FOREGROUND IMAGE EXTRCTION IN THE HSV COLOR SPACE

Jyh-Yeong Chang, Jia-Jye Shyu, and Yi-Cheng Luo

Department of Electrical and Control Engineering
National Chiao Tung University
Hsinchu, Taiwan, ROC

ABSTRACT

When the foreground color is different from the background color, the foreground subject can be extracted easily by the luminance component. When the foreground color is similar to the background color, we cannot extract the foreground image completely by the luminance component. To solve this, we utilize the HSV color space to build the background model, in line with similar spirit of W^4 segmentation algorithm, which can not only extract foreground image but also be helpful to shadow removal. Since H and S components are not reliable in some conditions, we make use of three criteria to obtain reliable and static hue values. In our experiment, extracting the foreground image in the HSV space improves the accuracy of the extracted foreground image.

***Index Terms*—**Foreground extraction, HSV color space, shadow removal

1. INTRODUCTION

Foreground subject extraction is an important step of the vision-based systems. such as automatic surveillance systems, human-machine interface, home care system and smart home applications. Many authors have developed methods of detecting foreground subject in images and videos [1]-[3]. Park and Aggarwal subtracted foreground pixels from background by computing Mahalanobis distance in each pixel in the HSV color model [1]. Jabri and Duric [2] used color and edge information to improve the quality and reliability of the results. Background subtraction is widely used for detecting moving objects from image frames of static cameras. Most of this work has been based on background subtraction using color or luminance component. In these approaches, difference between the coming frame and the background image is performed to detect foreground objects. W^4 [3] is a typical one with some modifications. It records the maximum and minimum luminance and the maximum interframe difference in every position of a frame in a background video. Then every image frame subtracts the maximum and minimum luminance at each position. If the pixel's absolute value of the subtraction operation is over the maximum interframe difference, the pixel is a foreground one. W4 only consider the luminance change to subtract the background. Under the W^4 framework, we cannot detect a foreground pixel correctly when it is similar in color to the background pixel. To make fully use of the spectrum of a pixel, it is imperative to do the segmentation in the color domain. To this end, foreground subject extraction is done in the HSV color space. We can have both the luminance information and the chromatic information in the background subtraction task. Furthermore, the moving cast shadows mostly exhibit a challenge for accurate foreground subject detection. A lot of attempts have been developed to tackle the shadow suppression [4]-[8] encountered in background subtraction. Horprasert *et al.* [4] and Cucchiara *et al*. [5] utilized the rationale that shadows have similar chromaticity, but lower brightness than the background model. Under the proposed framework in the HSV color space, we can effectively identify the shadow existent in our detected foreground subject.

Our foreground subject extraction model is composed of three components. The first component is foreground subject detection by luminance. The second component deals with the shadow suppression. And the final the component is the color compensation procedure to refine the foreground.

2. THE HSV COLOR SPACE

The HSV (hue, saturation and value) color space corresponds closely to the human perception of color. The hue parameter is the value which represents color information without brightness. Therefore, the hue is not affected by change of the illumination brightness and direction. Although hue is the most useful attribute, there are three problems in using hue attribute for color segmentation: (1) hue is meaningless when the intensity value is very low; (2) hue is unstable when the saturation is very low; and (3) saturation is meaningless when the intensity value is very low [7]. Accordingly, Ohba et al. [9] used three criteria (intensity value, saturation, and hue) to obtain the hue value reliably.

- **Intensity Threshold Value:**

If $V < V_t$, then $H = 0$, where V , V_t , and H are an intensity value, the intensity threshold value, and a hue value, respectively. If measured color is not bright enough, the color is discarded. Then, the hue value is set to a predetermined value, i.e., 0.

- **Saturation Threshold Value:**

If $S < S_t$, then $H = 0$, where S, S_t, and H are an saturation value, the saturation threshold value, and a hue value, respectively. Using this equation, measured color close to gray is discarded in the image.

- **Hue Threshold Value:**

If $H < \Delta P_t$ or $\|H - 2\pi\| < \Delta P_t$, then $H = 0$. The range of hue value is from 0 to 2π, and it has discontinuity at 0 and 2π. We use the hue threshold value ΔP_t to avoid the discontinuity effect.

3. FOREGROUND SUBJECT EXTRACTION

If we only use the luminance component to do background subtraction, we cannot detect reliably those foreground pixel whose luminance component close to background pixel. In order to solve this problem, we build our background model in the HSV color space. The HSV color space corresponds closely to the human perception of color. We can have the luminance information and the chromatic information simultaneously. Hue is unreliable in some condition, so we use the three criteria described in Section 2 to obtain the hue value reliably.

We build our background model with the minimum value ($[n^H(x,y),\ n^S(x,y), n^V(x,y)]$) and maximum value ($[m^H(x,y), m^S(x,y), m^V(x,y)]$) in each HSV domain. Besides, we also record the inter-frame ratio ($d^V(x,y)$) in the brightness component and the inter-frame different ($[d^H(x,y), d^S(x,y)]$) in the chromatic component.

After building background model, foreground objects can be segmented from every frame. Our framework of foreground subject extraction is composed of three components. The first component is foreground subject extraction. The second component is the shadow suppression. And the final component is the color compensation to recover the foreground pixels wrongly classified to the background due to their high luminance similarly.

A. Foreground Subject Detection by Luminance

Each pixel of the video frame is classified to either a background or a foreground pixel by the ratio between the background model and a captured image frame.

$$B(x,y) = \begin{cases} 0, & \text{if } I_i^V(x,y)/m^V(x,y) < k_V d^V(x,y) \\ & \text{or } I_i^V(x,y)/n^V(x,y) < k_V d^V(x,y) \\ 255, & \text{otherwise} \end{cases} \tag{1}$$

where $I_i^V(x,y)$ is the intensity of a pixel which is located at (x,y), $B(x,y)$ is the gray level of a pixel in a binary image, and k_V is a threshold, determined by light sufficiency of the scene. The value of k_V is normally set to 1.3 for normal light condition, and k_V will be reduced for in-sufficient light condition and increased otherwise.

B. Shadow Suppression

The pixels of the moving cast shadows are easily detected as the foreground pixel in normal condition. Because the shadow pixels and the object pixels share two important visual features: motion model and detectability. For this reason, the moving shadows cause object merging and object shape distortion. Horprasert *et al.* [4] and Cucchiara *et al.* [5] utilize the rationale that shadows have similar chromaticity, but lower brightness than the background model. Hence, we can detect the shadow from foreground subject in the HSV color space.

We define a shadow mask S for each (x,y) point as follows:

$$S(x,y) = \begin{cases} \text{shadow}, & \text{if } I_i^V(x,y) - n^V(x,y) < 0 \\ & \text{and } \left|I_i^H(x,y) - m^H(x,y)\right| < k_H d^H(x,y) \\ & \text{and } \left|I_i^S(x,y) - m^S(x,y)\right| < k_S d^S(x,y) \\ \text{object}, & \text{otherwise} \end{cases} \tag{2}$$

where $I_i^H(x,y)$, $I_i^S(x,y)$, and $I_i^V(x,y)$ are respectively the HSV channel of a pixel located at (x,y). Values k_S and k_H are selected threshold values used to measure the similarities of the hue and saturation between the background image and the current observed image. We can utilize the shadow mask $S(x,y)$ to change the shadow pixels into background in $B(x,y)$ and obtain $B_S(x,y)$ being shadow suppressed.

C. Color Compensation

Some colors such as yellow, pink, and light blue have similar luminance value. If we only use the luminance component to do background subtraction, we cannot detect foreground pixel correctly when its luminance is similar to that of a background pixel. In order to improve detectability, background subtraction is computed by taking into account not only a pixel's luminance, but also its chromaticity. We want to use the chromaticity to enhance the accuracy of the foreground object. Based on the amount of the chromaticity change, we reanalyze its background in B_S to be changed to a foreground of object, by

$$B_f(x,y) = \begin{cases} 255, & \text{if } \left|I_i^S(x,y) - m^S(x,y)\right| > k_S d^S(x,y) \\ & \text{or } \left|I_i^H(x,y) - m^H(x,y)\right| > k_H d^H(x,y) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $I_i^H(x,y)$ and $I_i^S(x,y)$ are respectively the hue and saturation components of a pixel at (x,y), k_S and k_H are selected threshold values. B_f is the final foreground object after the refined step of Eq. (3).

4. EXPERIMENTAL RESULTS

In our experiment, we tested our system on videos taken by digital camera. We took the video in our laboratory at the 5th Engineering Building in our university campus. The camera has a frame rate of thirty frames per second and image resolution is 320×240 pixels. The background is not complex and we equipped a table in the scene. The light source is fluorescent lamps and is stable.

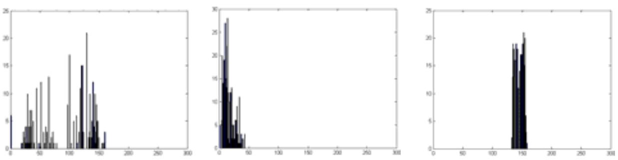
We test the foreground detection capability in two cases depending on the color of clothing worn by action subjects. That the action subject wore the clothing with color different from that of background is the first case. And the second case is that the action subject wore the clothing with color similar to that of background. When the color of clothing and background are similar in the second case, a moving object, such as human body, may not be segmented easily from image frame. We compare the result in these two cases and the color compensation in our proposed system demonstrates eminent improvement in the segmentation quality.

A. Background Model Construction

We built the background model in the HSV color space. The value of H or S or V is between 0 and 255. Figs. 1(a), 1(b), and 1(c) show the H, S, and V component of the background image, respectively. We can find from these three figures that the hue value is relatively unstable when the saturation is close to zero. We make an experiment to test the changes in the HSV components in constructing the background model. Fig. 2 represents the H, S, and V variations a pixel at coordinate $(x, y) = (120,\ 160)$ during the first 300 frames of the background video. From Fig. 2, we can see that V component is the most stable of the background model. H and S components are less stable.

(a) (b) (c)

Fig. 1. Background image. (a) The H component, (b) The S component, and (c)The V component.



(a) (b) (c)

Fig. 2. H, S, and V variations of the pixel at (120,160) versus frame index of background video from frame 1 to frame 300. (a) H, (b) S, and (c) V.

Hence, we use three criteria (V_t, S_t, H_t) to obtain the hue value reliably in building the background model. In our experiment, we set three criteria of Sec. 2 by

$$V_t = 50,\ S_t = 50, \text{ and } H_t = 25 \quad (4)$$

to make hue value reliably.

Fig. 3 shows the background image in the H color components after we use criterion to redefine it. We can find that the hue values in the background image are almost be set to zero. The reason is that our background is simple and the color is similar to the gray tones.

Fig. 3. Background image in the redefined H component.

B. The Experiment of Foreground Subject Extraction

We use the shadow mask in Eq. (2), with k_V = 1.3, k_H = 2, k_S = 2, and k_V = 1.3 in Eq. (1), to classify the pixels whether it is a shadow point or not. Fig. 4 shows the process result in shadow detection. Fig. 4(a) is input image. Fig. 4(b) is the foreground subject without shadow suppression. The foreground subject with shadow suppression is shown in Fig. 4(c).

(a) (b) (c)

Fig. 4. The example of the shadow suppression.

In what follows, the effectiveness of color compensation in obtaining a more accurate foreground is described. In Fig. 5, the left column contains three input images; the middle column contains the resulting foreground images, without color compensation step; and the right column is the foreground images detected with color compensation step. From the Fig. 5, we have found that we can get good compensation when the clothing color is light blue and yellow, but cannot obtain good compensation when the clothing color is pink. The reason is that when pink color pixels are transformed from RGB color space to HSV color space, the saturation of pink is lower than the set criterion S_t. Hence, we cannot recover those pixels from background to foreground for such small chromaticity difference in this space.

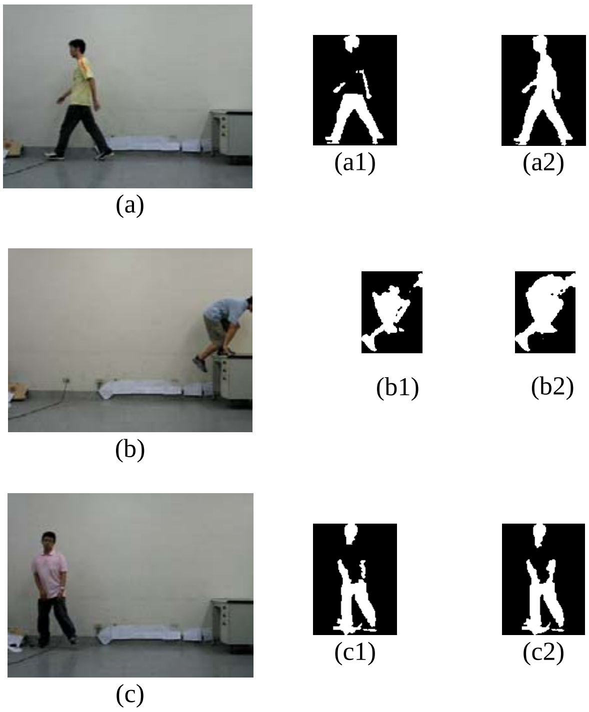


Fig. 5. Foreground detection without/with color compensation. (a) − (c) is the input images; (a1) − (c1) the extracted foreground images without color compensation; (a2) − (c2) the extracted foreground images with color compensation.

We randomly selected 100 frames from the video sequence of the model with a subject wearing clothing similar to the background color. The "foreground subject ground truths" of these 100 frames were generated manually. Let A be the detected foreground subject region and B be the corresponding "ground truth." Then we test the pixel accuracy by the following two metrics. To this end, accuracy rate$_1$, called *recall accuracy* [10] is calculated by

$$\text{Accuracy rate}_1 = \frac{N_s}{N_{total}} \times 100\% , \tag{5}$$

where N_{total} is the pixel number of segmented foreground image, i.e. true positive and false positive pixels and N_s is the true positive pixels segmented in A. accuracy rate$_2$ is adopted from [11] by

$$\text{Accuracy rate}_2 = \frac{A \cap B}{A \cup B} \times 100\% \tag{6}$$

This measure counts the percentage of the mutual positive pixels to expanded positive pixels. Table I shows the accuracy rates of 100 frames, and demonstrates the improvement of color compensation over that without color compensation.

TABLE I.
COMPARISON RESULT OF THE PIXEL ACCURACY RATES OVER 100 IMAGES

	Without color compensation	With color compensation
Acc_1	78.81%	89.13%
Acc_2	59.61%	81.23%

5. CONCLUSION

In this paper, we proposed the foreground subject extraction in the HSV color space. In the HSV color space, we can utilize not only the luminance component but also the chromatic component existent in the background image. In this way, we can reliably extract the foreground subject, even when the foreground luminance is similar to that of the background. Experiment results have shown that extracting the foreground image in the HSV space and improve the pixel accuracy of the foreground image segmented.

Some subjects wearing light color clothing, e.g., pink, still cannot be extracted well, which deserves to be investigated further. In addition, extensions of various test environments and more complicated surrounding are our future work.

ACKNOWLEDGEMENT

This research was supported in part by the grants NSC 95-2752-E-009-011-PAE, 95-EC-17-A-02-S1-032, and NSC NSC 95-2221-E-009-212, Taiwan, ROC.

REFERENCES

1. S. Park and J. K. Aggarwal, "Segmentation and Tracking of Interacting Human Body Parts under Occlusion and Shadowing," in *Proc.Workshop on Motion and Video Computing*, pp.105 — 111, 2002.
2. S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and Location of People in Video Images Using Adaptive Fusion of Color and Edge Information," in *Proc. Int. Conf. Pattern Recognition*, pp. 627 — 630, 2000.
3. I. Haritaoglu, D. Harwood, and L. S. Davis, "W^4: Real-Time Surveillance of People and Their Activities," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 809 — 830, 2000.
4. T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection," in *Proc. IEEE ICCV '99*, 1999.
5. R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," in *Proc. IEEE Intell. Transport. Sys. Conf.*, pp. 334 — 339, 2001.
6. A. Prati, I. Mikic, M. Trivedi and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," *IEEE Trans.Pattern Anal. Machine Intell.*, vol. 25, no. 7, pp. 918 — 923, 2003.
7. B. Chen and Y. Lei, "Indoor and Outdoor People detection and Shadow Suppression by Exploiting HSV Color Information," in *Proc. Fourth Int. Conf. Computer and Information Technol.*, pp. 137 — 142, 2004.
8. S. Vitabile, G. Pilato, G. Pollaccia, and F. Sorbello, "Road Signs Recognition Using a Dynamic Pixel Aggregation Technique in the HSV Color Space," in *Proc. 11th Int. Conf. Image Anal. and Processing*, pp. 572 — 577, 2002.
9. K. Ohba, Y. Sato, and K. Ikeuchi, "Appearance-based visual learning and object recognition with illumination invariance," *Machine Vision and Applications*, Vol. 12, No. 4, pp. 189 — 196, 2000.
10. J. Fan, Y. Gao, H. Luo, "Multi-level annotation of natural scenes using dominant image components and semantic concepts", in *Proc. ACM Multimedia*, pp. 540 — 547, 2004.
11. L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," *IEEE Trans.Image Processing*, vol.13, no.11, pp.1459 — 1472, 2004.