



A new cast shadow detection method for traffic surveillance video analysis using color and statistical modeling[☆]

Hang Shi^{*}, Chengjun Liu

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA

ARTICLE INFO

Article history:

Received 8 April 2019

Received in revised form 26 September 2019

Accepted 5 December 2019

Available online 13 December 2019

Keywords:

Traffic video analysis

Shadow detection

Global Foreground Modeling (GFM)

New chromatic criteria

Shadow region detection method

Statistical shadow modeling

ABSTRACT

In traffic surveillance video analysis systems, the cast shadows of vehicles often have a negative effect on video analysis results. A novel cast shadow detection framework, which consists of a new foreground detection method and a cast shadow detection method, is presented in this paper to detect and remove the cast shadows from the foreground. The new foreground detection method applies an innovative Global Foreground Modeling (GFM) method, a Gaussian mixture model or GMM, and the Bayes classifier for foreground and background classification. While the GFM method is for global foreground modeling, the GMM is for local background modeling, and the Bayes classifier applies both the foreground and the background models for foreground detection. The rationale of the GFM method stems from the observation that the foreground objects often appear in recent frames and their trajectories often lead them to different locations in these frames. As a result, the statistical models used to characterize the foreground objects should not be pixel based or locally defined. The cast shadow detection method contains four hierarchical steps. First, a set of new chromatic criteria is presented to detect the candidate shadow pixels in the HSV color space. Second, a new shadow region detection method is proposed to cluster the candidate shadow pixels into shadow regions. Third, a statistical shadow model, which uses a single Gaussian distribution to model the shadow class, is presented for classifying shadow pixels. Fourth, an aggregated shadow detection method is presented for final shadow detection. Experiments using the public video data 'Highway-1' and 'Highway-3', and the New Jersey Department of Transportation (NJDOT) real traffic video sequences show the feasibility of the proposed method. In particular, the proposed method achieves better shadow detection performance than the popular shadow detection methods, and is able to improve the traffic video analysis results.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

In traffic video analysis, shadows are often detected as part of the foreground, as they share the similar motion patterns to the foreground objects [1–3]. The cast shadows are always strong and occupy large areas, especially during the sunny days. These cast shadows often adversely affect the video analysis performance in various applications, such as vehicle tracking and vehicle classification [4]. Many algorithms have been published to detect the moving foreground objects in video [5–16]. Some methods like the Gaussian Mixture Modeling (GMM) estimate the background for each pixel using a number of Gaussian distributions [5–7,9,11,12]. Other methods apply a classification method, such as the support vector machine (SVM) to classify the foreground and the background

pixels [17,11,12]. Yet, the cast shadows are usually classified into the foreground class as they have the similar motion patterns to their foreground objects, which deteriorates traffic video analysis performance.

We present a novel cast shadow detection framework based on color and statistical modeling to detect and remove the cast shadows from the foreground region in order to improve video analysis performance. The novelty of our proposed framework comes from the following methods: a new foreground detection method and a novel cast shadow detection method. The new foreground detection method applies an innovative Global Foreground Modeling (GFM) method, a Gaussian mixture modeling or GMM method, and the Bayes classifier for foreground and background classification. The motivation for developing the GFM method is based on the observation that the foreground objects often appear in recent frames and their trajectories often lead them to different locations in these frames. In contrast to the uniform distribution or a local foreground modeling of the foreground [5–7], our GFM method models all the foreground objects globally using multiple Gaussian distributions.

Thus, the statistical models used by the GFM method to characterize the foreground objects are not locally defined. As a result, the GFM method is for global foreground modeling, the GMM method is for local background modeling, and the Bayes classifier applies both the foreground and the background models for foreground detection.

The novel cast shadow detection method contains four hierarchical steps, whose contributions are summarized below. First, we present a set of new chromatic criteria to detect the candidate shadow pixels in the HSV color space. We use the HSV color space for shadow detection due to its property of separating the chromaticity from intensity [2,18–21]. Our new chromatic criteria are more robust than the criteria used by other popular methods for shadow detection [2,21]. Second, we present a new shadow region detection method to cluster the candidate shadow pixels into shadow regions. Many shadow detection methods can not solve the shadow outlines problem: the outlines of the shadow regions are often classified to the foreground. As a result, after removing the shadow pixels from the foreground, the shadow regions are only partially removed, and the shadow outlines are often classified to the foreground. Our new shadow region detection method is able to solve this problem by applying the prior knowledge that both the foreground objects and their cast shadows should define continuous regions. Third, we present a statistical shadow modeling method, which uses a single Gaussian distribution to model the shadow class, to classify shadow pixels. The shadow pixels detected by both the new chromatic criteria and the new shadow region detection method tend to be more reliable shadow pixels, we therefore use these shadow pixels to estimate the Gaussian distribution for the shadow class. Finally, we present an aggregated shadow detection method that integrates the detection results using the new chromatic criteria, the new shadow region detection method, and the new statistical shadow modeling method. A gray scale shadow map is obtained by calculating a weighted summation of the candidate shadow pixels. A shadow free foreground may be derived by thresholding the gray scale shadow map.

We implement experiments using the public data ‘Highway-1’ and ‘Highway-3’ videos, and the New Jersey Department of Transportation (NJDOT) real traffic video sequences to show the feasibility of the proposed method. In particular, the experimental results (both qualitative and quantitative results) show that our proposed method achieves better shadow detection performance than some popular shadow detection methods [21–25], and is able to improve the vehicle tracking performance.

2. Related work

In traffic video analysis, our goal is to be able to analyze the behavior of the moving objects (vehicles, pedestrians etc.) and distinguish the categories of the objects, but the cast shadows of the objects often affect our analysis of video. Many methods have been published for cast shadow detection [1–3]. As color often provides useful information for shadow detection, some methods apply color information to detect shadows [18,20,26–28]. Many shadow detection methods assume that the shadow areas are darker in intensity but relatively invariant in chromaticity [2,18–21]. The color spaces that separate chromaticity from intensity are thus often used for shadow detection. Some example such color spaces are the HSV color space [19], and the YUV color space [20]. We apply a set of chromatic criteria by assuming that the cast shadows have similar hue to the background, but a lower value than the background [2,21].

Statistical shadow modeling is applied for shadow detection as well [29,24,30]. The major assumption of these methods is that the light source is pure white and the attenuation of the illumination is linear. Generally speaking, these statistical shadow modeling

methods are able to predict color changes of the shadow pixels better than the color based methods, but the shadow detection accuracy in outdoor scenes tends to deteriorate.

There are methods that use the shape, size, and orientation information for shadow detection [22,31,32]. These methods are designed to deal with some objects that have specific shapes. The advantage of these methods is that they do not need to estimate the background color of the shadow, but the disadvantage is that they have difficulty in dealing with multiple types of objects in complex scenes.

There are methods that utilize texture for shadow detection, such as classifying a region into the shadow region or the object region based on the texture correlation between the foreground and the background [23,25,33–36]. These methods extract the texture information in different sizes of the regions. The advantage of these methods is that they are more robust to illumination changes than the color based methods, but the disadvantage is that the computation efficiency of matching the texture features is low.

There are also methods that use machine learning techniques for shadow detection. Guo et al. proposed a paired region based shadow detection algorithm [34]. Vidente et al. presented a kernel least-squares SVM method for separating shadow and non-shadow regions [37,35]. Many shadow detection algorithms using the deep neural network are presented recently [38–43].

3. A new foreground detection method using global foreground modeling and local background modeling

In video analysis, the cast shadows are often detected with their foreground objects, as these shadows share the similar motion patterns to their objects. Our proposed method will first detect the foreground regions that contain both the foreground objects and their shadows, and then remove the cast shadows from the detected foreground regions. Towards that end, we first present a new foreground detection method, which applies an innovative Global Foreground Modeling (GFM) method [15,16], a Gaussian mixture model (GMM), and the Bayes classifier for foreground and background classification. While the GFM method is for global foreground modeling, the GMM is for local background modeling, and the Bayes classifier applies both the foreground and the background models for decision making.

Our GFM method applies multiple Gaussian distributions to model all the foreground objects globally. Compared with the traditional GMM method, which estimates a mixture Gaussian distribution for every location in a frame, the GFM method only estimates one mixture Gaussian distribution for the whole frame. The rationale of the GFM method stems from the observation that the foreground objects often appear in recent frames and their trajectories often lead them to different locations in these frames. For example, a vehicle may appear in several continuous frames but at different locations. As a result, the statistical models used to characterize the foreground objects should not be pixel based or locally defined. In order to adapt to the dynamic nature of the foreground objects in video, we propose to globally model the foreground objects as follows:

$$p(\mathbf{x}) = \sum_{k=1}^K \alpha_k p(\mathbf{x}|\omega_k) \quad (1)$$

$$p(\mathbf{x}|\omega_k) = \frac{\exp \left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{M}_k)^t \Sigma_k^{-1}(\mathbf{x} - \mathbf{M}_k) \right\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \quad (2)$$

$$\sum_{k=1}^K \alpha_k = 1 \quad (3)$$

where $\mathbf{x} \in \mathbb{R}^d$ is the input feature vector, and $\mathbf{M}_k$, Σ_k , and α_k are the mean vector, the covariance matrix, and the weight of the k -th Gaussian density $p(\mathbf{x}|\omega_k)$, respectively. The feature vector $\mathbf{x}$ can be defined



Fig. 1. (a) A video frame from an NJDOT traffic video. (b) The background derived using the GMM model. (c) The foreground (with shadow) detected using our new foreground detection method.

by the color components in a specific color space, or one of some innovated feature vectors [15,16]. K is the number of the Gaussian distributions used to model the foreground objects globally. We also create a counter n_k for each Gaussian density to count the number of input vectors $\mathbf{x}$ and use it in the updating steps.

The initialization of the GFM method involves initializing all the parameters in the K Gaussian distributions to zero: the mean vectors, the covariance matrices, and the weights. During the initialization process, the counters n_k for the K Gaussian distributions are also set to zero. Then, we apply an online learning strategy to update our foreground model. Before all the K Gaussian distributions are associated with a non-zero weight, we use a hard thresholding method to determine which Gaussian distribution should be updated. If an input feature vector $\mathbf{x}$ is within the threshold of the conditional probability density function of its corresponding background, we regard it as a background pixel and do not update the foreground model. Otherwise, we check if $\mathbf{x}$ can fall in the threshold of any non-zero weight foreground Gaussian distributions. If it can, we use $\mathbf{x}$ to update that Gaussian density and increase the corresponding counter by one. If it cannot, we find a Gaussian distribution with zero weight in the foreground model, and use $\mathbf{x}$ to define the mean vector, set the diagonal values of the covariance matrix to a predefined value σ_0 , and change the corresponding counter to one. The updating strategy is summarized as follows [44]:

$$\mathbf{M}'_k = (n_k \mathbf{M}_k + \mathbf{x}) / (n_k + 1) \quad (4)$$

$$\Sigma'_k = (n_k \Sigma_k + (\mathbf{x} - \mathbf{M}_k)(\mathbf{x} - \mathbf{M}_k)^T) / (n_k + 1) \quad (5)$$

$$n'_k = n_k + 1 \quad (6)$$

After updating the density function, we also update the weights of all the Gaussian densities:

$$\alpha'_k = n'_k / \sum_{k=1}^K n'_k \quad (7)$$

The GFM method learns multiple Gaussian distributions to globally model the foreground in video. For each pixel in a frame, the most significant Gaussian distribution is chosen as the conditional probability density function for the foreground class. Specifically, the probability density function from the GFM model chosen as the conditional probability density function for each pixel is determined using the Bayes decision rule for minimum error [45] as follows:

$$p(\mathbf{x}|\omega_f)P(\omega_f) = \max_{i=1}^K p(\mathbf{x}|\omega_i)P(\omega_i) \quad (8)$$

where ω_f represents the foreground class, and $p(\mathbf{x}|\omega_f)$ is the foreground conditional probability density function for that pixel. Note that the input feature vector $\mathbf{x}$ is also used to update the Gaussian distribution.

In contrast to the global foreground modeling, the background is locally modeled using a Gaussian mixture model (GMM), which defines L Gaussian density functions for each location in a frame [5–7]. At each location (i, j) , we select the most significant Gaussian density among the L Gaussian density functions as the conditional probability density function $p_{ij}(\mathbf{x}|\omega_b)$ for the background and use the mean values of that density to estimate the background color. Let ω_b be the background class.

After the global foreground modeling and the local background modeling, our new foreground detection method applies the Bayes classifier for decision making. In particular, let $p(\mathbf{x}|\omega_f)$ and $p(\mathbf{x}|\omega_b)$ be the foreground and background conditional probability density functions, respectively, and let $P(\omega_f)$ and $P(\omega_b)$ be the prior probabilities for the foreground and the background, respectively. The discriminant function is as follows:

$$h(\mathbf{x}) = p(\mathbf{x}|\omega_f)P(\omega_f) - p(\mathbf{x}|\omega_b)P(\omega_b) \quad (9)$$

Note that the prior probability for the background, $P(\omega_b)$, is estimated using the weight of the most significant Gaussian distribution in the GMM model, and the prior probability for the foreground, $P(\omega_f)$, is estimated as $1 - P(\omega_b)$. The pixel will be classified to the foreground class if $h(\mathbf{x}) > 0$, and the foreground mask is defined by the foreground pixels.

Fig. 1 (a) shows a video frame from an NJDOT traffic video. The background, which is derived using the GMM model, is shown in Fig. 1 (b), and the foreground (with shadow), which is detected using our new foreground detection method, is displayed in Fig. 1 (c). As shown in Fig. 1, the foreground includes both the foreground objects and their cast shadows. Next, we will present our novel cast shadow detection method to detect and remove the cast shadows from the foreground.

4. A novel cast shadow detection method

In video analysis, shadows are often detected as part of the foreground, which deteriorates the performance of many video analysis tasks. We therefore present in this section a novel cast shadow detection method that is able to detect and remove the cast shadows from the foreground. In particular, Fig. 2 shows the system architecture of our proposed cast shadow detection method in video using color and statistical modeling. First, we apply the new foreground detection method introduced in the previous section to detect the foreground,

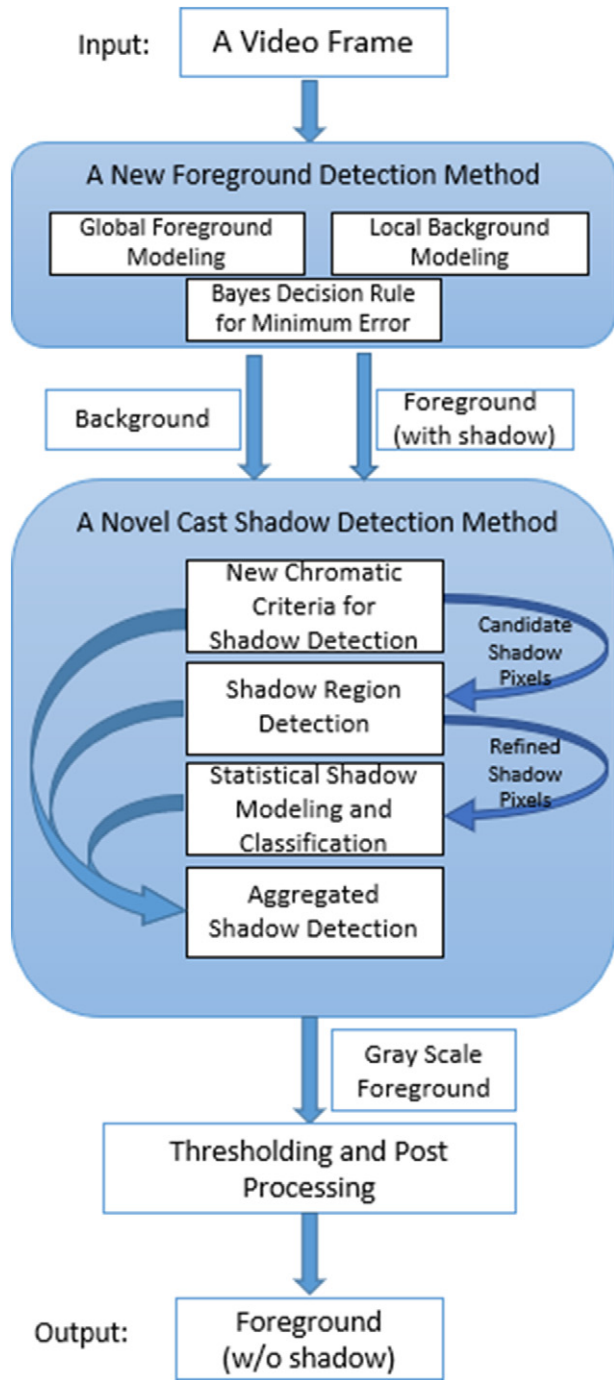


Fig. 2. The system architecture of our proposed cast shadow detection method.

which contains both the foreground objects and their cast shadows. Second, we present a cast shadow detection method with the following novelties: (i) A new method based on new chromatic criteria is presented for candidate shadow pixel detection. (ii) A shadow region detection method is proposed to cluster the candidate shadow pixels into shadow regions. (iii) A statistical shadow model is presented for shadow pixel classification. (iv) An aggregated shadow detection method is presented for final shadow detection.

4.1. New chromatic criteria for shadow pixel detection

As color provides useful information for shadow detection, we present in this section a new method based on a set of new chromatic criteria for shadow pixel detection. After foreground detection, we

need to detect the cast shadow pixels in the foreground region. Our new method will apply the new chromatic criteria to detect candidate shadow pixels. As the HSV color space is widely used in shadow detection due to its property of separating the chromaticity from intensity, we choose this color space for shadow detection. Let H , S , and V be the H (hue), S (saturation), and V (value) components in the HSV color space.

For every frame, we estimate a corresponding background picture using the GMM method. Let S_f and V_f be the S and V components of a pixel in the frame, respectively, and S_b and V_b be the S and V components of the corresponding pixel in the background picture, respectively. Our new chromatic criteria are defined as follows:

$$\begin{cases} \tau_{sl} < S_f - S_b < \tau_{sh} \\ \tau_{vl} < V_b - V_f < \tau_{vh} \end{cases} \quad (10)$$

where τ_{sl} , τ_{sh} , τ_{vl} , and τ_{vh} represent the thresholds. If a pixel in the foreground region satisfies these chromatic criteria, it is classified as a candidate shadow pixel.

To illustrate the rationale of our new chromatic criteria, we show the difference of the S component between the frame and the estimated background picture, and the difference of the V component between the estimated background picture and the frame, respectively. In particular, Fig. 3 (a) shows a color video frame,

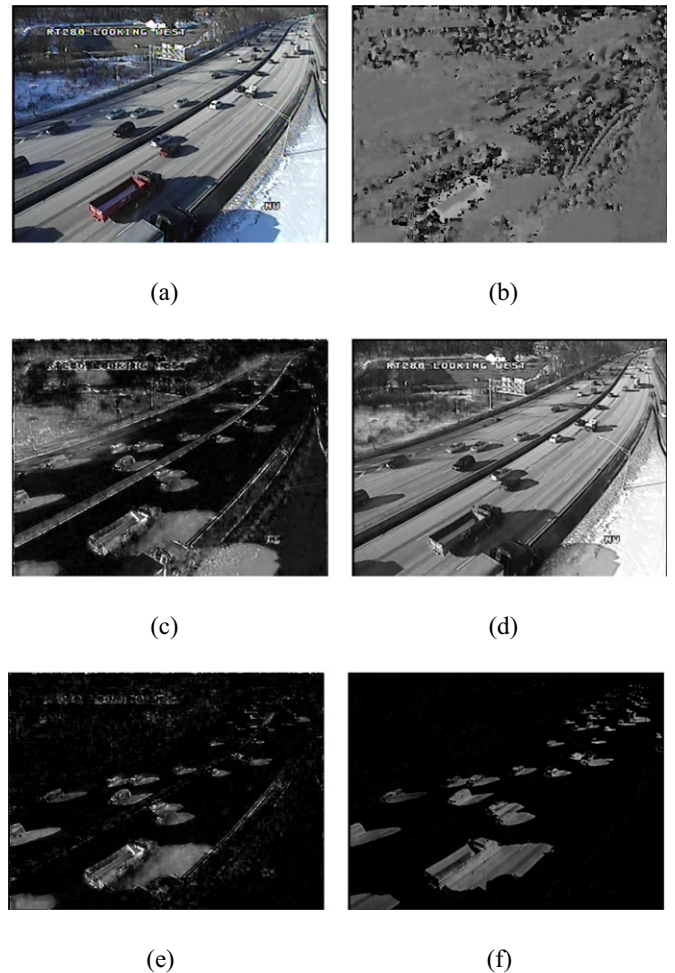


Fig. 3. (a) A video frame from an NJDOT traffic video. (b) The H (hue) component of a video frame. (c) The S (saturation) component of the video frame. (d) The V (value) component of the video frame. (e) The difference of the S component between the frame and the background. (f) The difference of the V component between the background and the frame.

Fig. 3 (b)–(d) display the H (hue), S (saturation), and V (value) components in the HSV color space, Fig. 3 (e) shows the difference of the S component between the frame and the estimated background picture, and Fig. 3 (f) shows the difference of the V component between the estimated background picture and the frame. From Fig. 3 (e), we can see that for the shadow pixels the difference values of the S component between the frame and the background are within a range that can be bounded by two threshold values τ_{sl} and τ_{sh} as shown in Eq. (10). From Fig. 3 (f), we can see that for the shadow pixels the difference values of the V component between the background and the frame also fall into a range that can be bounded by two threshold values τ_{vl} and τ_{vh} as shown in Eq. (10).

Note that many shadow detection methods assume that the shadow areas are darker in intensity but relatively invariant in chromaticity [2,18–21]. As a result, some color spaces that separate chromaticity from intensity are applied to detect shadows, such as the HSV color space [18], the c1c2c3 color space [19], and the YUV color space [20]. Some popular methods [2,21] apply a different set of chromatic criteria: $|H_f - H_b| \leq \tau_H$, $S_f - S_b \leq \tau_S$, $\beta_1 \leq V_f/V_b \leq \beta_2$, where H_f , S_f , V_f , H_b , S_b , and V_b represent the hue, saturation, and value of a pixel of the frame and the background, respectively. τ_H , τ_S , β_1 and β_2 are the thresholds that are chosen empirically. The pixels that satisfy these three criteria are classified as shadow pixels. These chromatic criteria assume that the cast shadows have similar hue to the background, but a lower S (saturation) and a lower V (value) than the background [2].

In contrast, our new chromatic criteria are more robust than these chromatic criteria. In our research, we find that the assumption that the cast shadows have similar hue to the background is often not satisfied. For example, Fig. 3 (b) shows that the H values of the cast shadows are not similar to the background. As a result, in our new chromatic criteria the H values are excluded as they vary a lot especially for the background. The S values, however, are relatively stable for the background and the cast shadows comparatively, but vary for the foreground objects. Thus, the difference of the S component between the shadow and the background often falls into a fixed range. Another characteristic of cast shadows is that the shadows are always darker than the background, but they cannot be exactly black. Based on these observations, we present our new chromatic criteria for candidate shadow pixel detection as shown in Eq. (10).

Fig. 4 shows the shadow detection results using our new chromatic criteria and the criteria in [2,21]. Specifically, Fig. 4 (a) shows a video frame from an NJDOT traffic video, Fig. 4 (b) displays the shadow detection results using the chromatic criteria in Refs.[2,21], and Fig. 4 (c) shows the shadow detection results using our new chromatic criteria. Note that the shadow pixels are represented using gray scale value of 128. We can see from Fig. 4 (b) and (c) that our proposed method using the new chromatic criteria is able to detect the shadow pixels more reliably.



(a)

(b)

(c)

Fig. 4. (a) A video frame from an NJDOT traffic video. (b) The shadow detection results (shadow pixels are represented using gray scale value of 128) using the chromatic criteria in Refs. [2,21] (c) The shadow detection results using our new chromatic criteria.

4.2. A new shadow region detection method

One inherent problem in shadow detection is that the outlines of the shadow region are often classified to the foreground class. As a result, after removing the shadow pixels from the foreground, the shadow regions are only partially removed, and the shadow outlines are often classified to the foreground. Fig. 5 (b) and (c) show the partially removed shadow regions and the shadow outlines that are not removed. These unremoved shadow regions and outlines often deteriorate the performance of video analysis tasks, such as video tracking and incident detection.

To solve this problem, we present a new shadow region detection method based on the prior knowledge that both the foreground objects and their cast shadows should define continuous regions. Note that in each frame, the detected foreground often consists of several foreground regions, each of which contains both the foreground objects and their cast shadows. In each foreground region, all the shadow pixels are on one side and all the foreground object pixels are on the other side. As a result, each foreground region may be divided into two regions: the shadow region and the foreground object region. As the candidate shadow pixels inside each foreground region are detected using the new chromatic criteria introduced in Section 4.1, the remaining pixels are the foreground object pixels.

The idea of our new shadow region detection method is to cluster the shadow pixels and the foreground object pixels into two classes using the centroids of the two classes. Our idea is similar to the K-means clustering algorithm but without any iteration steps. Specifically, in each foreground region B , we first find the centroid of the candidate shadow pixels $Cent_s(B)$ and the centroid of the foreground pixels $Cent_o(B)$. We then compute the Euclidean distances between each pixel and the two centroids. We finally classify the pixel into a foreground object class or a shadow class based on the Euclidean distances: if the distance to the foreground object class is smaller, the pixel is assigned to the foreground object class, and vice versa. In particular, for the pixel $\mathbf{x}$ at location (i, j) in each foreground region B , we calculate the distance between the pixel and the shadow centroid $Dist(\mathbf{x}_{ij}, Cent_s(B))$ and the distance between the pixel and the foreground object centroid $Dist(\mathbf{x}_{ij}, Cent_o(B))$, respectively. If $Dist(\mathbf{x}_{ij}, Cent_s(B))$ is smaller, then we classify $\mathbf{x}_{ij}$ into the shadow class. Otherwise, we classify it into the foreground object class. The new shadow region detection method thus detects the candidate shadow regions.

Fig. 5 (a) displays a video frame from an NJDOT traffic video, Fig. 5 (b) shows the shadow detection results using Huang and Chen's method [24], Fig. 5 (c) shows the shadow detection results using the new chromatic criteria introduced in Section 4.1, and Fig. 5 (d) shows the shadow detection results using the new shadow region detection method. Fig. 5 (b) and (c) reveal that the outlines of the shadow region are often classified to the foreground class leading to

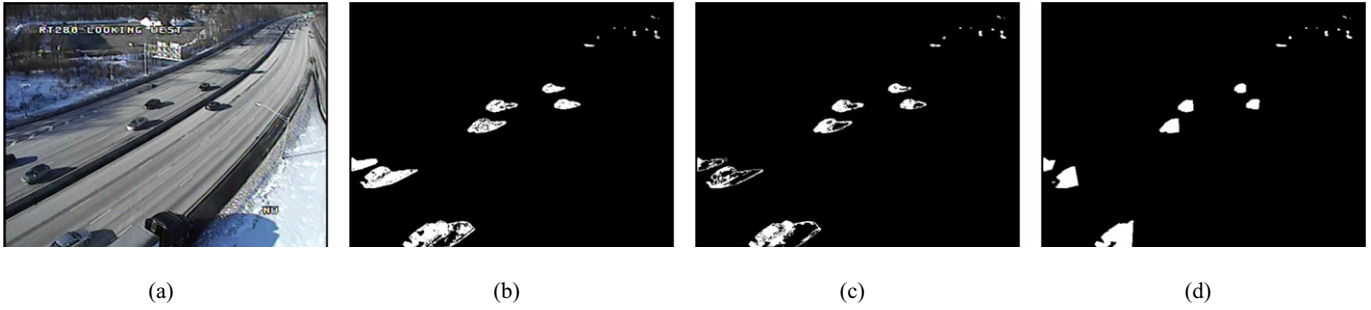


Fig. 5. (a) A video frame from an NJDOT traffic video. (b) The shadow detection results using Huang and Chen's method [24]. (c) The shadow detection results using our new chromatic criteria. (d) The shadow detection results using our shadow region detection method.

the incorrect shadow detection. In contrast, Fig. 5 (d) shows that our proposed new shadow region detection method is able to detect the whole shadow regions including their outlines.

4.3. A new statistical shadow modeling and classification method

We present in this section a new statistical shadow modeling and classification method. For statistical modeling, we use a single Gaussian distribution to model the shadow class. In the previous two sections, our proposed method using the new chromatic criteria detects candidate shadow pixels and our new shadow region detection method detects the candidate shadow regions. As the shadow pixels detected in both methods tend to be more reliable shadow pixels, we apply these shadow pixels to estimate the Gaussian distribution for the shadow class.

Specifically, let S_c and S_r be the candidate shadow pixel sets detected by our proposed method using the new chromatic criteria and our new shadow region detection method, respectively. For each pixel $\mathbf{x}$ in the foreground, if $\mathbf{x} \in S_c$ and $\mathbf{x} \in S_r$, we will use $\mathbf{x}$ to update the Gaussian distribution $N_s(\mathbf{M}, \Sigma)$ as follows:

$$\mathbf{M}' = \mathbf{M} - \alpha(\mathbf{M} - \mathbf{x}) \quad (11)$$

$$\Sigma' = \Sigma + \alpha((\mathbf{M} - \mathbf{x})(\mathbf{M} - \mathbf{x})^T - \Sigma) \quad (12)$$

where $\mathbf{M}$ and Σ are the mean vector and the covariance matrix of the shadow Gaussian distribution, respectively. α is a small number which influences the model updating speed.

For shadow pixel classification, we apply the following discriminant function for each pixel $\mathbf{x} \in \mathbb{R}^d$ in the foreground:

$$s(v_i) = (\mu_i - v_i)^2 - p \sigma_i \quad i \in \{1, 2, \dots, d\} \quad (13)$$

where v_i is the i -th element of the input vector $\mathbf{x}$, μ_i is the i -th element of the mean vector $\mathbf{M}$, σ_i is the i -th diagonal element of the covariance matrix Σ , and p is the parameter which determines the threshold. If $s(v_i)$ is greater than zero for any $i \in \{1, 2, \dots, d\}$, we classify $\mathbf{x}$ into the foreground object class. Otherwise, we classify it as a shadow pixel. Our new statistical shadow modeling and classification method thus detects the candidate shadow pixels.

the pixels in the shadow class a gray scale value of 128, and the pixels in the foreground object class a gray scale value of 255. We then define three weights for the three methods to indicate their significance for the final cast shadow detection: w_c for the new chromatic criteria, w_r for the shadow region detection, and w_s for the statistical modeling. The weights are normalized so that their summation equals one:

$$w_c + w_r + w_s = 1 \quad (14)$$

Note that the larger a weight is, the greater impact the corresponding method exerts to the final shadow detection results. These weights may be learned from the data, but without any prior information, they may be set to equal values.

For each location (i, j) in the foreground, the gray level $G(i, j)$ is calculated as follows:

$$G(i, j) = w_c C(i, j) + w_r R(i, j) + w_s S(i, j) \quad (15)$$

where $C(i, j)$, $R(i, j)$ and $S(i, j)$ are the values at location (i, j) derived by using the new chromatic criteria, the shadow region detection, and the statistical modeling method, respectively.

In the gray scale image, the smaller value a pixel has, the more likely it is a shadow pixel. We use a threshold T_s to generate a shadow free binary foreground mask. The binary value $B(i, j)$ at location (i, j) is calculated as follows:

$$B(i, j) = \begin{cases} 0, & \text{if } G(i, j) < T_s \\ 255, & \text{otherwise} \end{cases} \quad (16)$$

Fig. 6 shows the results of our novel cast shadow detection method step by step. Fig. 6 (a) is a video frame from an NJDOT traffic video. Fig. 6 (b) shows the foreground detected using our new foreground detection method. Fig. 6 (c) –(e) show the shadow detection results using the chromatic criteria detection, the shadow regions detection, and the statistical modeling detection. Note that the shadow pixels are indicated using the gray scale value of 128. Fig. 6 (f) shows the gray scale image generated by the aggregated shadow detection method. Fig. 6 (g) shows the foreground after removing the shadows. Fig. 6 (h) displays the video frame with the foreground in red color.

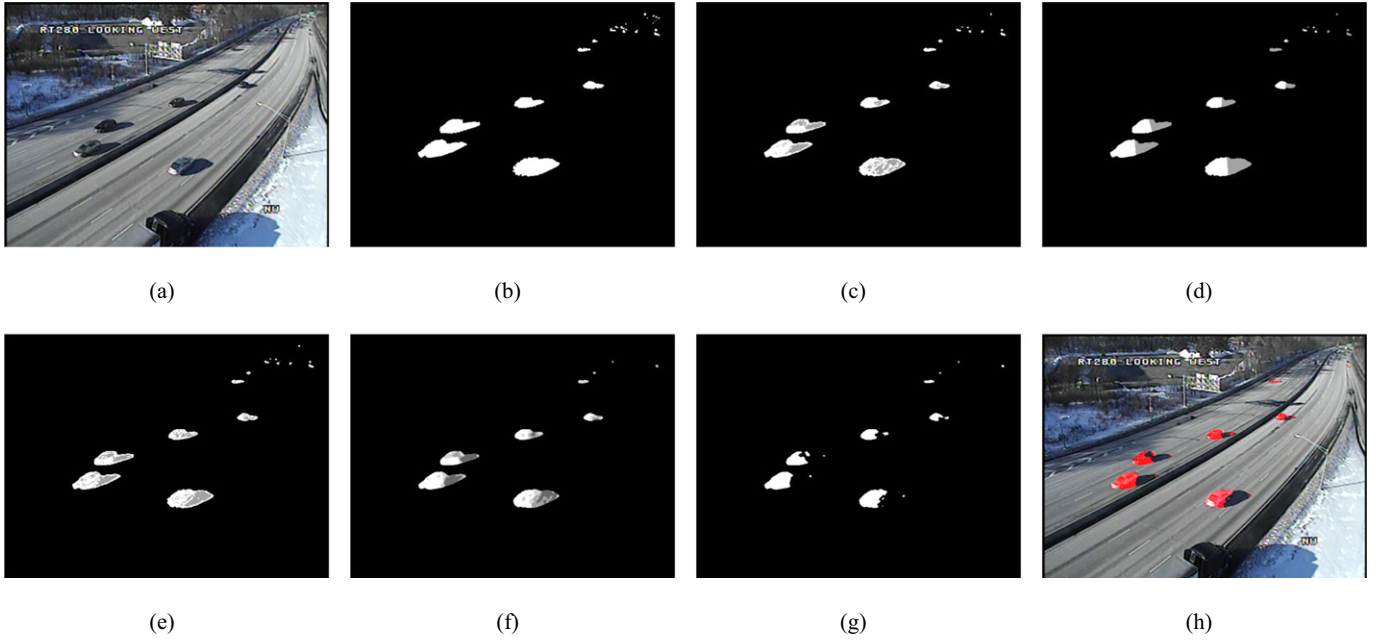


Fig. 6. (a) A video frame from the NJDOT traffic video. (b) The detected foreground (with shadow) using the new foreground detection method. (c) The detected shadow pixels using the new chromatic criteria. (d) The detected shadow regions using the shadow region detection method. (e) The detected shadow pixels using statistical shadow modeling and classification. (f) The detected shadow pixels using the aggregated shadow detection method. (g) The shadow free foreground. (h) The video frame with the foreground in red color.

5. Experiments

We first show the quantitative evaluation results using the 'Highway-1' and 'Highway-3' videos [2]. The 'Highway-1' and 'Highway-3' videos have the spatial resolution of 320×240 with a duration of 440 frames and 2227 frames, respectively. These videos, which are publicly available and broadly used, facilitate the comparative evaluation of our proposed method with other representative shadow detection methods published in the literature. We then use the New Jersey Department of Transportation (NJDOT) traffic video sequences to evaluate our proposed method qualitatively. Specifically, we apply four NJDOT traffic videos, each of which is 15 min with a frame rate of 15 frames per second or fps, and with a spatial resolution of 640×482 . We demonstrate the improvement for the vehicle tracking performance by using our proposed shadow detection method with these videos.

The thresholds τ_{sl} , τ_{sh} , τ_{vl} and τ_{vh} in Eq. (10) are defined by the saturation and value components of some manually selected shadow pixels in the first frame of the video. The weights w_c , w_r , and w_s are defined as 0.25, 0.25, and 0.5, respectively. The threshold T_s in Eq. (16) used in our experiment is 192. For fair comparison, we only provide one frame of each video for parameter initialization before testing. The model used for Zhu et al.'s method [43] is a pre-trained model.

The computer we use is a DELL XPS 8900 PC with a 3.4 GHz Intel Core i7-6700 CPU, an NVIDIA GeForce GTX 745 GPU and 16 GB RAM. All the running time is tested on the above computer. As shown in Table 1, it takes 9 ms to process each frame in the 320×240 video, and takes 39 ms to process each frame in the 640×482 videos using our method. The running speed of our proposed method is comparable to, or even faster than the other methods. As a result, our proposed shadow detection method is able to perform real time analysis of these videos.

The shadow detection rate η , the shadow discrimination rate ξ , and the F-measure are popular metrics used to evaluate shadow detection performance quantitatively [46], which are defined as follows:

$$\eta = \frac{TP_s}{TP_s + FN_s} \quad (17)$$

$$\xi = \frac{TP_o}{TP_o + FN_o} \quad (18)$$

$$F - measure = \frac{2\eta\xi}{\eta + \xi} \quad (19)$$

where TP_s and FN_s represent the number of true positive and false negative shadow pixels, respectively, and TP_o and FN_o stand for the number of true positive and false negative foreground object pixels, respectively.

Fig. 7 shows the shadow detection results of the 'Highway-1' and 'Highway-3' videos [2]. Specifically, Fig. 7 (a) shows one video frame from the 'Highway-1' video, and Fig. 7 (b) shows one video frame from the 'Highway-3' video. From left to right, top to down, each sub-figure is showing one video frame of 'Highway-1'/'Highway-3' video [2], the ground truth of foreground mask (the white parts are the foreground objects and the gray parts are the cast shadows),

Table 1

The comparative running time (in milliseconds) of our proposed method and some popular shadow detection methods.

	320 × 240	640 × 482
Sanin et al. [25]	10	44
Zhu et al. [43] (with GPU)	9	39
Proposed method	68	282
	10	43
	13	54
	421	1069
	9	39

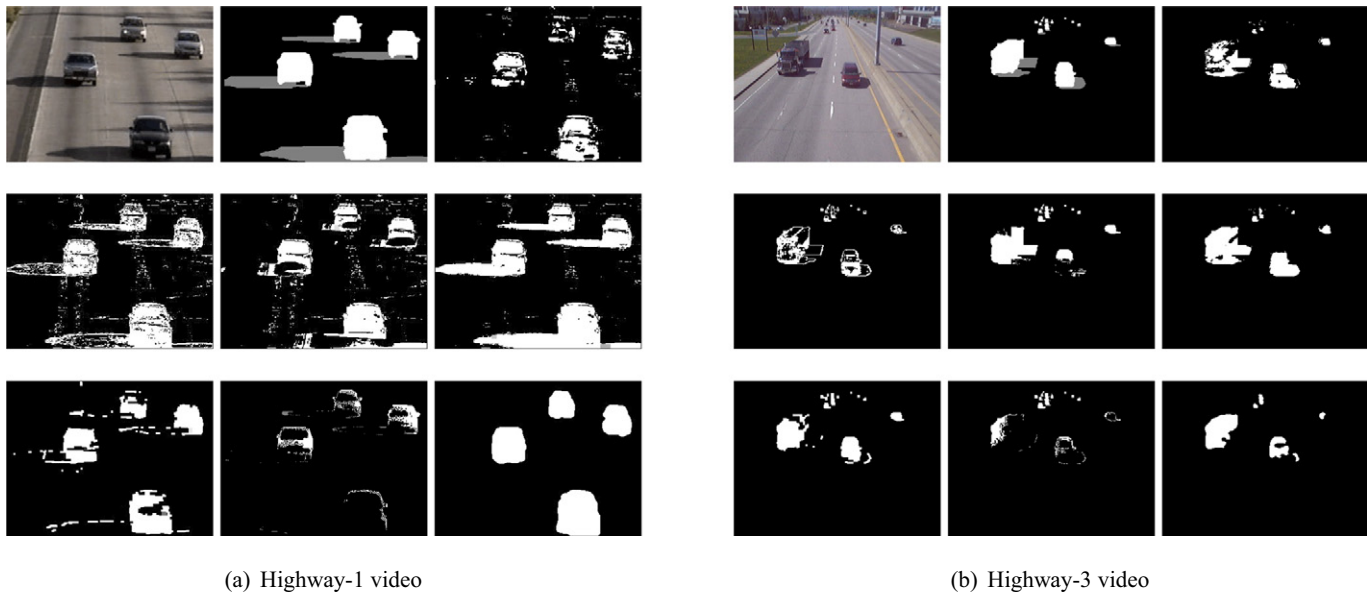


Fig. 7. The foreground masks obtained by different methods. (a). The results of ‘Highway-1’ video. (b). The results of ‘Highway-3’ video. From left to right, top to down, each sub-figure is showing one video frame of ‘Highway-1’/‘Highway-3’ video [2], the ground truth of foreground mask (the white parts are the foreground objects and the gray parts are the cast shadows), the shadow free foreground mask of Cucchiara et al.’s method [18], Huang and Chen’s method [24], Hsieh et al.’s method [22], Leone and Distant’s method [23], Sanin et al.’s method [25], Zhu et al.’s method [43], and our proposed method, respectively.

the shadow free foreground mask of Cucchiara et al.’s method [18], Huang and Chen’s method [24], Hsieh et al.’s method [22], Leone and Distant’s method [23], Sanin et al.’s method [25], Zhu et al.’s method [43], and our proposed method, respectively. We can see from Fig. 7 that our proposed shadow detection method achieves better shadow detection and removal results than the other popular shadow detection methods.

Table 2 shows the comparative shadow detection performance of our proposed method and some other popular shadow detection methods using the publicly available ‘Highway-1’ and ‘Highway-3’ videos. In particular, our proposed method achieves the highest F-measure score of 83% for ‘Highway-3’ video, compared with the 74%, 54%, 69%, 55%, 75%, and 46% F-measure scores by the Sanin et al. [25] shadow detection method, the Lalonde et al. [47] shadow detection method, the Bullkich et al. [48] shadow detection method, the Guo et al. [34] shadow detection method, the Gomes et al. [21] shadow detection method, and the Zhu et al. [43] shadow detection method, respectively. Our proposed method also achieves the highest F-measure score of 91% for ‘Highway-1’ video, which is comparable with the 91% achieved by the Gomes et al. [21] shadow detection method, and is better than all the others.

Another dataset we apply in our experiments is the NJDOT real traffic video sequences. The videos in this dataset have stronger

cast shadows and lower video quality than the ‘Highway-3’ video. These videos are the real traffic surveillance videos which include a lot of unexpected situations, such as camera jitter, network fluctuation, etc. Many shadow detection methods fail to detect shadows in these videos, but our proposed method is able to achieve good shadow detection performance on these videos. The first column in Fig. 8 shows several frames in the NJDOT traffic video, and each column shows the shadow free foreground mask by using one method. From left to right are the Cucchiara et al.’s method [18], Huang and Chen’s method [24], Hsieh et al.’s method [22], Leone and Distant’s method [23], Sanin et al.’s method [25], Zhu et al.’s method [43], and our proposed method, respectively. The significance of shadow detection in these videos is to improve the performance of video analysis tasks such as tracking and object detection. In particular, Fig. 9 shows comparatively the vehicle tracking performance using the NJDOT traffic videos: the vehicle tracking results without shadow detection and the vehicle tracking results with shadow detection using our proposed shadow detection method. We can see in the left figure that two vehicles are connected together by their cast shadows and fall into one tracking block when no shadow detection algorithm is applied. After applying our shadow detection algorithm, these two vehicles are separated into two tracking blocks. As a result, the tracking performance is more accurate.

Table 2
The quantitative shadow detection result of some popular methods [21] and our proposed method.

	Highway-1			Highway-3		
	η	ξ	F-measure	η	ξ	F-measure
Sanin et al. [25]	82%	94%	88%	62%	91%	74%
Guo et al. [34]	61%	76%	67%	39%	86%	54%
Gomes et al. [21]	72%	95%	82%	80%	61%	69%
Zhu et al. [43]	68%	75%	71%	42%	82%	55%
Proposed method	88%	94%	91%	65%	90%	75%
	95%	36%	53%	88%	32%	46%
	89%	94%	91%	90%	76%	83%

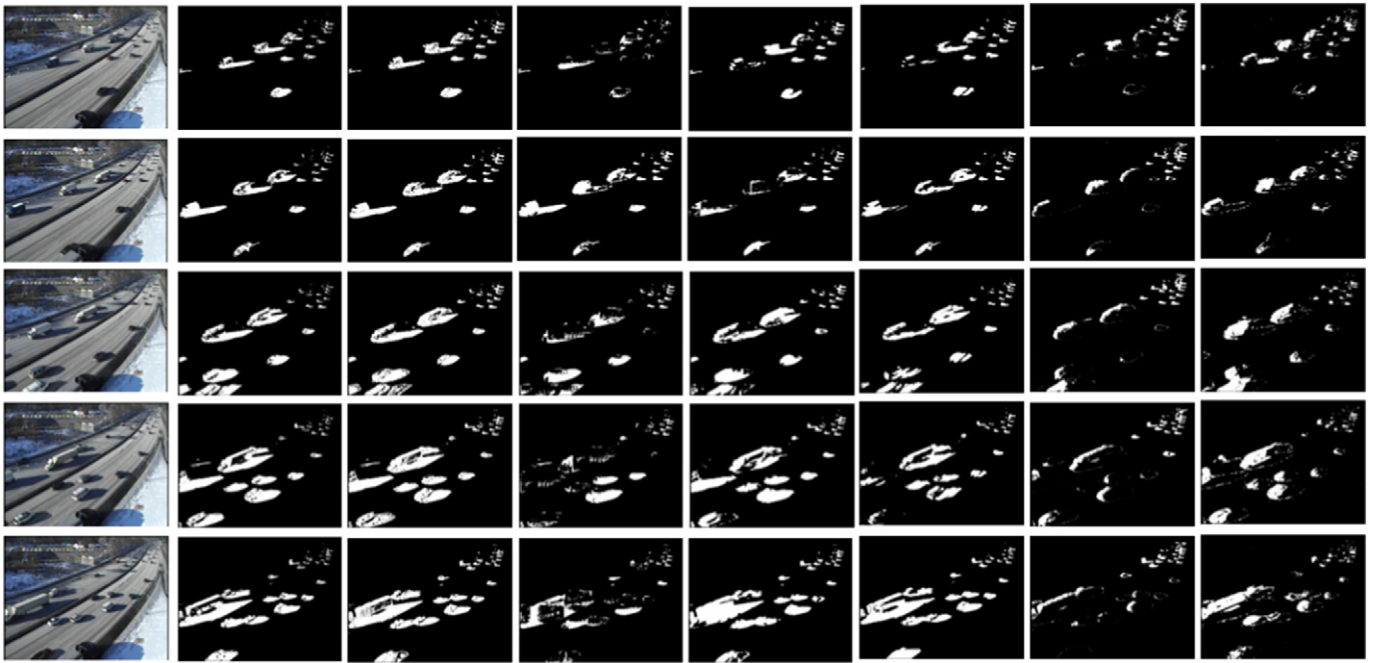


Fig. 8. The comparison of shadow detection performance of different methods. From left to right are the original video frames, the shadow free foreground masks of Cucchiara et al.'s method [18], the shadow free foreground masks of Huang and Chen's method [24], the shadow free foreground masks of Hsieh et al.'s method [22], the shadow free foreground masks of Leone and Distanto's method [23], the shadow free foreground masks of Sanin et al.'s method [25], the shadow free foreground masks of Zhu et al.'s method [43], and the shadow free foreground masks of our proposed method, respectively.

6. Conclusion

We have presented in this paper a novel cast shadow detection method for traffic video analysis. The major contributions of our proposed method are five-fold. First, we introduce a new foreground detection method, which integrates the novel Global Foreground Modeling (GFM) method, the Gaussian Mixture Model (GMM), and the Bayes decision rule for minimum error to obtain the foreground with shadows. Second, we propose a set of new chromatic criteria for shadow pixels differentiation. Third, we use a shadow region detection method to detect the continuous shadow regions based on the property of cast shadows. Fourth, we build a statistical shadow model to model and classify the shadow pixels with a single Gaussian distribution. The model keeps learning and updating to adapt to the changes of the environment. Fifth, we use an aggregated shadow detection method to combine the shadow detection results from the previous three steps. A weighted summation strategy is used to aggregate the candidate shadow detection results.

The experimental results using the publicly available data 'Highway-1' and 'Highway-3' videos, and the NJDOT real video sequences have shown that (i) our proposed method achieves better

shadow detection performance than other popular shadow detection methods, (ii) our proposed method is able to detect cast shadows in low quality videos, such as the NJDOT videos, while in comparison other methods fail to detect the shadows, and (iii) our method can help improve the results of traffic video analysis.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This paper is partially supported by the NSF grant 1647170.

References

- [1] A. Prati, I. Mikic, M.M. Trivedi, R. Cucchiara, Detecting moving shadows: algorithms and evaluation, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (7) (2003) 918–923.
- [2] A. Sanin, C. Sanderson, B.C. Lovell, Shadow detection: a survey and comparative evaluation of recent methods, *Pattern Recognit.* 45 (4) (2012) 1684–1695.
- [3] R. Mahajan, A. Bajpayee, A survey on shadow detection and removal based on single light source, 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), IEEE, 2015, pp. 1–5.
- [4] H.-S. Song, S.-N. Lu, X. Ma, Y. Yang, X.-Q. Liu, P. Zhang, Vehicle behavior analysis using target motion trajectories, *IEEE Trans. Veh. Technol.* 63 (8) (2014) 3580–3591.
- [5] C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, *CVPR*, IEEE, 1999, pp. 2246.
- [6] E. Hayman, J. Eklundh, Statistical background subtraction for a mobile observer, 14–17 October 2003, 9th IEEE International Conference on Computer Vision (ICCV 2003), IEEE Computer Society, Nice, France, 2003, pp. 67–74.
- [7] Z. Zivkovic, Improved adaptive Gaussian mixture model for background subtraction, August 23–26, 2004, 17th International Conference on Pattern Recognition, ICPR 2004, IEEE Computer Society, Cambridge UK, 2004, pp. 28–31.



Fig. 9. The comparison of vehicle tracking performance using a frame from the NJDOT traffic videos. Left: the vehicle tracking results without shadow detection. Right: the vehicle tracking results with shadow detection using our proposed shadow detection method.

- [8] J. Zhou, D. Gao, D. Zhang, Moving vehicle detection for automatic traffic monitoring, *IEEE Trans. Veh. Technol.* 56 (1) (2007) 51–59.
- [9] A. Sobral, A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, *Comput. Vis. Image Underst.* 122 (2014) 4–21.
- [10] H. Lee, H. Kim, J.-I. Kim, Background subtraction using background sets with image-and color-space reduction, *IEEE Trans. Multimedia* 18 (10) (2016) 2093–2103.
- [11] Y. Xu, J. Dong, B. Zhang, D. Xu, Background modeling methods in video analysis: a review and comparative evaluation, *CAAI Trans. Internet Technol.* 1 (1) (2016) 43–60.
- [12] T. Bouwmans, A. Sobral, S. Javed, S.K. Jung, E. Zahzah, Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset, *Computer Sci. Rev.* 23 (2017) 1–71.
- [13] T. Bouwmans, C. Silva, C. Marghes, M.S. Zitouni, H. Bhaskar, C. Frelicot, On the role and the importance of features for background modeling and foreground detection, *Computer Sci. Rev.* 28 (2018) 26–91.
- [14] A. Boulmerka, M.S. Allili, Foreground segmentation in videos combining general Gaussian mixture modeling and spatial information, *IEEE Trans. Circuits Syst. Video Technol.* 28 (6) (2018) 1330–1345.
- [15] H. Shi, C. Liu, A new foreground segmentation method for video analysis in different color spaces, 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 2899–2904.
- [16] H. Shi, C. Liu, A new global foreground modeling and local background modeling method for video analysis, *International Conference on Machine Learning and Data Mining in Pattern Recognition*, Springer, 2018, pp. 49–63.
- [17] M. Pandey, S. Lazebnik, Scene recognition and weakly supervised object localization with deformable part-based models, November 6–13, 2011, in: D.N. Metaxas, L. Quan, A. Sanfeliu, L.J.V. Gool (Eds.), *IEEE International Conference on Computer Vision, ICCV 2011*, IEEE Computer Society, Barcelona, Spain, 2011, pp. 1307–1314.
- [18] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, Detecting moving objects, ghosts, and shadows in video streams, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (10) (2003) 1337–1342.
- [19] E. Salvador, A. Cavallaro, T. Ebrahimi, Cast shadow segmentation using invariant color features, *Comput. Vis. Image Underst.* 95 (2) (2004) 238–259.
- [20] C.-T. Chen, C.-Y. Su, W.-C. Kao, An enhanced segmentation on vision-based shadow removal for vehicle detection, 2010 International Conference on Green Circuits and Systems (ICGCS), IEEE, 2010, pp. 679–682.
- [21] V. Gomes, P. Barcellos, J. Scharcanski, Stochastic shadow detection using a hypergraph partitioning approach, *Pattern Recognit.* 63 (2017) 30–44.
- [22] J.-W. Hsieh, W.-F. Hu, C.-J. Chang, Y.-S. Chen, Shadow elimination for effective moving object detection by Gaussian shadow modeling, *Image Vis. Comput.* 21 (6) (2003) 505–516.
- [23] A. Leone, C. Distanto, Shadow detection for moving objects based on texture analysis, *Pattern Recognit.* 40 (4) (2007) 1222–1233.
- [24] J.-B. Huang, C.-S. Chen, Moving cast shadow detection using physics-based features, *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. CVPR 2009, IEEE, 2009, pp. 2310–2317.
- [25] A. Sanin, C. Sanderson, B.C. Lovell, Improved shadow removal for robust person tracking in surveillance scenarios, 20th International Conference on Pattern Recognition (ICPR), 2010, IEEE, 2010, pp. 141–144.
- [26] B. Sun, S. Li, Moving cast shadow detection of vehicle using combined color models, 2010 Chinese Conference on Pattern Recognition (CCPR), IEEE, 2010, pp. 1–5.
- [27] A. Amato, M.G. Mozerov, A.D. Bagdanov, J. Gonzalez, Accurate moving cast shadow suppression based on local color constancy detection, *IEEE Trans. Image Process.* 20 (10) (2011) 2954–2966.
- [28] I. Huerta, M.B. Holte, T.B. Moeslund, J. González, Chromatic shadow detection and tracking for moving foreground segmentation, *Image Vis. Comput.* 41 (2015) 42–53.
- [29] S. Nadimi, B. Bhanu, Physical models for moving shadow and object detection in video, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (8) (2004) 1079–1087.
- [30] Y. Wang, Real-time moving vehicle detection with cast shadow removal in video based on conditional random field, *IEEE Trans. Circuits Syst. Video Technol.* 19 (3) (2009) 437–441.
- [31] L.Z. Fang, W.Y. Qiong, Y.Z. Sheng, A method to segment moving vehicle cast shadow based on wavelet transform, *Pattern Recognit. Lett.* 29 (16) (2008) 2182–2188.
- [32] C.-C. Chen, J.K. Aggarwal, Human shadow removal with unknown light source, 2010 International Conference on Pattern Recognition, IEEE, 2010, pp. 2407–2410.
- [33] D. Xu, J. Liu, X. Li, Z. Liu, X. Tang, Insignificant shadow detection for video segmentation, *IEEE Trans. Circuits Syst. Video Technol.* 15 (8) (2005) 1058–1064.
- [34] R. Guo, Q. Dai, D. Hoiem, Paired regions for shadow detection and removal, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (12) (2013) 2956–2967.
- [35] T.F.Y. Vicente, M. Hoai, D. Samaras, Leave-one-out kernel optimization for shadow detection and removal, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (3) (2018) 682–695.
- [36] X. Hu, L. Zhu, C.-W. Fu, J. Qin, P.-A. Heng, Direction-aware spatial context features for shadow detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7454–7462.
- [37] Y. Vicente, F. Tomas, M. Hoai, D. Samaras, Leave-one-out kernel optimization for shadow detection, *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3388–3396.
- [38] L. Shen, T. Wee Chua, K. Leman, Shadow optimization from structured deep edge detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2067–2074.
- [39] S.H. Khan, M. Bennamoun, F. Sohel, R. Togneri, Automatic shadow detection and removal from a single image, *IEEE Trans. Pattern Anal. Mach. Intell.* 3 (2016) 431–446.
- [40] L. Qu, J. Tian, S. He, Y. Tang, R.W. Lau, Deshadownet: a multi-context embedding deep network for shadow removal, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1, 2017, pp. 3.
- [41] Y. Wang, X. Zhao, Y. Li, X. Hu, K. Huang, N. CRIPAC, Densely cascaded shadow detection network via deeply supervised parallel fusion., *IJCAI*, 2018, pp. 1007–1013.
- [42] H. Le, Y. Vicente, F. Tomas, V. Nguyen, M. Hoai, D. Samaras, A+ D net: training a shadow detector with adversarial shadow attenuation, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 662–678.
- [43] L. Zhu, Z. Deng, X. Hu, C.-W. Fu, X. Xu, J. Qin, P.-A. Heng, Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection, *The European Conference on Computer Vision (ECCV)*, 2018.
- [44] A. Mittal, D.P. Huttenlocher, Scene modeling for wide area surveillance and image synthesis, 13–15 June 2000, Hilton Head, SC, USA. 2000 Conference on Computer Vision and Pattern Recognition (CVPR 2000), IEEE Computer Society, 2000, pp. 2160–2167.
- [45] A.R. Webb, *Statistical Pattern Recognition*, John Wiley & Sons, 2003.
- [46] W. Ji, Y. Zhao, Moving cast shadow detection using joint color and texture features based on direction and distance, 2016 2nd IEEE International Conference on Computer and Communications (ICCC), IEEE, 2016, pp. 439–444.
- [47] J.-F. Lalonde, A.A. Efros, S.G. Narasimhan, Detecting ground shadows in outdoor consumer photographs, *European Conference on Computer Vision*, Springer, 2010, pp. 322–335.
- [48] E. Bullick, I. Ilan, Y. Moshe, Y. Hel-Or, H. Hel-Or, Moving shadow detection by nonlinear tone-mapping, 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 2012, pp. 146–149.