AUTOMATIC OBJECT SEGMENTATION USING MULTIPLE IMAGES OF DIFFERENT LUMINOUS INTENSITIES

Jae-Kyun Ahn, Dae-Youn Lee, and Chang-Su Kim

Media Communications Laboratory
School of Electrical Engineering, Korea University, Seoul, Korea
E-mail: demian@korea.ac.kr, inomi@korea.ac.kr, changsukim@korea.ac.kr

ABSTRACT

This paper represents an efficient algorithm to segment objects from the background using multiple images of distinct luminous intensities. The proposed algorithm obtains images with different luminous intensities using a camera flash. From the multiple intensities for a pixel, a saturated luminous intensity is estimated together with the slope of intensity rate. Then, we measure the sensitivities of pixels from their slopes. The sensitivities show different patterns according to the distances from the light source. Therefore, the proposed algorithm segments near objects using the sensitivity information by minimizing an energy function. Experimental results on various objects show that the proposed algorithm provides accurate results without any user interaction.

Keywords: object, segmentation, flash, graph cut

1. INTRODUCTION

Object segmentation is one of the most practical and meaningful method of image processing with various applications in research and market. Object segmentation is applied to various applications with diverse form, most representative applications are surveillance, video telecommunication, mobile imaging, computer graphics, and video compression. Also, this includes a wide range of possibilities to develop novel algorithms due to pre-processing role in many graphic editing algorithms. Though many conventional researches have been progressed with many advantages, object segmentation is still challenging research of image processing and computer vision. In addition, extracting accurate and sophisticated result became more significant by increasing demand in market due to recent growth of communications.

Many algorithms have been proposed for object segmentation. Usually object segmentation is classified into auto method and semi-auto method depends on user's interaction. An approach for semi-auto method is Grab Cut, which uses user's interaction as occasion demands [1]. In [1], a user defines an approximate rectangular box including objects, and segmentation is applied by energy minimization method to compare colors in side of the box and outside of the box. If the initial segmentation is not sufficient, user marks more precise regions and energy minimization is applied again. Grab Cut is a robust and make an accurate result, but it is semi-auto method, which has constraints of various applications.

Another approach of object segmentation for still images is Cosegmentation [2]. In [2], two images which have the same objects and different background are used for segmentation. Color histograms of the both images are made and compared to each other. From histograms, colors found in common are estimated as objects and the other colors as background. If the extraction is not accurate or precise, user interaction is regarded as in Grab Cut. Final results are obtained using energy minimization. This algorithm is a mixture of auto and semi-auto method, and provides auto method with semi-auto method effectively. However this method is not robust in some cases, since histogram depends on image highly. Also this method is not fully automatic.

Flash Cut algorithm segments objects using flash/noflash image pair [3]. If an image is taken with a flash, objects close to light source reacts sensitively and objects far from the source reacts insensitively. Using this property, by comparing flash image with noflash image makes it possible to separate objects and background. Like Cosegmentation, histogram and intensity differences are main clue of objects and background features, and final result is achieved using energy minimization method like the others. Flash Cut algorithm is auto method, however, it is proper for only some special images which have insensitive background such as outdoor scenes.

Another approach is afforded by disparity estimation of stereo images [4]. If disparity is estimated precisely, foreground layer is easily extracted, that is, objects which have shallow depth are foreground and objects which have deep depth are background. But disparity estimation is another difficult problem and also challenging. If the image has homogeneous region, disparity is not estimated well due to wrong matching points. In addition, computational complexity of disparity estimation is too high to apply to practical object segmentation.

In this paper, we propose an algorithm which uses different luminous intensities. By controlling flash intensities, image array are obtained. Those matching points for each pixel show a pattern which depends on distance based on intensity rate of pixel. Based on intensity rate, sensitivity is defined as main cue of segmentation. To extract final result, sensitivity and boundary conditions are imposed as cost and energy minimization method is applied.

[1] This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. R01-2008-000-20292-0).

The rest of this paper is organized as follows. Section 2 describes image array of the luminous intensity in the image, and Section 3 explains a sensitivity modeling, and Section 4 details the energy minimization scheme based on the graph cut. Section 5 discusses experimental results. Finally, Section 6 concludes the paper.

2. LUMINOUS INTENSITY IN IMAGE

The luminous intensity is a measure of the light power in a direction per unit solid angle, based on the luminosity function. In general, the candela (cd) SI unit, is used as unit of luminous intensity. However, if a camera flash is used as a light source, candela can be replaced by Guide Number which is a flash unit. The Guide Number is luminous intensity of a flash and it is practical measure of the light quantity to take a picture with a flash. In practice, the Guide Number means proper distance of objects from the flash for fixed aperture called f-number, when take a picture. Since Guide Number depends on the distance and aperture, we control Guide Number for fixed f-number. Also, Guide Number is linear to distance, and the linearity make it easy to solve segmentation problem.

Fig. 1 shows an array of images for different luminous intensities captured by an unstationary DSLR camera with fixed camera parameters. As luminous intensities are increased, most of pixels are saturated at some luminous intensity due to limited intensity range of the image. To figure out the relationship between the luminous intensity and the image intensity, the luminous intensity for saturation should be found for each pixel, and we assume all pixels are independent to each other. Saturation points can be easily estimated using intensity differences of an image and the next image which is only 1-step brighter. For each pixel, intensity of the image is increase as Guide Number increase, and when Guide Number reaches saturation point, intensity of image is saturated. Therefore, range of luminous intensities can be divided by increasing range and saturation range on saturation point. Approximately intensity differences of unsaturated range and saturated range are uniform, and the average intensity differences of two ranges are not in accordance. A difference value of average intensity difference for two ranges has maximum value at saturation point. So, the saturation point can be estimated as a Guide Number which makes a maximum difference of average intensity differences, given by

$$GN_{sat} = \max_i \left(\left| \frac{1}{(i-j+1)} \sum_{j=1}^{i} (I_j - I_{j-1}) - \frac{1}{(GN_{\max} - i)} \sum_{j=i+1}^{GN_{\max}} (I_j - I_{j-1}) \right| \right) \tag{1}$$

where GN_{sat} is a Guide Number that saturation begins, I_i is intensity of image at Guide Number i. Because all pixels in the image are independent to each other, saturation point seems to have no pattern. However, most of pixels in foreground have lower saturation points than those in background, and this is one of the main cues to define sensitivity. Note that some pixels don't saturate even at maximum Guide Number. In this case, saturation point of (1) is not what we want. To prevent this, a simple constraint comparing intensity difference is applied to saturation region. The intensity difference of saturated region should be low and uniform.

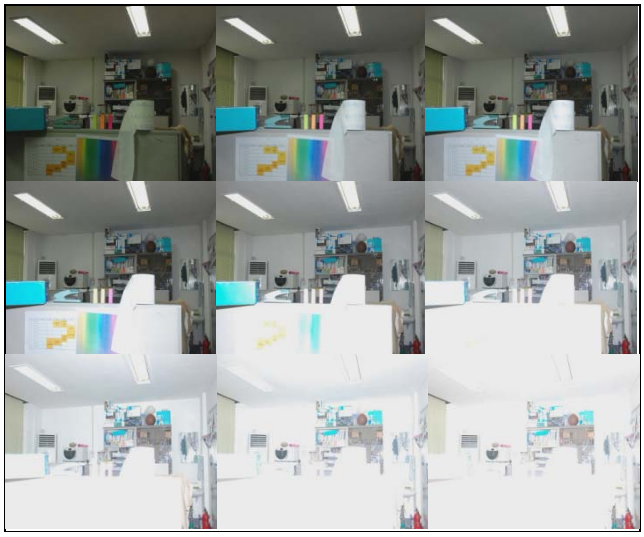


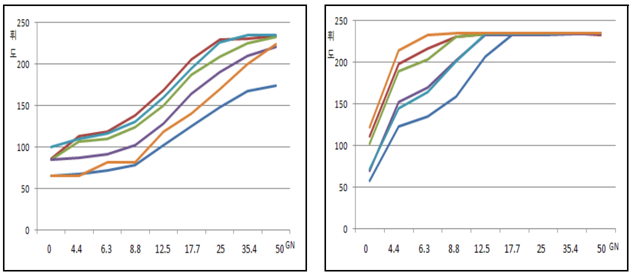
Fig. 1: An multiple images of different luminous intensities.

(a) (b)

Fig. 2: Enlarged images of region in Fig. 1.

3. SENSITIVITY MODEL

Intuitively, objects close to a flash are more sensitive than objects in background, and also dark objects are more sensitive than bright one. To make this concept quantitative value, we define a sensitivity which depends on distance from the light source with other properties. To compare the intensity of image for objects and background obviously, the same objects which have the same color and shape lie in both foreground and background. Fig. 2 shows enlarged images region in Fig. 1 to confirm the same objects. Fig. 3 shows intensity changes of the images for different luminous intensities. Fig. 3 (a) shows intensity change of foreground, (b) shows intensity change of background against Guide Number. Pixel samples selected in Fig. 3 is objects in Fig. 2. It shows that objects of the foreground are saturated more rapidly than background. In other words, objects in foreground are more sensitive than objects in background. So the intensity of image rate for luminous intensity is significant cue to define sensitivity which is



(a) (b)

Fig. 3: Intensity change of (a) foreground and (b) background against Guide Number

main parameter to segment objects from background. The intensity rate (R) is simply obtained by a slope in Fig. 3 to saturation point, given by

$$R = \frac{GN_{sat}}{I_{GN_{sat}} - I_0} \qquad (2)$$

where I_0 is a intensity of the original image. Note that in Fig. 3 the intensity rate is not uniform for each pixel, even though they lie in same distance from the light source. This depends on whether the intensity of the original image, I_0, is bright or dark, and it is inherent character of objects called luminance. Without respect to the medium, bright objects have high reflectivity phase, on the other hands, dark objects have high absorption phase. This is remarkable, especially in foreground than background due to proximity. The intensity rate for each pixel is highly related to intensity of the original image, and rate and this intensity are in inverse proportion in Fig. 3. As explained in section 2, Guide Number is linear to distance from the light source. So, all cues related to sensitivity are applied as linear form. Based on these cues of intensity of the original image, slope and distance, the sensitivity (S) is defined as,

$$S = \frac{1}{255R(I_0 + \lambda)} \qquad (3)$$

where λ is a parameter determined by another camera setting. If an exposure time is long, the sensitivity becomes high. Then it is difficult to segment objects and background. So, the sensitivity can be normalized by reducing λ. It is difficult to define λ due to many camera parameters. Also, camera parameters are pretty different to the manufacturer of the camera. Fig. 4 (a) is a normalized sensitivity image of Fig. 1. We see that foreground is more sensitive than background.

4. SEGMENTATION BY ENERGY MINIMIZATION

Though the objects have the same distance and the same original intensity, the sensitivity may not be the same due to camera problem. Also, the angle between the surface and light direction is not always perpendicular, and inaccurate result can be estimated. These problems are difficult to be estimated directly. So, final segmentation result using sensitivity can be obtained by energy minimization method. Energy minimization is a method to extract results by imposing data as costs. The constraint that we impose on boundary and sensitivity are described by the cost function , given by

$$E(\alpha) = E_{data}(\alpha) + E_{smooth}(\alpha) \qquad (4)$$

where each donates the data term means the sensitivity, and smoothness term is boundary property term, α is label which indicates foreground and background. To impose data cost, the sensitivity is modeled as probability form, since energy cost should be soft. In some case, boundary condition is lower than data cost, and it may be accurate label. If the data cost is discrete or difficult to normalize, energy minimization doesn't perform correctly.

A few sensitivities are too large or too small due to limited range of image. For example, if the intensity of the original image is 255, sensitivity is not estimated correctly. To prevent these problems, those pixels are limited to boundary value. Then, the sensitivities are distributed in the form of Gaussian Mixture Model (GMM). In other words, sensitivities are separated as two Gaussian distribution using the Expectation-maximization algorithm. Based on this distribution, probabilities of objects and background can be obtained for a given sensitivity. Fig. 4 (b) shows a clustered sensitivity distribution of Fig. 1 in the form of GMM. In general, sensitivity can be easily clustered as two groups, objects and background. Probabilities of objects and background for given sensitivity are used to define data tern in (4). The data term is given by

$$E_{data}(\alpha) = -\sum_{p \in S} \log p(s_p \mid \alpha) \qquad (5)$$

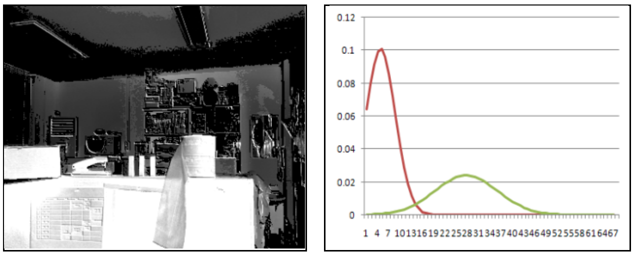
where s_p is sensitivity of the pixel p, and S is a set of all pixels in the image.

Smoothness term in (4) is boundary property to reduce noise and wrong sensitivity. If a neighbor pixel has similar intensity, it also has the same label. Also, if it doesn't have similar intensity, then different label should be assigned. This is called discontinuity preserving, and all smoothness term for energy minimization should satisfy this condition. One of the most effective smoothness terms is Gaussian distribution form for neighbor pixels [5]. Smoothness term for this algorithm is defined by

$$E_{smooth}(\alpha) = \sum_{\substack{(p,q) \in N \\ \alpha_p \neq \alpha_q}} \frac{1}{\| p - q \|} \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2} \right) \qquad (3)$$

where N is neighbor pixel set of pixel p and σ is variance of intensity.

The smoothness term is normalized form for the distance and intensity difference for neighboring pixels. If intensity difference is replaced by norm of color vector, more correct cost can be imposed. Using the two energy



(a) (b)

Fig. 4: (a) Sensitivity image and (b) distribution

term, energy minimization in (4) is applied to get final segmentation result. Energy minimization is done by Graph cut which is a robust and fast method [6]. Graph cut is a popular energy minimization method based on graph. Graph for the energy minimization has nodes and edge. Data costs and labels are set as node and smoothness cost as edge. From source to sink, max-flow is calculated to find min-cut. In this research, the structure of the Graph is not grid form but 8-neighbor form to impose more correct boundary condition.

5. EXPERIMENTAL RESULTS

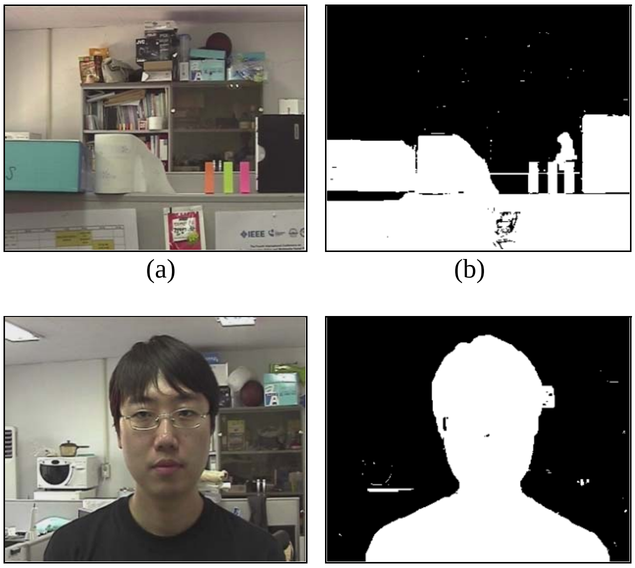
The proposed object segmentation algorithm is implemented in the C++ language with CIF size (352X288) images. Fig. 5 (b) is a final segmentation result of Fig. 1. As it shows, foreground is extracted precisely. However there are some noises which are came from mainly two problems. First one is matching point error of image array due to fine camera motion. In this case, intensity rate is not correct, and sensitivity becomes perfectly different value which is difficult to recover energy minimization. Another problem is reflectivity of medium of object. In Fig. 5 (a), there is a mirror in the middle of right side. Though it is a part of background, it is recognized as foreground due to high-reflective surface with high sensitivity. Fig. 6 is another experimental result using a video camera. By recording the scene, flash was lighted with range of luminous intensities. Then some frames influenced by flash were extracted to segment objects, and proposed algorithm is applied. In practice, the video camera is operated differently relative to the DSLR. Usually, it's difficult to control exposure, aperture and any other camera parameters. However, basically it also have similar feature to images captured by the DSLR, and as it show, the result is clear. Fig. 6 (d) is an example applied to a man, and this image also captured through video camera. The result shows that this algorithm is robust to most medium.

6. CONCLUSION

We proposed an image segmentation algorithm using various luminous intensities in this work, The proposed algorithm can provide the sensitivity for each pixel based on intensity rate of image. Also, it uses energy minimization method to get sophisticated result. Extensive simulation results show that the proposed algorithm provides a high quality segmentation results. For more correct results, sensitivity should be revised including reflection of medium.

(a) (b)

Fig. 5: (a) Original image (b) Segmentation result



(c) (d)

Fig. 6: (a)(c) Frame from video (b)(d) Segmentation result

7. REFERENCES

[1] C. Rother, V. Kolmogorov, and A. Blake, "Grab Cut: interactive foreground extraction using iterated graph cuts," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 309-314, Aug. 2004.

[2] C. Rother, V. Kolmogorov, T. Minka, and A. Blake, "Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs," *in Proc. CVPR*, Dec. 2006.

[3] J. Sun, J. Sun, S.B. Kang, Z. Xu, X. Tang, H. Shum, "Flash Cut: Foreground extraction with flash and no-flash image pairs," *in Proc. CVPR*, Jun. 2007.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, 47(1/2/3):7-42, April-June 2002.

[5] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in Proc. *ICCV*, vol. 1, pp. 105-112, Jul. 2001.

[6] Y. Boykov, O. Veksler, and R. Zabih. "Fast approximate energy minimization via graph cuts," In *ICCV*, volume I, pages 377–384, 1999.