Moving Shadow Detection Using a Physics-based Approach

Sohail Nadimi and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, California, 92521, USA
{sohail, bhanu}@cris.ucr.edu

Abstract

Moving object detection systems generally detect shadows cast by the moving object as part of the moving object. In this paper the problem of separating moving cast shadows from the moving objects in outdoor environment is addressed. Unlike other previous work, we provide a method that does not use any geometrical information. Our physics-based approach is based on a spatio-temporal albedo normalization test and a dichromatic reflection model. The physics based model is used both in the estimation and verification phases. We provide results for several different video sequences representing a variety of materials and shadows. We achieve excellent results in distinguishing moving objects from their shadows. The results indicate that our approach is robust to a variety of background and foreground materials and varying illumination conditions.

1. Introduction

Over the past several decades, many approaches have been developed for moving object detection for indoor and outdoor scenes. Moving object detection methods fall into the following general categories: 1) background subtraction, 2) temporal differencing, 3) optical flow, and 4) statistical modeling. The first two methods clearly detect moving shadows as well as moving objects. This is due to the fact that intensity changes in the scene caused by moving shadows are as large as those of the new objects in the scene. Optical flow on the other hand, measures the spatio-temporal image gradient to estimate motion in the scene. Since the shadow of a moving object potentially moves, the shadow also contributes to the optical flow. In statistical modeling, the probability of observing values for a pixel are estimated based on previous observations. Like a new object, its shadow has not been observed in the past; thus, probability of observing shadow of a new (moving) object is the same as observing the object itself; hence, detection of an object includes detection of its shadow as well.

Shadows are typically divided into static and dynamic shadows. Static shadows are shadows due to static objects such as buildings, parked cars, swaying trees, etc. Shadows due to static objects and objects with repetitive motion can be modeled by statistical-based approaches [5]. In this paper we concentrate on dynamic shadows due to moving objects such as pedestrians, cars, trucks, etc. In addition, shadows can have umbra, penumbra or both. Under the cloudy sky, the shadows are either weak (mostly penumbra) or non-existent. Our research focuses on outdoor scenes where we have a far away point source (sun) and a diffuse source (sky) contributing to the illumination in the scene. Since the distance between the objects and the background is negligible compared to the distance of illumination sources to the objects, most or all of the shadows are umbra. Outdoor scenes provide challenging problems in the sense that there is no control over illumination in the scene and limited or no knowledge may be available about the scene geometry.

We have developed a physics-based algorithm for distinguishing a moving object and its shadow in outdoor scenes. The salient features of our approach are: a) *Integration of different physical models*: sound physical models such as albedo and dichromatic reflection models are used in an integrated manner; b) *Temporal improvement*: spatial albedo-ratio test is extended temporally, called spatio-temporal albedo-ratio test, and it is utilized for surface segmentation; c) *Experiments:* the results of the algorithm are tested using a variety of real world data; and d) *No geometrical assumptions* are made.

2. Previous Work

There have been several approaches to segment shadows of objects. In [1] knowledge of the object geometry is used to provide proximity, collinearity and convexity constraints to separate airplanes from their shadows in aerial photographs. Image projections of binary images are used [2] to detect pedestrians and their shadows assuming that objects are erect and knowledge of the light source is given. By knowing the location of the sun with respect to the camera and assuming that shadows are larger than a certain size, [3] utilized a histogram based approach to separate shadows of ground vehicles. Binocular vision is also used for shadow detection [4]. By assuming that the

1051-4651/02 $17.00 © 2002 IEEE

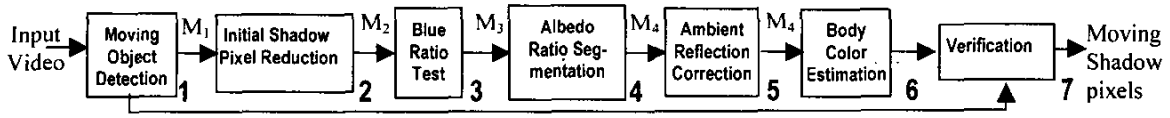

Figure 1. Different steps of physics-based shadow detection algorithm. M_i is the binary Mask of potential shadow pixels updated after each step.

shadows are formed on the ground plane and ground plane being flat, one could project the image formed on one camera back to the second camera plane via the ground plane. In this manner, shadows tend to occupy the same location on both images. All the above approaches either utilize geometric information or restrict the classes of the objects, while none exploits the temporal information present in video data.

Our approach is different from the previous work in that we solely rely on physics-based models, which can incorporate variety of classes of materials. We make no assumption about geometry including the sun position, types of objects and background and whether the background is already in shadow or not. We only require that the background image does not contain any moving objects. Note that moving objects do not cast shadow in regions where background is already shadowed. This is due to the fact that the same illumination source (e.g., sun) is blocked for both the object's shadow and the background shadowed region. We use the fact that the sky is blue due to the scattering in the upper atmosphere; therefore, we assume that the surface reflectance due to sky illumination is shifted toward the blue. We assume that all pixels in the shadow region are illuminated by the sky only; we consider interreflections due to nearby objects as negligible. This is a fair assumption since according to the power law of reflectance, the energy of reflected light decays exponentially for each reflection.

3. Technical Approach

The following is a summary of general observations with respect to shadow and background:
(I) Shadow pixels fall on the same surface as the background, that is they have the same reflectance properties as the background pixels. (II) Shadow pixels are darker than their background in all three channels R, G, B. (III) Background surfaces are modeled. (IV) Background surfaces are generally matte and they are different than moving object surfaces. These observations are turned into assumptions that are used at different stages of our algorithm. We do not expect these assumptions to be violated in a wide variety of scenarios. Our approach is outlined in Figure 1. In the following we describe each of the components and its input and output.

• **Moving Object Detection** (Step 1): Given a video sequence, we utilize our previously developed detection system that uses a mixture model for background modeling to detect moving objects [5]. The output of this stage is a binary mask of the detected moving pixels obtained after noise clean up. Background models (μ_i,σ_i) for each pixel, and the current frame are also available at this stage.

• **Initial Shadow Pixel Reduction** (Step 2): Once a binary mask is available, we apply the second observation (II) that the shadow pixels are darker than their background in all three channels R, G, B. It excludes some object pixels thus reducing the binary mask and computation at later stages. During training only, the user selects areas of typical backgrounds when they are shadowed. This can be done either by selecting a frame when the object is in the scene and casts a shadow or selecting a frame at a time when the scene is naturally shadowed. Initial estimation of the diffuse background color is then obtained by using the algorithm in steps 5 and 6.

• **Blue Ratio Test** (Step 3): Shadow regions are illuminated by the sky. We assume the sky is blue and it is the only source of illumination on shadowed regions. Although all RGB values are lower in the shadow region, the amount by which this reduction occurs is not proportional. The blue ratio test hypothesizes shadow regions. It states that the reflectance change in shadowed regions is greater in the red and green channels than the blue and is calculated as $(I_{Fr}/I_{Br}, I_{Fg}/I_{Bg}) < I_{Fb}/I_{Bb}$ where the subscripts r, g, b indicate the channel and F and B indicate foreground (object) and background. Output of this stage is a further reduction in the number of pixels from previous test.

• **Albedo Ratio Segmentation** (Step 4): Assuming that our first observation (I) is valid, then a method based on the albedo-ratio of neighboring pixels can provide a uniformity test for surface segmentation. It is shown that the albedo-ratio test is invariant to viewing and illumination geometry and spectral power distribution of the illumination, as long as the spectral response of the camera and spectral distribution of illumination are constant [6]. This is true for shadow regions since shadow regions are illuminated by the blue sky and sky is also a constant and diffuse source of illumination. Briefly, the albedo-ratio is derived as follows. The intensity of a pixel is described as $I = \int s(\lambda)\, e(\lambda)\, r(\theta, \lambda) d\lambda$, where θ is a vector including surface normal, viewing direction and illumination direction, $s(\lambda)$ and $e(\lambda)$ are the spectral response of the camera and the spectral distribution of the illuminating source. Assuming that $s(\lambda) = s$ and $e(\lambda) = e$ are constant and integrating over the spectrum λ, $I = k \times \rho \times R(\theta)$ where $k = s \times e$, and $\rho \times R(\theta)$ is the integrated result where ρ is the reflection coefficient referred to as surface albedo. Assum-

ing that for two neighboring pixels the illumination direction and viewing direction are the same, and that two neighboring pixels represent surfaces in the scene with near identical normals, then $R(\theta)$ for both pixels is the same. We develop the spatio-temporal reflectance (albedo) ratio as follows. Let A_b and B_b be the background intensity of the two neighboring pixels (the values are obtained from the background image). Let A_f and B_f be the intensity values of the foreground corresponding to A and B respectively. Now define the following ratios,

$$Rt_A = \frac{A_f - A_b}{A_f + A_b} \quad Rt_B = \frac{B_f - B_b}{B_f + B_b} \quad P(A,B) = \frac{Rt_A - Rt_B}{Rt_A + Rt_B}$$

Since the first two ratios Rt_A and Rt_B are temporal ratios, P(A,B) is the spatio-temporal ratio of albedo. If two neighboring pixels belong to the same surface they will have temporal ratios that are close together; hence, the spatio-temporal ratio will be close to zero. P(A,B) is used as the connectivity criterion in a connected component labeling procedure and a threshold value of 0.05 is fixed. After the above segmentation, assuming that background is of fairly uniform material, we select large segments. Smaller segments (smaller than 20% of the largest segment) are discarded since they do not provide reliable input for diffuse color estimation.

- **Ambient Reflection Correction** (Step 5): We consider reflection due to sky, ambient reflection, as an additive component. If the segments obtained from the previous step are shadows, they represent reflection due to sky only, this reflection must be accounted before the "true" color estimation is carried out. We simply subtract the foreground pixel values from the background over the masked area.
- **Body Color Estimation** (Step 6): We utilize dichromatic reflection model [7] to estimate the body color of the surface, which is an inherent physical property of the material. According to this model, total radiance of the reflected light is the sum of diffuse (body) and specular (surface/interface) reflections. $L(\lambda,\theta) = m_i(\theta)\, c_i(\lambda) + m_b(\theta)\, c_b(\lambda)$, where m_i and m_b are magnitude of reflection also called scale factor due to interface and body geometry and c_i and c_b are dichromatic reflection axis indicating reflection due to interface and body. The color of a pixel C_L is described by tristimulus integration of L, $C_L = m_i \times C_i + m_b \times C_b$. (Note that both L (in step 6) and r (in step 4) are functions of λ and θ. L is used to distinguish the reflection according to dichromatic model. r is normally used to describe the BRDF.) It is not possible to measure L for every λ; therefore, only RGB channels are used. The dichromatic model can be extended to account for ambient illumination by an additive term C_a. Thus, $C_L = m_i \times C_i + m_b \times C_b + C_a$. Several solutions for estimating the body component (C_b) such as plane fitting, clustering and principal component analysis are available. We have utilized the singular value decomposition (SVD) approach. One advantage of SVD in this analysis is that the principal direction tends to naturally align in the direction of the body component. Furthermore, as the by-product of SVD, eigenvalues and eigenvectors provide error terms and possible existence of specularity. The SVD is applied to each segmented region that remains after steps 4 and 5. The size of the matrix supplied to SVD is n×3 where n is the number of pixels in each region.
- **Verification** (Step 7): Results of the previous step are in the form of diffuse (body) color vectors. These vectors are compared to initial estimations obtained in step 2. The regions corresponding to vectors within an acceptable threshold (1°) are classified as shadow regions.

4. Experiments

We have collected several videos showing a variety of background surfaces such as asphalt, concrete, and grass

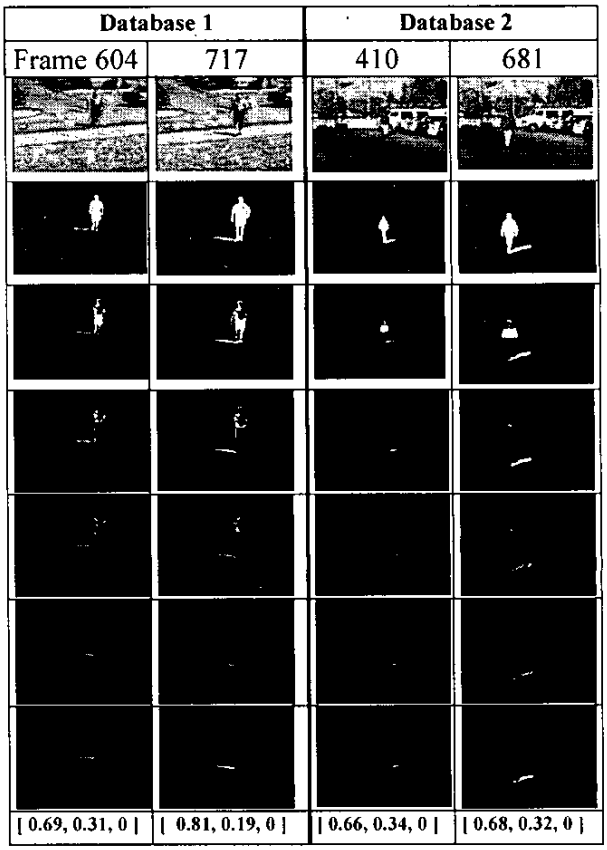

Figure 2. Row 1=Typical frames, rows 2 to 5 show results after steps 1 to 4, resp., row 6=after size filter in step 4, and row 7=detected shadows after step 7. Last row of numbers correspond to % of shadow pixels correctly classified, and missed, and object pixels classified as shadows.

found in typical urban areas. The data was collected on different days when the sky was either clear or had small patches of cloud. The algorithm given in Figure 1 is independent of the initial detection; hence, any of the four methods discussed in Section 1 could be applied here. After the training phase, once the initial color estimation has been performed, the algorithm is applied to the test frames.

Figure 2 shows selected random frames of moving persons from two video databases. Different background surface types and textured and uniform object surfaces with variety of colors are shown. Frame 604 indicates the shadow on asphalt with the subject wearing shorts and textured T-shirt. Frame 717 of the same database has the shadow on a concrete surface with the subject closer to the camera. Similarly, frames 401 and 681 belong to the second database where the subject is wearing a red shirt and white pants and the background surface is inclined. The subject is closer to camera in frame 681.

To evaluate the results, the ground truth is obtained by drawing a contour around the object and its shadow. The numbers at the bottom of the last row of images in Figure 2 indicate the classification percentages. Generally, as the subject approached the camera the algorithm performed better. This is due to several reasons. The camera signal is inversely proportional to the squared distance. Moreover, as shadows become larger, the body (diffuse) color estimation provides more robust results. Since each stage of the algorithm assures inclusion of potential object pixels as well as shadow pixels, we expect to see high correct object classification (low object misclassification) rates. Our results indicate that the percentage of object pixels misclassified as shadows is very low, typically below 0.01%.

Figure 3. Shadow of Object (pants) has similar colors as background.

Figure 3 shows a frame and the shadow extracted by the algorithm. It indicates a difficult case for our approach. In this scene, the person's pants has a neutral gray color, the same as the background, and are self shadowed. As a result, part of the pants is labeled as shadow. The self-shadowed region (pants), however, has higher luminance than the real shadow on the ground; hence, a statistical luminance test based on the histogram can be used to further classify these two regions.

Since our algorithm is modular, any improvements can be easily incorporated in the system. Two parameters affect the performance of our algorithm. In step 3, the blue ratio test is sensitive to the sensor and the background color saturation. This step can be bypassed if the background color is highly saturated or sky is cloudy. In step 5, by choosing large segments we may be eliminating smaller segments that are actually the shadow pixels. This can be overcome if the reflected color is estimated for each pixel rather than patches of pixels. This can be done by periodically imaging the scene as illumination direction of the sun changes at the cost of increased computation. Note also that we do not consider puddles; the dichromatic reflection model is not suitable for puddles since it does not account for transmittance and refractance due to water. Highly reflective surfaces such as puddles need to be handled by a more suitable physical model.

5. Conclusions

We presented an algorithm that is effective in distinguishing moving objects from their shadows cast on a background. The two main features of our physics based algorithm are segmentation based on spatio-temporal albedo test and utilization of a dichromatic reflection model. In our future work we plan to investigate robust methods for estimating color at the pixel level. Our physics based method can also be extended to infrared sensors where shadow regions can be modeled based on the thermal properties.

Acknowledgement – This work was supported in part by grants F49620-97-1-0184 and DAAD19-01-0357; the contents and information do not necessarily reflect the position or policy of U.S. Government.

References

[1] S. Das and B. Bhanu, "A system for model-based object recognition in perspective aerial images," *Pattern Recognition*, 3(34), pp. 465-491, 1998.

[2] Y. Sonada and T. Ogata, "Separation of moving objects and their shadows, and application to tracking of loci in the monitoring images," *Proc. Intl. Conf. on Signal Procesing*, 2(2), pp. 1261-4, 1998.

[3] I. Mikic,et al., "Moving shadow and object detection in traffic scenes," *Proc. Intl. Conf. on Pattern Recognition*, Vol. 1, pp. 321-324, 1998.

[4] K. Onoguchi, "Shadow elimination method for moving object detection," *Proc. Intl. Conf. on Pattern Recognition*, Vol. 1, pp. 583-587, 1998.

[5] S. Nadimi and B. Bhanu, "Multistrategy fusion using mixture model for moving object detection," *Proc. Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, pp. 317-322, Aug 2001.

[6] S. K. Nayar and R.M. Bolle, "Reflectance based object recognition," *Intl. J. of Comp. Vision*, 17(3), pp. 219-240, 1996.

[7] S. A. Shafer, "Using color to separate reflection components," *COLOR Research and Application*, 10(4), pp. 210-218, 1985.