

FOREGROUND OBJECT SEGMENTATION FROM VIDEO SEQUENCES USING SPEEDED-UP ROBUST FEATURE

¹Indhu R, ²Hemanandhini S, ³M.Akila

KPR Institute of Engineering and Technology

¹indhu.r@kpriet.ac.in, ²hemanandhini.kpriet.ac.in, ³akila.m@kpriet.ac.in

Abstract: Video Object Segmentation is one of the most imperative tasks in recent times. It is used in many applications such as object detection, object tracking, surveillance monitoring and behavior analysis of moving object in a video sequence. Unsupervised video object segmentation is to automatically segment the foreground objects in the video without prior knowledge. The main objective of this paper is enabling the object segmentation by extracting the object of interest from a series of successive video frames. Segmenting objects manually requires a tool and lot of time and user needs to select an object area which is to be segmented. In this paper, an unsupervised approach for foreground object segmentation has been proposed by generating a saliency map. Saliency map is computed based on contrast cue and spatial cue. Global cues are computed by employing kmeans clustering algorithm to cluster individual pixel. Finally foreground object is segmented based on cluster, global cues and saliency map.

Keywords: Object Segmentation, surveillance monitoring, object tracking, Saliency map, contrast cue and spatial cue.

1. Introduction

The incredible growth of vision based application in recent years such as video surveillance, content based video analysis, traffic monitoring and analysis of sports video, etc., has many problems in segmenting object of interest. Segmentation of object of interest from a video sequence is one of the basic and essential operation in computer vision. Video object segmentation has been defined as the problem of sub dividing the video sequence into coherent regions with respective to features and properties of the object. Foreground object segmentation is useful for several computer vision tasks including video analysis, object tracking, object recognition, video retrieval and activity recognition.

A common issue in the Video Foreground Object Segmentation is categorized in two. Supervised method

of object segmentation in which user needs to select an object area which is to be segmented. Unsupervised method of object segmentation which extracts the object without having any prior knowledge about the object. There are many approaches for supervise object segmentation whereas intensively less approaches focusing on unsupervised method of object segmentation. Unsupervised method of object segmentation is more trivial and it can be applied in many fields such as video analysis, video indexing, video retrieval and activity understanding.

Recently many approaches has been developed to extract the foreground object. Gu et al. [1] and Wang et al. [2] proposed a video scene segmentation approaches with content coherence and contextual dissimilarity. Brox and Malik [3] propose a method based on foreground object trajectory. A series of trajectories are extracted and then processed by spectral clustering. Ochs and Brox [4] extend the trajectory method to acquire a denser foreground region. Lee et al. [5] showed that these methods wants an explicit notion of how a foreground object should look like in video data and low level grouping of pixels, which results in over segmentation. Visual saliency has been used to form the foreground object segmentation. [6] This Method explore object-based segmentation in static image and achieve significant progress. Those methods generate multiple object hypotheses and rank hypotheses according to their scores. The model is learned for the generic foreground object using image features such as color, texture and boundary which then object category independent.

In this paper, a saliency based method to segment the foreground object from the unlabeled data has been proposed. First key frames are generated by using the method [7]. Then color, texture features are extracted and given as input to kmeans algorithm. Based on the cluster generated by kmeans clustering algorithm contrast cue and spatial cues are computed and element wisely multiplied to generate the saliency map based on the method [8]. Saliency map shows the foreground object which is need to be segmented.

2. Related Works

The most contemporary research done by Xiaochuncao et al [9] and Zhang et al [10] proposed a method to extract foreground object automatically without any prior knowledge about the object. First object like regions are segmented as an object candidate, then based on the corresponding map between the objects in the successive frames the object segmentation is converted into graph model. Correspondence map is calculated between objects in successive frames based on their appearance and motion. Then the shortest path algorithm explored to get global optimum solution from the graph model. GMM is applied in iterative manner to produce more refine results, when the GMM is matched with the results, then the iteration gets stopped and returns the foreground object.

Lei fan et al [11], which focuses video segmentation is a task of grouping pixels in successive video frames into perceptually coherent regions is achieved. Video segmentation is a useful tool for extracting relevant and interesting content from a video sequence for analysis and understanding. This is graph based approach which effectively models the data in a high dimensional feature space, which emphasizes the correlation between similar pixels while reducing the inter class connectivity between different objects. The graph model combines appearance, spatial and temporal information to break a large video sequence into semantic spatio temporal key segments. Continuous grouping of video sequence able to extract a moving object of interest from a video sequence based on its unique properties.

Rother et al [12] addressed the problem of an efficient extraction of a foreground object in a problematic situation background cannot be subtracted easily. GrabCut is an extension of the graph cut approach it has iterative version of optimization. The power of the iterative algorithm is used to simplify the user interaction needed for a given quality of result.

Ma and Latecki [13] attempt to address this video object segmentation problem by utilizing relationships between object proposals in adjacent frames. The object region candidates are selected simultaneously to construct a weighted region graph.

Dong Zhang et al [14], proposed a novel approach to extract primary objects in videos. The extracted primary object regions are then used to build object models for optimized video segmentation. DAG based framework is explored in this approach for detection and segmentation of the primary object in video. Objects are spatially cohesive and characterized by locally smooth motion trajectories. To extract the primary object from the set of all available proposals motion, appearance and shape properties are used.

3. Method

The problem of finding moving foreground objects in an unsupervised manner is a tedious task in surveillance monitoring, traffic monitoring system videos and also in moving object in a static background video. Low-level segmentation is done in previous approaches whereas in this approach object level segmentation is done by extracting keyframes. All keyframes are processed to extract feature vector based on color and texture. Then kmeans algorithm is employed to cluster the pixel based on the feature vector. Cluster map is generated based on clustering algorithm. Contrast and Spatial cues are computed to generate a saliency map. Finally foreground object is segmented based on the saliency map generated.

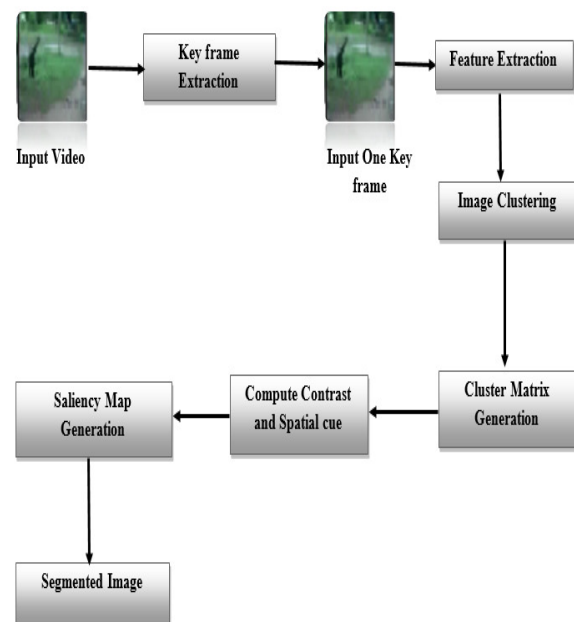


Figure 1. Proposed Architecture

The Fig.1 shows the proposed architecture for segmenting foreground objects from video in an unsupervised method of object segmentation. In this paper the system is implemented using FBMS dataset. It is an extension of the BMS dataset with 33 additional video sequences. A total of 720 frames is annotated.

A. Keyframe Extraction

Keyframe is the reference frame of the video sequence which is representative and able to reflect the summary of a video content. The key-frame is used to express the main content of video data clearly and reduce the amount

of memory needed for video data processing and complexity greatly. Video data which contains multiple shots it is necessary to identify individual shots for key-frame extraction. Visual features and temporal information are used to determine the keyframe. Key frames are extracted to reduce the number of frames in the given video sequence. Key frames are extracted based on setting a threshold value and comparing all frames with the threshold value, frames which satisfies the threshold condition are considered as key frame. Threshold is computed using Eq. (1).

$$T = \mu + \sigma \quad (1)$$

Where, μ and σ is mean and standard deviation of absolute difference. The following are the steps for key frame extraction.

Steps

1. Extract frames one by one
2. Compute histogram difference between two consecutive frames
3. Calculate mean and standard deviation of absolute difference
4. Compute threshold
5. Compare the mean with threshold
6. If it is greater than threshold set it as a keyframe
7. Continue till end of a video.

B. Feature Extraction

Feature extraction includes minimizing the amount of resources used to denote a huge set of data. When performing analysis of complex data one of the major problems is the number of variables involved. Exploring with a huge number of variables generally requires a more amount of energy and power. This process is contains a feature vector containing color and texture features. Gabor filter is employed to get a feature vector with texture features. Distance vector for each pixel is also calculated and original key frame is resized for further processing. Distance vector is calculated based on the Euclidean distance. The Euclidean distance is given in the Eq. (2).

$$d(p, q) = \sqrt{(q_1 - p_1)^2} + \sqrt{(q_1 - p_1)^2} \dots + \sqrt{(q_1 - p_1)^2} \quad (2)$$

Where, p and q are Euclidean vectors.

C. Image Clustering and Cluster Matrix Generation

Image clustering is the task of grouping individual pixels to the corresponding cluster based on the color, texture feature of pixels and distance vector is also calculated to

find out the centroid of the image cluster. Image clustering can be achieved using many algorithms for grouping similar pixels in an unsupervised manner. In this paper Centroid based clustering is done to group the image pixels based on the feature vector and distance vector. Kmeans algorithm is applied to cluster the image. The following are the steps of Kmeans algorithm.

Steps

- 1: select k .
- 2: select k points as an initial centroid.
- 3: From initial centroid and observation value calculate Euclidean distance.
- 4: Assign to the cluster based on the maximum distance.
- 5: Update the initial centroid based on the cluster assignment.
- 6: Again compute Euclidean distance for updated centroid and next observation.
- 7: Repeat the steps until all observations fit to the k cluster.

Cluster matrix is generated using Kmeans clustering algorithm. It is one of the simplest unsupervised clustering algorithm and its steps shown above. The grouping is done by minimizing the sum of squared distance between items and corresponding centroid. Euclidean distance for kmeans algorithm is calculated using Eq. (2) and cluster centroid is updated using Eq. (3)

$$cp(x_1, x_2, \dots, x_n) = \frac{\sum_{j=1}^n x_1}{n}, \frac{\sum_{j=1}^n x_2}{n}, \dots \quad (3)$$

D. Computation of Contrast and Spatial Cues

The contrast cue and spatial cue are called as global cues. The biological vision system is sensitive to contrast in visual signaling. So saliency of the pixels as its color contrast to all other pixels in the image is well defined in Eq. (4).

$$S_{cc}(I_k) \propto \sum_{\forall I_i \in I} D(I_k, I_i) \quad (4)$$

Where $D(I_k, I_i)$ is the color difference metric between pixels I_k and I_i in Lab color space. Pixels with same color have the same saliency value based on the equation defined in color space. To compute same color saliency pixel value another equation is defined in Eq. (5).

$$S_{cc}(I_k) \propto \sum_{j=1}^n f_i D(C_i, C_j) \quad (5)$$

where c_i is the color value of pixel I_k , n is the number of distinct pixel colors and f_j is the occurring probability of pixel color c_j in the image I . The color difference metric is proposed as a Gaussian system as given in Eq. (6).

$$D(C_i C_j) = 1 - e^{-\frac{d(C_i C_j)^2}{2\sigma}} \quad (6)$$

Spatial cue information can be used to distinguish foreground and background in the given frame. The pixels of salient object often lays more compactly and the pixels of background often distribute more dispersedly. The colors distributing more compactly would have higher probability belonging to the salient object. Spatial cue is defined in Eq. (7).

$$U_{sv}(I_k) \propto \frac{1}{m} \sum_{j=1}^n (x_{cl}^{(i)} - x_{cl})^2 + (y_{cl}^{(i)} - y_{cl})^2 \quad (7)$$

where cl is the color value of pixel I_k , m is the number of pixels that have color value $(x_{cl}^{(i)}, y_{cl}^{(i)})$ is the coordinate of one of these pixels and x_{cl}, y_{cl} is the average coordinate of these pixels. $U_{sv}(I_k)$ represents the degree of non-salient.

E. Saliency Map Generation and Object Segmentation

Saliency map is generated to segment the foreground object. Saliency map is generated by computing spatial cue and contrast cue. Both cues are computed and element wise multiplication is done to generate the saliency map. Elementwise multiplication notation is given in Eq. (8).

$$z = x.*y \quad (8)$$

4. Experimental Analysis

The dataset used for the development of the proposed system is Freiburg-Berkeley Motion Segmentation Dataset. It is an extension of the BMS dataset with 33 additional video sequences. A total of 720 frames is annotated. The data set contains ground truths and preprocessed video frames. This Freiburg-Berkeley Motion Segmentation Dataset is an open source and publically available for all research purposes. The dataset has been invented by scholars of the University of Freiburg

A. Key frame Extraction

Keyframes are extracted to reduce the number of frames in the given video sequence. Keyframes are extracted based on setting a threshold value and comparing all frames with the threshold value, frames which satisfy the threshold condition are considered as the key frame. The Fig. 2 and Fig. 3 shows the key frame extraction and it stored in a separate folder.



Figure 2. Threshold Calculation

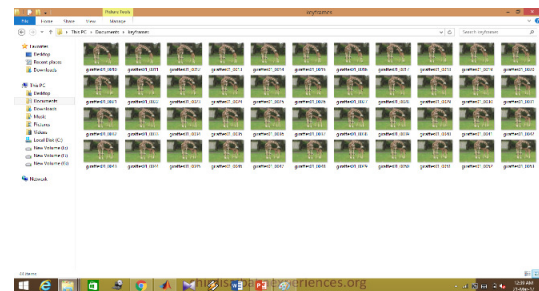


Figure 3. Extracted Frames are stored in Separate Folder

B. Feature Extraction and Image Clustering

This process contains a feature vector containing color and texture features. Gabor filter is employed to get a feature vector with texture features. Distance vector for each pixel is also calculated and original key frame is resized for further processing.

Image clustering is done using Kmeans algorithm. This algorithm clusters n objects based on attributes into k -partitions, where $k < n$. cluster bin number is given to cluster the pixels. The grouping kmeans algorithm is done by maximizing the sum of squared distance between the items and the corresponding initial centroid.

C. Cluster Matrix Generation

Cluster matrix is generated to cluster each and every individual pixel based on the kmeans algorithm. Based on the number of bins and feature vector, cluster matrix is generated.

D. Computation of Cues and Saliency Map Generation.

The contrast cue is computed based on the Eq. (4) and it is normalized between 0 and 1 based on the Gaussian normal distribution.

The spatial cue is computed on the Eq. (7) and it is also normalized between 0 and 1 based on the Gaussian normal distribution. It is computed based on cluster index, bin number, distance vector and width of the frame. And it is stored in an array.

Saliency map is generated based on the contrast cue and spatial cue. The global cue are combined to produce a saliency weight array for generating saliency map. Cluster map, saliency weight array and bin number are given as an input to generate saliency map.

E. Object Segmentation

The object segmentation is done based on the saliency map generated and it is also normalized to produce fine results of object segmentation. The Fig. 4 show the single object segmentation, by computing contrast and spatial cue. Based on the cuesingle saliency map is generated to segment the object.



Figure 4. Single Object Segmentation

The Fig.4shows the multiple object segmentation on sequential keyframes. Contrast and spatial cues for all frames are computed and combined element wisely to generate a saliency map. Based on the saliency map generated object is segmented on sequential keyframes.

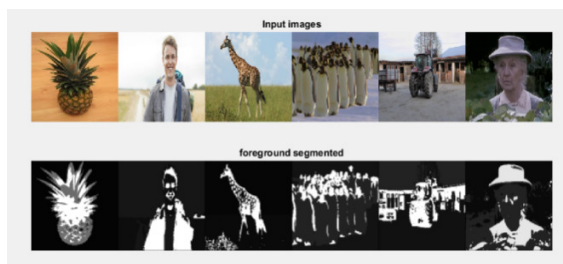


Figure 5. Multiple Object Segmentation

5. Evaluation

The dataset used for the development of the proposed system is Freiburg-Berkeley Motion Segmentation Dataset. It is an extension of the BMS dataset with 33 additional video sequences. Dataset is evaluated based on

proposed algorithm. The Fig. 5.5 shows the performance evaluation for the dataset which has been used for developing the proposed system. It also shows that the proposed system produces average precision, recall and fscore of 83%, 92% and 87 %. The accuracy of the system is about 95.43%.

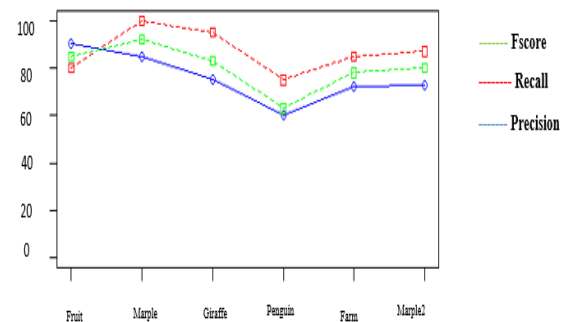


Figure 6. Performance Evaluation

The proposed system is evaluated on publically available dataset ECSSD and MSRA and their performance and results are given in Fig .6, 7, 8, 9,10.

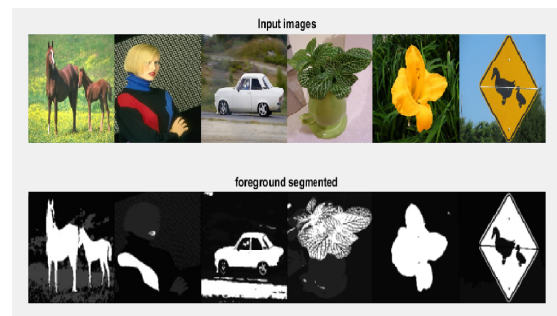


Figure 7. Results of ECSSD Dataset

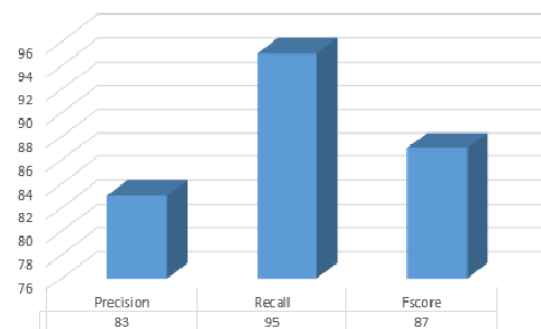


Figure 8. Evaluation of ECSSD Dataset

The results and evaluation of ECSSD Dataset shows that the system produces precision of 83%, recall of 95%, Fscore of 87% and accuracy of 94.01%.

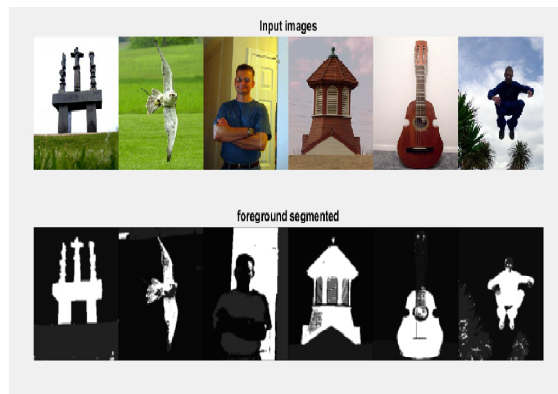


Figure 9. Results of MSRA Dataset

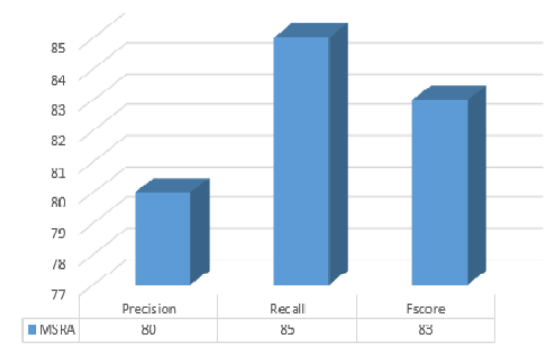


Figure 10. Results of MSRA Dataset

The result and evaluation of MSRA Dataset shows that the system produces precision of 80%, recall of 85%, Fscore of 83% and accuracy of 91.05%. The output and evaluation results show that the proposed system works well for all state of art dataset and produces most accuracy for all dataset.

6. Conclusion

An Unsupervised approach for video foreground object segmentation system is designed in such way that it provides efficient foreground object segmentation. The result shows that segmentation is done based on the saliency-based clustering method. An effective clustering method is involved to avoid serious learning. The contrast and spatial cues are computed effectively for the given dataset to segment the foreground object. The video foreground object segmentation approach is an

automatic and effective preprocessing step for many vision applications.

References

- [1] Z.Gu, T.Mei, X. S.Hua, X.Wu, S.Li, Ems: Energy Minimization Based Video Scene Segmentation, ICME, pp. 520–523, 2007.
- [2] J.Wang, X.Tian, L.Yang, Z.J.Zha, X.S.Hua, Optimized video scene segmentation, IEEE International Conference on Multimedia and Expo, pp.301–304, 2008.
- [3] T.Brox, J.Malik, Object segmentation by long Term Analysis of Point Trajectories, ECCV, pp.282–295, 2010
- [4] P.Ochs, T.Brox, Object segmentation in video: a hierarchical variational approach forturning point trajectories into dense regions, ICCV, pp. 1583–1590, 2011.
- [5] Y.J.Lee, J.Kim, K.Grauman, Key segments for video object segmentation, ICCV, pp.1995–2002, 2011.
- [6] I.Endres, D.Hoiem, Category independent object proposals, ECCV, pp. 575–588, 2010.
- [7] Sheena C Va, N. K. Narayanana, Key-frame Extraction by Analysis of Histograms of Video Frames using Statistical Methods, 4thInternational Conference on Eco-friendly Computing and Communication Systems, Procedia Computer Science, vol. 70, pp. 36 40, May 2015.
- [8] Huazhu Fu, Xiaochun Cao, and ZhuowenTu, Cluster-Based Co-Saliency Detection, IEEE Transactions On Image Processing, vol. 22, no. 10, October 2013.
- [9] XiaochunCao, FengWang, BaoZhang, HuazhuFu, ChaoLi, Unsupervised Pixel-Level Foreground Object Segmentation via ShortestPath Algorithm, Elsevier, Neurocomputing172, pp. 235-243, May 2015.
- [10] Bao Zhang, Handong Zhang, Xiaochun Cao, Video Object Segmentation with Shortest Path, ACM Transactions, pp. 801–804, November 2012.
- [11] Lei Fan and Alexander C. Loui, A Graph-based Framework for video object segmentation and extraction in feature space, Multimedia (ISM), IEEE International conference, December 2015.

- [12] Carsten Rother, Vladimir Kolmogorov and Andrew Blake, GrabCut-Interactive Foreground Extraction Using Iterated Graph Cuts, ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, vol. 23, pp. 309–314, August 2004.
- [13] T.Ma, L.J.Latecki, Maximum weight cliques with mutex constraints for video object segmentation, CVPR, pp.670–677, 2012.
- [14] Dong Zhang, Omar Javed and Mubarak Shah, Video Object Segmentation through Spatially Accurate and Temporally Dense Extraction of Primary Object Regions, IEEE Transactions, vol. 32, pp. 628–635, March 2013.
- [15] K. Ramash Kumar, D.Kalyankumar, DR.V.Kirbakaran” An Hybrid Multi level Inverter Based DSTATCOM Control, Majlesi Journal of Electrical Engineering, Vol. 5. No. 2, pp. 17-22, June 2011, ISSN: 0000-0388.
- [16] S.V.Manikanthan and T.Padmapriya “Recent Trends In M2m Communications In 4g Networks And Evolution Towards 5g”, International Journal of Pure and Applied Mathematics, ISSN NO: 1314-3395, Vol-115, Issue -8, Sep 2017.
- [17] S.V.Manikanthan and K.srividhya "An Android based secure access control using ARM and cloud computing", Published in: Electronics and Communication Systems (ICECS), 2015 2nd International Conference on 26-27 Feb. 2015, Publisher: IEEE, DOI: 10.1109/ECS.2015.7124833.
- [18] T. Padmapriya, V.Saminadan, “Performance Improvement in long term Evolution-advanced network using multiple input multiple output technique”, Journal of Advanced Research in Dynamical and Control Systems, Vol. 9, Sp-6, pp: 990-1010, 2017.

