solution for manual and automatic segmentation. Semi-automatic method involves user interaction providing semantic information in the form of scribbles in the initial stages of the segmentation process, for segmented object in accordance to the users' wishes. It is performed for creating "key frame", which is applied as a reference for the extraction process in the next frames. In the creation of key frames, user intervention is directly performed, necessarily to provide a constraint as initialization in the segmentation process. In this paper, the constraints are defined in the form of scribbles (i.e. white represents the foreground and black for the background). Furthermore, matting approach is applied in the object separation process [14][15][8].

The segmentation process in next frames follows the temporal transformation mechanism after the creation of key frames. It assumes that the movement of the object is coherence so that the movement occurring between the current and next frames is smooth and not abrupt. Therefore, the motion estimation approach is applied to estimate the movement and determine the motion vector which is describing a 2D transformation from one frame to another. The Block Matching Algorithm (BMA) is used to estimate the motion vector in which the results are applied to estimate the motion among frames (the value of motion vectors to drive the scribbles from one frame to another).

In this paper, one of approaches in block matching is used to estimate the motion vector [16][17][18], which is exhaustive or full search. However, when the value of the motion vector was used to move constraint, the scribble pixels visible spread out and away from the object. Thus, the scribble is uncorrelated with the pixels on the desired object. Consequently, the extraction is declined significantly. It occurs because the motion vector value of each block may be dissimilar, so that, when it is added to coordinate's pixels constraint on current frame, the scribble will spread on next frame. We assume that the accuracy of the constraint movement is only affected by a block which correlates to the scribble. Therefore, the motion vector prediction is conducted on the block which correlates to scribble only. The size of predicted block between 1 - ∞, and the absolute difference is used as a cost function to get the motion vector. By applying this algorithm, the results are proven to have a significant improvement for the extraction quality

This paper defines the writing structure as follows: section 2 explains the previous study which becomes the reference of a system manufacture and also describes the process of how features are constructed. Section 3 describes the framework of the segmentation process. The detailed discussion as well as evaluation of development as explained in section 4. And last section illustratesthe conclusion of this study and the plan of future studies.

2 Related Work

Most of existing study for the stages of video object segmentation was consisted by extracting object on spatial-domain and tracking object on the temporal domain (patio-temporal) [19]. In several approaches for object extraction, they were applying matting techniques with tramp images [20][21][22] as companion input. Tri map was used as a label on each pixel representing of foreground, background and unknown. It was intended to solve the problems of unknown pixels as in the composition equation (1). Furthermore, the separation of the object in next frames was possible to be performed by applying an automatic tracking since the correlation of two frames in the video sequence was visually resembled. This can be solved by applying motion estimation in which the movement direction of the object could be calculated by block matching [16][17][18]. It is inaccurate when the block matching calculation for driving the vector with exhaustive search [16]was applied to drive scribble from current frame tonext frame. Thus, the scribble movement with exhaustive search was performed by creating blocks around its area [23]. To simplify the computation complexity, estimation process was performed in color room of HSV (Hue, Saturation and Value).

2.1 Matte Extraction

The alpha channel was mathematically introduced by Porter and Duff [24] which was firstly applied to produce the linear interpolation between the object and background colors in the image. Generally, in the matting algorithm [14][15][8], it was assumed that each pixel I_i was linear combination between the foreground F_i and background colors B_i denoted as follows.

$$I_i = \alpha_i F_i + (1-\alpha_i) B_i \quad \text{with } 0 \le \alpha \le 1 \qquad (1)$$

Furthermore, it assumed that every pixel was a convex combination of the K layer of the image $F^1, \ldots, F^K$ represented as:

$$I_i = \sum_{k=1}^{K} \alpha_i^k F_i^k \qquad (2)$$

For the matte extraction [14], user-constraints in matte were supplied by the scribble-based GUI or trimap. Herein, the user was using the foreground pencil (white scribble) showing the pixel foreground represented by $\alpha = 1$, while the background pencil (black scribble) was indicating the background pixels defined by $\alpha = 0$. Matte extraction matching to the user constraints was resolved by:

$$\alpha = \arg\min \alpha^T L \alpha + \lambda(\alpha^T - b_S^T) D_S (\alpha - b_S) \qquad (3)$$

Where L is the matrix of size $N\ x\ N$, λ is some of the large numbers, and D_S Is a diagonal matrix in which diagonal elements are worth one for the pixel constraints and zero for the other pixels. Where, b_S is a vector comprising an alpha value which are specifically used for the pixel constraints and zero for the other pixels. Above, in the quadratic cost function in the alpha, the global minimum was found by differentiating in equation (3) and arranging derivatives up to zero by completing the following sparse linear systems that is:

$$(L + \lambda D_S)\alpha = \lambda b_S \qquad (4)$$

2.2 Block Matching

Temporal domain was the concern in this study since the pixels movements in video sequences affected the segmentation results on next frames. In order to extract the object on next frames, it assumed that matting technique should be conducted (such as extraction process on the "key frame"), so that the quality of the extracted object in next frames was strongly influenced by the constraint movement.

It was assumed that the movement of the object in two sequential frames was possible to be predicted by applying the motion estimation and was calculated by applying the motion vector. At this point, the motion vector computation was performed by BMA [16]. It drew the reason that a formation pattern of object relationships in next frames occurred between the object and the background of current frame of video sequences. Block-matching in the current frame was assumed to be a matrix, divided into a number of 'macro blocks' which was, later on, compared to ‘search block' of the previous frame.

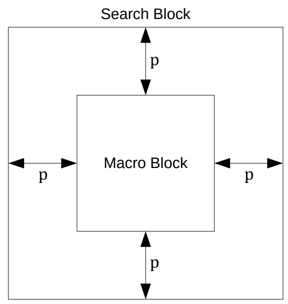

Fig. 1 Block matching structure

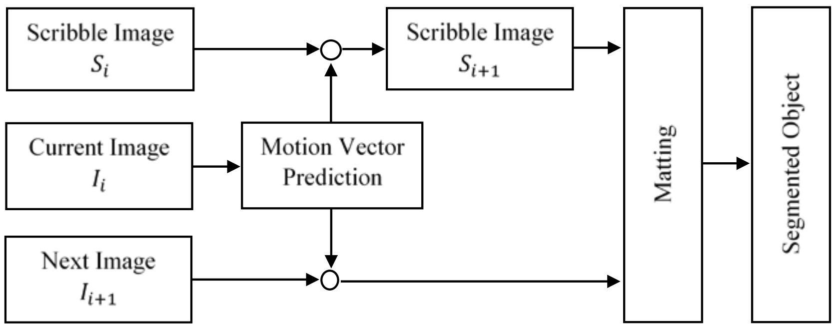

Fig. 2 Video object segmentation diagram

Table 1.The comparison result of the estimated constraint

Sequences	Block matching			Our approach		
	Min error	Mean errror	Max error	Min error	Mean errror	Max error
Artem	558.48	2,433.60	16,680.00	510.23	1,050.40	2,826.93
Alex	449.30	2,148.53	19,099.18	423.92	1,644.22	5,605.54
Vitaliy	633.80	6,422.57	14,878.52	549.99	2,565.87	9,668.96
Dmitriy	-	3,029.50	10,166.30	-	872.94	3,099.16

The invention process, a "macro block" having the highest similarity to the "search block" was conducted by applying tracking matrix from left to right and from top to the bottom. The constraint was given to the search block to 'p' pixels here in after referred to as search parameters for achieving an optimal search. The large value of 'p' in computation was most expensive in process of motion estimation.

Macro block was completed in squares measuring 8 x 8pixels on each side, while the search boundary coordinates (left, right, top and bottom) was represented by search parameter 'p' size of 7 pixels [16] (illustrated in fig. 1). The process of matching blocks was conductedseparately, based on the result of cost function. In this experiment, exhaustive search algorithm was applied to obtain a motion vector by performing block matching process on the entire frames. The movement direction was following the motion vector obtained from the matching blocks between the macro and search block, whereby the highest Peak Signal Noise to Ratio (PSNR) value was considered as blocks having similarity to the other blocks. Thus, the movement direction in these conditions was considered as the motion vectors value. PSNR denoted as follows:

$$PSNR = 10\log_{10}\left[\frac{(P)^2}{MSE}\right] \tag{5}$$

With P is the highest pixel value of the processed frame, in which

$$MSE = \frac{1}{B^2}\sum_{i=0}^{B-1}\sum_{j=0}^{B-1}\left(CF_{ij} - RF_{ij}\right)^2 \tag{6}$$

Whereby B is the block size, CF_{ij} and RF_{ij} (CF_{ij} = current frame, RF_{ij} = reference frame) is the block compared.

3 Video Object Segmentation System

In this section, we describe framework of a semi-automatic video object segmentation depicted in Fig. 2.In the semi-automatic segmentation, user constraint is required as initialization to define the area that represents the foreground and background areas. Scribble image is a user constraint in the form scrawl of the hand (white color to define the foreground area, and black to define the background area).

Since it is closely related to the pixels, the placement of scribble deeply affects the quality of the separated objects. Meanwhile, in the video data, it may not probably be scribble given by the user in all frames (in video sequences, the frame is a still image). We assume that the provision of scribble on the next image (S_{i+1}) can be performed automatically by predicting the distance of vector between the current image (I_i) and the next image (I_{i+1}), so that the result value of motion vector calculation can be applied to move the scribble position from (S_i) to (S_{i+1}). Furthermore, the object segmentation process on the next image (I_{i+1}) is conducted by matting techniques.

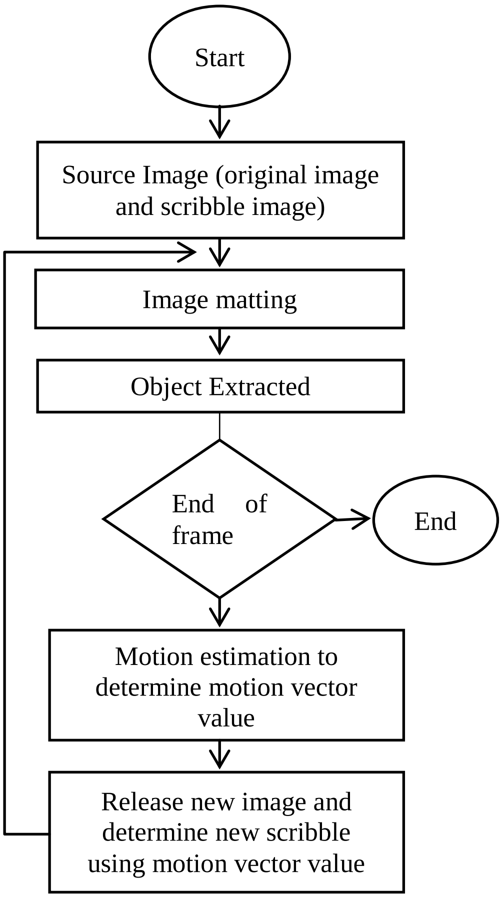

Fig 3. Matting technicque flow chart

4 Experiment and Evaluation

4.1 Motion Vector Prediction

The results of the extraction on the next frame were highly dependent on a scribble which should be placed. However, when the scene moving from current frame into next frame, there some pixel value which were moved. This caused an effect that the scribble was not associated with the expected pixels, so the quality of matte was also decreased (illustration in Fig 3). To solve this problem, we estimated the motion vector value in addition to the result attached at the coordinate’s position of earlier frame object.

In this experiment, reference frame is divided into block of size 8x8 pixels (called macro blocks). In order to find the minimum value, equation (6) is conducted by iteration ranging from 'p-7’ up to 'p + 7', from left to right and from top to bottom. The cost function computed block matching process concerning the macro block and the search block (5), and applying the highest PSNR value as the value of motion vector. New scribble in the next frame was acquired by adding the value of the motion vector at the position coordinates of pixels which had value 0 or 1 of the previous frame denoted as follows:

$$CF(i+J_{mx}, j+J_{my}) = PF(i,j,D)$$

With

$$J_{mx} = (0....MV_x) \text{ And } J_{my} = (0....MV_y) \quad (7)$$

Wherein i, j the coordinate position of frame is, D is the color channel, CF is the current frame, PF is the previous frame, MV_x MV_y and is the motion vector value.

From the 100 frames trialled, key frames were refreshed in a few scenes consistently. We performed the evaluation process over matte extraction results in each frame by estimating value of absolute difference between matte extracted with ground truth (9).

Block calculation performed in the entire frame became the weakness of this algorithm, since the value of motion vector in adjacent block was likely different. When the motion vector value was added, scribble position of the current frame would overlap on the next frame. For this reason, we assume that the extraction quality of temporal segment was only affected by a block of pixels which correlated to the scribble. Therefore, the motion estimation was only conducted in areas predicted to be moving, in order to increase the accuracy of object movement and to decrease the cost of computing. It assumed that every scribble defined by the user was an area representing the object movement, so that the motion vector was calculated only on that area by using cost function as follow.

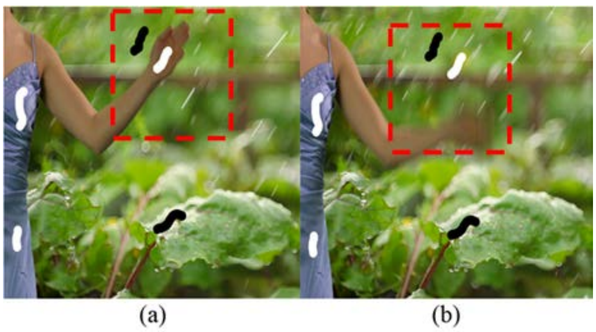

Fig. 3 Illustration of: (a) suitable constraint, (b) error constraint

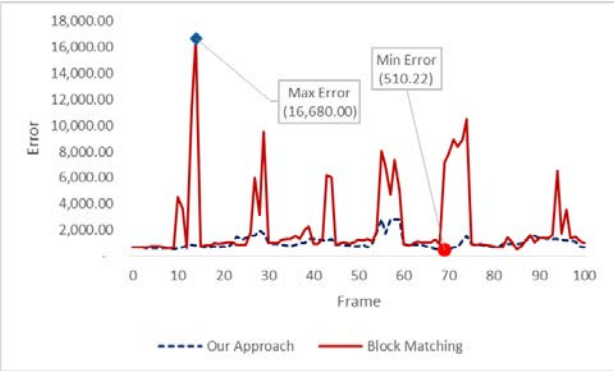

Fig. 4 Error value of *Artem* video dataset

$$f(u,v)=\sum_{w_{xy}}\left(CF\left(x+u,y+x\right)-PF\left(x+u,y+x\right)\right) \quad (8)$$

Whereby u, v is motion vector, w is the weight of value 1 up to ∞, and x, y is the coordinate position of pixel. In this way proved that the accuracy of the matte extraction increased significantly (see table 1).

4.2 Matting

In the next frame, we executed the object segmentation by applying matting techniques. The extraction process was applying the input image (Fig. 8) as well as scribble image (Fig. 8b. input image with user scribble as a constraint) as companion.

The users used type of brushes to determine constraint on the area which contained the mixed pixels. Black scribble indicated the background pixels (α = 0) and white scribble indicated the foreground pixels (α = 1) (as illustrated in Fig. 8b). Simple constraint explicitly could be applied to determine the value of F and B which were correlated to the scribble, so the constraint on α can could be calculated directly from equation (1). In extracting alpha matte which appropriate to user constraints, it could be solved by equation (3). Equation (3), made it possible to global minimum value found which was different. In order to arrange a derivative up to zero, it was conducted by following an equation based on sparse linear system (4).

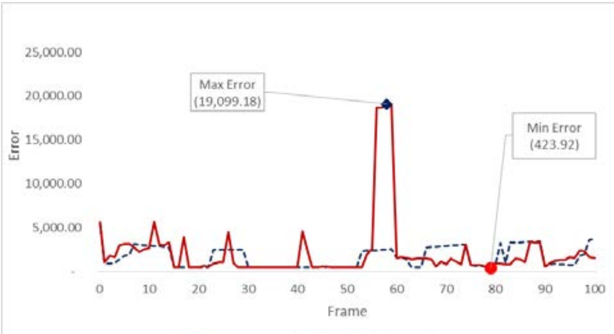

Fig. 5Error value of *Alex* video dataset

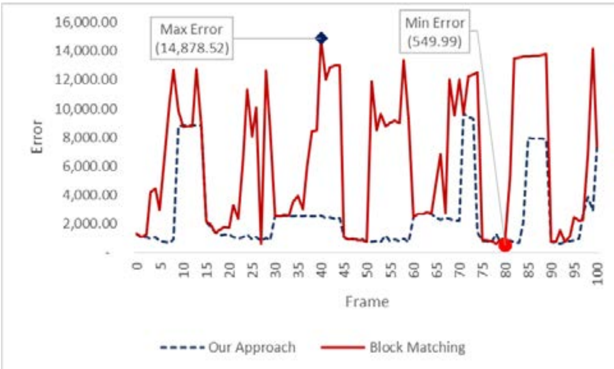

Fig. 6 Error value of *Vitaliy* video dataset

In order to evaluate the extraction quality, we calculated the value of distinction between the matte extractions produced by system (shown in Fig. 8d) with matte reference (obtainedfrom video matting dataset as described in Fig. 8c) in the following formula:

$$error=\frac{abs(MR-ME)\wedge 2}{M\times N} \quad (9)$$

With MR is the matte reference, ME is matte extraction results of an algorithm suggested. Whilst $M \times N$ is the image size.

4.3 Evaluation

In our experiments, we evaluated the proposed algorithm applied on the video matting dataset: *Artem*, *Alex*, *Vitally*, and *Dmitri* in 100 frames for each. The Dataset isframe video asdata input andground truth as reference data to calculate the accuracy level of segmentation results.In order to determine the scribble direction; we predicted the motion vector on the next frame, whereas the object segmentation was conducted by matting techniques.

Fig. 5 shows the evaluation results of the "*Artem*". The video data has the characteristics of a rotating object with a stationary background. The constraint movement estimation was performed by motion vector prediction. A block matching algorithm was used [16] and for the comparison, we applied in difference absolute as a cost function (8). Results of the comparison were calculated by the equation (9), including min error, max error, and mean error. By applying our approach, the average error was decreasing by 56.84%.

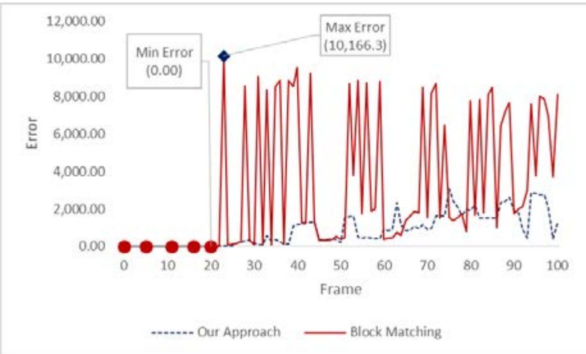

Fig. 7 Error value of *Dimitry* video dataset

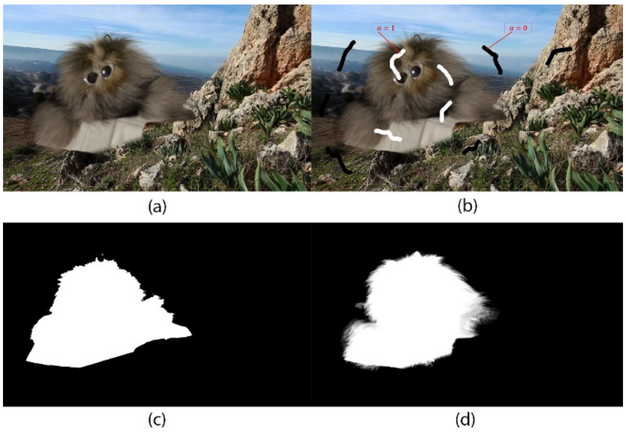

Fig. 8 Matte extraction process (a) Frame input, (b) scribble/label, (c) matte reference, (d) extraction result

The same evaluation was also performed on the "*Alex*", "*Vitally*", and "*Dimitriy*". For the test sequences "Alex" having the characteristics of rotating objects and a simple move (left and right) as well as the background move. It was resulted in an error decrease about 23.47%.As for the test sequences of "*Vitally*" which had a rotating feature object and the parts of background having similar color to the object which produce an average error decrease of 60.05%.Finally, the test sequences "*Dmitry*" which had a feature object rotates and the camera moving produced an average error decline up to 71.19% (described in Table. 1).

The experiment result in Table 1 showed that the algorithm suggested was able to improve the accuracy (lower down the pixel error level of segmentation result) compared to Exhaustive Search in Block Matching Algorithm [16] in the effort to drive scribble fromvideo matting dataset in RGB space.

5 Conclusion and Future Plan

In this paper, we proposed a semi-automatic video object segmentation using constraint estimation. First, the pixels constraint was defined by providing a scribble on the area representing object and background. Since the constraint moves from the current frame to the next frame, the pixels constraint direction estimation is calculated by predicting the motion vector value. Initially, we applied BMA to estimate value of the motion vector. However, since the calculation of block matching performed in entire block, the value of motion vector in adjacent blocks had different values. As result, when the motion vector value is used as parameter to drive the constraint, pixels scribble on next frame were spread out and away from the object targeted, so it was dramatically declined the quality of the object segmented.

To solve this problem, we calculated the value of motion vector in block correlated to pixels constraint only. Motion vector calculation was conducted by an absolute difference, i.e. by comparing the block in the current frame and the search block on the next frame (by iteration ranging from 'p-7' up to 'p + 7', from left to right and from top to bottom).The smallest value of absolute difference was considered to be the motion vector values and results were applied to drive the constraint. This algorithm was applied to the test sequences of: *Artem, Alex, Vitally* and *Dmitry* in our experiments. After the evaluation, this approach indicated the error decrease up to 71.19% in the implementation process of datasethaving similar scene.

Even though the proposed algorithm was proven for its function to improve the accuracy, pixels constraint on some final frames, before they were refreshed, were seemed spread out and far from the object. Therefore, future work is aimed to incorporate the Self Organized Map (SOM) method to increase the robustness of the algorithm which developed. In addition to the deployment of pixels constraint problems, experiment on a dataset having a different scene in video sequence became assignment in the next study.

References:

[1] H. Tsai-Tsung and P. Yu-Nan, "High

Efficiency Architecture Design of Real-Time QFHD for H.264/AVC Fast Block Motion Estimation," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1446–1658, 2011.

[2] A. Yong-Jo, H. Tae-Jin, S. Dong-Gyu, and H. Woo-Jin, "Implementation of Fast HEVC Encoder Based on SIMD and Data-level Parallelism," *EURASIP J. Image Video Process.*, vol. 16, pp. 1–19, 2014.

[3] O. Jens-Rainer, G. J. Sullivan, H. Schwarz, T. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, 2013.

[4] G. Sullivan, O. Jens-Rainer, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[5] K. Chen, Y. Duan, L. Yan, J. Sun, and Z. Guo, "Efficient SIMD Optimization of HEVC Encoder Over X86 Processors," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012.

[6] G. Clare, F. Henry, and S. Pateux, "Wavefront parallel processing for HEVC encoding and decoding," *ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-F274*, 2011.

[7] A. Bovic, *The Hand Book of Image and Video Processing*, 2nd ed. San Diego: Academic Press, 2000.

[8] R. Basuki, M. Soeleman, M. Hariadi, M. Purnomo, A. Yogananti, and R. Pramunendar, "Spectral-Based Video Object Segmentation using Alpha Matting and Background Subtraction," in *IIEEJ the 4th International Workshop on Image Electronics and Visual Computing*, 2014.

[9] C. Toklu, A. Tekalp, and A. Erdem, "Semi-Automatic Video Object Segmentation in the Presence of Occlusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 624–629, 2000.

[10] M. Hariadi, H. Loy, and T. AOKI, "Semi-Automatic Video Object Segmentation using LVQ with Color and Spatial Features," *IEICE Trans. Inf. Syst.*, vol. E88D, no. 7, pp. 1553–1560, 2005.

[11] Y. Tsaig and A. Averbuch, "Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 597–612, 2002.

[12] H. Li and K. Ngan, "Automatic Video Segmentation and Tracking for Content-Based Applications," *Adv. Vis. Content Anal. Adapt. Multimed. Commun.*, pp. 27–33, 2007.

[13] T. Meier and K. Ngan, "Automatic Segmentation of Moving Objects for Video Object Plane Generation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 525–538, 1998.

[14] A. Levin, D. Lischinski, and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 1–15, 2008.

[15] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral Matting," *Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1699–1712, 2008.

[16] B. Arroh, "Block Matching for Motion Estimation," *IEEE DIP 6620 Spring 2004 Final Project Paper*, 2004.

[17] C. Yong-Sheng, H. Yi-Ping, and F. Chiou-Shann, "Fast Block Matching Algorithm Based on the Winner-Update Strategy," *IEEE Trans. Image Process.*, vol. 10, no. 8, 2001.

[18] H. Seung-Ryong, T. Yamasaki, and K. Aizawa, "Time-Varying Mesh Compression Using an Extended Block Matching Algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1506–1518, 2007.

[19] H. Jiang, G. Zhang, H. Wang, and H. Bao, "Spatio-Temporal Video Segmentation of Static Scenes and Its Applications," *IEEE Trans. Multimed.*, vol. 17, no. 1, pp. 3–15, 2015.

[20] N. Apostoloff and A. Fitzgibbon, "Bayesian Video Matting Using Learnt Image Priors," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2004.

[21] Y. Chuang, A. Agarwala, B. Curless, D. Salesin, and R. Szeliski, "Video Matting of Complex Scenes," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 243–248, 2002.

[22] J. Sun, J. Jia, C. Tang, and H. Shum, "Poisson Matting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 315–321, 2004.

[23] R. S. Basuki, M. Hariadi, E. M. Yuniarno, and M. H. Purnomo, "Spectral-Based Temporal-Constraint Estimation for Semi-Automatic Video Object Segmentation," *Int. Rev. Comput. Softw.*, vol. 10, no. 9, pp. 959–965, 2015.

[24] T. Porter and T. Duff, "Compositing digital images," *Comput. Graph. (ACM).*, vol. 18, no. 3, pp. 253–259, 1984.