2021 IEEE Intelligent Transportation Systems Conference (ITSC)
Indianapolis, USA. September 19-21, 2021

A New Online Approach for Moving Cast Shadow Suppression in Traffic Videos

Hadi Ghahremannezhad[1], Hang Shi[2], and Chengjun Liu[3]

Abstract— **In applications of traffic video analysis, moving vehicles can induce cast shadows that have negative impacts on the system performance. Here, a new online cast shadow removal method is proposed which integrates pixel-based, region-based, and statistical modeling techniques to detect shadows. Specifically, the global foreground modeling(GFM) method is first applied in order to segment the moving objects along with their cast shadows from the stationary background. The potential shadow pixels are identified by considering the physics-based properties of reflection and comparing the changes in color values in the corresponding background and foreground locations in terms of brightness and chromaticity. A new region-based shadow detection method is proposed using an illumination invariant feature as the input to the k-means clustering method in order to partition each foreground component into separate segments. Each segment is classified into object and shadow based on its portion of potential shadows, the amount of gradient information introduced, and the number of extrinsic terminal points contained. Afterward, the background and foreground values in the RGB and HSV color-spaces are utilized to construct six-dimensional feature vectors which are modeled by a mixture of Gaussian distributions to classify the foreground pixels into shadows and objects. Lastly, the results of the previous steps are integrated for final shadow detection. Experiments using public video data 'Highway-1' and 'Highway-3', and real traffic video data provided by the New Jersey Department of Transportation (NJDOT) demonstrate the effectiveness of the proposed method.**

I. Introduction

Extracting moving objects from the stationary background is one of the fundamental steps in video analysis systems. In many scenarios, the blockage of light from a light source by the moving objects causes shadows to be cast on the stationary objects. These shadows follow a similar dynamic pattern to the moving objects and are often detected as part of the foreground in the background subtraction process. In the case of traffic videos, misclassifying cast shadows as foreground objects can result in object merging and shape alternation which has a negative effect on further video analysis tasks such as segmentation [5]–[7], [19], vehicle counting [18], classification [4], and tracking [8].

Many studies have addressed the problem of cast shadow removal in recent years. A large number of cast shadow detection methods utilize color information in order to detect shadowed pixels [9], [27]. A common assumption among many approaches is the shadowed region being darker in intensity while being invariant in chromaticity [3], [9]. Other methods have utilized texture features [10], [17], statistical modeling [12], [15], or a combination of features [22]–[24], [27] to detect and remove shadows from the foreground mask. In recent years, deep learning approaches have also been applied for removing the cast shadows [26], [29].

The existing methods have a number of limitations, including dependence on presumptions that are limited to specific scenarios, misclassification due to similarities among dark objects and shadows, and high computational complexity. In this paper, a real-time approach for moving cast shadow removal is proposed with three main contributions: First, we introduce an effective approach to select potential shadow candidates with general criteria that can be applied in a wide range of outdoor videos. Second, an efficient region-based segmentation method is applied in order to take the geometric properties into account and improve the results of the pixel-wise classification. Third, a set of six color and temporal features are extracted for each pixel and a Gaussian mixture model is applied to model the shadow characteristics and classify the feature vectors.

The remainder of this paper is organized as follows. In section II the various steps of the proposed method are described in order. Section II-A describes how the potential shadow candidates are extracted. Section II-B contains the description of the k-means clustering approach for grouping the pixels at each component of the foreground. Section II-C contains details on modeling the shadow samples by using the GMM method. Section II-D explains the integration process of the previous steps. The performance of the proposed method is evaluated in section III, and the paper is concluded in section IV.

II. A new cast shadow detection method

In this section, the steps of the proposed shadow detection method are discussed in detail. First, the moving objects are identified by applying the innovative global foreground modeling (GFM) method [20], [21] in order to extract the moving objects along with their shadows as foreground. Then the candidate shadow pixels are extracted based on the physical properties of shadows. As the next step, the K-means algorithm is applied to each blob in the foreground mask to segment the shadow regions. Lastly, local gradient and color features are modeled by a mixture of Gaussian distributions in order to detect the shadow pixels from the moving objects. Figure 1 illustrates the general workflow of the proposed method.

[1]Hadi Ghahremannezhad is with Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA hg255@njit.edu

[2]Hang Shi is with Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA hs328@njit.edu

[3]Chengjun Liu is with Faculty of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA cliu@njit.edu

2021 IEEE International Intelligent Transportation Systems Conference (ITSC) | 978-1-7281-9142-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITSC48978.2021.9565049

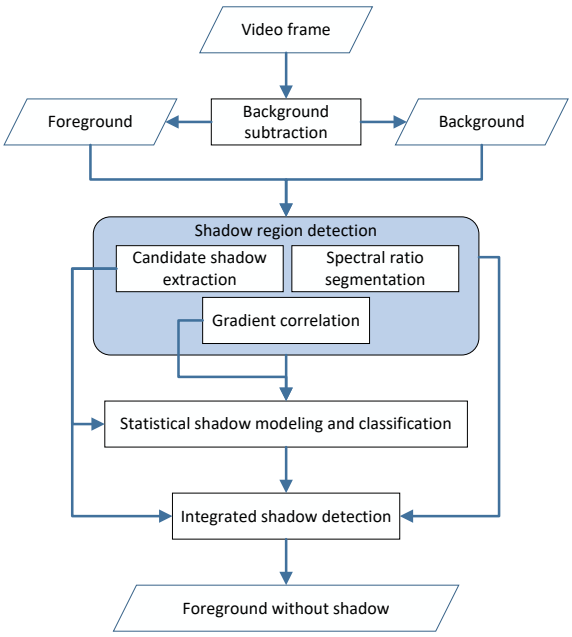


Fig. 1: The general workflow of the proposed method.

A. Extracting candidate shadow pixels

The initial candidate shadow pixels are identified as the pixels with semi-proportional attenuation in different color channels when shadowed. As a valid presumption, pixels with increased luminance values or drastically different in the proportions of the chromaticity values are considered to be moving objects. On the other hand, since in traffic videos we are dealing with outdoor scenes the direct and ambient illuminants are the sun and the sky, respectively. The atmosphere around the earth results in the Rayleigh scattering effect of the white radiation emitted from the sun which is inversely proportional to the fourth power of wavelength. This means the shorter wavelengths such as blue are scattered more strongly than the longer ones such as red. This observation is considered as a chrominance property that can be formulated as follows:

$$\Delta E(\lambda_B) < \Delta E(\lambda_G) \leq \Delta E(\lambda_R) \tag{1}$$

where $\Delta E(\lambda_R), \Delta E(\lambda_G), \Delta E(\lambda_B)$ represent the attenuation in incident illumination at wavelengths corresponding to red, green, and blue components, respectively.

Assuming $Q_k(\lambda) \in \{Q_R(\lambda), Q_G(\lambda), Q_B(\lambda)\}$ to be the spectral sensitivities of the red, green, and blue camera sensors, respectively, the color components of the reflected intensities reaching the sensors at each point (x, y) in the two-dimensional image plane are represented as follows:

$$C_i(x, y) = \int_\Lambda E(\lambda, x, y) S(\lambda, x, y) Q_k(\lambda) \mathrm{d}\lambda \tag{2}$$

where $C_k(x, y), k \in \{R, G, B\}$ are the sensor responses, $E(\lambda, x, y)$ and $S(\lambda, x, y)$ are the incident illumination and surface reflectance at location (x, y), respectively, and λ is the wavelength [13]. The interval of the summation is determined by $Q_k(\lambda)$, which is non-zero over the visible spectrum range of wavelengths Λ. If we consider the camera filters to have infinitely narrow bandwidth [3], they can be approximated as impulse functions centered on the filter's characteristics and be represented as Dirac delta functions, e.g., $Q_k(\lambda) = q_i \delta(\lambda - \lambda_k)$, and the sensor responses are defined as:

$$C_k(x, y) = E(\lambda_k, x, y) S(\lambda_k, x, y) q_k \tag{3}$$

where $\lambda_k, k \in \{R, G, B\}$ represents the central frequency of the k-th channel filter, and q_ks, $k \in \{R, G, B\}$ are the spectral sensitivities of the three color camera sensors. When shadow is cast over a pixel of the image at location (x, y), the incident illumination drops due to the blockage of the sunlight as opposed to the surface material and the camera spectral sensitivities which remain unchanged. Therefore, the chrominance property of Equation (1) can be represented as follows:

$$\Delta C_B(x, y) < \Delta C_G(x, y) \leq \Delta C_R(x, y) \tag{4}$$

where $\Delta C_k(x, y), k \in \{R, G, B\}$ is the contribution of the sunlight at the location (x, y) in the image plane when it is not under cast shadow.

For extracting initial candidate shadows we consider only a portion of the conic region in the RGB space between the values that the pixel takes when it belongs to the background among different frames and the origin. Since the chromaticity values of the shadowed pixel are semi-proportional to the lit pixel the apex angle of the cone is limited by a threshold depending on the ambient illumination. Let $\mathbf{f}(x, y) = [f_R, f_G, f_B]^T$ and $\mathbf{b}(x, y) = [b_R, b_G, b_B]^T$ be the vectors representing foreground and background pixels at spatial location (x, y) in the RGB space where f_R, f_G, f_B and b_R, b_G, b_B denote the red, green and blue components, respectively. The attenuation of the three components are also taken into account (Figure 2) to set up criteria for extracting potential shadows in a binary mask $\mathcal{P}$ as follows:

$$\mathcal{P}(x, y) = \begin{cases} 1, & \text{if } (b_B - f_B < b_G - f_G \leq b_R - f_R) \\ & \wedge (\tau_{al} < \theta_d < \tau_{ah}) \\ & \wedge (\tau_{rl} < r_d < \tau_{rh}) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

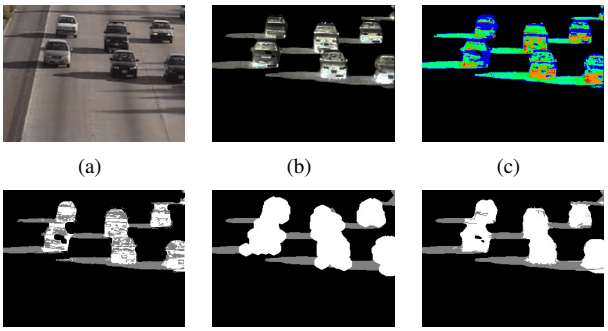
where $\theta_d = cos^{-1} \frac{\mathbf{f} \cdot \mathbf{b}}{\|\mathbf{f}\| \times \|\mathbf{b}\|}$ is the angular distance between the two pixel vectors when it belongs to the foreground and background, $r_d = \frac{\|\mathbf{f}\| \times cos(\theta_d)}{\|\mathbf{b}\|}$ is the ratio of the vector magnitudes, and $\tau_{al}, \tau_{ah}, \tau_{rl}$, and τ_{rh} denote the lower and upper thresholds for the angular distance and norm ratios, respectively. All the pixels in the foreground class that satisfy these criteria are considered to be initial shadow candidate samples (Figure 3(d)).

B. Spatial clustering for detecting shadow regions

As illustrated in Figure 2, separating the shadows and dark objects solely based on the variations in the pixels' values is not possible due to the similar attenuation caused by dark objects. The results of the pixel-wise classification can be

(a) (b) (c)

Fig. 2: Histogram of RGB norm ratios in a sample traffic video [1]. (a) video frame. (b) Lighter, darker, and shadowed samples represented by orange, brown, and gray, respectively. (c) Histogram of the RGB norm ratios.



(a) (b) (c) (d) (e) (f)

Fig. 3: Region based classification. (a) Original video frame. (b) Spectral ratio. (c) Segmentation. (d) Potential shadows ($\mathcal{P}$). (e) Gradient variations ($\mathcal{E}$). (f) Region-based classification ($\mathcal{S}$).

improved by applying region-based techniques that consider the relationship between each pixel and its neighborhood. In the case of traffic videos where the shadow is usually cast on the road with a rough Lambertian surface, we can safely assume a single dominant illumination source, e.g., the sun, uniform reflection parameters, and negligible specular reflection.

If we express the incident illumination in terms of direct and ambient irradiances, indicated by l_d and l_a, and represent the achromatic and chromatic aspects of the body reflection by m_b and c_b, respectively, we can express the camera sensor responses at location (x, y) in Equation (3) according to the Bi-Illuminant Dichromatic Reflection (BIDR) model [16] as follows:

$$\begin{aligned} C_k &= \alpha C_k^d + C_k^a = \alpha q_k E_k^d S_k^d + q_k E_k^a S_k^a \\ &= \alpha q_k l_d^k(\theta_L, \phi_L) m_b(\theta_e, \phi_e, \theta_i, \phi_i) c_b^k \\ &+ q_k c_b^k \int_{\theta_i, \phi_i} l_a^k(\theta_i, \phi_i) m_b(\theta_e, \phi_e, \theta_i, \phi_i) d\theta_i d\phi_i \end{aligned} \tag{6}$$

where C_k^d and C_k^a, $k \in \{R, G, B\}$ are the contributions of the direct and ambient components of illumination on the sensor responses, respectively, $\alpha \in [0, 1]$ is the attenuation factor that accounts for the unoccluded proportion of the direct light, E_k^d, S_k^d, E_k^a, and S_k^a are the incident illumination and surface reflectance of the direct and ambient components, respectively, c_b^i is the chromatic aspect of the body reflection, (θ_L, ϕ_L) are the pan and tilt angles indicating the direction of the direct light source relative to the local surface normal, and (θ_e, ϕ_e) and (θ_i, ϕ_i) are the emittance and incidence angles, respectively. By representing the integral in the ambient term as M_a^k, and considering the scene angles to be constants in a specific geometry, we can express the camera responses at each location (Equation (6)) in a simplified form as follows:

$$C_k = \alpha q_k l_d^k m_b c_b^k + q_k c_b^k M_a^k, k \in \{R, G, B\} \tag{7}$$

If we assume the variations in M_a^k due to the blockage of ambient light in some angles are negligible, we can obtain an near-invariant illumination feature, henceforth referred to as spectral ratio $\overrightarrow{S} = [S_R, S_G, S_B]^T$, as follows:

$$S_k = \frac{BG - FG}{FG} = \frac{(1-\alpha) q_k l_d^k m_b c_b^k}{\alpha q_k l_d^k m_b c_b^k + q_k c_b^k M_a^k} \tag{8}$$

where $k \in \{R, G, B\}$ indicates the sensor bands. Since there is little to no direct illumination in the umbra region of the shadow ($\alpha = 0$), the spectral ratio in this region can be indicated as follows:

$$S_k = \frac{l_d^k m_b}{M_a^k} \tag{9}$$

In traffic videos shadows are cast on the road surface with a homogeneous material where the body reflection does not change and therefore, the spectral ratio is near-constant across the shadow region and the changes are mostly due to the variations in the ambient illumination (Figure 3(b)). The binary motion mask, $\mathcal{M}(x, y)$ is divided into a number of independent regions $\mathcal{R} = \{r_1, r_2, ..., r_k\}$ by applying component analysis. The k-means clustering algorithm is applied on the spectral ratios of each region in R in order to partition each moving component $r_k \in R$ into a number of segments s_l^k, such that:

$$\bigcup_{l=1}^{n_k} s_l^k = r_k, \quad \bigcap_{l=1}^{n_k} s_l^k = \varnothing, \quad \bigcup_{k=1}^{K} r_k = R \tag{10}$$

where n_k is the total number of segments s_l^k at each region, i.e., object r_k (Figure 3(c)).

One of the observations in traffic videos is the difference between the amount of gradient information introduced by the vehicles and shadows. Therefore, we apply the Canny edge detection method to extract the edges of the background, foreground, and the binary motion mask, which are respectively referred to as E_{BG}, E_{FG}, and $E_{\mathcal{M}}$. Then the subtraction of background edges and the dilated $E_{\mathcal{M}}$ from the foreground edges at the location of moving objects are dilated to retrieve a binary mask $\mathcal{E}(x, y)$ representing the location of objects based on the edge variations as follows:

$$\begin{aligned} \mathcal{E}(x, y) = \Big(&\big(E_{FG}(x, y) \wedge \mathcal{M}(x, y)\big) \\ &- \big(\big(E_{\mathcal{M}(x,y)} \oplus N\big) + E_{BG}(x, y)\big) \Big) \oplus N \end{aligned} \tag{11}$$

where N is a dilation kernel (Figure 3(e)).

Another observation in traffic videos is the spatial distribution of shadow pixels around the objects which results in shadow segments of each region containing a considerable number of extrinsic terminal points T. Taking all these observations into account, we classify the segments of each region into shadow and object groups as follows:

$$\begin{aligned}
\mathcal{S}_p(s_l^k) &= \begin{cases} 1, \text{if } \frac{|\mathcal{P} \cap s_l^k|}{|s_l^k|} > \tau_p \\ 0, \text{otherwise} \end{cases} \\
\mathcal{S}_e(s_l^k) &= \begin{cases} 1, \text{if } \frac{|\mathcal{E} \cap s_l^k|}{|s_l^k|} < \tau_e \\ 0, \text{otherwise} \end{cases} \\
\mathcal{S}_t(s_l^k) &= \begin{cases} 1, \text{if } \frac{|T(r_k) \cap T(s_l^k)|}{|T(r_k)|} > \tau_t \\ 0, \text{otherwise} \end{cases} \\
\mathcal{S}(s_l^k) &= \begin{cases} 1, \text{if } \left(\mathcal{S}_p(s_l^k) \cap \mathcal{S}_e(s_l^k) \cap \mathcal{S}_t(s_l^k)\right) \\ 0, \text{otherwise} \end{cases}
\end{aligned} \tag{12}$$

where $S_p(s_l^k)$, $S_e(s_l^k)$, and $S_t(s_l^k)$ are indicator functions stating that the segment s_l^k belongs to the shadow class if more than τ_p of its pixels are classified as potential shadows, less than τ_e of its pixels are classified as object edges, and contains more than τ_t of terminal points of the region r_k, respectively. The superposition of these criteria defines the final shadow classification results indicated by $\mathcal{S}(s_l^k)$ which is 1 when the segment s_l^k is classified as shadow and 0 if it is classified as part of the moving object. Figure 3 illustrates the steps of the segmentation method in a sample video frame. The white and gray colors represent the 1 and 0 values in the binary masks, respectively.

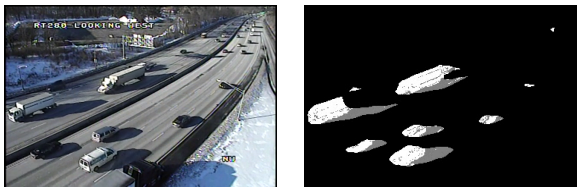
C. Statistical modeling based on shadow features

Although the candidate shadow samples include all possible shadow pixels at each frame some non-shadow samples of the foreground such as self shadows and dark objects may be detected as cast shadows due to their similarities. Here, a feature vector is constructed in order to model the shadow samples and only keep the cast shadows in the final shadow mask. A set of six-dimensional feature vectors is denoted by:

$$\mathcal{Z}^t = \{\mathbf{z}_1^t, \mathbf{z}_2^t, ..., \mathbf{z}_N^t\} = \{\mathbf{z}_i^t\}_{i=1}^N \tag{13}$$

where $\mathbf{z}_i^t$ is a six dimensional feature vector for pixel i at frame t. Let $\mathbf{u}(x,y) = [u_R, u_G, u_B]^T$ be the vector from a foreground pixel at location (x,y) to its corresponding pixel in the background image. The six features are defined as follows:

$$\begin{cases}
\mathfrak{z}_1 = \frac{V_f}{V_b + \epsilon} \\
\mathfrak{z}_2 = S_f - S_b \\
\mathfrak{z}_3 = min\left(\frac{\|\mathbf{u}\|}{\|\mathbf{b}\|}, 1\right) \\
\mathfrak{z}_4 = tan^{-1}\left(f_G / f_R\right) \\
\mathfrak{z}_5 = cos^{-1}\left(f_B / \|\mathbf{u}\|\right) \\
\mathfrak{z}_6 = \nabla(BG) / \left(\nabla(BG) + \nabla(FG)\right)
\end{cases} \tag{14}$$



(a) (b)

Fig. 4: The detected shadows using statistical modeling. (a) Sample video frame. (b) Classification results.

where V_f, V_b, S_f, and S_b are the value and saturation components of foreground and background in HSV colorspace, respectively, $\mathfrak{z}_3$ is the attenuation, $\mathfrak{z}_4$ is the green to red direction, $\mathfrak{z}_5$ is the blue direction, and $\mathfrak{z}_6$ is the gradient intensity of the pixel which is not influenced by cast shadows so much as it is affected by moving objects.

The feature vectors are modeled by K Gaussian distributions via a GMM for each pixel whenever it is identified as a potential shadow sample in the previous step. The Gaussian modeling of the feature vectors is described as follows:

$$\begin{aligned}
P(\mathbf{z}) &= \sum_{k=1}^{K} W_k N(\mathbf{z}|\omega_k) \\
N(\mathbf{z}|\omega_k) &= \frac{\exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{M}_k)^t \Sigma_k^{-1} (\mathbf{z} - \mathbf{M}_k)\right\}}{(2\pi)^{d/2} \mid \mathbf{\Sigma}_k \mid^{1/2}} \\
&\sum_{k=1}^{K} W_k = 1
\end{aligned} \tag{15}$$

where $\mathbf{z} \in \mathbb{R}^d$ is the six-dimensional feature vector, K is the number of Gaussian distributions, W_k is the weight of the k_{th} Gaussian distribution $N(\mathbf{z}|\omega_k)$. $\mathbf{M}_k$ and $\mathbf{\Sigma}_k$ are the mean vector and the covariance matrix of the k_{th} Gaussian density $N(\mathbf{z}|\omega_k)$. Note that the Gaussian model of each pixel is updated only when it is identified as a candidate shadow sample. Based on the Bayes classification method each feature vector is classified into either the object or the shadow class according to the following discriminant function:

$$h(\mathbf{z}) = p\left(\mathbf{z}|\omega_{obj}\right) P\left(\omega_{obj}\right) - p\left(\mathbf{z}|\omega_{sh}\right) P\left(\omega_{sh}\right) \tag{16}$$

where $p\left(\mathbf{z}|\omega_{obj}\right)$ and $p\left(\mathbf{z}|\omega_{sh}\right)$ are the conditional probability density function for the object and shadow classes, respectively. Each feature vector $\mathbf{z}$ is assigned to either one of the shadow or object classes based on the value of the discriminant function and a binary mask $\mathcal{H}$ for statistical modeling results is obtained as follows:

$$\mathcal{H}(\mathbf{z}) = \begin{cases} 1, \text{if } h(\mathbf{z}) > 0 \\ 0, \text{otherwise} \end{cases} \tag{17}$$

where $\mathcal{H}(\mathbf{z}) = 1$ indicates that the feature vector $\mathbf{z}$ is classified as part of an object and $\mathcal{H}(\mathbf{z}) = 0$ indicates that it is classified as shadow. Figure 4 is a an example of statistical shadow modeling results.

TABLE I: The shadow detection results compared to other methods.

	Highway-1			Highway-3			Interstate-280		
Methods	η	ξ	F-measure	η	ξ	F-measure	η	ξ	F-measure
Zhu et al. [29]	**95**%	36%	53%	88%	32%	46%	—	—	—
Cucchiara et al. [3]	74%	75%	75%	68%	62%	65%	45%	64%	53%
Sanin et al. [17]	82%	94%	88%	62%	**91**%	74%	43%	62%	50%
Gomes et al. [9]	88%	94%	**91**%	65%	90%	75%	—	—	—
Huang and Chen. [12]	73%	82%	78%	73%	71%	72%	58%	**74**%	65%
Hsieh et al. [11]	70%	72%	71%	60%	70%	65%	36%	58%	44%
Amato et al. [2]	81%	85%	83%	72%	75%	73%	21%	52%	30%
Wang et al. [25]	78%	93%	85%	66%	72%	69%	—	—	—
Zhang et al. [28]	86%	94%	90%	82%	**91**%	87%	—	—	—
Proposed method	90%	93%	**91**%	**90**%	86%	**88**%	**63**%	71%	**67**%

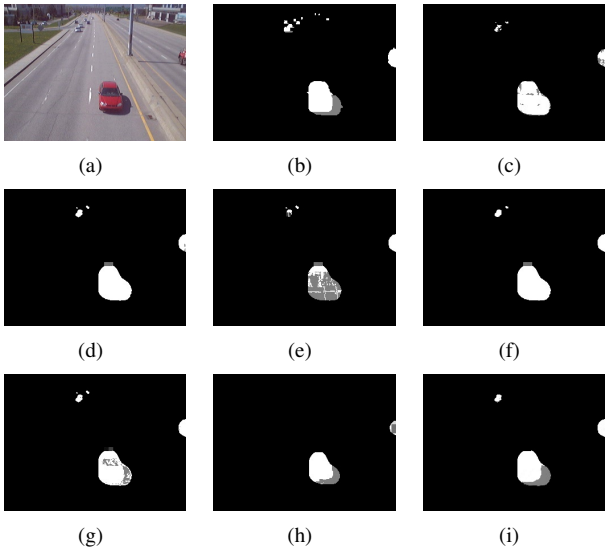


Fig. 5: The foreground masks and detected shadows in different methods. (a) Original video frame. (b) Ground truth. (c),(d), (e), (f), (g), (h), and (i) are the results of the Cucchiara et al. [3], Leone and Distante [14], Hsieh et al. [11], Sanin et al. [17], Huang and Chen [12], Amato et al. [2], and our proposed method, respectively.

D. Final shadow detection based on integration

As the last step, we integrate the shadow detection results of the previous steps by taking the weighted summation, as follows:

$$\mathcal{W}(x,y) = w_{\mathcal{P}}\mathcal{P}(x,y) + w_{\mathcal{S}}\mathcal{S}(x,y) + w_{\mathcal{H}}\mathcal{H}(x,y) \quad (18)$$

where $w_{\mathcal{P}} \in [0,1]$, $w_{\mathcal{S}} \in [0,1]$, and $w_{\mathcal{H}} \in [0,1]$ are the weights indicating the significance of the shadow detection results based on chromatic criteria, region-based classification, and statistical modeling, respectively. The three weights should sum up to one as they are normalized:

$$w_{\mathcal{P}} + w_{\mathcal{S}} + w_{\mathcal{H}} = 1 \quad (19)$$

Since the $\mathcal{P}$ and $\mathcal{H}$ are pixel-based classification methods and fail to differentiate some of the dark objects from the shadows, we have considered $w_{\mathcal{S}}$ to be twice the value of $w_{\mathcal{P}}$ and $w_{\mathcal{H}}$. The weighted sum values are thresholded in order to obtain a binary mask $\mathcal{F}$ representing the final shadow detection results as follows:

$$\mathcal{F}(x,y) = \begin{cases} 1, \text{if } \mathcal{W}(x,y) > \tau_f \\ 0, \text{otherwise} \end{cases} \quad (20)$$

where τ_f is a threshold, $\mathcal{F}(x,y) = 1$ indicates that the pixel at location (x,y) belongs to the moving objects and $\mathcal{F}(x,y) = 0$ means it belongs to the shadow class.

III. Experiments

In this section, the qualitative and quantitative performance of the proposed method is analyzed using the 'Interstate-280' data provided by the New Jersey Department of Transportation (NJDOT) and publicly available data 'Highway-1' and 'Highway-3' videos [1]. Each video sequence has a spatial resolution of 320×240 or 640×482 and has a frame rate of 15 fps. The hardware used to implement the method is a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. The average running speed of our proposed method is reported in Table II for each video frame of size 320×240 and 640×482 pixels that shows the method is fast enough to be used as a pre-processing step in real-time traffic video analysis tasks. In Figure 5, a sample frame from the Highway-3 video is demonstrated along with the foreground mask obtained by different methods after removing the cast shadow. In our experiments, the thresholds τ_p, τ_e, τ_t, and τ_f are all set to 0.5, empirically.

In order to evaluate the quantitative results of the shadow detection method we have utilized three performance measures as follows:

$$\begin{cases} \xi = TP_o/(TP_o + FN_o) \\ \eta = TP_s/(TP_s + FN_s) \\ F_1 = 2 \times (\eta \times \xi)/(\eta + \xi) \end{cases} \quad (21)$$

where TP_o and TP_s represent the true positive rates of object and shadow pixels, FN_o, and FN_s represent the false negative rates of the object and shadow pixels, respectively. ξ, η, and F_1 are the shadow discrimination rate, shadow detection rate, and F-measure that indicate the performance of the method. Table I demonstrates the comparative performance results of the proposed approach and some of the popular methods tested on three traffic sequences.

TABLE II: The average running time (in milliseconds) for each video frame in different methods

Methods	Running time (ms)	
	320×240	640×482
Cucchiara et al. [3]	23	141
Zhu et al. [29] (with GPU)	421	1069
Huang and Chen. [12]	16	81
Sanin et al. [17]	61	244
Hsieh et al. [11]	5	16
leone and Distante. [14]	135	284
Amato et al. [2]	16	102
Proposed method	12	37

IV. Conclusion

In this study, a new statistical moving cast shadow detection method is proposed which is based on physical properties. Specifically, the moving objects along with their cast shadows are segmented from the stationary background. Then physical properties of shadows are utilized in order to extract the potential candidate shadow samples. The k-means clustering approach is applied to group the pixels of each moving component into object and shadow regions. A six-dimensional feature vector is constructed for each pixel which is modeled by a mixture of Gaussian distributions whenever the pixel is a potential shadow sample and is in a shadow segment. Each pixel is classified to either the object or the shadow class by applying the Bayes classifier. Finally, the results of the previous steps are integrated for final shadow detection. The experimental results demonstrate the feasibility of the proposed method for shadow detection and removal in real-time applications.

Acknowledgment

This paper is partially supported by the NSF grant 1647170.

References

[1] Aton datasets by ucsd. [Online]. Available: http://cvrr.ucsd.edu/aton/shadow/

[2] A. Amato, M. G. Mozerov, A. D. Bagdanov, and J. Gonzalez, "Accurate moving cast shadow suppression based on local color constancy detection," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2954–2966, 2011.

[3] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.

[4] M. O. Faruque, H. Ghahremannezhad, and C. Liu, "Vehicle classification in video using deep learning," in *the 15th International Conference on Machine Learning and Data Mining*. IBAI, 2019, pp. 117–131.

[5] H. Ghahremannezhad, H. Shi, and C. Liu, "Robust road region extraction in video under various illumination and weather conditions," in *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*. IEEE, 2020, pp. 186–191.

[6] ——, "Automatic road detection in traffic videos," in *10th IEEE International Conference on Big Data and Cloud Computing*. IEEE, 2020, pp. 777–784.

[7] ——, "A new adaptive bidirectional region-of-interest detection method for intelligent traffic video analysis," in *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*. IEEE, 2020, pp. 17–24.

[8] ——, "A real time accident detection framework for traffic video analysis," in *the 16th International Conference on Machine Learning and Data Mining*. IBAI, 2020, pp. 77–92.

[9] V. Gomes, P. Barcellos, and J. Scharcanski, "Stochastic shadow detection using a hypergraph partitioning approach," *Pattern Recognition*, vol. 63, pp. 30–44, 2017.

[10] R. Guo, Q. Dai, and D. Hoiem, "Paired regions for shadow detection and removal," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 12, pp. 2956–2967, 2012.

[11] J.-W. Hsieh, W.-F. Hu, C.-J. Chang, and Y.-S. Chen, "Shadow elimination for effective moving object detection by gaussian shadow modeling," *Image and Vision Computing*, vol. 21, no. 6, pp. 505–516, 2003.

[12] J.-B. Huang and C.-S. Chen, "Moving cast shadow detection using physics-based features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2310–2317.

[13] M. J. Ibarra-Arenado, T. Tjahjadi, and J. Pérez-Oria, "Shadow detection in still road images using chrominance properties of shadows and spectral power distribution of the illumination," *Sensors*, vol. 20, no. 4, p. 1012, 2020.

[14] A. Leone and C. Distante, "Shadow detection for moving objects based on texture analysis," *Pattern Recognition*, vol. 40, no. 4, pp. 1222–1233, 2007.

[15] Z. Liu, D. An, and X. Huang, "Moving target shadow detection and global background reconstruction for videosar based on single-frame imagery," *IEEE Access*, vol. 7, pp. 42 418–42 425, 2019.

[16] B. A. Maxwell, R. M. Friedhoff, and C. A. Smith, "A bi-illuminant dichromatic reflection model for understanding images," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[17] A. Sanin, C. Sanderson, and B. C. Lovell, "Improved shadow removal for robust person tracking in surveillance scenarios," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 141–144.

[18] N. Seenouvong, U. Watchareeruetai, C. Nuthong, K. Khongsomboon, and N. Ohnishi, "A computer vision based vehicle detection and counting system," in *2016 8th International Conference on Knowledge and Smart Technology (KST)*. IEEE, 2016, pp. 224–227.

[19] H. Shi, H. Ghahremannezhadand, and C. Liu, "A statistical modeling method for road recognition in traffic video analytics," in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2020, pp. 000 097–000 102.

[20] H. Shi and C. Liu, "A new foreground segmentation method for video analysis in different color spaces," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2899–2904.

[21] ——, "A new global foreground modeling and local background modeling method for video analysis," in *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, 2018, pp. 49–63.

[22] ——, "Moving cast shadow detection in video based on new chromatic criteria and statistical modeling," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 196–201.

[23] ——, "A new cast shadow detection method for traffic surveillance video analysis using color and statistical modeling," *Image and Vision Computing*, vol. 94, p. 103863, 2020.

[24] M. Shilpa, M. T. Gopalakrishna, and C. Naveena, "Approach for shadow detection and removal using machine learning techniques," *IET Image Processing*, vol. 14, no. 13, pp. 2998–3005, 2020.

[25] B. Wang, Y. Zhao, and C. P. Chen, "Moving cast shadows segmentation using illumination invariant feature," *IEEE Transactions on Multimedia*, vol. 22, no. 9, pp. 2221–2233, 2019.

[26] Y. Wang, X. Zhao, Y. Li, X. Hu, K. Huang *et al.*, "Densely cascaded shadow detection network via deeply supervised parallel fusion," in *27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 1007–1013.

[27] M. Wu, R. Chen, and Y. Tong, "Shadow elimination algorithm using color and texture features," *Computational Intelligence and Neuroscience*, vol. 2020, 2020.

[28] H. Zhang, S. Qu, H. Li, J. Luo, and W. Xu, "A moving shadow elimination method based on fusion of multi-feature," *IEEE Access*, vol. 8, pp. 63 971–63 982, 2020.

[29] L. Zhu, Z. Deng, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, "Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 121–136.