Background Subtraction and Shadow Detection in Grayscale Video Sequences

Julio Cezar Silveira Jacques Jr Cláudio Rosito Jung
Soraia Raupp Musse

CROMOS Laboratory - PIPCA
University of Vale do Rio dos Sinos
Av. Unisinos 950, 93022-000 São Leopoldo, RS, Brazil
phone: +55 51 5908161 and fax: +55 51 5908162
julioj@turing.unisinos.br, crjung@unisinos.br, soraiarm@exatas.unisinos.br

Abstract

Tracking moving objects in video sequence is an important problem in computer vision, with applications several fields, such as video surveillance and target tracking. Most techniques reported in the literature use background subtraction techniques to obtain foreground objects, and apply shadow detection algorithms exploring spectral information of the images to retrieve only valid moving objects. In this paper, we propose a small improvement to an existing background model, and incorporate a novel technique for shadow detection in grayscale video sequences. The proposed algorithm works well for both indoor and outdoor sequences, and does not require the use of color cameras.

1 Introduction

A relevant problem in computer vision is the detection and tracking of moving objects in video sequences. Possible applications include surveillance [6, 7, 12], traffic monitoring [8] and athletic performance analysis [1], among others.

In applications using fixed cameras with respect to the static background (e.g. stationary surveillance cameras), a very common approach is to use background subtraction to obtain an initial estimate of moving objects. Basically, background subtraction consists of comparing each new frame with a representation of the scene background: significative differences usually correspond to foreground objects. Ideally, background subtraction should detect real moving objects with high accuracy, limiting false negatives (objects pixels that are not detected) as much as possible; at the same time, it should extract pixels of moving objects with the maximum responsiveness possible, avoiding detection of transient spurious objects, such as cast shadows, static objects, or noise.

In particular, the detection of cast shadows as foreground objects is very common, producing undesirable consequences. For example, shadows can connect different people walking in a group, generating a single object (typically called *blob*) as output of background subtraction. In such cases, it is more difficult to isolate and track each person in the group.

There are several techniques for shadow detection in video sequences [2–4, 10, 13, 15–17], and the vast majority of them are based on color video sequences. Although color images indeed provide more information for shadow detection, there are still several scenarios where monochromatic video cameras are utilized In this paper, we improve the background subtraction technique described in [6], and propose a new shadow detection algorithm for grayscale images.

The remainder of this paper is organized as follows. Section 2 presents related work concerning background subtraction and shadow detection. The proposed technique is described in Section 3, and some experimental results are provided in Section 4. Finally, conclusions are given in Section 5.

2 Related Work

Several techniques for background subtraction and shadow detection have been proposed in the past years. Background detection techniques may use grayscale or color images, while most shadow detection methods make use of chromaticity information. Next, some of these techniques are described.

The car tracking system of Koller et al. [9] used an adaptive background model based on monochromatic images filtered with Gaussian and Gaussian derivative (vertical and

horizontal) kernels. McKenna et al. [11] proposed a background model that combines pixel *RGB* and chromaticity values with local image gradients. In their W4 system, Haritaoglu and collaborators [6] used grayscale images to build a background model, representing each pixel by three values; its minimum intensity value, its maximum intensity value and the maximum intensity difference between consecutive frames observed during the training period. Elgammal et al. [4] used a nonparametric background model based on kernel based estimators, that can be applied to both color or grayscale images. KaewTrakulPong and Bowden [7] used color images for background representation. In their method, each pixel in the scene is modelled by a mixture of Gaussian distributions (and different Gaussians are assumed to represent different colors). Cucchiara's group [3] used a temporal median filtering in the *RGB* color space to produce a background model.

Shadow detection algorithms have also been widely explored by several authors, mostly based on invariant color features, that are not significatively affected by illumination conditions. McKenna et al. [11] used pixel and edge information of each channel of the normalized *RGB* color space (or *rgb*) to detect shadowed pixels. Elgammal et al. [4] also used the normalized *rgb* color space, but included a lightness measure to detect cast shadows. Prati's and Cucchiara's groups [3,13] used the *HSV* color space, classifying as shadows those pixels having the approximately the same hue and saturation values compared to the background, but lower luminosity. KaewTrakulPong and Bowden [7] used a chromatic distortion measure and a brightness threshold in the *RGB* space to determine foreground pixels affected by shadows. Salvador et al. [15] adopted the $c_1c_2c_3$ photometric invariant color model, and explored geometric features of shadows. A few authors [2, 14, 17, 20] have studied shadow detection in monochromatic video sequences, having in mind applications such as indoor video surveillance and conferencing. Basically, they detect the penumbra of the shadow, assuming that edge intensity within the penumbra is much smaller that edge intensity of actual moving objects. Clearly, such hypothesis does not hold for video sequences containing low-contrast foreground objects (specially in outdoors applications). More about background subtraction and shadow removal can be found in [18, 19].

A revision of the literature indicates that several background models are available, applicable for color and/or grayscale video sequences. Also, there are several shadow detection algorithms to remove undesired segmentation of cast shadows in video sequences. However, in accordance with other authors [3, 6], we chose to use a background model based on median filtering, because it is effective and requires less computational cost than the Gaussian or other complex statistics. More specifically, we improved the background model proposed in [6], and included a novel shadow detection algorithm that is effective for both indoor and outdoor applications.

3 The proposed algorithm

In this Section, we describe the background model of W4 [6] and propose a small improvement to the model. We also propose a novel method for shadow segmentation of foreground pixels, based on normalized cross-correlations and pixel ratios.

3.1 Background Scene Modeling

W4 uses a model of background variation that is a bimodal distribution constructed from order statistics of background values during a training period, obtaining robust background model even if there are moving foreground objects in the field of view, such as walking people, moving cars, etc. It uses a two stage method based on excluding moving pixels from background model computation. In the first stage, a pixel wise median filter over time is applied to several seconds of video (typically 20-40 seconds) to distinguish moving pixels from stationary pixels (however, our experiments showed that 100 frames $\approx$ 3.3 seconds are typically enough for the training period, if not too many moving objects are present). In the second stage, only those stationary pixels are processed to construct the initial background model. Let V be an array containing N consecutive images, $V^k(i,j)$ be the intensity of a pixel (i,j) in the *k-th* image of V, $\sigma(i,j)$ and $\lambda(i,j)$ be the standard deviation and median value of intensities at pixel (i,j) in all images in V, respectively. The initial background model for a pixel (i,j) is formed by a three-dimensional vector: the minimum $m(i,j)$ and maximum $n(i,j)$ intensity values and the maximum intensity difference $d(i,j)$ between consecutive frames observed during this training period. The background model $\mathbf{B}(i,j) = [m(i,j), n(i,j), d(i,j)]$, is obtained as follows:

$$\begin{bmatrix} m(i,j) \\ n(i,j) \\ d(i,j) \end{bmatrix} = \begin{bmatrix} \min\limits_z V^z(i,j) \\ \max\limits_z V^z(i,j) \\ \max\limits_z |V^z(i,j) - V^{z-1}(i,j)| \end{bmatrix}, \quad (1)$$

where z are frames satisfying $|V^z(i,j) - \lambda(i,j)| \leq 2\sigma(i,j)$. According to [6] This condition guarantees that only stationary pixels are computed in the background model, i.e., $V^z(i,j)$ is classified as a stationary pixel.

After the training period, an initial background model $\mathbf{B}(i,j)$ is obtained. Then, each input image $I^t(i,j)$ of the video sequence is compared to $\mathbf{B}(i,j)$, and a pixel (i,j) is classified as a background pixel if:

$$I^t(i,j) - m(i,j) \leq k\mu \quad \text{or} \quad I^t(i,j) - n(i,j) \leq k\mu, \quad (2)$$

where μ is the median of the largest interframe absolute difference image $d(i,j)$, and k is a fixed parameter (the authors suggested the value $k=2$). It can be noted that, if a certain pixel (i,j) has an intensity $m(i,j) < I^t(i,j) < n(i,j)$ at a certain frame t, it should be classified as background (because it lies between the minimum and maximum values of the background model). However, Equation (2) may wrongly classify such pixel as foreground, depending on k, μ, $m(i,j)$ and $n(i,j)$. For example, if $\mu = 5$, $k = 2$, $m(i,j) = 40$, $n(i,j) = 65$ and $I^t(i,j) = 52$, Equation (2) would classify $I^t(i,j)$ as foreground, even though it lies between $m(i,j)$ and $n(i,j)$. To solve this problem, we propose an alternative test for foreground detection, and classify $I^t(i,j)$ as a foreground pixel if:

$$I^t(i,j) > (m(i,j) - k\mu) \quad \text{and} \quad I^t(i,j) < (n(i,j) + k\mu) \quad (3)$$

Figure 1 illustrates an example of background subtraction (using $k = 2$, as in all other examples in this paper). The background image (median of frames across time) is shown in Figure 1(a), a certain frame of the video sequence is shown in Figure 1(b), and detected foreground objects are shown in Figure 1(c). It can be noticed that two kinds of shadows were detected: on the left, shadow was caused by obstruction of indirect light; on the right, shadow was produced by direct sunlight blocking.

3.2 Shadow identification

In shadowed regions, it is expected that a certain fraction α of incoming light is blocked [4]. Although there are several factors that may influence the intensity of a pixel in shadow [15], we assume that the observed intensity of shadow pixels is directly proportional to incident light; consequently, shadowed pixels are scaled versions (darker) of corresponding pixels in the background model.

As noticed by other authors [5], the normalized cross-correlation (NCC) can be useful to detect shadow pixel candidates, since it can identify scaled versions of the same signal. In this work, we use the NCC as an initial step for shadow detection, and refine the process using local statistics of pixel ratios, as explained next.

3.2.1 Detection of shadow pixel candidates

Let $B(i,j)$ be the background image formed by temporal median filtering, and $I(i,j)$ be an image of the video sequence. For each pixel (i,j) belonging to the foreground, consider a $(2N+1) \times (2N+1)$ template T_{ij} such that $T_{ij}(n,m) = I(i+n, j+m)$, for $-N \le n \le N$, $-N \le m \le N$ (i.e. T_{ij} corresponds to a neighborhood of pixel (i,j)). Then, the NCC between template T_{ij} and image B at pixel (i,j) is given by:

$$NCC(i,j) = \frac{ER(i,j)}{E_B(i,j) E_{T_{ij}}}, \quad (4)$$

where

$$\begin{aligned} ER(i,j) &= \sum_{n=-N}^{N} \sum_{m=-N}^{N} B(i+n, j+m) T_{ij}(n,m), \\ E_B(i,j) &= \sqrt{\sum_{n=-N}^{N} \sum_{m=-N}^{N} B(i+n, j+m)^2}, \text{ and} \quad (5) \\ E_{T_{ij}} &= \sqrt{\sum_{n=-N}^{N} \sum_{m=-N}^{N} T_{ij}(n,m)^2}. \end{aligned}$$

For a pixel (i,j) in a shadowed region, the NCC in a neighboring region T_{ij} should be large (close to one), and the energy $E_{T_{ij}}$ of this region should be lower than the energy $E_B(i,j)$ of the corresponding region in the background image. Thus, a pixel (i,j) is pre-classified as shadow if:

$$NCC(i,j) \ge L_{\text{ncc}} \text{ and } E_{T_{ij}} < E_B(i,j), \quad (6)$$

where L_{ncc} is a fixed threshold. If L_{ncc} is low, several foreground pixels corresponding to moving objects may be misclassified as shadows. On the other hand, selecting a larger value for L_{ncc} results in less false positives, but pixels related to actual shadows may not be detected. In fact, the influence of the threshold L_{ncc} for shadow detection can be observed in Figure 2. This Figure illustrates the application of our shadow detector in the foreground image of Figure 1(c) using $N = 4$, for different thresholds L_{ncc}. Black pixels are foreground pixels, and gray pixels correspond to shadowed pixels according to Equation (6). Our experiments indicated that choosing $L_{\text{ncc}} = 0.95$ results in a good compromise between false positives and false negatives, and that $N = 4$ is a good neighborhood size.

3.2.2 Shadow refinement

The NCC provides a good initial estimate about the location of shadowed pixels, by detecting pixels for which the surrounding neighborhood is approximately scaled with respect to the reference background. However, some background pixels related to valid moving objects may be wrongly classified as shadow pixels. To remove such false positives, a refinement stage is applied to all pixels that satisfy Equation (6).

The proposed refinement stage consists of verifying if the ratio $I(i,j)/B(i,j)$ in a neighborhood around each shadow pixel candidate is approximately constant, by computing the standard deviation of $I(i,j)/B(i,j)$ within this neighborhood. More specifically, we consider a region R

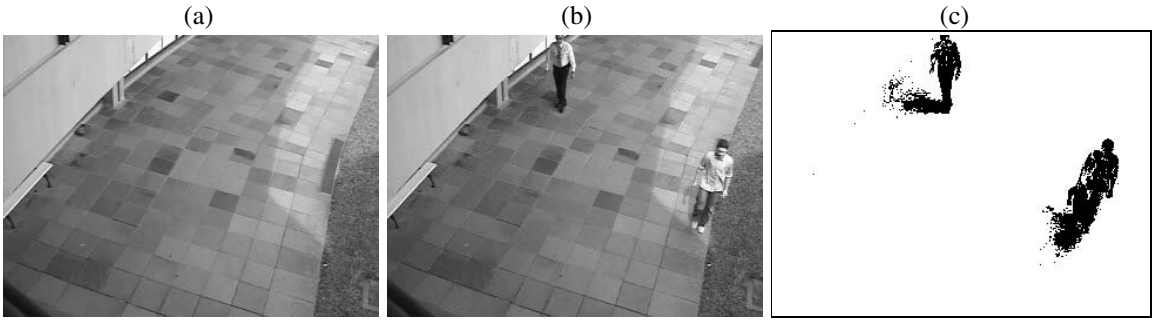

Figure 1. (a) Background image. (b) A certain frame of the video sequence. (c) Detected foreground objects.

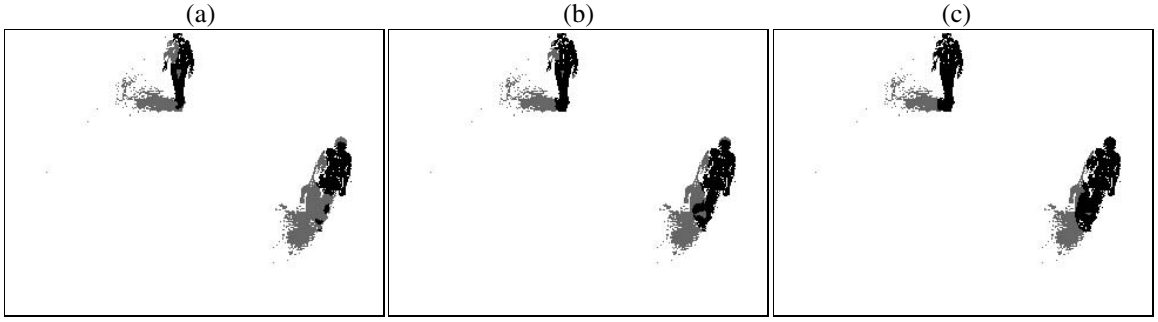

Figure 2. Shadow detection using different thresholds L_{ncc}. (a) $L_{\text{ncc}} = 0.90$ (b) $L_{\text{ncc}} = 0.95$ (c) $L_{\text{ncc}} = 0.98$

with $(2M+1) \times (2M+1)$ pixels (we used $M = 1$ in all experiments) centered at each shadow pixel candidate (i, j), and classify it as a shadow pixel if:

$$\text{std}_R\left(\frac{I(i,j)}{B(i,j)}\right) < L_{\text{std}} \text{ and } L_{\text{low}} \leq \left(\frac{I(i,j)}{B(i,j)}\right) < 1, \quad (7)$$

where $\text{std}_R\left(\frac{I(i,j)}{B(i,j)}\right)$ is the standard deviation of quantities $I(i,j)/B(i,j)$ over the region R, and $L_{\text{std}}, L_{\text{low}}$ are thresholds. More precisely, L_{std} controls the maximum deviation within the neighborhood being analyzed, and L_{low} prevents the misclassification of dark objects with very low pixel intensities as shadowed pixels. To determine values for L_{std} and L_{low}, we conducted the following experiment. We printed a chart with several graytones and analyzed its pixel values under direct sunlight, building a background model. We evaluated these pixels across time, when a moving cloud caused progressive light occlusion, and computed values $\text{std}_R\left(\frac{I(i,j)}{B(i,j)}\right)$. Experimentally obtained values were $L_{\text{std}} = 0.05$ and $L_{\text{low}} = 0.5$ (however, we believe that further studies on the selection of L_{std} and L_{low} are needed). It should be noticed that in sunny days shadows may be very strong, and information about pixel intensity in the umbra may be completely lost. Is such cases, $I(i,j)/B(i,j)$ is usually very small, and shadows may be misclassified as valid foreground objects. Also, we apply morphological operators to foreground pixels after shadow removal, to complete empty spaces and remove isolated pixels. We apply sequentially a closing and an opening operator with a 5×5 diamond-shaped structuring element.

Stauder and colleagues [17] also used the local variance for shadow detection. However, they did not compare each frame of the video sequence with a background model; in their approach, pixel ratios were computed for consecutive frames, which may cause erroneous detection in rotating objects.

An example of the shadow refinement technique applied to the initial shadow detection of Figure 2 is depicted in Figure 3(a). In this Figure, darker gray pixels correspond to the initial shadow detection, and lighter gray pixels correspond to the final shadow detection. Figure 3(b) shows all foreground pixels after shadow removal, and Figure 3(c) shows

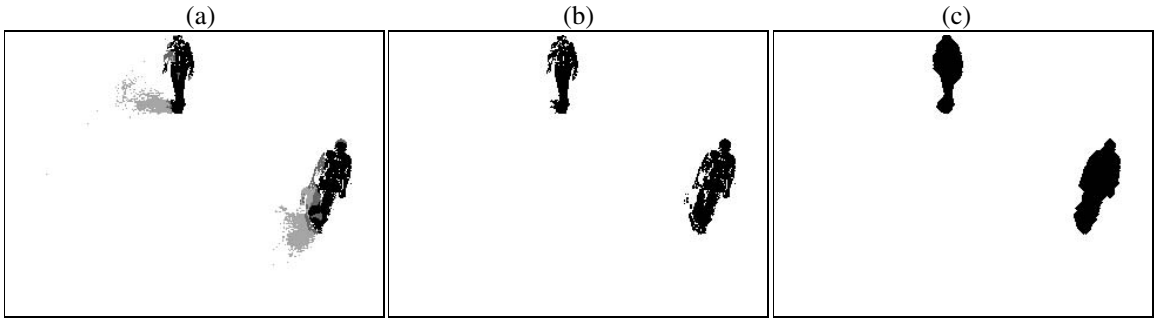

Figure 3. (a) Final shadow detection (shadow pixels are represented by light gray). (b) Foreground objects after shadow removal. (c) Elimination of gaps and isolated pixels through morphological operators .

the final result after applying morphological operators.

4 Experimental Results

In this Section, we analyze the performance of our background subtraction and shadow detection algorithms for indoor and outdoor grayscale video sequences. For example, Figure 4 shows four frames of an outdoor video sequence, in which some parts of the image are illuminated by direct sunlight, while in other parts there is only indirect light incidence. As a consequence, different kinds of shadows are produced (weak and strong shadows). In the first row, original grayscale images are displayed; the second and third rows show, respectively, foreground pixels before and after shadow removal; foreground objects after morphological operators are shown in the fourth row. It can be noticed that shadows were effectively detected and removed in both weak and strong shadow regions.

Another example in an outdoor environment is shown in Figure 5, corresponding to a video sequence acquired in a cloudy day, producing weak shadows. Figure 5(a) shows a certain frame, and Figure 5(b) shows detected foreground objects. It can be observed that a large foreground blob was produced, and it is very difficult to identify the person on the lower part of the image. Figures 5(c) and 5(d) illustrate the result of our shadow removal technique, before and after applying morphological operators. In Figure 5(d), all three persons are completely identifiable.

Figure 6 shows the performance of the proposed technique for the Hall video sequence[1]. Although this was originally a color video sequence, it was transformed to grayscale using MATLAB's command `rgb2gray` at each frame. Shadows were also correctly detected and removed in this indoor footage, and valid foreground moving objects were correctly segmented.

One drawback of the proposed technique is the misclassified of valid foreground objects as shadows in video sequences containing a homogeneous background with homogeneous (and darker) foreground objects. Such problem may happen because the NCC can be very high within such objects, and the standard deviation low. An example of misclassification is shown in Figure 7, that illustrates a person with a homogeneous shirt in front of a homogeneous white wall (the background). Cast shadows were correctly identified around the person, but the shirt and some parts of the skin were misclassified as shadows. Fortunately, the background presents some texture in most applications, and shadow misclassification is not common.

5 Conclusions

In this work, we improved an existing method for background subtraction and proposed a novel technique for shadow detection in grayscale video sequences. In our approach, the normalized cross-correlation is applied to foreground pixels, and candidate shadow pixels are obtained. A refinement process is then applied to further improve shadow segmentation.

Experimental results showed that the proposed technique performs well in video sequences containing strong shadows (occlusion of direct sunlight) and weak shadows (occlusion of indirect light), being suited for both indoor and outdoor applications. Other shadow detection techniques based on grayscale images [2, 17, 20] assume smooth gradient variation in shadowed regions, and are more appropriate to indoor applications only. With the proposed technique, persons walking close to each other connected by shadows can be successfully tracked individually.

[1] available for download at `http://www.ics.forth.gr/cvrl/demos/NEMESIS/hall_monitor.mpg`

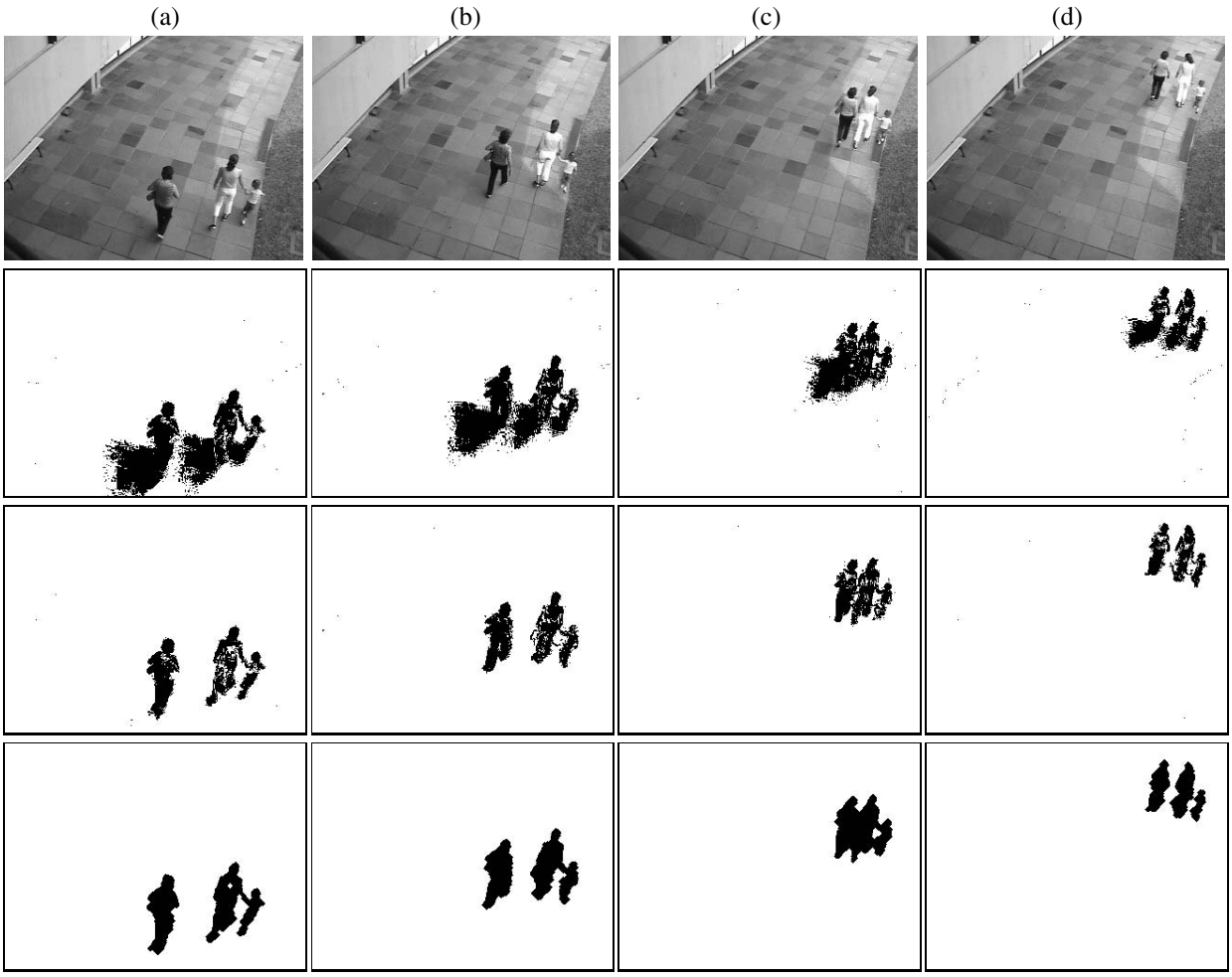

Figure 4. Top row: frames of a video sequence. Second row: detected foreground objects. Third row: foreground objects with shadow removal. Bottom row: result after morphological post-processing.

Figure 5. (a) Grayscale image. (b) Foreground pixels. (c) Shadow removal. (d) Morphological post-processing.

Figure 6. Top row: frames of a video sequence. Second row: detected foreground objects. Third row: foreground objects with shadow removal. Bottom row: result after morphological post-processing.

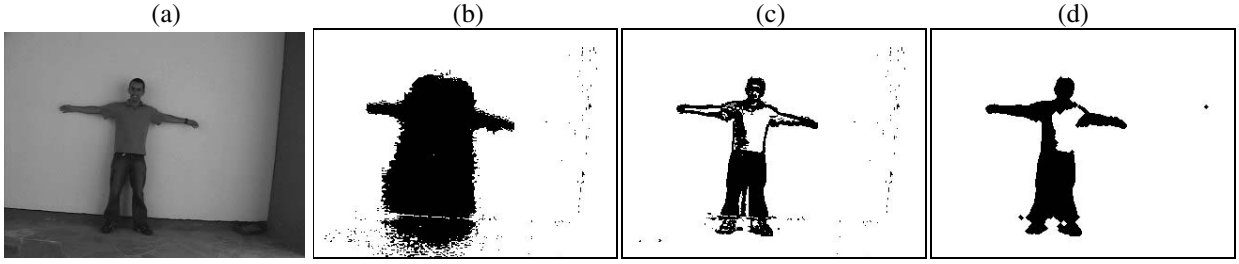

Figure 7. Example of unsuccessful shadow detection. (a) Grayscale image. (b) Foreground pixels. (c) Shadow removal. (d) Morphological post-processing.

Future work will concentrate on extending our approach for robust shadow detection in color sequences. We also intend to further investigate the selection of thresholds L_{ncc}, L_{std} and L_{low}, and to impose spatio-temporal constraints to improve shadow detection.

6 Acknowledgements

This work was developed in collaboration with HP Brazil R&D.

References

[1] J. Agbinya and D. Rees. Multi-object tracking in video. *Real-Time Imaging*, 8(5):295–304, October 1999.

[2] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7):577–586, 2002.

[3] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, October 2003.

[4] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceeedings of the IEEE*, 90(7):1151–1163, 2002.

[5] D. Grest, J.-M. Frahm, and R. Koch. A color similarity measure for robust shadow removal in real time. In *Vision, Modeling and Visualization*, pages 253–260, 2003.

[6] I. Haritaoglu, D. Harwood, and L. Davis. W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000.

[7] P. KaewTrakulPong and R. Bowden. A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. *Image and Vision Computing*, 21(9):913–929, September 2003.

[8] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis. A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21(4):359–381, April 2003.

[9] D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In *European Conference on Computer Vision*, pages A:189–196, 1994.

[10] P. Kumar, K. Sengupta, and A. Lee. A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system. In *IEEE International Conference on Intelligent Transportation Systems*, pages 100–105, 2002.

[11] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *Computer Vision and Image Understanding*, 80(1):42–56, October 2000.

[12] A. Mittal and L. Davis. M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, 51(3):189–203, February 2003.

[13] A. Prati, I. Mikic, C. Grana, and M. M. Trivedi. Shadow detection algorithms for traffic flow analysis: a comparative study. In *IEEE International Conference on Intelligent Transportation Systems*, pages 340–345, 2001.

[14] P. L. Rosin and T. Ellis. Image difference threshold strategies and shadow detection. In *6th British Machine Vision Conf., Birmingham*, pages 347–356, 1995.

[15] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, August 2004.

[16] B. Shoushtarian and H. E. Bez. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1):91–99, 2005.

[17] J. Stauder, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65–76, 1999.

[18] J. J. Wang and S. Singh. Video analysis of human dynamics: a survey. *Real-time imaging*, 9(5):321–346, 2003.

[19] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36:585–601, 2003.

[20] D. Xu, X. Li, Z. Liu, and Y. Yuan. Cast shadow detection in video segmentation. *Pattern Recognition Letters*, 26(1):5–26, 2005.