ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Special Issue 11, March 2016

Video Segmentation of Static Scenes using Spatio-Temporal Segmentation

G.Daisy
Department of Computer Science and Engineering
VINS Christian Womens College of Engineering.
E-mail id: daisygeorge93@gmail.com

Abstract
Video segmentation became popular and most important in the digital media storage. Extracting spatio-temporally consistent segments from a video sequence is a challenging problem due to the complexity of color, motion and occlusions. This paper presents a novel framework for spatio temporal segmentation. The spatio - temporal optimization is based on projecting the pixels to other frames for collecting the boundary and segmentation. To effectively perform segmentation process, iterative optimization scheme is used to independently link to the correspondence among different frame and iteratively refine them with the collected data. The first step of this segmentation process is done by pointing out the mean shift spatial boundary and spatial segmentation process. The next step is used to follow the correlation and connection of the collected data that should be applied and then proposed the methodology for retrieving the segmented object in image. The retrieval object maintains the temporal consistency in images. Finally the set of spatio-temporally consistent volume segments are achieved. Thus the effectiveness and usefulness of the video segmentation will be obtained and are demonstrated via its applications for Brightness, Enhancement, Sharpness and noise removal in the retrieval image.

Keyword:- Object Retrieval, Spatio-Temporal Segmentation

1. INTRODUCTION

With the increasing prevalence of digital cameras and intelligent mobile phones, more and more videos are shared and broadcasted over the Internet. Video segmentation is one of the most important processes performed to achieve video stylization, enhancement, reconstruction, semantic segmentation. However, compared to image segmentation, video segmentation is much more challenging due to much larger data and difficulty of maintaining the temporal coherence. To thoroughly solve these problems, need a powerful tool to segment the video into a set of temporally consistent layers. Spatio-temporal video segmentation is very challenging due to the large number of unknowns, possible colors and motion ambiguities, and complicated geometric structure of the captured scenes. Utilizing depth information for video segmentation is becoming an important issue. So far there are not many works done for video segmentation utilizing depth information. This paper, propose a novel depth-based video segmentation method which can be used for large-scale video reconstruction and many other applications. The objective of the video segmentation method is that the extracted segments not only preserve object boundaries but also maintain the temporal consistency in different images. The video segmentation is an imperative technique used for the improvement of video quality on the basis of segmentation. The function of video segmentation is to segment the static objects in video sequences. A novel segmentation algorithm is applied to test the object sequences in the image and the corresponding individual object retrieval is evaluated. This provides an automatic method for producing a temporally coherent output to retrieve the object in image. The spatio-temporal segmentation results can also facilitate many other applications, such as Brightness, Enhancement, Sharpness, noise removal and semantic segmentation. The structure of the paper is organized as follows: A brief review of the researches related to the video segmentation is given in Section 2. The proposed video segmentation technique is given in section 3. The experimental results of the proposed approach are presented in Section 4. Finally, the conclusions and future work are given in Section 5.

2. RELATED WORK

Image/Video Segmentation

During the past decades, many image segmentation methods have been proposed, such as normalized cuts [1], mean shift [2], segmentation via lossy compression, and segmentation by weighted aggregation (SWA). For a video sequence, if directly use these image-based segmentation methods to segment each frame

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Special Issue 11, March 2016

independently, the segmentation results will be inconsistent for different images due to the lack of necessary temporal coherence constraints.

Some spatio-temporal segmentation methods have been proposed to extend segmentation from single image to video. Two main types of segmentation criteria (motion and color/ texture) were generally used alone or in combination for video segmentation. Motion-based segmentation methods [3] aimed to group pixels which undergo similar motion, and separate them into multiple layers. Pure motion-based method are difficult to achieve high-quality segmentation results and usually produce inaccurate object boundaries due to the motion ambiguity and the difficulty of accurate optical flow estimation. Besides, all these methods are sensitive to large displacement with significant occlusions. In comparison, this paper method can achieve highly consistent spatio-temporal video segmentation without any learning priors. By utilizing the depth redundancy in multiple frames, spatio-temporal segmentation is rather robust to occlusions and out-of-view.

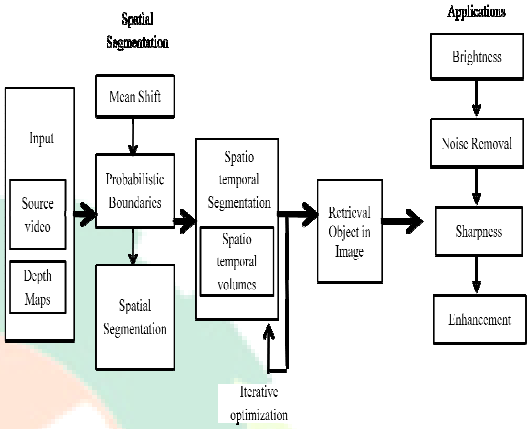
In summary, spatio-temporal segmentation is still a very challenging problem. Previous approaches generally have difficulties in handling large displacement with significant occlusions. This paper show that by associating multiple frames on the inferred dense depth maps, surprisingly spatio-temporal consistent segments can be obtained from video sequences. The high-quality segmentation results can benefit many other applications, such as video editing, and non-photorealistic rendering.

Disadvantage

- Framework procedure is complicated.
- Complexity of color, motion and occlusions.
- Difficulties and sensitive to handling large displacement with significant occlusions.
- Does not provide clear video output.
- Cannot retrieve the object in image.

3. PROPOSED METHOD

Given a video sequence of frames, the objective is to estimate a set of spatio-temporal volume segments, and fuse the volume segments to reconstruct a complete scene. The system overview is shown in the Fig. Assume that a depth map [5] is available for each frame of the input video. Structure from motion (SFM) method [6] proposed to recover the camera motion parameters from the input video sequence. With the recovered camera poses, employ the multi-view stereo method to recover a set of consistent depth maps. With the computed depth maps, first perform spatial segmentation for each frame with probabilistic boundary, and then iteratively optimize the segmentation results by enforcing the temporal coherence constraints among multiple temporal frames. The spatio-temporal segmentation can be used for many applications such as brightness, noise removal, enhancement, sharpness, and semantic segmentation.



3.1 SPATIAL SEGMENTATION WITH PROBABILISTIC BOUNDARY

Directly obtaining spatio-temporal volume segments in a video is difficult due to the large number of unknowns and the possible geometric and motion ambiguities in the segmentation. Therefore, design and iterative optimization scheme to achieve spatio–temporal video segmentation. For initialization, instead of directly segmenting each frame independently, first compute the probabilistic boundary map by collecting the statistics of segment boundaries among multiple frames. Then perform spatial segmentation for each frame independently with the computed probabilistic boundary maps. Experimental results demonstrate that much more consistent segmentation results can be obtained than those of directly using mean shift algorithm.

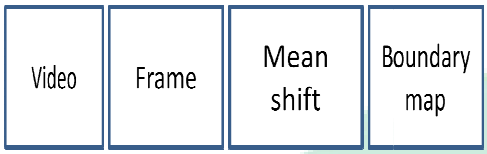
Probabilistic Boundary

First use mean shift algorithm to segment each frame independently with the same parameters. The segmentation results of the selected frames, which are not consistent in different images. The segmented boundaries are quite flickering, and a segment may span over multiple layers, which is obviously not good enough as a starting point for spatio-temporal segmentation. With the computed depths, project each pixel to other frames to find the correspondences. Compared to the traditional segmentation boundaries in a single image, probabilistic boundary map is computed with multiple frames, which is robust to image noise and occasional segmentation errors. The computed probabilistic boundary maps which are surprisingly consistent among different frames. The reason is that mean shift segmentation can preserve object boundaries well. Although the generated segment boundaries by

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

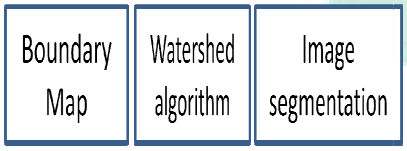
International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Special Issue 11, March 2016

mean shift may be occasionally in accurate in one frame, it still has large chance to be accurate in other frames. By collecting the boundary statistics in multiple frames, the computed probabilistic boundaries can naturally preserve the object boundaries and maintain consistency in neighboring frames.



Spatial Segmentation

With the computed probabilistic boundary map, use the watershed algorithm to segment the image. Compute a topographic surface of the maximal probabilistic boundary over the 4 connected probabilistic edges for each pixel, and apply watershed transformation on the surface. The topological map is clipped with a threshold value to avoid over segmentation. Notice that some quite small segments appear around the areas with strong probabilistic boundaries, most of us have segmentation noise and do not consistently appear in neighboring frames. So, eliminate segments that are too small (with less than 30 pixels), and set the pixels in these segments as unlabeled ones. By using this address the limitation of memory space and dramatically accelerate BP optimization, without affecting the segmentation results. The spatial segmentation results in different images are rather consistent, which provide a good starting point for the following spatio-temporal segmentation.



3.2 SPATIO - TEMPORAL SEGMENTATION

Due to the lack of explicit temporal coherence constraint, the spatial segmentation results may contain inconsistent segments. In addition, the segments in different images are not matched. In the following stage, perform spatio-temporal segmentation to achieve a set of pixel volumes. First, need to match the segments in different images and link them to initialize volume segments**.**

Initializing Spatio - Temporal Volumes

Without loss of generality, consider each segment can be projected to other frames, to find its matched segments in other frames. With these correspondences, build a matching graph. It is an undirected graph. Each segment corresponds to a vertex,and every pair of matched segments has an edge connecting them. Each connected component represents a volume segment. The initialized volume segments are already quite consistent. Then perform an iterative optimization to further improve the results. The initialized volume segments are used as candidate labels for further optimization.

Iterative Optimization

Due to segmentation error, the segment labels of pixels may be different. This experiments, found that most of these projected segments are the same, which indicates that an initialized volume segments are already quite good. Denote the set of segment candidates for pixel, which includes these projected volume segments and then define the probability of each segment label for pixel. These experiments, are sufficient to produce spatially and temporally coherent volume segments. Compared to the initialized volume segments, the refined volume segments become more consistent and better preserve object boundaries.

3.3 OBJECT RETRIEVAL

The object whose segmentation quality [12] should be evaluated is selected. After selected the object, then retrieve the particular object in the image using adaptive threshold based segmentation. The temporal features are selected for individual object retrieval and then the corresponding metrics are evaluated. If there is no occlusion or out-of-view, each projection should correspond to a valid segment. Then the retrieval object can be enhanced in the image. This can provide a simple and elegant retrieval image from the video.

Advantage

- Not a complicated task. Simple and efficient.
- Reconstruct the video for large scale scene.
- Robustly retrieve the object in the frame.
- Avoid accumulation of errors.
- Used to provide correct retrieval object in the image.
- Very easy and convenient for all the users.

4. EXPERIMENTAL RESULTS

Experimented with several challenging examples where the sequences are taken by a moving camera. The tested sequences generally contain frames. For the sequence with resolution, computing probabilistic boundary requires 26 seconds per frame on a desktop PC with Intel 4-Core 2.83GHzCPU. The spatial segmentation requires 2 minutes per frame, and each pass of spatio-temporal optimization requires 1.5 minutes per frame. The performance of this method is acceptable for many video applications, and allows further acceleration using GPU. The configuration of the parameters in this system is easy. For spatio-temporal segmentation, mean shift

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Special Issue 11, March 2016

allows the control of segmentation granularity, obtain different numbers of volume segments by adjusting the parameters of mean shift in the initialization stage.

Segmentation Results of Ordinary and Low-Frame-Rate Video

The segmentation results still preserve accurate object boundaries and are quite temporally consistent in the whole sequence. The reason is that this method mainly uses the depth information to connect the correspondences among multiple frames and collect the statistics information (such as probabilistic boundaries and the segment probability) for spatio-temporal segmentation, which is more robust than directly using depth information as an additional color channel. Generally, this method can achieve more temporally consistent segments, Besides complex occlusions are better handled by the method. Though the method is developed to solve the video segmentation problem, it can also handle low-frame-rate sequences that contain a relatively small number of frames with moderately wide baselines between consecutive frames. Segmentation results still preserve fine structures and faithfully maintain the coherence in different images.

Quantitative Evaluation of Segmentation

To further demonstrate the effectiveness of the proposed method, use the metrics similar(i.e., intra-object homogeneity, depth uniformity and temporal stability) and these method maintain the temporal coherence among temporally neighboring frames to objectively evaluate the quality of the segmentation results.

5. CONCLUSION AND FUTURE WORK

This paper, have proposed a novel video segmentation method, which can extract a set of spatio-temporal volume segments from a depth-inferred video. Most previous approaches rely on pairwise motion estimation, which are sensitive to large displacement with occlusions. By utilizing depth information, connect the correspondences among multiple frames, so that the statistics information, such as probabilistic boundaries and the segment probability of each pixel, can be effectively collected. From the segmented image, the particular boundared object can be retrieved. By incorporating these statistics information into the segmentation energy function, this method can robustly handle significant occlusions, so that a set of spatio temporally consistent segments can be achieved. This method uses a single handheld camera and multiview stereo method to recover the depth maps and collect multi-frame statistics, which is restricted to videos of a static scene. Believe this method can be improved in the future to handle dynamic scenes. For moving objects, the temporal correspondences among the different frames should be built by motion estimation with a depth camera or multiple synchronized video cameras, so that the temporally coherence constraint can be reliably enforced.

REFERENCES

[1] 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.
[2] D.Comaniciu, P. Meer, and S. Member, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 603–619, May 2002.
[3] M. P. Kumar, P. H. S. Torr, and A. Zisserman, "Learning layered motion segmentations of video," Int. J. Comput. Vis., vol. 76, no. 3, pp. 301–319, 2008.
[4] P. Sandand S. J. Teller, "Video matching," ACM Trans. Graphics, vol. 23, no. 3, pp. 592–599, 2004.
[5] G. Zhang, J. Jia, T.-T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," IEEE Trans. Pattern Anal. Mach. In- tell., vol. 31, no. 6, pp. 974–988, Jun. 2009.
[6] G.Zhang, X.Qin, W.Hua, T.-T.Wong, P.-A.Heng, an H.Bao,"Robust metric reconstruction from challenging video sequences," in Proc. CVPR, 2007, pp.1-8.
[7] S. Zhang, X. Li, S. Hu, and R. R. Martin, "Online video stream abstraction and stylization," IEEE Trans. Multimedia, vol. 13, no. 6, pp. 1268–1294, Dec. 2011.
[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[9] Wenzhuo Yang,Guofeng Zhang,Hujun Bao, Jiwon Kim,Ho Young Lee''Consistent Depth Maps Recovery from a Trinocular Video Sequence''IEEE conf 2012.
[10] Jeroen van Baar, Paul Beardsley ''Interactive video segmentation supported by multiple modalities, with an application to depth maps ''IEEE conf 2012.
[11] Yuliya Tarabalka, Member, IEEE, Guillaume Charpiat, Ludovic Brucker, and Bjoern H. Menze''Spatio-Temporal Video Segmentation With Shape Growth or Shrinkage Constraint'' IEEE Trans on Image processing, vol. 23, NO. 9, September 2014.
[12] Paulo Lobato Corriea and Fernando Pereira, Senior Member, IEEE'' Objective Evaluation of Video segmentation Quality'' IEEE Trans on Image Processing ,vol 12, NO.12,February 2003.