Computers and Electrical Engineering 70 (2018) 813–825

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Automatic removal of complex shadows from indoor videos using transfer learning and dynamic thresholding[☆]

Xiaohui Yuan[a,b,*], Daniel Li[b], Deepankar Mohapatra[b], Mohamed Elhoseny[c,b]

[a] Faculty of information engineering, China University of Geosciences, Wuhan, China
[b] Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA
[c] Faculty of Computers and information, Mansoura University, Mansoura, Daqahlia, Egypt

ARTICLE INFO

Article history:
Received 11 September 2017
Revised 11 December 2017
Accepted 12 December 2017
Available online 23 December 2017

Keywords:
Shadow removal
Transfer learning
Thresholding
Video analysis
Video surveillance

ABSTRACT

In video-based tracking and recognition applications, shadows are usually mis-classified as foreground or part of it due to its close associative to the objects. Shadows in indoor scenarios are more challenging and usually characterized by multiple light sources that produce complex patterns. In this article, we present a learning-based method for removing shadows. Our method suppresses light shadows with a dynamically computed threshold and removes dark shadows using an online learning strategy that is fine-tuned with the automatically identified examples in the new videos. Our experiments demonstrate that the proposed method adapts to the videos and remove shadows effectively. The average accuracy exceeds 97%. The sensitivity of shadow detection varies slightly with different confidence levels used in example selection for retraining and high confidence usually yields better performance with less retraining iterations. In the evaluation of efficiency, updating kNN imposes little impact on the processing time.

1. Introduction

Background subtraction is a critical step in many computer vision applications ranging from object tracking to action recognition [1,2], which requires accurate foreground objects. However, the foreground object is usually distorted by non-stationary shadows of the moving object. Due to its nature of dynamically emerging with objects, the shadow is usually misclassified as foreground object or part of it. There have been many methods developed to handle shadow removal in a variety of outdoor scenarios, e.g., traffic monitoring [3] and surveillance [4]. However, these methods are facing difficulties in indoor lighting, where multiple light sources combine to produce complex shadows. Research has been conducted for indoor scenarios [5], in which a manually specified threshold is used.

Shadows in indoor scenarios are usually characterized by multiple light sources. An example is shown in Fig. 1(a), which shows that part of the shadow appears brighter than the others. Without removing the shadow, the foreground object tends to be erroneously segmented, as shown in Fig. 1(b); and with shadow removal, the optimal body silhouette contains no shadow component, as shown in Fig. 1(c). The inconsistent hue and intensity of shadows make automatic removal a challenging task; simple color-based methods are ineffective and could cause the shattered object of interest [5].

[☆] Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Gustavo Ramirez Gonzalez.

[*] Corresponding author at: Faculty of information engineering, China University of Geosciences, Wuhan, China.
E-mail address: xiaohui.yuan@unt.edu (X. Yuan).

https://doi.org/10.1016/j.compeleceng.2017.12.026
0045-7906/© 2017 Elsevier Ltd. All rights reserved.

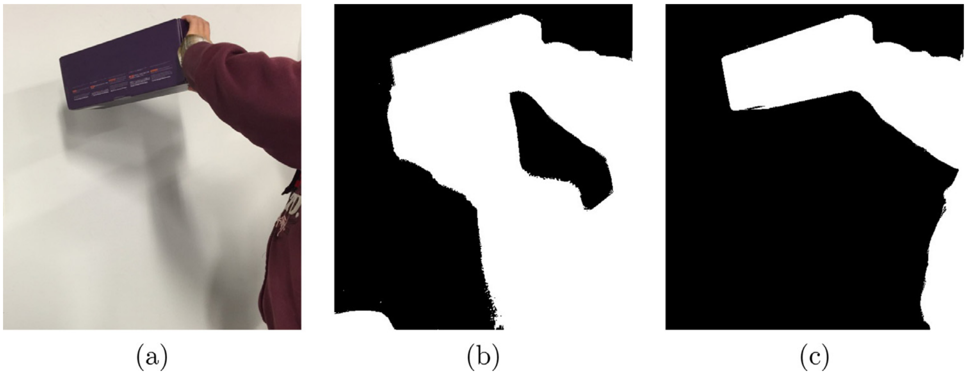


Fig. 1. Complex shadow and the background subtraction results. (a) a frame showing complex shadow of different shades. (b) background subtraction result. (c) background subtraction with shadow removal.

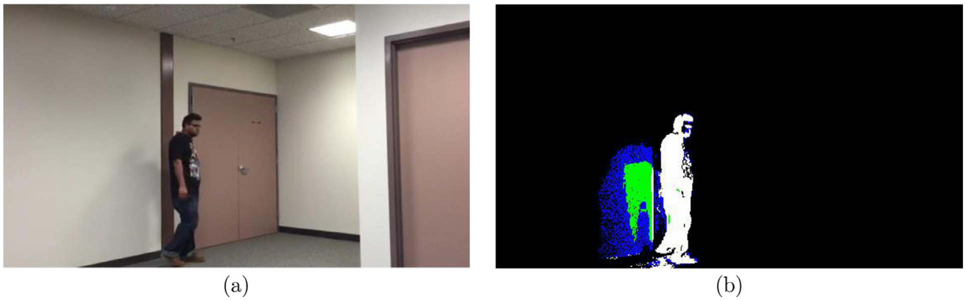


Fig. 2. An example of shadows in an indoor scenario.

In this article, we present a learning-based shadow removal method to suppress shadows in indoor videos that contain complex shadows. Our method categorizes shadows into light shadows and dark shadows based on the color changes with respect to the background model. In dealing with light shadows, chroma of a pixel has little changes but its intensity is slightly reduced. Hence, a threshold is dynamically determined by searching for pixels of the same color but darker in intensity in contrast to the background model. For dark shadows, an online transfer learning-based method is proposed to identify the unwanted regions. A base classifier is initially trained with manually annotated examples and refined with the automatically identified examples in the new videos on-the-fly to adapt to the video under process and to classifier dark shadow pixels.

The rest of this paper is organized as follows: Section 2 presents the related work of shadow removal in videos and, in particular, methods to handle indoor scenarios. Section 3 describes our proposed method in detail. Section 4 discusses the experimental results using several indoor videos. A comparison study is conducted to demonstrate the improvements in our method. Section 5 concludes this paper with a summary and future work.

2. Related work

Shadow removal is a challenging problem in both still images [6] and videos [7]. Although methods that deal with still image can be applied to video frames, their performance degrades and the computational complexity is usually too high for practical applications [8]. To remove shadows from videos, various color models have been explored to characterize their dynamic changes. Cucchiara et al. [9] proposed an HSV color space model for shadow removal from videos. The idea is that shadow changes the hue and the saturation components in a certain range while reduces the brightness. The thresholds are derived from the average image luminance and gradient. Gallego and Pardas [10] implemented a Bayesian method using brightness and color distortion model for shadow removal. Amato et al. [11] developed a method that employs local color constancy. The values of the background image are divided by the values of the current frame in the RGB space. The method assumes that in the luminance ratio space, a low gradient constancy is present in all shadowed regions due to local color constancy. A chroma difference model in RGB space was also developed in [12]. A 3D cone-shaped illumination model was proposed in [13] for background subtraction with shadow removal in indoor surveillance. The work explores the challenges of illumination changes in indoor environments. Gomes et al. [14] integrated color and gradient information with image segmentation using a cascade classifiers. Chromatic and texture features of the foreground objects and their shadows were extracted and classified and a stochastic majority voting scheme was used to detect the shadow regions. Huerta et al. [15] leveraged temporal similarity between textures and spatial similarities between color angle and brightness distor-

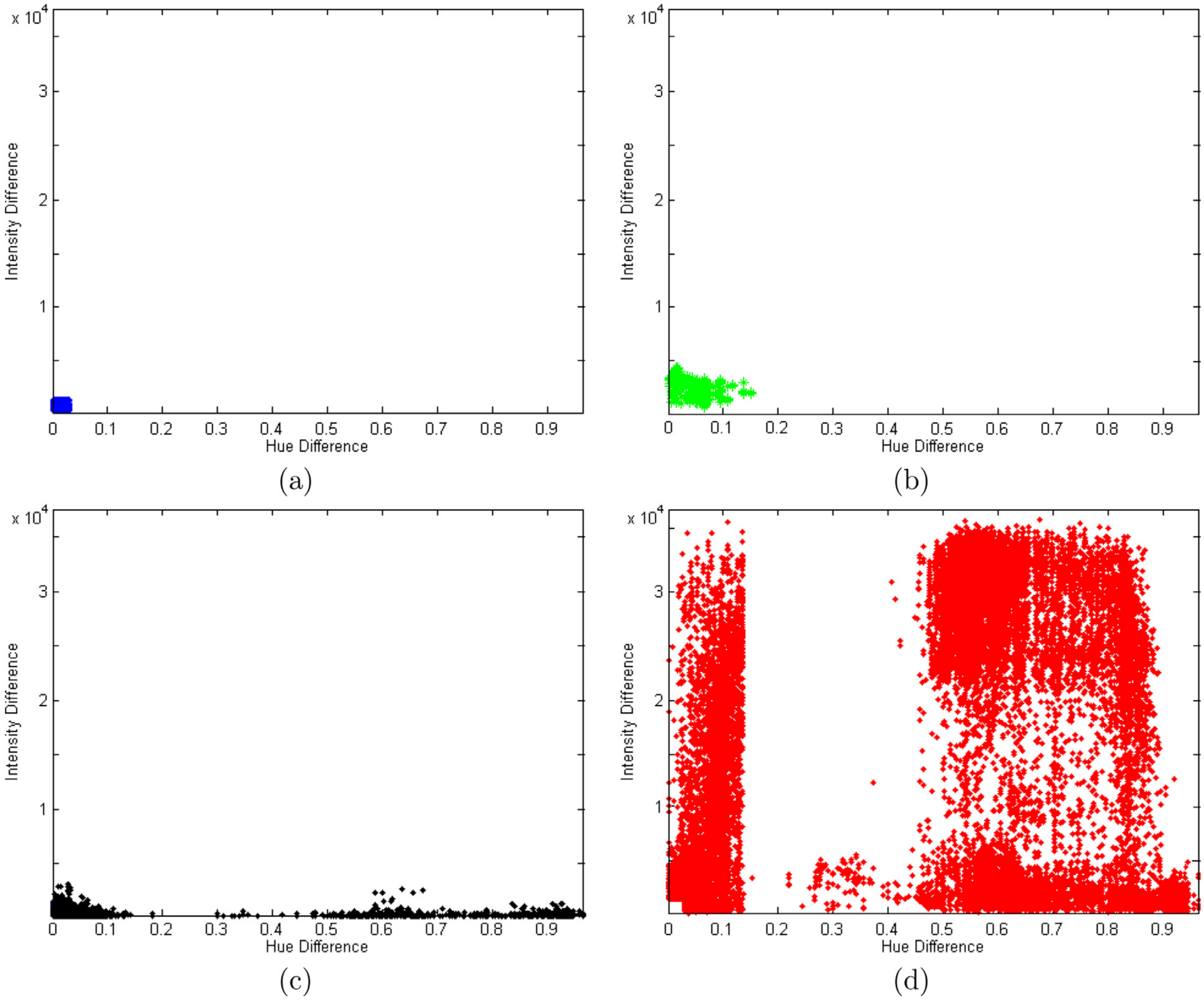
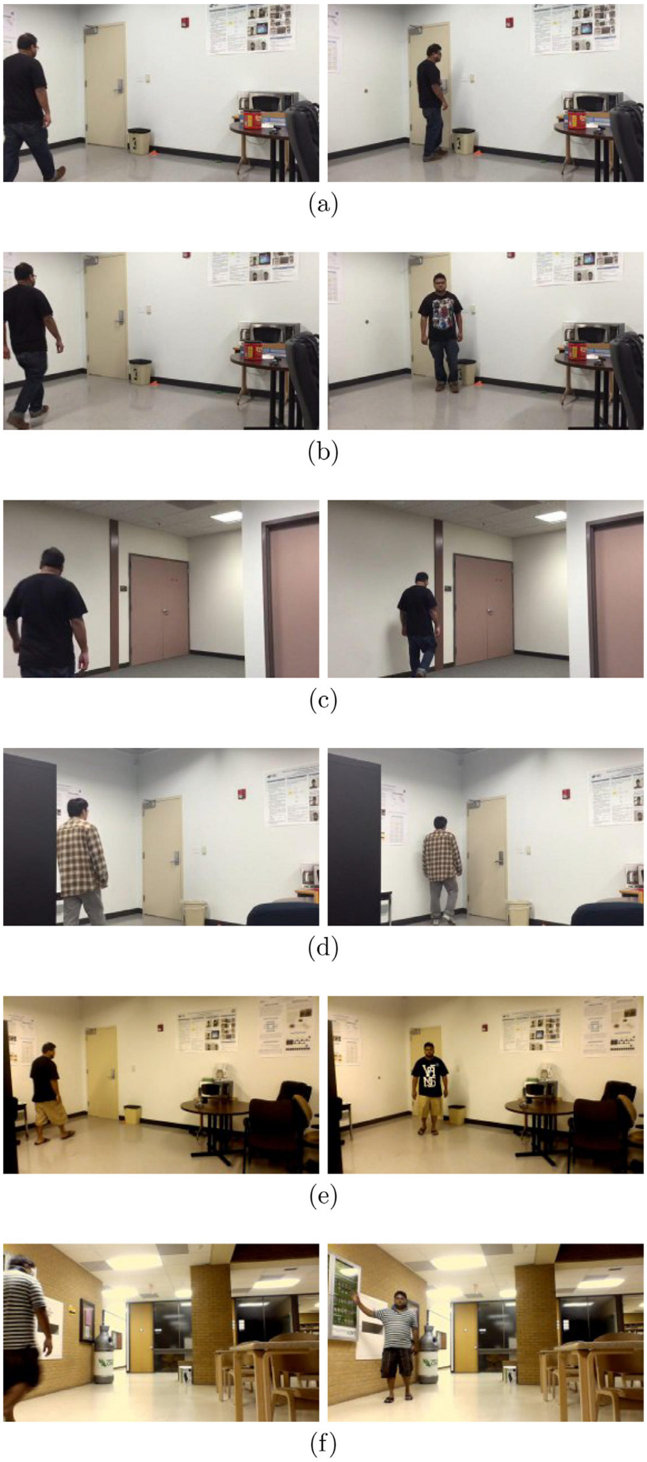


Fig. 3. Pixel distribution in the color difference space. The frames used to create the plots are from one video. (a) is the distribution of light shadow pixels. (b) is the distribution of dark shadow pixels. (c) is the distribution of background pixels. (d) is the distribution of foreground object pixels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tions in a tracking-based approach that recovers missed shadow pixels. A combination of motion filters in a data association framework was proposed, which exploits the temporal consistency between objects and shadows to increase the shadow detection rate. Nghiem et al. [5] employ chromaticity consistency, texture consistency and range of shadow intensity to remove shadows. However, the sensitivity and efficiency are in question [16].

Homogeneity and texture are also employed in shadow detection and removal. Asaril et al. [17] developed a shadow removal method based on the homogeneity property of the shadow. Thresholding and boundary removal are used for removing shadows followed by a validation step that checks the percentage of area that has been removed. Bian et al. [3] implemented a method that uses texture autocorrelation to extract the shadow of a vehicle. Later statistical discrimination is used to analyze the extracted portions. Error correction is performed using integer wavelength transform. Lu et al. [18] proposed a shadow removal method based on the direction of shadows using patch-based comparison on geometrical properties. The algorithm assumes that the shadow will start at the edge of the object. This is true if the whole object is visible from the camera, otherwise, the chance of a disjoint shadow arises. Disjoint shadows are shadows which are not connected spatially to the body.

Learning-based approaches have been developed to model and remove shadows. Wang et al. [19] proposed a dynamic conditional random field model for shadow segmentation in indoor video scenes that uses intensity and gradient features. Temporal and spatial dependencies are unified by the conditional random field. An approximate filtering algorithm is derived to recursively estimate the segmentation field from the observed images. Martel-Brisson and Zaccarin [20] proposed a Gaussian Mixture Model learning algorithm for detecting shadows. Physical properties of light sources and surfaces are employed in order to identify a direction in RGB space at which background surface values under cast shadows are found.


(a)

(b)

(c)

(d)

(e)

(f)

Fig. 4. Exemplar frames from our testing videos. (a)–(f) correspond to the videos listed in Table 1.

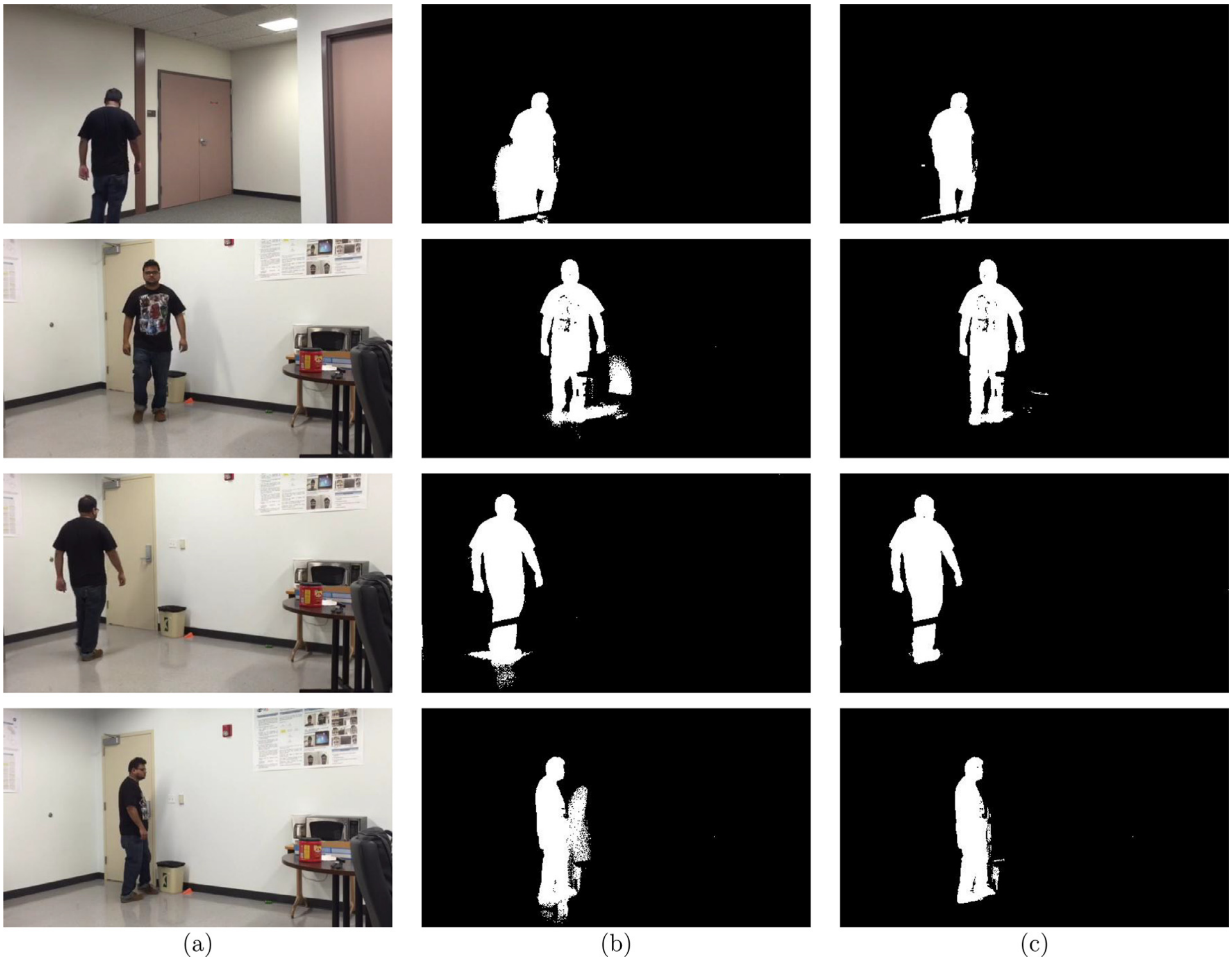


Fig. 5. Exemplar results. (a) are the original video frames. (b) are the background subtraction results using ViBe. (c) are the shadow removed results using our method.

However, the method is affected by the training phase and the computational complexity results in a long learning time. Joshi and Papanikolopoulos [21] proposed a dynamically adapting algorithm that applies co-training to create a classifier with a small number of manually labeled data. Semi-supervised learning helps in adapting to new environments. Intensity, color, and edge features are used to train a support vector machine for shadow removal. Qin et al. [22] employed a clustering method to remove shadows. However, complex indoor lighting conditions have not been discussed at length.

3. Shadow removal using dynamic thresholding and transfer learning

Depending on the position of the imaging device, the shadow appears in different shapes, which is complicated when multiple light sources are present. Fig. 2 illustrates an exemplary frame of an indoor human tracking scenario. The original frame is shown in Fig. 2(a) and the segmented result is shown in Fig. 2(b). In the resulting image, we separate the shadows into two kinds: light shadow and dark shadow. In this figure, the human silhouette is depicted in white, the dark shadow is in green, and the light shadow is in blue. Light shadow usually occurs when the human subject is at a distance to the background wall or there exist other light sources that brighten part of the shadow. The dark shadow, on the other hand, occurs with total (or near total) obstruction of light. The great variation in the shadow intensity makes it difficult to differentiate background from foreground object.

Since light shadows alter the background by slightly dimming its brightness, the hue of the affected pixels has no change, whereas the intensity value decreases slightly, which is proportional to the lighting conditions [4]. Dark shadows, however, greatly alter the background color, which impacts the hue, intensity, and saturation of the affected pixel. We can model the dark shadow in a similar way as the light shadow by setting the lower bound in brightness and hue, yet this could include the color range of true foreground object in dark colors. An example of Pixel distributions in color difference space is shown in Fig. 3. The plots show the distributions of a collection of pixels from several frames, and there are significant overlaps in this color and intensity difference space among shadows, background, and foreground object. It is clear that color or color difference is an unreliable feature to differentiate dark shadow from the rest in a video frame.

To address the issues in complex shadow removal, we propose a learning-based method based on Dynamic Thresholding and Transfer Learning (DyTTL) that deals with light and dark shadows differently based on the aforementioned properties. In summary, a video frame is processed with background subtraction and results in a foreground silhouette that encloses the moving foreground object and possibly a variety of shadows in different shades. Based on the color variation of the foreground object, thresholds are dynamically decided to remove the light shadows. Using pixels from annotated video frames as training examples, a classifier is developed as the initial model for the dark shadows in the video under processing. Using the spatial correlation of image pixels, the most likely neighboring pixels are recruited as training examples to update the classifier.

3.1. Calculating thresholds for light shadow removal

Light shadow changes only the intensity of a pixel. Yet, due to noise, the hue (i.e., color) of a pixel varies slightly over time. Hence, an upper bound for hue difference (denoted by τ_h) and brightness decrement (denoted by τ_i) are used to model light shadows. In this model, the intensity of a shadow-affected pixel decreases, and the difference is within τ_i, subject to the maximum hue change of τ_h induced by noise. Hence, τ_h can be estimated by computing the average hue difference of pixels in the background. To compute τ_h for a video, the candidate pixels are determined with background subtraction. Only those pixels that are in the background in a temporal range are used. Alternatively, if there are many initial frames that contain only stationary objects (i.e., background), the entire frame can be used in the estimation of τ_h.

Following the above idea, the upper bound of intensity difference τ_i to the background pixels can be computed by averaging the pixel intensity in a temporal and spatial neighborhood. Since shadow always reduces the brightness, it is plausible to assume the subtraction of the shadow affected pixel from the corresponding background pixel is always positive. This intensity difference accounts for the variations induced by the imaging factors such as noise, quantization error, etc., as shown in Fig. 3(c). It is also representative of the changes made by light shadows. Note that Fig. 3(c) depicts a broad range of hue difference for the background pixels. Given that the background is stationary, it is expected that both the intensity and hue differences are fairly small. The existence of large hue difference is caused by noise and quantization error. For a typical video frame with 169,016 background pixels, the number of pixels with hue difference greater than 0.2 is 2798, greater than 0.5 is 2765, and greater than 0.9 is 1837. It is clear that the percentage of light shadow pixels with large hue difference is very low (in the range of 0.01%).

3.2. A learning method for dark shadow removal

In contrast to light shadow, dark shadow introduces much brightness and hue changes, which makes it difficult to be separated from the foreground object using thresholding method (as shown in Fig. 3(b) and (d)). By increasing the threshold for intensity and raising the tolerance factor for hue variation, erroneous removals of the foreground object is likely to happen. To address this issue, supervised learning methods have been used [20,21]. Many machine learning methods work well under an assumption that the training and testing data are drawn from the same distribution. When this distribution changes, the existing models need to be rebuilt from scratch, which is expensive and inefficient. The open challenge is the capability to adapt to processing videos that are in different lighting conditions from the training examples.

To remove dark shadows, we propose a learning-based method based on k-Nearest Neighbor (kNN) classifier. In this learning method, a general model $\mathcal{H}$ for dark shadow is first developed using manually segmented video frames. This model $\mathcal{H}$ is used as the base classifier for videos. For each instance in *X*, a set of features are extracted from the video frame as follows:

- *Intensity and hue difference (d_i and d_h)*
  Different from background noise and light shadow, dark shadow introduces much greater changes to the intensity and hue of a pixel. In particular, the brightness of a shadowed area is reduced. These differences are computed with respect to the average intensity and hue of the background model:

  $$d_i(u,v) = \bar{H}(B(u,v)) - H(I(u,v)),$$

  $$d_h(u,v) = \bar{V}(B(u,v)) - V(I(u,v)),$$

  where $H(\cdot)$ and $V(\cdot)$ denote the hue and intensity, respectively; $\bar{H}(\cdot)$ and $\bar{V}(\cdot)$ denote the average hue and intensity components of the HSV space, respectively; $I(u, v)$ and $B(u, v)$ denote an image pixel and a pixel in the background model, respectively.
- *Pixel color in RGB space (r, g, and b)*
  Comparing to the intensity and hue difference, RGB color gives an approximate range of the shadow, which complements the difference feature.

Table 1
Videos acquired for our experimental evaluations.

Videos	Resolution	Lighting condition
A	320 × 568	Bright
B	320 × 568	Moderate
C	320 × 568	Dim
D	320 × 568	Bright
E	720 × 1280	Moderate
F	720 × 1280	Variable

- *Local entropy (e)*
  Local entropy, $e(u, v)$, is used to differentiate the foreground object that might have similar color to the shadows:

  $$e(u,v) = -\sum_i p_i log p_i,$$

  where p_i is the probability of a color in a M by M window. Due to the greater homogeneity of the shadow region, its entropy is lower than that of the object.

When a new video is processed, $\mathcal{H}$ is applied to identify dark shadow pixels in the video frames, and the neighboring pixels of the most confident shadow are recruited as training examples to update this model, which make $\mathcal{H}$ fine-tuned to the variations of the new video such as brightness and tone changes. Our assumption is that the close neighboring pixels of a dark shadow pixel are most likely to be a dark shadow pixel as well. The new examples $I(u, v)$ must satisfy the following criteria to be selected for updating the base classifier $\mathcal{H}$:

- The distance to the most confident dark shadow pixel is less than τ_d.
- The difference of the intensity of the candidate example pixel to the background pixel is greater than τ_i.
- The difference of the hue of the candidate example pixel to the background pixel is greater than τ_h.

Algorithm 1presents our learning-based dark shadow removal method. In this algorithm, $\Sigma_{u,\,v} y(u, v)$ gives the total number of dark shadow pixels and ϵ_s is the minimum number of dark shadow pixel to trigger classifier update. Set S holds the new training examples and is initialized with an empty set. Given a dark pixel $I^t(u, v)$, an instance $I^t(u', v')$ is added to S when it satisfies the distance and confidence criteria. Function $C(\cdot)$ gives the confidence of the prediction of an instance, and the minimum confidence of a dark pixel to serve as a start point for finding new examples is ϵ_c.

Algorithm 1 Transfer learning-based dark shadow detection.

```
for $t \leftarrow \{1, 2, \ldots, T\}$ do
    $\mathcal{H}(I^t(u,v)) \rightarrow y(u,v)$
    if $\sum_{u,v} y(u,v) \geq \epsilon_s$ then
        $S \leftarrow \emptyset$
        for $I^t(u,v) : \mathcal{H}(I^t(u,v)) = 1$ and $C(I^t(u,v)) \geq \epsilon_c$ do
            Get $I^t(u',v') : ||I^t(u',v') - I^t(u,v)||_2 \leq \tau_d$
            if $I^t(u',v')$ satisfies the above criteria and $\mathcal{H}(I^t(u',v')) \neq 1$ then
                $S \leftarrow S \cup I^t(u',v')$
            end if
        end for
        Update $\mathcal{H}$ with examples in $S$
    end if
end for
```

4. Experimental results

4.1. Experimental data and settings

To evaluate our method, we acquired 6 indoor videos in rooms and corridors using two cameras (camera on an iPhone 6 and camera on an ASUS laptop) with different lighting conditions. Table 1 lists the properties of videos used in our experiments. The color depth of the videos is 24 bits and the frame rate is 30 frames per second. Exemplar frames are depicted in Fig. 4.

In our implementation, we adopt ViBe [12] as the background subtraction method for its simplicity and efficiency. However, our method can be combined with any similar method for shadow removal from videos. In ViBe, each pixel in the

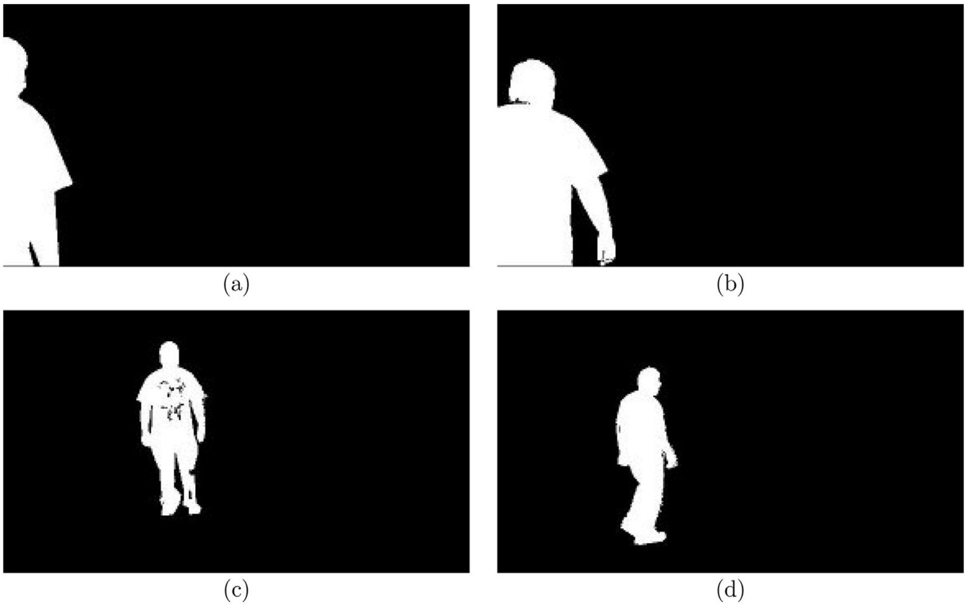


Fig. 6. Ground truth of human silhouette. (a) and (b) are the ground truth images with little shadows. (c) and (d) are the ground truth images with significant amount of shadows.

background model consists of a set of values that describe the possible color range, which is updated randomly in the process of background subtraction. The size of this set is suggested to be 20 based on empirical evaluations of the efficiency and accuracy [12], which is adopted in our implementation.

In our experiments, a minimum number of dark shadow pixels in a video frame is used to control if and when the classifier retraining starts, which is set to 300. When selecting pixels as new training examples, dark pixel confidence is at 100% and pixels in the 4-neighborhood, i.e., $\tau_d = 1$, of the most confident dark pixel are candidate training pixels. The distance metric of kNN classifier is Euclidean distance.

Fig. 5illustrates exemplary frames of our shadow removed foreground segmentation results. The left column depicts the original frames from our videos; the middle column depicts the background subtraction results using ViBe method; the right column depicts the shadow removed foreground segmentation results using our method. It is clear that when shadows are present the foreground object is greatly distorted in the background subtraction results. The shadow caused erroneous foreground regions could be connected to or disconnected from the human silhouette and vary in size and shape. It is demonstrated that our proposed method successfully removes the shadows and introduces little distortions to the foreground object. Note that there are voids (dark pixels) inside the human silhouette or imperfect foreground segmentation in the final results, which are, however, inherited from the background subtraction outcomes. Also shown in these examples is that the lighting conditions in these video frames are clearly different and hence the brightness of shadow varies. Our method is able to adapt to the videos and remove shadows correctly.

4.2. Accuracy analysis

Since we evaluate the performance of shadow removal, it is needed to have reference images of shadows only. However, it is extremely challenging to delineate the shadow region in a video frame even for manual tracing. Alternatively, we prepared the ground truth images of the human silhouette. Another consideration is to exclude errors from the background subtraction process. Due to noise and similar color of the human figure to the background, the output of background subtraction usually contains erroneous segmentation. To suppress the impact of such error to our evaluation of shadow removal, our ground truth is based on the output of the background subtraction procedure that excludes the shadow areas by manual tracing on the resulted foreground object. Fig. 6 depicts a few examples of our ground truth of human silhouette. Fig. 6(c) depicts a ground truth frame that contains errors (black pixels in the upper body) from the ViBe method [12]. In our experiments, we created 60 reference images with the manually segmented human silhouette, among which 25 contains very little shadows and 35 contains a significant amount of light shadows, dark shadows, or a mixture of both.

A key factor for dark shadow removal is the local entropy that differentiates grayish or dark foreground object from the shadow. In our experiments, we adopt sensitivity and specificity to quantify the classification errors to the moving human and to the shadow:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Table 2
Average sensitivity and specificity of classifying shadows with different window sizes for entropy calculation. The values in parenthesis are the corresponding standard deviation.

Window size		Shadow size	
		Small	Large
3 by 3	Sensitivity	58.5% (36.1)	88.3% (5.5)
	Specificity	98.3% (0.9)	98.2% (1.5)
5 by 5	Sensitivity	55.2% (34.6)	87.6% (5.8)
	Specificity	98.5% (0.7)	98.2% (1.6)
7 by 7	Sensitivity	48.1% (33.2)	87.1% (6.2)
	Specificity	98.6% (0.7)	98.1% (1.8)

Table 3
Average sensitivity and specificity of classifying shadows with different number of neighbors in kNN classifier. The values in parenthesis are the corresponding standard deviation.

Number of		Shadow size	
neighbors		Small	Large
k = 3	Sensitivity	59.3% (37.6)	87.3% (5.6)
	Specificity	98.5% (0.8)	98.4% (1.4)
k = 11	Sensitivity	58.5% (35.1)	88.3% (5.5)
	Specificity	98.3% (0.9)	98.2% (1.5)
k = 15	Sensitivity	58.4% (35.7)	88.6% (5.3)
	Specificity	98.2% (1)	98.1% (1.5)

$$\text{Specificity} = \frac{TN}{TN + FP},$$

where TP is the true positive (the correctly classified shadow pixels), TN is the true negative (the correctly classified non-shadow pixels), FP is the false positive (the wrongly classified shadow pixels), and FN is the false negative (the wrongly classified non-shadow pixels). Table 2 lists the average sensitivity and specificity and the corresponding standard deviation of detecting dark shadows using three window sizes. In our comparison, we examine the performance of detecting dark shadows in frames with a few shadows and the ones with a significant amount of shadows separately. The specificity for all cases is above 98% with very little variations. The sensitivity, however, varies greatly, especially for the small shadow cases. However, for frames with large shadow regions, the sensitivity achieves 88.3% and 87.1%, respectively. The lower sensitivity of the small shadow case is mostly due to the small denominator in computing sensitivity. It is evidential that window size of 3 by 3 exhibits the best performance for both small and large shadows. In the rest of our experiments, we use $M = 3$ for computing local entropy.

Table 3lists the average sensitivity and specificity of shadow detection with different numbers of neighbors in kNN classifier. The results of shadow detection when the size of shadow region is small exhibit much lower sensitivity and greater variation compared to the cases with large shadow regions. When the shadow region is small, small mis-detection is likely to have a greater impact to the sensitivity metric because the numerator (i.e., the total count of shadow pixels) is small; whereas the same number of mis-detection of a large shadow appears insignificant. The disparity of sensitivity with respect to the number of neighbors, however, is trivial. In addition, the average accuracy regardless of the shadow size is about 97%. With the goal of maximizing the correct detection of shadow pixels, $k = 15$ yielded the greatest overall average sensitivity of 76.5%.

4.3. Classifier retraining

In our experiments, we used four frames from two videos (videos C and D listed in Table 1) and selected 4221 dark shadow pixels and the equal number of non-dark shadow pixels as the training examples to create a base kNN classifier. The retraining process starts only if there are a significant number of dark pixels (i.e., ϵ_s) identified by the base classifier (or the updated classifier). However, the retraining continues for a certain number of frames. Fig. 7 depicts the average number of new training examples recruited to update the base classifier in the early stage of processing a new video. The maximum number of frames used is 100. The error bar marks the standard deviation among the training for all videos. The average clearly shows the declining number of new examples recruited for retraining, which implies a training convergence.

The average sensitivity and specificity of shadow removal using different confidence levels are presented in Table 4. The average specificities in all cases are very close to 98.2% with small variations (standard deviation at 1.3). The average sensitivity improves slightly as the confidence level increases, and when the confidence level is at 100% the retrained classifier exhibits the greatest sensitivity at 76.6% with a standard deviation of 27. It is clear that a greater confidence level helps to identify the most likely dark shadow pixels. On the other hand, the confidence level poses almost no effects to the perfor-

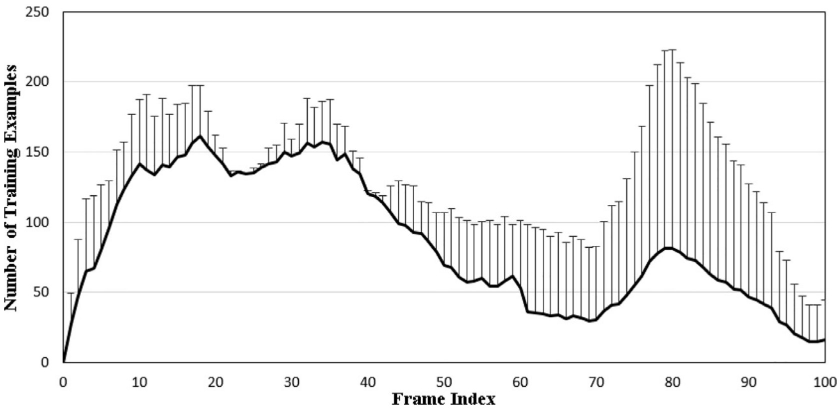


Fig. 7. The average number of new training examples recruited to update the base classifier in the process of a new video. The error bar shows the standard deviation among our testing cases.

Table 4
Average sensitivity and specificity using different confidence levels. The values in parenthesis are the corresponding standard deviation.

Confidence	Sensitivity	Specificity
50%	75.6% (28.3)	98.2% (1.3)
60%	76.0% (27.5)	98.2% (1.3)
70%	75.7% (27.5)	98.2% (1.3)
80%	76.3% (27.1)	98.2% (1.3)
90%	76.3% (27.4)	98.2% (1.3)
100%	76.6% (27.0)	98.2% (1.3)

Table 5
The average time (in second per frame) used for background subtraction and shadow removal. The standard deviation is in parenthesis.

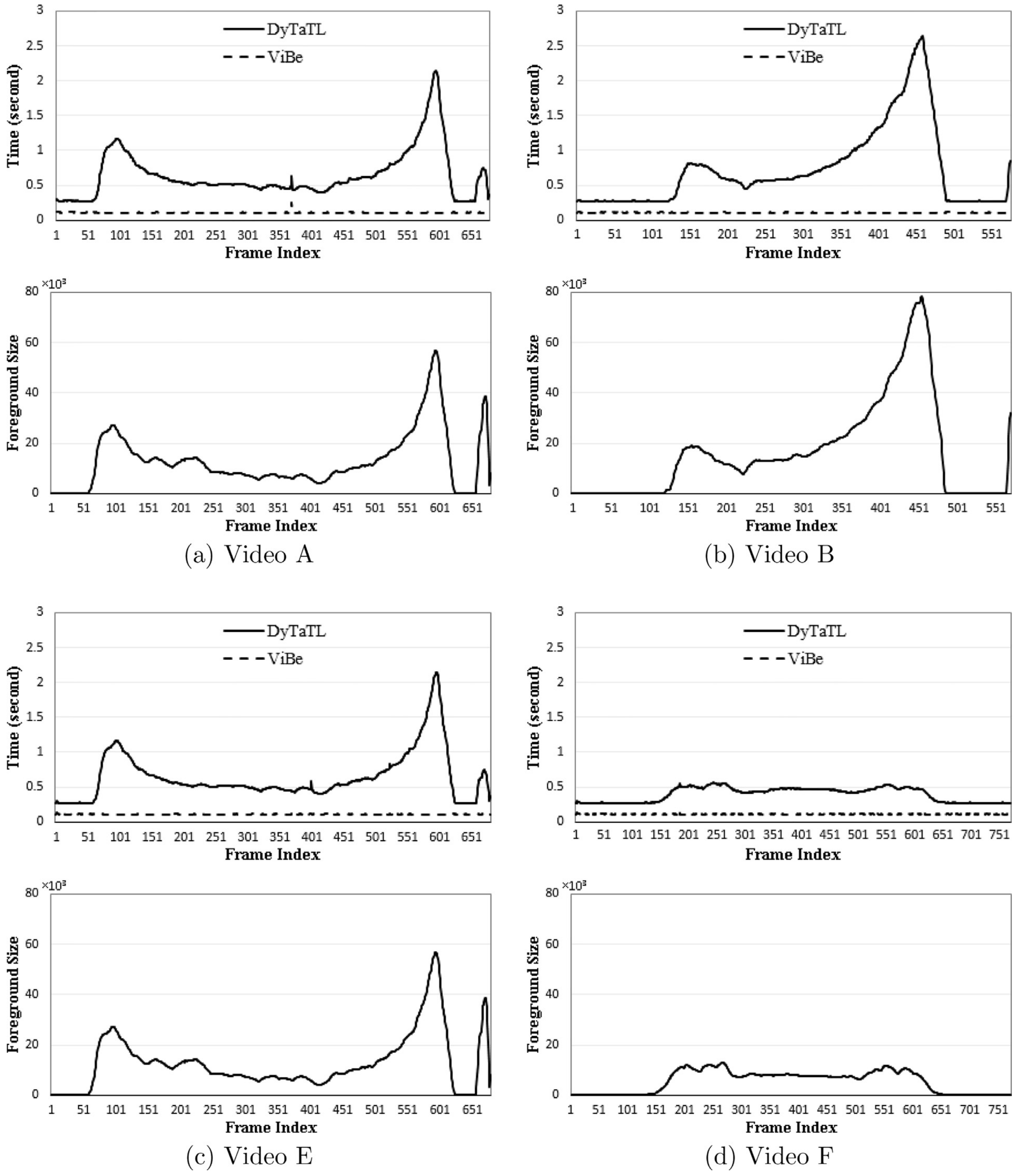
Videos	A	B	E	F
ViBe	0.110 (0.006)	0.109 (0.003)	0.109 (0.002)	0.111 (0.003)
DyTTL	0.532 (0.349)	0.610 (0.558)	0.532 (0.348)	0.284 (0.100)

mance of non-shadow pixels. This might sound counter-intuitive that a lower confidence level usually results in more dark pixels and hence it suppresses false negative results. However, the incorrect choice of dark pixels affects modeling the dark shadow and hence lower the count of true positive, which inherently degrades the sensitivity of detecting dark shadow.

4.4. *Efficiency analysis*

Our algorithm and ViBe method are implemented with MATLAB and tested on a PC system with Intel Core i7-4770 CPU at 3.40 GHz and 16GB memory. Among the six videos used in our experiments, two of them (videos C and D and see Table 1 for additional information about the videos.) were used to generate training examples for transfer learning. Hence, they were excluded from evaluations. Table 5 lists the average time to process a frame in videos using ViBe for background subtraction and using our method for shadow removal. The average time of our method is in the range of half a second and varies greatly between videos, whereas ViBe takes an average of 0.11 s.

In our experiments, we observe that the time for shadow detection within a frame was heavily affected by the size of the foreground including shadow. Fig 8 illustrates the plots of the processing time of each frame and the size of the foreground. In the time plots (the top panel of each sub-figure), the dashed line depicts the time used by ViBe method for background subtraction and the solid line depicts the total time used. The difference between these two lines gives the time used for shadow detection. In videos A, B, and E, the human subject walked away from the camera and then turn back to approach the camera; in video F, the human subject walked across the field of view and the size changed very little. By comparing with the time used to process each frame, it is clear that there is a strong correlation between time used for shadow removal and foreground size.



(a) Video A

(b) Video B

(c) Video E

(d) Video F

Fig. 8. Processing time (top) and size of the foreground (bottom) for the video frames.

5. Conclusion

Shadows in indoor scenarios are usually characterized by multiple light sources that produce complex shadow patterns of a single object. The inconsistent hue and intensity of shadows make automatic removal a challenging task. In this article, we present a hybrid method that leverages transfer learning and dynamic thresholding for removing complex shadows from multiple light sources in indoor environments. Our method suppresses light shadows with a dynamically computed threshold and removes dark shadows using an online learning strategy that is built upon a base classifier trained with manually annotated examples and refined with the automatically identified examples in the new videos.

Our experimental results demonstrate that our proposed method is able to adapt to the videos and remove shadows effectively despite variation of lighting conditions in the environment. The average accuracy reaches more than 97%. The sensitivity of shadow detection changes slightly with different confidence levels used in example selection for classifier retraining, and high confidence level usually yields better performance with less retraining iterations. The shadow detection accuracy using different window sizes for computing local entropy is very close. The window size of 3 by 3 exhibits the satisfactory performance for both small and large shadow regions. In the evaluation of efficiency, updating kNN imposes little impact on the processing time of a frame. Yet, more examples in a kNN model increase the decision time, which can be circumvented with parallel processing because computing the distance of a new instance to each example is independent.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.compeleceng.2017.12.026

References

[1] Zhou Y, Han J, Yuan X, Wei Z, Hong R. Inverse sparse group lasso model for robust object tracking. IEEE Trans Multimedia 2017;19(8):1798–810.
[2] Yuan X, Kong L, Feng D, Wei Z. Automatic feature point detection and tracking of human action in time-of-flight videos. IEEE/CAA J Autom Sin 2017;4(4):677–85.
[3] Bian J, Yang R, Yang Y. A novel vehicle's shadow detection and removal algorithm. In: International conference on consumer electronics, communications and networks. Yichang; 2012. p. 822–6.
[4] Ariel A, Huerta I, Mozerov MG, Roca FX, Gonzez J. Moving cast shadows detection methods for video surveillance applications. Augmented Vision Reality 2012;6:23–47.
[5] Nghiem AT, Bremond F, Thonnat M. Shadow removal in indoor scenes. In: IEEE international conference on advanced video and signal based surveillance, Santa Fe, NM; 2008. p. 291–8.
[6] Arbel E, Hel-Or H. Shadow removal using intensity surfaces and texture anchor points. IEEE Trans Pattern Anal Mach Intell 2010;33(6):1202–16.
[7] Choudhury SK, Sa PK, Bakshi S, Majhi B. An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios. IEEE Access 2016;4:6133–50.
[8] Benedek C, Sziranyi T. Bayesian foreground and shadow detection in uncertain frame rate surveillance videos. IEEE Trans Image Process 2008;17(4):608–21.
[9] Cucchiara R, Grana C, Piccardi M, Prati A. Detecting moving objects, ghosts, and shadows in video streams. IEEE Trans Pattern Anal Mach Intell 2003;25(10):1337–42.
[10] Gallego J, Pardas M. Enhanced bayesian foreground segmentation using brightness and color distortion region-based model for shadowremoval. IEEE international conference on image processing, Hong Kong; 2010.
[11] Amato A, Mozerov MG, Bagdanov AD, Gonzalez J. Accurate moving cast shadow suppression based on local color constancy detection. IEEE Trans Image Process 2011;20(10):2954–66.
[12] Barnich O, Droogenbroeck MV. Vibe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 2011;20(6):1709–24.
[13] Hu J, Su T, Jeng S. Robust background subtraction with shadow and highlight removal for indoor surveillance. In: IEEE/RSJ international conference on intelligent robots and systems, Beijing; 2006. p. 4545–50.
[14] Gomes V, Barcellos P, Scharcanski J. Stochastic shadow detection using a hypergraph partitioning approach. Pattern Recognit 2017;63:30–44.
[15] Huerta I, Holte MB, Moeslund TB, Gonzlez J. Chromatic shadow detection and tracking for moving foreground segmentation. Image Vis Comput 2015;41:42–53.
[16] Sofka M. Commentary paper on "shadow removal in indoor scenes". In: IEEE International conference on advanced video and signal based surveillance; 2008. p. 299–300.
[17] Asari MA, Sheikh UU, Abu-Bakar S. Object's shadow removal with removal validation. In: IEEE international symposium on signal processing and information technology, Giza; 2007. p. 841–5.
[18] Lu Y, Xin H, Kong J, Li B, Wang Y. Shadow removal based on shadow direction and shadow attributes. In: International conference on intelligent agents, computational intelligence for modelling, control and automation and international conference on web technologies and internet commerce, Sydney, NSW; 2006. p. 37.
[19] Wang Y, Loe K-F, Wu J. A dynamic conditional random field model for foreground and shadow segmentation. IEEE Trans Pattern Anal Mach Intell 2006;28(2):279–89.
[20] Martel-Brisson N, Zaccarin A. Learning and removing cast shadows through a multidistribution approach. IEEE Trans Pattern Anal Mach Intell 2007;29(7):1133–46.
[21] Joshi A, Papanikolopoulos N. Learning to detect moving shadows in dynamic environments. IEEE Trans Pattern Anal Mach Intell 2008;30(11):2055–63.
[22] Qin Y, Sun S, Ma X, Hu S, Lei B. A background extraction and shadow removal algorithm based on clustering for ViBe. In: International conference on machine learning and cybernetics, Lanzhou; 2014. p. 52–7.

Xiaohui Yuan received a Ph.D. degree from Tulane University in 2004. He is an Associate Professor in the University of North Texas, Denton, USA and a Visiting Professor in the China University of Geosciences, Wuhan, China. He is a recipient of Ralph E. Powe Junior Faculty Enhancement award in 2008. His research interests include computer vision, machine learning, and artificial intelligence.

Daniel Li is currently a student at the Texas Academy of Mathematics & Science at the University of North Texas. He also is a member of the Computer Vision and Intelligent Systems Lab at UNT. He is a recipient of the Undergraduate Research Fellowship in 2017.

Deepankar Mohapatra received a Master degree from the University of North Texas in 2015. He was a member of the Computer Vision and Intelligent Systems Lab at the University of North Texas. His research interests include computer vision, data mining, and artificial intelligence.

Mohamed Elhoseny is an Assistant Professor at Faculty of Computers and Information, Mansoura University, Egypt. He received his Ph.D. in Computer and Information Sciences from Mansoura University (in a scientific research channel with Department of Computer Science and Engineering, University of North Texas, USA). Collectively, he co-authored over 50 International Journal articles, Conference papers, Book Chapters, and 2 Springer edited books.