Moving Cast Shadows Segmentation Using Illumination Invariant Feature

Bingshu Wang, *Student Member, IEEE,*, Yong Zhao, *Member, IEEE,* and C. L. Philip Chen, *Fellow, IEEE*

Abstract—This paper presents an effective framework for removing moving cast shadows. Taking the reflection property of object surface for shadow regions under static and fixed scenes, an approximation estimation strategy of bidirectional reflectance distribution function as illumination invariant feature is proposed. It is valid for different types of shadow scenes. In this paper, we propose a new multiple ratios-based technique to justify shadow type for each frame: intensity ratio, area ratio and edge ratio of shadow regions are introduced. According to shadow types, several specified strategies are designed. For weak shadows, multiple features fusion strategy is employed, including color constancy, texture consistency and illumination invariant. For strong shadows, illumination invariant is utilized to detect the umbra and color constancy is utilized to detect the penumbra. Moreover, a suite of shadow direction features is firstly proposed to identify penumbra. The proposed approach is verified in fourteen video sequences varying from weak to strong shadows. The experimental results demonstrate the effectiveness and robustness of the proposed method for both indoor and outdoor scenes compared with some state-of-the-art approaches.

Index Terms—Moving shadows segmentation, bidirectional reflectance distribution function, illumination invariant, shadow direction features.

I. INTRODUCTION

MOVING cast shadows segmentation is a crucial and fundamental issue for accurate object detection in the field of video surveillance applications [1]–[4]. They are commonly generated by the occlusion of moving objects from illumination source. It is generally observed that moving cast shadows share two important visual properties with the moving objects [5], [6]: a) shadow points are easy to be detected as foreground since their significant difference from the corresponding background points; b) shadows share the very similar motion property with the objects casting them.

Bingshu Wang is with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, 999078 China, (e-mail: yb77408@umac.mo).

Yong Zhao is now with the Key Laboratory of Integrated Microsystems, School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China (e-mail: zhaoyong@pkusz.edu.cn).

C. L. Philip Chen is with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China; with Dalian Maritime University, Dalian 116026, China; and is also with School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: philip.chen@ieee.org).
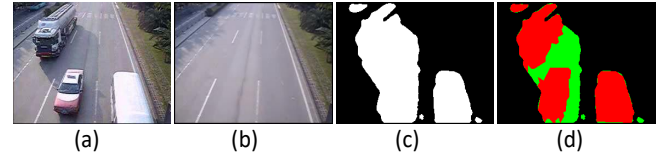

Fig. 1. The relation between background subtraction and moving cast shadows segmentation. (a) is an original frame of a vehicle scene, (b) is an estimated background by [3], (c) is a mask of moving objects generated by background subtraction [3], (d) is our shadow segmentation mask, the red and the green represent objects and shadows, respectively.

These properties attribute accurate moving cast shadows segmentation to a challenging task.

The influence caused by moving shadows should not be negligible, e.g., objects merging, silhouette distorting, erroneous location, which may easily lead to misclassification, erroneous object tracking and degraded matching of image retrieval in intelligent video analysis. Therefore, the detected moving objects in video sequences should be segmented into two classes: object and shadow. This is illustrated by Fig. 1. Object merging occurs in the detected mask due to moving shadows' effect. These two vehicles should be separated in realistic applications such as object location, recognition and tracking tasks. Figure 1. (d) is the segmentation result by the proposed method.

The difficulties for shadow segmentation may be summarized that the irregularity of illumination, whether one or multiple illumination sources, leads to the diversity and uncertainty of shadows in video sequences. Color and texture features are usually utilized within the assumption that the shadow regions show lower intensity than the corresponding background regions, but the chromacity changes little [6] and the texture is distributed uniformly [7], i.e., color constancy and texture consistency. They are always suitable for weak shadow scenes but not strong shadows. Illumination invariant model explored by [8]–[10] provide new means to segment moving cast shadows.

Our motivation is to detect different types of moving cast shadows on surveillance video sequences acquired from fixed and static cameras [11]–[13]. In this paper, a framework is proposed to address different types of shadows based on local illumination invariant by the estimation of Bidirectional Reflectance Distribution Function (BRDF) [14].

The main contributions of this paper include several parts:

- Unlike the assumptions of color constancy and texture consistency, illumination invariant of object reflectance is innovatively utilized for shadow detection. It is estimated

using pixel's neighboring information in an approximation manner, which provides an effective technique to tackle problematic situations of moving cast shadows (Section III-B).

- This paper explicitly introduces three ratios to justify the shadow types. It is capable of estimating shadow type accurately (Section III-D).
- A suite of direction features consisting of eight items is firstly proposed to help identify the penumbra of the very strong shadows. This may provide an effective means to detect the boundary of strong shadows (Section III-F).
- Quantitative and qualitative experiments on fourteen video sequences validate the proposed method's effectiveness and robustness compared with some state-of-the-art approaches.

II. THE RELATED WORK

Many algorithms or models aimed at separating moving shadows from foreground mask have been proposed [6], [12], [15], [16] in past decades. Prati *et al.* [6] proposed a survey earlier to divide shadows segmentation methods into four categories by a two-layer taxonomy. Sanin *et al.* [12] presented a feature-based taxonomy consisting of chromacity-based, physical-based, geometry-based and texture-based approaches. Al-Najdawi *et al.* [15] gave a novel taxonomy according to object or environment dependency, and implementation domain: pixel or transform; monochrome or color. Russel *et al.* [16] divided shadow detection techniques based on properties.

Chromacity-based method proposed by [17], based on the color constancy hypothesis, firstly chooses a color space with a good separation of intensity and chromacity. Then, the pixels whose intensities are lower than those of the corresponding background are selected and classified by chromacity components. The method is easy to implement and has low running time. However, it is sensitive to pixel-level noise and may fail under the lower saturation scenes.

Geometry-based method explored by Hsieh *et al.* [18] uses a histogram-based technique to separate pedestrian from moving foreground mask. Then a rough-to-fine strategy by Gaussian shadow modeling is utilized to remove the unwanted shadows. It is able to detect umbra accurately when the orientations of objects and shadows are different under single light source. However, the penumbra is hard to eliminate. The strict assumption may limit its extensibility.

Physical-based method presented by Huang *et al.* [19] does not require any assumption about illumination source and ambient light. For each pixel a Gaussian Mixture Model is created to learn its local color features. This method can reach a high accuracy on shadow detection, but it might fail when the spectral properties of objects are similar with background.

Texture-based methods [7], [12], [20], [21] are commonly based on the texture consistency hypothesis. They are operated at region level: small or large. A small region texture-based approach proposed by Leone *et al.* [7] constructs Gabor functions for texture feature extraction in a pixel's neighborhood. It is robust to pixel-level noise and effective for textured background. However, the computation is expensive.

Large region texture-based method proposed by Sanin *et al.* [12] has multiple processing stages: candidate shadows extraction, connected components analysis, the computation of gradient magnitude, and the region-level estimation of gradient direction correlations. The core is to classify each candidate shadow region into two classes: object or shadow. This method could obtain good performance on indoor scenes, but dark shadows are unmanageable for it.

Multiple features fusion [22] utilizes texture, gradient and chromacity to detect moving shadows. It shows comparative performance on both indoor and outdoor scenes. The method proposed by [10] exploits multiple features fusion to deal with shadows and achieves impressive results. These methods are selected for comparisons in this paper.

Deep learning technique was explored by Lee *et al.* [23] for moving shadow detection. It learns the shadow features automatically from input frames and background images by convolutional neural network. Notably, this method requires sufficient labeled data for training and super-pixel segmentation operation, which results in more computational time. Benedek *et al.* [24] employed spatial statistic of neighboring pixels, Markov random field model and microstructure analysis to improve shadow detection accuracy. Amato *et al.* [25] described a novel approach by means of the property of local color constancy for shadow suppression. Vitor *et al.* [26] brought about a stochastic hypergraph partitioning algorithm using chromatic and gradient information.

Wu *et al.* [27] proposed a method using patch-based edge detection for strong shadow removal. It is used for clear path detection because the images are captured by single in-vehicle camera. Russell *et al.* [28] presented an algorithm of feature-based image patch classification. It extracts features from image patches and constructs over-complete dictionaries for objects and shadows, respectively. In addition, there are many other methods, such as edge based [29], [30], statistic based [31]–[33], single image based [34]–[37], multiple features based [38]–[40], and other features [41]–[44], or techniques [45]–[50] proposed to address the problematic situations. It is not our goal to give the details of all approaches.

Based on our prior work [51], this paper proposes pixel-level reflection estimation for moving shadows segmentation. Differing from prior work, this paper explicitly designs a strategy for shadow classification. Moreover, most of the previous methods are only tested on several scenes, our method is validated on fourteen scenes varying from weak to strong, which are diverse real-world scenes.

III. THE PROPOSED METHOD

The overview of our method is as follows. The input frame, foreground mask and reference background image are given in advance. The main flowchart consists of three parts: the first is to detect candidate shadows and candidate objects (Section III-A). The second part is shadow type estimation for each frame (Section III-B, Section III-C, and Section III-D). The third part presents strategies to deal with different types of shadows (Section III-E and Section III-F). Post-processing is also designed to enhance the integrity of final objects.
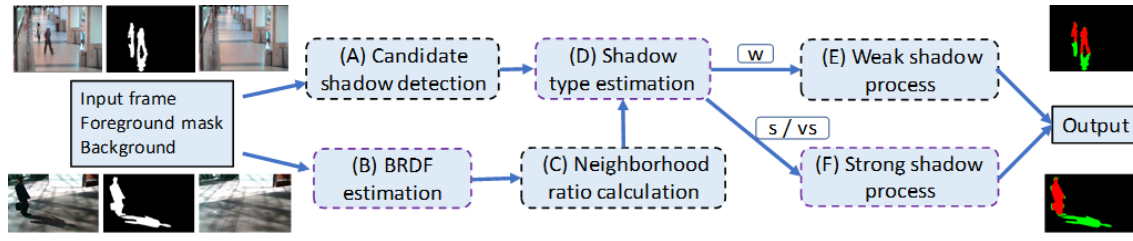
Fig. 2. The framework of the proposed method. Two samples are given as the instructions of the input and output. The upper is Caviar belonging to indoor scene and the lower is PeopleInShade belonging to outdoor scene. The purple dotted lines (B, D, and F) indicate the main contributions in this paper.
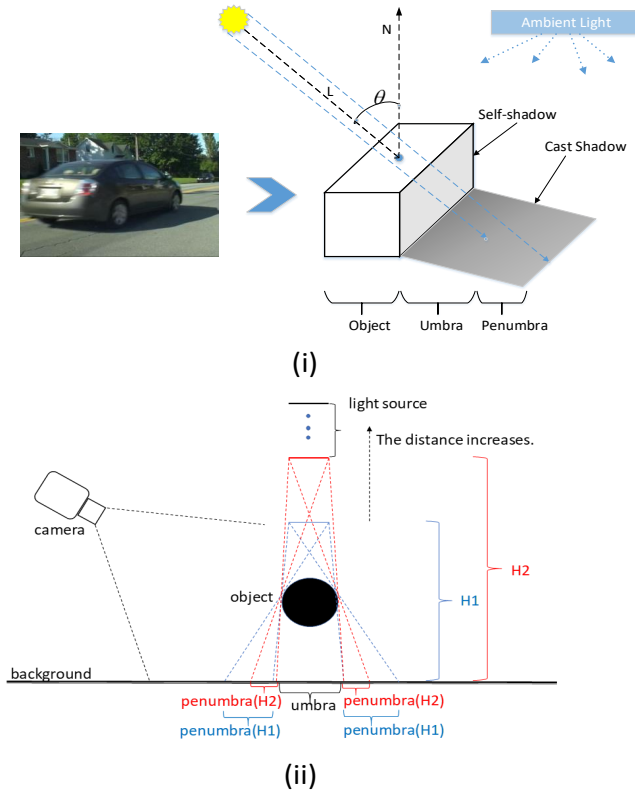

Fig. 3. The illumination changing model of moving cast shadows. (i) shows the cast shadow model, (ii) is the width changing of umbra and the penumbra.

A. Candidate Shadow Detection

Since the irradiance of illumination source is totally or partially occluded by moving objects, the cast shadows are darker than the referenced background image [6], [17]. Thus, the candidate shadows can be detected if pixels are satisfied by:

$$I_B < B_B + \varepsilon \wedge I_G < B_G + \varepsilon \wedge I_R < B_R + \varepsilon \qquad (1)$$

where I_B, I_G, I_R represent the three channel values (BGR) of input frame while B_B, B_G, B_R represent the three channel values of referenced background image. Parameter ε is used to ensure that the result includes all real shadow points.

B. The Proposed BRDF Estimation Methodology

In this section, shadow appearance and basic assumptions will be presented. Cast shadow model is shown in Fig. 3 (i). When an object is placed between the light source and the background, shadows are generated where the light is occluded, consisting of two categories: self-shadow and cast shadow [8]. Self-shadow belongs to object class while cast shadow not. When an object is moving, its cast shadow is called moving shadow. Thus, the moving shadows are darker than corresponding background and always associated with the objects that cast them [6], [17]. Physically, cast shadow is composed of umbra and penumbra. Umbra locates inner of shadow regions while the penumbra locates outer of shadow regions, i.e., the boundary of cast shadows. The distance between illumination source and object has significant effect on the width of penumbra as shown in Fig. 3 (ii). As the distance increases, the width of penumbra becomes narrower, which results in penumbra smaller. Hence, moving cast shadow detection can be transferred to the identification of umbra and penumbra.

In this paper, simplifying assumptions for the most generalized scenes are added as follows:

1)White illumination, the illumination spectrum is constant for each wavelength based on Kubelka-Munk theory [52].

2)The background is static.

3)The background surface is planar and diffuse reflection.

With the assumption of a static background, the reflectance of background does not change with time due to the material dependence. When the surface is occluded by moving objects, the reflectance will be different significantly, but when surface is projected by shadows, the surface reflectance will keep unchanged. This is called the illumination invariant of surface reflectance, which can be presented in terms of BRDF:

$$f\left(\theta_i, \phi_i, \theta_r, \phi_r\right) = \frac{L^{surface}\left(\theta_r, \phi_r\right)}{E^{surface}\left(\theta_i, \phi_i\right)} \qquad (2)$$

where $E^{surface}\left(\theta_i, \phi_i\right)$ is the irradiance in direction (θ_i, ϕ_i) and $L^{surface}\left(\theta_r, \phi_r\right)$ is the radiance in direction (θ_r, ϕ_r). BRDF has two important properties: rotational symmetry and Helmholtz Reciprocity [14]. For rotational symmetry, appearance does not change when surface is rotated about the normal; for Helmholtz Reciprocity, appearance does not change when illumination source and viewing directions are swapped.

Physically, surface reflection includes diffuse reflection and specular reflection. For images captured from surveillance cameras, due to the rough surfaces, only the diffuse reflection is taken into account in general and the specular reflection is simply disregarded as noise in most scenes. Based on Lambert's Law, the Lambertian BRDF is simply a constant.

In this section, we propose an approximation methodology for BRDF estimation. For simplification, we denote $R_k(x,y)$ as the surface reflection at the position (x,y) with k^{th} channel:

$$R_k(x,y) = \frac{L_k(x,y)}{E_k(x,y)} \in [0,1] \qquad (3)$$

where $L_k(x,y)$ and $E_k(x,y)$ are the radiance and irradiance at the position (x,y) with k^{th} channel, respectively. The intensity $L_k(x,y)$ for each pixel can be obtained from camera while the irradiance $E_k(x,y)$ is unknown in advance. According to Eq.3, it is obvious that the irradiance value must be greater than or equal to the pixel's luminance. Thus, a BRDF estimation through the pixel's neighborhood is explored as follows.

$$L_{max}(x,y) = max\{L(i,j)\} \in \mathbf{W}$$
$$L_{min}(x,y) = min\{L(i,j)\} \in \mathbf{W} \qquad (4)$$

$$\alpha = (L_{max}(x,y) - L_{min}(x,y)) / (L_{max}(x,y) + 1) \qquad (5)$$

$$E_k(x,y) = L_{max}(x,y) \cdot (1-\alpha) + L_{min}(x,y) \cdot \alpha \qquad (6)$$

The $\mathbf{W}$ represents the neighboring window of pixel (x,y) in RGB color space. In Eq. 5, α is defined as the fusion factor of illumination correction to overcome the underestimate of quick changing area under illumination source. Visual results of an example for illumination invariant image are shown in Fig. 4 (d) and (e) by normalization operation of $R_k(x,y)$. Each channel is operated independently. The inner of shadows (umbra) tends to retain the similar property with the corresponding background, whereas the objects differ from background significantly. The outer part of shadows (penumbra) are the changing regions from background to the inner of shadows.

It should be noted that BRDF estimation is different from edge detection although the image of BRDF estimation holds obvious edge information. The fundamental difference is that BRDF estimation focuses on pixel's reflection change while edge detection like Sobel or Canny detects intensity changes between pixels. Thus, edge detection can not serve the same purpose that BRDF estimation aims for. It has been verified through experiments that the segmentation results of edge detection (Sobel) are 5%-8% less than those of BRDF estimation.

C. Neighborhood Ratio Calculation

Based on the results of BRDF estimation above, they can be utilized to discriminate shadows from moving foreground. The difference between the illumination invariant image and that of background can be used to remove shadows and generate object mask. Each pixel is conducted like:

$$\Omega(x,y) = \begin{cases} 1, & if\ |I_k(x,y) - B_k(x,y)| < \lambda \\ 0, & otherwise \end{cases} \qquad (7)$$

where $I_k(x,y)$ represents the illumination invariant of the input frame at (x,y) position with the kth channel and $B_k(x,y)$ is the illumination invariant of corresponding background point. The λ is the range threshold parameter. However, it is sensitive to isolated pixels. '1' represents shadow class while '0' represents object class. Therefore, a small neighborhood

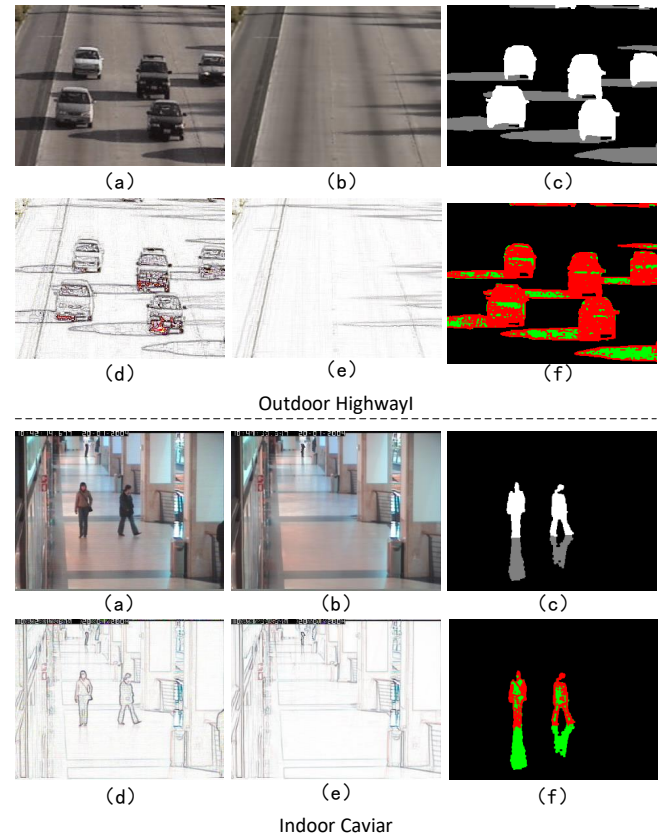

Outdoor HighwayI

Indoor Caviar

Fig. 4. The illumination invariant estimation. (a) input frame, (b) reference background, (c) ground truth, (d) and (e) are the images of illumination invariant estimated from (a) and (b) respectively, (f) is the result of neighborhood ratio calculation, the red represent objects and the green represent shadows.

region (e.g., 3×3 patch) is employed to reduce the influence of noise. The shadow detection result is determined by:

$$S_1 = \begin{cases} 1, & if(\sum_{j=1}^{n} \sum_{k \in (b,g,r)}(\Omega_k^j(x,y))) > \gamma \\ 0, & otherwise \end{cases} \qquad (8)$$

where γ is a threshold. This operation provides rough results as shown in Fig. 4 (f).

Umbra is mainly distributed in the inner of the cast shadows while penumbra in the boundary. It can be seen from Fig. 4 that for indoor scene, the penumbra can be classified correctly because of the effect of multiple light sources. For outdoor scene, the outer of shadows is easily classified as object incorrectly. In addition, the result is also affected by the matter that some object points are misclassified as shadow class, which will be addressed solely in subsequent steps.

D. The Proposed Shadow Type Estimation Strategy

To estimate the shadow type, we exploit three ratios: average intensity difference ratio, area ratio and edge ratio of potential shadows. For the simplification, some variables or expressions are predefined as follows:

I as the input frame; B as the background image; FM as the provided foreground mask; CS and CO are the candidate shadows and candidate objects from candidate shadow detection step, respectively; PS as the result of FM subtracted by CO, i.e., the potential shadow points.

Estimation of shadow type is based on three ratios:

1) Average intensity difference ratio of potential shadows:

$$\psi_{SDR} = \frac{\sum_{j=1}^{n} \sum_{k \in (b,g,r)} \left(B_k^j(x,y) - I_k^j(x,y) \right)}{3 \times n \times 255} \quad (9)$$

The n represents the number of pixels belonging to PS mask. The ψ_{SDR} is regarded as a component to reflect the dark extent of potential shadows. Its range is from zero to one. When it tends to be larger, the potential shadows are darker, which means the shadow strength is more likely to be strong. On the contrary, when it is smaller, the shadow strength is more likely to be more weak.

2) Area ratio of potential shadows:

$$\psi_{SAR} = \frac{N_{PS}}{N_{FM}} \quad (10)$$

The N_{PS} and N_{FM} represent the pixel number of potential shadow points in PS and foreground pixel points in FM, respectively. The range of ψ_{SAR} is from zero to one. When it is near to zero, FM belongs to object class overall. When it tends to one, FM is more likely to be shadow totally or dark object. Therefore, ψ_{SAR} may be used to determine whether the shadows exist or not.

3) Edge ratio of potential shadows: The edge image of potential shadows (named PSE) is acquired by $PSE = CS \cap CO \cap FME$, where FME is the edge of FM. The edge ratio is calculated by:

$$\psi_{SER} = \frac{N_{PSE}}{N_{FME}} \quad (11)$$

where the N_{PSE} and N_{FME} represent the number of pixels in PSE mask and FME mask, respectively. The range of ψ_{SER} is from zero to one. Based on the illumination invariant and neighborhood ratio calculation, the umbra is almost detected correctly but the detection of penumbra fails for strong shadow scenes like Fig. 4 HighwayI (f). The penumbra of weak shadows is detected correctly in Fig. 4 Caviar (f). Thus, when ψ_{SER} increases or is near to one, shadow type is more likely to be strong or very strong, whereas when it tends to be zero, the shadow type is more likely to be weak (see Table I).

TABLE I
THE AVERAGE (Ψ_{SER}) OF VIDEO SCENES AND THE TENDENCY.

Scene	Ψ_{SER}	Tend to be	Scene	Ψ_{SER}	Tend to be
Room	0.295	weaker	HighwayIII	0.714	strong
Lab	0.167	weaker	PeopleInShade	0.832	stronger
Campus	0.406	weak	Bungalows	0.534	medium
Caviar	0.385	weak	BusStation	0.592	strong
Hallway	0.256	weaker	Seam	0.670	strong
Cubicle	0.595	strong	Senoon	0.781	stronger
HighwayI	0.829	stronger	Sepm	0.564	medium

Joint estimation rule:

$$ST = \begin{cases} w, & if(\Psi_{SDR} < \tau_l \vee (\Psi_{SDR} < \tau_h \wedge \Psi_{SAR} > \Psi_{SER}))) \\ s, & if(\Psi_{SDR} \geq \tau_h \vee (\Psi_{SDR} \geq \tau_l \wedge \Psi_{SAR} \leq \Psi_{SER}))) \quad (6) \\ vs, & if(\Psi_{SER} > \kappa) \end{cases}$$

The symbols τ_l, τ_h and κ are three parameters for estimation of shadow type. In Eq. 6, w represents weak, s represents strong and vs represents very strong. The detailed analysis of these three parameters will be presented in Section IV.

E. Weak Shadow Removal

To detect weak shadows, the fusion strategy of chromacity, gradient, texture and other features tends to achieve good performance. Thus, we choose this strategy.

1) Shadow Boundary Removal by Pixel-level Chromacity: This part depends on Candidate Objects which includes real moving objects and might include the boundary of shadows. The pixels distributed in or near the penumbra are darker than the corresponding background but chromacity changes little. HSV color space fits the perception of peoples eyes with good separation of intensity and chromacity. HSV color space-based method achieves comparative results in the field of shadow detection with less computation. Therefore, chromacity-based method [17] is chosen for the detection and removal of penumbra from CO.

2) Shadow Removal by Region-level Gradient Information: This part is based on the result of candidate shadows in Section III-A. Connected components are extracted from candidate shadow mask, i.e., each connected component is a candidate shadow region that will be classified as shadow or object [12]. Then for each pixel in a blob, its gradient magnitude $GM(x,y)$ and gradient direction $GD(x,y)$ are computed. The difference of gradient direction between input frame and background image is represented by Δ_{GD}. To avoid noise effect, only those pixels with $GM(x,y)$ greater than certain threshold are taken into consideration for the computation of Δ_{GD}. Then, the direction correlation is estimated by $H(\tau_{GD} - \triangle_{GD}) = 1$ if $\Delta_{GD} < \tau_{GD}$.

For a connected region, it is considered as a shadow blob only if at least the proportion of the number of $H(\cdot) = 1$ is greater than τ_c. Otherwise, it will be regarded as object class.

3) Shadow Removal by Pixel-level Improved Local Ternary Pattern: Improved local ternary pattern, as an effective texture descriptor, was proposed by the previous work [22]. We take advantage of local texture at patch-level to discriminate shadows from objects. For each frame, threshold parameters is estimated by previous segmented result.

After the classification by three feature detectors above, for each pixel belonging to CS, it will be classified as shadow category only if at least two component results are classified as shadow category. Then the detected shadows may be removed from foreground mask, and three object masks are generated and used for voting process: the best of three sets.

F. The Proposed Direction Features for Strong Shadow Removal

Strong shadows tend to have significant discrimination between shadow regions and corresponding background regions [24]. It is observed from the Fig. 4(f) that the boundary of shadows is distributed in the illumination changing areas from background to shadow, which are weaker than the inner parts of strong shadows. What needs to be pointed out is that
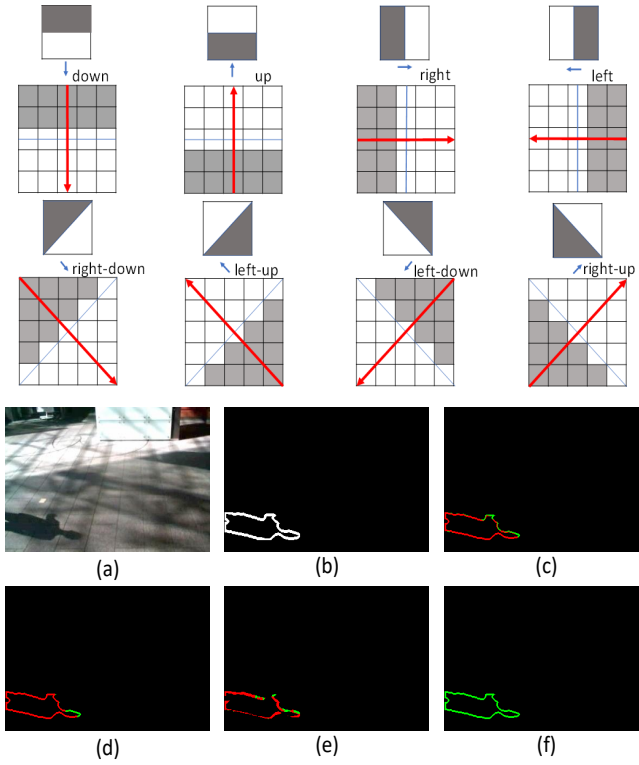
Fig. 5. The proposed direction features with eight categories and an example of very strong shadow scene. (a) is the 434th frame from PeopleInShadow, (b) is the edge of shadows, (c) is the result of chromacity-based method [17], (d) is the result of texture-based method [12], (e) is the result of multiple features method [22], and (f) is the result of the proposed method, where the green stands for shadow class and the red stands for object class.

the smaller the size of connected region is, the larger the percentage of the boundary accounting for whole connected region is. It may be configured according to the contour size of connected region ($CSCR$). If $CSCR$ is more than half of the sum of one frame's height and width, the boundary width is set to three pixels. If it is less than one-fourth of the sum, the boundary width is set to one pixel. For the rest of situations, the boundary width is set to two pixels.

For each connected region, the boundaries consist of object points and shadow points. Shadow points belong to penumbra and keep the color information that is similar with the reference background, whereas object points are not. Therefore, chromacity-based method [17] can be used to detect the boundaries of shadows and remove them.

For very strong shadows, it seems difficult for chromacity-based methods to cope with the shadow boundaries. This may be resolved by physical property of penumbra. For outdoor scenes, the irradiance $E(x,y)$ can be expressed [6] by

$$E(x,y) = C_a + T(x,y) \cdot C_s \cdot cos(\omega) \qquad (7)$$

where C_a and C_s represent the intensity of light source (sunlight) and the ambient light, respectively. The ω represents the incident angle of dominant light source. The $0 \leq T(x,y) \leq 1$ stands for the percentage of receiving the dominant illumination source (sunlight). Let p and $p + \triangle_p$ be two points

distributed at the inner and outer of the penumbra. We denote $\Lambda(p) = E^I(p)/E^B(p)$, then

$$\Delta(p, p + \triangle_p) = \frac{E^I(p)}{E^B(p)} - \frac{E^I(p + \triangle_p)}{E^B(p + \triangle_p)} \qquad (8)$$

where $E^I(\cdot)$ and $E^B(\cdot)$ are the irradiance values of input frame and background image, respectively. When $C_s \gg C_a$, Eq. 8 can be expressed as:

$$\Delta(p, p + \triangle_p) \approx \frac{T^I(p)}{T^B(p)} - \frac{T^I(p + \triangle_p)}{T^B(p + \triangle_p)} \qquad (9)$$

For outside scenes, the neighboring pixels share the very similar energy receiving percentage, i.e., $T^B(p) \approx T^B(p + \triangle_p)$. In this regard, Eq. 9 can be expressed by:

$$\Delta(p, p + \triangle_p) \approx \frac{1}{T^B(p)}(T^I(p) - T^I(p + \triangle_p)) \qquad (10)$$

It is observed that the inner of penumbra is darker than the outer of penumbra, i.e., $T(p) < T(p + \triangle_p)$. So, $\Delta(p, p + \triangle_p) < 0$.

$$\Lambda(p) < \Lambda(p + \triangle_p) \qquad (11)$$

The $T(\cdot)$ is not easy to obtain. Actually, the original images can be exploited in a proper way. Based on Eq. 3, the $E(p) = L(p)/R(p)$. BRDF $R(p)$ is constant for umbra and penumbra in Section III-B, i.e., $R^I(p) \approx R^B(p)$. Thus,

$$\begin{aligned} \Delta(p, p + \triangle_p) &= \frac{L^I(p)/R^I(p)}{L^B(p)/R^B(p)} - \frac{L^I(p + \triangle_p)/R^I(p + \triangle_p)}{L^B(p + \triangle_p)/R^B(p + \triangle_p)} \\ &\approx \frac{L^I(p)}{L^B(p)} - \frac{L^I(p + \triangle_p)}{L^B(p + \triangle_p)} < 0 \end{aligned} \qquad (12)$$

Given that five points p_1, p_2, p_3, p_4, p_5, they belong to penumbra in the direction of shadow from inner to outer. According to Eq. 12, we can conclude the relation $\Lambda(p_1) < \Lambda(p_2) < \Lambda(p_3) < \Lambda(p_4) < \Lambda(p_5)$. This property is invalid for the boundary of objects. Thus, it can be utilized to discriminate shadows from objects under very strong shadow scenes. The determination of direction is important. Eight direction features with $0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°$ are proposed to construct a suite of direction features shown in Fig. 5. The gray represents the possible shadow region while the white represents the background. The main advantage of direction features is that it evaluates the directions fast. To improve the classification accuracy, in our implementation the shadow category is determined by $\Lambda(p_1) < \Lambda(p_3) \wedge \Lambda(p_2) < \Lambda(p_4) \wedge \Lambda(p_3) < \Lambda(p_5)$.

In addition, the post-processing process is also of importance to maintain the structures of real objects. In this paper, it is implemented by polygonal fitting and convex hull operations. The final visual results are given in Section IV.

IV. EXPERIMENTAL RESULTS

Qualitative and quantitative experiments are performed on various sequences that are from [11]–[13]. All the methods including Chr [17], Geo [18] , Phy [19], SR [7], LR [12], MFF [22], ORI [10] and ours are conducted on PC 3.5GHz Xeon machine with C++ implementation. Each method utilizes one suite of parameters provided by its original publication.

TABLE II
THE INFORMATION OF VIDEO SEQUENCES FOR EXPERIMENTS.

	Size	Length	Labeled frames	Scene type	Shadow stength	Objects	Object size
IntelligentRoom	320×240	750	22	Indoor	Weak	People	Medium
Laboratory	320×240	2217	14	Indoor	Weak	People	Medium
Campus	352×288	2947	53	Outdoor	Weak	People/Vehicle	Medium/Small
Caviar	384×288	1113	1113	Indoor	Weak	People	Variable
Hallway	320×240	800	13	Indoor	Weak	People	Large/Medium
Cubicle	352×240	7400	3210	Indoor	Weak/Strong	People	Large/Medium
HighwayI	320×240	733	8	Outdoor	Strong	Vehicle	Large
HighwayIII	320×240	2227	7	Outdoor	Strong	Vehicle	Small
PeopleInShade	380×244	1199	666	Outdoor	Very Strong	People	Large
Bungalows	360×240	1700	636	Outdoor	Strong	Vehicle	Large
BusStation	360×240	1250	873	Outdoor	Strong	People	Medium
Seam	320×240	460	151	Outdoor	Strong	People/Vehicle	Medium/Small
Senoon	320×240	1501	216	Outdoor	Very Strong	People/Vehicle	Medium/Small
Sepm	320×240	251	37	Outdoor	Strong	People/Vehicle	Medium/Small

A. Datasets and Evaluation Metrics

To validate the effectiveness of the proposed method, four-teen video sequences consisting of five indoor scenes and nine outdoor scenes are selected in TABLE II. The scenes are all real-world, and the dataset is standard and used by previous researchers. The authors of [11]–[13] make a good contribution to provide ground-truths that are widely used to test the methods. The shadow type of the dataset ranges from weak to very strong. Single light source and multiple light sources are covered. Foreground-background camouflage and foreground-shadow camouflage appear. The size and type of moving objects are also diverse. Compared with weak shadows under indoor conditions, the (very) strong shadows under outdoor conditions are more challenging, which accounts for a large proportion. Hence, a variety of scenes cover the problematic issues of moving shadows as much as possible.

The most widely used evaluation metrics are proposed by [6]: shadow detection rate (η) and discrimination rate (ξ).

$$\eta = \frac{TP_s}{TP_s + FN_s} \qquad \xi = \frac{TP_o}{TP_o + FN_o} \qquad (13)$$

where TP_s represents the number of shadow points that are classified correctly, FN_s is the number of shadow points that are classified as object, TP_o is the number of object points that are classified correctly and FN_o is the number of object points that are classified as shadow.

Shadow detection rate is a metric which measures one method's performance of identifying shadow points from all real shadow points successfully. Shadow discrimination rate is a metric which measures one method's performance of identi-fying object points from all real object points successfully. One algorithm will be regarded as good if the two metrics are both high. A comprehensive metric called $F - measure = \frac{2*\eta*\xi}{\eta+\xi}$ [22] is suitable and reliable for evaluation of various methods.

B. Parameters Analysis

In this paper, we propose to exploit illumination invariant of surface reflection for initial rough classification. Neighborhood

ratio calculation is the fundamental and core step for the whole framework. Its parameters should be analyzed in detail: the range threshold parameter λ and determination parameter of neighborhood ratio γ. In addition, shadow type estimation is essential and its parameters should be detailed.

1) λ : It controls the pixel-level shadow/object classification from the perspective of intensity difference. When λ becomes larger, more object points might be classified as shadow, which results in low shadow discrimination rate. When λ becomes smaller, those shadow points that have larger difference values might be classified as object, which reduces the shadow detection rate. A trade-off should be adjusted by a proper value or dynamic range. Figure 6 shows the changing tendency of average values of three metrics with λ varying from 1 to 49 for 14 scenes. With the increase of λ value, the shadow detection rate keeps an increasing tendency until 40 while the shadow discrimination rate reduces all the time until 45. The F-measure firstly increases to highest level and then decreases. A fixed value $\lambda = 23$ could provide a potential good shadow detection and discrimination performance for all scenes. For indoor scenes, λ is advised to select from 15 to 24 and for outdoor scenes it is advised to select from 20 to 27.

Therefore, based on the analysis above we select $\tau_l = 0.2, \tau_h = 0.5$ and $\kappa = 0.9$ to judge shadow types. The fixed parameters may provide satisfactory results for the 14 video sequences. The parameter ε for candidate shadow detection is set to 7 for all scenes. Its range is recommended from 4 to 10. Other parameters for weak shadow removal are referred to [12], [22] which had been optimized by the original authors.

2) γ: It is the decision threshold of neighboring ratio, which decides whether the pixel should be classified as object or shadow through a patch of neighborhood pixels. A larger γ value implies a strict limitation for shadow classification, which leads to the reduced shadow detection rate. Smaller value of γ may increase the shadow detection rate but cause the misclassification of object pixels as shadow. Fig. 7 shows the changing tendency of average values of three evaluation metrics for all the video sequences. When γ is small enough
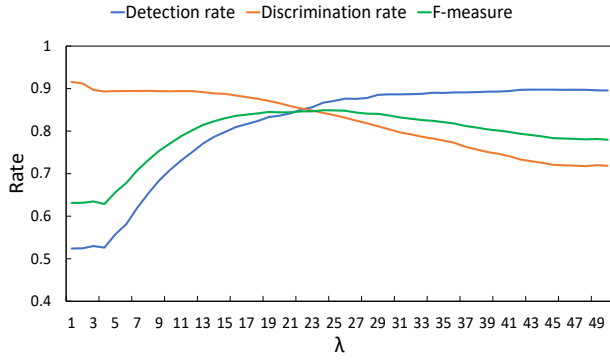
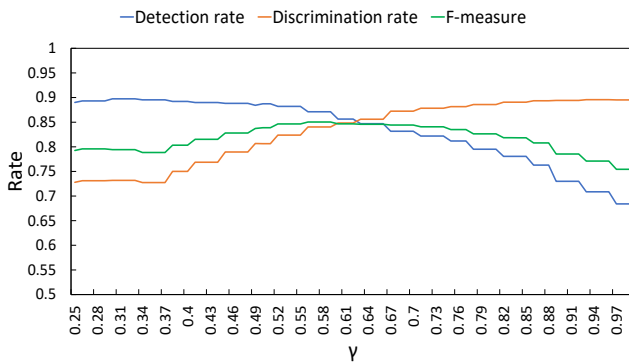Fig. 6. Shadow detection rate, shadow discrimination rate and F-measure changing with the varying range threshold parameter (λ).

Fig. 7. Shadow detection rate, shadow discrimination rate and F-measure changing with the varying range of neighborhood ratio parameter (γ).

(e.g., $\leqslant 0.34$), the shadow detection rate and discrimination rate both retain almost unchanged. With the increase of γ value, shadow detection rate goes up while shadow discrimination rate goes down. The two metrics crosses at 0.6 more or less. The F-measure fluctuates little between 0.56 and 0.65. A fixed value 0.58 is selected to provide satisfactory results for all scenes.

3) τ_l, τ_h and κ: three parameters are used to estimate shadow type. Their values stem from the varying characteristics of all the shadow scenes. Fig. 8 shows typical weak (Caviar) and strong (BusStation) scenes. The ψ_{SDR} measures how the shadows are darker than the corresponding background.

For weak shadows, the boundary of shadows can be fuzzy. The ψ_{SDR} of weak shadow tends to have a smaller value while that of strong shadow tends to have a higher value. For the very strong shadow scene of PeopleInShadow, ψ_{SDR} is near to one in many frames. Therefore, ψ_{SDR} values may distinguish the very strong shadow scenes. What makes estimation difficult is that there is not a clear demarcation between weak and strong types. The boundary of weak shadows can be identified easily by illumination invariant of surface reflection, which leads ψ_{SER} values smaller. The ψ_{SAR} should be higher than ψ_{SER}, e.g., Fig. 8. Caviar. For strong shadows the edges of shadows are not easy to detect. The ψ_{SER} values tend to be larger than ψ_{SAR}, e.g., Fig. 8 BusStation, and Fig. 9. It should be noted that there are extreme situations like Fig. 9 Bungalows. The

ψ_{SAR} values are very high in multiple frames, which can be explained when all the foreground regions are shadows.

C. Experimental Results and Discussion

To evaluate the performance of comparative methods, all the video sequences are divided into types according to shadow strength. The weak shadow class consists of six scenes: IntelligentRoom, Laboratory, Campus, Caviar, Hallway and Cubicle. The rest of video scenes belong to strong shadow class. Quantitative results are presented in TABLE III and visual results are given in Fig. 10 and Fig. 11. It can be concluded from TABLE III that the proposed method outperforms the comparative approaches on the average shadow detection rate. Meanwhile, our method has a competitive shadow discrimination rate. For the average F-measure, our method achieves higher value than those of compared approaches.

The methods proposed by SR [7] and Phy [19], providing the lowest average values of shadow detection, it seems that these methods are unable to handle diverse video sequences. The algorithms provided by Chr [17] and Geo [18] have the lowest average shadow discrimination results, i.e., they are more likely to classify objects as shadow category incorrectly. The LR [12] approach reaches a higher discrimination rate than other approaches, however, its shadow detection rate is low. For Caviar scene, the approach MFF [22] shows better performance than the proposed method. For Laboratory scene, our method achieves similar performance with that of MFF. The average shadow discrimination rate of MFF is 2.6% higher than that of our method, but in terms of average shadow detection rate ours is 21.6% higher than that of MFF. The method ORI [10] has advantages over others on the indoor scenes like Laboratory, Campus, Caviar, and Hallway.

It seems that the comparative methods have difficulty in tackling strong shadows, which is illustrated by TABLE III. For instance, a typical strong shadow scene: Bungalows, the shadow detection rates of Chr, Phy, SR, and MFF are all lower than 20%. The proposed method achieves 77% detection rate. For the very strong scene: PeopleInShadow, Geo method achieves a higher shadow detection rate (57.34%) among the compared methods. Although the scene meets Geo's assumptions basically, it is sill 26.72% lower than that of our method (84.06%). Although ORI achieves good performance on weak shadow scenes, but it fluctuates much on some strong shadow scenes like Seam, Senoon, and Sepm.

The segmented foreground masks with color labeled are shown in Fig. 10 and Fig. 11 with presenting one frame per video sequences. As seen from the figures, the results of Chr, Phy, SR are not stable. The method Chr can obtain good performance on the scenes with sufficient chromacity (Cubicle), however, it fails when the objects are darker (Caviar and BusStation) and objects have similar colors to the background (Bungalows). The method Geo shows good performance when the orientations of objects and shadows are different (PeopleInShadow and HighwayI) but might fail when the shadows are distributed separately under multiple light sources (IntelligentRoom and Laboratory). The approach presented in Phy may present good results (IntelligentRoom
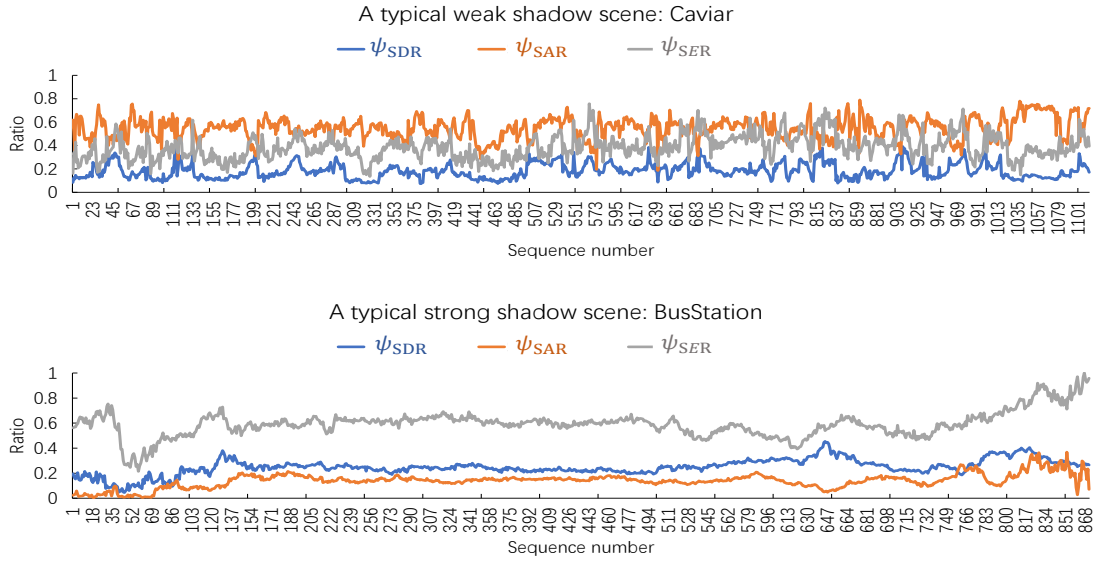
Fig. 8. The analysis of three ratios for shadow type estimation. (a) A typical soft shadow scene: Caviar. (b) A typical strong shadow scene: BusStation.
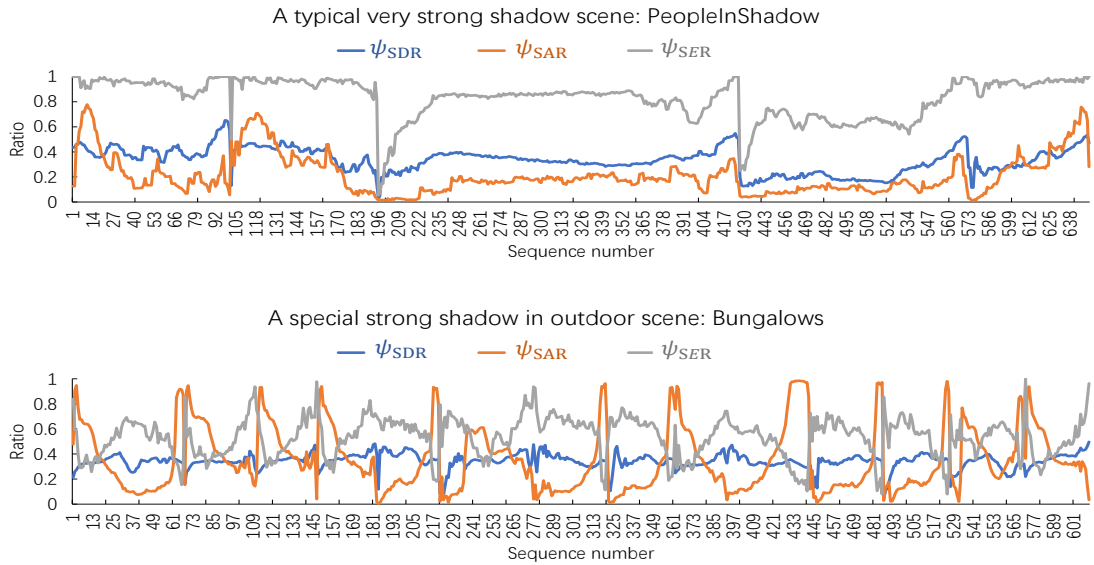

Fig. 9. The analysis of three ratios for shadow type estimation. (a) A typical very strong shadow scene: PeopleInShade. (b) A spercial strong shadow scene: Bungalows.

and Seam) when provided with enough examples to learn and adapt to the changing of moving shadows. However, it is sensitive to the pixel-level noise (Hallway and Caviar). The method proposed by SR is robust to various illumination conditions (Cubicle), but it requires that the background is textured. It is hard to deal with non-texture surface of foreground or background (HighwayI). The method LR achieves good results for indoor scenes (Caviar and Hallway). It is also robust to pixel-level noise. However, it depends on region segmentation algorithms too much and may fail when objects are similar with shadows or the edges of objects and shadows

are fuzzy (HighwayI and Seam).

The approaches explored by MFF and ORI exploit multiple features to obtain good results for weak shadows (Fig. 10). However, they face challenges in adapting to the strong shadows because of the insufficient chromacity and texture information (e.g., Fig. 11 PeopleInShadow). Visual results on fourteen sequences validate that the proposed method can adapt to various illumination conditions.

Considering average running time per frame, the compared approaches and the proposed method are operated on all the sequences. They are Chr (6.3ms), Geo (2.4ms), Phy (5.1ms),
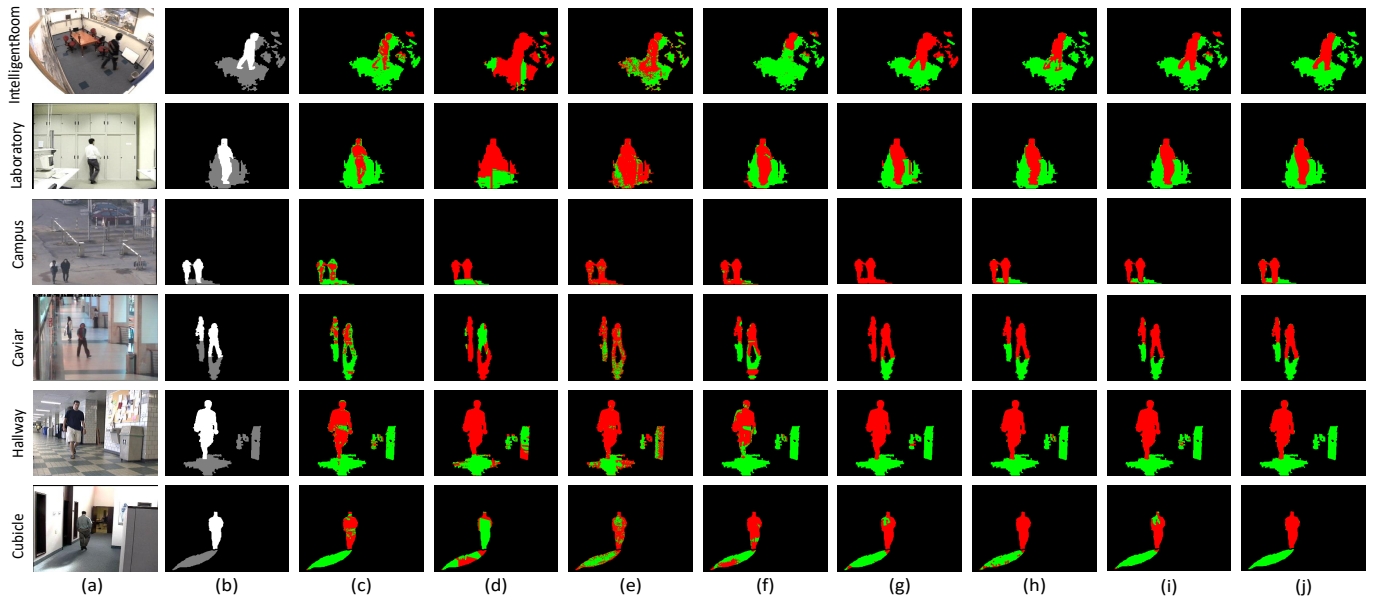
Fig. 10. Visual results of six weak shadow scenes. The red represent objects and the green represent shadows. (a) input frames; (b) ground-truth masks; (c) segmentation masks generated by Chr [17]; (d) segmentation masks generated by Geo [18]; (e) segmentation masks generated by Phy [19]; (f) segmentation masks generated by SR [7]; (g) segmentation masks generated by LR [12]; (h) segmentation masks generated by MFF [22]; (i) segmentation masks generated by ORI [10]; (j) segmentation masks generated by our method.

Fig. 11. Visual results of eight (very) strong shadow scenes. The red represent objects and the green represent shadows. (a) input frames; (b) ground-truth masks; (c) segmentation masks generated by Chr [17]; (d) segmentation masks generated by Geo [18]; (e) segmentation masks generated by Phy [19]; (f) segmentation masks generated by SR [7]; (g) segmentation masks generated by LR [12]; (h) segmentation masks generated by MFF [22]; (i) segmentation masks generated by ORI [10]; (j) segmentation masks generated by our method.
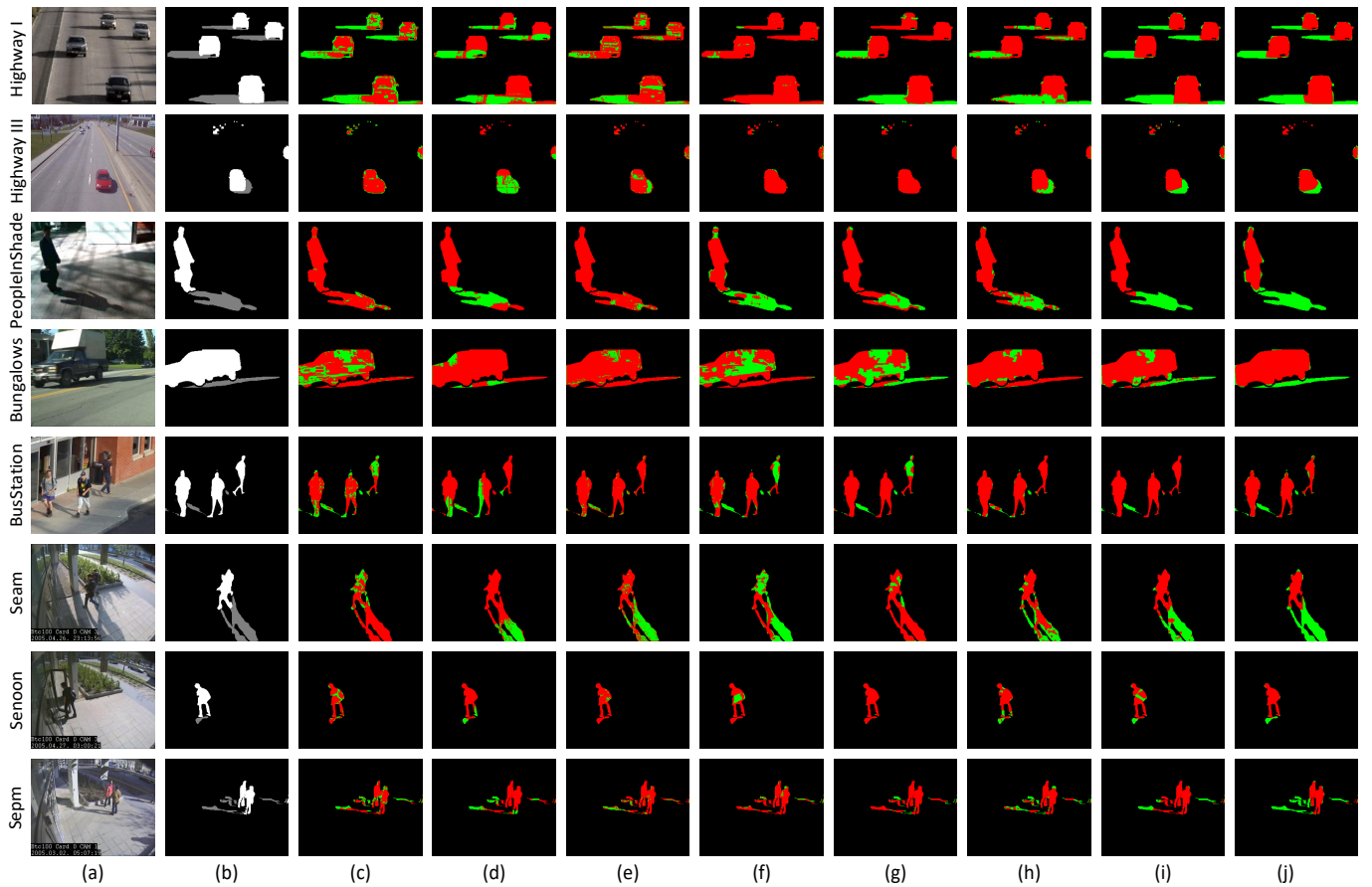
TABLE III
THE SHADOW DETECTION RATE AND SHADOW DISCRIMINATION RATE COMPARISONS BETWEEN THE PROPOSED METHOD AND THE STAT-OF-THE-ART APPROACHES ON FOURTEEN VIDEO SEQUENCES.

|	%	Chr [17]	Geo [18]	Phy [19]	SR [7]	LR [12]	MFF [22]	ORI [10]	Ours
IntelligentRoom	η	96.63	54.77	58.03	93.82	81.80	94.00	91.13	95.04
	ξ	65.65	69.50	93.43	62.84	96.62	92.16	96.92	94.22
	$F-m$	78.18	61.26	71.59	75.27	88.59	93.07	93.94	94.63
Laboratory	η	93.65	47.76	38.45	81.51	64.76	85.83	87.46	84.84
	ξ	76.70	69.20	93.31	88.13	98.19	94.96	95.28	96.51
	$F-m$	84.33	56.51	54.46	84.69	78.05	90.16	91.20	90.30
Campus	η	53.86	60.86	45.90	55.37	50.39	78.16	77.64	86.53
	ξ	52.61	57.03	69.37	86.05	99.11	80.11	91.83	75.11
	$F-m$	53.22	58.88	55.24	67.39	66.81	79.12	84.14	80.42
Caviar	η	95.32	46.59	40.69	66.87	73.23	92.55	94.91	95.71
	ξ	56.33	70.73	86.68	85.35	98.83	84.88	94.55	77.84
	$F-m$	70.82	56.17	55.38	74.99	84.13	88.55	94.73	85.86
Hallway	η	93.56	48.67	66.05	96.09	95.08	95.20	94.85	96.66
	ξ	80.18	75.56	89.17	65.53	96.13	97.97	98.79	93.85
	$F-m$	86.35	59.20	75.89	77.92	95.60	96.57	96.78	95.24
Cubicle	η	92.37	59.27	67.11	26.95	89.40	69.64	89.81	98.44
	ξ	91.56	68.66	91.10	89.64	94.66	84.39	87.19	90.13
	$F-m$	91.96	63.62	77.28	41.44	91.96	76.31	88.48	94.10
HighwayI	η	76.01	66.17	42.47	16.93	60.46	65.69	84.62	78.13
	ξ	65.49	75.56	85.69	94.00	95.48	91.88	93.63	92.72
	$F-m$	70.36	70.55	56.79	28.69	74.04	76.61	88.90	84.80
HighwayIII	η	45.08	42.73	36.28	5.77	38.08	52.24	67.75	66.01
	ξ	65.49	75.63	74.47	90.35	90.20	88.34	93.49	71.58
	$F-m$	53.40	54.61	48.79	10.85	53.56	65.66	78.57	68.68
PeopleInShade	η	9.83	57.34	8.96	51.88	32.99	35.56	75.90	84.06
	ξ	97.10	52.50	97.31	79.64	97.37	90.32	94.40	93.08
	$F-m$	17.86	54.81	16.42	62.83	49.29	51.02	84.15	88.34
Bungalows	η	3.81	59.91	15.79	8.13	70.23	3.57	85.02	77.30
	ξ	74.75	55.43	89.79	78.26	64.16	94.57	78.08	85.94
	$F-m$	7.25	57.58	26.85	14.73	67.06	6.87	81.40	81.39
BusStation	η	50.72	20.52	60.15	51.62	36.00	63.49	65.18	87.64
	ξ	83.05	64.45	94.30	84.95	94.33	89.72	95.00	89.09
	$F-m$	62.98	31.13	73.45	64.22	52.11	74.36	77.31	88.36
Seam	η	19.22	57.39	56.04	69.06	22.66	73.44	79.10	97.91
	ξ	59.92	64.61	87.11	67.75	79.08	77.65	85.70	69.52
	$F-m$	29.11	60.78	68.20	68.40	35.23	75.49	82.27	81.30
Senoon	η	18.45	53.42	64.92	3.22	4.76	48.76	52.69	95.71
	ξ	59.15	58.41	81.72	86.28	98.22	60.40	68.54	70.48
	$F-m$	28.12	55.80	72.35	6.20	9.08	53.96	59.58	81.18
Sepm	η	46.66	60.33	35.47	35.50	18.77	58.38	44.95	75.70
	ξ	66.91	63.13	86.56	79.74	98.10	85.62	83.73	76.29
	$F-m$	54.98	61.70	50.32	49.13	31.51	69.42	58.50	76.00
Average	η	56.80	52.55	45.45	47.34	52.76	65.46	77.93	87.12
	ξ	71.06	65.74	87.14	81.32	92.89	86.64	89.80	84.03
	$F-m$	56.35	57.33	57.36	51.91	62.64	71.23	82.85	**85.04**
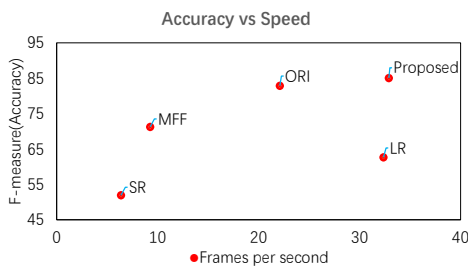
Fig. 12. Accuracy vs Speed Trade-off comparsion of five methods.

SR (157ms), LR (30.9ms), MFF (108ms), ORI (45.25ms) and Ours (30.4ms) per frame. Figure 12 shows that our method can reach a good trade-off between accuracy and speed.

In addition, our method also compares with a deep learning approach [23] which exploits deep convolutional neural network to learn shadow features from input frames and background images. The [23] only provides detection results of five scenes belonging to weak shadows. Figure 13 shows that our average F-measure is 4.2% higher than [23]. However, Lee's approach also requires sufficient labeled training data and its effectiveness on strong shadows needs to be invested.

In real-world applications, the scenes are diverse. To verify our method's effectiveness further, the proposed method is conducted on four typical transport scenes as shown in Fig. 14. These datasets are without manual ground-truth. It's also notable that the masks detected by SubSENSE [3] are not always ideal and easily disturbed by moving shadows. The proposed method can obtain a good shadow segmentation.
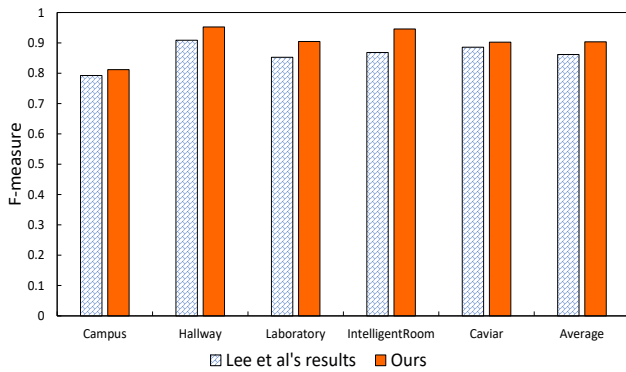
Fig. 13. F-measure comparisons between our method and convolutional deep neural network approach proposed by Lee *et al.* [23].

(i) Original frames

(ii) Detected masks by SubSENSE
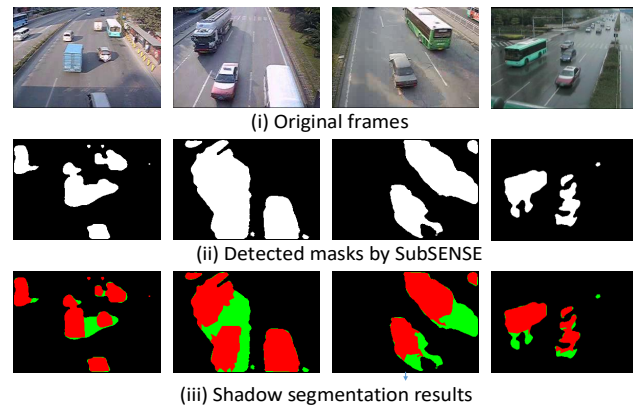
(iii) Shadow segmentation results

Fig. 14. The proposed method are validated on four real-world scenes which are captured from transport surveillance. The foreground is generated by SubSENSE [3]. For the third row, the red and the green represent objects and shadows, respectively.

V. CONCLUSION

In this paper, we develop a framework capable of segmenting moving cast shadows for various scenes. It receives input frame, referenced background image and foreground mask as inputs, removes moving cast shadows from foreground mask, and finally produces a mask with only objects as output.

Several contributions are highlighted. The first is the introduction of illumination invariant feature for moving cast shadows segmentation. Secondly, shadow strength is estimated for each frame using intensity difference ratio, area ratio, and edge ratio of potential shadows. Thirdly, we propose different strategies to handle various shadows. Multiple features fusion strategy is designed to detect weak shadows using color constancy, texture consistency, and illumination invariant. For strong shadows consisting of umbra and penumbra, umbra is detected using illumination invariant feature and penumbra is detected using color constancy. For the very strong shadows, a suite of innovative shadow direction features is proposed to segment penumbra. Experiments conducted on multiple indoor and outdoor sequences shows that the proposed method is expected to have promising results in accuracy and processing speed compared with some state-of-the-art approaches. In the future, further speed optimization needs to be invested and more labeled datasets should be provided for test.

REFERENCES

[1] T. Bouwmans, F. Porikli, B. Hferlin, and A. Vacavan, *Background modeling and foreground detection for video surveillance*. CRC press, 2014.

[2] X. Zhang, C. Zhu, H. Wu, Z. Liu, and Y. Xu, "An imbalance compensation framework for background subtraction," *IEEE Transactions on Multimedia*, vol. 19, no. 11, pp. 2425–2438, 2017.

[3] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, 2015.

[4] H. Lee, H. Kim, and J. I. Kim, "Background subtraction using background sets with image-and color-space reduction," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2093–2103, 2016.

[5] J. Stander, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 65–76, 1999.

[6] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: formulation, algorithms and evaluation," 2003.

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

[8] K. Yang, S. Gao, and Y. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2254–2263.

[9] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Computer Vision Image Understanding*, vol. 95, no. 2, pp. 238–259, 2004.

[10] B. Wang and C. L. Chen, "Optical reflection invariant-based method for moving shadows removal," *Optical Engineering*, vol. 57, no. 9, 2018.

[11] C. Benedek and T. Szirnyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 608–621, 2008.

[12] 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.

[13] Y. Wang, P. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnet 2014: An expanded change detection benchmark dataset," in *Computer Vision Pattern Recognition Workshops*, 2014.

[14] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur, "Color subspaces as photometric invariants," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 13–30, 2008.

[15] N. Al-Najdawi, H. E. Bez, J. Singhai, and E. A. Edirisinghe, "A survey of cast shadow detection algorithms," *Pattern Recognition Letters*, vol. 33, no. 6, pp. 752–764, 2012.

[16] 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.

[17] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 505–516, 2003.

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

[19] J. Huang and C. Chen, "A physical approach to moving cast shadow detection," in *IEEE International Conference on Acoustics*, 2009.

[20] K. Jiang, A. H. Li, Z. G. Cui, T. Wang, and Y. Z. Su, "Adaptive shadow detection using global texture and sampling deduction," *IET Computer Vision*, vol. 7, no. 2, pp. 115–122, 2013.

[21] Z. Zhu, X. You, C. L. Chen, D. Tao, W. Ou, X. Jiang, and J. Zou, "An adaptive hybrid pattern for noise-robust texture analysis," *Pattern Recognition*, vol. 48, no. 8, pp. 2592–2608, 2015.

[22] B. Wang, Y. Yuan, Y. Zhao, and W. Zou, "Adaptive moving shadows detection using local neighboring information," pp. 521–535, 2016.

[23] J. T. Lee, K. Lim, and Y. Chung, "Moving shadow detection from background image and deep learning," in *Pacific-rim Symposium on Image Video Technology*, 2015.

[24] C. Benedek and T. Szirnyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 608–621, 2008.

[25] 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.

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

[27] Q. Wu, W. Zhang, and B. V. Kumar, "Strong shadow removal via patch-based shadow edge detection," in *IEEE International Conference on Robotics Automation*, 2012.

[28] M. Russell, J. Zou, G. Fang, and W. Cai, "Feature-based image patch classification for moving shadow detection," *IEEE Transactions on Circuits Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[29] W. Zhang, X. Z. Fang, X. K. Yang, and Q. J. Wu, "Moving cast shadows detection using ratio edge," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1202–1214, 2007.

[30] J. Wang, Y. Wang, M. Jiang, X. Yan, and M. Song, "Moving cast shadow detection using online sub-scene shadow modeling and object inner-edges analysis," *Journal of Visual Communication Image Representation*, vol. 25, no. 5, pp. 978–993, 2014.

[31] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," in *IEEE Conference on Computer Vision Pattern Recognition*, 2005.

[32] B. N. Subudhi, S. Ghosh, S. Cho, and A. Ghosh, "Integration of fuzzy markov random field and local information for separation of moving objects and shadows," *Information Sciences*, vol. 331, no. C, pp. 15–31, 2016.

[33] Z. Liu, K. Huang, and T. Tan, "Cast shadow removal in a hierarchical manner using mrf," *IEEE Transactions on Circuits Systems for Video Technology*, vol. 22, no. 1, pp. 56–66, 2012.

[34] V. Nguyen, T. F. Y. Vicente, M. Zhao, M. Hoai, and D. Samaras, "Shadow detection with conditional generative adversarial networks," in *IEEE International Conference on Computer Vision*, 2017.

[35] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, "Automatic shadow detection and removal from a single image," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 38, no. 3, pp. 431–446, 2016.

[36] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau, "Deshadownet: A multi-context embedding deep network for shadow removal," in *IEEE Conference on Computer Vision Pattern Recognition*, 2017.

[37] L. Liu, C. L. Chen, Y. Zhou, and X. You, "A new weighted mean filter with a two-phase detector for removing impulse noise," *Information Sciences*, vol. 315, pp. 1–16, 2015.

[38] I. Huerta, M. B. Holte, T. B. Moeslund, and J. Gonzlezi, "Chromatic shadow detection and tracking for moving foreground segmentation," *Image Vision Computing*, vol. 41, no. C, pp. 42–53, 2015.

[39] Y. Su, A. Li, Y. Cai, G. Feng, and G. Jin, "Moving shadow detection with multifeature joint histogram," *Journal of Electronic Imaging*, vol. 23, no. 5, p. 053015, 2014.

[40] Y. Lin, B. Wang, and Y. Zhao, "Moving shadow detection using fusion of multiple features," pp. 487–494, 2018.

[41] R. Qin, S. Liao, Z. Lei, and S. Z. Li, "Moving cast shadow removal based on local descriptors," in *International Conference on Pattern Recognition*, 2010.

[42] J. Peng, Y. Zhou, and C. L. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4810–4824, 2015.

[43] L. Liu, L. Chen, C. L. Chen, and Y. Tang, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 600–611, 2016.

[44] P. Zhang, X. You, W. Ou, C. L. Chen, and Y. M. Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognition*, vol. 52, no. C, pp. 249–259, 2016.

[45] C. R. Jung, "Efficient background subtraction and shadow removal for monochromatic video sequences," *IEEE Transactions on Multimedia*, vol. 11, no. 3, pp. 571–577, 2009.

[46] S. Feng and C. L. Chen, "Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification," *IEEE Transactions on Cybernetics*, 2018.

[47] M. Khare, R. K. Srivastava, and A. Khare, "Moving shadow detection and removal - a wavelet transform based approach," *IET Computer Vision*, vol. 8, no. 6, pp. 701–717, 2014.

[48] Y. Zhou, J. Peng, and C. L. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*, vol. 8, no. 6, pp. 2351–2360, 2017.

[49] H. Asaidi, A. Aarab, and M. Bellouki, "Shadow elimination and vehicles classification approaches in traffic video surveillance context ," *Journal of Visual Languages Computing*, vol. 25, no. 4, pp. 333–345, 2014.

[50] C. L. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks Learning Systems*, vol. 29, no. 1, pp. 10–24, 2018.

[51] B. Wang, C. Chen, Y. Li, and Y. Zhao, "Hard shadows removal using an approximate illumination invariant," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 1628–1632.

[52] P. Kubelka, "New contributions to the optics of intensely light-scattering materials. part i," *Journal of the Optical Society of America*, vol. 38, no. 5, p. 448, 1948.

Bingshu Wang received the B.S. degree in computer science and technology from Guizhou University, Guiyang, China, in 2013 and M.S. degree in electronic science and technology (Integrated circuit system) from Peking University, Beijing, China, in 2016. Now he is a Ph.D student of the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include digital image processing, intelligent video analysis and machine learning.

Yong Zhao received Ph.D degree in Automatic Control and Applications from Southeast University, Nanjing, China, 1991. After that he joined Zhejiang University, Hangzhou, China, as an assistant researcher. In 1997, he went to Concordia University, Montreal, Canada, as a post-doctoral fellow. He was a senior Audio/Video compression engineer of Honeywell Corporation in May 2000. Since 2004, he became an associate professor of Peking University Shenzhen Graduate School, and he is now the header of the lab of Mobile Video Networking Technologies. He is currently working on computer vision, machine learning, video analytics and video compression with special focus on applications of these new theories and technologies to various industries. His team has developed many innovative products and projects which have been successful in the market.

C.L. Philip Chen (S88M88SM94F07) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988. He is a chair professor of the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China. His current research interests include systems, cybernetics, and computational intelligence.

Dr. Chen was the IEEE SMC Society President from 2012 to 2013 and is now a Vice President of Chinese Association of Automation (CAA). He is a Fellow of IEEE, AAAS, IAPR, CAA, and HKIE, and a member of Academia Europaea. He is the editor-in-chief of the IEEE Transaction on Systems, Man, and Cybernetics: Systems, and an associate editor of several IEEE Transactions. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control (2015-2017), and also a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) of the U.S. for computer engineering, electrical engineering, and software engineering programs. He received 2016 Outstanding Electrical and Computer Engineers award from his alma mater, Purdue University after he graduated from the University of Michigan, Ann Arbor, Michigan, USA.