Illumination-Aware Image Segmentation for Real-Time Moving Cast Shadow Suppression

Hadi Ghahremannezhad
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
Email: hg255@njit.edu

Hang Shi
Innovative AI Technologies
Newark, NJ 07103, USA
Email: hang@iaitusa.com

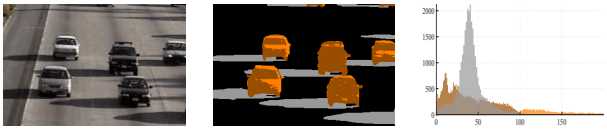
Chengjun Liu
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102, USA
Email: cliu@njit.edu

***Abstract*—One of the main challenges facing foreground detection methods is the performance deterioration due to shadows cast by moving objects. In this paper, a new real-time method is proposed that integrates various cues for region-wise classification to deal with achromaticity and camouflage issues in suppressing cast shadows. Specifically, after background subtraction, a locally near-invariant illumination feature is used as input for watershed segmentation approach to extract a number of superpixels. The superpixels are further merged according to three illumination criteria with the purpose of constructing segments that are locally homogeneous in terms of illumination variations. These segments are then classified according to the number of potential shadow candidates, gradient direction correlation, and the number of external boundary points. The potential shadow candidates are extracted by establishing a set of chromatic criteria in the HSV color-space. The gradient correlation is considered due to the fact that shadows do not impose considerable variations in the gradient directions. On the other hand, shadow segments contain a notable number of extrinsic boundary points which is used as an additional cue. Final shadow detection is achieved by integrating the outputs of the previous steps. The experimental results using publicly available videos from ATON dataset show the feasibility of our proposed method for real-time applications. The code is publicly available at: http://github.com/hadi-ghnd/ShadowDetection.**

I. Introduction

Detecting moving objects is a fundamental step in many applications, such as video surveillance, traffic monitoring, content-based video coding, gesture recognition, and human-computer interaction [1]. One of the main challenges in foreground detection is the shadows cast by moving objects on the background, which are often classified as part of the foreground as a result of their similar movement patterns to the moving objects. This misclassification can have severe negative effects on the performance of the further steps in the video analysis systems, such as object classification [2], segmentation [3]–[6], and object tracking [7], [8].

The task of shadow removal has been addressed in many studies, which have been grouped into seven categories based on the methodologies and exploited features [10], such as color [11] and texture features [12], statistical modeling [13], or a combination of features [14], [15]. Recently, some methods have applied deep convolutional neural networks (DCNNs) for shadow detection [16]–[18]. However, these techniques are not suitable for many real-world applications due to requiring a large amount of training data and high demand for computational resources. There are a number of other shortcomings in the existing shadow removal methods, such as being limited to specific applications or the requirement for manually specifying sensitive parameters.



(a) (b) (c)

Fig. 1: Histogram of RGB norm ratios. (a) Sample video frame [9]. (b) Lighter, darker, and shadowed samples represented by orange, brown, and gray, respectively. (c) Histogram of the RGB norm ratios.

In this paper, a real-time method is proposed to detect and suppress moving shadows with minimal manual involution. First, the global foreground modeling (GFM) method [19] is applied for foreground segmentation due to its efficiency and robustness. Therefore, we employ a region-based classification method, which is capable of dealing with achromaticity and camouflage issues. The watershed segmentation approach [20] is applied in order to extract superpixels. A locally near-invariant illumination feature is applied to merge correlated superpixels and segment the foreground into a number of regions. These regions are then classified based on the number of candidate shadow samples, foreground-background gradient direction correlation, and the number of external terminal points. At the end, the results of all the three steps are integrated for final shadow detection. This integration results in an accurate and robust shadow detection method for real-time video analytics applications. Figure 2 shows the system architecture of the proposed shadow detection method.

The remainder of this paper is organized as follows. In section II the major steps of the proposed method, including image segmentation (section II-A) and region classification (section II-B) are described in detail. The performance of the proposed method is evaluated on publicly available data in section III and IV is the conclusion of the paper.

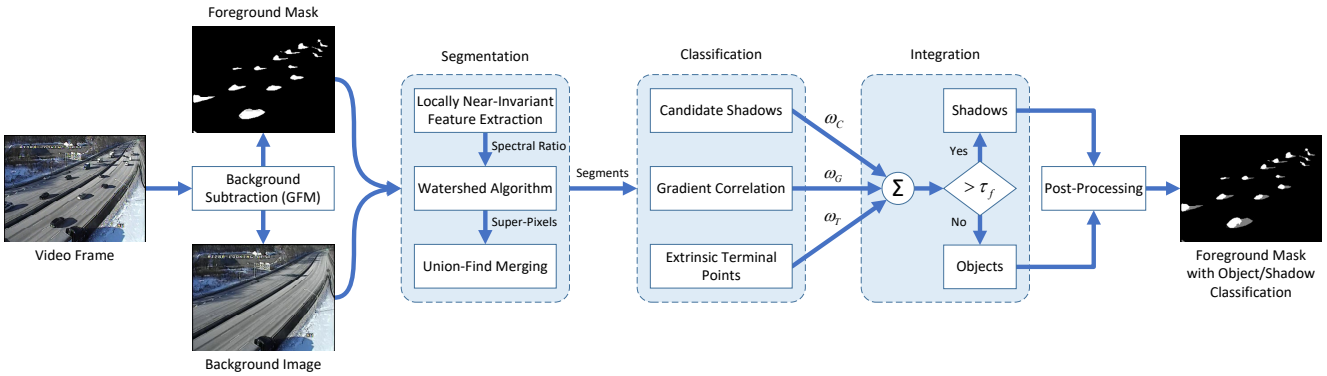


Fig. 2: The general overview of the system architecture of the proposed shadow detection method

II. A NEW CAST SHADOW DETECTION METHOD

In order to subtract the background GFM method [19] is applied which results in a binary motion mask $\mathcal{M}(x,y)$ where $\mathcal{M}(x,y) = 1$ indicates there is significant motion at location (x,y), either caused by an object or moving cast shadow and $\mathcal{M}(x,y) = 0$ means the location (x,y) belongs to the stationary background. The goal here is to classify the foreground pixels into object and shadow classes in order to disregard the moving cast shadows in the further tasks of video analytics. The details of the proposed multi-layer shadow detection method are discussed in this section.

A. Image segmentation based on locally near-invariant illumination feature

Pixel-wise approaches fail to differentiate between shadows and dark objects that have similar color values (see Figure 1) as they are limited only to the variations in the RGB values and do not take the spatial relations among each pixel and its neighborhood into account. Therefore, a combination of pixel-based and region-based techniques can help with locating the dark objects and reduce the misclassification errors. Here, we first apply component analysis [21] in order to partition the binary motion mask $\mathcal{M}(x,y)$ into a set of independent components $\mathbb{R} = \{r_1, r_2, ..., r_k\}$.

By assuming that most locations in the scene have rough Lambertian surfaces with negligible specular reflection, there is a single dominant illumination source, there is a specific geometry with constant scene angles, and the camera filters have infinitely narrow bandwidth [11] we can express the camera sensor responses at location (x,y) as follows:

$$C_k(x,y) = q_k\, E(\lambda_k,x,y) S(\lambda_k,x,y) \tag{1}$$

where $\lambda_k, k \in \{R,G,B\}$ represents the central frequency of the k-th channel camera filter, q_k, $k \in \{R,G,B\}$ indicates the spectral sensitivities of the three color camera sensors, and $E(\lambda,x,y)$ and $S(\lambda,x,y)$ are the incident illumination and surface reflectance at location (x,y), respectively [22]. This response can be expressed by the contributions of the direct C_k^d and ambient C_k^a illumination components [23] as follows:

$$C_k = \alpha C_k^d + C_k^a = \alpha q_k E_k^d S_k^d + q_k E_k^a S_k^a, k \in \{R,G,B\} \tag{2}$$

where $\alpha \in [0,1]$ is the attenuation factor that accounts for the unblocked 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.

With the assumption of $\alpha = 1$ in the background and negligible variations in the ambient illumination, we can define spectral ratio $\overrightarrow{S} = [S_R, S_G, S_B]^T$ as a near-invariant illumination feature:

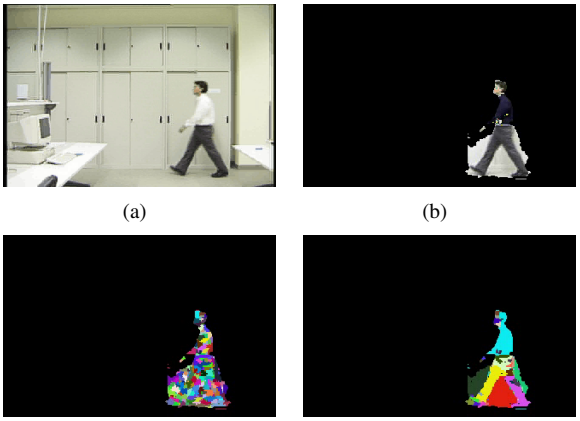
$$S_k = \frac{FG}{BG} = \frac{q_k E_k^d S_k^d}{\alpha q_k E_k^d S_k^d + q_k E_k^a S_k^a} \tag{3}$$

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 ($\alpha = 0$) and the surface material is the same at location (x,y) in the foreground and background when shadowed ($S_k^d = S_k^a$), the spectral ratio in this region can be indicated as follows:

$$S_k = \frac{E_k^d}{E_k^a} \tag{4}$$

which is near-constant among neighboring pixels across the umbra region and the changes are mostly because of the variations in the ambient illumination (Figure 3(b)).

We apply the watershed segmentation approach [20] on the spectral ratios of each region in $\mathbb{R}$ to obtain the superpixels. Afterward, correlated superpixels are merged by applying the union-find algorithm [24]. Due to the ratio-invariance property of shadows, two neighboring superpixels are merged if their spectral ratio differences are less than a small threshold across all three color channels. In addition, the edge between two superpixel may have been caused by intersecting shadows, which are difference-invariant [25]. Therefore, two neighboring segments are merged if the difference between the foreground values is close to the difference between their background values. Another possible scenario is if the moving



(a) (b) (c) (d)

Fig. 3: The segmentation process of each frame. (a) Original video frame. (b) Spectral ratio. (c) superpixels. (d) Merged segments based on eq. (5).

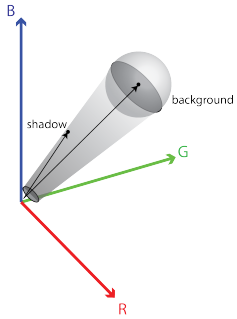


Fig. 4: The initial shadow candidate detection. Pixels falling in the conic region are considered as potential shadow samples.

shadow is cast over an existing stationary shadow. In this case, the background values are different, but the foreground values are similar and close to the background value of the darker segment.

Two neighboring superpixels/segments s_i and s_j are merged according to three criteria:

$$\begin{cases} {}^{FG_i}/_{BG_i} \approx {}^{FG_j}/_{BG_j} \\ FG_i - FG_j \approx BG_i - BG_j \\ BG_j \approx FG_i \approx FG_j, \quad BG_i \gg BG_j \end{cases} \tag{5}$$

If any of the above conditions hold, the two segments will be merged. Figure 3 shows an example of the segmentation and merging process. At this point, each foreground component $r_k \in \mathbb{R}$ is partitioned into a number of segments s_l^k, such that:

$$\bigcup_{l=1}^{n_k} s_l^k = r_k, \quad \bigcap_{l=1}^{n_k} s_l^k = \varnothing, \quad \bigcup_{k=1}^{K} r_k = \mathbb{R} \tag{6}$$

where n_k is the total number of segments s_l^k at each region r_k (Figure 3(c)). Note that the efficiency of this method is much higher than pixel-wise segmentation methods [26] due to the use of superpixels and applying the union-find algorithm. This way if two pixels belong to the same superpixel/segment there is no need for calculating the merging criteria. Otherwise the dissimilarity measures are calculated in order of priority only for the neighboring pixels of two separate superpixels/segments. If the two superpixels/segments are decided to be merged, all the pixels corresponding to them will be merged at the same time. Figure 3 illustrates the steps of the segmentation method in a sample video frame. The white and gray colors represent the 0 and 1 values in the binary masks, respectively.

B. Segment classification based on various heuristic cues

In order to improve the robustness and accuracy of the object/shadow classification of the foreground, we employ four various heuristic cues simultaneously, including thresholds, gradient correlation, and the number of extrinsic boundary points. The classification results of all three steps are aggregated for final classification. In this section, the steps in the segment classification process are explained in detail.

1) Extracting candidate shadow pixels: Since the HSV color-space separates the chromaticity from the intensity to a good level and is useful to distinguish the variations in illumination from the changes in material. Figure 4 illustrates the potential shadow zone in the RGB color-space which is a portion of the conic region in the RGB space Since shadows have little to no effect on the H(hue) component of the HSV color-space, we choose the S(saturation) and V(value) components to set the criteria. The value ratio can roughly specify the attenuation which is represented by the vector magnitudes and the saturation component can determine the apex angle of the cone which depends on the ambient illumination. By assuming S_f, V_f, S_b, and V_b to be the saturation and value components of the foreground and background, respectively, the chromatic criteria can be formulated as follows:

$$\mathcal{P}(x,y) = \begin{cases} 1, & (\tau_{vl} < {}^{V_f}/_{V_b} < \tau_{vh}) \\ & \wedge (\tau_{sl} < S_f - S_b < \tau_{sh}) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\mathcal{P}$ is a binary mask where $\mathcal{P}(x,y) = 1$ indicates that pixel at location (x,y) is a potential shadow sample, and $\tau_{vl}, \tau_{vh}, \tau_{sl}$, and τ_{sh} denote the lower and upper thresholds for the value ratio and saturation variation, respectively. All the foreground pixels that satisfy these criteria are considered to be potential shadow candidate samples. Figure 5 illustrates an example of potential shadows represented by gray color.

Each segment s_l^k of region r_k is classified to object or shadow according to its intersection with potential shadow candidates. If most pixels inside a segment are classified as potential shadow candidates that segment is likely to belong to the shadow class. This can be expressed as follows:

$$\mathcal{C}(s_l^k) = \begin{cases} 1, \text{if } \frac{|\mathcal{P} \cap s_l^k|}{|s_l^k|} > \tau_p \\ 0, \text{otherwise} \end{cases} \tag{8}$$

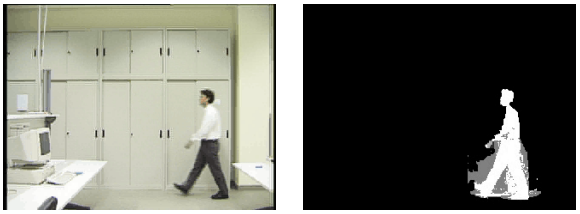


Fig. 5: Extracting potential shadow candidates. (a) Sample video frame. (b) Potential shadows.

where $\mathcal{C}(s_l^k)$ is an binary mask where $\mathcal{C}(s_l^k) = 1$ if more than τ_p of the pixels in segment s_l^k are classified as potential shadows.

2) Calculating the gradient direction correlation: The amount of gradient information introduced by the objects is generally more than the amount introduced by shadows. The dominant edges are extracted by applying the Canny edge detection method and the difference in gradient direction between the frame and the background is calculated as follows:

$$\Delta\theta(x,y) = cos^{-1}\frac{\overrightarrow{\nabla} f(x,y)\cdot\overrightarrow{\nabla} b(x,y)}{\left\|\overrightarrow{\nabla} f(x,y)\right\|\left\|\overrightarrow{\nabla} b(x,y)\right\|} \tag{9}$$

where $\overrightarrow{\nabla} f(x,y)$ and $\overrightarrow{\nabla} b(x,y)$ are the gradient vectors at location (x,y) in the frame and the background, respectively, and $\Delta\theta(x,y)$ is the angular distance of two vectors. If the gradient direction is highly correlated between the frame and the background in a segment it has a higher probability of belonging to the shadow class. This criterion is expressed as follows:

$$\mathcal{G}(s_l^k) = \begin{cases} 1, \text{if } \frac{1}{|s_l^k|}\sum_{i=1}^{|s_l^k|} H(\Delta\theta_i - \tau_a) > \tau_e \\ 0, \text{otherwise} \end{cases} \tag{10}$$

where $|s_l^k|$ is the number of pixels in the segment s_l^k, $H(.)$ denotes the unit step function which is one if the angular distance is larger than or equal to a threshold τ_a, and $\mathcal{G}(s_l^k)$ is a binary mask which is one if a fraction more than τ_e of the pixels in segment s_l^k have similar gradient direction in the frame and the background.

3) Computing the number of extrinsic terminal points: Another observation about shadow samples is their spatial distribution around the objects which results in shadow segments of each region containing a considerable number of extrinsic terminal points. Such criterion can be expressed in a binary mask $\mathcal{S}_t$ as follows:

$$\mathcal{T}(s_l^k) = \begin{cases} 1, \text{if } \frac{\left|\mathbb{T}(r_k)\cap\mathbb{T}(s_l^k)\right|}{\left|\mathbb{T}(s_l^k)\right|} > \tau_t \\ 0, \text{otherwise} \end{cases} \tag{11}$$

where $\mathbb{T}(r_k)$ and $\mathbb{T}(s_l^k)$ are the sets of external boundary points of the foreground component r_k and each of its segments s_l^k, respectively, and $\mathcal{T}(s_l^k)$ is a binary mask which is 1 if a fraction more than τ_t of terminal points are external.

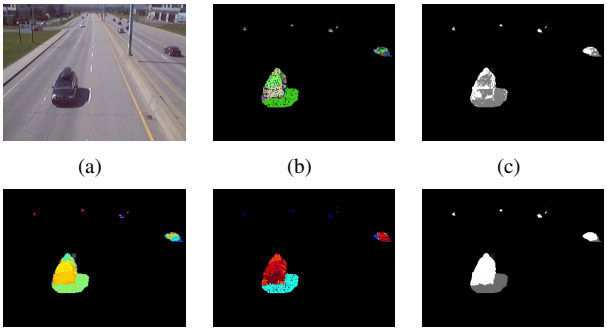


Fig. 6: The classification process of a sample frame from Highway 3 sequence. (a) Original video frame. (b) segmentation. (c) Potential shadows. (e) Heatmap of gradient correlation. (d) Heatmap of external terminal points. (f) Region-based classification ($\mathcal{S}$).

4) Final shadow detection based on integration of the previous steps: For the final object and shadow classification, the results of the previous steps are integrated by calculating the weighted summation, as follows:

$$\mathcal{W}(x,y) = w_C\mathcal{C}(x,y) + w_G\mathcal{G}(x,y) + w_T\mathcal{T}(x,y) \tag{12}$$

where $w_C \in [0,1]$, $w_G \in [0,1]$, and $w_T \in [0,1]$ are the weights indicating the significance of the shadow detection results based on chromatic criteria, gradient correlation, and extrinsic terminal points, respectively. The three weights are normalized and sum up to one:

$$w_C + w_G + w_T = 1 \tag{13}$$

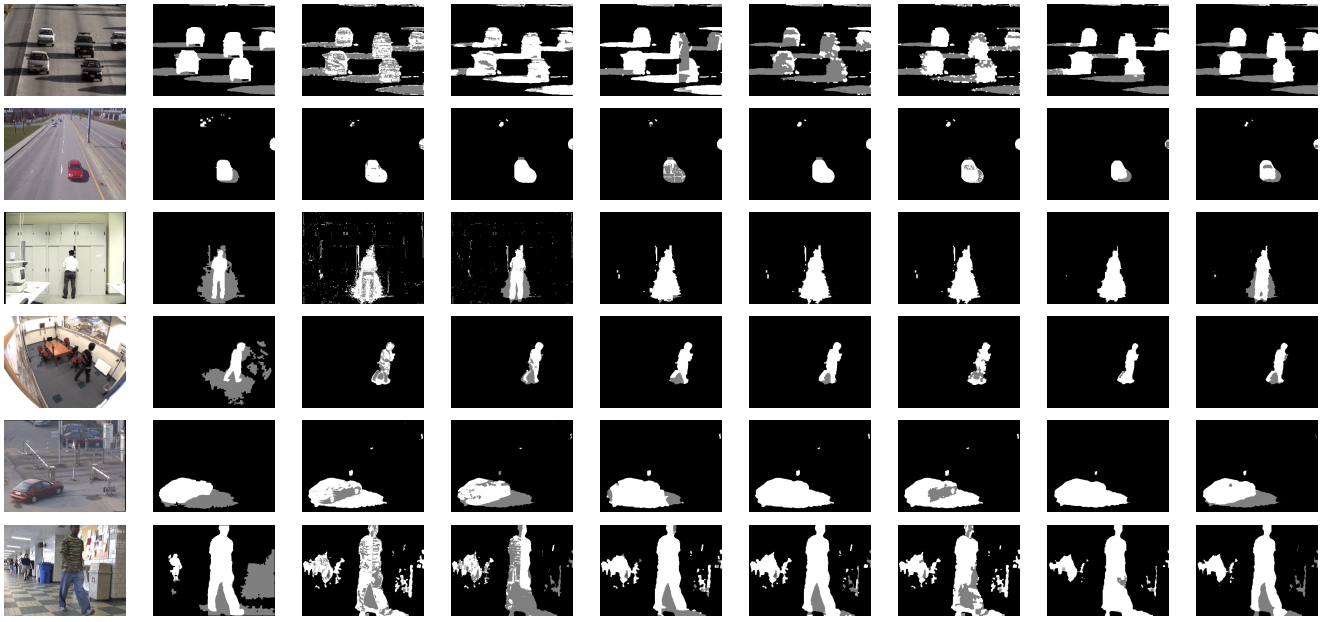
We have considered w_G to be twice the value of w_C and w_T. By thresholding the weighted sum values we obtain a binary mask $\mathcal{F}$ which represents the final shadow detection results as follows:

$$\mathcal{F}(x,y) = \begin{cases} 1, \text{if } \mathcal{W}(x,y) > \tau_f \\ 0, \text{otherwise} \end{cases} \tag{14}$$

where τ_f is a threshold, $\mathcal{F}(x,y) = 1$ indicates that the pixel at location (x,y) belongs to shadow and $\mathcal{F}(x,y) = 0$ means it belongs to moving objects. Subtracting $\mathcal{F}$ from $\mathcal{M}$ will result in a shadow-free foreground mask. Figure 6 shows an example of the described steps in the classification procedure. In the heatmaps, the warmer colors represent the objects and the colder colors represent shadows.

III. Experiments

The quantitative and qualitative results of the proposed method are evaluated using publicly available video data [9]. The spatial resolution of each video sequence is 320×240 pixels and each video contains 15 frames per second. The underlying system hardware is a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. The processing time is on average 5.48 milliseconds for each frame which is consistent with the efficiency requirements of real-time applications.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 7: The foreground masks and detected shadows in different methods using sequences from the ATON dataset [9]. Note that the ground-truth data provided in the dataset is obtained with a different background subtraction technique. (a) Original video frame. (b) Ground truth. (c),(d), (e), (f), (g), (h), and (i) are the results of the Cucchiara et al. [11], Leone and Distante [27], Hsieh et al. [28], Sanin et al. [29], Huang and Chen [13], Amato et al. [26], and our proposed method, respectively.

TABLE I: The average runtime (in milliseconds) for each video frame in different methods

Methods	Runtime (ms)	
	320×240	640×482
Cucchiara et al. [11]	23	141
Zhu et al. [30] (with GPU)	421	1069
Huang and Chen. [13]	16	81
Sanin et al. [12]	61	244
Hsieh et al. [28]	**5**	**16**
leone and Distante. [27]	135	284
Amato et al. [26]	16	102
Proposed method	**5**	**16**

TABLE II: The average runtime (in milliseconds) of the main steps in shadow detection after background subtraction

Steps	Runtime (ms)	
	320×240	640×482
pre-processing	0.32	0.55
segmentation	2.85	11.44
candidate shadows	0.07	0.28
gradient correlation	1.40	3.36
terminal points	0.27	0.56
post-processing	0.37	0.73
Total	5.29	16.93

Table I compares the run-time with some of popular shadow detection methods for video frames of size 320×240 and 640×482 pixels. Table II contains detailed run-time for each step of the process. The pre-processing step involves removing the fringe of the shadow segments and smoothing each image by Gaussian blurring. The post-processing is a noise correction step that assigns a shadow/object class to each foreground pixel according to the majority of its surrounding pixels. In Figure 7, a sample frame from some videos is illustrated along with the shadow detection results of some of the representative methods. The thresholds τ_p, τ_e, τ_t, and τ_f are all empirically set to 0.5 and show low sensitivity when experimented with various videos.

Three performance measures are calculated for quantitative evaluation of the shadow detection method as follows:

$$\begin{cases} \xi = TP_o/(TP_o + FN_o) \\ \eta = TP_s/(TP_s + FN_s) \\ F_1 = 2 \times (\eta \times \xi)/(\eta + \xi) \end{cases} \tag{15}$$

where TP_o and TP_s denote the true positive rates of the object and shadow pixels and FN_o, and FN_s are the false negative rates of the object and shadow pixels, respectively. η, ξ, and F_1 denote the shadow detection rate, shadow discrimination rate, and F-measure, respectively. In Table III the calculated measures for the performance evaluation are reported along with some of the popular methods [29].

IV. Conclusion

This paper proposes a new moving cast shadow detection method to separate the moving objects from their cast shadows in real-time applications of video analytics. After applying the global foreground modeling (GFM) method for background subtraction the foreground class contains the moving objects along with their cast shadows. At first, a set of chromatic

TABLE III: The shadow detection results compared to other methods in terms of F-measure.

	IntelligentRoom	Laboratory	Highway-1	Campus	Highway-3	Hallway
Cucchiara et al. [11]	78.18	84.33	70.36	53.22	53.40	86.35
Hsieh et al. [28]	61.26	56.51	70.55	58.88	54.61	59.20
Huang et al. [31]	71.59	54.46	56.79	55.24	48.79	75.89
Leone et al. [27]	75.27	84.69	28.69	67.39	10.58	77.92
Sanin et al. [29]	88.59	78.05	74.04	66.81	53.56	**95.60**
Wang et al. [32]	**94.63**	**90.30**	84.80	80.42	68.68	95.24
Proposed method	92.68	84.22	**88.14**	**89.92**	**84.09**	95.55

criteria in the HSV color space is applied in order to extract the potential shadow candidates. Then a segmentation technique is used based on the physical properties of the surface reflections to group the correlated pixels at each foreground component and classify the segments according to a set of three criteria. The final decision about shadow and object classification is made through an integration process of the previous steps. The experimental results demonstrate the effectiveness of the proposed method in real-time video analytics applications.

REFERENCES

[1] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Computer Science Review*, vol. 35, p. 100204, 2020.

[2] 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*, pp. 117–131, IBAI, 2019.

[3] 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)*, pp. 186–191, IEEE, 2020.

[4] H. Ghahremannezhad, H. Shi, and C. Liu, "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)*, pp. 17–24, IEEE, 2020.

[5] 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)*, pp. 000097–000102, IEEE, 2020.

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

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

[8] H. Shi, H. Ghahremannezhad, and C. Liu, "Anomalous driving detection for traffic surveillance video analysis," in *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6, IEEE, 2021.

[9] "Aton datasets by ucsd."

[10] M. Russell, J. J. Zou, and G. Fang, "An evaluation of moving shadow detection techniques," *Computational Visual Media*, vol. 2, no. 3, pp. 195–217, 2016.

[11] 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.

[12] 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*, pp. 141–144, IEEE, 2010.

[13] 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*, pp. 2310–2317, IEEE, 2009.

[14] H. Shi and C. Liu, "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.

[15] H. Ghahremannezhad, H. Shi, and C. Liu, "A new online approach for moving cast shadow suppression in traffic videos," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3034–3039, IEEE, 2021.

[16] 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)*, pp. 1007–1013, 2019.

[17] Z. Chen, L. Wan, L. Zhu, J. Shen, H. Fu, W. Liu, and J. Qin, "Triple-cooperative video shadow detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2715–2724, 2021.

[18] T. Wang, X. Hu, C.-W. Fu, and P.-A. Heng, "Single-stage instance shadow detection with bidirectional relation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1–11, 2021.

[19] H. Shi and C. Liu, "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*, pp. 49–63, Springer, 2018.

[20] F. Meyer, "Color image segmentation," in *1992 international conference on image processing and its applications*, pp. 303–306, IET, 1992.

[21] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.

[22] 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.

[23] 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*, pp. 1–8, IEEE, 2008.

[24] D. Xiang, W. Wang, T. Tang, D. Guan, S. Quan, T. Liu, and Y. Su, "Adaptive statistical superpixel merging with edge penalty for polsar image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2412–2429, 2019.

[25] F. A. Kingdom, "Perceiving light versus material," *Vision research*, vol. 48, no. 20, pp. 2090–2105, 2008.

[26] 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.

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

[28] 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.

[29] A. Sanin, C. Sanderson, and B. C. Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern recognition*, vol. 45, no. 4, pp. 1684–1695, 2012.

[30] 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)*, pp. 121–136, 2018.

[31] J.-B. Huang and C.-S. Chen, "A physical approach to moving cast shadow detection," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 769–772, IEEE, 2009.

[32] 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.