Title
Moving shadow and object detection in traffic scenes

Journal
Pattern Recognition, 2000. Proceedings. 15th International Conference on, 1

Authors
Mikic, I
Cosman, P C
Kogut, G T
et al.

Publication Date
2000

Peer reviewed

Moving Shadow and Object Detection in Traffic Scenes

Ivana Mikić, Pamela C. Cosman, Greg T. Kogut, Mohan M. Trivedi
Department of Electrical and Computer Engineering, University of California, San Diego, USA
e-mail: ivana, pcosman, gkogut, trivedi @ece.ucsd.edu

Abstract

We present an algorithm for segmentation of traffic scenes that distinguishes moving objects from their moving cast shadows. A fading memory estimator calculates mean and variance of all three color components for each background pixel. Given the statistics for a background pixel, simple rules for calculating its statistics when covered by a shadow are used. Then, MAP classification decisions are made for each pixel. In addition to the color features, we examine the use of neighborhood information to produce smoother classification. We also propose the use of temporal information by modifying class a priori probabilities based on predictions from the previous frame.

1. Introduction

This work is motivated by the need for a robust segmentation algorithm to be used in a traffic monitoring and incident detection system [1]. Of course, extracting positions of moving objects in image sequences is an important component for many other applications. Background subtraction is a common approach to this problem. The background model is built from the data and objects are segmented if they appear significantly different from the background. Unfortunately, moving shadows are usually extracted along with the objects. This can result in large errors in object localization and can cause serious problems for algorithms that use segmentation results as their basic measurements.

Many algorithms that detect shadows take into account the location of the light source, geometry of the scene and models of moving objects [2]. Our aim was to avoid using any such knowledge in detecting shadows. One such algorithm, proposed by Strauder et al. [3], instead assumes that static edges caused by background texture remain in regions covered by shadows and that shadows have penumbra, a soft luminance transition at the contour of the shadow. However, this is rarely true for outdoor scenes, where shadows usually have sharp edges and background is often non-textured. Even with textured background, our experience is that texture is almost invisible in the shadow regions due to the properties of the imaging process.

Without using scene models and assumptions mentioned above, we can identify three sources of information that can help in detecting objects and shadows. The first is local, based on the appearance of the individual pixels. A point covered by a shadow gets darker compared to its appearance when illuminated. The second source of information is spatial: objects and shadows inhabit compact regions in the image, and the third is temporal: object and shadow positions can be predicted from previous frames.

We propose an algorithm that uses all three sources of information to classify pixels into the shadow, object and background classes. In Section 2 we present the segmentation algorithm based on pixel appearance, in Section 3 we examine the use of spatial information, and in Section 4 the results are presented. We propose a way of incorporating temporal information and other directions of future work in Section 5.

2. Pixel appearance based segmentation

If we assume that a pixel in a given frame belongs to either the background, shadow or vehicle, we need the estimates of probability density functions of the three classes (i.e. the parameters of the three-component mixture pdf) for reliable classification.

Friedman and Russell [4] used the incremental version of the EM algorithm to estimate the parameters of a three-component mixture pdf for each pixel. We implemented this algorithm to classify pixels using a three component feature vector (the three color components) and found that parts of cars that appear darker than the background get classified as shadows. Since the EM algorithm automatically "groups" measurements, those pixels that belong to dark cars contribute to the component of the mixture that represents shadows, resulting in inaccurate parameters of the mixture.

0-7695-0750-6/00 $10.00 © 2000 IEEE

We believe that a more accurate estimate of the pdf of a shadowed pixel can be computed using a model of the change in appearance of a pixel when shadowed given its appearance when illuminated. Using this approach, the pdf of an illuminated background pixel is estimated from the data, and the parameters of the pdf of the same pixel when shadowed are derived from it using the model of appearance change. In the following section, we describe the estimation of such a model from the labeled example sequence.

2.1. Color change under shadow

The effects of surface reflectance and properties of the illuminant on the appearance of a surface in an image have been successfully modeled [5]. Since a shadowed pixel represents the same surface under different illumination, we are interested in the effects of illumination on pixel appearance. We have found the approximation of this effect by a diagonal matrix to be satisfactory. See [6, 7] for details. In other words, if $\mathbf{v} = [R\ G\ B]^T$ is the camera response for a point on a surface when illuminated, then $\mathbf{Dv}$ is the camera response for the same point when shadowed, where $\mathbf{D}$ is a diagonal matrix.

Figure 1 shows the appearance change for manually segmented background and shadow pixels in one traffic video sequence. The slopes of lines fitted to plots for three color components shown in the figure determine the values of corresponding coefficients in the matrix $\mathbf{D}$. In this example, $\mathbf{D} = diag\ (0.48, 0.47, 0.51)$. The parameter that corresponds to the blue color component is the largest, which agrees with our observation that shadowed surfaces appear bluer in traffic video scenes. $\mathbf{D}$ is approximately constant over flat surfaces. If the background is not flat over the entire image, we can divide the image into subregions where this assumption is more likely to hold and model each subregion separately.

With this model of appearance change under shadow, we easily derive the rules for estimating means and variances for the three color components under shadow ($\mu_{SH}^R, \mu_{SH}^G, \mu_{SH}^B, \sigma_{SH}^R, \sigma_{SH}^G, \sigma_{SH}^B$), given those parameters for the same pixel when illuminated ($\mu_{IL}^R, \mu_{IL}^G, \mu_{IL}^B, \sigma_{IL}^R, \sigma_{IL}^G, \sigma_{IL}^B$). If $\mathbf{D} = diag(d_R, d_G, d_B)$, we have:

$$\begin{aligned} \mu_{SH}^i &= \mu_{IL}^i d_i \\ \sigma_{SH}^i &= \sigma_{IL}^i d_i,\ i \in \{R,G,B\} \end{aligned} \tag{1}$$

2.2. Pixel classification

We call the three classes background, shadow and foreground. It should be noted, however, that shadow pixels also belong to static surfaces in the scene but are at the current frame shadowed by a moving object. Gaussian distributions are assumed for illuminated and shadowed states of a pixel, since the only reason their appearance is not constant is noise, which we assume is Gaussian. Uniform distribution is assumed for a pixel covered by a moving vehicle since there is no particular reason to prefer one color over any other.

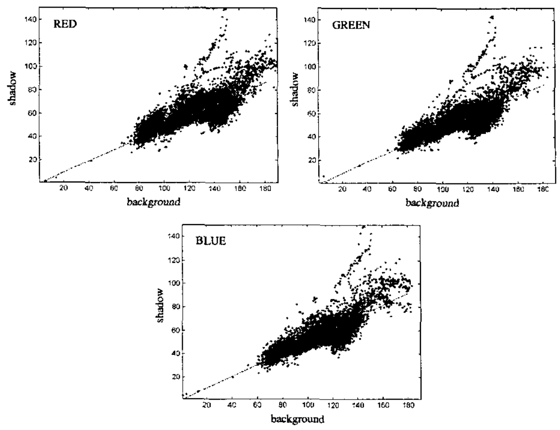


Figure 1. Plots of background vs. shadowed color components for a set of example data and the linear approximation to the data. The slope of the line determines the corresponding element of matrix $\mathbf{D}$, red: d_R=0.48, green: d_G=0.47, blue d_B=0.51

The feature vector for each pixel contains the three color components. A fading memory estimator [8] calculates background means and variances for all pixel locations. Using the rules presented in the previous section, we derive statistics for same pixels when shadowed. We start the segmentation by comparing the feature vector for each pixel to the mean at that location in the background model. If not significantly different (difference less than 10% of the mean), the pixel is classified into the background class. Otherwise, we assign to that location the a priori probabilities p_{BG}, p_{SH}, and p_{FG} of belonging to background, shadow and foreground classes, respectively. We use p_{BG} = 0.3, p_{SH} = 0.4, and p_{FG} = 0.4 Then, we classify each pixel by maximizing the a posteriori probability of the class membership (C_1 = background, C_2 = shadow and C_3 = foreground):

$$p(C_i/\mathbf{v}) = \frac{p(\mathbf{v}/C_i)p(C_i)}{\sum_{j=1,2,3} p(\mathbf{v}/C_j)p(C_j)} \tag{2}$$

where $\mathbf{v}$ is the feature vector for a given pixel, $p(C_i)$ the a priori probability of occurrence of the i-th class at that location and $p(\mathbf{v}/C_i)$ the probability of the observed feature values given that the pixel belongs to the i-th class.

3 Imposing spatial constraints

The majority of the pixels are classified correctly by the described appearance-based algorithm (73%, when compared to the hand-segmented images). However, object and shadow regions are very noisy due to misclassified pixels (See Figure 3a). The results can be significantly improved by imposing spatial smoothness. We investigated two approaches. First is simple post-processing by spatial filtering of the segmented images. We eliminate small gaps in foreground regions by performing one vertical and then one horizontal scan and assigning an encountered small line segment of non-foreground pixels to foreground if it is surrounded by foreground pixels in the direction of the scan. This is followed by morphological opening.

The second approach we investigated was performing an iterative probabilistic relaxation to propagate neighborhood information. In the first step, the a posteriori probability computations based on color are performed for all pixels. This is a local, appearance based computation. In the second step, we perform spatial propagation where the new class membership probabilities are computed for each pixel based on the results of the first step on the neighboring pixels. These are then used for a new computation of a posteriori probabilities in the first step and so on (Figure 2). The scheme converges quickly, and there is no noticeable change beyond the second iteration.

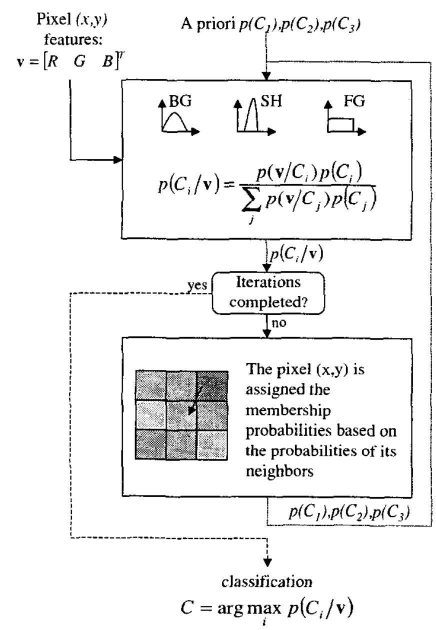


Figure 2. Iterative procedure that integrates appearance based and spatial information

We have found that the probabilities have to be dramatically changed in the spatial propagation step to change the final classification of a given pixel. We got the best results by assigning to the central pixel in a window the probabilities associated with its neighbor with the highest membership probability for the class that dominates the neighborhood. We found that the results are slightly improved (78% of pixels correctly classified – see Figure 3b). However, there is still a need for post-processing that is of similar complexity to the post-processing described in the previous paragraph, which we used on original segmentation results. The final result is very similar (around 90% of pixels classified correctly – see Figure 3c). Also, performing these iterations reduces the speed and increases the memory requirements. We therefore conclude that the spatial smoothness is imposed most efficiently by a simple post-processing.

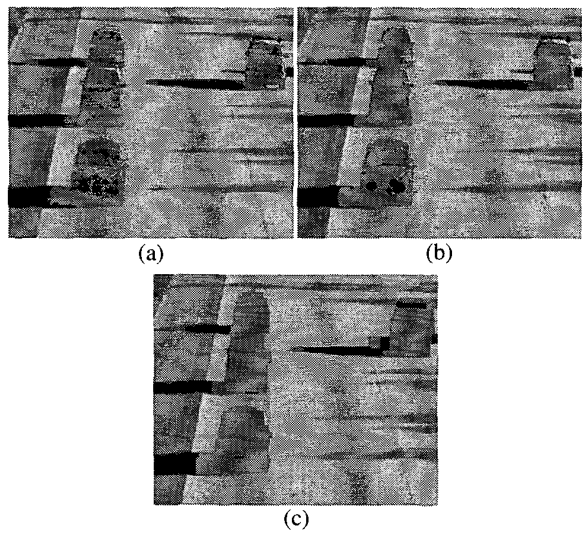
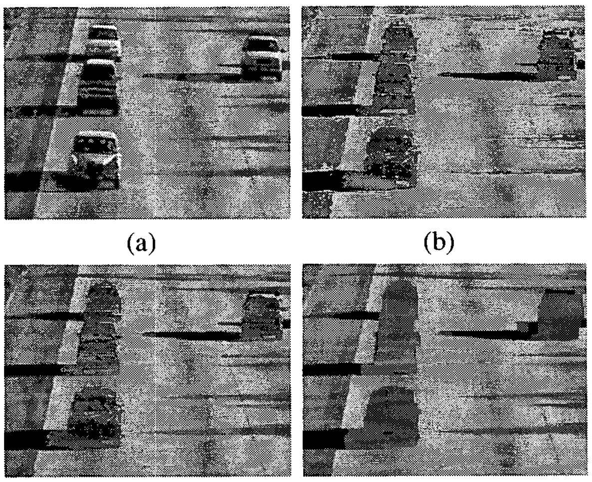


Figure 3. Imposing spatial smoothness. (a) result of the color based segmentation. (b) result of adding a smoothing component to the iteration loop. (c) Result of post-processing of (a).

4 Results

Figure 4 shows segmentation results for one frame from the video of a traffic scene. The pixels colored with red, blue and green are those that differed more than 10% from the means of the background pdf's. By correctly classifying shadows and flickering background pixels that simple background subtraction would classify as foreground, the accuracy of the calculated object locations is greatly improved, especially in scenes with long shadows. Note that static shadows are considered to be part of the background. Segmented shadows also provide an important clue for separating objects that are so close that they are segmented as one object. Often in those cases, the shadows of such objects will be distinct and help us separate the objects (see Figure 4d). Figure 5 shows results on several video frames.



(a) (b) (c) (d)

Figure 4. Moving shadow and object detection. (a) the original image frame. (b) Classification results. Red pixels are classified as foreground, blue as shadow and green as background. (c) Same as in (b), with background pixels not shown. (d) final result after post-processing by a spatial filter

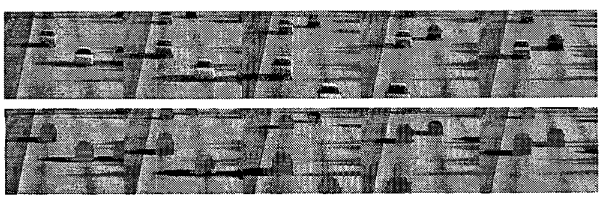


Figure 5. Five frames from the video of a traffic scene. Top row shows the raw video data and the bottom row shows the results of the algorithm

5 Conclusions and future work

We have presented a real time algorithm for segmentation of moving objects and cast shadows in image sequences. Our final aim is to have a deployable algorithm able to operate over extended periods of time and provide robust measurements for a traffic monitoring system.

We propose several improvements and areas of further study. First, including temporal information could significantly improve the performance of the algorithm without much speed degradation. We could use predicted object locations to select a priori probabilities in the current frame. Locations where we expect objects of one class would be assigned high corresponding a priori probabilities.

Another important direction of future work is analysis of the relationship between the scene illumination and the matrix **D**. As the algorithm adapts background statistics to the slow changes in the scene conditions, it could also collect statistics for shadow pixels identified with high confidence and modify matrix **D** accordingly. By measuring illumination of the scene, we should be able to build a lookup table for the components of **D** indexed by the scene illumination and use it to recover from sudden changes in scene conditions.

Acknowledgements

This work was supported in part by the California Digital Media Innovation Initiative and by an NSF Career Award MIP-9624729

References

[1] M. Trivedi , S. Bhonsle, A. Gupta "Database Architecture for Autonomous Transportation Agents for On-scene Networked Incident Management (ATON)", *ICPR 2000*

[2] D. Koller, K. Danilidis, H. H. Nagel, "Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes", *Int. Journal of Computer Vision*, 10:3, 257-281, 1993

[3] J. Stauder, R. Mech, J. Ostermann, "Detection of Moving Cast Shadows for Object Segmentation", *IEEE Trans. Multimedia*, 1:1, 65-76, 1999

[4] N. Friedman, S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach", *Proc. Thirteenth Conf. on Uncertainty in Artificial Intelligence, 1997*

[5] B. A. Wandell, "The Synthesis and Analysis of Color Images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:1, 2-13, 1987

[6] G. D. Finlayson, M. S. Drew, B. V. Funt, "Diagonal Transforms Suffice for Color Constancy", *IEEE Int. Conf. on Computer Vision*, 164-171, 1993

[7] K. Barnard, G. Finlayson, B. Funt, "Color Constancy for Scenes With Varying Illumination", *Computer Vision and Image Understanding*, 65:2, 311-321, 1997

[8] E. Sudderth, E. Hunter, K. Kreutz-Delgado, P. Kelly, R. Jain, "Adaptive Video Segmentation: Theory and Real-Time Implementation", *Proc. DARPA Image Understanding Workshop*, 177-181, 1998