International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 01
January 2015

FCM Based Enhanced Approach for Object Extraction from Videos

Jasleen Kaur
M.tech Student, CSE Department, ACE & AR, Devsthali, Near Mithapur, Haryana, India
Ashok
Assistant Professor, CSE Department, ACE & AR, Devsthali, Near Mithapur, Haryana, India

Abstract—
In Video segmentation huge number of research progress has done, which includes variety of algorithms for specific applications. Videos cover wide variations in object with multiple frames capturing different pose, illumination and other variations. This diverse information can be aggregated together for efficient object segmentation. Till now, it remain challenging problem to accurately extract the target object from the video because of the variations of intensity, view, brightness, motion and other complex backgrounds. Also, accuracy and fast object tracing and object's motion updating is necessary. In this work, we proposed a method to extract object from video sequence. It involves tracking through training phase, test phase then update the appearance model. To do this, we start with an initialized location of the object for the first frame. We do simple tracking for the first 4 frames, and use the corresponding target-background regions to train the discriminative appearance model. For each frame, we first use the appearance model to get a target confidence map, and then find the target candidate with the largest confidence. The target-around region is saved for updating appearance model every several frames. On the computed confidence map saved images, simple threshold segmentation method is applied and segmented objects are saved. Then on segmented object, apply Fuzzy C-Means clustering. This will result in improved segmentation quality. Output is shown as the segmented images of object in video sequence.

Keywords—Segmentation; FCM; PSNR; MSE; Mean-Shift; SLIC

1. INTRODUCTION

Video segmentation becomes a key technique for visual information extraction and plays an important role in pattern recognition, computer vision and digital video processing. Several advances of video segmentation will benefit in wide range of video-based applications including security and surveillance, personal entertainment, bank transactions monitoring, and video conferencing [5]. In a video, each object can be considered as single pattern represented by temporal and/or spatial features [8].

Segmentation is the process of partitioning a piece of information into meaningful elementary parts termed segments [6]. In terms of similar characteristics, it is the process of partitioning data into groups of potential subsets. Whereas video segmentation refers to the process of splitting videos into homogenous temporal,spatial segments meaningful from a semantic point of view [5]. The term segmentation is also used to describe background/foreground detachment in video, which can be seen as a special case of spatio-temporal segmentation. Over all the decision space, i.e. 1D, 2D or 3D for spatial, temporal and temporal-spatial segmentation, respectively. There exists no unique solution to the segmentation problem of video and images, the temporal, spatial or spatio-temporal segments are applications specific and depend on the subjective view of each human observer [6].

Video segmentation is an integral part of many video analysis and coding tasks, including: (i) video summarization and indexing, (ii) advanced image/video coding and rate allocation, (iii) improved motion estimation, (iv) 3-D motion and structure estimation with multiple objects, (v) video surveillance/understanding, (vi)video indexing and summarization, and (vii) video authoring and editing. In the last two decades, the problem of segmenting video/image data had significant impact both on applications and new pattern recognition [5].

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 01
January 2015

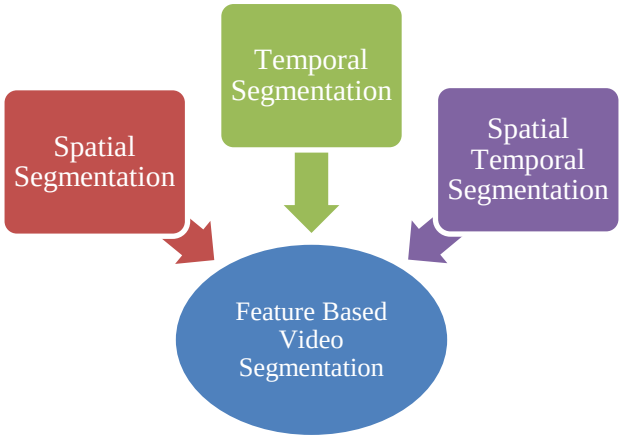

Figure 1: Application specific video segmentation types.

Spatial segmentation: Region-based and boundary-based are two methods for segmenting 2D images. To locate object boundaries, boundary-based methods use primarily gradient information. Region-based approaches uses position, texture, and intensity features [6].If we treat the motion cue as one of the low level features such as intensity, color, and texture, many image segmentation algorithms can be easily extended to video segmentation. For example, to segment a moving object out from a video clip, a 3D graph cut was presented to partition watershed presegmentation regions into foreground and background while preserving temporal coherence [5]. For each frame, the segmentation in each tracked window is refined using a 2D graph cut based on a local color model [4].

Temporal segmentation: Temporal video segmentation is to partition the video into primary image sequences termed shots and scenes. Temporal segmentation is usually based on change detection followed by motion analysis.Temporal segmentation between frames is performed by calculating the transition from one frame to the next. Transitions between shots, may be detectible by examining two consecutive frames. Transition between more than two frames being usually more difficult to detect, depending upon the actual transition type (e.g. dissolve, wipe, fade, etc.). In uncompressed video transition is detected by means of pair-wise pixel comparisons between distant or successive frames or by finding the color histogram corresponding to different frames. Other techniques to temporal segmentation consider block-wise comparisons, changes in blocks are detected by means of edge-based, and thresholding methods. Other methods, have also been proposed, including the comparison of motion features at different time instances. For the homogeneity-based spatial segmentation, several methods use clustering algorithms [6].

Spatio-temporal segmentation: Segmentation in a 3D decision space, several methods have also been proposed (i.e. Spatio-temporal video segmentation), both supervised and unsupervised as in case of 1D and 2D decision spaces. Temporal segmentation is followed by spatio-temporal segmentation to shots. Some cases initially spatial segmentation method is applied to each frame independently. Tracking of regions play an important role in 3D segmentation. In temporal tracking algorithms, two key factors (i.e. illumination conditions and noisy data) are limiting the efficiency. To solve these issues, a color model that is robust against changing illumination and noisy data is used. By foreground/background separation the problem of video segmentation could limit. Later performing segmentation using feature spaces,or by using primarily motion information and applying rule-based processing to enhance the motion segmentation result [6].

II. Related work

In existence for over several years, video segmentation has undergone vast technological progress, which has resulted in a great variety of algorithms.

Yadav *et al*. [9] have discussed about the segmentation of video as deformation Video quality improvement and blurring of video. Many authors have improved the quality of video using different noise filter and sliding window technique but all these methods were not accurate and lack for object prediction in noise video. Li Bing, Xu De, Wang Fangshi [10] advanced a semiautomatic algorithm of video segmentation based on object tracing. In this algorithm video segmentation makes use of the temporal and spacial information, with the participating of users. A good result can be obtained when the color of foreground is similar to the color of the background. But when the video object's shape changes obviously the error caused by motion error need to be improved. An idea about the mean shift algorithm [2] used for mode detection and clustering proposed by Fukunaga and Hostetler in 1975 and rekindle by Cheng. Mean shift is a nonparametric, iterative procedure for searching the mode of a density function. More specifically, mean shift estimates the local density gradient of similar pixels via finding the peaks in the local density. In [7], they use efficient and effective mid-level visual cues for object tracking with superpixels. Involves training phase, test phase then update the appearance model. Confidence map is compu ted using the appearance model to obtain the best candidate by maximum a posterior estimate. A discriminative appearance model is used to distinguish the target and the background with mid-level cues. Used SLIC [3] (simple linear iterative clustering), which is the new version of k-means for superpixel generation. Fuzzy C-Means algorithms to extract the segmented object from video. The main advantage of Fuzzy C-Means technique is that it yields regions which are more homogeneous also it reduces spurious blobs. Further it is less sensitive to noise [8].

III. PROBLEM STATEMENT

The limitations and challenges of image segmentation motivates towards the extraction of object from video. Video segmentation has undergone vast technological progress, which has resulted in a great variety of algorithms. In video segmentation, the first main step is tracking the object. The superpixel tracking is one of the algorithms used for object tracking. This algorithm computes a confidence map in the image based on the color histogram of the object. By applying threshold method on the computed confidence map, the segmented object can be obtained.The quality of segmented object can be enhancedby using some clustering technique. Fuzzy c-means clustering is one of techniques that can be used to enhance the quality of segmented object. So, we have decided to use Fuzzy c-means Clustering to enhance the quality of segmented object.Video segmentation has undergone vast technological progress, which has resulted in a great variety of algorithms. For video segmentation, if we use more than one method together, then improved results can be obtained. So, we implemented multiple methods for video object segmentation which are: Superpixel Tracking, Mean-Shift and Fuzzy C-Means is used. Then Comparative analysis is done to show the quality enhancement the proposed work.

IV. METHODOLOGY FOLLOWED

To show the methodology various steps have been followed as shown in Figure 2. In the first step, the video is input in the .avi format. For example, in this work we used the video "#bird2". The next step, from the given video frames are extracted and stored for further accesses in different functions. Frames are converted from rgb to grayscale color space. The motion need to be extracted from frames. Tracking of target object are saved in directory, frames shown with bounding box around the target object. The parameters for tracking and location of target object are initialized, which are used in further process (grid_size = 64;).The size of template for simple tracking in first several frames (train_frame_num = 4;).

Initially specify the frames for training; we used first 4 frames for training. For the bird, spacing, frequency etc. are initialized. And location of target object in the first frame are initialized by p=[px, py, sx, sy, theta]. Where px and py are the coordinates of the center of the bounding box, sx and sy are the size of the box in the x (width) and y (height) dimensions, before rotation. And theta is the rotation angle, which is currently set to 0.0. Process start by initializing the location of the target object. In the first frame the target object parameters are specified. For the training of frames, initially we have taken first 4 frames for training. To construct an appearance model for both the target and the background, each pixel can be learned from a set of m training frames. That is, for a certain pixel at location (i, j) in the t-th frame pixel (t, i, j). Here m=4, i.e we have taken first four frames for the training process and the training information is used by the tracking process. Assume that the target object can be represented by a set of superpixels without significantly destroying the boundaries between target and background. The surrounding region of the target is segmented into superpixels. First, we segment the surrounding region of the target in the t-th training frame into N_t superpixels. Each superpixel sp(t, r) (t = 1, ..., m, r = 1, ...,N_t) is represented by a feature vector f_t^r. Basically in this we calculate the number of superpixels of the frame, then calculate superpixel histogram and then calculate the HSI color histogram of this frame.

Next, we apply the mean shift clustering algorithm on the total feature pool $F = \{f_t^r | t = 1,...,m ; r = 1, ...,N_t\}$, mean shift return the n different clusters. Mean shift represents a general non-parametric mode finding/clustering procedure. The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. In this case, a Gaussian kernel is chosen instead of a flat kernel, then every point will first be assigned a weight which will decay exponentially as the distance from the kernel's center increases. The bandwidth of the mean shift clustering is set to the range of 0.15 and 0.20. . The cluster centre ,cluster radius and cluster members are obtained. In the feature space, each cluster clst(i) (i = 1, ..., n) is represented by its cluster center fc(i), its cluster radius rc(i) and its own cluster members. We give each cluster a target-background confidence measure between 1 and -1. We assign every pixel in the superpixel sp(t, r) with superpixel confidence 1 to N_t , and every pixel outside this surrounding region with -1. Superpixel appearance model is constructed based on the four factors: cluster confidences, cluster radius, cluster centres and cluster members which are used for determining the cluster for a certain superpixel. This appearance model is used to compute confidence map in further step.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 01
January 2015

Figure 2: Proposed Methodology for Object Based Video Segmentation.

For each frame, we first use the appearance model to get a target confidence map. When a new frame arrives, we first extract a surrounding of the target and segment it into N_t superpixels. To compute a confidence map for current frame, we evaluate every superpixel and compute its confidence measure. The confidence measure of a superpixel depends on two factors: the cluster it belongs to, and the distance between this superpixel and the corresponding cluster center in the feature space. The farther the feature of a superpixel f_t^r lies from the corresponding cluster center fc(i) in feature space, the less likely this superpixel belongs to clst(i). We obtain a confidence map for each pixel on the entire current frame as follows. Then we normalize the confidence, for the target candidates with positive confidence values, the ones with larger area size should be weighted more. For the target candidates with negative confidence values, the ones with larger area size should be weighted less. We then normalize the final confidence of all targets within the range of [0, 1]. The confidence maps facilitate the process of determining the most likely target location. In the last step of tracking, when the target-background region is saved and used for updating every several frames. We update the appearance model with the retained sequence every frames, and this process is the same as the training process.

Then for target-background segmentation simple adaptive threshold method is applied. Threshold is the simplest Segmentation method. The pixels are partitioned depending on their intensity value. Global thresholding using an approximate threshold T.

$$g(x, y) = \begin{cases} 1 & ,if\ f(x,y) > T \\ 0 & ,if\ f(x,y) \leq T \end{cases} \qquad (1)$$

level = graythresh (I); Level is a normalized intensity value that lies in the range [0, 1]. The graythresh function uses Otsu's method, which chooses the threshold to minimize the intra class variance of the black and white pixels.

The FCM clustering algorithm was first introduced by Dunn [10] in 1973 and later extended by Bezdek [1] in 1981. This algorithm has been used as one of the popular clustering techniques for image segmentation in pattern recognition. In the FCM, each image pixel has certain membership degree associated with each cluster centroid. These membership

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 01
January 2015

degrees have values in the range, indicating the strength of the association between that pixel and a particular cluster centroid. The FCM algorithm attempts to partition every image pixel into a collection of the fuzzy cluster centroids.

$$J_m(U,C) = \sum_{i=1}^{N}\sum_{j=1}^{K} u_{ji}^{m}\, d_{ji}^{2}$$

subject to (1)

$$\sum_{j=1}^{K} u_{ji}^{m} = 1, 1 < j < K$$

$$\sum_{i=1}^{N} u_{ji}^{m} < N, 1 \le i \le N$$

$$\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ji}^{m} = \mathrm{N} \qquad (2)$$

where N is the total number of pixels in image, u_{ji} is the membership degree of i th pixel x_i to j th cluster centroid c_j , m is the exponential weight of membership degree which controls the fuzziness of the resulting partition, and $d_{ji} = \| x_i - c_j \|$ is the distance between x_i and c_j . Let $U_i = (u_{1i}, u_{2,}\ldots,u_{ki})^T$ be the set of membership degree of x_i associated with each cluster centroids, then U = (U_1, U_2, ……,U_N) is the membership degree matrix and C = (c_1, c_2, ……,c_k) is the set of cluster centroids.

$$u_{ji} = \frac{1}{\sum_{k=1}^{k}\left(d_{ji}/d_{ki}\right)^{(2/m-1)}} \qquad (3)$$

where $1 \le j \le K$ and $1 \le i \le N$.

$$c_j = \frac{\sum_{i=1}^{N}\left(u_{ji}^{(b)}\right)^{m} X_i}{\sum_{i=1}^{N} u_{ji}^{m}} \qquad (4)$$

where $1 \le j \le K$ and X_i is the multidimensional feature vector of i th pixel x_i.

Steps For Fuzzy C-Means:

1.Set values for k, m and ϵ.
2. Initialize the fuzzy partition matrix U = [u_{ji}].
3. Set the loop counter b = 0.
4. Calculate the c cluster centers c_j using equation 4.
5. Calculate the membership using equation 3 for b+1. For i=1 to N.
6. $\|U^{b+1} - b\| \le \epsilon$ stop iteration. Otherwise b=b+1 and go to step 4.

V. EXPERIMENTAL RESULTS

Implementated in MATLAB 2013 (a). In the FCM ϵ=0.0001 and 4 cluster centers are initialized. After iteration the value of cluster centres are changed and final cluster centres obtained with segmented image. The SLIC algorithm [17] is applied to segment frames into superpixels where the spatial proximity weight and number of superpixels are set to 10 and 300, respectively. The bandwidth of the mean shift clustering [6] is set to the range of 0.15 and 0.20. The σ_c and σ_s in are set 3.04 in anticipation of the fastest motion speed or changing scale of the target objects.

We calculated the PSNR and MSE for the 12 different frames. The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics used to compare image compression quality.

$$MSE = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M * N}$$

The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error. The higher the PSNR, the better the quality of the compressed, or reconstructed image.

$$PSNR = 10\, log_{10}\left(\frac{R^2}{MSE}\right)$$

The PSNR and MSE calculated to show the quality enhancement. MSE and PSNR is calculated for the original and segmented frame (after tracked result) and the original and enhanced frame(FCM). As shown in table 5.1 and 5.2 resp.

Table 5.1 Calculated MSE for different frames for original frame with segmented result 3(b) and enhanced frame 3(d).

Frame number	MSE Original v/s segmented frame	MSE Original v/s fuzzy frame
7	2.4771e+4	2.2886e+4
21	2.7019e+4	2.4951e+4
26	2.7400e+4	2.4759e+4
30	2.6576e+4	2.4428e+4
35	2.6835e+4	2.5095e+4
38	2.7467e+4	2.6103e+4
45	2.8556e+4	2.6824e+4
46	2.9123e+4	2.7034e+4
51	2.9280e+4	2.7391e+4
52	2.9127e+4	2.7146e+4
61	2.7369e+4	2.5009e+4
80	2.5272e+4	2.3882e+4

As MSE is calculated in this table between the original frame and segmented frames 3(b) and with enhanced frame 3(d) obtained using FCM algorithm. The original , segmented and fuzzy frame are shown in figure 3. As the MSE of original vs

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 01
January 2015

fuzzy is less, means error is less. And we can say that the quality is improved after applying another technique.

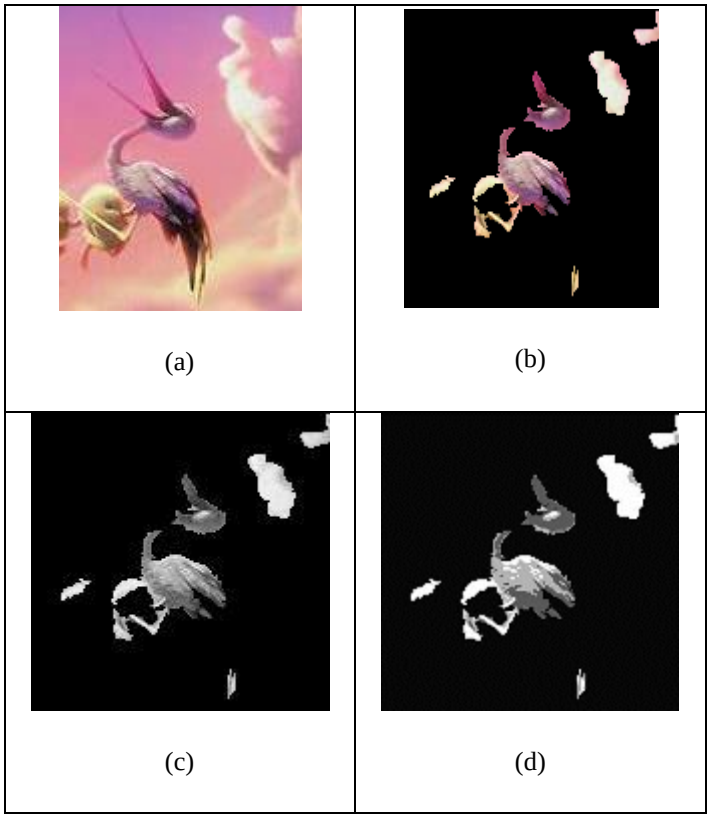

Figure3. 3(a) the otiginal frame with the target object is shown. 3(b) frame obtained from tracking and applied threshold method for segmentation. 3(c) 3(b) converted in grayscale. 3(d) enhancement of 3(b) using FCM Clustering algorithm. Only one of the frames is shown here.

Table 5.2 Calculated PSNR for different frames for original frame with segmented result 3(b) and enhanced frame 3(d).

Frame number	Original v/s segmented frame PSNR	Original v/s fuzzy frame PSNR
7	4.1914	4.5351
21	3.8141	4.1600
26	3.7533	4.1934
30	3.8858	4.2520
35	3.8438	4.1350
38	3.7427	3.9639
45	3.5738	3.8456
46	3.4885	3.8117
51	3.4650	3.7547
52	3.4873	3.7937
61	3.7581	4.1498
80	4.1044	4.3501

The PSNR is calculated in this table between the original frame and segmented frames 3(b) and with enhanced frame 3(d) obtained using FCM algorithm. The original , segmented and fuzzy frame are shown in figure 3.The PSNR value is higher with the enhanced frames. The higher the PSNR, higher is the quality. So, we proved the enhancement of our work.

VI. FUTURE SCOPE

We have Input video in .avi format only. So, there is a scope to convert original video images format from .avi to any other image formats like .mp4, .3gp, .flv, .mkv. We can use quality parameters like precision and recall for the quality measurement. This approach can be used to track different moving objects like woman, car etc. This can be enhanced to extract the accurate target object.

VII.REFERENCES

[1] Shaoping Xu, Lingyan Hu, Xiaohui Yang and Xiaoping Liu," A Cluster Number Adaptive Fuzzy c-means Algorithm for Image Segmentation ".International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, No.5 (2013), pp.191-204.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. PAMI, 24(5):603–619, 2002.

[3] A. Radhakrishna, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels. Technical Report 149300, EPFL,2010.

[4] Li Y. Sun J., Shum H.-Y.: Video object cut and paste. SIGGRAPH 2005, 24, 595–600 (2005).

[5] Li, Hongliang, and King Ngi Ngan. "Image/video segmentation: Current status, trends, and challenges." Video segmentation and its applications. Springer New York, 2011. 1-23.

[6] Encyclopedia Segmentation-of-Images-and-Video.html.

[7] Wang, Shu, et al. "Superpixel tracking." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[8] M. Arfan Jaffar, Bilal Ahmed, Nawazish Naveed, Ayyaz Hussain, and Anwar M. Mirza," Color Video Segmentation using Fuzzy C-Mean Clustering with spatial Information". WSEAS Transactions on Signal Processing 2009.

[9] Kuo-Liang Chung, Yah-Syun Lai and Pei-Ling Huang, "An Efficient Predictive Watershed Video Segmentation Algorithm Using Motion Vectors",Journal of Information Science and Engineering,Vol.26, pp. 699-711, 2010.

[10] Li Bing, Xu De, Wang Fangshi, “A video segmentation algorithm based on object tracing,” Journal of Beijing Jiaotong University. vol. 29, no. 5, pp. 89-91, 110, October 2005.