Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences

Article *in* IEEE Transactions on Multimedia · May 2009

DOI: 10.1109/TMM.2009.2012924 · Source: IEEE Xplore

Claudio Rosito Jung
Universidade Federal do Rio Grande do Sul

Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences

Cláudio Rosito Jung, *Member, IEEE*

Abstract

This paper presents a new method for background subtraction and shadow removal for grayscale video sequences. The background image is modeled using robust statistical descriptors, and a noise estimate is obtained. Foreground pixels are extracted, and a statistical approach combined with geometrical constraints are adopted to detect and remove shadows.

Index Terms

Background subtraction, shadow removal, monochromatic video sequences, object tracking

I. Introduction

Moving object detection is an active research subject in computer vision, with several applications such as surveillance, face tracking and videoconferencing. When static cameras are used, a popular approach is background subtraction, which consists of obtaining a mathematical model of the static background and comparing it with every new frame from the video sequence.

Despite the widespread use of color cameras in the past years, the development of algorithms focused on monochromatic video sequences is still relevant. First, there are several archived monochromatic sequences (e.g. old news broadcast videos), that may be processed in present days. Second, there are some cheap color cameras that present poor chromatic resolution, particularly in developing countries (e.g. low-cost surveillance cameras). Finally, background subtraction algorithms focused on monochromatic sequences may be applied to the luminance component of a color sequence, which is usually faster than using all color components. Furthermore, a complementary algorithm based only on the chromaticity component could be added to the monochromatic module to achieve a more robust result, if there is enough computational power.

This work proposes a new background subtraction algorithm with shadow identification suited for monochromatic video sequences (i.e. it does not rely on color cues). In the training stage, robust estimators are used to model the background, and a fast test is used to detect foreground pixels in the evaluation stage. A statistical model is combined with expected geometrical properties for shadow identification and removal. Finally, morphological operators are applied to remove isolated foreground pixels. The main contribution of the paper is the use of a

Cláudio R. Jung is with PIPCA (Graduate School of Applied Computing), Universidade do Vale do Rio dos Sinos. Avenida Unisinos, 950. São Leopoldo, RS, Brazil 93022-000. Phone: +55 51 3591 1122 ext. 1626. Fax: +55 51 3590 8162. E-mail: crjung@unisinos.br

This work was partially supported by HP Brazil R&D and Brazilian agency CNPq.

EDICS: 7-CONF Videoconferencing and Collaboration Environment; 1-FACE Face Image Analysis and Synthesis

simple and fast foreground test, that combined with a statistical approach for shadow detection and morphological post-processing, can effectively detect foreground objects and remove shadows in monochromatic video sequences.

The remainder of this paper is organized as follows. Section II provides an overview of existing background subtraction models and shadow identification. The proposed technique is described in Section III, and experimental results are presented in Section IV. A comprehensive discussion on several aspects of the proposed method is provided in V, and Section VI presents the conclusions and ideas for future work.

II. Related Work

Several techniques for background subtraction dealing with shadow detection and illumination changes have been proposed in the past years, some requiring only luminance information, and others exploring color information. Next, some of these techniques are briefly described.

Cucchiara's group [1] used a temporal median filtering in the *RGB* color space to produce a background model, and explored the *HSV* color space for shadow detection, classifying as shadows those pixels having the approximately the same hue and saturation values compared to the background, but lower luminosity. Salvador et al. [2] presented a method for shadow identification suited for both still images or video sequences. In particular, their approach for video sequences consists of an initial hypothesis based on *RGB* differences between each frame and the reference image, and a validation stage by exploiting photometric and geometric properties of shadows. Mertel-Brisson and Zaccarin [3] proposed an adaptive Gaussian mixture shadow model based on the *YUV* color space to detect different types of shadows that may arise in a scene, depending on the illumination complexity. More recently, the same authors [4] presented a broader analysis including shadow models in other color spaces, such as *HSV* and Brightness-Chromaticity. Zhang and collaborators [5], [6] used the MoG model for background subtraction and explored "ratio edges" for shadow identification and removal. The core of their approach for shadow identification is to compute local ratios, which are modeled as a chi-squared distributions in shadowed regions. Their quantitative results showed to be one of the best among competitive approaches.

Another class of methods use only luminance information for background removal and shadow identification. Horpraset et al. [7] proposed a color model based on the *RGB* color space that separates the brightness from the chromaticity component to deal with shadows and illumination changes. Stander and colleagues [8] computed pixel ratios in consecutive frames, and explored local variance of pixel ratios within small neighborhoods for shadow detection. Chien et al. [9] built a background image from the accumulated frame difference information, and computed a morphological gradient of foreground objects. Then, regions with small gradients were characterized as shadows. As stated by the authors, the shadow detection stage may find limitations when strongly textured background regions are present. Wang et al. [10] proposed a probabilistic approach for background subtraction and shadow removal. In their approach, a combined intensity and edge measure is used for shadow detection, and temporal continuity is used to improve detection results. Results shown by the authors are good, but the determination of several parameters needed by their model increase the computational cost of the method. Tian et al. [11] used an adaptive background model based on Gaussian mixtures, a local normalized cross-correlation metric to detect shadows, and a texture similarity metric to detect illumination changes. Jacques Jr. et al. [12] used a simple statistical model for background/foreground separation, and explored the standard deviation of pixel ratios in small neighborhoods for shadow identification. Leone and Distante [13] proposed a shadow detection algorithm for intensity images based on texture analysis. In their approach, patches of each new frames are compared with the respective patches of the background model, and Gabor functions are used to detect if textural information remains the same (meaning shadowed regions).

In general, approaches that use only luminance information tend to be faster and less accurate, while methods that explore color information are usually slower but present better results. The proposed method explores statical properties of shadowed pixel combined with geometrical features to obtain an efficient method for shadow identification in grayscale video sequences, presenting accuracy (measured quantitatively in terms of shadow detection and shadow discrimination rates [14]) comparable to other competitive approaches that require color information. The proposed method is explained next.

III. The Proposed Model

The core of the proposed algorithm consists of a simple (but effective) background subtraction scheme, in which a background model is extracted in a training stage, and foreground pixels are obtained in the operation (or test) stage. After the initial foreground extraction, morphological operators are applied to remove isolated responses, and the remaining foreground pixels are processed for shadow identification/removal. Finally, morphological operators are applied again, for further noise cleaning and hole filling.

A. Background Subtraction

This work presents a simple and fast model similar to that presented in [13], but improving the estimate of the background by using a metrically trimmed mean (which produces a more robust model without requiring a large number of training frames), and also including a local spatial coherence for foreground detection, as explained next.

A common simplified model for camera noise is the zero-mean Gaussian distribution [15]. Under this assumption, the temporal series of each background pixel captured by a static camera should be normally distributed, and its mean is the background model. However, moving objects generate outliers in the distribution, so that robust mean estimators are required. A simple, fast and (most of the times) efficient estimate of the mean is the median, as adopted in [12], [13]. However, Kim [16] showed that a better estimate of the mean for a symmetric distribution with asymmetric contamination (which is usually the case when moving objects are present) is the metrically trimmed mean, briefly described next.

Let us denote $\boldsymbol{x} = (x, y)$, let $\{I_1(\boldsymbol{x}), I_2(\boldsymbol{x}), ..., I_T(\boldsymbol{x})\}$ be T image frames used in the training period, and $M(\boldsymbol{x})$ be the image containing the temporal median of each pixel $\boldsymbol{x}$. For $0 \leq \alpha < 1$, the α-metrically trimmed mean $\lambda_\alpha(\boldsymbol{x})$ for each pixel is obtained by computing the temporal average of $I_t(\boldsymbol{x})$, disregarding the largest $[\alpha T]$ deviations from the median (here, $[\ \cdot\ ]$ denotes the greatest integer function). Formally, let us consider an ordering of the differences $|I_t(\boldsymbol{x}) - M(\boldsymbol{x})|$, and define an integer function $1 \leq f(\boldsymbol{x}, t) \leq T$ that returns the position of $|I_t(\boldsymbol{x}) - M(\boldsymbol{x})|$ in such ordering. The α-metrically trimmed mean λ_α is given by

$$\lambda_\alpha(\boldsymbol{x}) = \frac{1}{T - [\alpha T]} \sum_{t \in S_\alpha(\boldsymbol{x})} I_t(\boldsymbol{x}), \tag{1}$$

where $S_\alpha(\boldsymbol{x}) = \{t : f(\boldsymbol{x}, t) \leq T - [\alpha T]\}$. When $\alpha = 0$, the trimmed mean is exactly the average (which is affected by moving objects), and as α gets closer to one, the trimmed mean gets closer to the median value (which is a snapshot of the background image at a given frame). In this work, we determined experimentally that $\alpha = 0.3$ presented good results, averaging the noise and at the same time removing outliers. From this point on $\lambda(\boldsymbol{x})$, will denote the α-metrically trimmed mean value for pixel $\boldsymbol{x}$ using $\alpha = 0.3$.

A scale parameter (such as the standard deviation) is also important to evaluate the spread of the noise around the actual background value. A robust scale estimator is given by the Mean Absolute Deviation (MAD), defined as

$$\mathit{MAD}(\boldsymbol{x}) = \underset{t\in\{1,\dots,T\}}{\text{median}} \left\{ |I_t(\boldsymbol{x}) - M(\boldsymbol{x})| \right\}, \tag{2}$$

where $M(\boldsymbol{x})$ is the temporal median, as described above. In fact, if additive Gaussian noise is assumed, the relation $\sigma(\boldsymbol{x}) = 1.4826\mathit{MAD}(\boldsymbol{x})$ provides an estimate of the standard deviation for each pixel [17].

Pixels belonging to the foreground are probably far from the estimated mean of the distribution. Instead of performing a pixel-wise comparison for foreground detection, we analyze a small neighborhood around each pixel. In fact, foreground pixels usually do not appear isolated in an image: they tend to appear in blobs. A pixel $\boldsymbol{x}$ is assigned to the foreground if

$$\sum_{\boldsymbol{u}\in\Omega(\boldsymbol{x})} w(\boldsymbol{u})|I_t(\boldsymbol{u}) - \lambda(\boldsymbol{u})| > k \sum_{\boldsymbol{u}\in\Omega(\boldsymbol{x})} w(\boldsymbol{u})\sigma(\boldsymbol{u}), \tag{3}$$

where $\Omega(\boldsymbol{x})$ is a small neighborhood centered at $\boldsymbol{x}$, $w(\boldsymbol{u})$ is a weighting mask for each pixel $\boldsymbol{u} = (u, v)$, and k controls the maximum allowed deviation from the mean w.r.t. the standard deviation. In this work, w is the weighted average mask, whose central point has weight 4, vertical and horizontal neighbors weight 2, and diagonal neighbors weight 1. Smaller values for k tend to present more foreground pixels, and the opposite happens for larger values of k. Larger regions Ω tend to generate larger connected blobs, but may degrade the detection close to the boundaries of foreground objects. Experimental results indicated that using a 3×3 neighborhood for Ω and $k = 3$ present a good trade-off between false positives and false negatives.

In practical applications, the Gaussian assumption may not hold completely, particularly when compressed video sequences are analyzed (please, see Section V). In such cases, the detection of foreground pixels using Equation (3) may produce some isolated pixels, or holes in the interior of valid objects. To overcome this problem, we apply sequentially an opening and a closing morphological operator, where the structuring element for the closing is slightly larger than the opening, to fill small holes. The size of these structuring elements should depend on the video resolution, and experimentation indicated that 5×5 and 7×7 diamond-shaped structuring elements are adequate for the opening and closing operations, respectively, when 320×240 video sequences are employed. Clearly, larger structuring elements for the opening operation lead to the removal of larger blobs (more noise cleaning), but they may also destroy foreground objects with holes in the interior; larger elements for closing tend to fill holes better, but they may also connect neighboring blobs into a single one, and deform the shape of the blobs.

An example of background/foreground segmentation using Equation (3) and morphological operators is shown in Fig. 1. Fig. 1(a) shows frame 249 of the *Intelligent Room* video sequence, Fig. 1(b) illustrates the background model computed with the first 100 frames of the sequence, and foreground blobs are shown in Fig. 1(c). As it can be observed, the person was correctly segmented from the background, along with his shadow. However, some isolated false positives were also detected along the image, due mostly to compression issues. Shadow is identified and removed using a statistical approach, and residual noise is removed through some morphological operators, as described next.

B. Shadow Detection

A known drawback of background subtraction techniques is the undesired detection of shadows as foreground objects. Also, global illumination changes (e.g. a cloud covering sunlight) may produce a large number of erroneous foreground pixels. In [6], the authors explored a summation of ratios between pixels in a neighborhood and its central pixel as an illumination invariant feature. This work also explores pixel ratios, but by comparing pixel-wise ratios using the background model $\lambda(\boldsymbol{x})$ and each new frame $I(\boldsymbol{x})$ of the video sequence.

We assume that the intensity $I_s(\boldsymbol{x})$ of a pixel $\boldsymbol{x}$ in a shadowed region (penumbra region) is a scaled version of the background model plus additive Gaussian noise, i.e.

$$I_s(\boldsymbol{x}) = \alpha(\boldsymbol{x})\lambda(\boldsymbol{x}) + \eta(\boldsymbol{x}), \qquad \eta(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma(\boldsymbol{x})^2), \tag{4}$$

where $0 < \alpha(\boldsymbol{x}) < 1$ relates to the intensity of the shadow. If we consider that $\alpha(\boldsymbol{x}) \approx \alpha$ is constant within a small neighborhood $\Omega_s(\boldsymbol{x})$ of the penumbra centered at pixel $\boldsymbol{x}$, then

$$R(\boldsymbol{x}) = \frac{I_s(\boldsymbol{x})}{\lambda(\boldsymbol{x})} = \nu(\boldsymbol{x}), \tag{5}$$

where $\nu(\boldsymbol{x}) = \alpha + \frac{\eta(\boldsymbol{x})}{\lambda(\boldsymbol{x})} \sim \mathcal{N}\left(\alpha, \left(\frac{\sigma(\boldsymbol{x})}{\lambda(\boldsymbol{x})}\right)^2\right)$.

Also, let

$$\mu_R(\boldsymbol{x}) = \frac{1}{N_s} \sum_{\boldsymbol{u} \in \Omega_s(\boldsymbol{x})} R(\boldsymbol{u}), \tag{6}$$

$$\sigma_R(\boldsymbol{x}) = \sqrt{\frac{1}{N_s} \sum_{\boldsymbol{u} \in \Omega_s(\boldsymbol{x})} (R(\boldsymbol{u}) - \mu_R(\boldsymbol{x}))^2}, \tag{7}$$

denote the mean and standard deviation of $R(\boldsymbol{x})$ within $\Omega_s(\boldsymbol{x})$, where $N_s = \#\Omega_s$ is the number of pixels in Ω_s. Assuming that $\nu(\boldsymbol{x})$ are independent normally distributed random variables, then

$$D(\boldsymbol{x}) = R(\boldsymbol{x}) - \mu_R(\boldsymbol{x}) \sim \mathcal{N}\left(0, \sigma_D(\boldsymbol{x})^2\right), \tag{8}$$

where

$$\sigma_D(\boldsymbol{x})^2 = \frac{1}{N_s^2} \left((N_s - 1)^2 \sigma_R(\boldsymbol{x})^2 + \sum_{\substack{\boldsymbol{u} \in \Omega_s(\boldsymbol{x}) \\ (\boldsymbol{u}) \neq (\boldsymbol{x})}} \sigma_R(\boldsymbol{u})^2 \right). \tag{9}$$

The proposed approach for shadow identification consists of scanning each foreground pixel $\boldsymbol{x}$ detected according to Equation (3), and computing the mean $\mu_R(\boldsymbol{x})$ and standard deviation $\sigma_R(\boldsymbol{x})$ within a small neighborhood $\Omega_s(\boldsymbol{x})$ centered at $\boldsymbol{x}$ according to Equation (6). If $\boldsymbol{x}$ is indeed in a shadowed region, then $D(\boldsymbol{x}) = R(\boldsymbol{x}) - \mu_R(\boldsymbol{x})$ is a normally distributed random variable (as explained above), and we can determine a constant $k_s(\beta)$ as a function of the confidence level $0 < \beta < 1$ so that

$$\Pr\left\{|D(\boldsymbol{x})| > k_s(\beta)\sigma_D(\boldsymbol{x})\right\} < \beta. \tag{10}$$

Pixels that satisfy $|D(\boldsymbol{x})| > k_s(\beta)\sigma_D(\boldsymbol{x})$ are not likely associated with shadowed pixels at a confidence level β. However, pixels that do not satisfy this condition are not guaranteed to belong in shadowed regions, and require additional validation. In fact, if a pixel $\boldsymbol{x}$ is shadowed, then at least some of its neighbors are likely to be shadowed as well (except for pixels at shadow borders). Let $S(\boldsymbol{x}) = 0$ if $|D(\boldsymbol{x})| > k_s(\beta)\sigma_D(\boldsymbol{x})$, and $S(\boldsymbol{x}) = 1$ otherwise. The pixel $\boldsymbol{x}$ is identified as shadowed if:

$$I_{\text{low}} \leq \mu_R(\boldsymbol{x}) < 1 \quad \text{and} \quad \sum_{\boldsymbol{u} \in \Omega'_s(\boldsymbol{x})} S(\boldsymbol{u}) > T_s, \tag{11}$$

where I_{low} prevents very dark regions from being marked as shadows, and T_s is the minimum number of pixels within the neighborhood Ω'_s that must pass the shadow test. Clearly, the choice of Ω_s, Ω'_s, T_s, k_s and I_{low} influence the balance between false positives and false negatives in shadow detection. Our experimental results indicated that $I_{\text{low}} = 0.7$, $k_s = 3$ (which leads to a confidence level $\beta = 99.73\%$), Ω_s a 3×3 region, $\Omega'_s(\boldsymbol{x})$ a circular region with radius of 2 pixels (leading to a discretized circular region with 13 pixels), and $T_s = 7$ (so that at least half of the pixels in the neighborhood must also be considered shadowed) lead to a good compromise of false positives and false negatives.

An analogous reasoning can be made to detect highlighted pixels, and the only difference is that parameter $\alpha(\boldsymbol{x})$ in Equation (5) must be greater than one. Hence, a pixel $\boldsymbol{x}$ is considered highlighted if:

$$1 < \mu_R(\boldsymbol{x}) \leq I_{\text{high}} \quad \text{and} \quad \sum_{\boldsymbol{u} \in \Omega'_s(\boldsymbol{x})} S(\boldsymbol{u}) > T_s, \tag{12}$$

where I_{high} prevents very bright pixels from being detected as highlights.

As noticed by other authors (e.g. [6]), cast shadows should not occur in the interior of foreground objects. To remove actual foreground pixels falsely detected as shadows, a flood-fill algorithm is applied. Furthermore, the shadow removal procedure may also produce some isolated responses, and the sequence of opening and closing operators described in the end of Section III-A is applied to obtain the final background/foreground segmentation result.

An example of the proposed approach for shadow removal and noise cleaning is shown in Fig. 1. Fig. 1(d) shows the shadow removal algorithm applied to the result of Fig. 1(c), and the final foreground extraction after morphological post-processing is illustrated in Fig. 1(e). For sakes of comparison, the foreground extraction result using [6] is shown in Fig. 1(f).

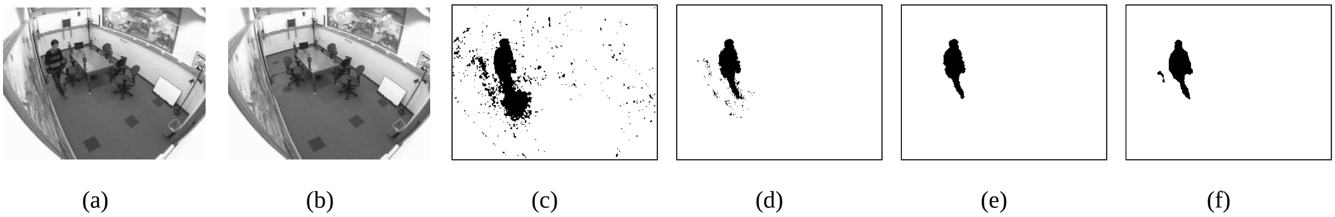


Fig. 1. (a) Frame 249 of the *Intelligent Room* video sequence. (b) Background model. (c) Initial extraction of foreground blobs. (d) Shadow removal. (e) Final result after morphological post-processing. (f) Foreground extraction using [6].

It is important to notice that the notion of "neighborhood" (used for shadow identification and morphological operators) depends on the resolution of the video sequence. The default values described in this Section were devised for 240×320 images, and higher resolution video sequences tend to require larger neighborhoods.

IV. Experimental Results

This Section presents some experimental results obtained with the proposed approach. Results were evaluated qualitatively, by visual inspection, and also quantitatively, by comparing the results with other competitive approaches using quantitative parameters. For color video sequences analyzed in this Session, only the luminance information Y in the YUV color space was employed.

In our analysis, we used three benchmark video sequences for background subtraction and shadow removal (*Campus*, *Intelligent Room* and *Laboratory*)[1], as well as other sequences produced by our personnel. Fig. 2 shows some frames of the *Laboratory* and *Campus* video sequences, along with the final foreground extraction result with the proposed approach. Despite a few false positive and false negative results, it can be observed that foreground objects are discriminated, and shadows are mostly removed.

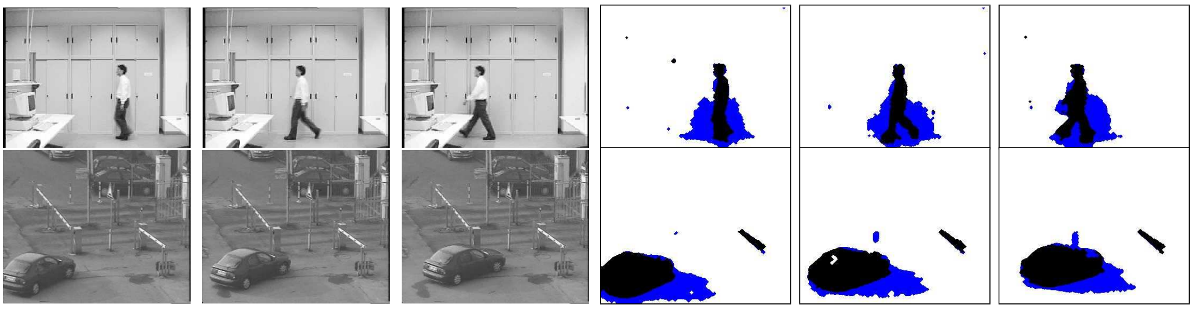


Fig. 2. Some frames extracted from the *Laboratory* (top) and *Campus* (bottom) sequences, with final foreground extraction (pixels before shadow removal and morphological post-processing are marked in blue).

An example of shadow and highlight detection (using $I_{\text{high}} = 1.2$) is shown in Fig. 3. This Figure relates to an outdoor video sequence in a partially cloudy day, which generates light shadows. In this sequence, one cloud moves away, increasing the global luminance of the scene, and generating several highlighted regions. Fig. 3 shows some frames of this video sequence, and the results obtained with the proposed technique. As it can be observed, several pixels were marked as foreground as the illumination changed, and most of them were correctly identified as highlights and removed from the final foreground detection.

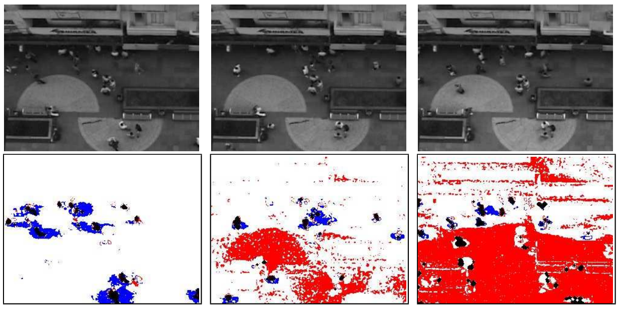


Fig. 3. Top: some frames extracted from the *Santa Maria* video sequence. Bottom: Final foreground extraction, with shadows (blue) and highlights (red) and valid foreground objects (dark).

[1] The sequences *Campus*, *Intelligent Room* and *Laboratory* are available for download at http://cvrr.ucsd.edu/aton/shadow/

A quantitative evaluation of shadow detection was also performed using the shadow detection rate η and the shadow discrimination rate ξ, introduced in [14]. The metric η describes how well shadows are detected by the algorithm, while ξ indicates how well the method discriminates shadows from actual foreground objects. More precisely, these metrics are given by

$$\eta = \frac{TP_S}{TP_S + FN_S}, \quad \text{and} \quad \xi = \frac{\overline{TP}_F}{TP_F + FN_F}, \tag{13}$$

where TP_F denotes true positive foreground detections (i.e., the number of correctly identified foreground pixels), and FN_F denotes false negative foreground detections (i.e. the number of actual foreground pixels that were incorrectly marked as background or shadows). TP_S and FN_S denote analogous parameters regarding shadow identification. The value $\overline{TP}_F$ is the number of ground-truth foreground pixels that were in fact marked as foreground (i.e. ground-truth pixels minus pixels of the foreground incorrectly marked as shadows).

We used three benchmark video sequences for background subtraction and shadow removal to evaluate the proposed method quantitatively in terms of η and ξ. We also compared our results with the statistical parametric approach (SP) described in [18], the statistical nonparametric approach (SNP [7]), two deterministic non-model based methods (DNM1 [19] and DNM2 [8]), the invariant color features (IFC) method given in [2], the Gaussian mixture shadow model (GMSM [3]) and the shadow suppression algorithm based on edge ratios (ER [6]). Parameter setting for these techniques was performed as in [6], and the proposed model was applied using default values described along this paper.

Comparison results are summarized in Table I, and the best result for each video sequence is shown in bold face. For sakes of further comparison, we also computed the mean values of η and ξ achieved by the different techniques considering all videos (these results are shown in the last two columns of Table I). Although it is difficult to provide an overall ranking of these methods (they depend on the video sequence, and there is also the compromise between η and ξ), it can be observed that the proposed approach presented the 2nd, 1st and 4th best results with respect to η, respectively, for the three sequences. Considering ξ, it ranked 2nd, 6th and 1st, respectively. Evaluating the mean values, the proposed method achieved the largest overall shadow discrimination rate, and the second largest shadow detection rate. It should be emphasized that the proposed method employs luminance only, while the remaining techniques explore full color information.

V. Discussion

A. Computational Cost

This section provides an estimate of the computational cost of the test stage in terms of the number of multiplications/divisions (MD) and additions/subtractions (AS). Let N_p be the total number of pixels at each frame of the video sequence.

The first step of the test stage is the initial extraction of foreground pixels using Equation (3), that involves $N_p\#\Omega$ MD and $N_p\#\Omega$ additions, where $\#\Omega$ is the number of pixels in the region. It should be noticed that right-hand size of this Equation is computed only once.

The following step is the concatenation of opening and closing morphological operators. A naïve analysis yields a cost of $\mathcal{O}(N_p\#S_e)$, where $\#S_e$ relates to the dimension of the structuring element S_e. However, very fast implementations of binary dilation and erosion can be achieved using Destination Word Accumulation (DWA)

	Campus		*Intelligent Room*		*Laboratory*		*Mean Values*	
	η	ξ	η	ξ	η	ξ	η	ξ
SP	78.4%	71.2%	79.85%	87.82%	65.4%	92.43%	83.82%	74.55%
SNP	69.1%	76.23%	85.07%	76.6%	87.7%	88.3%	80.38%	80.62%
DNM1	85.2%	80.06%	82.2%	88.2%	84.8%	86.12%	84.79%	84.07%
DNM2	86.3%	79.61%	78.6%	86.8%	69.9%	75.5%	80.64%	78.27%
ICF	72.4%	72.4%	73.45%	86.52%	**88.24**%	93.57%	84.16%	78.03%
GMSM	66.2%	73.2%	73.6%	79.1%	76.62%	75.14%	75.81%	72.14%
ER	**87.95**%	**97.74**%	88.63%	**88.91**%	86.28%	92.64%	**93.10**%	87.62%
Proposed	87.69%	92.18%	**97.67**%	86.21%	85.84%	**95.10**%	91.16%	**90.40**%

TABLE I

COMPARISON OF ANALYZED ALGORITHMS IN TERMS OF THE SHADOW DETECTION RATE η AND THE SHADOW DISCRIMINATION RATE ξ.

methods [20], by processing chunks of bits in 32 (or 64) bit words (in [9], the authors report a 5.5 speedup when using DWA for open-close operations).

The costliest step of the proposed method concerns the identification of shadowed pixels, which is applied only to those pixels considered as foreground according to the steps described so far. Let N_f denote the number of pixels classified as foreground. We must compute N_f pixel ratios according to Equation (5), and the estimation of local means and standard deviations, according to Equations (6) and (7) require $N_sN_f + 2N_f$ MD and $2N_sN_f$ AS, where $N_s = \#\Omega_s$ is the number of pixels in the neighborhood Ω_s (the square root is not taken, since σ_R^2 is needed in the following Equations). Equations (8) and (9) require $N_sN_f + N_f$ AD, and $2N_f$ MD.

Condition (10) requires N_f evaluations, and the final shadow test given by condition (11) requires $N_s'N_f$ summations, where $N_s' = \#\Omega_s'$. Also, the flood-fill algorithm and the same morphological operators used after the initial foreground pixel extraction are applied to obtain the final result.

Using the default regions Ω, Ω_s and Ω_s' described in the paper, the proposed approach requires $10N_p + 13N_f$ MD, and $9N_p + 41N_f$ AS, plus the morphological operators and hole filling. For sakes of comparison, the approach in [6] requires $23N_p + 78N_f$ MD, $17N_p + 42N_f$ AS, plus $3N_p$ 3-D Gaussian possibility operations (required in the initial foreground test using the mixtures of Gaussians) and hole-filling operations. The theoretical cost of the other methods listed in Table I can be found in [6].

A prototype of the proposed algorithm was implemented in MATLAB (with no optimizations such as DWA), and experiments were performed in a PC with a Pentium Core 2 Duo 2.13 GHz processor, 2GM RAM running on Windows. As explained before, the execution time is highly dependent on the number of foreground pixels, and average processing speeds for the *Campus*, *Intelligent Room* and *Laboratory* sequences were, respectively,

25.75, 30.56 and 20.31 frames per second. Clearly running times depend on the number of foreground pixels in the initial detection (because of the shadow removal procedure). In average, the processing time analyzing all video sequences is around 0.35ms for frames presenting less than 1500 foreground pixels, and around 0.67ms when $10,000$ foreground pixels are present. Clearly, an optimized implementation of the proposed method in a compiled language (such as C or C++) would reduce running times.

B. Compression Issues

One important issue hardly mentioned by papers related to background subtraction is that most video sequences are compressed. There are several compression algorithms that affect image quality in different aspects, but most of them tend to present artifacts close to image edges. Also, the Gaussian noise assumption made by several background subtraction techniques (including the proposed one) clearly does not hold when compressed videos are used. In fact, most benchmark sequences publicly available on the internet, such the ones used in this paper (I*ntelligent Room*, *Campus* and *Laboratory*) are compressed, which makes benchmark results difficult to evaluate. It should be also noticed that most commercial webcams send compressed data to the computer, which influences the Gaussian assumption as well.

Fig. 4 illustrates the standard deviation $\sigma(\boldsymbol{x})$ for the *Laboratory* and the *Intelligent Room* sequences. As it can be seen, the estimated standard deviation is significantly larger close to object edges, since intensity values of compressed videos tend to present more variations in these positions. As a consequence, the results of background subtraction tend to be worse close to the edges of the background model. To cope with this problem, a future approach could take into account also the edge map of the background model when estimating the noise variance. Fig. 4 also shows compression effects using different video capture devices. Figs. 4(c) and 4(d) illustrate the estimated noise standard deviation of the same scenario using two different cameras: an analog camera combined with a digitizing board (Pinnacle PCTV USB2.0), and a Logitech QuickCam Pro5000 webcam USB2.0. It is noticeable that the blockiness effect is more salient in the webcam (Fig. 4(d)), which affects the background subtraction and shadow identification processes.

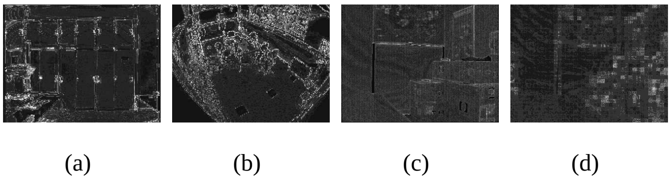


Fig. 4. Effect of video compression on the estimation of the noise standard deviation $\sigma(\boldsymbol{x})$.

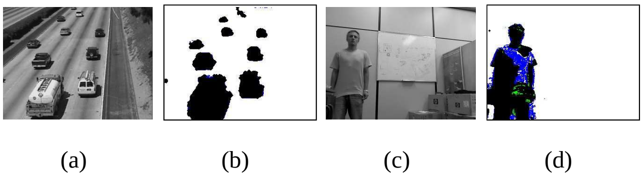
C. Limitations

The proposed approach was focused on the identification of weak shadows, indicating its use for indoor environments. For outdoor environments in sunny days, strong shadows are very similar to very dark objects, and the threshold I_{low} in Condition (11) would consider such shadows valid objects. One possibility would be to decrease the value of I_{low} to capture darker shadows, but the number of false positives (actual objects marked as shadows) would increase as well.

Another limitation of the proposed method arises when homogeneous foreground objects appear in front of homogeneous portions of the background. In this case, pixel ratios $R(\boldsymbol{x})$ in Equation (5) would be very similar, and the object may be erroneously classified as a shadowed pixel. If the false detection happens only in the interior of an object, the hole-filling algorithm may solve this problem, but sometimes the false detection "connects" with the background, and the hole-filling fails. It is important to note that similar situations also cause theoretical problems in competitive approaches, such as [6].

Finally, there may be some video sequences where foreground objects may present very distinct colors when compared to the background, but the luminance component of these objects may be similar (or exactly the same). In these cases, the proposed approach would fail (as any technique relying only on luminance would).

Fig. 5 illustrates examples where the proposed algorithm fails. Figs. 5(a)-(b) show an example of of the proposed method applied to the *Highway* video sequence. This outdoor sequence presents strong shadows, that are not captured by the proposed algorithm. Figs. 5(c)-(d) shows a situation of a homogeneous object in front of a homogeneous background, and some foreground pixels are falsely marked as shadows. In Fig. 5(d), pixels in green are those initially marked as shadows, but removed after flood-filling. Pixels in blue were marked as shadows in the final detection, and were not removed by flood-filling because they connect with the background.



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

Fig. 5. (a), (c) Original frames. (b), (d) Respective foreground objects.

VI. Conclusions

This paper presented a novel approach for background subtraction with shadow detection using only luminance information. For background subtraction, the metrically trimmed mean was employed as a robust estimate of the background model, and the MAD was adopted as a scale estimate. Local spatial coherence was also used to minimize the occurrence of isolated foreground pixel. For shadow (and optionally highlight) removal, a statistical model based on the relations of pixel ratios within small neighborhoods was proposed, and a hole filling algorithm was used to complete pixels wrongly detected as shadows. Finally, a morphological post-processing scheme was employed to remove residual noise.

The qualitative (visual) results shown in Section IV indicate that the proposed algorithm can effectively extract foreground pixels and remove shadows/highlights in static cameras. Quantitative results in terms of the shadow detection and the shadow discrimination rates showed that the proposed approach achieves a performance comparable to state-of-the-art algorithms that use the full color information for shadow identification.

For future work, a more thorough analysis of the effect of video compression on the hypothesized noise model is planned. Also, we plan to include color information in a complementary manner, so that chromaticity cues would be used when enough computational power is available.

ACKNOWLEDGMENTS

The author would like to gratefully thank Wei Zhang for providing ground truth data concerning the *Campus* and *Laboratory* video sequences. The author also thanks the anonymous reviewers, for their valuable comments.

REFERENCES

[1] 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*, vol. 25, pp. 1337–1342, October 2003.

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

[3] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Washington, DC, USA), pp. 643–648, IEEE Computer Society, 2005.

[4] N. Martel-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1133–1146, 2007.

[5] W. Zhang, X. Z. Fang, and X. Yang, "Moving cast shadows detection based on ratio edge," in *Proceedings of the 18th International Conference on Pattern Recognition*, (Washington, DC, USA), pp. 73–76, IEEE Computer Society, 2006.

[6] W. Zhang, X. Z. Fang, and X. K. Yang, "Moving cast shadows detection using ratio edge," *IEEE Transactions on Multimedia*, vol. 9, pp. 1202–1214, October 2007.

[7] T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *IEEE ICCV Frame-Rate Workshop*, 1999.

[8] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 65–76, 1999.

[9] 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*, vol. 12, no. 7, pp. 577–586, 2002.

[10] Y. Wang, T. Tan, K. Loe, and J. Wu, "A probabilistic approach for foreground and shadow segmentation in monocular image sequences," *Pattern Recognition*, vol. 38, pp. 1937–1946, November 2005.

[11] Y. Tian, M. Lu, and A. Hampapur, "Robust and efficient foreground analysis for real-time video surveillance," in *IEEE Computer Vision and Pattern Recognition*, pp. I: 1182–1187, 2005.

[12] J. C. S. Jacques Jr., C. R. Jung, and S. R. Musse, "A background subtraction model adapted to illumination changes," in *IEEE International Conference on Image Processing*, pp. 1817–1820, IEEE Press, 2006.

[13] A. Leone and C. Distante, "Shadow detection for moving objects based on texture analysis," *Pattern Recognition*, vol. 40, no. 4, pp. 1222–1233, 2007.

[14] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.

[15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley Publishing Company, 1992.

[16] S.-J. Kim, "The metrically trimmed mean as a robust estimator of location," *The Annals of Statistics*, vol. 20, no. 3, pp. 1534–1547, 1992.

[17] P. J. Huber, *Robust Statistics*. New York: John Wiley and Sons, 1981.

[18] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, "Moving shadow and object detection in traffic scenes," in *Proceedings of the International Conference on Pattern Recognition*, (Washington, DC, USA), pp. 321–324, IEEE Computer Society, 2000.

[19] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *IEEE International Conference on Intelligent Transportation Systems*, pp. 334–339, 2001.

[20] D. S. Bloomberg, "Implementation efficiency of binary morphology," in *Proceedings of the 4th International Symposium for Mathematical Morphology*, pp. 209–218, 2002.