Moving Cast Shadow Removal Based on Local Descriptors

Rui Qin, Shengcai Liao, Zhen Lei, and Stan Z. Li*

Center for Biometrics and Security Research,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
{rqin, scliao, zlei, szli}@cbsr.ia.ac.cn

Abstract

Moving cast shadow removal is a difficult problem in video analysis. This paper presents a novel algorithm for detection of moving cast shadows, based on a local texture descriptor called Scale Invariant Local Ternary Pattern (SILTP). An assumption is made that the texture properties of cast shadows bears similar patterns to those of the background beneath them. The likelihood of cast shadows is derived using information in both color and texture. An online learning scheme is employed to update the shadow model adaptively. Finally, the posterior probability of cast shadow region is formulated by further incorporating prior contextual constrains using a Markov Random Field (MRF) model. The optimal solution is found using graph cuts. Experimental results tested on various scenes demonstrate the robustness of the algorithm.

1. Introduction

Extracting moving objects is one of the key problems in video analysis applications, including visual surveillance, content-based video retrieval, etc. The problem is further plagued by moving cast shadows caused by e.g. sunlight. Misclassification of moving cast shadows as parts of foreground objects usually induces problems, such as silhouette distortions and merging of nearby objects, and hence mistakes in subsequent stages. An effective moving shadow detection method is therefore necessary for accurate extraction of moving objects.

There are a number of cues that provide information regarding properties and behavior of cast shadows. A direct way for modeling cast shadows is based on the assumption that shadow pixels should have lower luminance and the same chrominance as the corresponding background. This attenuation property has been employed in different color spaces like *RGB* [5] and *HSV* [3]. Unfortunately, such assumptions are difficult to justify in general, especially when pixels of foreground objects are darker than the reference surface they cover. Furthermore, it is not reliable to exploit only the color information of isolated point. Therefore, in addition to color properties, texture or gradient information extracted from the spatial domain is used to detect cast shadows [7, 13]. Some physical models [2] are also used to model cast shadows. The major limitation of these methods is that they require offline training and often estimate parameters again for each new scene. Hence, they cannot handle complex conditions like time-varying lighting. A comprehensive study of moving cast shadow detection approaches can be found in [11].

Recently, online approaches have been developed to learn moving cast shadows [6, 9, 10] in color space adaptively. Compared to the complexity and variability of cast shadows in color spaces, the distribution of texture differentia is relatively simple, hence we propose to also update the cast shadow model online in the texture space.

In this paper, we propose a novel method for shadow detection, using a local texture descriptor called *Scale Invariant Local Ternary Patterns (SILTP)* [8]. Global properties of cast shadows in both texture and color domains are learned through the use of *Mixture of Gaussian*, with an *online-EM* update scheme. Contextual constraint from *MRF* modeling is further incorporated to obtain the *MAP* estimation of the cast shadows. Experimental results demonstrate the effectiveness and robustness of the proposed method. The contributions are the following: Firstly, *SILTP* is used as a local texture descriptor for cast shadow detection, which can deal with the sudden changes of gray scale intensities caused by environmental illumination variations. Secondly, an online learning scheme is introduced to shadow learning process in both texture and color space, which makes the proposed method more robust to
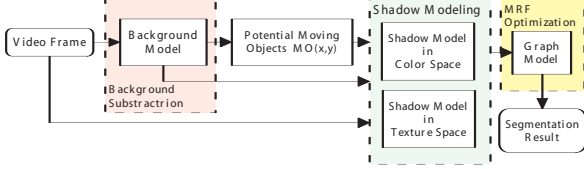
---

*Corresponding author.

Figure 1. Flow diagram of the algorithm

changes in environments.

2 Learning Cast Shadows

A flow diagram of the proposed algorithm is illustrated in Figure 1. For each pixel p, a background model is learned by the nonparametric *KDE* method in the RGB color space [4], from which the foreground probability can be estimated. After that, potential moving objects are segmented, and within it we evaluate the likelihood probability of cast shadows over both the color and texture domain as follows

$$P(MP|S,p) = \sum_{i=1,2} P(MP|D_i, S, p)P(D_i|S, p) \tag{1}$$

where MP denotes potential moving pixels, S denotes shadow, D_1 and D_2 represent the texture and color domains respectively. The estimation details are described in the following.

2.1 Shadow Model in Texture Space

With the assumption that the texture within the cast shadow tends to be similar with that in the corresponding background surface, in this work we propose to learn a texture shadow model to discriminate the shadow from moving objects and update it online.

Tan in [12] proposed a local image texture descriptor called *Local Ternary Pattern (LTP)* for face recognition. It is robust to image noises but not invariant to gray-scale changes. However, in practice, for surveillance scenario, there always exist sudden changes of gray scale intensities due to environmental illumination variations such as shadow. To address this problem, we extend the original *LTP* to the intensity scale invariant *LTP (SILTP)* as in [8] for handling cast shadows.

As shown in Figure 2, for any pixel location (x_c, y_c), *SILTP* can be encoded as

$$SILTP_{N,R}^{\tau}(x_c, y_c) = \bigoplus_{k=0}^{N-1} s_\tau(I_c, I_k), \tag{2}$$

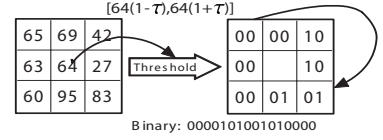where I_c is the gray intensity value of the center pixel, $I_k(k = 0, 1, ...N-1)$ are that of its N neighborhood

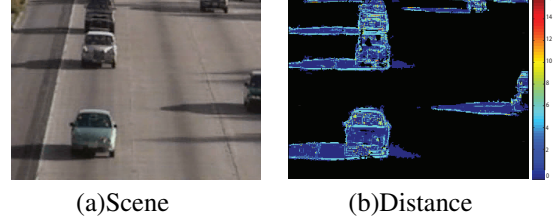**Figure 2. The SILTP operator**

(a)Scene   (b)Distance

Figure 3. Hamming distance of SILTP

pixels equally spaced on a circle of radius R^1, $\bigoplus$ is defined as concatenation operator of binary strings, and s_τ denotes a piecewise function defined as

$$s_\tau(I_c, I_k) = \begin{cases} 01, & if\ I_k > (1+\tau)I_c, \\ 10, & if\ I_k < (1-\tau)I_c, \\ 00, & otherwise. \end{cases} \tag{3}$$

In real scenarios, illumination variations always make the gray intensities of neighboring pixels to be changed simultaneously, from brighter to darker or conversely, which approximately causes a scale transform on neighboring pixels with a constant factor. In this case the proposed *SILTP* can well encode the illumination-invariant textures. Figure 3 shows the *Hamming distance of SILTP* between the potential moving objects of a frame and the corresponding background. As can be seen from Figure 3, the cast shadow regions are more similar with the the corresponding backgrounds (with lower distances), except that the boundaries have higher distances. Thus we apply *Gaussian mixture model (GMM)* with two states to learn a universal likelihood distribution of such distance as our shadow model in texture space. Consequently, the likelihood probability $P(MP|D_1, S, p)$ of a pixel p being moving cast shadow can be evaluated by the learned texture shadow model.

The *Expectation Maximization (EM)* algorithm is adopted to estimate the parameters of *GMM* from different scenes. Moreover, *online-EM* is employed to update the universal *GMM* model automatically for a specific scene in real-time video. Since the distribution of the distance based shadow likelihood probability in texture space is usually simple for various scenes, the *Online-EM* based adaptation can converge very quickly.

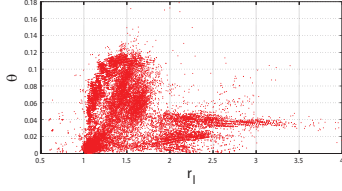[1]In this work, N=8 and R=1 are used for *SILTP*.

Figure 4. Distribution of (r_l, θ)

2.2 Shadow Model in Color Space

Figure 3 shows that SILTP can represent shadows similar with the corresponding backgrounds, thus discriminates them from moving objects. Yet it also shows that with SILTP some flat surfaces of moving objects are also similar with the flat background regions. However, in this case the surface colors of the two are different. Therefore, we also learn a color shadow model for a complement with textures.

Porikli and Thomton showed that in RGB color space shadow can be defined as a conic volume around the corresponding background [10]. Following their work, we also learn a shadow model in RGB color space. For a moving pixel *p*, the relationship of the observation pixel vector $\mathbf{z}_t(p)$ and the corresponding background pixel vector $\mathbf{b}_t(p)$ can be characterized by two parameters [10]: luminance ratio $r_l(p)$ and angle variation $\theta(p)$, which are defined as follows :

$$r_l(p) = \frac{\|\mathbf{b}_t(p)\|}{\|\mathbf{z}_t(p)\| cos(\theta(p))}, \qquad (4)$$

$$\theta(p) = arccos(\frac{< \mathbf{z}_t(p), \mathbf{b}_t(p) >}{\|\mathbf{z}_t(p)\| \cdot \|\mathbf{b}_t(p)\|}), \qquad (5)$$

where $\| \cdot \|$ is the norm of a vector, and $<, >$ is the inner product operator. Figure 4 illustrates the distribution of (r_l, θ) collected from shadows of some scenarios. It can be seen that the two parameters fall within several clusters. Therefore, we adopt *GMM* with five components to learn the above parameter distribution as a color shadow model. We also apply *EM* algorithm first to learn a universal *GMM* model with (r_l, θ) samples over shadows of various scenarios. And then, for a real-time video of a specific scene, we update it automatically by online learning based on *Online-EM* algotithm. Finally, the likelihood probability $P(MP|C_2, S, p)$ of cast shadows in color space is estimated by the updated *GMM* model.

3 Segmentation for the cast shadow

In the likelihood probability map of cast shadows, if we deal with each pixel independently, the segmen-

tation results may contain many small pieces. Consequently, we build the likelihood probability into an *MR-F* energy function which considers neighboring smooth information that will refine the final segmentation. The energy is defined as

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{p,q \in \mathcal{N}} V_{p,q}(f_p, f_q), \quad (6)$$

where $E(f)$ is the energy of a particular shadow/foreground labeling f, p and q are indexes over the pixels, $D_p(f_p)$ is the data cost of assigning the pth pixel to label f_p, and $V_{p,q}(f_p, f_q)$ represents the smoothness cost of assigning pixels p and q in a neighborhood $\mathcal{N}$ to respective labels f_p and f_q. In this work, the data cost assigning shadow is set as $-logP(MP|S, p)$, while that assigning foreground is defined as $log(1-\alpha P(MP|S, p))$, where α is a weighting factor. The smoothness cost term is defined as

$$V_{p,q} = (f_p - f_q)^2 e^{-\beta|I_p - I_q|} \qquad (7)$$

where I_p and I_q denote gray-scale intensities of pixels p and q, $|\cdot|$ denotes absolute difference, and β is a constant. To minimize the energy function of Equ. (6), we apply the *graph cut* algorithm [1] for an approximate *MAP* estimation of the labeling field, and hence obtain the final segmentation result.

4 Experiment Results

We run experiments on three benchmark video sequences containing moving cast shadows [2] to evaluate the effectiveness of the proposed method. Figure 5 illustrates the visual results of our method on these sequences. As shown in Figure 5, moving cast shadows can be almost completely detected by our approach, except for some thin mistakes presented around the boundary of cast shadows in the outdoor sequences. For the indoor sequence, the soft cast shadows of moving objects can be removed better by the texture descriptor *SILTP*. Also notice that, thanks to the new descriptor, the moving highlight reflected on the road is also removed (see Figure 5(1)c).

For a quantitative evaluation, we calculate the accuracy of the cast shadow detection by using two metrics proposed in [11]. The *shadow detection rate* η measures the percentage of correctly labeled shadow pixels among all detected ones, while the *shadow discrimination rate* ξ measures the discriminative power between foregrounds and shadows. The quantitative comparison with both the proposed and previous approaches are

---

[2] http://cvrr.ucsd.edn/aton/shadow.
http://vision.gel.ulaval.ca/ CastShadows

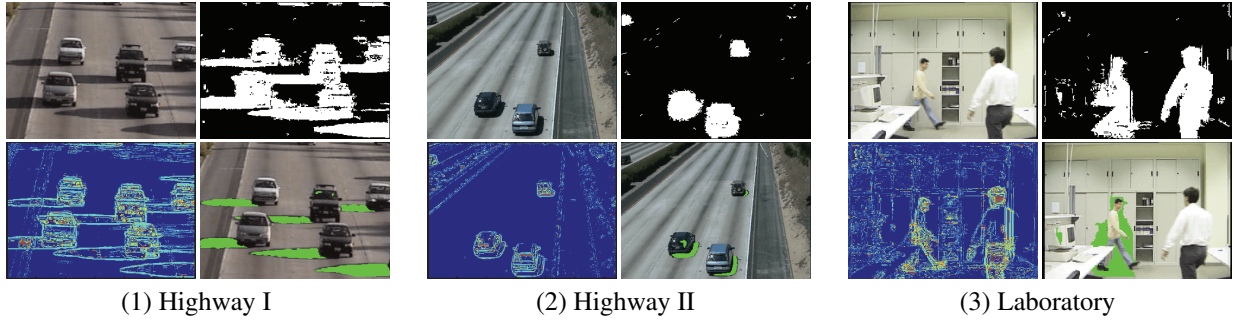(1) Highway I   (2) Highway II   (3) Laboratory

Figure 5. Visual results in various environments. $\frac{a|b}{c|d}$ (a) Frame from video sequence. (b) Potential foreground labels. (c) Hamming distance of SILTP. (d) Final result using MRF

Table 1. Quantitative evaluation results

Sequence	Highway I		Highway II		Hallway	
Method	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
Proposed	72.51	84.90	75.38	74.12	82.31	91.07
Physics[6]	70.83	82.37	76.50	74.51	82.05	90.47
Kernel[2]	70.50	84.40	68.40	71.20	72.40	86.70
GMSM[9]	63.30	71.30	58.51	44.40	60.50	87.00

given in Table 1. The results for other's approaches are taken directly from [6][2][9]. From Table 1, we can see that the proposed method achieves comparable performance as the state-of-the-art algorithms in the literature. Due to the illumination invariant texture descriptor *SILTP*, our approach performs better in indoor scene like *Hallway*, and outdoor scenario with large cast shadow regions, such as *HighwayI*.

5 Conclusions and Future work

A novel method for moving cast shadow removal is presented in this paper. Color and texture information using *SILTP* are built into a *MRF* energy function. Additionally, with the aid of *online-EM* process, the shadow model is updated dynamically. Qualitative and quantitative evaluation in various experiments validate the effectiveness of our method. Moreover, our method performs better in the indoor scenarios. The proposed pixel-based method is suitable for parallel computing, therefore it can be accelerated by multi-core and *GPU* implementations, which will be one of our future work.

References

[1] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. PAMI*, 26:359–374, 2004.

[2] N. M. Brisson and A. Zaccarin. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation. In *CVPR*, 2008.

[3] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti. Improving shadow suppression in moving object detection with hsv color information. In *Intelligent Transportation Systems, Proceedings of the IEEE*, pages 334–339, 2001.

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

[5] T. Horprasert, D. Harwood, and L. S. Davis. A robust background subtraction and shadow detection. In *ACCV*, 2000.

[6] J. B. Huang and C. S. Chen. Moving cast shadow detection using physics-based features. In *CVPR*, 2009.

[7] R. Leone, C. Distante, and F. Buccolieri. A shadow elimination approach in video-surveillance context. *PR Letters*, 27:345–355, 2006.

[8] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li. "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes". In *Computer Vision and Pattern Recognition*, San Francisco, CA, USA, June 2010.

[9] N. M-Brisson and A. Zaccarin. Learning and removing cast shadows through a multidistribution approach. *IEEE Trans. PAMI*, 29:1133–1146, 2007.

[10] F. Porikli and J. Thornton. Shadow flow: A recursive method to learn moving cast shadows. In *ICCV*, 2005.

[11] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Trans. PAMI*, 25:918–923, 2003.

[12] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *AMFG*, 2007.

[13] W. Zhang, X. Z. Fang, and X. Yang. Moving cast shadows detection based on ratio edge. In *ICPR*, 2000.