GRAPH-BASED FOREGROUND EXTRACTION IN EXTENDED COLOR SPACE

Hansung Kim and Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey, GU2 7XH, UK
{h.kim, a.hilton}@surrey.ac.uk

ABSTRACT

We propose a region-based method to extract semantic foreground regions from color video sequences with static backgrounds. First, we introduce a new distance measure for background subtraction which is robust against shadows. Then the foreground region is extracted with a graph-based region segmentation method considering background difference and spatial homogeneity. For efficient computation, the graph structure is optimized by the minimum spanning tree before segmentation. The main contribution is that the proposed algorithm improves on conventional approaches especially in strong shadow regions and does not require manual initialization. We have verified through experiments and comparison to state of the art methods that the proposed algorithm works well with various cameras and environment.

Index Terms— Background subtraction, Graph algorithm, Minimum spanning tree

1. INTRODUCTION

Conventional object segmentation algorithms are roughly classified into two categories based on their primary segmentation criteria. The first kind of approach uses spatial homogeneity as a criterion like graph-based segmentation [1-3] or active contours [4][5]. Recently, Levin et al. proposed a globally optimal alpha matting method by solving a sparse linear system of equations from local smoothness assumptions on foreground and background colors [6]. The segmentation results of these algorithms tend to track the object boundary more precisely than other methods, but the main drawback of these algorithms is high computational complexity. Moreover, they need manual initialization to choose semantic regions because the results just show regions with similar properties in the image. Mu et al. have proposed a combination method of block-based initial segmentation and refinement by Graph-cut for automatic segmentation [7]. The second category of approaches exploits change detection in video sequences. Background subtraction is one of the most common approaches for detecting semantic foreground regions [8][9]. This technique subtracts the current image from a static background image acquired in advance from multiple images over a period of time. Gaussian mixture models are the most representative background models in parametric approaches and have been widely incorporated with other image properties [10-12]. Non-parametric approaches such as using a Kernel density estimator have also been researched for a long time [13][14]. However, conventional methods based on colour models do not provide enough information to effectively segment foreground regions.

In this paper, we propose an automatic graph-based foreground extraction method which is robust against strong shadows and preserves object boundaries. We assume that the background is static and the semantic foreground is new objects in the scene. We propose a new distance measure for background subtraction in an extended colour space. The whole image is segmented into sub-regions by considering background subtraction, edge intensity and spatial homogeneity then each sub-region is classified into foreground/background groups.

2. PROPOSED ALGORITHM

2.1. Distance measure in an extended color space

The difference between a background model and observed pixel is measured by considering red, green, blue, brightness and hue components for each pixel. The extended colour space for the distance measure is shown in Fig. 1. The variation of each pixel over time shows different distribution in an image according to the material of objects or lighting conditions. In this paper, the background model for each pixel in RGB channels is constructed with simple Gaussian distributions. This can be replaced more complex models such as a Gaussian mixture model [10] or generalized Gaussian family model [15]. However, the independent RGB components may change drastically due to shadows of objects in the background. When we assume that the shadow region usually has similar colour but lower brightness to the background, we can remove this region by checking the brightness and hue information.

We measure the difference between background pixel and current pixel as follows.

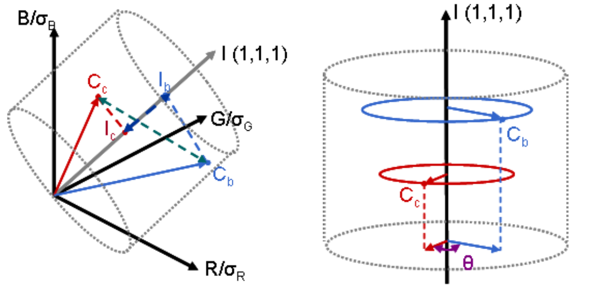


Fig. 1. Background subtraction in the extended colour space

$$BS = \left|\overline{C_c C_b}\right| - \lambda\left(1 - \frac{\theta}{180^\circ}\right)\cdot H\left(\left|I_b\right| - \left|I_c\right|\right) \tag{1}$$

$H(\bullet)$ is a unit step function, C_x indicates a pixel in RGB domain normalized by standard deviation of the background model, and I_x is projection of each RGB vector to the intensity vector $I(1,1,1)$. θ is angular distance in the range of 0~180º for hue difference calculated as:

$$\theta = \arccos\left(\frac{\overline{I_c C_c}\cdot\overline{I_b C_b}}{|\overline{I_c C_c}|\cdot|\overline{I_b C_b}|}\right)\times\frac{180^\circ}{\pi}. \tag{2}$$

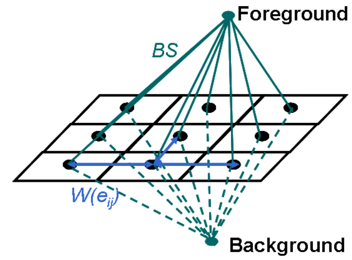
In Eq. (1), the first term $|\overline{C_c C_b}|$ is the Mahalanobis distance for background difference in RGB domain and the second term is the penalty by shadows. The second term works only if the current pixel is darker than a background pixel. The weighting factor λ is experimentally set to 45 for all results presented in this paper.

2.2. Minimum spanning tree and Region-cut

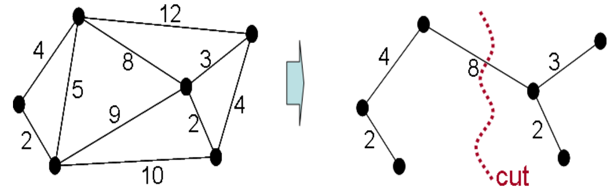
A graph $G=(V,E)$ is then constructed for the input image as Fig. 2(a), where $v_i \in V$ is the set of pixels and $(e_{ij}) \in E$ is the edge between neighbouring elements (v_i,v_j) with a weight $w(e_{ij})$. We set the edge weights according to the background subtraction BS, gradient of the current image ∇C_c and attenuated edge $G(i)=|Sobel(I_c(i))\text{-}Sobel(I_b(i))|$ as Eq. (3). The attenuated edge takes a role of suppressing edges in the background. The weighting factor λ_2 is set to 3 and T is fixed to 30.

$$w(e_{ij}) = |BS(i) - BS(j)| + \lambda_2|\nabla C_c(i_j)|\cdot\left(1 - \frac{1}{1+(G(i)/T)^2}\right) \tag{3}$$

However, most graph-cut algorithms have a problem of NP-hard computational complexity. To avoid this NP-hard problem for finding minimum cut through all nodes of a weighted graph and increase computation speed, the graph structure is firstly refined by the minimum spanning tree (MST) [16]. MST was originally developed for detecting clustered structure in arbitrary node sets, but we use it to optimize graph structure by connecting all components with minimum energy and eliminating other edges as shown in Fig. 2(b).



(a) Graph structure



(b) Minimum spanning tree (MST) and region-cut

Fig. 2. Graph construction and region-cut

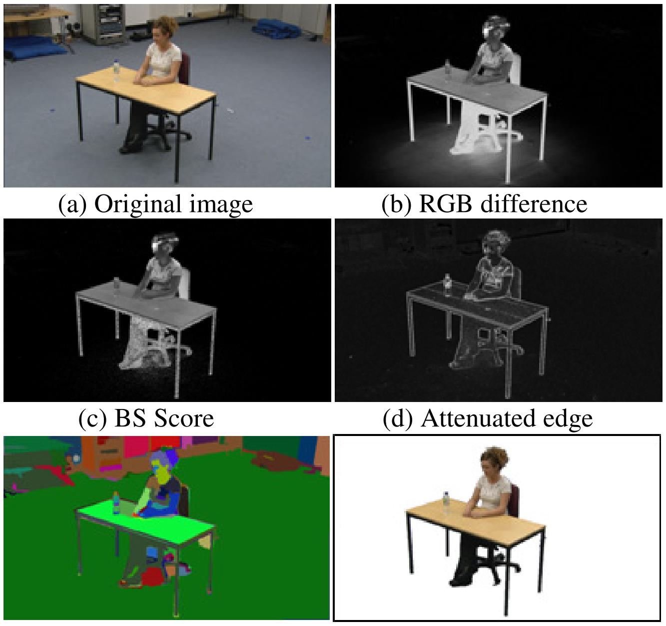
Then the tree is cut into sub-regions R so that it minimizes the internal difference $Int(R_r)$ and maximizes external difference $Ext(R_r,R_s)$ between different regions. Felzenszwalb has proposed an image segmentation method which uses very simple and intuitive concept that "The intensity differences across the boundary of two regions are perceptually important if they are large relative to the intensity difference inside at least one of the regions" [2]. We use his criteria for the region-cut of the graph. The method cuts edges if two neighbouring regions satisfy the condition in Eq. (4).

$$\begin{aligned} &Ext(R_r,R_s) > Int(R_r,R_s) \\ &\quad Ext(R_r,R_s) = \min_{i\in R_r,\, j\in R_s} w(e_{ij}) \\ &\quad Int(R_r,R_s) = \min\left(\max_{e\in R_r} w(e) + k/|R_r|,\ \max_{e\in R_s} w(e) + k/|R_s|\right) \end{aligned} \tag{4}$$

$|R|$ is the number of pixels in a region and the term $k/|R_x|$ indicates the preference for large components. Felzenszwalb's algorithm is good at detecting shadow regions because it can find slowly changing region which commonly occur due to shading. The separated sub-regions are finally allotted to foreground or background by the BS scores in each region. k is empirically set to 800 pixels for HD images and 300 for SD images.

3. SIMULATION RESULTS

The algorithm has been tested on nine videos with various cameras shown in Table 1 and compared with results from several state-of-the-art approaches.



(a) Original image (b) RGB difference

(c) BS Score (d) Attenuated edge

(e) Segmented sub-regions (f) Segmentation result

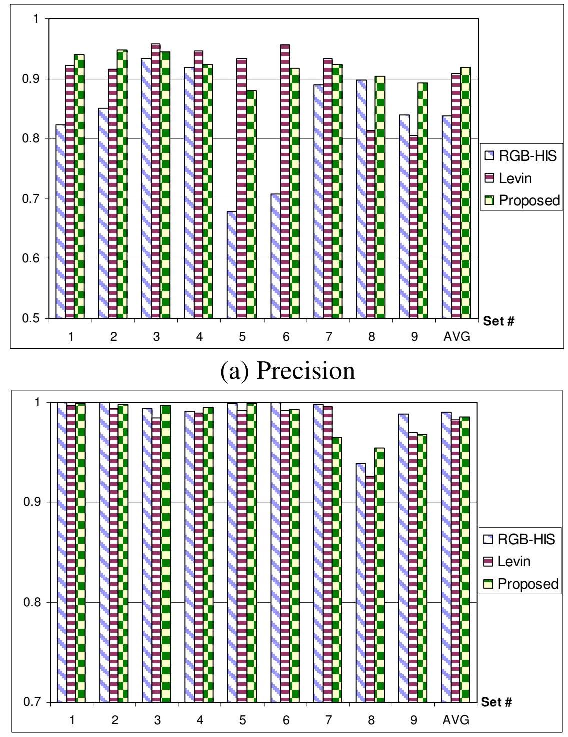
Fig. 3. Segmentation process

Table 1. Test sets for evaluation

#	Scene	Camera	Format	Characteristic
1	Office	HD cam 1920*1080	YUV422 raw	Complex scene with strong shadows
2	Office			
3	Dance			Fast motion with motion blur
4	Dance			
5	Office	SD cam 720*576	RGB 24bits raw	Low resolution
6	Dance			
7	Lobby	Compact dig. cam 640*480	YUV420 highly compressed	Blocking artifacts from compression. Unstable background
8	Street			
9	Street			

Figure 3 shows each step of the proposed algorithm for the test set #1. We can see that the BS score suppressed the effect of shadows in Fig. 3(c) and the attenuated edge simplified regions in the background in Fig. 3(d).

For objective evaluations, we randomly selected several frames from each test set and created ground-truth segmentation masks by manual segmentation. Then we compared the average segmentation error of the proposed algorithm with a RGB-HSI algorithm with a generalized Gaussian family model [15] and a Levin's closed form solution algorithm [6]. For fair evaluations, we estimated the optimal parameters by training as stated in [15] for the RGB-HSI algorithm, provided good manual initialization for the Levin's algorithm, and used the fixed parameter set of the proposed algorithm as stated in Section 2 for all test video sets. We evaluated the results by calculating precision and recall in Eq. (4), where the *False Positive* means the background region being falsely classified as the foreground while the *False Negative* is the foreground region falsely classified as the background region.



(a) Precision

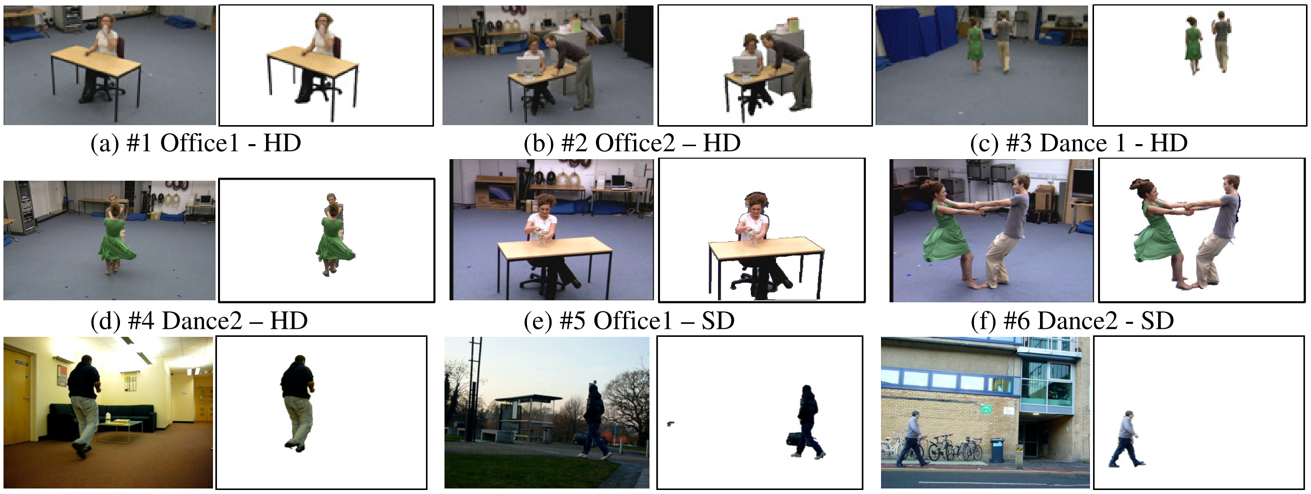
(b) Recall

Fig. 4. Objective evaluation

$$precision = \frac{\#of\ True\ Positive}{\#of\ True\ Positive + \#of\ False\ Positive} \qquad (4)$$
$$recall = \frac{\#of\ True\ Positive}{\#of\ True\ Positive + \#of\ False\ Negative}$$

Graphs in Fig. 4 show the objective evaluations of the proposed algorithm. The RGB-HIS algorithm is not good at removing strong shadows so it shows low precision ratio in office scenes because the shadows regions under the table and chair are included in *False Positive*. On the other hand, it shows high scores in the recall ratio. The Levin's algorithm shows good results both in precision and recall, but it shows low score with highly compressed outdoor scene of sets #8 and #9 because of the blocking artifacts. Another problem of the Levin's algorithm is the necessity of manual initialization for foreground and background. Actually, it shows good results against strong shadows because parts of shadow regions have already been initialized as background in the manual process. The proposed algorithm shows evenly good performance for various scenes and cameras.

Figure 5 shows snapshots of the results for the test sets. The proposed algorithm produces improved results especially in the strong shadow regions. A video clip displaying the results can be downloaded from the following URL.



(a) #1 Office1 - HD (b) #2 Office2 – HD (c) #3 Dance 1 - HD

(d) #4 Dance2 – HD (e) #5 Office1 – SD (f) #6 Dance2 - SD

(g) #7 Lobby – Compact camera (h) #8 Street1 - Compact camera (i) #9 Street2 – Compact camera

Fig. 5. Snapshots of segmentation results

http://www.3dkim.com/research/Segment0901.wmv

4. CONCLUSION

In this paper, we proposed a region-based algorithm for foreground extraction that is robust against shadows. The input image is classified into foreground and background by segmenting a graph constructed with background subtraction in the extended color space and spatial homogeneity. The experimental results show that the proposed approach can extract foreground regions even if the image has strong shadows, motion blur or artifacts from compression. However, flickering on the foreground boundaries are shown in the result videos because temporal consistency is not considered and the graph-based region segmentation produces subtly different cuts for each frame. Future work aims to track the stable object boundaries over frames.

5. ACKOWLEGEMENT

This research was executed with the financial support of the EU IST FP7 project i3Dpost.

6. REFERENCES

[1] S Vicente, V Kolmogorov, C Rother, "Graph cut based image segmentation with connectivity priors," *Proc. CVPR*, pp.1-8, 2008.

[2] P.F. Felzenszwalb et al. "Efficient Graph-Based Image Segmentation," *IJCV*, vol.59, pp.167-181, 2004,

[3] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," *IJCV*, vol.70, no.2, 109-131, 2006,

[4] C. Li, C. Xu, C. Gui, and M. Fox, "Level Set Evolution Without Re-initialization: A New Variational Formulation," *Proc. CVPR*, pp.430-436, 2005,

[5] C. Li, C. Kao, J. Gore and Z. Ding, "Implicit Active Contours Driven by Local Binary Fitting Energy," *Proc. CVPR*, pp.1-7, 2007.

[6] A. Levin, D. Lischinski and Y. Weiss "A Closed Form Solution to Natural Image Matting," *IEEE Trans. PAMI*, vol.30, pp.228-242, 2008,

[7] Y. Mu, H. Zhang, H. Wang and W. Zuo, "Automatic Video Object Segmentation Using Graph Cut," *Proc. ICIP*, 377-380, 2007,

[8] E.D. Gelasca, T. Ebrahimi, M. Karaman and T. Sikora, "A Framework for Evaluating Video Object Segmentation Algorithms," *Proc. CVPR Workshop*, pp.198-198, 2006.

[9] M. Piccardi, "Background Subtraction techniques: a review," *Proc. IEEE. SMC*, vol. 4 pp.3099- 3104, 2004

[10] Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. *Proc. CVPR,* pp.246-252, 1999.

[11] F. Porikli, and O. Tuzel, "Human body tracking by adaptive background models and mean-shift analysis," *Proc. PETS-ICVS*, 2003.

[12] A. Criminisi, G. Cross, A. Blake and V. Kolmogorov, "Bilayer Segmentation of Live Video," *Proc. CVPR*, pp.53-60, 2006

[13] B. Han, D. Comaniciu, and L. Davis, "Sequential kernel density approximation through mode propagation: applications to background modeling," *Proc. ACCV*, 2004

[14] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," *Proc. CVPR*, pp.302-309, 2004.

[15] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama and K. Kogure, "Robust Foreground Extraction Technique Using Gaussian Family Model and Multiple Thresholds," *Proc. ACCV*, pp.758-768, 2007.

[16] J. Eisner, *State-of-the-art algorithms for minimum spanning trees: A tutorial discussion*, University of Pennsylvania, 1997.