Study on Color Space Selection for Detecting Cast Shadows in Video Surveillance

Csaba Benedek and Tamás Szirányi

Pázmány Péter Catholic University, Department of Information Technology, Práter utca 50/A, Budapest, Hungary
Distributed Events Analysis Research Group, Computer and Automation Institute, Kende u. 13-17, Budapest, Hungary

Abstract

In this paper we address the color modelling problem of cast shadows in video sequences. We illustrate that the performance of shadow detection can be improved significantly through appropriate color space selection, if for practical purposes, we should keep the number of free parameters of the method low. Hence, we compare several well known color spaces with a six-parameter shadow model embedded into a globally optimal MRF framework. Experimental results on real-life videos show that CIE L*u*v* color space is the most efficient.

Key words: video surveillance, shadow detection, color spaces, MRF

1. Introduction

Detection of foreground objects is a crucial task in visual surveillance systems. If we are able to retrieve the accurate silhouettes of the objects or object groups, their high-level description could be more sophisticated, so it is favorable e.g. in detection of people (Havasi et al., 2006) or vehicles (Rittscher et al., 2000), and in activity analysis (Stauffer & Grimson, 2000). The presence of moving shadows on the background makes difficult to estimate shape or behavior of the objects, therefore, shadow detection is an important issue in the applications. However, we should not focus on self shadows (i.e. shadows appearing on the foreground objects), which are part of the foreground, and static shadows (cast shadows of the static objects), because they correspond to the background.

Apart from a few geometry based approaches which are suited to specific conditions (Havasi & Szirányi, 2006), (Yoneyama et al., 2003), shadow detection is usually done by color filtering. Still image based methods (Fredembach & Finlayson, 2005),(Finlayson et a., 2006) attempt to find and remove shadows in the single frames independently. However, these models have been evaluated only on high quality images where the background has a uniform color or texture pattern, while in video surveillance, we must expect images with poor quality and resolution. The authors in (Finlayson et a., 2006) note that their algorithm is robust when the shadow edges are clear, but artifacts may appear in cases of images with complex shadows or diffuse shadows with poorly defined edges. For practical use, the computational complexity of these algorithms should be decreased (Fredembach & Finlayson, 2005).

Some other methods focus on the discrimination of the shadow edges, and edges due to objects boundaries (Gevers & Stokman, 2003), (Khan & Reinhard, 2005). However, it may be difficult to extract connected foreground regions from the resulting edge map, which is often ragged (Gevers & Stokman, 2003). Complex scenarios containing several small objects or shadow-parts may be also disadvantageous for these methods.

(Prati et al., 2003) gives a thematic overview on shadow detection for video surveillance. The methods are classified into groups based on their *model structures*, and the performance of the different model-groups are compared via test sequences. The authors note that the methods work in different color spaces, like RGB (Mikic et al., 2000) and HSV (Cucchiara et al., 2001), 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: (Rittscher et al., 2000) used only gray levels for

shadow segmentation, other approaches were dealing with the CIE L*u*v* (Martel-Brisson & Zaccarin, 2005) and CIE L*a*b* (Rautiainen et al., 2001) spaces, respectively. For the above reasons, the main issue of this paper is to give an experimental comparison of different color models regarding cast shadow detection on the video frames. For the comparison, we propose a general model framework, which can work with different color spaces. During the development of this framework, we have carefully considered the main approaches in the state-of-the art. Our presented model is the generalization of our previous work (Benedek & Szirányi, 2006), while the preliminary version of this study was published in (Benedek & Szirányi, 2007). We note that an experimental evaluation of color spaces have been already done for edge classification in (Khan & Reinhard, 2005), but in the current paper, we address the detection of the shadowed and foreground regions, which is a fairly different problem.

2. Overview on different approaches

In (Prati et al., 2003), the authors distinguished *deterministic* methods (e.g. Cucchiara et al. (2001)), which use on/off decision processes at each pixel, and *statistical* approaches (see Mikic et al. (2000)) which contain probability density functions to describe the shadow-membership of a give image point. The classification of the methods whether they are deterministic or statistical depends often only on interpretation, since deterministic decisions can be done using probabilistic functions also. However, statistical methods have been widely distributed recently, since they can be used together with Markov Random Fields (MRF) to enhance the quality of the segmentation significantly (Wang et al., 2006).

The first model, which we introduce in this paper is a deterministic method: the pixels are classified independently and 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 processing 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 into an adaptive MRF framework for foreground-background-shadow segmentation which we developed earlier (Benedek & Szirányi, 2006). We compare the results after MRF optimization qualitatively and quantitatively.

Another important point of view regarding the categorization of the algorithms in (Prati et al., 2003) is the discrimination of the *non parametric* and *parametric* cases. Non parametric, or 'shadow invariant' methods convert the pixel values into an illuminant invariant feature space: they remove shadows instead of detecting them. This task is often performed by a color space transformation. The normalized rgb (Cavallaro et al., 2004),(Paragios & Ramesh, 2001)

and $C_1C_2C_3$ spaces (Salvador et al., 2004) [1] are supposed to fulfill color constancy through using only chrominance color components. We find an overview on these approaches in (Salvador et al., 2004) indicating that several assumptions are needed regarding the reflecting surfaces and the lightings. (Khan & Reinhard, 2005) emphasizes the limits of these methods: 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 sunlit. We have also found in our experiments that the shadow invariant methods fail outdoors several times, and they are rather usable indoors (Fig. 7). Moreover, since they ignore the luminance components of the color which is a very important factor both in the human and computer vision, these models become sensitive to noise. We show in a later part of this paper that the rg and $C_1C_2C_3$ spaces are also less effective in the parametric case.

For the above reasons, we develop a *parametric* model: first, we estimate the mean background values of the individual pixels trough a statistical background model (Stauffer & Grimson, 2000), then we extract feature vectors from the actual and the estimated background values of the pixels and apply shadow detection as solving a classification problem in that feature space. This approach is widespread in the literature, and the key points are the way of *feature extraction*, the *color space selection* and the *shadow-domain description* in the feature space. We will deal with these issues in the following sections of the paper. In Section 3, we introduce the feature vector which characterizes the shadowed pixels effectively. In Section 4, we describe the chosen shadow domain in the feature space, and define the deterministic pixel classification method. We show the quantitative classification results with the deterministic model regarding five real-world video sequences in Section 5.1. Finally, we introduce the MRF framework and analyse the segmentation results in Section 5.2.

We use a few assumptions in the paper: (1) The camera stands in place and has no significant ego-motion. (2) The background objects are static (e.g. there is no waving river in the background), and the topically valid 'background image' is available in each moment (e.g. by the method of (Stauffer & Grimson, 2000)). (3) 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. (4) We assume that the estimated background values of the pixels correspond usually to the illuminated surface points. Obviously, this is not the case if the pixel is covered by a static shadow, which is not handled here explicitly. However, according to our tests, our approach is able to handle a few static shadows robustly.

---

[1] We refer later to the normalized rgb as rg space, since the third color component is determined by the first and second.

Method	Color space	Number of parameters/ color channels*	Outdoor/ Indoor tests	Number of test sequences
Cavallaro et al. (2004)	rg	invariant	both	6
Salvador et al. (2004)	$C_1 C_2 C_3$	invariant	both	4
Paragios & Ramesh (2001)	rg	invariant	indoor	1
Mikic et al. (2000)	RGB	1	outdoor	1
Rittscher et al. (2000)	grayscale	2	outdoor	1
Wang et al. (2006)	grayscale	2	indoor	4
Cucchiara et al. (2001)	HSV	1.33	both	5
Martel-Brisson & Zaccarin (2005)	CIE L*u*v*	2	indoor	1
Rautiainen et al. (2001)	CIE L*a*b*/HSV	N.a.	outdoor	3
Siala et al. (2004)	RGB	N.a.‡	outdoor	1
Proposed	All from above	2	both	5

Table 1

Overview on different methods. †Only in case of parametric methods: the (average) number of parameters of the shadow detector module for one color channel. ‡Proportional to the number of support vectors after supervised training.

3. Feature vector

Here, we define features for a parametric case where a shadow model can be constructed including some challenging environmental conditions. First, we introduce a well-known physical approach on shadow detection with marking that its model assumptions may fail in real-world video scenes. Instead of constructing a more difficult illumination model, we overcome the appearing artifacts with a statistical description. Finally, the efficiency of the proposed model is validated by experiments.

3.1. Physical approach on shadow detection

According to the illumination model (Forsyth, 1990) the response $g(s)$ of a given image sensor placed at pixel s can be written as

$$g(s) = \int e(\lambda, s) \rho(\lambda, s) \nu(\lambda) d\lambda, \qquad (1)$$

where $e(\lambda, s)$ is the illumination function, $\rho(s)$ depends on the surface albedo and geometry, while $\nu(\lambda)$ 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 $e(\lambda, s)$. For example, 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 tingle (Lynch & Livingstone, 1995), and residual light components reflected from other non-emissive objects. In the shadow, the effect of the direct component is missing.

Although the validity of eq. (1) is already limited by several scene assumptions (Forsyth, 1990), in general, it is still too difficult to exploit appropriate information about the corresponding background-shadow values, since the components of the illumination function are unknown. Therefore, further strong simplifications are used in the applications. According to (Finlayson et a., 2006) the camera sensors must be exact Dirac delta functions: $\nu(\lambda) = q_0 \cdot \delta(\lambda - \lambda_0)$ and the illumination must be Planckian (Wyszecki & Stiles, 1982). In this case, eq.(1) implies the well-known 'constant ratio' rule. Namely, the ratio of the shadowed $g_{\mathrm{sh}}(s)$ and illuminated value $g_{\mathrm{bg}}(s)$ of a given surface point is considered to be constant over the image:

$$\frac{g_{\mathrm{sh}}(s)}{g_{\mathrm{bg}}(s)} = A, \qquad (2)$$

where A is the shadow 'darkening factor', and it does not depend on s.

In the CCD camera model (Forsyth, 1990) the RGB sensors are narrow banded and the constant ratio rule is supposed to be valid for each color channel independently. Accordingly, the shadow descriptor is a triple $[A_r, A_g, A_b]$ containing the ratios of the shadowed and illuminated background values for the red, green and blue channels. However, due to the deviation of the scene properties from the model assumptions (Forsyth, 1990), imprecise estimation of the background values (Stauffer & Grimson, 2000) and further artifacts caused by video compression and quantification, the ratio of the shadowed and estimated background values is not a constant in fact. On the other hand, to prescribe a *domain* instead of a single value for the ratios results a powerful detector (Siala et al., 2004). In this way, shadow detection is a one-class-classification problem in the three dimensional color ratio space.

3.2. Constant ratio rule in different color spaces

In this section, we examine, how we can use the previous physical 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 equations in (Tkalcic & J. Tasic, 2003). The ITU D65 standard is used for calibration of the CIE L*u*v* and L*a*b* spaces. In the HSV, CIE L*u*v* and L*a*b* spaces we should discriminate two types of color components. One component is directly related to the brightness of the pixel (V, respectively L*; we refer to them later as 'luminance' components), while the other components correspond to the 'chrominances'. We classify also 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'.

As we stated in the last section, the ratios of the shadowed and illuminated values of the R, G, B color channels regarding a given pixel are near to a global reference value $[A_r, A_g, A_b]$. In the following, we show by experiments that the 'constant ratio rule' is also a reasonable approximation regarding the 'luminance' components of other color spaces. While shadow may darken the 'luminance' values of the pixels significantly, the changes in the 'chrominances' are usually small. In (Cucchiara et al., 2001), the hue difference was modelled as a zero-mean noise factor. This approach is sometimes inaccurate: outdoors, due to the ambient light of the blue sky, the shadow shifts to the 'blue' color domain. We will show that modelling the offset between the shadowed and illuminated 'chrominance' values of the pixels with a Gaussian additive term is reasonable.

To sum it up, if the current value of a given pixel in a given color space is $[x_0, x_1, x_2]$ (the indices $0, 1, 2$ correspond to the different color components), the estimated (illuminated) background value is there $[m_0, m_1, m_2]$, we define the shadow descriptor $\overline{\psi} = [\psi_0, \psi_1, \psi_2]$ by the following, for $i = \{0, 1, 2\}$:

- If i is the index of a 'luminance' component:

$$\psi_i(s) = \frac{x_i(s)}{m_i(s)}. \qquad (3)$$

- If i is the index of a 'chrominance' component:

$$\psi_i(s) = x_i(s) - m_i(s). \qquad (4)$$

We classify s as shadowed point, if its $\overline{\psi}(s)$ value lies in a prescribed domain.

The efficiency of this feature selection can be observed in Fig. 1, 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. However, there is a significant overlap between the one dimensional foreground and shadow histograms, therefore, as we examine

in the next section, an efficient shadow domain description is needed.

We define the descriptor in grayscale and in the rg space similarly to eq. (3) and (4) considering that $\overline{\psi}$ will be a one and a two dimensional vector, respectively.

4. Shape of the shadow domain

The shadow domain is usually defined by a manifold having a prescribed number of free parameters, which fit the model to a given scene/situation. Previous methods use different approaches. The *domain of shadows* in the feature space is usually an interval for grayscale images (Wang et al., 2006). Regarding color scenes, this domain could be a three dimensional rectangular bin (Cucchiara et al., 2001): ratio/difference values for each channel lie between defined threshold; an ellipsoid (Mikic et al., 2000), or it may have general shape, like in (Siala et al., 2004). 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. there may be foreground points whose feature values are in the shadow domain. Therefore, the chosen shadow-domain should be not only large enough, containing 'almost all' the feature values corresponding to the occurring shadowed points, but also 'narrow' to decrease the number of the background or foreground points which are erroneously classified as shadows. 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, a corresponding problem appears: the shadow domain may alter significantly (and often rapidly) in time due to the changes in the illumination conditions, and adaptive models are needed to follow these changes. Sometimes, it is not possible to train a model with supervision regarding each forthcoming case of illumination. Therefore, those domains are preferred, which have less free parameters, and we can construct an update strategy regarding them.

For these reasons, we used an elliptical shadow domain descriptor having *parallel axes* with the xyz co-ordinate axes:

$$\text{Pixel } s \text{ is shadowed} \Leftrightarrow \sum_{i=0}^{2} \left(\frac{\psi_i(s) - a_i}{b_i} \right)^2 \leq 1, \qquad (5)$$

where $\{a_i, b_i \mid i = 0, 1, 2\}$ are the shadow domain parameters. For these parameters, a similar update procedure can be constructed to that we proposed in (Benedek & Szirányi, 2006). We explain briefly the adaption algorithm in Appendix A.

We note that with the SVM method (Siala et al., 2004), 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. For these reasons, we preferred the ellipsoid model, and in the following we examine
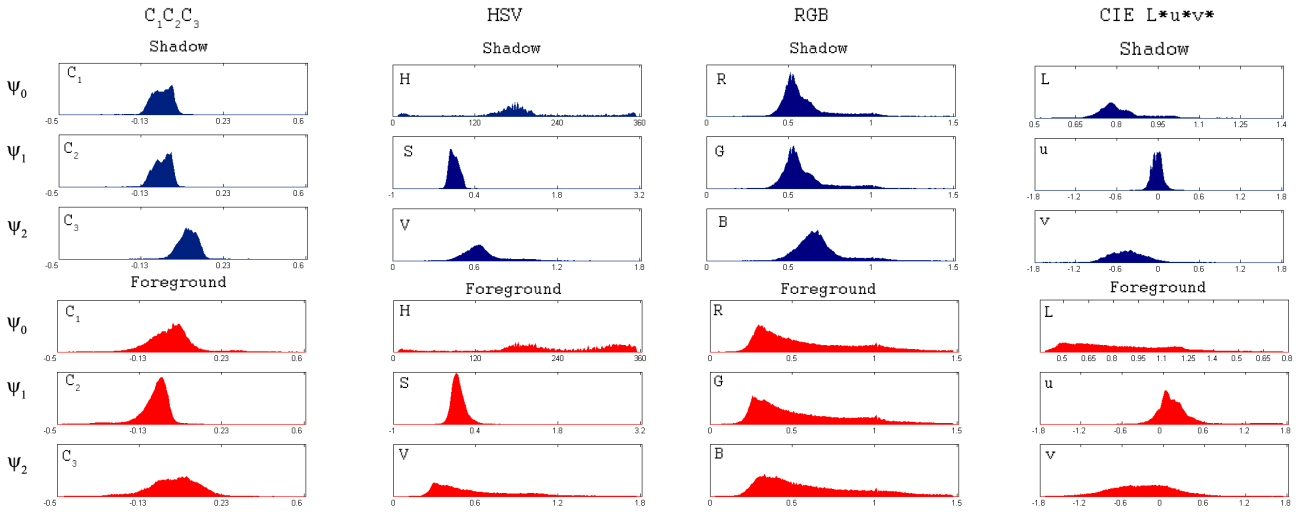
Fig. 1. One dimensional projection of histograms of foreground (red) and shadow (blue) $\overline{\psi}$ values in the 'Entrance pm' test sequence.
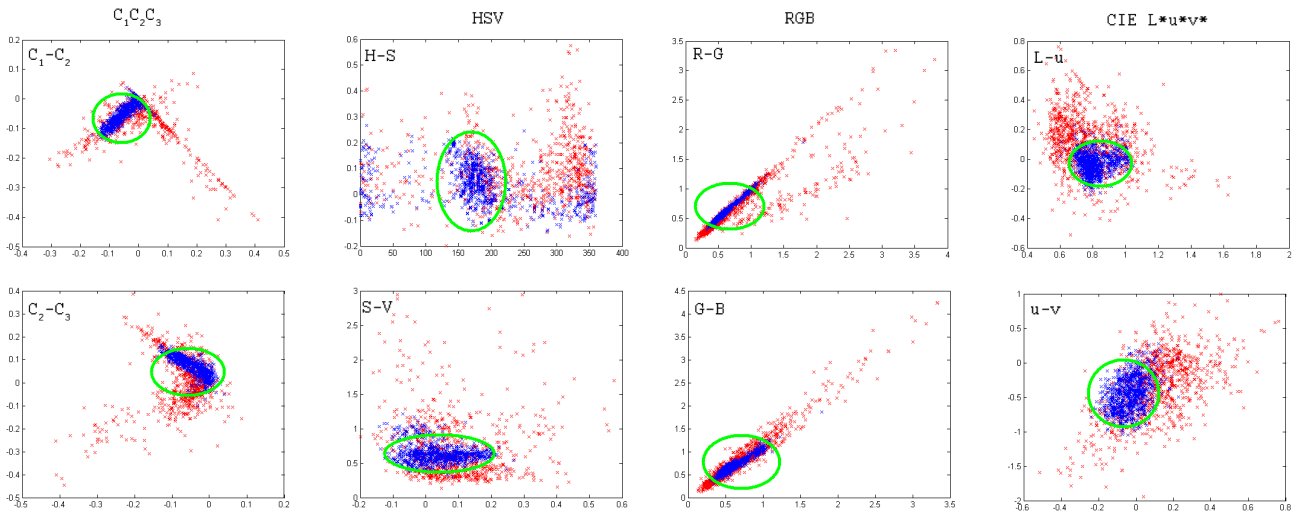

Fig. 2. Two dimensional projection of foreground (red) and shadow (blue) $\overline{\psi}$ values in the 'Entrance pm' test sequence. Green ellipse is the projection of the optimized shadow boundary.

its limits. For the sake of completeness, we note that the domain defined by eq. (5) becomes an interval if we work with grayscale images, and a two dimensional ellipse in the rg space.

We visualize the shadow domain of the 'Entrance pm' test sequence in Fig. 2, where the two dimensional projection of the occurring foreground and shadow $\overline{\psi}$ values are shown corresponding to different color space selections. We can observe that the components of vector $\overline{\psi}$ are strongly correlated in the RGB space (and also in $C_1C_2C_3$), and the previously defined ellipse cannot present a narrow boundary. (It would be better to fit an ellipse with arbitrary axes, but that choice would cause more free parameters in the system.) 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 next section, we support this statement by experimental results.

5. Evaluation

The evaluations were done through manually generated ground truth sequences regarding the following five videos:

- 'Laboratory' test sequence from the benchmark set (Prati et al., 2003). This shot contains a simple indoor environment. We generated ground truth for 20 frames.
- 'Highway' video (from the same benchmark set). This sequence contains dark shadows but homogenous background without illumination artifacts (52 frames).
- 'Entrance am' (20 frames), 'Entrance noon' (20 frames) and 'Entrance pm' (75 frames) sequences captured by the 'Entrance' (outdoor) camera of our university campus in different parts of the day. These sequences contain difficult illumination and reflection effects and suffer from sensor saturation (dark objects and shadows).

In the following, we present evaluations in two ways. In Section 5.1, we show the tentative limits of the elliptical shadow domain defined by eq. (5). The goal of these experiments is

to compare the foreground-shadow discriminating ability of the different color spaces purely based on the extracted $\overline{\psi}$ features. Here, we set the parameters manually, and do not take into consideration local connectivity or post processing. Although post processing could enhance the quality of segmentation, the improvements may depend on the scene-conditions. For example, in the 'Highway' sequence, we can estimate easily the appearing shape and size of the objects, and we may get reasonable results by specific post processing rules (Mikic et al., 2000) even if the color space selection is not optimal. However, these results may be not competent regarding other videos. Success of parameter estimation depends also on the environment rather than it would characterize the color space.

On the other hand, the above color model will be inserted into a foreground-background-shadow segmentation process for practical use (Section 5.2). Here, we suit the proposed model to an adaptive Bayesian model-framework, and show that the advantage of using the appropriate color space can be measured directly in the applications. In the second part of the experiments, we use the parameter-updating procedure of Appendix A.

5.1. Comparative evaluation with the ellipse model

In this experiment, we collect two sets of $\overline{\psi}$ values corresponding to manually marked foreground and shadowed pixels, respectively. We investigate how many pixels are classified properly by the ellipse model with different color spaces. Denote the number of correctly identified foreground pixels of the evaluation sequence by T_F. Similarly, we introduce T_S for the number of well classified shadowed points, M_F and M_S is the number of misclassified foreground, and shadowed ground truth points, respectively. First, we define the *Recall (R)* and *Precision (P)* rates of foreground detection (Sheikh & Shah, 2005):

$$\text{Recall}: R = \frac{T_F}{T_F + M_F} \qquad \text{Precision}: P = \frac{T_F}{T_F + M_S}.$$

For some optimized ellipse parameters, we plot the corresponding Precision and Recall values regarding the 'Laboratory' and 'Entrance pm' test sequences in Fig. 3. We can observe that the CIE L*a*b* and L*u*v* produce the best results in both cases (the corresponding P/R curve is 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 the least effective, while in the 'Entrance pm' sequence, the performance of the chrominance spaces is the poorest.

In the further tests, we will use the F-measure (Van Rijsbergen , 1979) which combines recall and precision in a single efficiency measure (it is the harmonic mean of precision and recall):

$$F = \frac{2 \cdot R \cdot P}{R + P}. \tag{6}$$

Video	Scene	Frames†	Dark‡	Worst	Best
Laboratory	indoor	20	0.73	gray, RGB	Luv, Lab
Entrance am	outdoor	20	0.50	gray, RGB	Luv, Lab
Entrance pm	outdoor	75	0.39	$C_1C_2C_3$, rg	Luv, Lab
Entrance noon	outdoor	20	0.35	$C_1C_2C_3$, rg	Luv, Lab
Highway	outdoor	52	0.23	$C_1C_2C_3$, rg	Luv, RGB

Table 2
Indicating the two most successful and the two less effective color spaces regarding each test sequence. We also denote †the number of frames which we evaluated through manually generated ground truth masks and ‡the mean darkening factor for shadows in grayscale.

We summarized the F rates in Fig. 4, regarding the test sequences. Also here, we can see that the CIE L*a*b* and L*u*v* spaces are 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 Lab/Luv results are similarly effective (see also Table 2). If shadow is brighter ('Entrance am', 'Laboratory'), the performance of the 'chrominance spaces' becomes reasonable, but the 'luminance spaces' are relatively poor. (In this case, shadow is characterized better by the illuminant invariant features than the luminance darkening domain). Since the hue coordinate in HSV is very sensitive to the illumination artifacts (Section 3), the HSV space is effective only in case of light-shadow.

5.2. Segmentation by using Bayesian optimization

5.2.1. Model description
The results in the previous section confirm that using the defined elliptical shadow domain, the CIE L*u*v* color space is the most effective to separate shadowed and foreground pixels only considering their colors, if we have enough training data. However, in several applications, we should consider the following facts:

- Representative ground truth foreground-shadow points are not available, the optimal ellipse parameters should be estimated somehow.
- The classification of a given pixel is usually done considering other effects than color, like neighborhood connection.

Since we want to test the color model itself, we use a Markov Random Field (MRF) optimization procedure (Geman & Geman , 1984) to get the globally optimal segmentation upon the above model.

The segmentation framework is an MRF-Potts model (Potts, 1952). An image S is considered to be a two-dimensional grid of pixels (sites), with a neighborhood system on the lattice. The procedure assigns a label ω_s to each pixel $s \in S$ form the label-set: $L = \{bg,sh,fg\}$ corresponding to the three classes: foreground (fg), back-
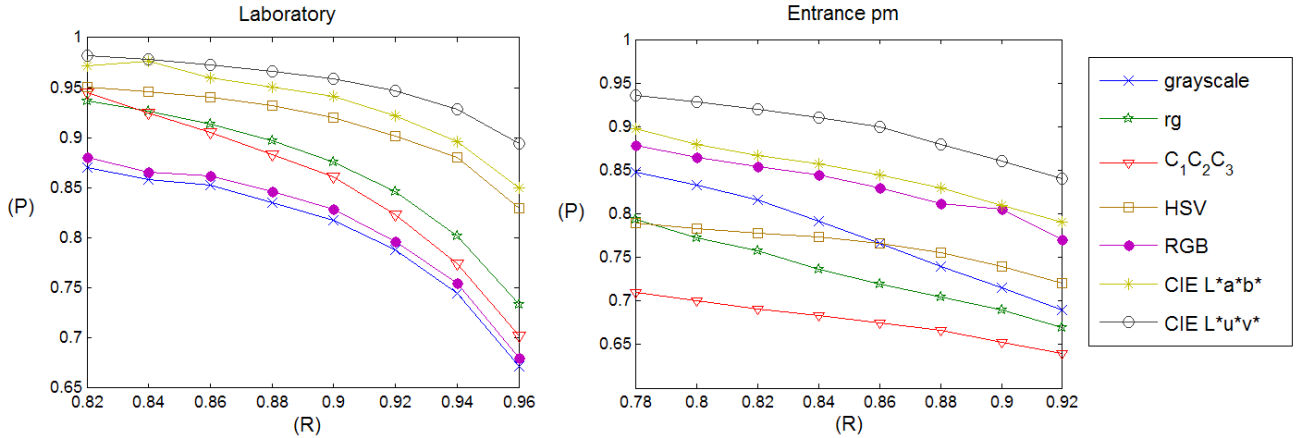
Fig. 3. **Evaluation of the deterministic model** Recall-precision curves corresponding to different parameter-settings on the 'Laboratory' and 'Entrance pm' sequences.
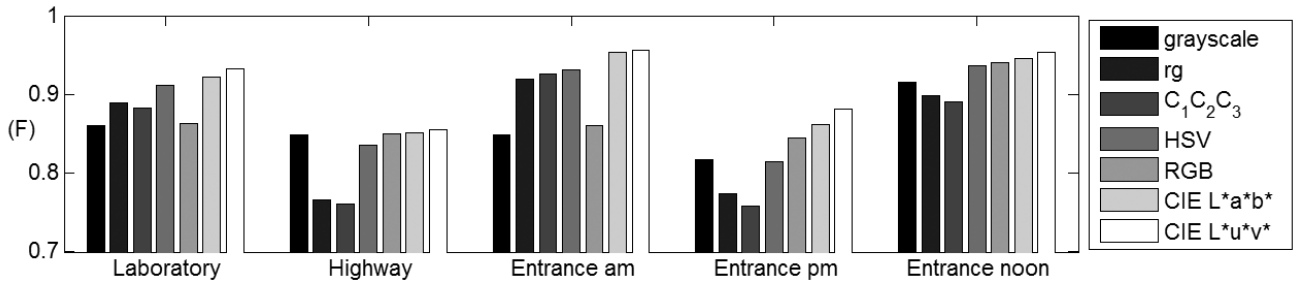

Fig. 4. **Evaluation of the deterministic model** F coefficient (eq. 6) regarding different sequences

ground (bg) and shadow (sh). Therefore, the segmentation is equivalent with a global labeling $\Omega = \{\omega_s \mid s \in S\}$. Each class at each pixel position is characterized by a conditional density function: $p_k(s) = P(x_s|\omega_s = k)$, $k \in L, s \in S$. Eg. $p_{\text{bg}}(s)$ is the probability of the fact that the background process generates the observed color value x_s at pixel s.

Following the Potts model, the optimal segmentation corresponds to the labeling which minimizes:

$$\widehat{\Omega} = \text{argmin}_\Omega \sum_{s \in S} -p_{\omega_s}(s) + \sum_{r,q \in S} \Phi(\omega_r, \omega_q), \qquad (7)$$

where the Φ term is responsible for getting smooth, connected regions in the segmented image. $\Phi(\omega_r, \omega_q) = 0$ if q and r are not neighboring pixels, otherwise:

$$\Phi(\omega_r, \omega_q) = \begin{cases} -\beta & \text{if } \omega_r = \omega_q \\ +\beta & \text{if } \omega_r \neq \omega_q \end{cases}$$

The definition of the density functions $p_{\text{bg}}(s)$ and $p_{\text{fg}}(s)$ $s \in S$ is the same, as we defined in (Benedek & Szirányi, 2006). We use a mixture of Gaussians model for the pixel values in the background, where the parameters are determined automatically (Stauffer & Grimson, 2000). Regarding the foreground, we exploit that the object-parts can usually be characterized by typical color distributions, hence, the color of a given foreground pixel matches to the color distribution of the set of the neighboring foreground pixels. Based on our experiments, this approach is more effective than using a simple uniform distribution for $p_{\text{fg}}(.)$(Rittscher et

al., 2000), and more general, since it does not need permanently high frame-rate like deriving the pixel-state transition probabilities in (Wang et al., 2006).

Before inserting our model into the previously defined MRF framework, we give to the shadow-classification step defined in Section 4, a probabilistic interpretation. We rewrite eq. (5): we match the current $\overline{\psi}(s)$ value of pixel s to a probability density function $f\left(\overline{\psi}(s)\right)$, and decide its class by:

$$\text{pixel } s \text{ is shadowed} \Leftrightarrow f\left(\overline{\psi}(s)\right) \geq t. \qquad (8)$$

The domains defined by eq. (5) and eq. (8) are equivalent, if f is a Gaussian density function (η):

$$f\left(\overline{\psi}(s)\right) = \eta(\overline{\psi}(s), \overline{\mu}_\psi, \overline{\overline{\Sigma}}_\psi) =$$

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

with the following parameters:

$$\overline{\mu}_\psi = [a_0, a_1, a_2]^T, \quad \overline{\overline{\Sigma}}_\psi = \text{diag}\{b_0^2, b_1^2, b_2^2\},$$

while

$$t = (2\pi b_0 b_1 b_2)^{-\frac{3}{2}} e^{-\frac{1}{2}}.$$

Using Gaussian distribution for the occurring feature values is also supported by the one dimensional marginal histograms in Fig. 1.

In the following, the use of the previously defined probability density functions in the MRF model is straightforward:

$$p_{\text{sh}}(s) = f\left(\overline{\psi}(s)\right).$$

The flexibility of this MRF model comes from the fact that we defined $\overline{\psi}(s)$ shadow descriptors for different color spaces differently in Section 3.

5.2.2. Test results

Fig. 5 shows the MRF-segmentation results for two frames of 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 $C_1C_2C_3$). We can observe that the CIE L*u*v* space outperforms significantly the other ones, while $C_1C_2C_3$ is very poor in cases of sharp shadows. A typical problem regarding the HSV and RGB spaces is that foreground 'glories' may appear around some dark shadow parts due to the penumbra of cast shadow (Salvador et al., 2004) and video compression. On the other hand, the proposed probabilistic model removes these artifacts with the other color spaces.

Next, we perform quantitative evaluations using the MRF model. In Section 5.1, 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 error rate between foreground and background, however, the crossover between shadow and background does not count for errors. Therefore, we modified the definition of the recall and precision rates:

$$\text{Recall}: R^* = \frac{T_F}{T_F + M_F^*} \qquad \text{Precision}: P^* = \frac{T_F}{T_F + M_S^*},$$

where M_F^* is the number of misclassified foreground points (i.e. foreground pixels classified as background *or* shadow), and M_S^* is the number of misclassified non-foreground points of the ground truth sequence. We derive F^*-measure similarly to eq. (6), with using R^* and P^* rates. We observe in Fig. 4 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 5.1. The reason for these differences is that due to the composite foreground model, MRF neighborhood conditions and errors in parameter estimation, the artifacts may appear differently in the different sequences. Therefore, we consider the numerical results from Section 5.1 to be more relevant to compare the capabilities of the color spaces for shadow separation. However, the experiments of this section confirm that appropriate color space selection is crucial in the applications, and the CIE L*u*v* space is preferred for this task.

5.2.3. Model validation

The main goal of this paper has been to find an appropriate color space for efficient shadow separation with a low number of free model-parameters. In Section 2, 3 and 4, we referred to the appearing problems of models which have
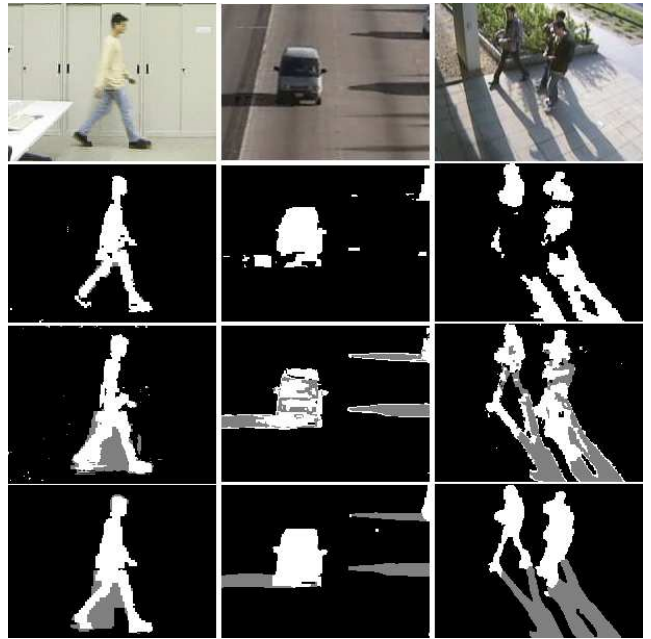

Fig. 7. Segmentation using different methods. Row 1: video frames, Row 2: segmentation with the $C_1C_2C_3$ based illumination-invariant method (Salvador et al., 2004), Row 2: a three-parameter method (Mikic et al., 2000), Row 3: proposed method with CIE L*u*v* color space selection

less than 2 free parameters for one color channel. Some corresponding artifacts are demonstrated in Fig. 7. For comparison with the proposed 6 (=3 × 2)-parameter model, we implemented a *nonparametric* method based on (Salvador et al., 2004), and a 3-parameter-model (Mikic et al., 2000). Although in the indoor 'Laboratory' sequence, all the three methods are similarly efficient, our model outperforms the other ones outdoors, especially in the surveillance shot. Note that the preliminary version of the proposed approach has been already compared to the state-of-the-art in (Benedek & Szirányi, 2006).

6. Conclusion

This paper examined the color modelling problem of shadow detection. We developed a model framework for this task, which can work with different color spaces. Meanwhile, the model can detect shadows under significantly different scene conditions and it has a few free parameters which is advantageous in practical point of view. In our case, the transition between the background and shadow domains is described by statistical distributions. With this model, we compared several well known color spaces, and observed that the appropriate color space selection is an important issue regarding the segmentation results. We validated our method on five 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 show that CIE L*u*v* color space is the most efficient both in the color based clustering of the individual pixels and in the case of Bayesian
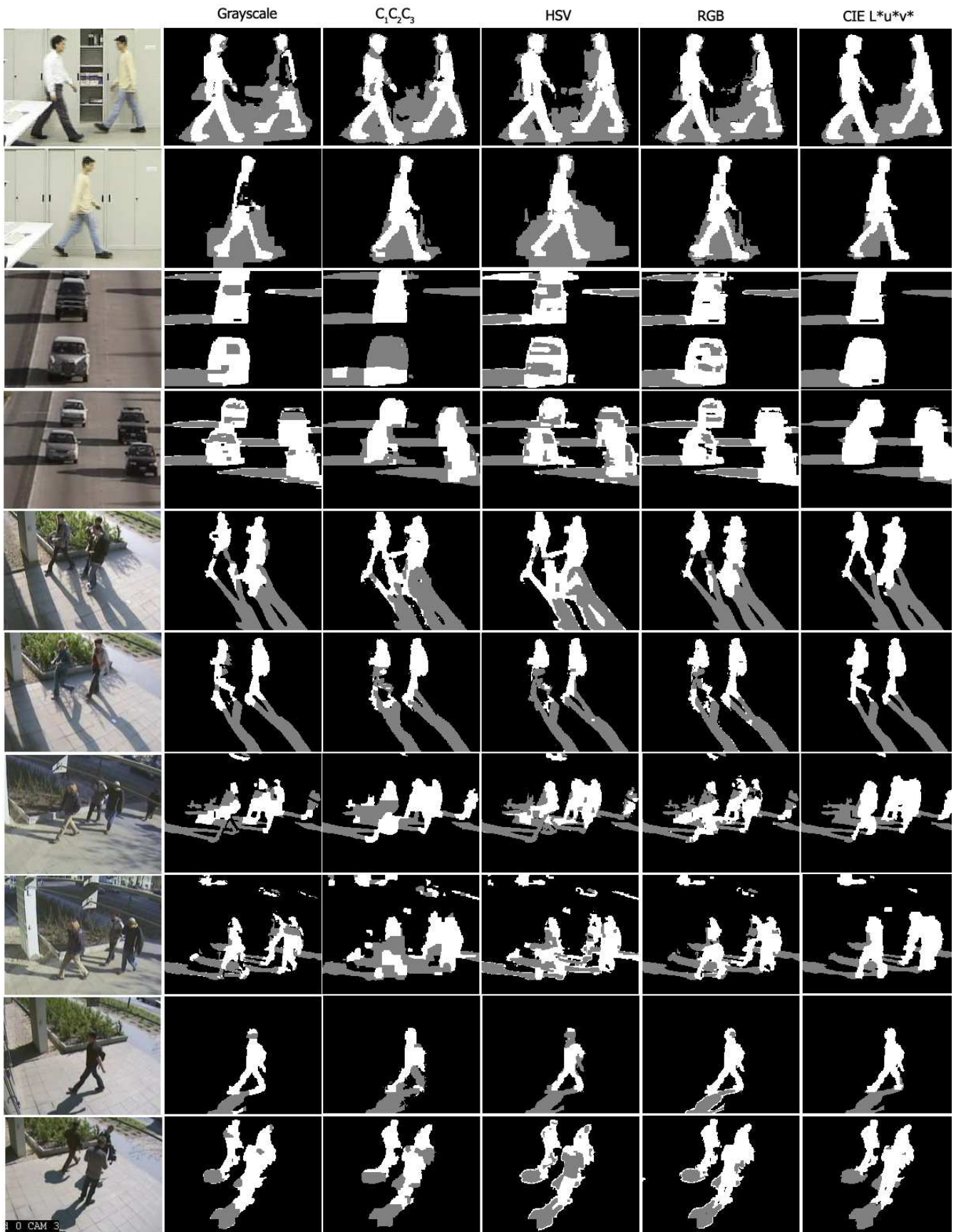
Fig. 5. MRF segmentation results with different color models. Test sequences (up to down): rows 1-2 'Laboratory', rows 3-4: 'Highway', rows 5-6: 'Entrance am', rows 7-8: 'Entrance pm', rows 9-10: 'Entrance noon'.
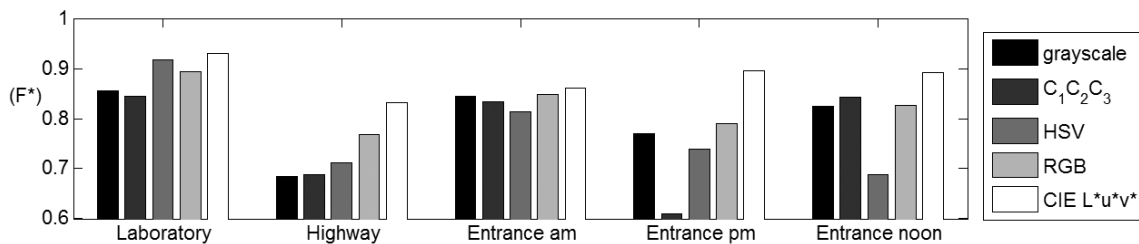
Fig. 6. **Evaluation of the MRF model** F^* coefficient regarding different sequences

foreground-background-shadow segmentation.

7. Acknowledgement

Appendix A. Adaption of the shadow parameters

Consider the shadow domain which can be described with the parameters of eq. (5). We found in the experiments, that the b_i parameters and the appropriate a_i-values corresponding to the 'chrominance' components are approximately constant in time letting us to estimate them once in a scene. However, the mean darkening ratio (a_i) of the 'luminance' components may change significantly between sunlit (0.3) and overcast weather (> 0.9), and these changes are often rapid. After a sudden change, the shadow parameters from the previous moments are not more reliable. Let $l \in \{0, 1, 2\}$ denote a 'luminance' channel. For a given image we may collect histogram from the $\psi_l(.)$ values of those pixels, which are marked as non background point by the adaptive background subtraction algorithm (Stauffer & Grimson, 2000). If the image contains considerable shadowed parts, a peak appears in the histogram near the desired a_l value. Figure A.1 shows 3 typical situations from the video 'Entrance pm', where the optimal a_l was definitely 0.68. On the first image, a large shadow is observable, and the peak in the histogram is very significant. On the second one, the peak is still in the right place, however it is smaller. On the third image there is small shadow and the histogram is flat. Denote $h[t]$ the location of the peak in the histogram of the t-th frame of the sequence, $v[t]$ is the maximum value, $\overline{v}[t]$ is the average value. $h[t]$ can be a good estimation for a_l, if peak-value $v[t]$ is high and significant: $\frac{v[t]}{\overline{v}[t]}$ is high. We define the update process by the following:

$$a_l[t+1] = \rho \cdot h[t] + (1 - \rho) \cdot a_l[t], \qquad \rho = \alpha \cdot v[t] \cdot \frac{v[t]}{\overline{v}[t]}$$

where $\alpha = 0.001$ is a constant factor, and we perform the parameter update only, if there are enough non-background points in the image.
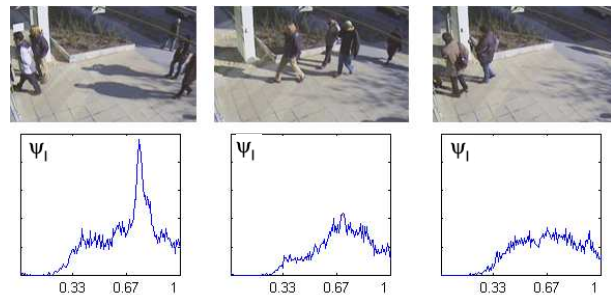

Fig. A.1. Three images from sequence 'Entrance pm' and the corresponding histograms for the ψ_l values of the non-background pixels, where l is a 'luminance' channel.

References

Cs. Benedek, and T. Szirányi, Markovian Framework for Foreground-Background-Shadow Separation of Real World Video Scenes, Proc. Asian Conference on Computer Vision (ACCV 2006), LNCS 3851, pp. 898907, Jan. 2006.

Cs. Benedek, T. Szirányi: Color Models of Shadow Detection in Video Scenes, In Proc. VISAPP, Barcelona, Spain, 2007, accepted.

A. Cavallaro, E. Salvador, and T. Ebrahimi, Detecting Shadows in Image Sequences, Proc. of IEEE Conference on Visual Media Production (CVMP), March 2004.

R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati, The Sakbot System for Moving Object Detection and Tracking, Video-Based Surveillance Systems-Computer Vision and Distributed Processing, pp. 145-157, 2001.

G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, On the Removal of Shadows From Images, IEEE Trans. PAMI, Vol.28, No.1, pp.59-68, 2006.

D. A. Forsyth: A novel algorithm for color constancy, International Journal of Computer Vision, 5(1):5-36, August 1990.

C. Fredembach and G. D. and Finlayson, Hamiltonian path based shadow removal, In Proc. BMVC, 2005.

S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattern Analysis and Machine Intelligence, pp. 721-741, 1984.

T. Gevers and H. Stokman, Classifying color edges in video into shadow-geometry, highlight, or material transitions, IEEE Trans. on Multimedia, vol 5, issue 2, pp 237–243, 2003.

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.

L. Havasi and T. Szirányi, Estimation of Vanishing Point in Camera-Mirror Scenes Using Video, Optics Letters, Vol. 31, No. 10, pp: 1367-1566, 2006.

E. A. Khan and E. Reinhard, Evaluation of Color Spaces for Edge Classification in Outdoor Scenes, IEEE International Conference on Image Processing, Genova, Italy, September 11-14, 2005.

D. K. Lynch and W. Livingstone, Color and Light in Nature, Cambridge University Press, 1995.

N. Martel-Brisson, and A. Zaccarin, Moving Cast Shadow Detection

from a Gaussian Mixture Shadow Model, IEEE CVPR 2005, Vol. 20-26, No. 2, pp. 643 - 648, June 2005.

I. Mikic, P. Cosman, G. Kogut and M. M. Trivedi, Moving Shadow and Object Detection in Traffic Scenes, Proc. ICPR, vol 1, pp. 321-324, Sept 2000.

N. Paragios and V. Ramesh, A MRF-based Real-Time Approach for Subway Monitoring, Proc. IEEE Conference in Computer Vision and Pattern Recognition (CVPR 2001), pp. I:1034-1040, 2001.

R. Potts, Some generalized order-disorder transformation, Proceedings of the Cambridge Philosophical Society, 48(106), 1952.

A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, Detecting moving shadows: algorithms and evaluation, IEEE Trans. Pattern Analysis and Machine Intelligence, No. 7, pp. 918–923, July 2003.

M. Rautiainen, T. Ojala, H. Kauniskangas, Detecting Perceptual Color Changes from sequential images for scene surveillance IEICE Transactions on Information and Systems E84-D:1676 - 1683, 2001

C. J. Van Rijsbergen, Information Retrieval, 2nd edition, London, Butterworths.

J. Rittscher, J. Kato, S. Joga and A. Blake, A Probabilistic Background Model for Tracking Proc. European Conf. on Computer Vision 2000.

E. Salvador, A. Cavallaro, and T. Ebrahimi, Cast shadow segmentation using invariant color features,Comput. Vis. Image Underst. 95, 2 pp. 238-259, Aug. 2004.

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, November 2005.

K. Siala, M. Chakchouk, F. Chaieb, and O.Besbes, Moving shadow detection with support vector domain description in the color ratios space, Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), 2004.

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.

M. Tkalcic and J. Tasic, Colour spaces - perceptual, historical and applicational background, Proc. Eurocon 2003, Sept. 2003.

Y. Wang, T. Tan, Adaptive Foreground and Shadow Detection in Image Sequences, Proceedings of the 16th International Conference on Pattern Recognition, pp: 983-986, 2002.

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, 28(2):279-89, 2006.

G. Wyszecki and W. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulas, 2nd edition, Wiley, 1982.

A. Yoneyama, Chia H. Yeh, C.-C. Jay Kuo, Moving Cast Shadow Elimination for Robust Vehicle Extraction Based on 2D Joint Vehicle/Shadow Models, IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.

Y. Zhou, Y. Gong, and H. Tao, Background segmentation using spatial-temporal multi-resolution MRF, IEEE Motion05, January 2005.

S. C. Zhu and A. L. Yuille, A Flexible Object Recognition and Modeling System, Int'l Journal of Computer Vision, Vol. 20, No. 3, Dec, 1996.