

Chapter 1

Shadow Detection in Digital Images and Videos

CSABA BENEDEK^{1,2} AND TAMÁS SZIRÁNYI¹

¹Distributed Events Analysis Research Group, Computer and Automation Research Institute

H-1111, Kende utca 13-17, Budapest, Hungary

Email: bcsaba@sztaki.hu, sziranyi@sztaki.hu

²Faculty of Information Technology, Péter Pázmány Catholic University

H-1083, Práter utca 50/A Budapest, Hungary

1.1 Introduction: Goals and Applications

Shadow detection is an important preprocessing task and hot topic of computer vision. We find numerous applications which need to address shadows, moreover the motivations of the investigations are also highly various. On one hand in *video surveillance* [1, 2], *aerial exploitation* [3], or *traffic monitoring* [4] shadows are usually mentioned as harmful effects, because they make difficult to separate and track the moving objects via background subtraction (Fig. 1.1). In *remote sensing*, shadows may corrupt change detection techniques [5] resulting in false differences. As well in *scene reconstruction* it is a fundamental problem to distinguish surfaces-edges from illumination differences [6], or shadow-free images should be generated for purely visual purposes [6].

On the other hand, shadows may be helpful phenomena in many situations. *Shape from shading* [8] methods derive the 3-D parameters of the objects based on the estimated shadowing effects. Shadows also provide general descriptors for the illumination conditions in the scenes, which can be used for image and



Figure 1.1: Results of background subtraction with the Stauffer-Grimson algorithm [7]. Object silhouettes are strongly corrupted, and multiple moving objects cannot be separated due to cast shadows.



Figure 1.2: Built-in area extraction using cast shadows. Left: input image, middle: output of a color based shadow filter (red areas are detected as shadows), right: built-in areas identified as neighbouring image regions of the shadows blobs considering the sun direction.

video indexing or event analysis [9]. For example, the darkness of shadow informs us whether an outdoor shot was taken in sunlit, or in overcast weather; meanwhile the size and orientation of the shadow blobs are related to the time and date of capturing the frame. If multiple shadows are observable with different darkness, we can expect several light sources in the scene. Object extraction in still images can be also facilitated by shadow detection. In aerial image analysis, it is often necessary to detect static scene objects, such as buildings [10, 11] or trees [12], which are in general challenging pattern recognition problems. However, if a noisy shadow map can be obtained by e.g. filtering pixels from the *dark-blue* color domain [11], the object candidate regions can be estimated as image areas lying next to the shadow blobs in the sun direction as demonstrated in Fig. 1.2.

As the large variety of applications shows, shadow detection is a wide concept: different classes of approaches should be separated depending on the environmental conditions and the exact goals of the systems. In this chapter, we have selected the video surveillance problem to demonstrate a few appearing challenges and proposed solutions related to shadow detection. In surveillance video streams, foreground areas usually contain the regions of interest, moreover, an accurate object-silhouette mask can directly provide useful in-

formation for several applications, for example people [13, 14, 15] or vehicle detection [4], tracking [16, 17], biometrical identification through gait recognition [18, 19] or activity analysis [7]. However, moving cast shadows on the background make it difficult to estimate shape [20] or behavior [14] of moving objects, because they can be erroneously classified as part of the foreground mask. Considering that under some illumination conditions more than half of the non-background image areas may belong to cast shadows, their filtering has a crucial role in scene analysis.

In the following part of this chapter, we will use a few assumptions for the scene and the input data. First, the camera is fixed and has no significant ego-motion. We expect static background objects (for example there is no waving river or flickering object in the background), therefore all the motions are caused either by moving objects or by shadows. We also use a topically valid image in each moment, which can be obtained by the conventional mixture of Gaussians method [7]. There is one emissive light source in the scene (the sun or an artificial source), but we consider the presence of additional effects (e.g. reflection), which may change the spectrum of illumination locally. We also assume that the estimated background values of the pixels correspond usually to the illuminated surface points.

On the other hand, we consider several properties of real situations. The background may change in time: due to varying lightning conditions, and changes of the static objects. ‘Crowded’ and ‘empty’ may scenarios alternate and we must expect background or shadow colored object parts. Due to the daily changes of the sun position and weather, the shadow properties may strongly alter as well.

1.2 Shadow Detection in Video Surveillance: an Overview

The shadow filtering problem has been handled in various ways in the literature. *Geometry based* approaches estimate the spatial transform between the objects and their cast shadows in the projected image plane [21, 22]. However, these methods are highly restricted to specific conditions and object types, therefore *color filtering* techniques are more widespread in practise. We can also distinguish methods working on single images or sequences. Still image based methods [6, 23] – which attempt to find and remove shadows in the single frames independently – are usually applied for high quality photos where the background has a uniform color or texture pattern. On the other hand, they are less efficient in video surveillance, where we must expect images with poor quality and resolution [6], and the computational complexity should be kept low for real time operation [23]. We find a few approaches which focus on the discrimination of shadow edges, and edges due to objects boundaries [24, 25]. However, it may be difficult to extract connected foreground regions from a ragged edge map of a noisy video frame [24]. Complex scenarios containing several small objects or shadow-parts may be also disadvantageous for these methods.

Considering the above reasons, we focus on video (instead of still image) and region (instead of edge)

based shadow modeling techniques in this chapter. Here the next important categorization issue deals with the description of the shadow-background color transform, which can be *non parametric* and *parametric* [26]. Non parametric techniques are often referred as ‘shadow invariant’ approaches, since instead of detecting the shadows they remove them by converting the pixel values into an illuminant invariant feature space. Usually a conventional color space transformation is applied to fulfill this task: the normalized rgb (or rg) [27, 28] and $C_1C_2C_3$ spaces [29] purely contain chrominance color components which are less dependent on luminance. Similar constancy of the hue channel in the HSV space is exploited in [30]. However, as [29] points out, illumination invariant approaches have several limitations regarding the reflecting surfaces and the lighting conditions of the scenes. Outdoors, shadows will have a blue color cast (due to the sky), while lit regions have a yellow cast (sunlight), hence the chrominance color values corresponding to the same surface point may be significantly different in shadow and in sunlight [25]. We have also found in our experiments that the shadow invariant methods fail outdoors several times, and they are rather usable indoors (see later in Fig. 1.7). Moreover, since they ignore the luminance components of the color, these models become sensitive to noise.

Consequently, we keep our investigations among the *parametric* models. First, we estimate the mean background values of the individual pixels through a statistical background model [7], then we extract feature vectors from the actual and the estimated background values of the pixels and model the feature domain of shadows in a probabilistic way. Parametric shadow models may be *local* or *global*.

In a *local* shadow model [31] independent shadow processes are proposed for each pixel. The local shadow parameters are trained using a second mixture model similarly to the background in [7]. This way, the differences in the light absorption-reflection properties of the scene points can be highly considered. However, each pixel should be shadowed several times till its estimated parameters converge under unchanged illumination conditions, which hypothesis is often not satisfied in surveillance videos.

In Section 1.3 we introduce a novel statistical shadow model which follows another approach: shadow is characterized with *global* parameters in an image, and the model describes the relation of the corresponding background and shadow color values. We consider this transformation as a random transformation affected by a perturbation, hence, we take several illumination artifacts into consideration. On the other hand, we derive the shadow parameters from global image statistics, therefore, the model performance is reasonable also on image regions where motion is rare.

Color space choice is a key issue in several corresponding methods, as it will be intensively studied in Section 1.4. For our initial model in Section 1.3 we will propose the CIE $L^*u^*v^*$ space, for reasons which will be detailed later. Here, we only mark two well known properties of the CIE $L^*u^*v^*$ space: we can measure the perceptual distance between colors with the Euclidean distance [32], and the color components are approximately uncorrelated with respect to camera noise and changes in illumination [33]. Since we

Table 1.1: Comparison of different corresponding methods and the proposed model.

Method	Shadow detection	Adaptive	Scenes
Mikic 2000 [35]	global, constant ratio	No	outdoor
Paragious 2001 [27]	illumination invariant	No	indoor
Salvador 2004 [29]	illumination invariant	No	both
Martel-Brisson 2005 [31]	local process	Yes	indoor
Sheikh 2005 [36]	No	-	both
Wang 2006 [37]	global, constant ratio	No	indoor
Benedek 2008 [1]	global, probabilistic	Yes	both

derive the model parameters in a statistical way, there is no need for accurate color calibration and we use the common CIE D65 standard. It is also not critical to consider exactly the physical meaning of the color components which is usually environment-dependent [29, 34]; we use only an approximate interpretation of the L, u, v components and show the validity of the model via experiments.

In Section 1.4, we will give a detailed qualitative and quantitative study on the color space selection problem of shadow detection. In this sense, this section can be considered both as premise and generalization of Section 1.3. Our previous choice for using the CIE $L^*u^*v^*$ space will be justified here, but on the other hand, experiments will refer to the previously introduced model elements, extending their validity to various color models. The reason for dedicating an independent part to this issue is that statistical feature modeling and color space analysis are two different and in themselves composite aspects of shadow detection. Although interaction between the two approaches will be emphasized several times, we hope that the separate discussion will help the clarity of presentation. Due to the various experiments the consequences of Section 1.4 may be more generally usable than in the context of the proposed statistical model framework.

For validation we use real surveillance video shots and also test sequences from a well-known benchmark set [26]. Table 1.1 summarizes the different goals and tools regarding some of the above mentioned state-of-the-art methods and the proposed model. For detailed comparison see also Section 1.3.6.

1.3 A Bayesian Approach for Modeling Shadows in Video Scenes

We solve the shadow detection problem in a Bayesian image segmentation framework, which separates foreground, background and shadow regions in the video frames. Denote by S the two dimensional pixel grid and we use a first ordered neighborhood system on S . The procedure assigns a label $\omega(s)$ to each pixel $s \in S$ form the label-set: $\Phi = \{fg, bg, sh\}$ corresponding to three possible classes: foreground (fg), background (bg) and shadow (sh). The label field is modeled by a Markov Random Field [38]: the segmentation is equivalent to a global labeling $\underline{\omega} = \{[s, \omega(s)] \mid s \in S\}$, and the probability of a given $\underline{\omega} \in \Omega$ follows

Gibbs distribution [38].

The observation at pixel s is the three dimensional color vector – at this point – in the CIE $L^*u^*v^*$ space: $\bar{o}(s) = [o_L(s), o_u(s), o_v(s)]^T$. Set $\mathcal{O} = \{\bar{o}(s) | s \in S\}$ refers to the global image data. The key point in the model is to define the conditional density functions $p_\phi(s) = P(\bar{o}(s) | \omega(s) = \phi)$, for all $\phi \in \Phi$ and $s \in S$. E.g. $p_{bg}(s)$ is the probability that the background process generates the observed feature value $\bar{o}(s)$ at s .

We do not address foreground modeling in this book chapter, since it has an exhaustive literature in itself. The simplest approach is using uniform foreground distributions $p_{fg} = u$ [35] which is equivalent to outlier detection. There have been also proposed more sophisticated models based on temporal foreground descriptions [36] or pixels state transition probabilities [37]. In our upcoming model we will use the spatial foreground calculus introduced in [1], which is not sensitive to the frame rate parameter of the video streams ensuring robustness in surveillance environments.

We define the background's and shadow's conditional density functions in Section 1.3.1-1.3.4, and the segmentation procedure will be presented in Section 1.3.6 in details. Before continuing, note that we minimize the minus-logarithm of the global probability term in fact. Therefore, in the following we use the $\epsilon_\phi(s) = -\log p_\phi(s)$ local energy terms, for easier notation.

1.3.1 General Probabilistic Models of the Background and Shadow Processes

We model the distribution of feature values in the background and in the shadow by Gaussian density functions, similarly to [26, 37, 39]. For simplicity, we approximate the joint distribution of the color components by a three dimensional Gaussian density function with diagonal covariance matrix: $\bar{\Sigma}_k(s) = \text{diag}\{\sigma_{k,L}^2(s), \sigma_{k,u}^2(s), \sigma_{k,v}^2(s)\}$ for $k \in \{bg, sh\}$. Accordingly, the distribution parameters are the $\bar{\mu}_k(s) = [\mu_{k,L}(s), \mu_{k,u}(s), \mu_{k,v}(s)]^T$ mean vector, and $\bar{\sigma}_k(s) = [\sigma_{k,L}(s), \sigma_{k,u}(s), \sigma_{k,v}(s)]^T$ standard deviation vectors. With this 'diagonal' model we avoid matrix inversion and determinant recovering during the calculation of the probabilities, and the $\epsilon_k(s) = -\log p_k(s)$ terms can be derived directly from the one dimensional marginal probabilities:

$$\epsilon_k(s) = C + \sum_{i=\{L,u,v\}} \log \sigma_{k,i}(s) + \frac{1}{2} \left(\frac{o_i(s) - \mu_{k,i}(s)}{\sigma_{k,i}(s)} \right)^2 \quad (1.1)$$

with $C = 2 \log 2\pi$. According to eq. (1.1), each feature contributes with its own additional term to the energy calculus. Therefore, the model is modular: the one dimensional model parameters, $[\mu_{k,i}(s), \sigma_{k,i}^2(s)]$, can be estimated separately.

The use of a Gaussian distribution to model the observed color of a single background pixel is well established in the literature, with the corresponding parameter estimation procedures such as in [7, 40]. We train the color components of the background parameters $[\bar{\mu}_{bg}(s), \bar{\sigma}_{bg}(s)]$ in a similar manner to the

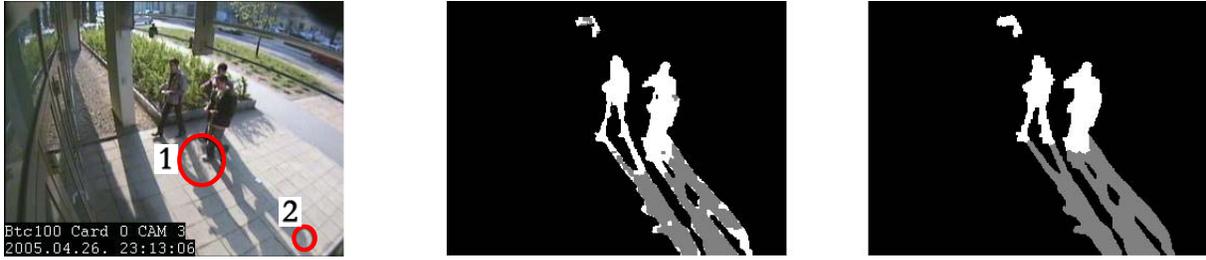


Figure 1.3: Illustration of two illumination artifacts (the frame has been chosen from the ‘Entrance am’ test sequence). 1: dark shadow part between the legs (more object parts change the reflected light). 2: penumbra artifact near the edge of the shadow. The constant ratio model (in the middle) causes errors, while the proposed model (right image) is more robust.

conventional online k-means algorithm [7]. $[\mu_{bg,L}(s), \mu_{bg,u}(s), \mu_{bg,v}(s)]^T$ vector estimates the mean background color of pixel s measured over the recent frames, while $\bar{\sigma}_{bg}(s)$ is an adaptive noise parameter. An efficient outlier filtering technique [7] excludes most of the non-background pixel values from the parameter estimation process, which works without user interaction.

As we have stated in Section 1.2, we characterize shadows by describing the background-shadow color value transformation in the images. The shadow calculus is based on the illumination-reflection model [41] introduced in Section 1.3.2. This model assumes constant lighting, flat and Lambertian reflecting surfaces, however our scene does not usually fulfill these requirements. As a novelty we use a probabilistic approach in Section 1.3.3 to describe the deviation of the scene from the ideal surface assumptions for getting a more robust shadow detection .

1.3.2 Shadow Description by Lambertian Color Features

According to the illumination model [41] the response $g(s)$ of a given image sensor placed at pixel s can be written as

$$g(s) = \int e(\lambda, s)\rho(\lambda, s)\nu(\lambda)d\lambda \quad (1.2)$$

where $e(\lambda, s)$ is the illumination function at a given wavelength λ , $\rho(s)$ depends on the surface albedo and geometric, $\nu(\lambda)$ is the sensor sensitivity. Accordingly, the difference between the shadowed and illuminated background values of a given surface point is caused only by the different local value of $e(\lambda, s)$. Outdoors, the illumination function observed in sunlit is the composition of the direct component (sun), the Rayleigh scattering (sky), causing that the ambient light has a blue tinge [42], and residual light components reflected from other objects. On the other hand, the effect of the direct component is missing in the shadow.

Although the validity of eq. (1.2) is already limited by several scene assumptions [41], in general, it is still too difficult to exploit appropriate information about the corresponding background-shadow values,

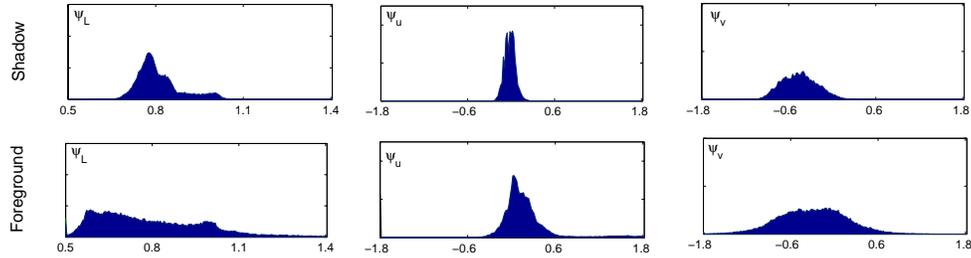


Figure 1.4: Histograms of the ψ_L , ψ_u and ψ_v values for shadowed and foreground points collected over a 100-frame period of the video sequence ‘Entrance pm’ (frame rate: 1 fps). Each column corresponds to a color component.

since the components of the illumination function are unknown. Therefore, further strong simplifications are used in the applications. According to [6] the camera sensors must be exact Dirac delta functions: $\nu(\lambda) = q_0 \cdot \delta(\lambda - \lambda_0)$ and the illumination must be Planckian [43]. In this case, eq.(1.2) implies the well-known ‘constant constant ratio’ rule. Namely, the ratio of the shadowed $g_{sh}(s)$ and illuminated value $g_{bg}(s)$ of a given surface point is considered to be constant over the image: $\frac{g_{sh}(s)}{g_{bg}(s)} = A$.

The ‘constant ratio’ rule has been used in several applications [35, 37, 39]. Here the shadow and background Gaussian terms corresponding to the same pixel are related via a globally constant linear density transform. In this way, the results may be reasonable when all the direct, diffused and reflected light can be considered constant over the scene. However, the reflected light may vary over the image in case of several static or moving objects, and the reflecting properties of the surfaces may differ significantly from the Lambertian model (See Fig. 1.3). The efficiency of the constant ratio model is also restricted by several practical reasons, like quantification errors of the sensor values, saturation of the sensors, imprecise estimation of $g_{bg}(s)$ and A , or video compression artifacts. Based on our experiments (Section 1.3.6), these inaccuracies cause poor detection rates in some outdoor scenes.

1.3.3 Proposed Shadow Model

The previous section suggests that the ratio of the shadowed and background luminance values of the pixels may be useful, but not powerful enough as a descriptor of the shadow process. Instead of constructing a more difficult illumination model, for example in 3D with two cameras, we overcome the problems with a statistical model. For each pixel s , we introduce the variable $\psi_L(s)$ by:

$$\psi_L(s) = \frac{o_L(s)}{\mu_{bg,L}(s)} \quad (1.3)$$

where, as defined earlier, $o_L(s)$ is the observed luminance value at s , and $\mu_{bg,L}(s)$ is the mean value of the local Gaussian background term estimated over the previous frames [7].

Thus, if the $\psi_L(s)$ value is close to the estimated shadow darkening factor, s is more likely to be a shadowed point. More precisely, in a given video sequence, we can estimate the distribution of the shadowed ψ_L values globally in the video parts. Based on experiments with manually generated shadow masks, a Gaussian approximation seems to be reasonable regarding the distribution of shadowed ψ_L values (Fig. 1.4 shows the global ψ statistics regarding a 100-frame period of outdoor test sequence ‘Entrance pm’). For comparison, we have also plotted the statistics for the foreground points, which follows significantly different, more uniform distribution.

Due to the spectral differences between the direct and ambient illumination, cast shadows may also change the u and v color components [25]. We have found an offset between the shadowed and background u values of the pixels, which can be efficiently modelled by a global Gaussian term in a given scene (similarly as for the v component). Hence, we define $\psi_u(s)$ (and $\psi_v(s)$) by

$$\psi_u(s) = o_u(s) - \mu_{bg,u}(s) \quad (1.4)$$

As Fig. 1.4 shows, the shadowed $\psi_u(s)$ and $\psi_v(s)$ values follow approximately normal distributions.

Consequently, the shadow color process is characterized by a three dimensional Gaussian random variable:

$$\forall s \in S : \bar{\psi}(s) = [\psi_L(s), \psi_u(s), \psi_v(s)]^T \sim N[\bar{\mu}_\psi, \bar{\sigma}_\psi] \quad (1.5)$$

Using eq. (1.3) and (1.4), the color values in the shadow at a given pixel position are also generated by Gaussian distribution,

$$[o_L(s), o_u(s), o_v(s)]^T \sim N[\bar{\mu}_{sh}(s), \bar{\sigma}_{sh}(s)] \quad (1.6)$$

with the following parameters:

$$\mu_{sh,L}(s) = \mu_{\psi,L} \cdot \mu_{bg,L}(s) \quad (1.7)$$

$$\sigma_{sh,L}^2(s) = \sigma_{\psi,L}^2 \cdot \mu_{bg,L}^2(s) \quad (1.8)$$

Regarding the u (and similarly to the v) component:

$$\mu_{sh,u}(s) = \mu_{\psi,u} + \mu_{bg,u}(s), \quad \sigma_{sh,u}^2(s) = \sigma_{\psi,u}^2 \quad (1.9)$$

1.3.4 Parameter Settings

Our method works with scene-dependent and condition-dependent parameters. *Scene-dependent* parameters can be considered constant in a specific field, and are influenced by, e.g. camera settings and prior knowledge about the appearing objects or reflection properties. We provide strategies on how to set these parameters if a surveillance environment is given. *Condition-dependent* parameters vary in time in a scene, therefore, we use adaptive algorithms to follow them. We emphasize that regarding the background and shadow processes, only the one dimensional marginal distribution parameters should be estimated (Section 1.3.1).



Figure 1.5: Different parts of the day on ‘Entrance’ sequence, segmentation results. Above left: in the morning (‘am’), right: at noon, below left: in the afternoon (‘pm’), right: wet weather

The *background* parameter estimation and update procedure is automated, based on the work in [7], which presents reasonable results, and it is computationally more effective than the standard EM algorithm.

The changes in the global illumination significantly alter the *shadow* properties (Fig. 1.5). Moreover, changes can be performed rapidly: indoors due to switch on/off different light sources, while outdoors due to the appearance of clouds. Regarding the shadow parameter settings, we discriminate parameter initialization and re-estimation. From a practical point of view, initialization may be supervised with marking shadowed regions in a few video frames by hand, once after switching on the system. Based on the training data, we can calculate maximum likelihood estimates of the shadow parameters. On the other hand, there is usually no opportunity for continuous user interaction in an automated surveillance environment, thus the system must adapt to the illumination changes raising a claim to an automatic re-estimation procedure. Therefore we use supervised initialization, and focus on the parameter adaption process in the following. The presented method is built in a 24-hour surveillance system of our university campus (Fig. 1.5).

According to section 1.3.3, the shadow process has six parameters: the 3-3 components of $\bar{\mu}_{\psi}$, respectively $\bar{\sigma}_{\psi}$ vectors. Fig. 1.6(a) shows the one-dimensional histograms for the occurring ψ_L , ψ_u and ψ_v values of shadowed points for each video shot. We can observe that while the variation of parameters $\bar{\sigma}_{\psi}$, $\mu_{\psi,u}$ and $\mu_{\psi,v}$ are low, $\mu_{\psi,L}$ varies in time significantly. Therefore, we update the parameters in two different ways.

Re-estimation of the Chrominance Parameters

The update procedure of parameters $[\mu_{\psi,u}, \sigma_{\psi,u}]$ and $[\mu_{\psi,v}, \sigma_{\psi,v}]$ is similar to that was used in [44]. We show it regarding the u component only, since the v component is updated in the same way.

We re-estimate the parameters at fixed time-intervals \mathcal{T} . Denote $\mu_{\psi,u}[t], \sigma_{\psi,u}[t]$ the parameters at time t . W_{t_2} is the set containing the observed ψ_u values collected over the pixels detected as shadow between

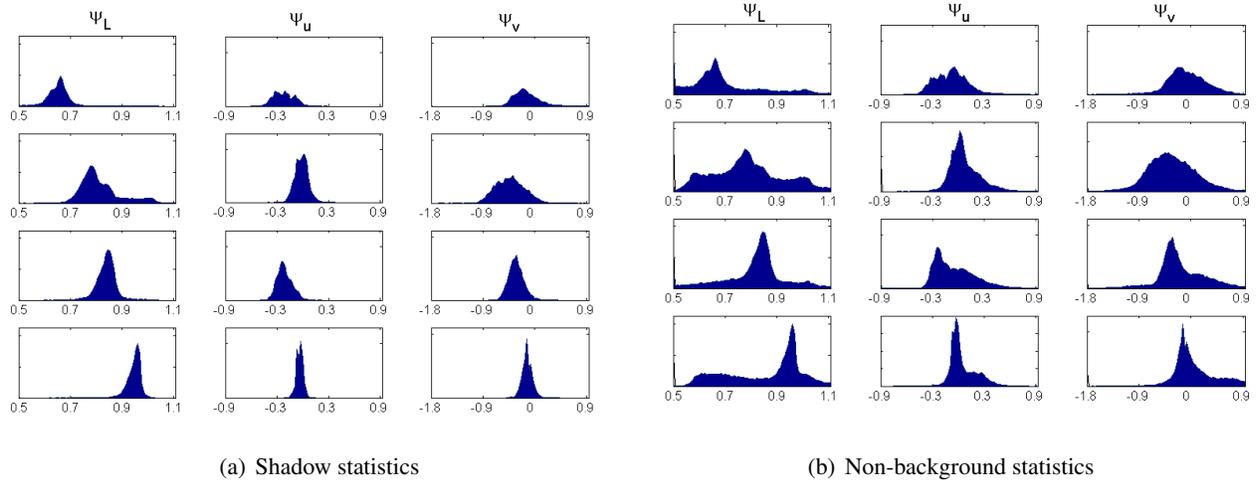


Figure 1.6: (a) Shadow $\bar{\psi}$ statistics on four sequences recorded by the ‘Entrance’ camera of our University campus. Histograms of the occurring ψ_L , ψ_u and ψ_v values of shadowed points. Rows correspond to video shots from different parts of the day. We can observe that the peak of the ψ_L histogram strongly depends on the illumination conditions, while the change in the other two shadow parameters is much smaller. (b) $\bar{\psi}$ statistics for all non-background pixels Histograms of the occurring ψ_L , ψ_u and ψ_v values of all the non-background pixels in the same sequences

time $t_1 = t_2 - \mathcal{T}$ and t_2 :

$$W_{t_2} = \{\psi_u^{[t]}(s) | t = t_1, \dots, t_2 - 1, \omega^{[t]}(s) = \text{sh}, s \in S\} \quad (1.10)$$

where upper index $[t]$ refers to time, $\#W_{t_2}$ is the number of the elements, M_{t_2} and D_{t_2} are the empirical mean and the standard deviation values of W_{t_2} . We update the parameters as:

$$\mu_{\psi,u}[t_2] = (1 - \xi^{[t_2]}) \cdot \mu_{\psi,u}[t_1] + \xi^{[t_2]} \cdot M_{t_2} \quad (1.11)$$

$$\sigma_{\psi,u}^2[t_2] = (1 - \xi^{[t_2]}) \cdot \sigma_{\psi,u}^2[t_1] + \xi^{[t_2]} \cdot D_{t_2}^2 \quad (1.12)$$

Parameter $\xi^{[t]}$ is a weighting term ($0 \leq \xi^{[t]} \leq 1$) depending on $\#W_t$, namely greater number of detected shadow points increase $\xi^{[t]}$ and the influence of the M_t respectively D_t^2 term. We use $\mathcal{T} = 60$ sec.

Re-estimation of the Luminance Parameters

Parameter $\mu_{\psi,L}$ corresponds to the average background luminance darkening factor of the shadow. Except from window-less rooms with constant lightning, $\mu_{\psi,L}$ is strongly condition dependent. Outdoors, it can vary between 0.6 in direct sunlight and 0.95 in overcast weather. The simple re-estimation from the previous section does not work in this case, since the illumination properties between time t and $t + \mathcal{T}$ may rapidly change a lot, which would result in absolutely false detected shadow values in set W_t presenting false M_t and D_t parameters for the re-estimation procedure.

For this reason, we gain the actual $\mu_{\psi,L}$ from the statistics of all non-background ψ_L -s (where the background filtering should be done by a good approximation only, we use the Stauffer-Grimson algorithm). In Fig. 1.6(b) we can observe that the peaks of the ‘non-background’ ψ_L -histograms are approximately in the same location as they were in Fig. 1.6(a). The videos of the first and second rows were recorded around noon where the shadows were relatively small, but the peak is still in the right location.

The previous experiments encourage us to identify $\mu_{\psi,L}$ with the location of the peak on the ‘non-background’ ψ_L -histograms for the scene. The description of the update-algorithm of $\mu_{\psi,L}$ is as follows. We define a data structure which contains a ψ_L value with its timestamp: $[\psi_L, t]$. We store the ‘latest’ occurring $[\psi_L, t]$ pairs of the non-background points in a set \mathcal{Q} , and update the histogram h_L of the ψ_L values in \mathcal{Q} continuously. The key point is the management of set \mathcal{Q} . We define MAX and MIN parameters which controls the size of \mathcal{Q} . The queue management algorithm has the following steps:

Algorithm for updating the $\mu_{\psi,L}$ shadow parameter

1. For each frame t we determine $\Psi_t = \{[\psi_L^{[t]}(s), t] \mid s \in S, \omega^{[t]}(s) \neq \text{bg}\}$
2. We append Ψ_t to \mathcal{Q} .
3. We may remove elements from \mathcal{Q} :
 - if $\#\mathcal{Q} < \text{MIN}$, we keep all the elements.
 - if $\#\mathcal{Q} \geq \text{MIN}$ we find the eldest timestamp t_e in \mathcal{Q} and remove all the elements from \mathcal{Q} with time stamp t_e .
4. If $\#\mathcal{Q} > \text{MAX}$ after step 3: in order of their timestamp we remove further (‘old’) elements from $\#\mathcal{Q}$ till we reach $\#\mathcal{Q} \leq \text{MAX}$.
5. We update the histogram h_L regarding \mathcal{Q} and apply: $\mu_{\psi,L}^{[t+1]} = \text{argmax}\{h_L\}$

Consequently, \mathcal{Q} contains always the latest available ψ_L values. The algorithm keeps the size of \mathcal{Q} between prescribed bounds MAX and MIN ensuring the topicality and relevancy of the data contained. The actual size of \mathcal{Q} is around MAX in case of cluttered scenarios. In the case of few or no motion in the scene, the size of \mathcal{Q} decreases until MIN: this fact increases the influence of the forthcoming elements, and causes quicker adaptation, since it is faster to modify the shape of a smaller histogram.

Parameter $\sigma_{\psi,L}$ is updated similarly to $\sigma_{\psi,u}$ but only in the time periods when $\mu_{\psi,L}$ does not change significantly. Note that the above update process may fail in shadow free scenarios. However, that case occurs mostly under artificial illumination conditions, where the shadow detector can be switched off.

Table 1.2: Comparing the processing speed of our proposed model to three recent reference methods (using the published frame-rates). Note that [31] does not use any spatial smoothing (like MRF), and [36] performs only a two-class separation.

	M-Brisson05 [31]	Sheikh05 [36]	Wang06 [37]	Proposed
Classes	3	2	3	3
MRF Opt	-	Graph cut	ICM	ICM
Frame-rate	10 fps	11 fps	1-2 fps	3 fps

1.3.5 MRF Optimization

The MAP estimator is realized by combining a conditional independent random field of signals and an unconditional Potts model [45]. The optimal segmentation corresponds to the global labeling, $\hat{\omega}$, defined by

$$\hat{\omega} = \arg \min_{\omega \in \Omega} \left\{ \sum_{s \in S} \underbrace{-\log P(\bar{o}(s) | \omega(s))}_{\epsilon_{\omega(s)}(s)} + \sum_{r, s \in S} \Theta(\omega(r), \omega(s)) \right\} \quad (1.13)$$

where the minimum is searched over all the possible segmentations (Ω) of a given input frame. The first part of eq. (1.13) contains the sum of the local class-energy terms regarding the pixels of the image (see eq. (1.1)). The second part is responsible for the smooth segmentation: $\Theta(\omega(r), \omega(s)) = 0$ if s and r are not neighboring pixels, otherwise:

$$\Theta(\omega(r), \omega(s)) = \begin{cases} -\delta & \text{if } \omega(r) = \omega(s) \\ +\delta & \text{if } \omega(r) \neq \omega(s) \end{cases} \quad (1.14)$$

As for optimization, we have found the deterministic Modified Metropolis (MMD) [46] relaxation method similarly efficient but significantly faster for this task than the original stochastic SA algorithm [47]: processing 320×240 images runs with 1 fps using it. If we use ICM [48] with our model, the running speed is 3 fps, in exchange for some degradation in the segmentation results. For comparison, frame-rates of three latest reference methods are shown in Table 1.2. We can observe that our model has approximately the same complexity as [37]. Although the speed of [31] and [36] is notably higher, one should consider that [31] does not use any spatial smoothing (like MRF), thus a separate noise filter must be applied there in the post-processing phase. On the other hand [36] performs only a two-class segmentation (background and foreground). That simplification enables using the quick graph cut based MRF optimization techniques, unlike in the three-class cases [49].

Table 1.3: Overview on the evaluation parameters regarding the five sequences. Notes: *number of frames in the ground truth set. ** *Frame rate of evaluation* (fre): number of frames with ground truth within one second of the video. ***Length of the evaluated video part. †fre was higher in ‘busy’ scenarios.

Video	Frames*	fre**	Duration (min) ***
Laboratory	205	2-4 fre [†]	1:28
Entrance am	160	2 fre	1:20
Entrance pm	75	1 fre	1:15
Entrance noon	251	1 fre	4:21
Highway	170	5-8 fre [†]	0:29

1.3.6 Experimental results

The goal of this section is to demonstrate the benefit of using the novel shadow model introduced in this chapter qualitatively and quantitatively. We have validated our method on several test sequences, here, we show results regarding the following 7 videos:

- ‘Laboratory’ test sequence from the ATON benchmark set [26]. This shot contains a simple environment where previous methods [37] have produced already accurate results.
- ‘Highway’ video (ATON benchmark set). This sequence contains dark shadows but homogenous background without illumination artifacts. Contrast to [35] our method reaches the appropriate results without post processing, which is strongly environment-dependent.
- ‘Corridor’ indoor surveillance video. Although, it is on the face of a simple office environment the bright objects and background elements often saturate the image sensors and it is hard to accurately separate the white shirts of the people from the white walls in the background.
- 4 surveillance video sequences captured by the ‘Entrance’ (outdoor) camera of our university campus in different lightning condition. (See Fig 1.5: ‘Entrance am’, ‘Entrance noon’, ‘Entrance pm’ and ‘Entrance overcast’). These sequences contain difficult illumination and reflection effects and suffer from sensor saturation (dark objects and shadows). Here, the presented model improves the segmentation results significantly versus previous methods.

Results of different shadow detectors are demonstrated in Fig. 1.7. For the sake of comparison we have implemented in the same framework an illumination invariant (‘II’) method based on [29], and a constant ratio model (‘CR’) [35]. We have observed that the results of the previous and the proposed methods are similar in simple environments, but our improvements become significant in the surveillance scenes. In

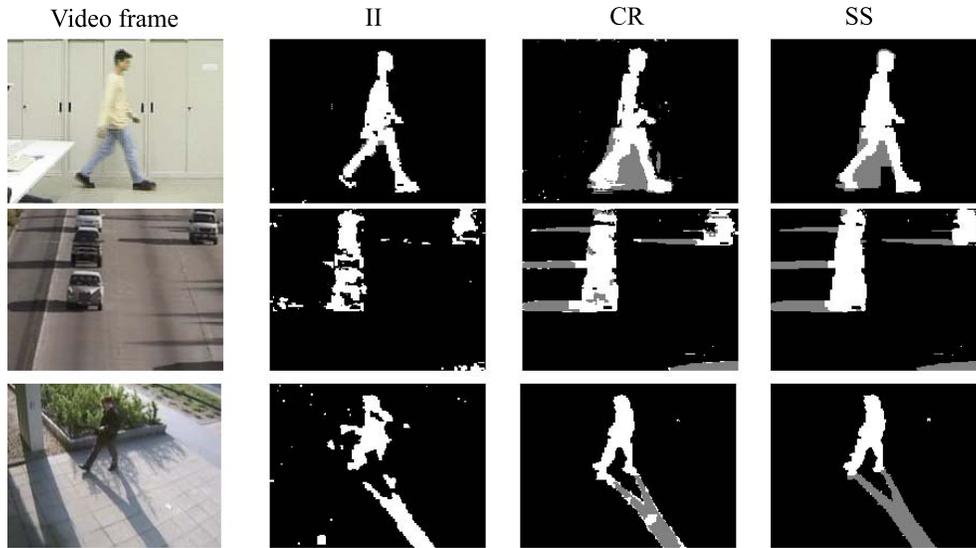


Figure 1.7: *Shadow model validation*: Comparison of different shadow models in 3 video sequences (From above: ‘Laboratory’, ‘Highway’, ‘Entrance am’) . Col. 1: video image, Col. 2: $C_1C_2C_3$ space based illumination invariants [29] (‘II’). Col. 3: ‘constant ratio model’ by [35] (‘CR’, without object-based post-processing) Col. 4: Proposed statistical shadow model (‘SS’)

the ‘*Laboratory*’ sequence, the ‘II’ approach is reasonable, while the ‘CR’ and the proposed method are similarly accurate. Regarding the ‘*Highway*’ video, although the ‘II’ and ‘CR’ find the objects without shadows approximately, but the results are much noisier than it is with our model. Finally on the ‘*Entrance am*’ surveillance video, the ‘II’ method fails completely: shadows are not removed, while the foreground component is also noisy due to the lack of using luminance features in the model. The ‘CR’ model produces poor results also: due to the long shadows and various field objects the constant ratio model becomes inaccurate. Our model handles these artifacts robustly.

The quantitative evaluations are done through manually generated ground truth sequences. Since the application’s goal is foreground detection, the crossover between shadow and background does not count for errors. Denote the number of correctly identified foreground pixels of the evaluation images by TP (*true positive*). Similarly, we introduce FP for misclassified background points, and FN for misclassified foreground points. The evaluation metrics consists of the *Recall* rate and the *Precision* of the detection:

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP} \quad (1.15)$$

Later on, we will also use the *F-measure* (FM) [50] which combines *Recall* (Rc) and *Precision* (Pr) in a single efficiency measure (it is the harmonic mean of Rc and Pr):

$$FM = \frac{2 \cdot Rc \cdot Pr}{Rc + Pr}. \quad (1.16)$$

Table 1.4: Quantitative evaluation results. Comparing the constant ratio (CR) and the proposed statistical (SS) shadow models.

Data Set	Recall R_c		Precision Pr		FM -measure	
	CR	SS	CR	SS	CR	SS
Laboratory	0.950	0.941	0.883	0.929	0.915	0.935
Highway	0.886	0.890	0.644	0.805	0.746	0.845
Entrance am	0.946	0.968	0.596	0.774	0.731	0.861
Entrance noon	0.980	0.963	0.742	0.833	0.845	0.894
Entrance pm	0.972	0.961	0.621	0.830	0.756	0.891

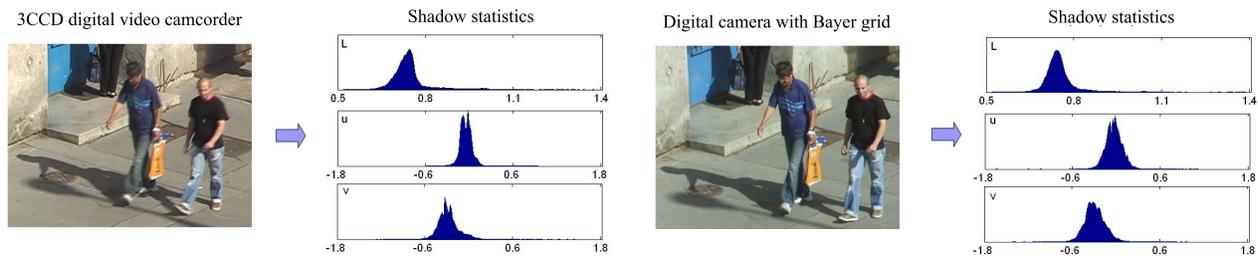


Figure 1.8: Distribution of the shadowed $\bar{\psi}$ values in simultaneous sequences from a street scenario recorded by different CCD cameras. Note: the camera with Bayer grid has higher noise, hence the corresponding u/v components have higher variance parameters.

Note that while R_c and Pr characterize a given algorithm only together¹, FM is in itself an efficient evaluation metrics.

For numerical validation, we used in the aggregate 861 frames chosen from the ‘Laboratory’, ‘Highway’, ‘Entrance am’, ‘Entrance noon’ and ‘Entrance pm’ sequences. Details about the test sets are given in Table 1.3. Table 1.4 shows the detection results of the proposed method compared to the constant ratio (CR) model, confirming that our shadow calculus improves the precision rate, since it decreases the number of false negative shadow pixels significantly, while it preserves the high foreground recall rate at the same time. Consequently, the FM -measure of the proposed model outperforms the CR method one in all sequences.

Influence of CCD Selection on the Shadow Domain

We have introduced a statistical shadow model without any knowledge about the technical details and embedded control/LUT of the different cameras. However, we have performed an additional experiment regarding this issue. We have recorded simultaneous videos from a street scenario with two different cameras: a 3CCD digital video camcorder and a conventional digital camera, which uses a Bayer grid. By examining

¹Consider an algorithm which classifies each pixel as foreground. It is obviously a weak segmenter, but its R_c is equal to 1. However, Pr is low in that case.

Table 1.5: Color space selection in the state-of-the-art methods. [†] In cases of parametric methods, the (average) number of shadow parameters for one color channel. [‡]Proportional to the number of support vectors after training.

Method	Color space	Number of param./ color channels [†]
Cavallaro 2004 [28]	<i>rg</i>	invariant
Salvador 2004 [29]	$C_1C_2C_3$	invariant
Paragios 2001 [27]	<i>rg</i>	invariant
Mikic 2000 [35]	RGB	1
Rittscher 2002[39]	grayscale	2
Wang 2006 [37]	grayscale	2
Cucchiara 2001 [51]	HSV	1.33
Martel-Brisson 2005 [31]	CIE L*u*v*	2
Rautiainen 2001 [52]	CIE L*a*b*/HSV	N.a.
Siala 2004 [53]	RGB	N.a. [‡]
Benedek 2007 [2]	All from above	2

the corresponding shadow domains (see Fig. 1.8), we can observe that the distributions of the shadowed $\bar{\psi}$ values are very similar. However, the higher noise of the Bayer grid camera results in higher variance parameters regarding the *u* and *v* components.

1.4 Color Space Selection in Cast Shadow Detection

In this section we focus on a particular aspect of shadow detection: we illustrate that the performance of segmentation can be significantly improved through appropriate color space selection, if for practical purposes, we should keep the number of free parameters of the method low. We show experimental results regarding the following questions:

- What is the gain of using color images instead of grayscale ones?
- What is the gain of using uncorrelated spaces instead of the standard RGB?
- Are chrominance (illumination invariant), luminance, or ‘mixed’ spaces more efficient?
- In which scenes are the differences significant?

We qualify the metrics both in color based clustering of the individual pixels and in the case of Bayesian foreground-background-shadow segmentation through generalizing the model introduced in Section 1.3.

Table 1.6: Luminance-related and chrominance channels in different color spaces

Color space	gray	rgcolor space!rg	$C_1C_2C_3$	HSV	RGB	$L^*a^*b^*$	$L^*u^*v^*$
luminance ch.	g	-	-	H	R,G,B	L^*	L^*
chrominance ch.	-	r,g	C_1,C_2,C_3	S,V	-	a^*,b^*	u^*,v^*

Experimental results on the test videos show that CIE $L^*u^*v^*$ color space is the most efficient in both cases.

1.4.1 Color Spaces: Significance of the Right Choice

Appropriate color space selection is a crucial step for many image processing problems [25, 54, 55]. Since the shadow model proposed in Section 1.3 is primarily based on describing the shadow’s color domain, issues on color spaces should also be investigated in this case. Although shadow detection is a very well examined problem and some comparative works [26, 56] have also been published in this topic, previous reviews classify and compare the existing methods based on their *model structures*. The authors [26] note that the methods work in different color spaces, like RGB [35] and HSV [51]. However, it remains open-ended, how important is the appropriate color space selection, and which color space is the most effective regarding shadow detection. Moreover, we find also further examples: [39] used only gray levels for shadow segmentation, other approaches were dealing with the CIE $L^*u^*v^*$ [31] and CIE $L^*a^*b^*$ [52] spaces, respectively (an overview is in Table 1.5). Note that an experimental evaluation of color spaces has been already done for shadow edge classification in [25], but in the current book chapter, we address the detection of the shadowed and foreground regions, which is a fairly different problem.

For the above reasons, the main issue of this section is to give an experimental comparison of different color models regarding cast shadow detection on the video frames. Of course, the validity of such experiments is limited to the examined model structures, thus it is important to make the comparison in a relevant framework. Taking a general approach, we consider the task as a classification problem in the space of the extracted features, describing the different cluster domains with relatively few free parameters.²

Models in the literature use usually *deterministic* (per pixel, e.g. [51]) or *statistical* (probabilistic, see [35]) approaches. Up to now, we have only dealt with statistical models, since they proved to be advantageous considering the whole segmentation process. On the contrary, here we introduce a deterministic method first, where the pixels are classified independently before the rate of the correct pixel-classification is investigated. That way, we can perform a relevant quantitative comparison of the different color spaces, since the decision for each pixel depends only on the corresponding local color-feature value; post process-

²Most models in Table 1.5 also contain 2 parameters for each color channel, drawbacks of methods using less parameters have been emphasized in Section 1.3.

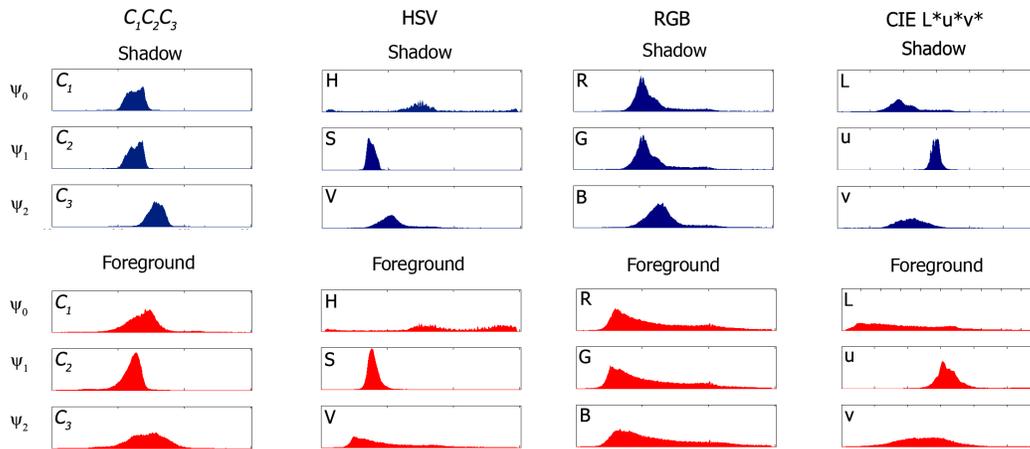


Figure 1.9: One dimensional projection of histograms of shadow (above) and foreground $\bar{\psi}$ values in the ‘Entrance pm’ test sequence.

ing and prior effects whose efficiency may be environment-dependent do not take account here. Thereafter, we give a probabilistic interpretation to this model and we insert it to the MRF framework which was introduced in Section 1.3. We also compare the results after MRF optimization qualitatively and quantitatively.

1.4.2 Generalized Feature Vector of Shadow Separation

Feature extraction is done similarly to Section 1.3, but here we give a generalization of the $\bar{\psi}$ shadow features to handle different color spaces.

We remind the Reader of the constant ratio model introduced in Section 1.3.2, where the ratio of the shadowed and illuminated sensor values have been considered near constant over the images. To handle the different artifacts, one can prescribe a *domain* [53] or a *distribution* (see Section 1.3.3) instead of a single value for the ratios, which results in a powerful detector.

Next, we should examine, how one can use this approach in different color systems. We begin the description with some notes. We assume that the camera presents the frames in the RGB space, and for the different color space conversions, we use the formulas of [57]. The ITU D65 standard is used again for the calibration of the CIE L*u*v* and L*a*b* spaces.

As we did in the CIE L*u*v* based model (page 8) we will separately handle the color components which are directly related to the brightness of the pixels (we refer to them later as ‘luminance’ components), and the remaining ones which correspond to ‘chrominances’ of the observed colors. Classification of channels regarding different color spaces can be found in Table 1.6. In this way, we can also classify the color spaces: since the normalized rg and $C_1C_2C_3$ spaces contain only chrominance components we will call them ‘chrominance spaces’, while grayscale and RGB are purely ‘luminance spaces’. In this terminology,

HSV, CIE L*u*v* and L*a*b* are ‘mixed spaces’.

The shadow descriptor is derived in an analogous manner to the approach of Section 1.3.3: the ‘probabilistic ratio’ method is used for the ‘luminance’ components, while the offsets between the shadowed and illuminated ‘chrominance’ values of the pixels are modeled by a Gaussian additive term. In summary, if the current value of a given pixel in a given color space is $[o_0, o_1, o_2]$ (indices 0, 1, 2 correspond to the different color components), the estimated (illuminated) background value is there $[\mu_{bg,0}, \mu_{bg,1}, \mu_{bg,2}]$, we define the shadow descriptor $\bar{\psi} = [\psi_0, \psi_1, \psi_2]$ by the following, for $i = \{0, 1, 2\}$:

- If i is the index of a ‘luminance’ component:

$$\psi_i(s) = \frac{o_i(s)}{\mu_{bg,i}(s)}. \quad (1.17)$$

- If i is the index of a ‘chrominance’ component:

$$\psi_i(s) = o_i(s) - \mu_{bg,i}(s). \quad (1.18)$$

We define the descriptor in grayscale and in the rg space similarly to eq. (1.17) and (1.18) considering that $\bar{\psi}$ will be a scalar and a two dimensional vector, respectively.

The efficiency of the proposed feature selection regarding three color spaces can be observed in Fig. 1.9, where we plot the one dimensional marginal histograms of the occurring ψ_0 , ψ_1 and ψ_2 values for manually marked shadowed and foreground points of a 75-frames long outdoor surveillance video sequence (‘Entrance pm’). Apart from some outliers, the shadowed ψ_i values lie for each color space and each color component in a ‘short’ interval, while the difference of the upper and lower bounds of the foreground values is usually greater.

1.4.3 Quantitative Comparison through a Deterministic Classifier

In this section, we temporarily put aside the MRF concept, and taking a deterministic approach, we consider the shadow detection problem as a simple classification task in the $\bar{\psi}$ -feature space. Considering Fig. 1.9, an important note should be taken here. While the $\bar{\psi}$ statistics characterizes the scene and illumination conditions, the foreground $\bar{\psi}$ histograms only correspond to the occurring foreground objects in the evaluated sequence. On the other hand, an efficient shadow model is expected to work with differently colored objects as well. Therefore, the upcoming discrimination process will follow a one-class-classification approach: pixel s will be classified as a shadowed point, if its $\bar{\psi}(s)$ value lies in the estimated *shadow domain*, and the *outlier points* will be labeled as foreground. As usual, the shadow domain is defined by a manifold having a prescribed number of free parameters, which fit the model to a given scene/situation. For grayscale images shadowed ψ features should be included by an interval [37], while regarding color scenes different domain

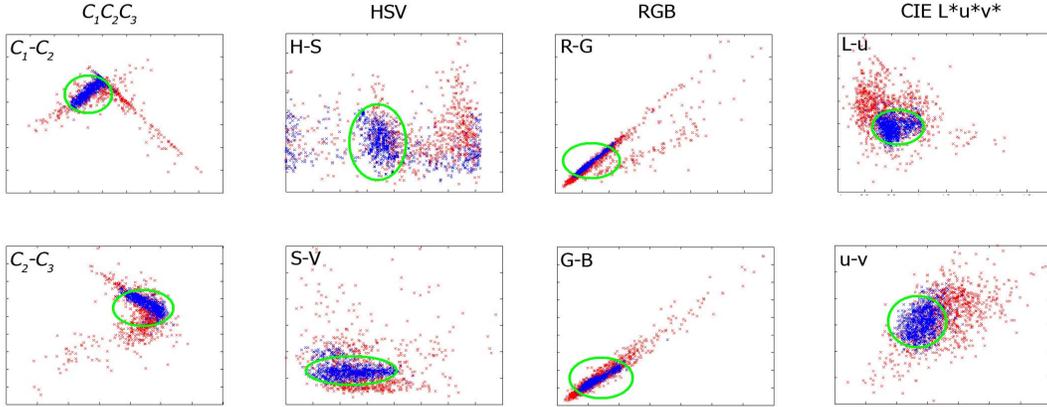


Figure 1.10: Two dimensional projection of foreground (red) and shadow (blue) $\bar{\psi}$ values in the 'Entrance pm' test sequence. Green ellipse is the projection of the optimized shadow boundary.

models are used in the literature, like a three dimensional rectangular bin [51] (ratio/difference values for each channel lie between defined threshold), an ellipsoid [35], or the domain may have general shape [53]. In the latter case a Support Vector Domain description is proposed in the RGB color ratios' space.

By each domain-selection we must consider overlap between the classes, e.g. foreground points may appear whose feature values lie in the shadow domain. Therefore, the optimal domain should be as narrow as possible meanwhile containing 'almost all' the feature values corresponding to the occurring shadowed points. Accordingly, if we 'only' prescribe that a shadow descriptor should be accurate, the most general domain shape seems to be the most appropriate. However, in practise, we also have to consider issues of parameter estimation and adaption (see Section 1.3.4). Therefore, we prefer the domains with relatively few free parameters, for which we can construct an automatic update strategy.

Observe that according to Fig. 1.9, the shadowed ψ_0 , ψ_1 and ψ_2 values follow approximately normal distributions, therefore, a 3D joint normal representation of the $\bar{\psi}$ features in shadows is straightforward (similarly to Section 1.3). Since the equipotential surfaces of the 3D Gaussian density functions are ellipsoids, a natural choice is using an elliptical shadow domain boundary. We will use the equation of a standard ellipsoid body having *parallel axes* with the coordinate axes in the $\psi_0 - \psi_1 - \psi_2$ Cartesian coordinate system:

$$\text{Pixel } s \text{ is shadowed} \Leftrightarrow \sum_{i=0}^2 \left(\frac{\psi_i(s) - a_i}{b_i} \right)^2 \leq 1, \quad (1.19)$$

where $[a_0, a_1, a_2]$ is the coordinate of the ellipsoid center and (b_0, b_1, b_2) are the semi-axis lengths. In other words, $[a_0, a_1, a_2]$ is equivalent to the mean $\bar{\psi}(s)$ value of shadowed pixels in a given scene, while b_0, b_1 and b_2 depend on the spatiotemporal variance of the $\bar{\psi}(s)$ measurements under shadows. Later on we will show that the similarity to the $\bar{\mu}_\psi$ and $\bar{\sigma}_\psi$ parameters from Section 1.3 is not by chance, thus, parameter adaption

can also be done in a similar manner.

Note that with the SVM method [53], the number of free parameters is related to the number of the support vectors, which can be much greater than the six scalars of our model. Moreover, for each situation, a novel SVM should be trained. Note as well that one could use an arbitrarily oriented ellipsoid, but compared to eq. 1.19, it is also more difficult to define, since it needs the accurate estimation of 9 parameters.

For the sake of completeness, we note that the domain defined by eq. (1.19) becomes an interval if we work with grayscale images, and a two dimensional ellipse in the rg space.

Fig. 1.10 shows the two dimensional scatter plots about the occurring foreground and shadow $\bar{\psi}$ values. We can observe here that the components of vector $\bar{\psi}$ are strongly correlated in the RGB space (and also in $C_1C_2C_3$), and the previously defined ellipse cannot present a narrow boundary. In the HSV space, the shadowed values are not within a convex hull, even if we considered that the hue component is actually periodical (hue = $k * 2\pi$ means the same color for each $k = 0, 1, \dots$). Based on the above facts, the CIE $L^*u^*v^*$ space seems to be a good choice. In the following, we support this statement by numerical results.

Evaluation of the Deterministic Model

The evaluations were done through manually generated ground truth sequences regarding five of the previously introduced test videos: namely the ‘Laboratory’, ‘Highway’, ‘Entrance am’, ‘Entrance noon’ and ‘Entrance pm’ sequences, with the same test parameters as before (see Table 1.3 in page 14 for details).

In this section, we show the tentative limits of the elliptical shadow domain defined by eq. (1.19). The goal of these experiments is to compare the foreground-shadow discriminating ability of the different color spaces purely based on the extracted per pixel $\bar{\psi}$ features. Therefore, we set here the parameters manually, and do not take into consideration local connectivity or post processing.

In the upcoming experiments, we collect for each test sequence two sets of $\bar{\psi}$ values corresponding to manually marked foreground and shadowed pixels, respectively. We investigate on the correct-classification-rates of the pixels by using the ellipse model (eq. 1.19) with different color spaces. We henceforward use the *Recall* (Rc), *Precision* (Pr) measures, which were introduced in Section 1.3.6.

For some optimized ellipse parameters, we plot the corresponding *Precision* and *Recall* values regarding the ‘Laboratory’ and ‘Entrance pm’ test sequences in Fig. 1.11. We can observe that the CIE $L^*a^*b^*$ and $L^*u^*v^*$ spaces produce the best results in both cases (the corresponding Pr/Rc curves are the highest). However, the relative performance of the other color systems is strongly different regarding the two videos. In the indoor scene, the grayscale and RGB segmentations are less efficient than the other ones, while regarding the ‘Entrance pm’ sequence, the performance of the chrominance spaces is prominently poor.

In the further tests, we will use the *FM*-measure (eq. 1.16). We summarized the *FM* rates in Fig. 1.12, regarding the test sequences. Also here, we can see that the CIE $L^*a^*b^*$ and $L^*u^*v^*$ spaces are

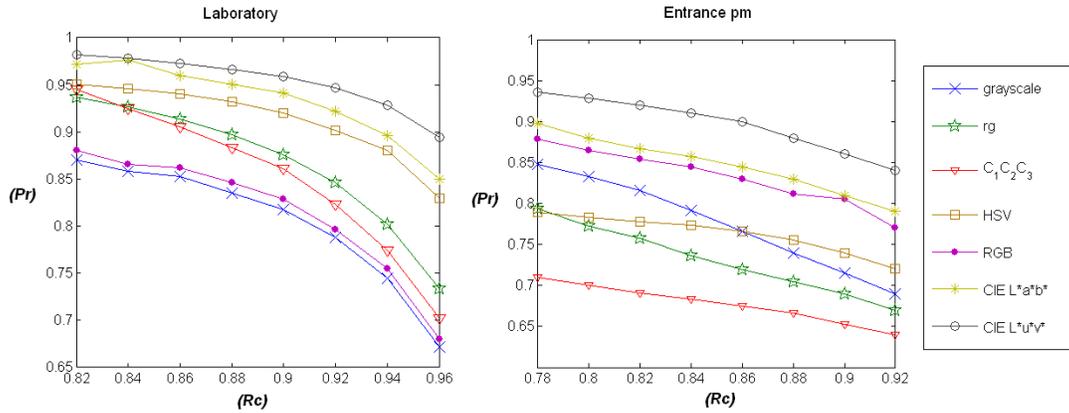


Figure 1.11: Evaluation of the deterministic model. Recall-precision curves corresponding to different parameter-settings on the ‘Laboratory’ and ‘Entrance pm’ sequences.

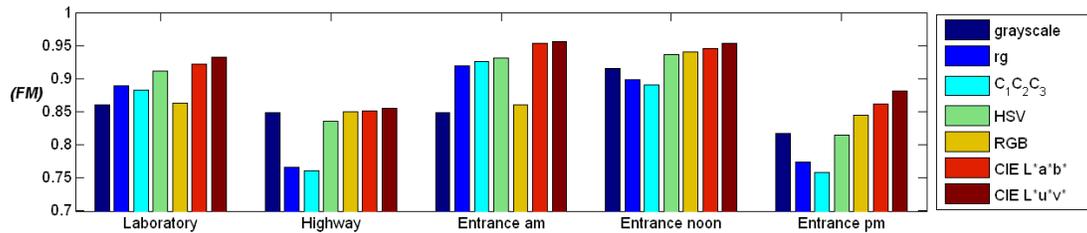


Figure 1.12: Evaluation of the deterministic model. FM coefficient (eq. 1.16) regarding different sequences

the most efficient. As for the other color systems, in sequences containing dark shadows (‘Entrance pm’, ‘Highway’), the ‘chrominance spaces’ produce poor results, while the gray, RGB and L*a*b*/L*u*v* results are similarly effective. If the shadow is brighter (‘Entrance am’, ‘Laboratory’), the performance of the ‘chrominance spaces’ becomes reasonable, but the ‘luminance spaces’ are relatively poor. In the latter case, the color constancy of the chrominance channels seems to be more relevant than the luminance-darkening domain. We have also observed that the hue coordinate in HSV is very sensitive to the illumination artifacts (see also Fig. 1.9), thus the HSV space is more efficient in case of light-shadows. We give a summary about the relationship between the darkness of shadow and the performance of color spaces in Table 1.7, where ‘darkness’ is characterized by the mean of the grayscale- ψ_0 values of shadowed points.

1.4.4 Segmentation with Different Color Spaces

The results in the previous section confirm that using the elliptical shadow domain defined by eq. 1.19, the CIE L*u*v* color space is the most efficient regarding the separation of shadowed and foreground pixels. However, those experiments needed manually evaluated training data to set the parameters. In the

Table 1.7: Indicating the two most successful and the two less efficient color spaces regarding each test sequence, based on the experiments of Section 1.4.3 (For numerical evaluation see Fig. 1.11 and 1.12). To compare the scenarios, we also denote † the mean darkening factor of shadows in grayscale.

Video	Scene	Dark [†]	Worst	Best
Laboratory	indoor	0.73	gray, RGB	Luv, Lab
Entrance am	outdoor	0.50	gray, RGB	Luv, Lab
Entrance pm	outdoor	0.39	$C_1 C_2 C_3$, rg	Luv, Lab
Entrance noon	outdoor	0.35	$C_1 C_2 C_3$, rg	Luv, Lab
Highway	outdoor	0.23	$C_1 C_2 C_3$, rg	Luv, Lab

following, we suit the above model to the adaptive Bayesian model-framework of Section 1.3, and show that the advantage of using the appropriate color space can be also measured directly in the applications.

First, we give to the shadow-classification step defined in Section 1.4.3, a probabilistic interpretation. We rewrite eq. (1.19): we match the current $\bar{\psi}(s)$ value of pixel s to a probability density function $f(\bar{\psi}(s))$, and decide its class by:

$$\text{pixel } s \text{ is shadowed} \Leftrightarrow f(\bar{\psi}(s)) \geq t. \quad (1.20)$$

Based on the one dimensional marginal histograms in Fig. 1.9, we model $f(\bar{\psi}(s))$ by a multi variate Gaussian density function, similarly to the CIE L*a*v* case introduced in Section 1.3. To keep the six-parameter shadow model, a diagonal covariance matrix will be used (i.e. the three element-mean value vector, and the three diagonal components of the covariance matrix should be defined). In this way, we model the variety of the $\bar{\psi}$ values observed in shadows, which variety is caused by camera noise, fine alterations in illumination, and differences in albedo and geometry of the different surface points. However, the changes in the different color components are considered to be independent exploiting that many color spaces (like CIE L*a*v*, CIE L*a*b*, HSV) have approximately uncorrelated basis [33]. As for the RGB space, this ‘diagonal’ approach is less accurate. However, we show later on that for most of the sequences the performance of this oversimplified RGB-model is also reasonable.

Note that as shown in [2], the domains defined by eq. (1.19) and eq. (1.20) are equivalent, if f is a Gaussian density function (η):

$$f(\bar{\psi}(s)) = \eta(\bar{\psi}(s), \bar{\mu}_\psi, \bar{\Sigma}_\psi) = \quad (1.21)$$

$$= \frac{1}{(2\pi)^{\frac{3}{2}} \sqrt{\det \bar{\Sigma}_\psi}} \exp \left[-\frac{1}{2} (\bar{\psi}(s) - \bar{\mu}_\psi)^T \bar{\Sigma}_\psi^{-1} (\bar{\psi}(s) - \bar{\mu}_\psi) \right] \quad (1.22)$$

with the following parameters:

$$\bar{\mu}_\psi = [a_0, a_1, a_2]^T, \quad \bar{\Sigma}_\psi = \text{diag}\{b_0^2, b_1^2, b_2^2\}, \quad t = (2\pi)^{-\frac{3}{2}} (b_0 b_1 b_2)^{-1} e^{-\frac{1}{2}}. \quad (1.23)$$



Figure 1.13: MRF segmentation results with different color models. Test sequences (up to down): row 1: ‘Laboratory’, row 2: ‘Highway’, row 3: ‘Entrance am’, rows 4: ‘Entrance pm’, row 5: ‘Entrance noon’.

In the following, we use the previously defined probability density functions in the MRF model in a straightforward way: $p_{sh}(s) = f(\bar{\psi}(s))$. The flexibility of this MRF model comes from the fact that we defined $\bar{\psi}(s)$ shadow descriptors for different color spaces differently in Section 1.4.2.

Test Results

Fig. 1.13 shows the MRF-segmentation results of two frames from each test sequence using five color spaces: grayscale, $C_1C_2C_3$, HSV, RGB and CIE L*u*v*. (Note that in the experiments, the results of the CIE L*a*b* space have been very similar to the L*u*v* outputs, while the *rg* has worked similarly to $C_1C_2C_3$, thus we skip them in this comparison). We can observe that the CIE L*u*v* space outperforms the other ones significantly, while we get the largest errors with $C_1C_2C_3$, especially in the cases of sharp shadows. We find a typical problem regarding the HSV and RGB spaces: foreground ‘glories’ may appear around some dark shadowed parts due to the penumbra of cast shadow [29] and video compression. These erroneous areas correspond to shadows, but they are lighter than the central areas, thus they get out of the shadow domain in the feature space. On the other hand, the proposed probabilistic model removes these artifacts with the other color spaces.

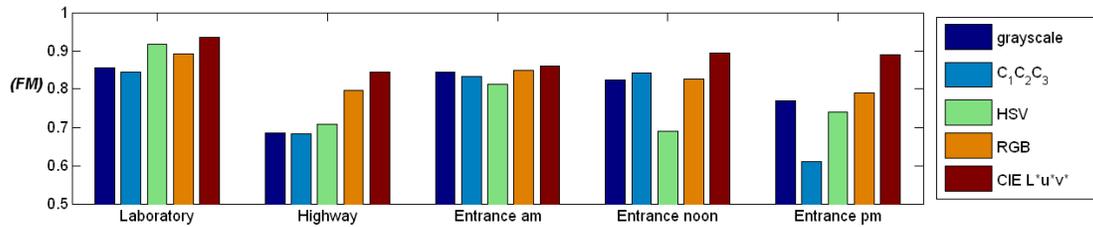


Figure 1.14: Evaluation of the MRF model. F^* coefficient regarding different sequences

Hereinafter, we perform quantitative evaluations using the MRF model. In Section 1.4.3, we measured purely the ability to discriminate foreground and shadowed pixels. Since the present model uses three classes and the goal is accurate foreground detection, we should also consider the confusion rate between foreground and background. However, similarly to Section 1.3.6 the crossover between shadow and background does not count for errors (both of them are non-foreground areas).

We observe in Fig. 1.14 the clear superiority of the CIE L*u*v* space. However, the relative performance of the color spaces does not show exactly the same tendencies as we have measured in Section 1.4.3. The differences between Fig. 1.12 and 1.14 are caused by effects of the composite foreground model, MRF neighborhood conditions and errors in parameter estimation, since the artifacts may appear differently in the different sequences. Therefore, we consider the numerical results from Section 1.4.3 to be more relevant to characterize the capabilities of the color spaces for shadow separation. However, the experiments of this section confirm that appropriate color space selection is also crucial in the applications, and the CIE L*u*v* space is preferred for this task.

1.5 Conclusion of the Chapter

This chapter has examined the color modeling problem of cast shadows, focusing on the video surveillance application. We have introduced a novel accurate and adaptive model for shadow segmentation without strong restrictions on a priori probabilities, image quality, objects' shapes and speed. Thereafter we have generalized the proposed model framework for different color spaces, and made a detailed comparison among them. We have observed that the appropriate color space selection is an important issue of the classification, and the CIE L*u*v* space is the most efficient both in the color based clustering of the individual pixels and in the case of Bayesian foreground-background-shadow segmentation. We validated our method on various video shots, including well-known benchmark videos and real-life surveillance sequences, indoor and outdoor shots, which contain both dark and light shadows. Experimental results have shown the advantages of our statistical approach versus earlier methods.

Contents

1	Shadow Detection in Digital Images and Videos	1
1.1	Introduction: Goals and Applications	1
1.2	Shadow Detection in Video Surveillance: an Overview	3
1.3	A Bayesian Approach for Modeling Shadows in Video Scenes	5
1.3.1	General Probabilistic Models of the Background and Shadow Processes	6
1.3.2	Shadow Description by Lambertian Color Features	7
1.3.3	Proposed Shadow Model	8
1.3.4	Parameter Settings	9
1.3.5	MRF Optimization	13
1.3.6	Experimental results	14
1.4	Color Space Selection in Cast Shadow Detection	17
1.4.1	Color Spaces: Significance of the Right Choice	18
1.4.2	Generalized Feature Vector of Shadow Separation	19
1.4.3	Quantitative Comparison through a Deterministic Classifier	20
1.4.4	Segmentation with Different Color Spaces	23
1.5	Conclusion of the Chapter	26

Bibliography

- [1] C. Benedek and T. Szirányi, “Bayesian foreground and shadow detection in uncertain frame rate surveillance videos,” *IEEE Trans. on Image Processing*, vol. 17, no. 4, pp. 608 – 621, 2008.
- [2] C. Benedek and T. Szirányi, “Study on color space selection for detecting cast shadows in video surveillance,” *International Journal of Imaging Systems and Technology*, vol. 17, no. 3, pp. 190–201, 2007.
- [3] C. Benedek, T. Szirányi, Z. Kato, and J. Zerubia, “Detection of object motion regions in aerial image pairs with a multi-layer Markovian model,” *IEEE Trans. on Image Processing*, vol. 18, no. 10, pp. 2303–2315, 2009.
- [4] J. Kato, T. Watanabe, S. Joga, L. Ying, and H. Hase, “An HMM/MRF-based stochastic framework for robust vehicle tracking,” *IEEE Trans. on Intelligent Transportation Systems*, vol. 5, no. 3, pp. 142–154, 2004.
- [5] C. Benedek and T. Szirányi, “Change detection in optical aerial images by a multi-layer conditional mixed Markov model,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.
- [6] D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, “On the removal of shadows from images,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, 2006.
- [7] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.
- [8] R. Z. Ping-Sing, R. Zhang, P. sing Tsai, J. E. Cryer, and M. Shah, “Shape from shading: A survey,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 690–706, 1999.
- [9] Z. Szlávik, L. Kovács, L. Havasi, C. Benedek, I. Petrás, A. Utasi, A. Licsár, L. Czúni, and T. Szirányi, “Behavior and event detection for annotation and surveillance,” in *Proc. International Workshop on Content-Based Multimedia Indexing (CBMI)*, (London, UK), pp. 117–124, IEEE, 2008.
- [10] A. Katartzis and H. Sahli, “A stochastic framework for the identification of building rooftops using a single remote sensing image,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, pp. 259–271, January 2008.
- [11] B. Sirmacek and C. Unsalan, “Building detection from aerial imagery using invariant color features and shadow information,” in *International Symposium on Computer and Information Sciences (ISCIS)*, (Istanbul, Turkey), October 2008.
- [12] G. Perrin, X. Descombes, and J. Zerubia, “2D and 3D vegetation resource parameters assessment using marked point processes,” in *Proc. International Conference on Pattern Recognition (ICPR)*, (Hong-Kong), August 2006.
- [13] R. Cutler and L. S. Davis, “Robust real-time periodic motion detection, analysis, and applications,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 781–796, 2000.
- [14] L. Havasi, Z. Szlávik, and T. Szirányi, “Higher order symmetry for non-linear classification of human walk detection,” *Pattern Recognition Letters*, vol. 27, pp. 822–829, 2006.

- [15] L. Havasi, Z. Szlávik, and T. Szirányi, "Detection of gait characteristics for scene registration in video surveillance system," *IEEE Trans. Image Processing*, vol. 16, no. 2, pp. 503–510, 2007.
- [16] L. Czúni and T. Szirányi, "Motion segmentation and tracking with edge relaxation and optimization using fully parallel methods in the cellular nonlinear network architecture," *Real-Time Imaging*, vol. 7, no. 1, pp. 77–95, 2001.
- [17] Z. Zivkovic, *Motion Detection and Object Tracking in Image Sequences*. PhD thesis, PhD thesis, University of Twente, 2003.
- [18] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Human identification by spatio-temporal symmetry," in *Proc. International Conference on Pattern Recognition*, vol. 1, (Washington, DC, USA), p. 10632, IEEE Computer Society, 2002.
- [19] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.
- [20] S. C. Zhu and A. L. Yuille, "A flexible object recognition and modeling system," *Int'l Journal of Computer Vision*, vol. 20, no. 3, 1996.
- [21] L. Havasi and T. Szirányi, "Estimation of vanishing point in camera-mirror scenes using video," *Optics Letters*, vol. 31, no. 10, pp. 1411–1413, 2006.
- [22] A. Yoneyama, C. H. Yeh, and C.-C. J. Kuo, "Moving cast shadow elimination for robust vehicle extraction based on 2D joint vehicle/shadow models," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, IEEE, 2003.
- [23] C. Fredembach and G. D. Finlayson, "Hamiltonian path based shadow removal," in *Proc. British Machine Vision Conference*, pp. 970–980, 2005.
- [24] T. Gevers and H. Stokman, "Classifying color edges in video into shadow-geometry, highlight, or material transitions," *IEEE Trans. on Multimedia*, vol. 5, no. 2, pp. 237–243, 2003.
- [25] E. A. Khan and E. Reinhard, "Evaluation of color spaces for edge classification in outdoor scenes," in *Proc. of International Conference on Image Processing*, vol. 3, (Genoa, Italy), pp. 952–955, IEEE, Sept. 2005.
- [26] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.
- [27] N. Paragios and V. Ramesh, "A MRF-based real-time approach for subway monitoring," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1034–1040, 2001.
- [28] A. Cavallaro, E. Salvador, and T. Ebrahimi, "Detecting shadows in image sequences," in *Proc. of European Conference on Visual Media Production*, pp. 167–174, March 2004.
- [29] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Computer Vision and Image Understanding*, no. 2, pp. 238–259, 2004.
- [30] F. Porikli and J. Thornton, "Shadow flow: a recursive method to learn moving cast shadows," in *Proc. IEEE International Conference on Computer Vision*, vol. 1, pp. 891–898, 2005.
- [31] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 643–648, June 2005.

- [32] Y. Haeghen, J. Naeyaert, I. Lemahieu, and W. Philips, "An imaging system with calibrated color image acquisition for use in dermatology," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 722–730, 2000.
- [33] M. G. A. Thomson, R. J. Paltridge, T. Yates, and S. Westland, "Color spaces for discrimination and categorization in natural scenes," in *Proc. Congress of the International Colour Association*, pp. 877–880, June 2002.
- [34] T. Gevers and A. W. Smeulders, "Color based object recognition," *Pattern Recognition (PR)*, vol. 32, pp. 453–464, 1999.
- [35] I. Mikic, P. Cosman, G. Kogut, and M. M. Trivedi, "Moving shadow and object detection in traffic scenes," in *Proc. International Conference on Pattern Recognition*, 2000.
- [36] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [37] Y. Wang, K.-F. Loe, and J.-K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 279–289, 2006.
- [38] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [39] J. Rittscher, J. Kato, S. Joga, and A. Blake, "An HMM-based segmentation method for traffic monitoring," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1291–1296, 2002.
- [40] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 827–832, 2005.
- [41] D. A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–36, 1990.
- [42] D. K. Lynch and W. Livingstone, *Color and Light in Nature*. Cambridge University Press, 1955.
- [43] G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*. John Wiley & Sons, second ed., 1982.
- [44] Y. Wang and T. Tan, "Adaptive foreground and shadow detection in image sequences," in *Proc. International Conference on Pattern Recognition*, pp. 983–986, 2002.
- [45] R. Potts, "Some generalized order-disorder transformation," in *Proceedings of the Cambridge Philosophical Society*, no. 48, p. 106, 1952.
- [46] Z. Kato, J. Zerubia, and M. Berthod, "Satellite image classification using a modified Metropolis dynamics," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, pp. 573–576, March 1992.
- [47] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chem. Physics*, vol. 21, pp. 1087–1092, 1953.
- [48] J. Besag, "On the statistical analysis of dirty images," *Journal of Royal Statistics Society*, vol. 48, pp. 259–302, 1986.
- [49] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [50] C. J. V. Rijsbergen, *Information Retrieval*. London: Butterworths, second ed., 1979.

- [51] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati, "The Sakbot system for moving object detection and tracking," in *Video-Based Surveillance Systems-Computer Vision and Distributed Processing*, pp. 145–157, 2001.
- [52] M. Rautiainen, T. Ojala, and H. Kauniskangas, "Detecting perceptual color changes from sequential images for scene surveillance," *IEICE Transactions on Information and Systems*, pp. 1676 – 1683, 2001.
- [53] K. Siala, M. Chakchouk, F. Chaieb, and O. Besbes, "Moving shadow detection with support vector domain description in the color ratios space," in *Proc. International Conference on Pattern Recognition*, vol. 4, pp. 384–387, 2004.
- [54] V. Meas-Yedid, E. Glory, E. Morelon, C. Pinset, G. Stamon, and J.-C. Olivo-Marin, "Automatic color space selection for biological image segmentation," in *Proc. International Conference on Pattern Recognition*, vol. 3, (Washington, DC, USA), pp. 514–517, IEEE Computer Society, 2004.
- [55] P. Guo and M. R. Lyu, "A study on color space selection for determining image segmentation region number," in *Proc. International Conference on Artificial Intelligence*, vol. III, (Las Vegas, Nevada, USA), pp. 1127–1132, June 2000.
- [56] A. Prati, I. Mikic, C. Grana, and M. Trivedi, "Shadow detection algorithms for traffic flow analysis: A comparative study," in *Proc. IEEE Intelligent Transportation Systems Conference*, (Oakland, CA, USA), pp. 340–345, 2001.
- [57] M. Tkalcic and J. Tasic, "Colour spaces - perceptual, historical and applicational background," in *Proc. Eurocon*, 2003.

Index

- background, 1–15, 17, 20, 26
- Bayesian, 5, 17, 24

- color space, 5, 17–20, 22–26
 - $C_1C_2C_3$, 17–19, 22, 24, 25
 - CIE $L^*a^*b^*$, 17, 18, 22, 24, 25
 - CIE $L^*u^*v^*$, 4, 17–20, 22–26
 - grayscale, 17, 19, 20, 22–25
 - HSV, 17, 18, 20, 22–25
 - rg, 19, 20, 22, 24
 - RGB, 17–19, 21–25

- foreground, 2, 3, 5, 6, 8, 9, 13, 15–23, 25, 26

- histogram, 10, 11, 19, 20, 24

- moving object, 1–3, 8
- MRF, 13, 19, 20, 23, 25, 26
 - Potts model, 13

- optimization, 13, 19

- shadow, 1–12, 14–26
 - constant ratio, 5, 7, 8, 14–16
 - invariant, 4, 5, 14, 15, 17
 - shadow detection, 2, 5, 7, 17, 18, 20
- Stauffer-Grimson algorithm, 2, 12