

Comparative Analysis of Different Clustering Techniques for Video Segmentation



Tunirani Nayak and Nilamani Bhoi

Abstract Video segmentation is an extremely challenging and active area in the field of video processing and computer vision. Video segmentation techniques can be classified basically into two approaches: one approach for which there are preassigned thresholds and another clustering approach for which the number of clusters has been used, which is known. Here, we have studied and analyzed the cluster-based techniques such as mean-shift, K -means, and fuzzy C -means segmentation algorithms. We have evaluated and compared the performances of segmentation methods qualitatively and also quantitatively. To calculate the different quantitative metrics, the images and ground truth of the CDnet 2014 database have been used.

Keywords Video segmentation · Mean-shift · K -means · Fuzzy C -means

1 Introduction

The video segmentation refers to decomposing a video data into meaningful elementary parts that have strong application and correlation with the real world contained in the video data. The collective numbers of segments resulted from video segmentation cover the entire real-time video data. The video signal keeps temporal information, which introduces the object motion and includes camera motion concept. Therefore, video has both temporal nature and spatial (static) nature [1]. Segmentation of video can be set with respect to temporal, spatial, or spatiotemporal [2–5]. In a video, the static image can be obtained by segmenting a frame in spatial domain. In temporal segmentation, the sequence of video frames is segmented.

T. Nayak (✉) · N. Bhoi

Department of Electronics and Telecommunication Engineering, VSSUT, Burla, Sambalpur, Odisha, India

e-mail: tuniraninayak@gmail.com

N. Bhoi

e-mail: nilamanib@gmail.com

© Springer Nature Singapore Pte Ltd. 2019

H. S. Saini et al. (eds.), *Innovations in Electronics and Communication*

Engineering, Lecture Notes in Networks and Systems 65,

https://doi.org/10.1007/978-981-13-3765-9_3

The segmentation criteria may be motion-based or color/texture-based which can be used both individually and combine form. In motion-based segmentation methods [6], there is a group of pixels with similar motion, and they are separated by multiple layers. In most of the methods [7–9], the optical flow estimation is preprocessed, and then the pixels are divided based on the pattern of motion. Some methods [6, 10, 11] are transformed and together have an effect on motion and performance evaluation.

In the last few years, many ways of segmentation have been analyzed for different images as normalized cuts [12], mean-shift [13], segmentation based on lossy compression [14], weighted aggregation (SWA) [15], and the fuzzy C -means technique [16]. Many works have been done in application of skin tumor border in medical images using segmentation which is unsupervised [17]. In these methods [16, 17], adaptive thresholding has been implemented by using scale-space filter for the extraction of the required number of clusters using histograms. Another method is proposed for fuzzy C -means algorithm which is based on histogram [18, 19] and also purposed for simplification of mean-shift filter and K -means clustering both by using background modeling. One of the commonly used models is mixture of Gaussian (MOG), kernel density estimation (KDE), etc., for background estimation.

Here, we have studied and analyzed the mean-shift algorithm [13], for the segmentation of each frame independently with the given predefined parameters. Fuzzy K -means (also called fuzzy C -means) and K -means are two simple clustering methods. K -means finds complex clusters (fusion of single clusters), and fuzzy C -means is the more formalized statistical method and finds a soft cluster with more than one block with a certain probability.

The paper is prepared as follows: Sect. 2 deals with methodology of video segmentation methods, i.e., mean-shift, K -means, and fuzzy C -means; Sect. 3 gives the performance evaluation metrics. The experimental and discussion results are presented in Sect. 4, and the conclusions of this study are given in Sect. 5.

2 Study of Cluster-Based Video Segmentation Methods

We have analyzed and studied three video segmentation algorithms such as mean-shift, K -means, and fuzzy C -means that are discussed in the following sections.

2.1 Video Segmentation Using Mean-Shift

Mean-shift is a simplest procedure where each data point moves to the center of data in neighboring countries, introduced by Fukunaga and Hostetler [13]. The kernel density function was introduced by Cheng's research into the mean-shift procedure.

In video segmentation, mean-shift algorithm is a direct resumption to protect the smoothness of interruptions. Each pixel is in its neighborhood. Then $\{x_1, \dots, x_n\}$ is a

set of data points that are based on geometric data that measure the dimension whose function dimension is d , and then the mean-shift vector of x is defined as follows:

$$M(x) = \frac{\sum_{i=1}^n x_i K_w(x_i - x)}{\sum_{i=1}^n K_w(x_i - x)} \quad (1)$$

where x is the center of the kernel and $K_w(x) = |W|^{-1/2} K(W^{-1/2}x)$. W is the approximate density of the dimension $d \times d$, and matrix values determine the solution, the detection, the mode, and the smallest number of consecutive pixels. W is usually a diagonal matrix and W is a symmetric positive bandwidth matrix in which the values determine the smallest number of contiguous pixels. $K(x)$ is a d -variate nucleus and is regarded as nonnegative free radical symmetric functions centered at zero.

The normal kernel defined below was used in our segmentation algorithm:

$$K(x) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\|x\|^2\right) \quad (2)$$

Because of the mean-shift, the $M(x)$ axis always shows the direction of the maximal density, and the average intermediate procedure is obtained by calculating the core change and changing the vectors of the vector $K(x)$. At the end of the approximate procedure, it is guaranteed at the nearest point where the estimate is zero gradients. Define as

$$y_{j+1} = \frac{\sum_{i=1}^n x_i K_w(y_j - x_i)}{\sum_{i=1}^n K_w(y_j - x_i)} \quad j = 1 \quad (3)$$

Equation (2) can be modified as

$$y_{j+1} = \frac{\sum_{i=1}^n x_i \exp\left(-\frac{1}{2} \sum_{l=1}^d \left(\frac{y_j^l - x_i^l}{w_l}\right)^2\right)}{\sum_{i=1}^n \exp\left(-\frac{1}{2} \sum_{l=1}^d \left(\frac{y_j^l - x_i^l}{w_l}\right)^2\right)} \quad (4)$$

where y_j^l and x_i^l are the l th component of y_j and x_i , respectively. Thus, mean-shift algorithm is an update procedure for y_j .

The mean-shift segmentation algorithm is as follows:

Step 1: Two-dimensional image size is a map of the two-dimensional space and looks at each pixel by $x_i \in R^d$.

Step 2: Initialize the value $j = 1$ and $y_{ij} = x_i$.

Step 3: Calculate the value of $y_{i,j+1}$ according to Eq. (2) when it is approaching a certain point and then set $z_i = y_{i,c}$, where $y_{i,c}$ is the convergence value of y_{ij} .

Step 4: Imagine domain sharing $\{C_p\}_{p=1,\dots,m}$ by grouping together all z_i than w_1 to the l th function ($l = 1, 2, \dots, d$).

Step 5: For each $i = 1, 2, \dots, n$, assign $L_i = \{p | z_i \in C_p\}$.

Step 6: Neglect the spatial regions having less than N pixels.

2.2 Video Segmentation Using K-Means

The simplest algorithm for learning without follow-up are the K -means, which can solve cluster problems. This is one of the easiest ways to divide datasets provided by a number of clusters (assume k clusters) that are supposed to fix priority. The main purpose is to set up a wider group for all (k centroids). These centroids are placed in the way different locations lead to different results. Therefore, everything is put as far as possible to get a good result. So each point in the dataset is selected at one time and connects it to the nearest center. When each point is processed, the first step is complete. So, at this point, the centroid is recalculated by clusters that came from the previous result. When we have these new centroids, we have to create a new connection between the same data and the nearest new center. Since it is a method, a loop is created and uses this cycle; we can notice that centroids k change their position step by step to further changes. Finally, we minimized an objective function. The objective function is given as

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (5)$$

$\sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$ is the objective function within group I . x_i is a vector and c_j the corresponding cluster center.

The K -means segmentation algorithm is as follows:

Step 1: Consider the first vector of the K function in clusters.

Step 2: Sample vector is calculated using the minimum distance principle.

Step 3: New average is a new cluster for each cluster.

Step 4: Repeat step if the center is moved to Step 2, or stop.

2.3 Video Segmentation Using Fuzzy C-Means

Clear compilation plays an important role in problem-solving in recognizing patterns and sample identities and organizing video images. Different separation methods are provided by clusters, and most of them are based on distance criteria. The widely used algorithm is the C -means (FCM) delaying algorithm. This is useful when the cluster count is predetermined. In this way, the algorithms try to place each data point on a cluster. What makes FCM different is that it cannot resolve the absolute membership in the data points provided in the cluster. Instead, it calculates the probability

(membership) that the data point given will belong to a cluster. So depending on the accuracy of grouping required to exercise proper tolerance, it may be advisable. Since absolute membership is not calculated, FCM can be fast because of the number of changes required to achieve special components.

In the following FCM algorithm iteration, the following goals (objective function) are minimized:

$$J = \sum_{i=1}^N \sum_{j=1}^C \|x_i - c_j\|^2 \quad (6)$$

Here N is the number of data points, C is the necessary number of clusters, c_j is the central vector j and δ_{ij} is the membership level for the data point i in the j group. The $\|x_i - c_j\|$ standard measures the (or near) approximation of the x_i data point with the central vector c_j of j clusters. Note that in each iteration, the rules maintain a vector center for each cluster. These data are calculated on the average weight of the weighted data displayed by membership.

The fuzzy C -means segmentation algorithm is as follows:

Step 1: Start the membership-matrix U with values between 0 and 1.

Step 2: Calculate the center cluster c .

Step 3: Calculate the function that costs according to the equation. Stop if either it is below a certain tolerance value or its improvement, than the previous repetition is below certain threshold levels.

Step 4: Calculate the new U . Please go to Step 2

3 Performance Evaluation Metrics

The quality rating of the video segmentation algorithm is by comparing the result of the segment with part of the ground truth and obtaining the following statistics:

True Positive (TP): Intersection of pixels in the foreground of the segmented output and ground truth.

False Positive (FP): Intersection of pixels in the foreground of the segmented output and pixels in the background of ground truth.

True Negative (TN): Intersection of pixels in the background of the segmented output and ground truth.

False Negative (FN): Intersection of pixels in the background of the segmented output and pixels in the foreground of ground truth.

Using these statistics, assessments including FPR, FNR, Precision (Pr), Recall (Re), Accuracy (Ac), and Specificity (Sp) are calculated.

$$\text{FPR} = \frac{\text{number of FP}}{(\text{number of FP} + \text{number of TN})}$$

$$\begin{aligned} \text{FNR} &= \frac{\text{number of FN}}{(\text{number of FN} + \text{number of TP})} \\ \text{Precision(Pr)} &= \frac{\text{number of TP}}{(\text{number of TP} + \text{number of FP})} \\ \text{Recall(Re)} &= \frac{\text{number of TP}}{(\text{number of TP} + \text{number of FN})} \\ \text{Specificity(Sp)} &= \frac{\text{number of TN}}{(\text{number of FP} + \text{number of TN})} \\ \text{Accuracy(Ac)} &= \frac{\text{number of TP} + \text{number of TN}}{(\text{number of TP} + \text{TN} + \text{FP} + \text{FN})} \end{aligned}$$

These above parameters describe the performance [20–22] of the moving object.

4 Experimental Results and Discussion

Analysis of video segment algorithms has been tested and implemented using Matlab. We have selected some of the testing segments, the standard video clips called CDnet 2014 [23] in the section of the video. These sets of videos differ in size with corresponding ground truth. The mean-shift, K -means, and fuzzy C -means segmentation algorithms are applied over the above dataset and got the segmentation result.

In mean-shift algorithm, we have taken the `spatial_bandwidth` = 40, `color_bandwidth` = 3 and convergence factor = 3. Different values of bandwidth W can be taken for the performance measurement.

In K -means algorithm, we have taken the cluster number = 8 and we choose the cluster center randomly.

In fuzzy C -means algorithm, we choose the cluster center randomly (Fig. 1).

To measure the quality of the segmentation algorithms, we calculate different metric measurements [20–22]. The metrics were reported for CDnet for 2014 [23]. We have provided a comparative analysis of mean-shift, K -means, and fuzzy C -means clustering-based segmentation. We can see that for baseline category, the accuracy is more for K -means method. The baseline and shadow categories consist of simple sequences where the normal order of pedestrian and car is the focus. With different types of fraud and variation related to Intern.Object motion, fuzzy C -means method gives very good recall compared to other. In shadow category, the accuracy is more than 90% for mean-shift segmentation. But the recall is high for fuzzy C -means method. In dynamic background category, the accuracy is more than 90% both in mean-shift and K -means methods but recall is more in fuzzy C -means method (Tables 1, 2, 3 and 4).

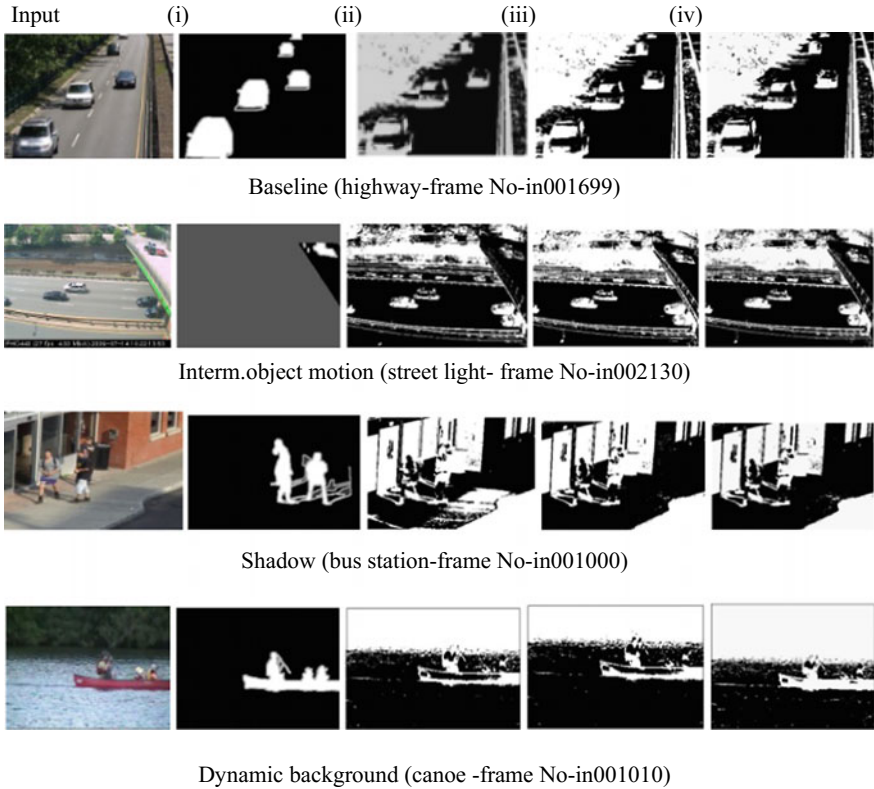


Fig. 1 Segmentation results obtained for various sequences of the CDnet dataset: (i) ground truth maps; (ii)–(iv) output results of mean-shift, K -means, and fuzzy C -means methods, respectively

Table 1 Performance measures (FPR, FNR, Pr, Re, Ac, Sp) of different methods for baseline image

Methods	FPR	FNR	Pr	Re	Ac	Sp
Mean-shift	0.0562	0.9389	0.0161	0.0611	0.9307	0.9438
K -means	0.0168	0.9864	0.0123	0.0136	0.9685	0.9832
Fuzzy C -means	0.8484	0.1469	0.0145	0.8531	0.1618	0.1516

Table 2 Performance measures (FPR, FNR, Pr, Re, Ac, Sp) of different methods for Interm. Object motion image

Methods	FPR	FNR	Pr	Re	Ac	Sp
Mean-shift	0.1642	0.8929	0.1200	0.1071	0.7099	0.8356
K -means	0.1481	0.7692	0.2000	0.2308	0.7660	0.8519
Fuzzy C -means	0.3523	0.3333	0.2051	0.6667	0.6500	0.6477

Table 3 Performance measures (FPR, FNR, Pr, Re, Ac, Sp) of different methods for shadow image

Methods	FPR	FNR	Pr	Re	Ac	Sp
Mean-shift	0.0508	0.9067	0.0505	0.0933	0.9251	0.9492
<i>K</i> -means	0.0535	0.9385	0.0307	0.0615	0.9227	0.9465
Fuzzy <i>C</i> -means	0.8768	0.1095	0.0264	0.8905	0.1431	0.1232

Table 4 Performance measures (FPR, FNR, Pr, Re, Ac, Sp) of different methods for dynamic background image

Methods	FPR	FNR	Pr	Re	Ac	Sp
Mean-shift	0.0531	0.9620	0.0131	0.0380	0.9304	0.9469
<i>K</i> -means	0.0566	0.9447	0.0182	0.0553	0.9269	0.9434
Fuzzy <i>C</i> -means	0.8599	0.1554	0.0157	0.8446	0.1514	0.1401

5 Conclusion

In the process of visualizing a video and a computer, one of the most difficult and most active areas of the research is the part of the video material. One of the most important issues for successful use of successful video sequences is the video segment that focuses on dividing video files into essential video frames. Various techniques are being used by the developers in reviewed. We analyzed three video segmentation algorithms based on the discovery of the changes in the video clip. We have used mean-shift, *K*-means, and fuzzy *C*-means clustering-based segmentation algorithm for video segmentation and measured the performance evaluation metrics. Mean-shift and *K*-means methods provide good accuracy in comparison to fuzzy *C*-means. But the recall is generally more in case of fuzzy *C*-means method.

References

1. Jiang H, Zhang G, Wang H, Bao H (2015) Spatio-temporal video segmentation of static scenes and its applications. *IEEE Trans Multimed* 17(1)
2. Wu GK, Reed TR (1999) Image sequence processing using spatiotemporal segmentation. *IEEE Trans Circ Syst Video Technol* 9(5):798–807
3. Kim EY, Hwang SW, Park SH, Kim HJ (2001) Spatiotemporal segmentation using genetic algorithms. *Pattern Recognit* 34(10):2063–2066
4. Koprinska I, Carrato S (2001) Temporal video segmentation: a survey. *Signal Process Image Commun* 16(5):477–500
5. Megret R, Dementhon D (2002) A survey of spatio-temporal grouping techniques. In: *Language and media process*, University of Maryland, College Park, MD, USA, Tech. Rep. LAMP-TR-094/CS-TR-4403
6. Kumar MP, Torr PHS, Zisserman A (2008) Learning layered motion segmentations of video. *Int J Comput Vis* 76(3):301–319
7. Shi J, Malik J (1998) Motion segmentation and tracking using normalized cuts. In: *Proceedings of the ICCV*, pp 1154–1160

8. Fowlkes C, Belongie S, Malik J (2001) Efficient spatio temporal grouping using the nystrom method. In Proceedings of the CVPR, pp 231–238
9. Khan S, Shah M (2004) Object based segmentation of video using color, motion and spatial information. In: Proceedings of the CVPR, vol 2, pp 746–750
10. Remers D, Oatto S (2003) Variational space-time motion segmentation. In: Proceedings of the ICCV, pp 886–893
11. Itnick CL, Jojic N, Kang SB (2005) Consistent segmentation for optical flow estimation. In Proceedings of the ICCV, vol 2, pp 1308–1315
12. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22(8):888–905
13. Comaniciu D, Meer P, Member S (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5):603–619
14. Mobahi H, Rao S, Yang AY, Sastry SS, Ma Y (2011) Segmentation of natural images by texture and boundary compression. *Int J Comput Vis* 95(1):86–98
15. Sharon E, Galun M, Sharon D, Basri R, Brandt A (2006) Hierarchy and adaptively in segmenting visual scenes. *Nature* 442(7104):719–846
16. Lim YW, Lee SU (1990) On the color image segmentation algorithm based on the thresholding and the fuzzy *c*-means techniques. *Pattern Recognit* 23(9):935–952
17. Hance GA, Umbaugh SE, Moss RH, Stoecker WV (1996) Unsupervised color image segmentation with application to skin borders. *IEEE Eng Med Biol*, 104–111
18. Devikar MM, Jhac MK (2013) Segmentation of images using histogram based FCM clustering algorithm and spatial probability, Department of Telecommunication Engineering, CMRIT, Bangalore, India. *Int J Adv Eng Technol*
19. Ali SM, Abood LK, Abdoon RS (2013) Clustering and enhancement methods for extracting 3D brain tumor of MRI images. Remote Sensing Research Unit, Department of Computer Science, University of Baghdad, Department of Physics, University of Babylon, Volume 3, Issue 9
20. Senior A, Hampapur A, Tian Y, Brown L, Pankanti S, Bolle R (2000) Appearance models for occlusion handling. In: Proceedings of the 2nd IEEE workshop performance evaluation of tracking and surveillance
21. Erdemand CE, Sankur B (2000) Performance evaluation metrics for object based video segmentation. In: Proceedings of 10th European Signal Processing Conference, vol 2, pp 917–920
22. Marichal X, Villegas P (2000) Objective evaluation of segmentation masks in video sequences. In: Proceedings of 10th European Signal Processing Conference, vol 4
23. Wang Y, Jodoin P-M, Porikli F, Konrad J, Benezeth Y, Ishwar P (2014) CDnet 2014: an expanded change detection benchmark dataset. In: Proceedings of the IEEE conference on workshops of computer vision and pattern recognition (CVPR), pp 387–394