Effective Moving Cast Shadow Detection for Monocular Color Image Sequences

George S.K. Fung, Nelson H.C. Yung, Grantham K.H. Pang, Andrew H.S. Lai

Laboratory for Intelligent Transportation Systems Research
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong SAR
E-mail: skfung, nyung, gpang, hslai@eee.hku.hk

Abstract

For an accurate scene analysis in monocular image sequences, a robust segmentation of moving object from the static background is generally required. However, the existence of moving cast shadow may lead to an inaccurate object segmentation, and as a result, lead to further erroneous scene analysis. In this paper, an effective detection of moving cast shadow in monocular color image sequences is developed. Firstly, by realizing the various characteristics of shadow in luminance, chrominance, and gradient density, an indicator, called shadow confidence score, of the probability of the region classified as cast shadow is calculated. Secondly, canny edge detector is employed to detect edge pixels in the detected region. These pixels will then be bounded by their convex hull, which estimate the position of the object. Lastly, by analyzing the shadow confidence score and the bounding hull, cast shadow is identified as those regions outside the bonding hull and with high shadow confidence score. A number of typical outdoor scenes are evaluated and it is shown that our method can effectively detect the associated cast shadow from the interested object.

1. Introduction

Object segmentation in image sequence analysis has been actively studied in recent years [1]. In many application areas, including visual surveillance, video compression and others, stationary background estimation is one of the most common approaches in segmenting the moving object from the scene. However, in real outdoor daylight scenes, shadows cast by moving objects such as vehicles or pedestrians are often detected as a part of the moving objects since shadows move in accordance with the movement of objects. When the detected objects contain shadows, large errors may occur with respect to the estimation of the number of objects from their sizes,

analysis of their shape, and estimation of their locations. This also creates a multitude of problems associated with occlusion. Thus, it is important to separate cast shadows from the objects.

Being aware of the importance of shadow detection, numerous shadow detection methods have been proposed in the last decade [1-6]. In [1], by using the intra-frame difference and explicitly detecting the penumbra and umbra properties of shadow, a detection method for ideal indoor shadow was proposed. The major limitations of this method are: weak shadows could not be detected; shadows could be entirely detected only if they entirely cover a new background; and the assumptions are not realistic for real outdoor scene. In [2], a method for partly eliminating shadows accompanied with pedestrian-like moving objects in outdoor visual surveillance systems was presented. To detect the zero height property of the shadow, two cameras were used with their common visual fields on the surveillance area. By subtracting the images transformed from one camera and acquired from the other camera, shadows could be removed since shadows that exist on the road plane would occupy the same areas in these images. The major limitations of this method are: only shadow on the road plane can be removed, and there is possible shadow misclassification caused by overlapping of adjacent objects in transformed image. In [3], the author stated that, in general, a shadow is edge-less, but object, such as vehicle, has significant edges especially at their borders. It was claimed that by using high-level knowledge about the geometry of the scene (heading of the observed road), about global data (date and time), and so called subtracted edge histograms of the interested areas, cast shadow was partially separated from the corresponding detected vehicle. The major limitations of this method are: unrealistic requirement of environment knowledge; only partial shadow was detected by bounding box; and only edge property was used to guide the search. In [4], an illumination model or shadow formation model was defined to allow the shadow of the interested object to be

taken into account. As reported, if the illumination model is omitted, incorrect model interpretations resulted in experiments with real-world traffic scenes in which vehicles exhibited salient shadow edges. The incorrect model interpretations would further induce error in the estimation of the position and orientation of the vehicle model. To overcome this problem, a simple illumination model, which assumed parallel incoming light, was used and the visible contour of the 3-D vehicle model projected onto the street plane was computed. This illumination model approach may be feasible only for limited real outdoor surveillance purpose. It is limited by a priori settings of the parameters for the illumination direction, the unrealistic assumptions of the simplified illumination model, and highly complex interpretation required. In [5], the authors realized two different shadow detection approaches: looking for shadows from still objects in the scene, and relying on the consideration that shadows were more similar than the actual objects between corresponding targets. By exploiting the Hue, Luminosity, Saturation (HLS) color components, an algorithm that extracted a shadow model from a monocular color scene was presented. The limitations of this method are: the analysis does not work well when there are not enough shadows in the reference image, all the shadow need to be cast on the same kind of surface, and only HLS information were used.

Although numerous shadow detection methods have been proposed, they all suffer from certain limitations that make them ineffective in real outdoor environment. In this paper, we propose an effective detection method based on the various aspects of shadow properties. In the next section, the basic concept and methodology of proposed detection method is highlighted. Then, in section 3, by realizing the various characteristics of cast shadow in luminance, chrominance, and gradient density, the calculation of an indicator, called shadow confidence score, which gives a probability of the region classified as cast shadow is defined. In section 4, the process of defining the object by bounding convex hull on the selected edge pixels is described. Thus, cast shadow can be identified as those pixels outside the object. In section 5, a number of typical outdoor scenes are evaluated and analyzed. Finally, we conclude that our method can effectively separate the associated cast shadow from the interested object.

2. Concept and Methodology

In essence, shadows occur when objects partially or completely occlude direct light from a light source. As defined in [6], there are two parts in a shadow: the self-shadow and the cast shadow. The self-shadow is the part

of object, which is not illuminated by direct light. The cast shadow is the region projected by the object in the direction of direct light. In this paper, our objective is to detect the cast shadow from the object. Although the formation of cast shadow depends on various factors of the environment, with the estimation of the background, there are four generic properties of cast shadow that could be used to guide our proposed shadow detection method:

Property 1: The luminance of the cast shadow is lower than the background.
Property 2: The chrominance of the cast shadow is identical or slightly shifted when compared with background.
Property 3: The difference in gradient density between the cast shadow and background is lower than the difference in gradient difference between the object and background.
Property 4: The cast shadow is at the boundary region of the moving foreground mask. That is, the cast shadow can be formed in any direction of the object, but not inside the object.
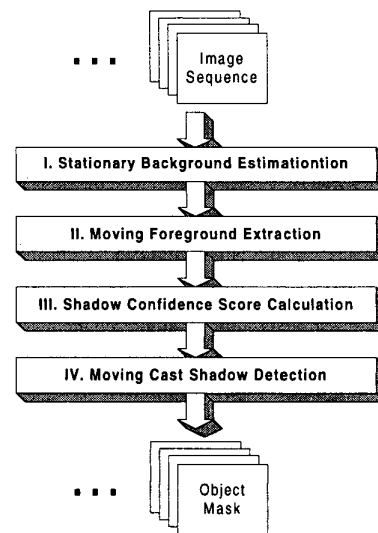

Figure 1. Proposed Method

In order to extract the moving object without the cast shadow from the stationary background in an image sequence, our proposed methodology consists four-steps as depicted in Figure 1: I. Stationary background estimation, II. Moving foreground extraction, III. Shadow confidence score calculation and IV. Moving cast shadow detection.

In Stage I, any background estimation algorithms can be used to generate a stationary background from a sequence of images. Lately, in [7], a fast and accurate

scoreboard based algorithm for estimating stationary background was discussed. This algorithm has been adopted for Stage I computation. Then, in Stage II, the moving foreground mask is identified by the subtraction of the background image from the input image. Mathematical morphological closing was employed to join the disjoint regions that belong to the same object in the resulting image. In Stage III, by realizing the various characteristics of cast shadow in luminance, chrominance, and gradient density, the shadow confidence score is calculated from various mapping functions defined according to the cast shadow's characteristics stated above. In Stage IV, based on the shadow confidence score calculated and the significant edge detected in the input image, the object and cast shadow are separated accordingly.
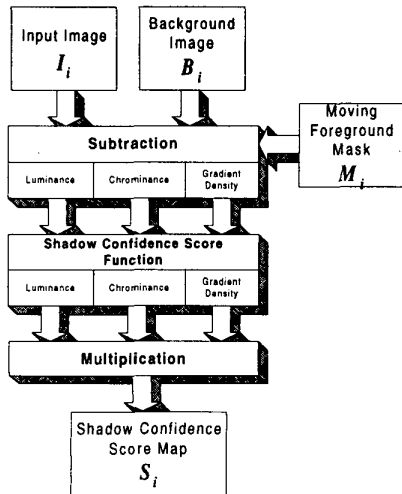
3. Shadow Confidence Score Calculation

Figure 2. Shadow Confidence Score Calculation

Let the current image and the background image be defined respectively as follows:

$$I_i(x,y) = \begin{Bmatrix} l_{I,i}(x,y) \\ c_{I,i}(x,y) \\ g_{I,i}(x,y) \end{Bmatrix}, (1) \qquad B_i(x,y) = \begin{Bmatrix} l_{B,i}(x,y) \\ c_{B,i}(x,y) \\ g_{B,i}(x,y) \end{Bmatrix}, (2)$$

where

$x = 0,...,W-1$, $y = 0,...,H-1$,

i is the frame number, W is the width of the image, H is the height of the image, $l_{I,i}(x,y)$ is the luminance of the image of pixel (x,y), $c_{I,i}(x,y)$ is the chrominance of the image of pixel (x,y), $g_{I,i}(x,y)$ is the gradient density of the image of pixel (x, y). Let the foreground mask be defined as

$$M_i(x,y) = \begin{cases} 1, & I_i(x,y) \text{ is a foreground pixel} \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

To indicate whether the region should be classified as cast shadow, a shadow confidence score, S, is defined. If the region is likely to be a cast shadow, a high score S will be given to that region. On the other hand, if the region is likely to be object or background, a lower score will be given. The score is a probability value ranged from 0 to 1 inclusive. Only the shadow confidence scores within the extracted region in the moving foreground mask generated in Stage II are calculated.

As shown in Figure 2, the luminance, chrominance, and gradient density levels for each pixel in the input and background images are calculated. Then, the subtraction of the two images is calculated in the luminance, chrominance, and gradient density dimensions. To calculate the score S, the three mapping functions are defined, which are Luminance Score (S_L) vs Luminance Difference, Chrominance Score (S_C) vs Chrominance Difference and Gradient Density Score (S_G) vs Gradient Density Difference. Then, the overall score (S) is computed by combining these three individual scores.

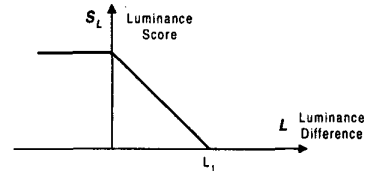## A. Luminance Score (S_L) vs Luminance Difference (L) Function

Figure 3. S_L vs L Function

Let $L_i(x,y) = l_{I,i}(x,y) - l_{B,i}(x,y)$

$$\forall(x,y) \text{ where } M_i(x,y) = 1, \qquad (4)$$

$$S_L = \begin{cases} 1, & L_i(x,y) \le 0 \\ (L_1 - L_i(x,y))/L_1, & 0 < L_i(x,y) < L_1 \\ 0, & L_i(x,y) \ge L_1 \end{cases} \qquad (5)$$

The mapping function of Luminance Score (S_L) against Luminance Difference (L) is depicted in Figure 3. L_1 is a predefined parameters to accommodate the acquisition noise. As defined in **Property 1**, shadows are formed when light is occluded by objects, the luminance level must be lower in input image comparing with the background image at the shadow. Therefore, for negative luminance difference value, the cast shadow criteria is satisfied and the region is most likely to be a cast shadow. In the opposite, when the luminance level is higher in input image comparing with the background image (positive luminance difference value), it does not satisfy the shadow criteria and S_L tends to zero. For L between 0 and L_1, a linear mapping from 0 to 1 is chosen to provide a smooth transition.

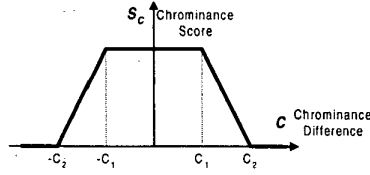## B. Chrominance Score (S_C) vs Chrominance Difference (C) Function

Figure 4. S_C vs C Function

Let $C_i(x, y) = c_{I,i}(x, y) - c_{B,i}(x, y)$

$$\forall(x, y) \text{ where } M_i(x, y) = 1, \tag{6}$$

$$S_C = \begin{cases} 1, & |C_i(x,y)| \le C_1 \\ (C_2 - C_i(x,y))/(C_2 - C_1), & C_1 < |C_i(x,y)| < C_2 \\ 0, & |C_i(x,y)| \ge C_2 \end{cases} \tag{7}$$

The mapping function of Chrominance Score (S_C) against Chrominance Difference (C) is depicted in Figure 4. C_1 and C_2 are predefined parameters to accommodate the tolerance to chrominance change. As defined in **Property 2**, the chrominance of the input and background images should be the same except in cast shadow where there is less light shining on it. Thus, we have observed that the change will only occur in luminance dimension and there should be very small or no change in chrominance level. Therefore, for C between $-C_1$ and $+C_1$, S_C is set to 1 since it satisfies the shadow criteria (small change) in the chrominance dimension. For C larger than C_2 or C smaller than $-C_2$, S_C is set to 0 since there is large change in chrominance dimension. Smooth transition from 0 to 1 is implemented for the rest of the range of chrominance difference.

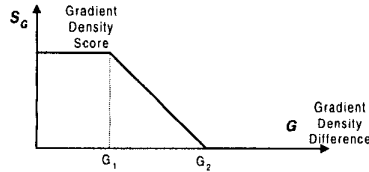## C. Gradient Density Score (S_G) vs Gradient Density Difference (G) Function

Figure 5. S_C vs G Function

Let $G_i(x, y) = g_{I,i}(x, y) - g_{B,i}(x, y)$

$$\forall(x, y) \text{ where } M_i(x, y) = 1, \tag{8}$$

$$S_G = \begin{cases} 1, & G_i(x,y) \le G_1 \\ (G_2 - G_i(x,y))/(G_2 - G_1), & G_1 < G_i(x,y) < G_2 \\ 0, & G_i(x,y) \ge G_2 \end{cases} \tag{9}$$

The mapping function of Gradient Density Score (S_G) against Gradient Density Difference (G) is depicted in Figure 5. G1 and G2 are predefined parameters. As defined in **Property 3**, after subtraction of gradient density in the input and background images, the gradient density level is mostly cancelled out in the cast shadow. However, in the object region, there is significant difference between the input and background images in gradient density. Therefore, for small gradient density difference value, the region is more likely to be shadow and S_G is set to 1. For high gradient density value, the region is likely to be an object and S_G is set to 0. Smooth transition from 0 to 1 is implemented for the rest of the range.

D. Combined Shadow Confidence Score (S)

$$S_i(x, y) = S_{L,i}(x, y) \times S_{C,i}(x, y) \times S_{G,i}(x, y). \tag{10}$$

Finally, after the three scores, S_L, S_C and S_G, are calculated for the three difference dimensions, the total S should be computed by combining all the three scores. Since each dimension is a necessary requirement for the region to be classified as cast shadow, hence a direct multiplication of S_L, S_C and S_G.
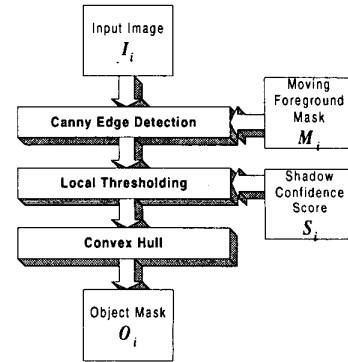
4. Moving Cast Shadow Detection

Figure 6. Moving Cast Shadow Removal

In Stage III, the overall shadow confidence score is calculated, where in Stage IV, the cast shadow is separated from the object based on S and significant edges of the input image. Firstly, with the input image and moving foreground mask, all the pixels with significant gradient level are detected using the canny edge detector within the moving foreground region. Secondly, for each pixel with high gradient level, a local thresholding test is applied to filter out the pixel with low shadow confidence score level. This test removes noise and edge pixels inside and around the border of shadow. If the high gradient pixel has high local shadow confidence score level, the pixel will be retained; otherwise, it will be discarded. Thirdly, as defined in **Property 3** and **Property 4**, the object can be segmented out from the foreground mask by bounding convex hull on the selected pixels. Based on these properties, the cast shadow can be identified as the moving foreground excluding the object.

5. Experimental Results and Analysis

To test the effectiveness of the proposed method, three different typical traffic image sequences captured under different lighting conditions, including sunny and cloudy, have been tested. For the first image sequence, the output images of different stages of the proposed method are presented in Figure 7. An image of a red truck with light gray trunk, which captured on a cloudy day, is shown in Figure 7(a). In Figure 7(b), the background color image was generated by the background estimation method of [7]. After subtraction of the input and background images, the moving foreground mask after closing transform is shown in Figure 7(c), where the black region is the background. In Figure 7(d), (e) and (f), S_L, S_C and S_G for the foreground region are shown accordingly. In Figure 7(d), except the roof of the trunk, most region of the truck is recognized as shadow since the red color of the truck and the gray color of some part of the cargo have similar luminance level as the background image. Therefore, luminance can only provide limited indication

on the shadow confidence level. In Figure 7(e), the chrominance level of the red part of the front of the truck clearly classifies it as non-shadow. In Figure 7(f), the regions with high gradient density difference are clearly marked as non-shadow. By multiplying the S_L, S_C and S_G, the S is shown in Figure 7(g), in which, only the cast shadow has high S value while the other part of the foreground mask is mostly covered with low S value. In Figure 7(h), the background, object and shadow are shown in gray, white and black colors respectively after performing the black pixels within the object are the significant edge pixels. The results of the second and third image sequences are depicted in Figures 8 and 9 respectively. From these results, it can be concluded the shadows of all three image sequences have been successfully detected. However, since the lower part of the vehicle and the tires exhibit cast shadow properties on the road surface, the shadow is overly segmented which covers part of the object. For the purpose of vehicle tracking and classification, this over segmentation error will not significantly affect the accuracy of the system.
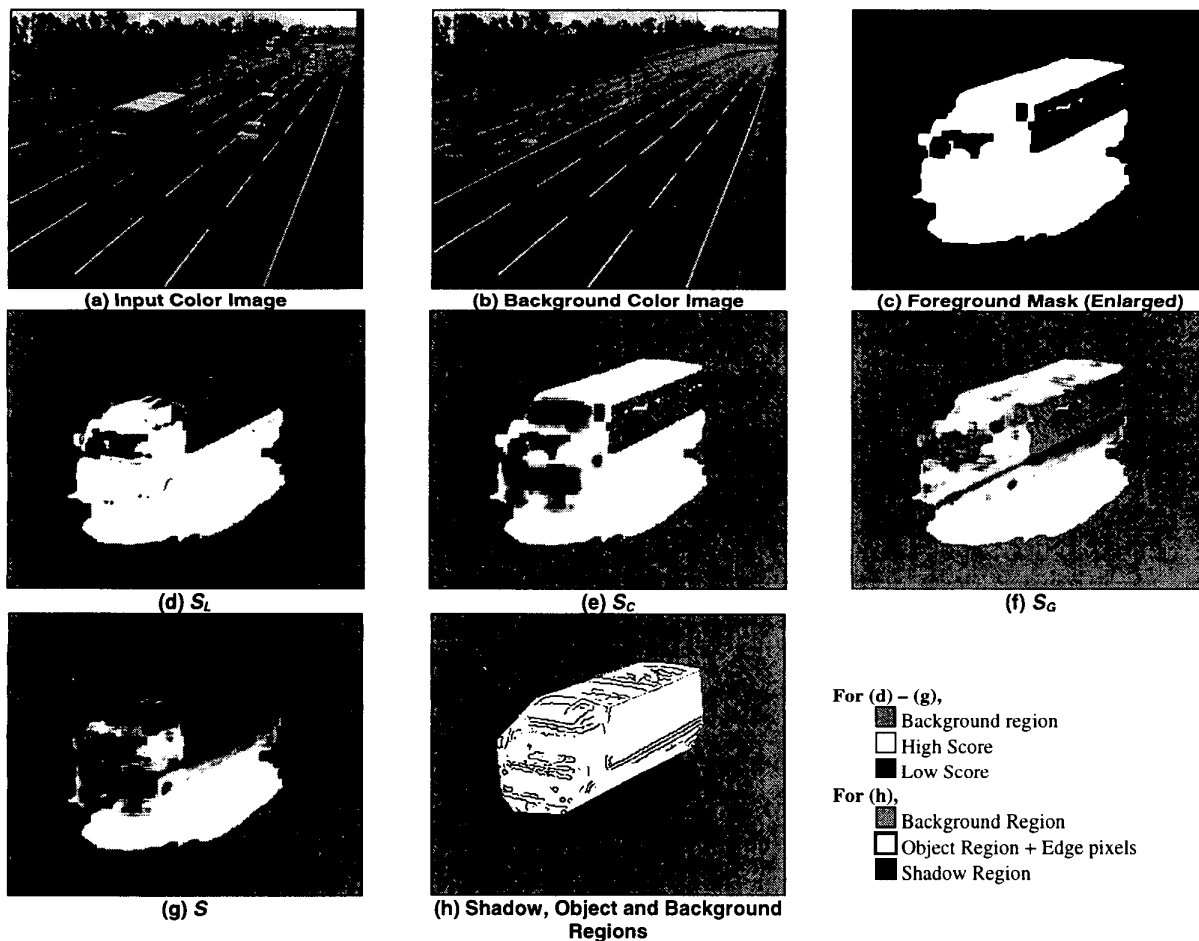

(a) Input Color Image

(b) Background Color Image

(c) Foreground Mask (Enlarged)

(d) S_L

(e) S_C

(f) S_G

(g) S

(h) Shadow, Object and Background Regions

For (d) – (g),
  ▨ Background region
  ☐ High Score
  ■ Low Score
For (h),
  ▨ Background Region
  ☐ Object Region + Edge pixels
  ■ Shadow Region

Figure 7. Experimental Results on Image Sequence 1

408

(a) Input Color Image   (b) Foreground Mask (Enlarged)   (c) Shadow, Object and Background Regions
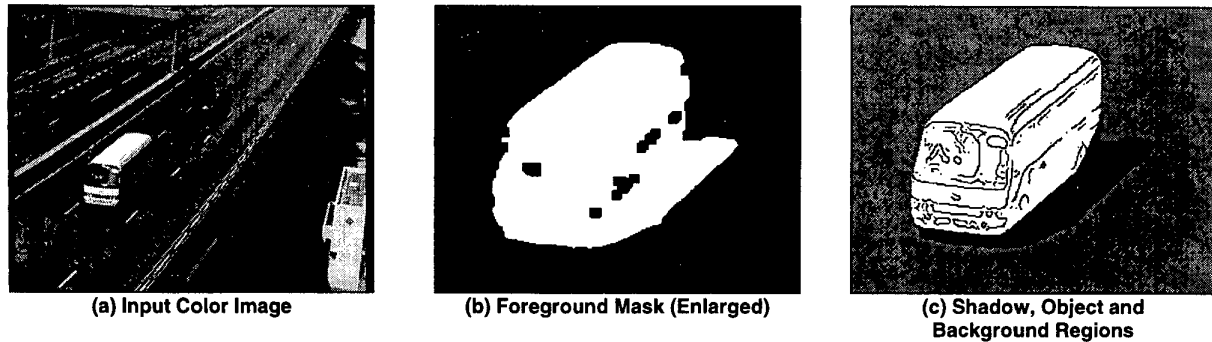
Figure 8. Experimental Results on Image Sequence 2

(a) Input Color Image   (b) Foreground Mask (Enlarged)   (c) Shadow, Object and Background Regions
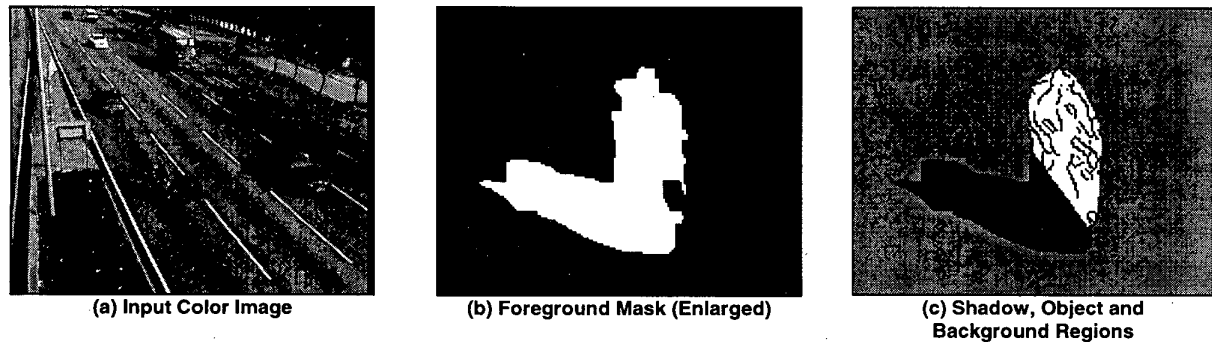
Figure 9. Experimental Results on Image Sequence 3

6. Conclusions

From the generic properties of cast shadow, an effective moving cast shadow detection method for monocular color image sequences has been presented. Firstly, the stationary background is estimated from a sequence of images. Secondly, the moving foreground mask is determined by the difference between the input and background image with mathematical morphology transformation applied. Thirdly, a shadow confidence score is computed in three different dimensions of the input and background difference image. Finally, based on the shadow confidence score and the edge pixel, the cast shadow is separated from object in the foreground mask. From the experimental results and analysis of the testing scene, our proposed method can successfully detect the cast shadow in the foreground mask. The method is able work in environment with different object types, in different lighting condition, in outdoor environment and without illumination model. However, the shadow is overly segmented in our results.

Further testing on traffic and pedestrian image sequences under different lighting conditions are in progress to further study the feasibility of our method, such as the impact of object's color and other physical properties. In additional, a set of fuzzy inference rules for computing the shadow confidence score can be developed to further improve the robustness of our method.

7. References

[1] Stauder, J., Mech, R., and Ostermann, J., "Detection of Moving Cast Shadows for Object Segmentation," *IEEE Transactions on Multimedia*, Vol. 1, No. 1, p. 65-76, 1999.

[2] Onoguchi, K., "Shadow Elimination Method for Moving Object Detection," *Proceedings of Fourteenth International Conference on Pattern Recognition*, Vol. 1, p. 583-587, 1998.

[3] Kilger, M., "A Shadow Handler in a Video-based Real-time Traffic Monitoring System," *Proceedings of IEEE Workshop on Applications of Computer Vision*, p. 11-18, 1992.

[4] Koller, D., Danilidis, K., and Nagel, H.H., "Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes," *International Journal of Computer Vision*, Vol. 10, No. 3, p. 257-281, 1993.

[5] Gamba, P., Lilla, M., and Mecocci, A., "A Fast Algorithm for Target Shadow Removal in Monocular Colour Sequences," *Proceedings of International Conference on Image Processing*, p. 436-439, 1997.

[6] Jiang, C., and Ward, M.O., "Shadow Identification," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 606-612, 1992.

[7] Lai, A.H.S., and Yung, N.H.C., "A Fast and Accurate Scoreboard Algorithm for Estimating Stationary Backgrounds in an Image Sequence," *Proceedings of IEEE International Symposium on Circuits and Systems*, Vol. 4, p. 241-244, 1998.