Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation

Nicolas Martel-Brisson and André Zaccarin
Computer Vision and Systems Laboratory
Department of Electrical and Computer Engineering
Laval University, Québec, Qc, Canada
(nmartel,andre.zaccarin)@gel.ulaval.ca

Abstract

In background subtraction, cast shadows induce silhouette distortions and object fusions hindering performance of high level algorithms in scene monitoring. We introduce a nonparametric framework to model surface behavior when shadows are cast on them. Based on physical properties of light sources and surfaces, we identify a direction in RGB space on which background surface values under cast shadows are found. We then model the posterior distribution of lighting attenuation under cast shadows and foreground objects, which allows differentiation of foreground and cast shadow values with similar chromaticity. The algorithms are completely unsupervised and take advantage of scene activity to learn model parameters. Spatial gradient information is also used to reinforce the learning process. Contributions are two-fold. Firstly, with a better model describing cast shadows on surfaces, we achieve a higher success rate in segmenting moving cast shadows in complex scenes. Secondly, obtaining such models is a step toward a full scene parametrization where light source properties, surface reflectance models and scene 3D geometry are estimated for low-level segmentation.

1. Introduction

In contrast with other fields such as 3D modeling where cast shadows provide information about geometry, in background subtraction cast shadows are often considered a nuisance. Labeled as foreground, cast shadows induce silhouette distortions and object fusions, hindering the performance of high level algorithms in scene monitoring, target counting, and object recognition. Recently, algorithms to remove cast shadows from foreground data have grown in complexity and statistical methods modeling moving cast shadows have been proposed. Adopting this philosophy, we develop a nonparametric framework based on an illumination model and surface properties to model surfaces' appearance under moving cast shadows. Without *a priori* knowledge of scene geometry, light sources' positions and 3D shape of foreground objects, the model is built from scene activity in an unsupervised environment, and is used to segment cast shadows in complex scenes with a high success rate. Modeling surface appearance under cast shadows increases our understanding of the scene and close the knowledge gap toward a full scene parametrization where light source properties, surfaces reflectance models and scene 3D geometry can be estimated from a video sequence.

2. Cast shadows

Shadow detection algorithms are either property-based or model-based algorithms. Property-based approaches use features like geometry, brightness, or color to identify shadowed regions. Unlike model-based techniques, they minimize any *a priori* knowledge of scene geometry, foreground objects, or light sources. The model we introduce is property-based.

When an object casts a shadow on a surface, it deprives the surface of direct illumination from a light source, hence inducing a variation of its appearance. This variation is more or less severe as a function of the scene composition, such as the presence of other light sources and the reflectivity properties of other scene objects. This variation is one of the main properties used in the literature to segment cast shadows in background-foreground segmentation algorithms. Given the value (in RGB space or other) of a surface under the absence of cast shadows (the cyan sphere labeled BG - background - in Fig. 1), many algorithms assume that the value of the surface under cast shadows will be linearly attenuated from the BG value, and thus fall on or near the line between the BG value and the origin of the RGB cube. This type of modeling has been used in different color space [1] and some authors have used training sequences [2] or

statistical models [3, 4] to capture variations of this model or to adapt it to the observed scene [5]. Frequently, this linear model has also been used with other shadow-induced properties such as edge information and spatial gradients [6, 7].

Algorithms based on this linear hypothesis will falsely label pixels as cast shadows when foreground objects have chromaticity values similar to that of the background. Furthermore, this hypothesis does not hold when light sources are not exclusively white or when the objects' chromatic properties diffuse on their neighbors (color bleeding). This situation occurs, for example, in outdoor scenes when blocking a surface from direct sunlight since the light scattered by the sky has a spectrum which differs from that of the sun [8]. Consequently, RGB background values under a cast shadow will not necessary be proportional to RGB values under direct light. This non-proportionality is addressed in [8] via a dichromatic model. The approach however is supervised and mainly suited for outdoor situations.

The appearance of a shadowed surface shows a certain regularity even in scenes with complex illumination conditions. This regularity is caused by several factors: light sources are generally fixed and have a time invariant spectral power distribution (SPD); the foreground objects circulating in the scene have a similar scale factor and they move following physical constraints like walls, ground, roads, hallways, etc. Since different foreground objects block light sources in a similar manner, the shadows cast on the background surfaces are relatively similar at the pixel level. This phenomenon is particularly strong in busy hallways or highways where different people or different cars induce the same intensity fluctuation on a surface when blocking a light source. Statistical learning techniques can then be used to capture the surface appearance under cast shadows, such as Gaussian mixture models (GMM) [3, 4].

In this paper, we first propose in section 3 a new model to describe the value of a background surface under cast shadow. Unlike the so-called linear model, our model introduces an ambient illumination term which determines the direction, in color space, along which the value of a background surface will be found under cast shadows (see Fig. 1). We call this direction the cast shadow direction. This model is more general than the linear model and explains off-axis attenuation as described or observed in [3, 8]. In section 4, we present the unsupervised kernel-based approaches used to estimate this cast shadow axis for each pixel and the illumination attenuation observed along the same axis under cast shadows. Finally, we develop in section 5 the probability of observing a cast shadow given a sample value using our model and correlation of spatial gradients. Using three different video sequences, we present in section 6 results illustrating the performance and validity of our model.

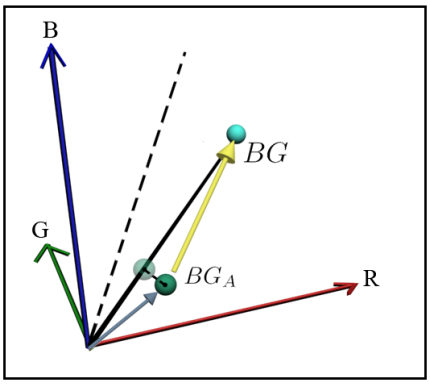

Figure 1. The black line from the BG value to the origin of the RGB space is the cast shadow axis in models assuming linear attenuation. Our model defines an ambient value BG_A which defines a different direction (yellow line) on which cast shadow values are modeled.

3. Physical cast shadow model

A surface appearance depends on its reflectivity properties and its total incident energy. We limit ourselves to a Lambertian image formation model [9] where surfaces showing specular reflection are neglected and light energy is scattered uniformly. Under these considerations, the camera sensor response at the pixel level, ρ_k, depends on the SPD denoted $E(\lambda)$, the surface reflectance function $R(\lambda)$, the sensor spectral sensitivity $Q_k(\lambda)$ and a shading factor σ (dot product between the surface normal and the illumination direction):

$$\rho_k = \sigma \int E(\lambda)R(\lambda)Q_k(\lambda)d\lambda. \quad (1)$$

We propose to approximate the illumination incident to a surface point by an ambient illumination $A(\lambda)$ term and the contribution of M punctual light sources L_M (where $M \rightarrow \infty$ for large area lights). In this model, the ambient illumination is independent from the surface normal ($\sigma_a = 1$), thus we can write:

$$\rho_k = \int \left[A(\lambda) + \sum_{m=1}^{M} \sigma_m L_m \right] (\lambda)R(\lambda)Q_k(\lambda)d\lambda. \quad (2)$$

The ambient illumination is highly dependent on the scene (indoor/outdoor) but also on the objects surrounding a surface point. For example, lights reflected by highly colored surfaces will "bleed" on their surrounding, thus influencing the surface point appearance.

To simplify the model, we propose a heuristic about the nature of the punctual light sources: for a given surface point, we assume that they share the same SPD profile but with different power factor $L_m = \alpha_m L$. Equation 2 then becomes

$$\rho_k = A_k + \sum_{m=1}^{M} \sigma_m \alpha_m \int L(\lambda) R(\lambda) Q_k(\lambda) d\lambda \quad (3)$$

where $A_k = \int A(\lambda) R(\lambda) Q_k(\lambda) d\lambda$ is the ambient illumination contribution. Therefore, for a 3 sensor camera with a linear response ($\rho_{1,2,3} = \rho_{R,G,B}$), the response to each light source shares the same direction in RGB space, and the light sources' total contribution follows a line in the color space (see Fig. 2a). Hence, blocking one or many light sources or a fraction of an area light induces a sensor response following the cast shadow direction, $\hat{S}$. The darkest color value BG_A (green sphere in Fig. 2) is obtained when the surface point is illuminated only with the ambient illumination and the brighter color value BG (cyan sphere in Fig. 2) represents the background color value.

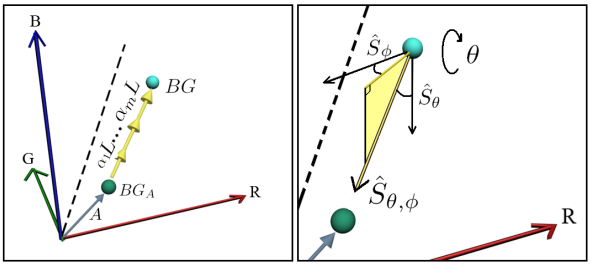

Figure 2. (a) Light contributions on a surface point appearance, from ambient illuminated appearance (green) to direct illumination of all light sources (cyan); (b) Cast shadow direction, $\hat{S}$, on which shadowed background value will be found, expressed in spherical coordinates.

Note that in this model, the ambient lighting can have a different SPD than the light source(s), as is the case for example in an outside scene where the ambient light (sky) has a different SPD than the sun. Our model is therefore more general than modeling the surface value under cast shadows as a linear reduction of the background value, which assumes that the ambient illumination and lights share the same SPD. Finally, since our model is pixel-based, the SPD of ambient and punctual sources can vary from one pixel to the next. With this pixel-based approach, the heuristic imposing a unique SPD for punctual light sources at a pixel becomes significantly less restrictive.

4. Estimation of model parameters

The physical model in equation 3 supports the existence of a direction $\hat{S}$, in color space, along which the value of a background surface will be found under cast shadows. The first step in our approach is to estimate this direction (sect. 4.1). Once estimated, we then project the sample values of the observed shadows on this axis to model illumination attenuation (sect. 4.3). Since pixels represent surfaces with different reflectance functions, the cast shadow parameters could differ for each pixel and are therefore estimated on a pixel-basis.

Our starting point is an N frame video sequence captured by a static camera observing a scene with significant foreground activity. For a given pixel X, we therefore have a set of N observations $x_i, i = 1, ..., N$ and these observations are from both the foreground and background. We use Kernel Density Estimators [11] to model the background and the foreground. From these models, we can then compute the posterior probability that a pixel belongs to either the background $P(B|x)$ or the foreground $P(FS|x)$ (i.e. foreground or cast shadow). It is also possible to obtain such posteriors by other low-level algorithms such as Gaussian mixture models [14].

4.1. Estimation of cast shadow direction ($\hat{S}$)

For a given pixel, we want to estimate the cast shadow directions $\hat{S}$ associated to its background values. To reduce the computational cost, the estimation is limited to the most likely background value BG for each pixel, but the approach could be generalized to model the cast shadow behavior for multimodal backgrounds. $\hat{S}$ is a unitary vector in a coordinate system centered on **BG**, and we use $\hat{S}_{\theta,\phi}$ to denote its direction in spherical coordinates. Similarly, for each sample x_i, we can compute its direction $\mathbf{s}_i = (\theta_i, \phi_i)^T$ relative to **BG**. We propose to compute a non-parametric distribution of the direction $\mathbf{s}$ from samples that are likely to represent cast shadow values, and set $\hat{S}_{\theta,\phi}$ as the maximum likelihood of that distribution.

We use KDE to compute this non-parametric distribution $p(\mathbf{s})$ from the set $\{\mathbf{s}_i\}$. However, to accurately capture the cast shadow direction, samples from the background and the foreground must be filtered out. Thus, we weight each sample $\mathbf{s}_i$ using two parameters, $(FS|x_i)$ and ω_i. The first parameter, $FS|x_i$, assigns to each sample a weight equal to its posterior probability that it does not belong to the background. The second parameter, ω_i, is used to separate foreground values from cast shadow values and its computation is described in sections 4.1.1 and 4.1.2.

We compute the non-parametric distribution of the cast shadow direction as

$$p(\mathbf{s}) = \frac{1}{\sum_{i=1}^{N} \omega_i P(FS|x_i)} \sum_{i=1}^{N} \omega_i P(FS|x_i) K_{H_{\theta,\phi}}(\mathbf{s} - \mathbf{s}_i) \quad (4)$$

and $\hat{S}_{\theta,\phi} = \arg\max_{\mathbf{s}} p(\mathbf{s})$ is obtained using binned kernels [10]. We use a bivariate Gaussian kernel function given by

$$K_H(\mathbf{s}) = |H_{\theta,\phi}|^{-1/2} (2\pi)^{-1} exp\left(-\frac{1}{2}\mathbf{s}^T H_{\theta,\phi}^{-1}\mathbf{s}\right). \quad (5)$$

where the kernel bandwidth $H_{\theta,\phi}$ captures the uncertainty related to the each direction sample $(\mathbf{s}_i)$. This uncertainty is

function of the uncertainty on the sample value (x_i) as well as the sample's distance from the BG value, $\|\mathbf{BG} - \mathbf{x}_i\|$.

The uncertainty on the sample values is mostly due to camera noise and is modeled by a 3-variate gaussian kernel with diagonal matrix bandwidth $H_{RGB} = diag(h^2(1), h^2(2), h^2(3))$ where $\mathbf{h} = (h_r, h_g, h_b)^T$. These three parameters are estimated following the method described in [11] where the median of the frame differences $|x_i - x_{i-1}|$ allows separation of camera noise from background modes.

As illustrated in Fig. 3, the RGB bandwidth of samples close to the BG value (blue dot) induce a larger subtended arc on the parametrization sphere than samples at a greater distance (red dot). The bandwidth reflecting this relationship is symmetric and sample dependent; it is given by the diagonal matrix $H_{\theta,\phi} = diag(h^2_{DH}, h^2_{DH})$ where

$$h_{DH} = atan\left[(\|h\| \,/\, \|\mathbf{BG} - \mathbf{x}_i\|\right]. \tag{6}$$

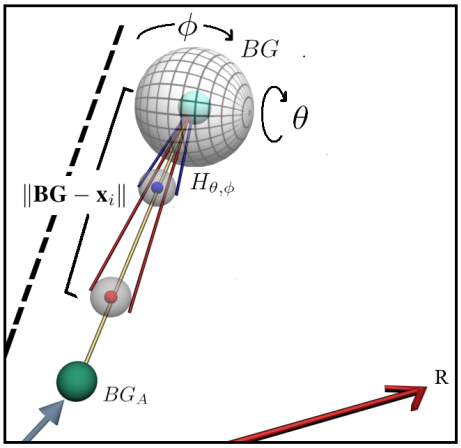

Figure 3. The angular bandwidth $H_{\theta,\phi}$ is function of the RGB sample bandwidth and its distance from the BG value. The angular bandwidth of close samples to BG (blue dot) is larger than the bandwidth of farther samples (red dot).

As explained above, samples contributing to the KDE must be weighted proportionally to the likelihood that these samples are cast shadow samples and not foreground. We compute this likelihood $\omega_i = \omega_{li}\, \omega_{\nabla i}$ by multiplying two uncorrelated shadow characteristics: an RGB linear reduction (ω_{li}) and a factor ($\omega_{\nabla i}$) which measures how well background spatial gradients are preserved under cast shadows.

4.1.1 Linear reduction (ω_{li})

The term ω_{li} measures how parallel the angular direction is to the line from BG to the RGB color space origin:

$$\omega_{l,i} = \frac{\langle \mathbf{BG}, \mathbf{BG} - \mathbf{x}_i \rangle}{\|\mathbf{BG}\| \, \|\mathbf{BG} - \mathbf{x}_i\|} \tag{7}$$

As argued earlier, the cast shadow direction is often not perfectly parallel to the linear approximation found in the literature. Here ω_{li} is defined to only penalize angular directions that deviate strongly from this relation. Negative dot products from samples brighter than the background are set to zero since shadow samples exempt of noise are always darker than the surface under direct illumination [13].

4.1.2 Spatial information ($\omega_{\nabla i}$)

Although our cast shadow model is pixel based, significant information can be extracted from the pixel spatial context. As shown in [12], strong edges derived from surface textures and the scene 3D shapes are well preserved during illumination variations; hence, gradient magnitudes of shadowed surfaces are modified but their directions are kept constant. We take advantage of this property by comparing the 2D direction of the gradients computed in the luminance image $\nabla(I)$ of a given frame to the gradients computed on the background luminance $\nabla(BG)$ image:

$$\omega_{\nabla(I,BG)} = \frac{\langle \nabla(BG), \nabla(I) \rangle}{\|\nabla(BG)\| \, \|\nabla(I)\|} \frac{\|\nabla(BG)\|}{max(\|\nabla(BG)\|, \|\nabla(I)\|)}. \tag{8}$$

The second part of eq.8 heavily penalizes samples where the gradient magnitude from a shadow sample is amplified instead of reduced. If both gradient magnitudes are smaller than the median of the gradients computed on the BG image, the gradient direction is subject to noise and the criteria must be relaxed: we therefore set $\omega_{\nabla(I,BG)} = 1$ in those cases.

If light sources are clustered in the scene and are relatively brighter than ambient illumination, blocking them induces a sharp shadow where a strong gradient is visible at the light/shadow border. This gradient strongly correlates with the gradient computed in the posterior image $\nabla(P(B|x))$. Sharp edges due to deep cast shadows are only visible if the surface luminance drops significantly. Hence, the projection of $\nabla(P(B|x))$ and $\nabla(I)$ is modulated by the luminance gradient magnitude itself

$$\omega_{\nabla(I,P(B|x))} = \frac{\langle \nabla(P(B|x)), \nabla(I) \rangle}{\|\nabla(P(B|x))\| \, \|\nabla(I)\|} \, |\nabla(I)| \tag{9}$$

The overall spatial gradient correlation term is given by

$$\omega_{\nabla i} = max\left[\omega_{\nabla(I,P(B|x))}, \omega_{\nabla(I,BG)}\right]. \tag{10}$$

Figure 4 illustrates the different gradient terms for a frame where deep shadows are present.

4.1.3 Example of angular estimation result

Fig. 5(a) shows the distribution $p(\mathbf{s})$ for the pixel circled in red in Fig. 4. The black continuous line drawn from the BG

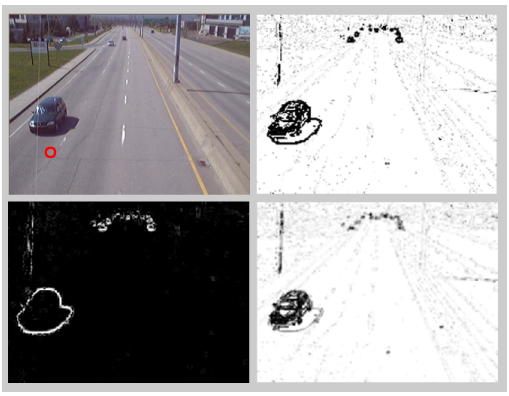

Figure 4. From top left to bottom right: frame of the sequence, gradient correlation $\omega_{\nabla(I,BG)}$, gradient correlation $\omega_{\nabla(I,P(B|x))}$ and the overall gradient correlation ω_{∇_i}.

value (blue circle) in Fig. 5(B) shows the estimated $\hat{S}$ in the RGB space. Due to the the illumination condition and acquisition setting, the shadow direction differs from the linear relation represented by the dashed line. Both foreground and shadow samples ($P(FS|x_i) > 0.5$) are drawn in red.

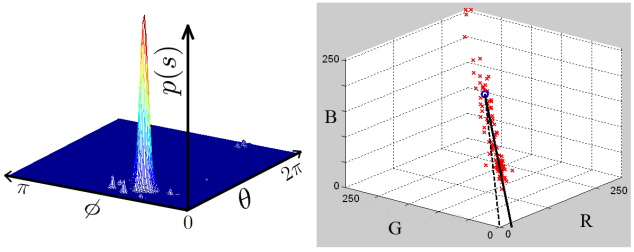

Figure 5. (a) Probability distribution $p(\mathbf{s})$ for one pixel; (b) $\hat{S}$ is drawn in black from the BG value (blue circle). Note that $\hat{S}$ differs from the linear relation. Foreground and shadow samples are plotted in red.

4.2. Illumination and Chromaticity Metric

Once the cast shadow direction has been estimated, a metric must be defined to evaluate how much a sample respects this criterion. Our decomposition is inspired by Horprasert, et al., [2] where the distance between a pixel value x and the BG value is defined in terms of illumination attenuation β and chromaticity distortion CD. The illumination attenuation is obtained by computing the projection of the vector $\mathbf{v}_i = \mathbf{x}_i - \mathbf{BG}$ on the cast shadow vector

$$\beta = \frac{\left\langle \hat{S}, \mathbf{v}_i \right\rangle \|\mathbf{v}_i\|}{\|\mathbf{BG}\|}. \tag{11}$$

The chromaticity distortion is the length of the perpendicular vector between the sample value and the cast shadow direction:

$$CD = \sqrt{\sum_{j=1}^{3} \left[\frac{(\mathbf{v}_i(j)) - \left\langle \hat{S}, \mathbf{v}_i \right\rangle \hat{S}(j)}{h(j)} \right]^2}. \tag{12}$$

The chromaticity distortion is an indicator of how much the color of sample x_i differs from the color value of the background under cast shadows.

As shown in [2], the CD distribution is not Gaussian due to the correlation between the color channels. Instead of imposing a binary threshold as in [2] based on cumulative distribution, we propose to use a decreasing function of CD

$$f(CD) = \begin{cases} -(\frac{CD}{CD_{max}})^2 + 1 & \text{if } (\frac{CD}{CD_{max}}) < 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where $CD_{max} = 3$ for all our results. Fig. 6 illustrates the illumination and chromaticity decomposition.

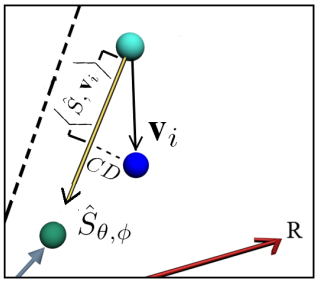

Figure 6. Illumination attenuation and chromaticity distortion.

4.3. Illumination Attenuation Estimation

As argued in [3], even if samples are situated on the cast shadow direction, this does not mean than these samples represent cast shadows. For example, to base our decision only on this criterion would label cars with different shades of gray as shadow. For the general indoor situation where shadows are shallow, a threshold is often seen on the maximum illumination attenuation to prevent dark foreground objects to be labeled as shadow.

To overcome this issue, we model the illumination likelihood of cast shadows $P(\beta|S)$ and of foreground samples $P(\beta|F)$ using the samples x_i that are darker than the BG, (i.e., $< \hat{S}, \mathbf{s}_i >$ positive). We also weigh the samples inversely with their distance from the cast shadow direction. As a result, both illumination attenuation likelihoods are generated from samples that are either cast shadows or foreground samples sharing similar illuminance and color characteristics.

The nonparametric nature of illumination variation under cast shadows [3] is best suited to be modeled by a one dimensional KDE, and we generate the likelihood functions

as

$$P(\beta|S) = \rho\left[\frac{1}{\sum_{i=1}^{N}\psi_{i,S}}\sum_{i=1}^{N}\psi_{i,S}K_{H_\beta}(\beta-\beta_i)\right]+(1-\rho)U(\beta), \quad (14)$$

where

$$\psi_{i,S} = f(CD)P(FS|x_i)\omega_{\nabla i} \quad (15)$$

and $H_\beta = \|h\|$. The $\rho \approx 0.95$ factor and the uniform probability density function (PDF) $U(\beta)$ allows previously unseen illumination attenuation to be correctly labeled. Eq. (15) shows that more weight is given to samples which are 1) close to the cast shadow direction, 2) are not background values $(P(FS|x))$ and 3) for which the local spatial gradients are correlated (sec. 4.1.2).

Similarly, $P(\beta|F)$ can be obtained by substituting $\omega_{\nabla i}$ by $(1-\omega_{\nabla i})$ in (15) since more weight should be given to samples for which the spatial gradients are uncorrelated. Moreover, we use the same function $f(CD)$ since we only model the subset of foreground samples situated on the cast shadow direction.

We obtain priors on shadow and foreground values by summing the sample weights

$$P(S) = \sum_i \psi_{i,S}, \quad P(F) = \sum_i \psi_{i,F}. \quad (16)$$

Both priors are then normalized $(P(F)+P(S)=1)$ and finally from Bayes' theorem we get

$$P(S|\beta) = \frac{P(S)P(\beta|S)}{P(S)P(\beta|S)+P(F)P(\beta|F)} \quad (17)$$

Obviously, samples brighter than the background cannot be cast shadow. Therefore, $P(S|\beta<0)=0$ and $P(F|\beta<0)=1$.

All likelihoods, priors and posteriors in eq.(14), (16), and (17) are conditioned on samples being 1) darker than the background, 2) close to the cast shadow direction and 3) showing correlated spatial gradients. The notation has been simplified for clarity purposes. An example of illumination attenuation likelihoods weighted by their respective priors is shown in Fig. 7.

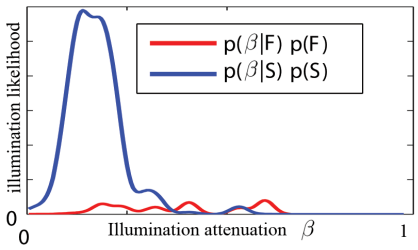

Figure 7. Non parametric shadow and foreground illumination attenuation likelihoods.

5. Cast Shadow and Foreground Posterior

So far in this paper, we have introduced a new physical model describing the appearance of a surface under cast shadow. This pixel-based model is ultimately characterized by 1) the posterior distribution of the illumination attenuation for both shadow and foreground hypothesis, and 2) the spatial position of a sample with respect to the cast shadow direction. We will now integrate these two elements with a spatial gradient measure to compute $P(S|x)$ and $P(F|x)$, respectively the posterior distributions of a sample x_i under shadow and foreground hypothesis. These two posteriors distributions can then be used directly to segment cast shadow samples from non-background samples.

5.1. Cast Shadow Posterior, $P(S|x)$

First, we decompose $P(S|x)$ over the (FS, BG) domain. Since $p(S|x, BG) = 0$ we have

$$P(S|x) = P(S|x, FS)P(FS|x). \quad (18)$$

Samples from the set foreground/shadow can be seen as belonging to two categories: on (C_1) or off (C_2) the cast shadow direction. We model the posterior distribution of a foreground or shadow sample belonging to the first category, $P(C_1|x, FS)$, as $f(CD)$, and $P(C_2|x, FS)$ as $1-f(CD)$. Under these considerations,

$$P(S|x, FS) = \sum_{i=1,2} P(S|x, C_i, FS)P(C_i|x, FS). \quad (19)$$

Since a cast shadow cannot be off the cast shadow direction, we have $P(S|x, C_2, FS) = 0$.

Next, we take into account the spatial gradient correlation. We also divide this last parameter in two subsets: spatial gradients from samples are either correlated (∇_c) or not (∇_{nc}). We can then write

$$P(S|x, C_1, FS) = \sum_{\nabla_c \nabla_{nc}} P(S|x, C_1, FS, \nabla)P(\nabla|x, C_1, FS). \quad (20)$$

Here, $\beta(x)$ is a sufficient statistic for x since x is on the cast shadow direction. Therefore we have $P(S|x, C_1, FS, \nabla_c) = P(S|\beta(x), C_1, FS, \nabla_c)$ which is simply what we called $p(S|\beta)$ in equation 17. We use $\omega_{\nabla i}$ for $P(\nabla_c|x, C_1, FS)$. To take into account the possibility that we could observe uncorrelated spatial gradients for cast shadow samples (e.g., due to noise), we give a small weight to $P(S|x, C_1, FS, \nabla_{nc})$ by setting it to $\epsilon\, p(S|\beta)$ $(\epsilon \approx 0.2)$, and we use $1-\omega_{\nabla i}$ for $P(\nabla_{nc}|x, C_1, FS)$.

5.2. Foreground Posterior, $P(F|x)$

Similarly, we have

$$P(F|x) = p(F|x, FS)P(FS|x) \quad (21)$$

and the computation of $P(F|x, FS)$ follows that of $P(S|x, FS)$. First, we have $P(F|x, C_2, FS) = 1$ since a sample not on the cast shadow direction is a foreground sample. Second, $P(F|x, C_1, FS, \nabla_{nc})$ is given by $p(F|\beta)$, and we also set $P(F|x, C_1, FS, \nabla_c) = \epsilon\ p(F|\beta)$ to take into account the possibility that we will observe correlated spatial gradients for foreground (not cast shadow) samples.

6. Results and discussion

The results presented here are obtained from challenging video sequences known in the literature. The physical model proposed in this paper is compared to other statistical models cited earlier when results are available. Since these models have greater accuracy when activity is present in the scene, the sequences chosen are relatively long and shadows are cast by many foreground objects. To demonstrate the validity of our model and parameter estimates, we only show results on frames from the training sequences. All sequences, posteriors of background, foreground and shadow obtained by our approach and hand-segmented ground truth are available at `http://vision.gel.ulaval.ca/en/Projects/Id_283/projet.php`.

6.1. Highway 1

The first sequence shows a highway (Fig. 8(a)) where the vast majority of car colors are shades of gray. Therefore, these cars respect the cast shadow linear relation, inducing a large number of false positives and dramatically affecting the performance of the shadow subtraction algorithm based on this criterion alone. The background posterior obtained from the background subtraction algorithm is shown in Fig. 8(b), while foreground and shadow posteriors obtained by the proposed approach are given in Fig. 8(c,d). Posteriors are the optimal way to present qualitative results since their robustness to different thresholds can also be evaluated by the intensity difference between the two categories.

6.2. Hallway

This sequence was shot in a busy hallway where people are walking or standing still. The scene shows cast shadows, specular reflections on the floor, and highlights. As shown in Fig.9(a), the large number of light sources induces large penumbra regions, and the surface illumination vary widely under cast shadows. Illumination attenuation likelihoods shown in Fig. 7 represent a floor pixel (red circle). As can be seen, both underlying $PDFs$ are non-Gaussian. Parametric models with a finite number of modes, such as the $GMSM$ [3], are therefore less suited for these conditions. Quantitative results show that our approach outperforms the $GMSM$. Note that a threshold on the maximal illumination attenuation value β_{max} was set to 0.5, similarly to the $GMSM$. Figures 9(c)(d) show the shadow and foreground posteriors computed with our approach.

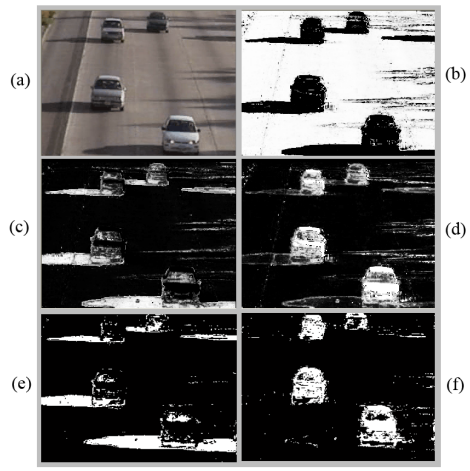

Figure 8. (a) Frame from a busy highway,(b) posterior values for background $P(B|x)$, (c) cast shadow posterior $P(S|x)$, (d) foreground posterior $P(F|x)$, (e) and (f) binary results obtained with $P(S|x) > 0.5$.

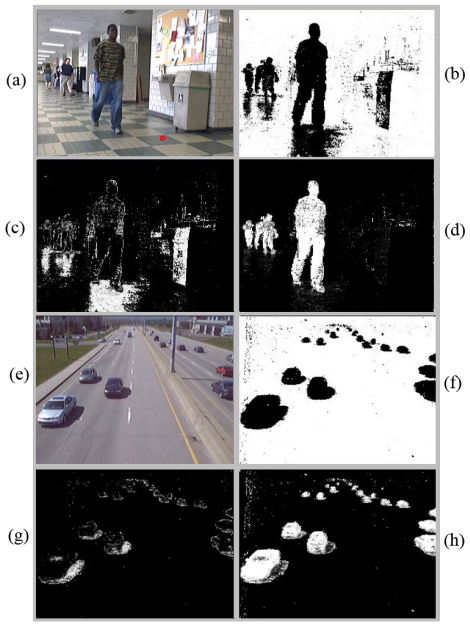

Figure 9. (b)(c)(d) Background, cast shadow and foreground posteriors for the hallway sequence. (e)(f)(g) Background, cast shadow and foreground posteriors for the Highway II sequence.

6.3. Highway II

The last scene shows a highway where there is typically a steady stream of vehicles (Fig. 9(e)). As seen in Fig. 5(b), cast shadows induced a significant color shift, therefore breaking the shadow linear approximation. Since this

color shift is modeled by our approach, we generate posterior distributions that are faithful to the scene (Fig. 9(g)(h)).

6.4. Quantitative results

This evaluation follows [15] where the global performance is given by two metrics. The shadow detection rate SR is related to the percentage of shadow pixels incorrectly labeled as foreground. As for the shadow discrimination rate SD, it is related to both incorrectly labeled foreground and shadow pixels. The reader should refer to [15] for exact equations. Quantitative results were obtained by thresholding the cast shadow posterior and show that our model performs better than a parametric approach based on Gaussian mixtures. Note that results for the $GMSM$ [3] and the GMM using local and global features on the sequence Highway I have been obtained directly from [4].

Method	Highway I		Hallway		Highway II	
	SR	*SD*	*SR*	*SD*	*SR*	*SD*
Physical	0.705	0.844	0.724	0.867	0.684	0.712
GMM LGf	0.721	0.797	-	-	-	-
GMSM	0.633	0.713	0.605	0.870	0.5851	0.444

Table 1. Results for different approaches

7. Conclusion

Qualitative and quantitative results presented in this paper validate the model we have introduced based on the physical properties of the light sources and surface behavior. The results also show that we can successfully learn in a completely unsupervised environment the parameters of this model, i.e., the cast shadow direction and the illumination attenuation with respect to a background sample, under both shadow and foreground hypothesis. When combined with a simple measure of spatial gradient correlation, it then becomes possible to differentiate foreground and moving cast shadow values with similar chroma. We should stress that the results shown in this paper are pixel-based and that integrating both posterior functions in segmentation algorithms using spatial and temporal coherence will yield impressive results. Finally, the descriptive model we have introduced is a first step toward a more elaborate parametrization of a scene from a video sequence where light sources, surface reflectance models and 3D geometry are estimated for low-level segmentation.

Acknowledgment

The authors would like to thank PRECARN and NSERC for their financial support, and to thank the reviewers for their comments which helped improve this paper.

References

[1] R. Cucchiara et al., "Improving shadow suppression in moving object detection with HSV color information", *Proc of Intel. Transportation Systems Conf.*, pp. 334-339, 2001.

[2] T. Horprasert et al., "A statistical approach for real-time robust background subtraction and shadow detection", *IEEE ICCV FRAME-RATE Workshop*, 1999.

[3] N. Martel-Brisson and A. Zaccarin, " Learning and Removing Cast Shadows through a Multidistribution Approach", *IEEE Trans. PAMI* vol. 29, no. 7, July 2007.

[4] Zhou Liu et al., " Cast Shadow Removal Combining Local and Global Features", *IEEE Conf. on CVPR*, 17-22 June 2007

[5] J.-M. Pinel and H. Nicolas, "Shadows analysis and synthesis in natural video sequences", *IEEE Proceedings of ICIP* , pp.285-288, vol.3, 24-28 June 2002.

[6] J. Stauder et al., "Detection of moving cast shadows for object segmentation", *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 65-76, Mar. 1999.

[7] W. Zhang et al., "Moving Cast Shadows Detection Using Ratio Edge", *IEEE Trans. on multimedia*, vol. 9, no. 6, pp. 1202-1214, Oct. 2007.

[8] S. Nadimi and B. Bhanu, "Physical models for moving shadow and object detection in video", *IEEE Trans. PAMI* vol. 26, no. 8, August 2004.

[9] G.D. Finlayson et al., "On the removal of shadows from images", *IEEE Trans. PAMI* vol. 28, no. 1, pp. 59-68, January 2006.

[10] M.P. Wand and M.C. Jones. "Kernel Smoothing."*Monographs on Statistics an Applied Probability*, Chapman and Hall, 1995.

[11] A. Elgammal et al., "Non-parametric Model for Background Subtraction", *6th European Conference on Computer Vision 2000*, pp. 751-767, June 2000.

[12] R. O'Callaghan and T. Haga, "Robust change-detection by normalized gradient-correlation," *IEEE Conf. on CVPR*, 17-22 June 2007.

[13] E. Salvador et al., "Cast shadow segmentation using invariant color features", *Computer Vision and Image Understanding*, pp. 238-259, 2004.

[14] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real- time tracking", *IEEE Trans. PAMI*, vol. 22, no. 8, pp 747-757, Aug. 2000.

[15] A. Prati et al., "Detecting moving shadows: algorithms and evaluation", *IEEE Trans. PAMI* vol. 25 no.7, pp. 918-923, 2003.