

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 100190, China
 {rqin, scliao, zlei, szli}@cbsr.ia.ac.cn

Abstract—Moving cast shadow removal is an important yet difficult problem in video analysis and applications. This paper presents a novel algorithm for detection of moving cast shadows, that 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 constraints 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.

Keywords-shadow removal, local texture descriptor, Markov Random Field

I. 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 behaviors 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* [1], *HSV* [2]. 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 [3] or gradient [4] information extracted from the spatial domain is used to detect cast shadows. Some physical models [5] are also used to model cast shadows. The major limitation of these methods is that they often require off-line training and need to re-estimate

parameters for each new scene. Consequently, they cannot handle complex conditions, such as time-varying lighting conditions, etc. A comprehensive study of moving cast shadow detection approaches can be found in [6].

Recently, online approaches have been developed to learn moving cast shadows [7], [8], [9] in color space adaptively. Compared with the complexity and variability of cast shadows in color spaces, the distribution of texture differentia is relatively simple, hence we propose to 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)*. 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 *Markov Random Field (MRF)* [10] 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 as follows: 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 changes in environments.

II. 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 [11], from which the foreground probability can be estimated. Potential moving objects can be extracted by simply thresholding this density distribution, and within the segmentation the likelihood probability of cast shadows can be evaluate over both the color and texture domain as follows

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

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

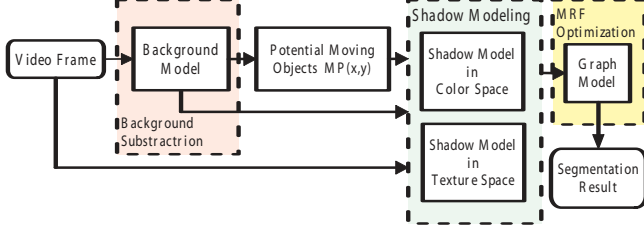


Figure 1. Flow diagram of the algorithm

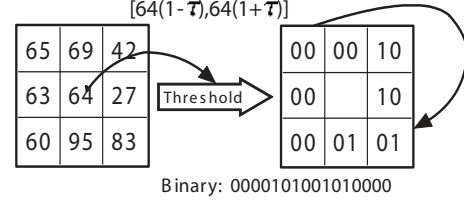


Figure 2. The SILTP operator

the following subsections.

A. Shadow Model in Texture Space

Under 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 dynamically.

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)* for handling cast shadows.

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

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

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

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

Since each comparison can result in one of three values, we encode *SILTP* with two bits, leaving the value of “11” undefined. τ is determined by the noise in the scene. The intensity scale invariant property can be easily verified from Equ. (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*

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

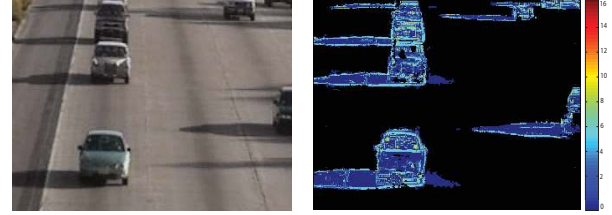


Figure 3. Hamming distance of SILTP

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. Therefore, 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 this universal *GMM* model dynamically 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.

B. Shadow Model in Color Space

Figure 3 shows that *SILTP* can represent the similarity between shadows and the corresponding backgrounds, which can be employed to discriminate cast shadows 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 as a complement for the previous texture discrimination.

Porikli and Thomton showed that in RGB color space shadow can be defined as a conic volume around the corresponding background [9]. Following their work, we also learn a shadow model in RGB color space. For a moving

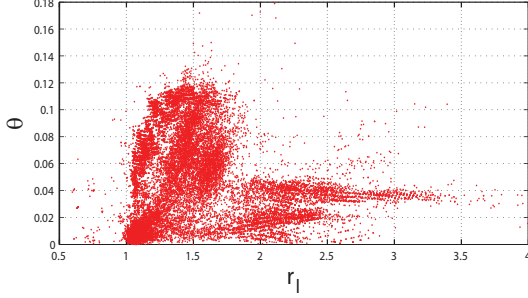


Figure 4. Distribution of (r_l, θ)

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 [9]: luminance ratio $r_l(p)$ and angle variation $\theta(p)$, which are defined as follows :

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

$$\theta(p) = \arccos\left(\frac{\langle \mathbf{z}_t(p), \mathbf{b}_t(p) \rangle}{\|\mathbf{z}_t(p)\| \cdot \|\mathbf{b}_t(p)\|}\right), \quad (5)$$

where $\|\cdot\|$ is the norm of a vector, and $\langle \cdot, \cdot \rangle$ 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. The *EM* algorithm also apply to learn a universal *GMM* model with (r_l, θ) samples over shadows of various scenarios. Then, for a real-time video of a specific scene, we update it automatically by online learning based on *Online-EM* algorithm. Finally, the likelihood probability $P(MP|D_2, S, p)$ of cast shadows in color space is estimated by the updated *GMM* model.

III. SEGMENTATION FOR THE CAST SHADOW

In the likelihood probability map of cast shadows, if we deal with each pixel independently, the segmentation results may contain many small pieces. Consequently, we build the likelihood probability into an *MRF* energy function [10] which considers neighboring smooth information that will refine the final segmentation. The energy function 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 p^{th} 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 $-\log P(MP|S, p)$, 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

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

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 [13] for an approximate *MAP* estimation of the labeling field, and hence obtain the final segmentation result.

IV. EXPERIMENT RESULTS

The results presented here are evaluated from challenging video sequences known in the literature². We run experiments only on three benchmark video sequences to evaluate the effectiveness of the proposed method, where the quantitative accuracy of other comparison are available. 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*. We can also notice that, thanks to the new descriptor, the moving highlight reflected on the road is also removed (see Figure 5(1)b).

For a quantitative evaluation, we calculate the accuracy of the cast shadow detection by using two metrics proposed in [6]. 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 given in Table I. The results of other's approaches are taken directly from [7][5][8]. From Table I, we can see that the proposed method achieves comparable performance as the state-of-the-art algorithms in the literature. By using the illumination invariant texture descriptor *SILTP*, our approach performs better in the indoor scene like *Hallway*, and the outdoor scenario with large cast shadow regions, such as *HighwayI*.

Table I
QUANTITATIVE EVALUATION RESULTS

Sequence	Highway I		Highway II		Hallway	
	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
Proposed	72.51	84.90	75.38	74.12	82.31	91.07
Physics[7]	70.83	82.37	76.50	74.51	82.05	90.47
Kernel[5]	70.50	84.40	68.40	71.20	72.40	86.70
GMSM[8]	63.30	71.30	58.51	44.40	60.50	87.00

²<http://cvrr.ucsd.edu/aton/shadow>.
<http://vision.gel.ulaval.ca/CastShadows>



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

V. ACKNOWLEDGEMENT

This work was supported by the Chinese National Hi-Tech (863) Program Projects #2008AA01Z124, and the AuthenMetric R&D Fund.

VI. 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] T. Horprasert, D. Harwood, and L. S. Davis, "A robust background subtraction and shadow detection," in *ACCV*, 2000.
- [2] 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*, 2001, pp. 334–339.
- [3] R. Leone, C. Distanto, and F. Buccolieri, "A shadow elimination approach in video-surveillance context," *PR Letters*, vol. 27, pp. 345–355, 2006.
- [4] W. Zhang, X. Z. Fang, and X. Yang, "Moving cast shadows detection based on ratio edge," in *ICPR*, 2000.
- [5] 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.
- [6] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Trans. PAMI*, vol. 25, pp. 918–923, 2003.
- [7] J. B. Huang and C. S. Chen, "Moving cast shadow detection using physics-based features," in *CVPR*, 2009.
- [8] N. M-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *IEEE Trans. PAMI*, vol. 29, pp. 1133–1146, 2007.
- [9] F. Porikli and J. Thornton, "Shadow flow: A recursive method to learn moving cast shadows," in *ICCV*, 2005.
- [10] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 2009.
- [11] 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*, vol. 90, no. 7, pp. 1151–1163, Jul 2002.
- [12] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *AMFG*, 2007.
- [13] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. PAMI*, vol. 26, pp. 359–374, 2004.