An Efficient Method for Object Segmentation in Video Scenes

Kanagamalliga.S[1],Vasuki. S[2], Manjupriya.C[3], SubhaDharshini.B[4]
Departmentof Electronics and Communication Engineering,
Viraganoor,Madurai 625009.

***Abstract*-- The system present a coherent, frame work for tracking multiple objects. A video locales is a sequences of image features that share similar features in the spatial-temporal domain videos. The main objective of the paper to track region of interest using segmentation. MRF&SLIC algorithm for the effective segmentation. (i.e) introduce new super pixel & MRF structures. By improving the segmentation accuracy, the accuracy of the objects tracking improved. It embeds the looks as auxiliary nodes and edges in the MRF structure can enhance the segmentation in one graph cut. In the addition experimental evaluation validate the superiority of the proposed approach over the state-of-the art methods in both efficient and effectiveness.**

Keywords: Object, Segmentation, MRF, SLIC, Super Pixel, Graph Cut.

I. INTRODUCTION

Efficient detection in videos has attracted increasing interest recently. Object detection is a challenging problem on account of scene complexity, camera motion, and action variability (the same action performed by different people may look quite different).Also, most video analysis applications, such as surveillance, require high computational efficiency. The object in a video sequence can be defined as the object that is locally present in most of the frames [12]. The target of video object segmentation is to segment out the primary object in a video sequence without any human intervention.Some examples are shown in Fig. a. The existing works of video object segmentation can be divided into two groups based on the amount of human intervention required: interactive segmentation [3], and fully automatic segmentation [17].

Succeeding the performance of Markov Random Field (MRF) based methods in image object segmentation [6], [13], many of the existing video object segmentation approaches also build spatio-temporal MRF graphs and show positive results [5].These approaches build a comprehensive graph by connecting spatially or temporally connected regions, *e.g.*, pixels [17]or super pixels [18], and cast the segmentation problem into a node labeling problem in a Markov Random Field.Such automatic video object segmentation methods: initial visual, comprehensive graph connection and foreground/background appearance modeling. Formally, with the presence of appearance restriction, there are two groups of segmentation labels xand appearance modelϕ.Commonly used appearance models such as Gaussian Mixture Models (GMM), it is inflexible to solve both parameters simultaneously. According to, many existing methods adapt an repetitive approach.

Recently, proposed a technique for appearance modeling by the graph based interactive object segmentation framework which can cure both the segmentation labels and appearance model parameters concurrently without iteration. In their method, they model each pixel as a node and quantize to a bin in the RGB histogram. It shows that adding equivalent auxiliary nodes and edges to the original MRF structure.

For video object segmentation super pixels are generally used due to the big data volume and more powerful features like SIFT or Textons are beneficial to better capture the viewpoint and lighting variations between different frames.Extend the efficient appearance modeling technique in [18] to video object segmentation solve by these challenges. The proposed appearance modeling technique is more usual than [18].The resultant auxiliary joints are also different from [2] because each super pixel node is connected to one auxiliary node. Experimental evaluations validate the superiority of the proposed approach over directly applying [18] for automatic segmentation.

In summary, the major contribution is that propose an efficient and effective appearance modeling technique in the MRF based segmentation framework for video object segmentation. It embeds the appearance constraint directly into the graph.so, the resultant graph-partition problem can be solved efficiently by one graph cut. The organization of the paper is given below. The next section describes the related work done with respect to the proposed method. Section III explains the overall methodology. Section IV presents overview of the experiment and the integration done. Section V addresses the results and discussions obtained from the algorithms conclusion and future work.

II RELATED WORK

A. Low level object segmentation

Low level video segmentation methods includesuper pixel segmentation [1], and super voxel segmentation.Super voxel segmentation is similar to super pixelsegmentation but also groups pixels temporally. Hence, itproduces spatio-temporal segments. Actually, super pixelsand super voxels are usually used as the primitive input inplace of pixels in the context of video object segmentation forefficiency [18].

Another type of low level segmentationis object proposal segmentation Many highlevel video object segmentation methods use these proposalsas the original input [17].

B. Object level video segmentation

The existing works related to video object segmentation[5] are given below,*i.e.*, interactive segmentation, automatic segmentation and video object co-segmentation. These approaches require theuser to provide a pixel-wise segmentation on the first fewframes for initialization[15],while othersrequire the user to continuously correct the segmentationerrors [3], [9].The most related approach is [10] as it alsorelies builds spatio-temporal graphby connecting neighborhood superpixels. Several papers [5], [17], use object proposal are the primitive input which contribute significantly to the inefficiency of these methods. The method in [19] first uses spectral clustering to group proposals with consistent appearance and then train foreground/background color GMMs and object location earlier. Pixel-wise graph cut is used to produce the final segmentation mask for each individual frame.

The method in [12] explores this problem in MPEG2 compressed domain. On the I-Frames, it computes the color-based segmentation by morphological approach. Video object co-segmentation is also automatic supervision by assuming the primary object is present in a batch of given videos [17]. Both [8] and [12] formulate the segmentation as node selection or labeling in spatio-temporal graph, while finds the maximum weighted clique in a completely connected graph. The method does not have an explicit global appearance model, and adapts the iterative appearance modeling.

C. MRF segmentation Framework

In the existing image or video object segmentation frameworks using MRF structure, the most commonly used appearance model is color GMM which models the foreground and background appearances separately [17], [5], [13]. Multiple instance learning on context features is also used to model the foreground and background appearance in a discriminative manner. Recently, [18] proposed to use color histograms to model the appearance non-parametrically for static image segmentation.

III. PROPOSED METHODOLOGY

Pre-processing consists of computing track-lets and computing frames are occurred by the single input video. Frame conversion is the process of converting the single video into the several number of images. By the frame conversion method, need not to process the video directly in to the process. So that the process is done by the image processing.Laplace Filter technique used for modifying or enhancing an image. For example, you can filter an image to emphasize certain features or remove other features.

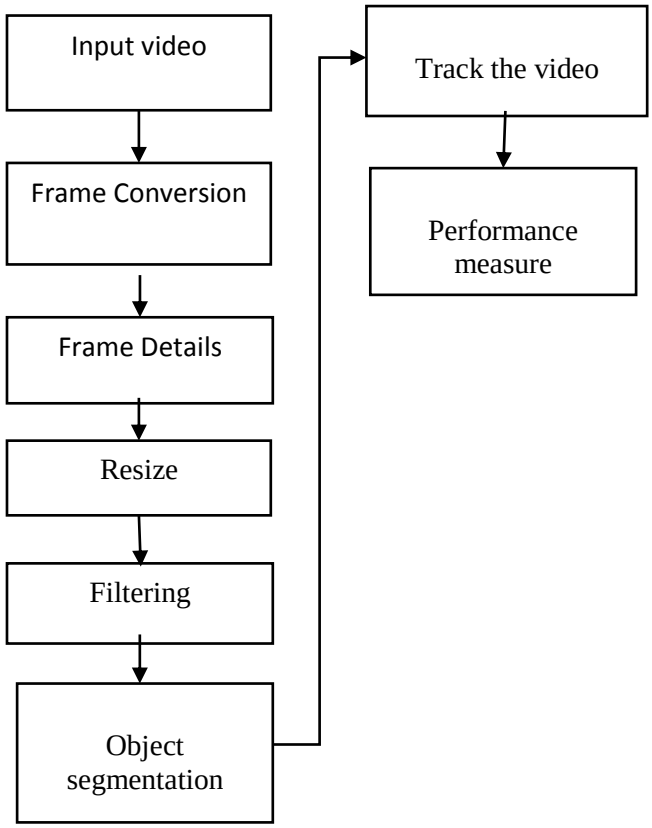


Fig. 1. Flow Diagram of proposed method

Filtering is a neighborhood operation, in which the value of a given pixel in the output image is determined by applying algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. To begin with, generate a set of match hypotheses for track-let association and a likely set of tracks the video by the super pixel segmentation technique. An observation potential is computed for each track-let using the features computed at the region of contour. Track-lets are grouped into activity segments using a standard baseline of the region present inside the bounding box.

Efficient and effective appearance modeling technique in the MRF based segmentation framework for primary video object segmentation in Fig. 1. It embeds the appearance constraint directly into the graph. Each pixel will now have multiple features and each node will correspond to multiple pixels.

IV.OVERVIEW OF SYSTEM MODULES AND INTEGRATION

Introduce the proposed approach for automatic primary video object segmentation. The input is a plain video clip without any annotations and the output is a pixel-wise spatio-temporal foreground vs. background segmentation of the entire sequence. Similar to many existing image and video object segmentation approaches, cast the segmentation to a two-class node labeling problem in a Markov Random Field. Within the MRF graph, each node is modeled as a super pixel, and will be labeled as either foreground or background in the segmentation process. The overall work flow is shown in Fig.

2.first segment each video frame into a set of super pixels using the SLIC algorithm [1] and then represent each node in theMRF as a super pixel. Meanwhile super pixels produced by SLIC [1] can preserve most of the boundaries, and over-segmentation is not a critical concern.

In the following, uses$s_{i,}^{j}$ to denote the j^{th} super pixel of i^{th}frame, N to denote the total number of frames and M^{i} to denote the number of super pixels in the i^{th}frame. The segmentation target is to assign each super pixels$s_{i,}^{j}$ label x_i^j indicating if it is foreground,x_i^j= 1, or background, x_i^j= 0.The overall optimization formulation in terms of the graphenergy minimization is expressed as

$$x* = \arg\min_{x,\theta} \min E(s,x,\theta) \quad (1)$$

where E(s, x,θ) is defined as,

$$E(s,x,\theta) = \alpha_p * \varphi_p(s,x) + \alpha_a * \varphi_a(s,x,\theta) \quad (2)$$

The vector x and θ denote the {0, 1} labeling of all the Super pixels and the appearance model parameters, respectively, s denotes the collection of all the super pixels and φp and φa denote pairwise potential and appearance constraint potential, respectively. $\alpha_a$andα_p are two weight parameters for linear combination.

A. Pairwise Potentials

There are two types of neighborhood relationships between superpixels in videos, *i.e.*, spatial neighborhoods and temporal neighborhoods. Two superpixels are spatially connected if they share a common edge and temporally connected if they have pixels linked by optical flow. In the MRF graph, only neighboring superpixels will have nonzero edge and the edge weight represents the cost induced by assigning different labels to the connected superpixels. Hence, the edge weight is usually measured as the inverse likelihood of the existence of a real edge between two superpixels. More specifically, ituse color and optical flow orientation histogram to compute the local similarity and the structural forest edge detector [7] to compute the edge strengths. To detect motion boundaries for each frame, first convert the XY dense flow vector of each pixel to a color representation using the method proposed in [10] and then apply the edge detection in the color domain. The appearance and motion edge maps are then combined by the maximum operation. Overall, the spatial and temporal pairwise potentials between neighboring superpixels are computed as given in Eq. (3).

$$\boldsymbol{\phi}_s(s_i^j, s_p^q) = \left(1 - e(s_{i,}^{j} s_p^q)\right) * \left(1 - \boldsymbol{\delta}(x_i^j, x_p^q)\right) * \exp(-\beta_s^{-1} ||F_{i,}^{j} F_p^q||^2$$

$$\boldsymbol{\phi}_t(s_i^j, s_p^q) = c(s_i^j, s_p^q) * \left(1 - \boldsymbol{\delta}(x_i^j, x_p^q)\right) * \exp(-\beta_s^{-1} ||H_{i,}^{j} H_p^q||^2 \quad (3)$$

Here, $e(s_{i,}^{j} s_p^q)$denotes the average edge strength betweensuperpixels$s_{i,}^{j}$and s_p^q, $c(s_i^j, s_p^q)$denotes the percentage of pixelin s_p^qthat are linked tos$_{i,}^{j}$by optical flow, and δ is thestandard Kronecker delta function.$F_{i,}^{j}$is the concatenation of colorand optical flow orientation histogram and $H_{i,}^{j}$is the colorhistogram.

B. Appearance Auxiliary Potential

In general, the appearance constraint$\varphi_a(s,x,\theta)$ in Eq.(2) can be written as Eq. (4).

$$\varphi_a(s,x,\theta) = f(\mathbf{s}, \mathbf{x}, g(\mathbf{s}, \mathbf{x})) \quad (4)$$

Where

*f*measures how consistent the current labeling

xis with the appearance model, and

*g*computes the appearance model parameters given the current labeling **x**.

Theoptimization scheme is usually employed to solveEq.(1), *i.e.*, fix the appearance model while solving **x** and fix **x** while optimizing the appearance model. Inspired by [11], in this work propose an appearance model for video object segmentation in which$\varphi_a(s,x,\theta)$ can be expressed analytically in terms of **x**, and Eq.(1) can be solved efficiently by one graph cut. In the following, first review the method of [20] on static image segmentation and then discuss the challenges in adapting the idea to videos and how overcome them. The method in [20] models each pixel as a node and represents each node as a single bin in the RGB histogram space for appearance modeling.

$$\Phi_a(\mathrm{x}, \theta) = \sum_{k=1}^{H} \min(\boldsymbol{\Omega}_F^k, \boldsymbol{\Omega}_B^k) \quad (5)$$

A naive extension of [18] to our superpixel based video object segmentation is to take the mean RGB color of each superpixel and assign it to one of the bins in the color histogram space,fromEq.(5).However raw color features alone may not be robust enough to accurately capture the viewpoint and lighting variations between frames Eq. (6).

$$\varphi_a(s,x,\theta) = \sum_i^N \sum_j^{Mi} \boldsymbol{\phi}_s(s_i^j) \quad (6)$$

However, in practice, a superpixel node will be connected to an appearance auxiliary node only if the corresponding bin is not empty and an appearance auxiliary node will be added to the graph only when it is connected to at least two different superpixels.

V.EXPERIMENTAL RESULTS

The given input video is converted into frames of images using MATLAB software. Generally, every video or animation that is seen on our television, computer, phone or any other electronic devices is made from a succession of still images. These images are then played one after the other several times a second which fools us upon thinking the object is moving. The faster the images are being played, the smoother and more sequential the movement looks. Mostly videos and movies are filmed at around 24-30 images per second whereby each individual image is called a frame. The input video appear in Fig.2.Each frame is seen in term of frames per second (FPS).In order to perform object segmentation, the video has to be converted into frames shown in Fig.3. Frame resizing is shown in Fig. 4. Each frame can be used specifically for the segmentation process shown in Fig. 5. The tracking video output is shown in Fig. 6.

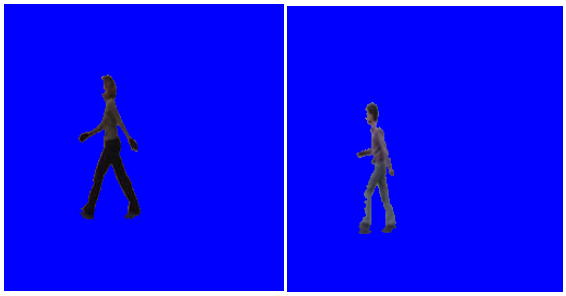
(a) (b)

Fig. 2. Input video "Weizman dataset" (a) 'lena-walk' (b) 'denis_walk'

Fig. 3. Frame conversion

(a) (b)

Fig. 4. Resized frames(a) 'lena-walk' (b) 'denis_walk'



(a) (b)

Fig. 5.object segmentation(a) 'lena-walk' (b) 'denis_walk'

(a) (b)

Fig. 6.Tracking videoa) 'lena-walk' (b) 'denis_walk'

Table. 1. Performance measure

Dataset	Walking Speed
Weizman 'lena-walk'	4.2450
Weizman 'denis_walk'	4.3634
Weizman 'ira-jump'	4.8160
Weizman 'ira-run'	6.0579

From the Table. 1. it can be seen that the performance of algorithm. When using this approach, the segmentation and superpixelwork well with Weizman data set. It can be overcome by using the MRF & SLIC methodology as proposed. Thisevaluate the proposed approach against several state-of-the- art methods including both MRF based method [5] and non-MRF based methods [17]. Also compare with several baseline methods in order to separate the contributions of the different components. Pixel-wise Jaccard similarity coefficient, *i.e.*, intersection over union ratio, is used to evaluate the segmentation accuracy of each video. The efficiency of the proposed method is because of its Simplicity, *i.e.*, one graph cut on a sparsely connected graph in which the pairwise and appearance potentials can be computed efficiently.

VI. CONCLUSION

The proposed system in this paper has combined two algorithm are the MRF framework and SLIC for automatic video object segmentation. The proposed method uses features to characterize the local regions and embed the global appearance constraint into the region by auxiliary nodes and connections. Compared with many existing appearance models, the optimization process of our method is non-iterative. Experimental evaluations show that our method is faster than many of the alternatives and the segmentation accuracy is also better than or comparable with the state-of-the-art methods

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slicsuperpixels compared to state-of-the-art superpixel methods," *IEEETrans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274 2282, Nov. 2012.

[2] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2189–2202, Nov. 2012.

[3] X. Bai, J. Wang, D. Simons, and G. Sapiro, "Video SnapCut: Robust video object cutout using localized classifiers," ACM Trans. Graph., vol. 28, no. 3, p. 70, 2009.

[4] Y. Boykov and V. Kolmogorov, "An experimental comparison of mincut/ max-flow algorithms for energy minimization in vision," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 9, pp. 1124–1137,Sep. 2004.
[5] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," IEEE Trans. Pattern Anal.Mach. Intell., vol. 34, no. 7, pp. 1312–1328, Jul. 2012.
[6] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 569–582, Mar. 2015.
[7] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 8,pp. 1558–1570, Aug. 2015.
[8] P. Kohli, L. Ladický, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," Int. J. Comput. Vis., vol. 82, no. 3,pp. 302–324, 2009.
[9] Y. Li, J. Sun, and H.-Y. Shum, "Video object cut and paste," ACM Trans. Graph., vol. 24, no. 3, pp. 595–600, 2005.
[10] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.
[12] F. Manerba, J. Benois-Pineau, R. Leonardi, and B. Mansencal, "Multiple moving object detection for fast video content description in compressed domain," EURASIP J. Adv. Signal Process., vol. 2008, p. 5, Jan. 2008.
[13] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004.
[14] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, Aug. 2000.
[15] D. Tsai, M. Flagg, A. Nakazawa, and J. M. Rehg, "Motion coherent tracking using multi-label MRF optimization," Int. J. Comput. Vis., vol. 100, no. 2, pp. 190–202, 2012.
[16] J. Yuan, G. Zhao, Y. Fu, Z. Li, A. K. Katsaggelos, and Y. Wu,"Discovering thematic objects in image collections and videos," IEEETrans. Image Process., vol. 21, no. 4, pp. 2207–2219, Apr. 2012.
[17] Y. J. Lee, J. Kim, and K. Grauman, "Key-segments for video object segmentation," in Proc. IEEE Trans. Comput. Vis., Nov. 2011,pp. 1995–2002.
[18] M.Tang, L. Gorelick, O. Veksler, and Y. Boykov, "GrabCut in one cut," in Proc. IEEE Trans. Comput. Vis., Dec. 2013, pp. 1769–1776.