Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi

**Vol. 7, No. 2, July 2013** **ISSN 0216 – 0544**

# QUALITY IMPROVEMENT OF OBJECT EXTRACTION FOR KEYFRAME DEVELOPMENT BASED ON CLOSED-FORM SOLUTION USING FUZZY C-MEANS AND DCT-2D

**[a]Ruri Suko Basuki, [b]Mochamad Hariadi, [c]Mauridhi Hery Purnomo**
[a,b,c]Faculty of Industrial Technology, Dept. of Electrical Engineering Institut Teknologi Sepuluh Nopember, Kampus ITS Keputih, Sukolilo, Surabaya, Jawa Timur, Indonesia
[a]Faculty of Computer Science, Dian Nuswantoro University
Jalan Imam Bonjol, Semarang, Indonesia
E-mail: [a]rurisb@research.dinus.ac.id

## Abstrak

Penelitian ini bertujuan untuk meningkatkan kualitas ekstraksi obyek pada citra tunggal hasil pemecahan frame dari video sekuensial yang terkompresi. Kualitas hasil ekstraksi obyek dengan algoritma *closed-form solution* menurun karena adanya beberapa perubahan nilai intensitas pada *channel* RGB. Sehingga di sekitar batas tepi obyek hasil ekstraksi terlihat kasar baik secara visual maupun hasil pengukuran dengan *Mean Squared Error* (MSE) antara obyek hasil ekstraksi dengan *ground truth*. Untuk meningkatkan kualitas hasil ekstraksi objek, nilai *threshold* pada *unknown region* ditentukan melalui *adaptive threshold* yang diperoleh dengan mengaplikasikan algoritma *Fuzzy C-Means* (FCM). Pemilihan algoritma FCM karena dalam penelitian sebelumnya algoritma ini menunjukkan hasil yang lebih *robust* dibandingkan algoritma Otsu untuk mendapatkan nilai *threshold* yang optimal. Sedangkan untuk menghaluskan obyek di sekitar daerah batas tepi digunakan filter *Discrete Cosine Transform* (DCT) – 2D. Dari 10 obyek yang digunakan dan dievaluasi dengan MSE menunjukkan peningkatan rata-rata sebesar 31.55%. Namun pendekatan ini tidak begitu *robust* pada citra yang memiliki kemiripan warna. Penggabungan pendekatan ini dengan optimasi *cost function* dalam *alpha region* pada basis *spectrum* diharapkan mampu meningkatkan kinerja algoritma ekstraksi obyek pada penelitian selanjutnya.

Kata kunci: *Closed-form Solution*, Algoritma *Fuzzy C-Means*, *Discrete Cosine Transform-2D.*

## *Abstract*

*The research is aimed to improve the quality of the extraction of the object in a single image resulted from frame's fragmentation of sequential compressed video. The quality of the extracted objects with closed-form solution algorithm decreased due to some changes in the intensity values on the RGB channel. Thus, the extraction result around the boundary edges of objects visually seemed to be rough and when it was measured with the Mean Squared Error (MSE) beween the object extraction results with ground truth. To improve the quality of the extracted object, the threshold value on unknown region was determined by adaptive threshold obtained by applying the Fuzzy C-Means algorithm (FCM). FCM algorithm is chosen since in the previous research this algorithm gives more robust results than Otsu algorithm to obtain the optimal threshold value. Meanwhile, to eliminate noise around the border area, this research applies Discrete Cosine Transform (DCT) - 2D filters. The result of 10 objects used and evaluated with the MSE showed an average increase of 31.55%. However, this approach is not so robust to images having similar color. Combination of this approach with optimization of the cost function on the alpha region based on spectrum is expected improving the performance of object extraction algorithm for the next research.*

*Key words: Closed-form Solution, Fuzzy C-Means Algorithm, Discrete Cosine Transform-2D.*

## INTRODUCTION

The emergence of digital television standard such as Digital Television (DTV) in America, Digital Video Broadcasting – Terrestrial (DVB-T) in Europe and Integrated Services Digital Broadcasting - Terrestrial (ISDB-T) in Japan rise the rapid demand of multimedia technology. The rapid increase of multimedia data exchange based on networking encourages the emerging of video coding standard.

Several methods of video coding standards such as H.261, H.263, MPEG-1 and MPEG-2 have been extensively used in DVD, digital television, video conferencing and some application in telemedicine [1]. The simple method in computation, however, has not been able to fulfill the users requirements.

Currently, the new model of video standard defined by MPEG-4 and MPEG-7 provides standard technology for representation and video data manipulation [2]. One of the important innovations in the MPEG-4 standard is the ability to manipulate object in image sequences. Scene description, a video or audio object that correlated with the technique of organizing objects in a scene can be encoded with this standard [3,4]. Meanwhile, the MPEG-7 standard supports indexed multimedia database and a structured meta data provides a rich media content in semantic view.

The function of the object-based technology can be used in computer vision applications such as object extraction, motion understanding, images recognition, and augmented reality. However, object extraction process on video data will be a difficult job, since it has no semantic information. Indeed, the process of object extraction is an ill-posed problem [5], whereby only human perception can understand of the semantic meaning of the video object. For instance, in the process of video editing, pullout of objects are often done by manual segmentation. Therefore, a video objects semantic can only be identified by human vision by considering video context. So it is not effective in applying the action to large volume video data.

In recent years, several algorithms have been developed to overcome the problems of video object segmentation. Based on user interaction, the technique can be classified into an automatic (unsupervised) and semi-automatic (supervised) category. An automatic object extraction doesn't require human intervention in the extraction process. It uses special characteristic of the scene or specific knowledge such as color, texture and movement. Therefore, it is difficult to do the semantics from the object segmentation, because it is possible to have some color, texture and movement [6]. For these reasons, a semi-automatic approach to extract the object (supervised) involving human intervention is proposed by combining human perception (manual segmentation) and automatic segmentation.

During the segmentation process on the semi-automatic object extraction method (supervised), human intervention is involved in several stages to provide a direct semantic information. When semantic of the video object is provided by user, the mechanism of object tracking is done by following the temporal transformation in subsequent frames. Generally, the tracking process are prone to error due to the degradation of intensity value of reference frame. Therefore, users need to make improvement on semantic information by refreshing the keyframe at specific locations in a video sequence.

In this paper, we propose the techniques to improve an object extraction quality for keyframe which can be applied in a semi-automatic video object extraction. Extraction process on the single image is performed by closed-form solution approach [7]. To distinguish the objects and background in the image (single frame), it is done by applying the interactive matting using a scribble technique as an interface [8]. This approach produces impressive results in the implementation in the raw image (uncompressed image). However, this method is not robust to handle the compressed image sincethe extraction result seems rough. It occurs due to the changes in the intensity values at each pixel, so that when the image is displayed in a histogram graph it turns towards the contrast. To improve the extraction quality, we propose the adaptive threshold for determination threhold value in the unknown area and remove the noise on extracted object. Furthermore, an accuration of the object extraction based on spectrum using closed-form solution is be improved.

The organization of this paper as follows : The second section explains the closed-form solution algorithms for basis system, and also describes the adaptive threshold determination and noise reduction quality improvement for object extraction. The third secgtion describes framework of the extraction system. The fourth section describes our experiments and also evaluate the performance of our algorithm. The fifth section provides conclussion and future work discussion of our paper.

## THEORITICAL BACKGROUND

### Matting Process

The object extraction in image and video is interesting object to be analyzed. Porter and Duff [7,9,10] introduce alpha channel used as tool to control linear interpolation of the foreground and background color. Then alpha channel is described as matting algorithm by assuming that every pixel $I_i$ in input image is linear combination of foreground $F_i$ and background $B_i$ color.

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

By compositing Equation (1), it is assumed that every pixel is a convex combination of layer $K$ image $F^1, \ldots, F^k$ as noted in Equation (2).

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

Vector $K$ from $\alpha^k$ is the matting component of image determining fractional contribution from each layer observed in each pixel.

### Spectral Analysis

Generally, spectral segmentation method is associated with image with $A$ matrix affinity size $N \times N$, which is assumed as $A_{(i,j)} = e^{-d_{ij}/\sigma^2}$ and $d_{ij}$ is the space among pixels (e.g. color and geodesic space), which is defined in Equation (3).

$$L = D - A \qquad (3)$$

where $D$ is matrix degree from graph that is shown in Equation (4).

$$G = (V, E)\, with V = n \qquad (4)$$

And $D$ is diagonal matrix as seen in Equation (5).

$$D_{(i,j)} = \sum_j A(i,j), \qquad (5)$$

$$\text{where}\, d_{i,j} = \begin{cases} \deg(v_i) & if\ i = j \\ 0 & Otherwise \end{cases}$$

$D_{(i,j)}$ is filled with degree information of each vertex (node) with $D$ for $G$ a s rectangular matrix size $n \times n$ that is described. Thus, $L$ is symmetric positive semi-definite matrix which the eigenvector catches many image structures. The affinity matrix $A$ is able to catch the information that an image consists of some different clusters or connected component. Subset $C$ in image pixel is the connected component of image $A_{(i,j)} = 0$ for each $(i,j)$ so $i \in C$ and $j \notin C$. Thus, there is no subset $C$ which is able to fulfill this property. If the indicator vector of component $C$ is noted as $m^C$, see Equation (6) thus

$$m_i^C = \begin{cases} 1 & i \in C \\ 0 & i \notin C \end{cases} \qquad (6)$$

$m^C$ represents 0-eigenvector (eigenvector with eigenvalue 0) from $L$. In the assumption that image consists of connected component $K$, $C_1, \ldots, C_K$ to $\{\mathbf{1}, \ldots, \boldsymbol{N}\} = \cup_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{C_k}$ with $C_k$ disjoint subset on pixel. The indicator vector $m^{C_1}, \ldots, m^{C_K}$ resulted from eigenvector calculation on $L$ is only reaching the rotation, since the rotation of matrix $R$ in size $K \times K$, and vector $[m^{C_1}, \ldots, m^{C_K}]R$ is the nullspace base on $L$. The different components extraction of the smallest eigenvector is called as “Spectral Rounding” and it becomes the concern for some researches [11-15]. The simple approach for clustering pixel image uses *K*-

*Means* algorithm [11] and perturbation analysis to limit algorithm mistake as connectivity function in and among clusters.

### *Matting Laplacian*

Matting *Laplacian* [7] is used for evaluating the matte quality without estimating foreground and background color as in Equation (7). It uses local window *w* forming two lines which is different in RGB domain. The $\alpha$ in *w* stated as linear combination of color channel.

$$\forall i \in w \;\; \alpha_i = a^R I_i^R + a^G I_i^G + a^B I_i^B + b \qquad (7)$$

The matte extraction problem becomes one of the findings in alpha matte minimizing the deviation of linear model (previous equation) in all image windows $w_q$ as shown in Equation (8).

$$J(\alpha, a, b) = \sum_{q \in I} \sum_{i \in w_q} \left( \begin{matrix} \alpha_i - a_q^R I_i^R + \\ a_q^G I_i^G + a_q^B I_i^B + b_q \end{matrix} \right)^2 + \varepsilon \left\| a_q \right\|^2 \qquad (8)$$

$\varepsilon \left\| a_q \right\|^2$ is the regularization requirement of $\alpha$. The coefficient of linear model $\alpha, b$ enables the elimination from Equation (8), and results quadratic cost in $\alpha$.

$$J(\alpha) = \alpha^T L \alpha, \qquad (9)$$

Cost function as seen in Equation (9) has minimum trivial that is the vector. Then, in framework user-assisted [13], $J(\alpha)$ is the subject minimized in user constraint. $L$ is matting *Laplacian*. Symmetric semi-definite positive matrix $N \times N$ is the matrix inserting input image function in local windows, depending on unknown foreground and background color in the coefficient of linear model. $L$ variable is defined by the sum of matrix $\boldsymbol{L} = \sum_q \boldsymbol{A_q}$ in which on each is filled with affinity among pixels in local window $w_q$ as seen in Equation (10).

$$A_q(i,j) = \begin{cases} \delta_{ij} - \dfrac{1}{|w_q|} \left( 1 + (I_i - \mu_q)^T \left( \sum_q + \dfrac{\varepsilon}{|w_q|} I_{3x3} \right)^{-1} (I_j - \mu_q) \right), \\ 0 \quad Otherwise \end{cases}$$
$$where\ (i,j) \in w_q \qquad (10)$$

$\delta_{ij}$ is Kronecker delta, $\mu_q$ is the average color vector in al pixel $q$, $\sum_q$ is matrix covariant size $3 \times 3$ in the same windows, $|w_q|$ is the sum of pixels in window, and $I_3$ is identity matrix size $3 \times 3$. By the occurrence of the smallest eigenvector, the other use of matting *Laplacian* property (eq. 10) is to catch information of job fuzzy cluster on image pixel, including the calculation before the limit determent by user is specified [13].

### *Linear Transformation*

Seeking linear transformation in eigenvector will result a set of vector in which the value is closed to binary. The formula is noted as $E = [e^1, \ldots, e^k]$ become matrix $N \times K$ of eigenvector. Then, in finding a set of linear combination $K$, vector $y^k$ minimizes Equation (11).

$$\sum_{i,k} \left| \alpha_i^k \right|^\gamma + \left| 1 - \alpha_i^k \right|^\gamma, \text{ where } \alpha^k = E y^k$$
$$\text{subject to} \sum_k \alpha_i^k = 1 \qquad (11)$$

If $0 < \gamma < 1$, thus, the value of $\gamma = 0{,}9$, then $\left| \boldsymbol{\alpha_i^k} \right|^\gamma + \left| \mathbf{1} - \boldsymbol{\alpha_i^k} \right|^\gamma$ is robust measurement value in matting component [10]. The result of Newton process depends on initialization process since the cost function (eq. 11) is not convex. The *K*-means algorithm is applicable in the initialization process on the smallest eigenvector in matting *Laplacian* and projects indicator vector of cluster resulted from eigenvector $E$ is shown in Equation (12).

$$\alpha^k = E E^T m^{C^k} \qquad (12)$$

The matting component result is then added so that it gives solution in (see Equation (11)).

### *Grouping Component*

The complete extraction result of foreground matte is determined by simple sum of components in foreground. For example, $\alpha^{k_1}, \ldots, \alpha^{k_n}$ is design as foreground component as seen in Equation (13).

$$\alpha = \alpha^{k_1} + \ldots + \alpha^{k_n} \quad (13)$$

If the smallest eigenvector is not equal to zero, the measurement of result quality $\alpha$-matte is done by $\alpha^T L \alpha$, in which $L$ is the matting *Laplacian*. The first calculation of correlation among matting component and $L$ and deviation in matrix $\Phi K \times K$ is defined in Equation (14).

$$\Phi(k,l) = \alpha^{kT} L \alpha^{l} \quad (14)$$

Then matte cost is calculated in Equation (15).

$$J(\alpha) = b^T \Phi b \quad (15)$$

where $b$ is the binary vector of *K-dimensional* indicating the chosen component.

## Fuzzy C-Means (FCM) Algorithm

Cluster center of the FCM algorithm initially was performed to mark the location of the average of every cluster. In these conditions, the accuracy of the cluster centers is inaccurate, since each data has a degree of membership in every cluster. Accuracy improvement at the cluster center is performed by minimizing the objective function; see Equation (16), which is performed repeatedly in order to move towards the cluster center in the right position. The objective function [16] is denoted in Equation (16) and Equation (17).

$$J_m(U,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c} (\mu_{ik})^{m'} (d_{ik})^2, m' \in (1,\infty) \quad (16)$$

where

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2} \quad (17)$$

The objective function $J_m$ can have a large number of values, however only the smallest one related to the best clustering, therefore it is needed to find the most optimum value of the large number of possible values. Euclidean distance is used to measure the distance between the ***i***th cluster center and the ***k***th data sets [16]. The matrix data is denoted in Equation (18).

$$x = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{nm} \end{bmatrix} \quad (18)$$

And matrix of cluster center defined in Equation (19).

$$v = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{m1} & \cdots & v_{mm} \end{bmatrix} \quad (19)$$

The minimum objective function indicates the best clustering result, therefore (see Equation (20)).

$$J_m^*(\overset{*}{U},\overset{*}{V};X) = \min J(U,V;X) \quad (20)$$

if $d_{ik} > 0, \forall i,k; m > 1$, and $X$ has at least $m$ elements, then $(U,V)$ can minimize $J_m$ with the condition as seen in Equation (20) and Equation (21).

$$\mu_{ik=} \frac{\left[\sum_{j=1}^{m} (X_{ij} - V_{kj})^{2}\right]^{\frac{-1}{m-1}}}{\sum_{k=1}^{m}\left[\sum_{j=1}^{m} (X_{ij} - V_{kj})^{2}\right]^{\frac{-1}{m-1}}}; 1 \le i \le m; 1 \le k \le n \quad (21)$$

and

$$V_{kj} = \frac{\sum_{i=1}^{n} \left((\mu_{ik})^m X_{ij}\right)}{\sum_{i=1}^{n} (\mu_{ik})^m}; 1 \le i \le m; 1 \le j \le m \quad (22)$$

## Discrete Courier Transform-2D

DCT - 2D (Discrete Cosine Transform - 2D) is applied to reduce noise around the edge boundary of the foreground, which is a direct extension of the 1-D case and is given in Equation (23)/

$$C(u,v) = \alpha(u)\alpha(v)\sum_{y=0}^{N-1}\sum_{x=0}^{N-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2n}\right] \cos\left[\frac{\pi(2y+1)v}{2n}\right] \quad (23)$$

for $u,v =$ 0, 1, 2, $, N-1$, $\alpha(u)$ and $\alpha(v)$ are defined in Equation (24).

$$\alpha(u) = \begin{cases} \sqrt{\dfrac{1}{N}} & \textit{for } u = 0 \\ \sqrt{\dfrac{2}{N}} & \textit{for } u \ne 0 \end{cases} \quad (24)$$

The inverse transformation is defined in Equation (25).

$$f(x,y)=\sum_{y=0}^{N-1}\sum_{x=0}^{N-1}\alpha(u)\alpha(v)C(u,v)\cos\left[\frac{\pi(2x+1)u}{2n}\right]\cos\left[\frac{\pi(2y+1)v}{2n}\right], \tag{25}$$

for $x, y = 0, 1, 2, .., N-1$.

The 2-D basis functions can be generated by horizontally multiplying the basic functions as in the 1-D with a vertical set oriented of similar functions [17]. Progressive increase in frequency occurs both in the vertical or horizontal direction as the basic function for $N$ = 8. The top left of the basic function is the result of multiplying the DC component

## OBJECT EXTRACTION SYSTEM

In this section, the proposed object extraction framework for compressed image was performed in the steps, as depicted in figure 1.

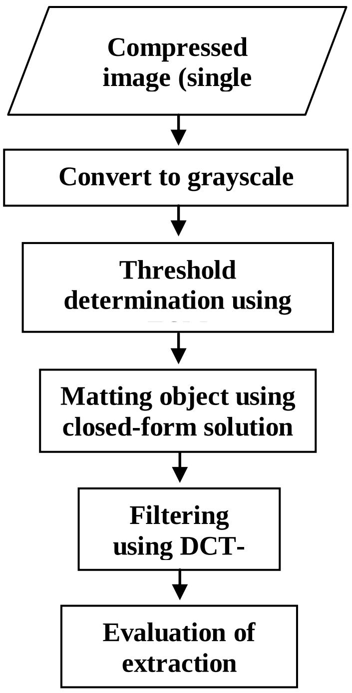



Figure 1. Object Extraction Stages.

The raw video data was fragmented into the frame and compressed in a single image (ilustrated in figure 1). The first image was converted into grayscale . Furthermore, the grayscale image was used as data matrix (eq. 18) and FCM [18] was applied to obtain the objective function (eq. 16).

The previous method [7, 10], the extraction process is performed by deriving the cost function in alpha, whereas the optimal cost function was obtained by a sparse linear system. In our work, the objective function of the FCM is applied to determine the best threshold instead of using a sparse linear system. Object extraction is performed by considering a user-spesified constraint on the foreground and background areas (white scribble representing foreground and black represents the background). A new closed-form solution approach, which is used as the basis in this study was applied to draw "matte" of the whole image.

Finally, the evaluation of the performance of the system is performed by comparing the extracted objects with ground truth.

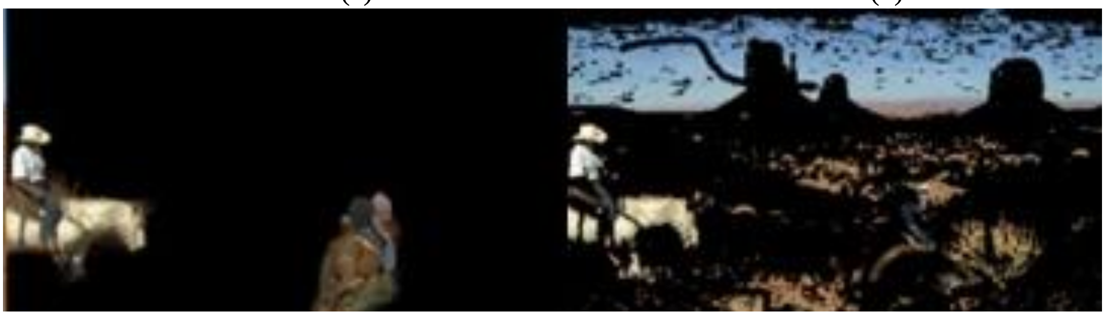
## RESULT AND DISCUSSION

In our experiments, we applied the MPEG-4 standard tests for video sequences obtained from the UCF Sports Action Data Set. At the first stage, we splited a video into several frames and used the first frame as the analyzed frame (see Figure 2). The first frame was extracted with a closed-form solution approach [7]. Initially, we used first frame as input image and scribble image as depicted in the Figure 3.a and 3.b. Object separation is performed by pulling "matte" of the whole image. It is performed by considering the alpha value in the Equation (1) which is obtained by the cost function in the Equation (9). The value of the cost function greatly affects accuracy of matte extraction. To optimize the cost function, [7,10] used a sparse linear system. The results are quite impressive when applied to the raw image data (describe in Figure 3.c).

However, the techniques were not robust as it was implemented in a compressed image (such as object extracted in Figure 3.d). It occured since the intensity values in each channel for some pixels in the compressed image, was change. The changes were presented in the form of a histogram which was comparing the frequency of intensity value between raw image and the compressed image. (shown in Figure 4.a, b, c).

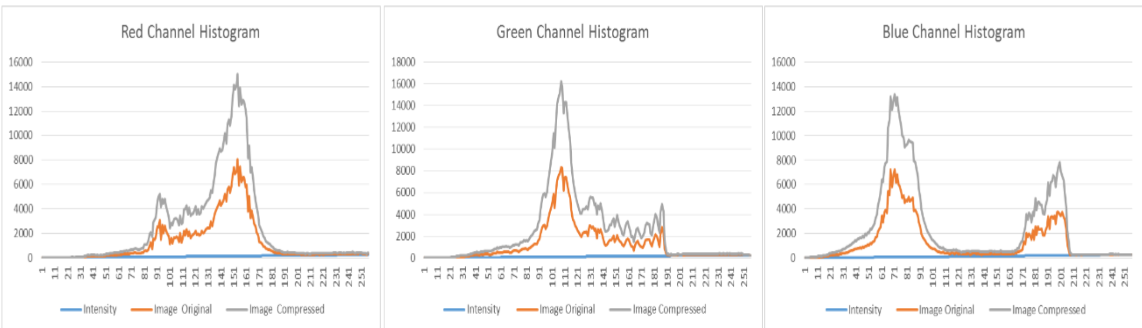

Figure 2. Frame separation of the video sequences

(a) (b)



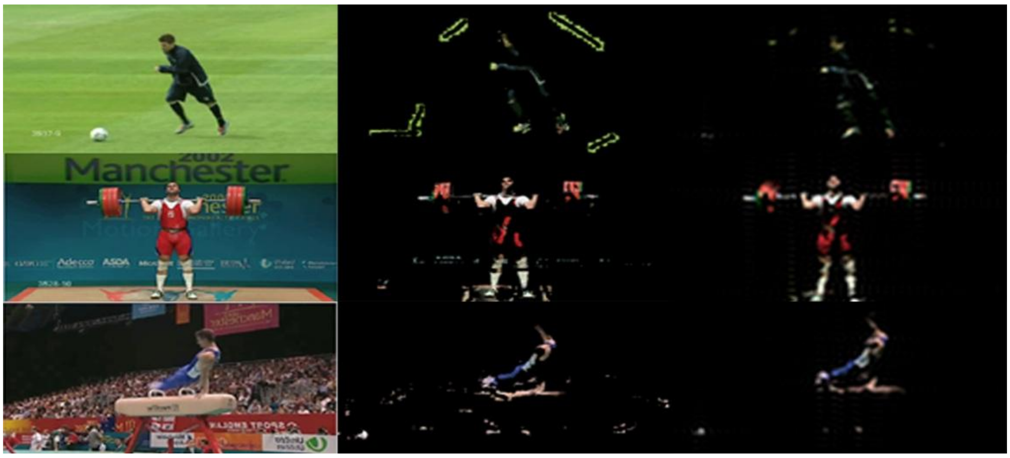
(c) (d)

Figure 3. (a) Input Image, (b) Scribble Image, (c) Object Extracted from the Raw Image, (d) Object Extracted from the Compressed Image.



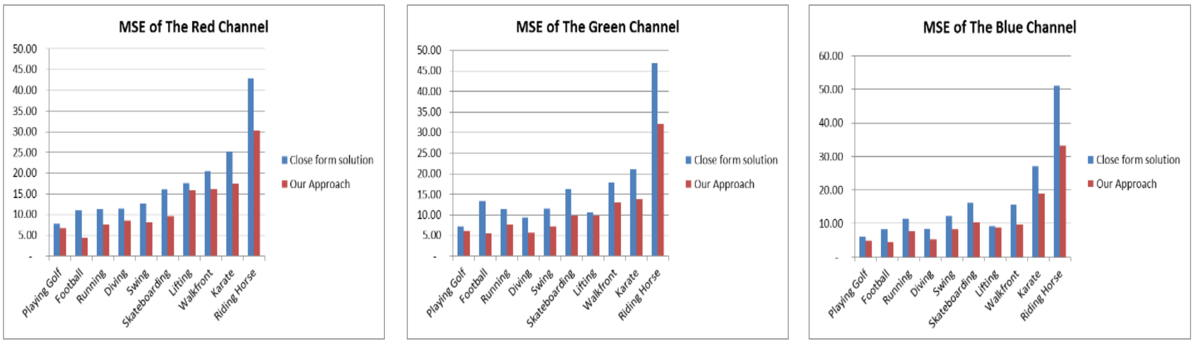

(a) (b) (c)

Figure 4. The Changes in Intesity : (a). Red Channel, (b). Green Channel, (c). Green Channel.



(a) (b) (c)

Figure 5. (a) Compressed Image, (b) Extracted Objects with Closed-Form Solution, (c) Extracted Objects with Our Approach.




(a) (b) (c)

Figure 6. MSE Value Between the Extracted Object and Ground Truth : (a). Red Channel, (b). Green Channel, (c). Blue channel

Table 1. The Value of MSE between the Extracted Object and Ground Truth

| Image | Closed-Form Solution | | | Our Approach | | |
|---|---|---|---|---|---|---|
| | **Red Channel** | **Green Channel** | **Blue Channel** | **Red Channel** | **Green Channel** | **Blue Channel** |
| Playing Golf | 7.79 | 7.17 | 5.94 | 6.69 | 6.07 | 4.85 |
| Football | 11.07 | 14.42 | 8.34 | 4.49 | 5.57 | 4.43 |
| Running | 11.37 | 11.45 | 11.37 | 7.69 | 7.76 | 7.71 |
| Diving | 11.49 | 9.36 | 8.47 | 8.52 | 5.69 | 5.23 |
| Swing | 12.74 | 11.50 | 12.13 | 8.07 | 7.21 | 8.19 |
| Skateboarding | 16.12 | 16.27 | 16.10 | 9.65 | 9.85 | 10.41 |
| Lifting | 17.54 | 10.56 | 9.28 | 15.99 | 9.80 | 8.76 |
| Walkfront | 20.53 | 17.95 | 15.51 | 16.30 | 13.14 | 9.71 |
| Karate | 25.22 | 21.19 | 27.03 | 17.51 | 13.90 | 19.00 |
| Riding Horse | 42.97 | 46.99 | 51.18 | 30.40 | 32.13 | 33.33 |

To improve the ability of the previous algorithm, we applied FCM [18] threshold which is applied in the cost function and DCT-2D to filter noise around edge boundary. FCM algorithm comprises the following stage:

1. Specify the data to be in the cluster, namely $n \times m$ matrix (*n* = number of sample data, *m* = attribute each data). $X_{ij}$ sample data $i$th (*i*=1,2,3 .... *n*), attribut *j*th (*j* = 1,2,3....*m*).
2. Determine :
   - Number of cluster ($c$)
   - Grade ($m$)
   - Maximum iteracy ($MaxIter$)
   - The smallest expected error ($\xi$)
   - Initial objective function ($P_0 = 0$)
   - Initial iteracy ($t = 1$)
3. Generate random numbers $\mu_{ik}$, *i=1,2,3....n, k=1,2,3....c* as initial elemen of $U$ matrix. Calculate the amount of each column by Equation (26).
$$Q_i = \sum_{k}^{c} \mu_{ik}, k = (1,2,3...c) \quad (26)$$
4. Estimation the cluster center using Equation (22), where *k=1, 2, 3...c* and *j=1, 2, 3...m*.
5. Calculate the objective function by Equation (27)
$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]\left(\mu_{ik}\right)^m\right) \quad (27)$$
6. Calculate the change of the matrix partition using Equation (21), where *i=1, 2, 3...n* and *k=1, 2, 3...c*.
7. Check the stop condition :
   - If $\left(\left|P_t - (P_t - 1)\right| < \zeta\right)$ or $\left(t > MaxIter\right)$ then stop
   - Otherwise : $\left(t = t + 1\right)$: go to step – 4.

Next, the value of the FCM threshold is considered as a cost function in alpha parameters of the closed-form solution approach. Extraction results are depicted in Figure 5.b. To reduce the noise around edge boundary of the extracted object, we applied DCT-2D as seen in Equation (23), Equation (24), and Equation (25). The final result is illustrated in Figure 5.c.

We evaluated the extraction results of the experiments by using MSE (Mean Squared Error) as in Equation (28).

$$MSE = \sqrt{\frac{(\sum_{i=1}\sum_{j=1}[Grd.img_{(i,j)} - Ext.Obj_{(i,j)}])}{NM}} \quad (28)$$

$Grd.img$ is the ground truth image resulted from the object extraction of the raw data. Whereas $Ext.Obj$ is the new image generated from the extraction process [19]. The value of MSE between the raw image and compressed image is described in Table 1. The improvement of extraction results of each channel was depicted in the Figure 6.

## CONCLUSION

In this paper, we proposed an approach to increase the object extraction quality by using the FCM algorithm. FCM is used to produce an adaptive threshold on the unknown region. To smooth the extracted object, we used DCT-2D filter to remove noises around the edge boundary. From our experiments in 10 video datasets and perfomance evaluation conducted by MSE, the result showed that an average increase in accuracy of 31.55%. However, we found that our approach was not so robust on similarity of the color. Therefore, we would combine this algorithm with optimization of the cost function on the alpha region based on spectrum which are expected to improve the performance of object extraction algorithm.

## REFERENCES


[1] M. Hariadi., L. Huichien, and T. Aoki., “Semi-Automatic Video Object Segmentaion Using LVQ with Color and Spatial Features,” *IEICE Trans on Information and System,* vol. E-88D, no. 7, pp. 1553-1556, 2005.

[2] Kosch. H, *Distributed Multimedia Database Technologies supported by*

*MPEG-7 and MPEG- 2*, Washington D.C. : CRC Press, 2003.

[3] L. Chiariglione, "The MPEG-4 Standard," *Journal of China Institute of Communications*, pp. 54-67, September 1998.

[4] R. Koenen, F. Pereira, and L. Chiariglione, "MPEG-4: Context and Objectives," *Signal Processing: Image Communication*, vol. 9, no.4, pp. 295-304, 1997.

[5] A. Bovic, *The Hand Book of Image and Video Processing*, San Diego : Academic Press, 1998.

[6] E. Şaykol, E. Güdükbay, and O. Ulusoy, "A Semi-Automatic Object Extraction Tool for Querying," *In Proceedings of the 7th Workshop on Multimedia Information Systems (MIS '01),*, Villa Orlandi, Capri, Italy, 2001.

[7] A. Levin, D. Lischinski, and Y. Weiss, "A Closed-Form Solution to Natural Image Matting," *IEEE Transactions on Pattern Analysis And Machine Intelligence*, vol. 30, no. 2, pp. 1-15, February 2008.

[8] J. Wang and M. Cohen, "An iterative optimization approach for unified image segmentation and matting," *in Proceedings of Tenth IEEE International Conference on Computer Vision (ICCV)*, Beijing, 2005.

[9] T. Porter and T. Duff, "Compositing digital images," *Computer Graphics*, vol. 18, no. 3, pp. 253-259, 1984.

[10] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1699-1722, 2008.

[11] K. Lang, "Fixing Two Weaknesses of the Spectral Method," *in Proceedings of Advances in Neural Information Processing Systems*, 2005.

[12] S. Yu and J. Shi, "Multiclass Spectral Clustering," *in Proceedings of Ninth IEEE Conference on Computer Vision*, Nice, France, 2003.

[13] L. Zelnik-Manor and P. Perona, "Self-Tuning Spectral Clustering," *in Proceedings of Advances in Neural Information Processing Systems*, 2005.

[14] A. Ng, M. Jordan and Y. Weiss, "Spectral Clustering: Analysis and an Algorithm," *in Proceedings of Advances in Neural Information Processing Systems*, Cambridge, 2001.

[15] D. Tolliver and G. Miller, "Graph Partitioning by Spectral Rounding: Applications in Image Segmentation and Clustering," *in Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.

[16] T. Roos, *Fuzzy Logic With Engineering Applications, Third Edition*, John Wiley & Sons, 2010.

[17] W. Pennebaker and J. Mitchell, *JPEG – Still Image Data Compression Standard*, Newyork: International Thomsan Publishing, 1993.

[18] R. S. Basuki, M. Hariadi, and R. Pramunendar, "Fuzzy C-Means Algorithm for Adaptive Threshold on Alpha Matting," *in Proceedings of Computer Society Coference on Information Technology and Electrical Engineering*, Yogyakarta, Indonesia, 2012.

[19] C. Mythili dan V. Kavitha, "Color Image Segmentation using ERKFCM," *International Journal of Computer Applications*, vol. 41, 2012.