Tricolor Attenuation Model for Shadow Detection

Jiandong Tian, Jing Sun, and Yandong Tang

***Abstract*—Shadows, the common phenomena in most outdoor scenes, bring many problems in image processing and computer vision. In this paper, we present a novel method focusing on extracting shadows from a single outdoor image. The proposed tricolor attenuation model (TAM) that describe the attenuation relationship between shadow and its nonshadow background is derived based on image formation theory. The parameters of the TAM are fixed by using the spectral power distribution (SPD) of daylight and skylight, which are estimated according to Planck's blackbody irradiance law. Based on the TAM, a multistep shadow detection algorithm is proposed to extract shadows. Compared with previous methods, the algorithm can be applied to process single images gotten in real complex scenes without prior knowledge. The experimental results validate the performance of the model.**

***Index Terms*—Outdoor scenes, shadow detection, single image, tricolor attenuation model (TAM).**

I. Introduction

SHADOWS may cause some undesirable problems in many computer vision and image analysis tasks, such as edge detection, image segmentation, object recognition, video surveillance, and stereo registration. Detecting and removing shadows in images are of great practical significance in image processing, which have attracted a great deal of attention recently. However, shadows are difficult to be detected especially in single outdoor images. The tricolor attenuation model (TAM) based algorithm presented in the paper can solve the problem to some extent.

In outdoor scenes, there are mainly two light sources: direct sunlight, which can be regarded as a point light source; diffuse skylight, which can be regarded as an area light source. Shadows will occur when direct light from a light source is partially or totally occluded. Shadow can be divided into two types: self shadow and cast shadow. The self shadow is the part of an object that is not illuminated by direct light; the cast shadow is the dark area projected by an object on the background. Cast shadow can be further divided into umbra and penumbra region. The umbra region is the part of a cast shadow where direct light is completely blocked; the penumbra region is the part of a cast shadow where direct light is only partially blocked. As shown in Fig. 1, the illumination on nonshadow region is daylight (direct sunlight and diffused skylight); that on penumbra is skylight and part of sunlight; that on umbra is only skylight. Since skylight is a component of daylight, pixel intensity in shadow must be lower than that in nonshadow background.

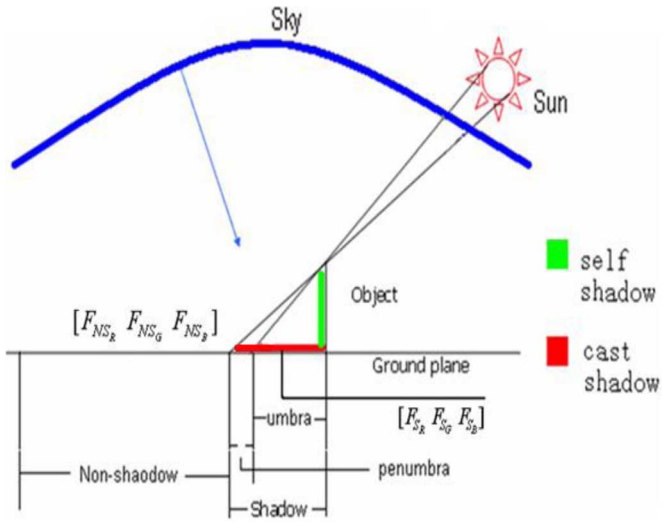


Fig. 1. Shadow will occur when direct light is occluded.

Denoting $[F_R\ F_G\ F_B]$ as a tricolor vector of a pixel value in a color image F, $[F_{\mathrm{NS}_R}\ F_{\mathrm{NS}_G}\ F_{\mathrm{NS}_B}]$ as a pixel value vector in a nonshadow background region, $[F_{S_R}F_{S_G}F_{S_B}]$ as a pixel value vector in the corresponding shadow region which has the same response of reflectance as $[F_{\mathrm{NS}_R}F_{\mathrm{NS}_G}F_{\mathrm{NS}_B}]$, and $[\Delta R\Delta G\Delta B]$ as the value attenuation vector, the relationship between $[F_{\mathrm{NS}_R}F_{\mathrm{NS}_G}F_{\mathrm{NS}_B}]$ and $[F_{S_R}F_{S_G}F_{S_B}]$ is

$$\begin{cases} F_{S_R} = F_{\mathrm{NS}_R} - \Delta R \\ F_{S_G} = F_{\mathrm{NS}_G} - \Delta G \\ F_{S_B} = F_{\mathrm{NS}_B} - \Delta B. \end{cases} \tag{1}$$

Equation (1) implies that if $\Delta R, \Delta G, \Delta B$ are different, the disparities of R, G, B channels of a shadow region are expected to be different from those of the corresponding nonshadow background region. Taking $\Delta R > \Delta G > \Delta B$ as an example, if we subtract B channel from R channel

$$\begin{aligned} &F_{S_R} - F_{S_B} \\ &\quad = F_{\mathrm{NS}_R} - \Delta R - (F_{\mathrm{NS}_B} - \Delta B) \\ &\quad = F_{\mathrm{NS}_R} - F_{\mathrm{NS}_B} + (\Delta B - \Delta R) < F_{\mathrm{NS}_R} - F_{\mathrm{NS}_B}. \end{aligned} \tag{2}$$

Obviously, in this example, the disparity between R and B channels of shadow is lower than that of the corresponding nonshadow background.

Manuscript received February 25, 2009; revised June 03, 2009. First published July 06, 2009; current version published September 10, 2009. This work was supported in part by the Natural Science Foundation of China (Grant Number: 60871078 and 60835004). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark (Hong-Yuan) Liao.

J. Tian and J. Sun are with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, and also with the Graduate School of the Chinese Academy of Sciences, Beijing, China, 100039 (e-mail: tianjd@sia.cn; sunjing@sia.cn).

Y. Tang is with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016 (e-mail: ytang@sia.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2026682

Our model originates from the idea that if we subtract the minimum attenuated channel from the maximum attenuated channel, the results in shadow regions will be lower than the results in nonshadow regions. This is very useful for shadow identification. In fact, in most cases, $\Delta R, \Delta G, \Delta B$ are different. The key problem is how to find the maximum and the minimum attenuated channels. It is analyzed and solved in the following sections. To begin with, we deduce the TAM based on the mechanism of image formation. To determine the parameters of TAM, we employ Planck's blackbody irradiance theory parameterized by the correlated color temperature (CCT) to estimate the SPDs of daylight and skylight. Finally, a multistep shadow extraction algorithm based on the proposed TAM is presented to test our model.

There are mainly three contributions in the paper.

1) Deducing the novel tricolor attenuation model, on which a completely data-driven shadow detection algorithm is proposed.
2) Presenting an approximate shadow invariant transformation from a RGB image to a gray image.
3) Giving a simple way, which we called T operator, to obtain an illumination invariant image from a color image.

The rest of the paper is organized as follows. In Section II, we review the research background; in Section III, we deduce TAM theoretically; in Section IV, we present the shadow detection algorithm based on the TAM; in Section V, we give and analyze some experimental results, followed by the discussion in Section VI and a brief conclusion in Section VII.

II. Background

Technically, shadow detection methods can be classified into property-based and physics-based. Physics-based techniques need some prior knowledge, such as light and geometry [1], camera calibration [2], or indoor scenes [3]. However, it is extremely difficult to obtain the accurate model for an arbitrary scene because the environment are complex and the light sources vary from time to time and from place to place. Hence, most of physics-based techniques are designed for specific applications, such as moving cast shadow detection [4], [5] and shadow detection in aerial images [6]. Physics-based methods exploit some knowledge of the scene, which result in these methods being only used in specific applications they are designed for. When the application environments are different, the algorithms may fail.

Property-based techniques identify shadows through shadow features. The most straightforward feature of a shadow is that it darkens the surface it cast on, and this feature is used by almost all methods. Other features like edge [7], [8], histograms [9], texture [10], geometry property [11], color ratios [12], and gradient [13], [14] are also widely adopted. Sometimes, only one feature is not enough. For example, shadows usually have lower pixel values, but pixels that have lower values may not be shadows. Computer vision cannot directly judge that a dark region is a shadow or is a black object. Therefore, most methods combine more than one feature. For example, in [15], color information is combined with geometric information to detect cast shadows. Property-based approaches are more flexible than physics-based ones. They can be applied to a wider class of scenes.

According to the number of images used, shadow detection can be further classified into multiimage and single-image methods. The prior literatures mainly focused on the shadows detection from image sequences, i.e., they used multiple images. Multiimage methods are mainly applied to detect moving shadows, such as [4], [5], [16], and [17]. Multiple images can provide more information than single image for shadow detection. In [18], shadows are detected according to the differences of adjacent frames. Finlayson *et al.* [19] employ a chromagenic camera to take two pictures of each scene to detect illumination and shadows. In [20], the authors employ a sequence of images to generate a 1-D illumination invariant image. The image is used together with the original image to locate shadows. Prati *et al.* [21] presented a good review for shadow detection methods in image sequences.

As detecting shadows from image sequences has made great progress, detecting them from a single image remains a difficult problem. In contrast to multiimages, shadow detection in still image is more difficult due to less information available. Wu *et al.* employ the Bayesian approach [22] and shadow matting [23] to extract shadows in a single image, but their method requires user interaction as input. Nielsen *et al.* [24] employ α-channel for soft shadow segmentation, but it requires to manually handpick a sunlit surface and its shadow counterpart to initialize the overlay color. The method proposed in [25] can identify and remove shadows from a sole outdoor image, but it requires user-supplied regions. Thus, these methods cannot be used in totally automatic computer vision tasks. In [26], the authors employ Markov random field model to detect shadows automatically in a single color image. However, this method cannot work on complex scenes.

The method called "color invariance" has been extensively researched in recent years. Color invariance features are not sensitive to illumination changes to some extent. Color invariance features mainly includes YUV [4], normalized RGB [7], hue (H) and saturation (S) [27], and $c_1c_2c_3$ [28]. Shadows mainly change the intensity of the surface that they cover with but seldom change the color invariance features. Therefore, using the shadow invariant image, hard shadow edge mask can be estimated by comparing the original image and the invariant image. Unfortunately, these methods cannot totally eliminate the illumination effect and, thus, are mostly applied in simple scenes.

In spite of these extensive studies, many proposed approaches are either designed for specific applications or need some assumptions about the environments. In addition, the majority of the proposed methods focus on detecting moving shadows in image sequences. Even if some methods can work on single still image, they have suffered from at least one of the following three problems.

1) Needing some prior knowledge, such as human's interaction [22].
2) Effective in specific application [6].
3) Failing on complex scenes [26].

In contrast, the method proposed in this paper is not designed for specific applications and can automatically extract

shadows from single still images, even those with complex outdoor scenes. Furthermore, our method does not need prior knowledge. It is completely data-driven.

III. Tricolor Attenuation Model

A. Image Formation Theory

Generally, a color image derived by a camera is determined by four main factors: illumination spectral power distribution, surface reflectance, spectral sensitivities of the camera, and the postprocessing on the RAW data. RAW data is unprocessed electric signal data converted from optical signal by CCD or CMOS sensors. Cameras apply some nonlinear post processing to the linear RAW data, such as white-balance, gamma correction, compression. For details, we refer the reader to [30].

A pixel value vector in a color image derived by a camera can be expressed as

$$F(x,y,k) = G\left[\sigma \int_{\lambda} E(\lambda,\Theta)S(\lambda,\Theta)Q(\lambda,k)d\lambda\right] \tag{3}$$

where:

- $F(x,y,k)$ is the output signal at location (x,y) in $k = \{R,G,B\}$ color channel;
- λ is wavelength;
- Θ is viewing geometry;
- σ is exposure time;
- E is SPD of incident light;
- S is surface reflecting function;
- Q is camera response function;
- G is postprocessing function.

This model is difficult to be analyzed in mathematics due to the nonlinear postprocessing. To reduce its complexity, we neglect the nonlinear G function. Furthermore, our method is not concerned about the pixel locations. A simplified model is as the following:

$$F_k = \sigma \int E(\lambda)S(\lambda)Q_k(\lambda)d\lambda, \quad k = \{R,G,B\}. \tag{4}$$

If the camera has narrowband sensitivity, i.e., the camera sensor response properties can be approximated by Dirac delta function: $Q_k(\lambda) = q_k\delta(\lambda - \lambda_k)$. Then, formula (4) is reduced to a simpler form

$$F_k = \sigma E(\lambda_k)S(\lambda_k)q_k. \tag{5}$$

B. Tricolor Attenuation Model

Denoting the SPD of the illumination on nonshadow region as E_1 and that on shadow region as E_2, substituting (5) into (1), we have

$$\sigma\begin{bmatrix} E_1(\lambda_R)S(\lambda_R)q_R \\ E_1(\lambda_G)S(\lambda_G)q_G \\ E_1(\lambda_B)S(\lambda_B)q_B \end{bmatrix} - \sigma\begin{bmatrix} E_2(\lambda_R)S(\lambda_R)q_R \\ E_2(\lambda_G)S(\lambda_G)q_G \\ E_2(\lambda_B)S(\lambda_B)q_B \end{bmatrix} = \begin{bmatrix} \Delta R \\ \Delta G \\ \Delta B \end{bmatrix}. \tag{6}$$

Dividing both sides of (6) by ΔB,[1] we get

$$\begin{cases} \frac{\Delta R}{\Delta B} = \frac{\sigma S(\lambda_R)q_R[E_1(\lambda_R) - E_2(\lambda_R)]}{\sigma S(\lambda_B)q_B[E_1(\lambda_B) - E_2(\lambda_B)]} \\ \frac{\Delta G}{\Delta B} = \frac{\sigma S(\lambda_G)q_G[E_1(\lambda_G) - E_2(\lambda_G)]}{\sigma S(\lambda_B)q_B[E_1(\lambda_B) - E_2(\lambda_B)]} \end{cases}. \tag{7}$$

From (5), we have

$$\begin{cases} \frac{\sigma S(\lambda_R)q_R}{\sigma S(\lambda_B)q_B} = \frac{F_{\mathrm{NS}_R}}{F_{\mathrm{NS}_B}} \cdot \frac{E_1(\lambda_B)}{E_1(\lambda_R)} \\ \frac{\sigma S(\lambda_G)q_G}{\sigma S(\lambda_B)q_B} = \frac{F_{\mathrm{NS}_G}}{F_{\mathrm{NS}_B}} \cdot \frac{E_1(\lambda_B)}{E_1(\lambda_G)} \end{cases}. \tag{8}$$

Substituting formula (8) into (7), we have

$$\begin{cases} \frac{\Delta R}{\Delta B} = \frac{F_{\mathrm{NS}_R}}{F_{\mathrm{NS}_B}} \cdot \frac{E_1(\lambda_B)}{E_1(\lambda_R)} \cdot \frac{E_1(\lambda_R) - E_2(\lambda_R)}{E_1(\lambda_B) - E_2(\lambda_B)} = m \cdot \frac{F_{\mathrm{NS}_R}}{F_{\mathrm{NS}_B}} \\ \frac{\Delta G}{\Delta B} = \frac{F_{\mathrm{NS}_G}}{F_{\mathrm{NS}_B}} \cdot \frac{E_1(\lambda_B)}{E_1(\lambda_G)} \cdot \frac{E_1(\lambda_G) - E_2(\lambda_G)}{E_1(\lambda_B) - E_2(\lambda_B)} = n \cdot \frac{F_{\mathrm{NS}_G}}{F_{\mathrm{NS}_B}} \end{cases} \tag{9}$$

where

$$\begin{cases} m = \frac{E_1(\lambda_B)}{E_1(\lambda_R)} \cdot \frac{E_1(\lambda_R) - E_2(\lambda_R)}{E_1(\lambda_B) - E_2(\lambda_B)} \\ n = \frac{E_1(\lambda_B)}{E_1(\lambda_G)} \cdot \frac{E_1(\lambda_G) - E_2(\lambda_G)}{E_1(\lambda_B) - E_2(\lambda_B)} \end{cases}. \tag{10}$$

Then, the vector $[\Delta R \Delta G \Delta B]$ can be represented by

$$\begin{bmatrix} \Delta R \\ \Delta G \\ \Delta B \end{bmatrix} = \begin{bmatrix} \Delta R/\Delta B \cdot \Delta B \\ \Delta G/\Delta B \cdot \Delta B \\ 1 \cdot \Delta B \end{bmatrix} = \begin{bmatrix} m \cdot F_{\mathrm{NS}_R}/F_{\mathrm{NS}_B} \\ n \cdot F_{\mathrm{NS}_G}/F_{\mathrm{NS}_B} \\ 1 \end{bmatrix} \cdot \Delta B. \tag{11}$$

We have obtained the TAM shown in formula (11). The model does, definitely, depend on the assumption of the infinitely narrow camera responsibility. However, in fact, the real camera sensors are not exactly narrow-band. This will be further discussed in Section VI. The maximum channel and the minimum channel in $[\Delta R\ \Delta G\ \Delta B]$ equal to those in $[m \cdot F_{\mathrm{NS}_R}/F_{\mathrm{NS}_B}\ n \cdot F_{\mathrm{NS}_G}/F_{\mathrm{NS}_B} 1]$ because of $\Delta B \geqslant 0$. To find the maximum attenuated channel and the minimum one, obtaining a probably value of parameters m and n becomes necessary.

C. Parameters Estimation

Denote E_{sun} as the SPD of direct sunlight and E_{sky} as that of diffuse skylight. As shown in Fig. 1, nonshadow region is lighted by daylight (sunlight plus skylight). Therefore, $E_1 = E_{\mathrm{sun}} + E_{\mathrm{sky}} = E_{\mathrm{day}}$; shadow region is lighted by skylight plus partially sunlight, and can be denoted as

$$E_2 = \alpha E_{\mathrm{sun}} + E_{\mathrm{sky}}. \tag{12}$$

Here, α in (12) is a proportional factor. In umbra region, α equals to 0; in nonshadow, α equals to 1; in penumbra region, $\alpha \in (0\ 1)$.

[1]Here, assume $\Delta B \neq 0$. Otherwise, using ΔR or ΔG which not equal zero instead. ΔR, ΔG, and ΔB will not be all zeros because the pixel intensity attenuates when shadow happens.

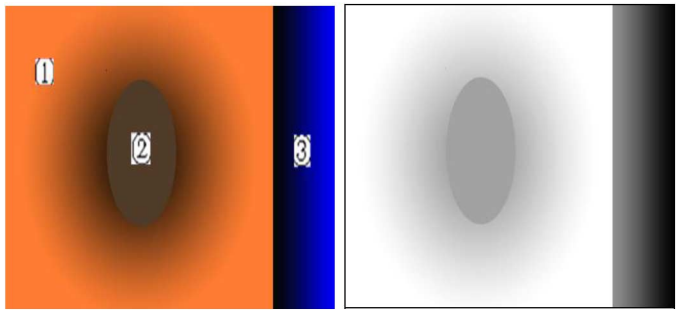


Fig. 2. Subtraction on whole image: (a) red color is the domain color; (b) the subtractive image using R-B. The black box around the image is for illustrative purpose only.

Rewriting (10), we get

$$\begin{cases} m = \frac{E_{\text{day}}(\lambda_B)}{E_{\text{day}}(\lambda_R)} \cdot (1-\alpha)\frac{E_{\text{sun}}(\lambda_R)}{E_{\text{sun}}(\lambda_B)} \\ n = \frac{E_{\text{day}}(\lambda_B)}{E_{\text{day}}(\lambda_G)} \cdot (1-\alpha)\frac{E_{\text{sun}}(\lambda_G)}{E_{\text{sun}}(\lambda_B)} \end{cases}. \tag{13}$$

We will use blackbody irradiance to approximate the SPDs of daylight and direct sunlight. According to Planck's law [29], the spectral of a blackbody at color temperature T, per unit wavelength interval, is given by

$$E(\lambda, T) = c_1 \lambda^{-5} \left(e^{c_2/T\lambda} - 1 \right)^{-1} \tag{14}$$

where $c_1 = 2\pi hc^2$ and $c_2 = (hc)/(k)$, in which:

- c is the velocity of light—equals to 3.0×10^8 m.s^{-1};
- h is the Plank constant -equals to 6.63×10^{-34} J.s;
- k is the Boltzmann constant -equals to 1.38×10^{-23} J.K^{-1}.

Therefore, we obtain

$$c_1 = 3.74 \times 10^{-16}\ (\text{W} \cdot \text{m}^2)$$
$$c_2 = 1.43 \times 10^{-2}\ (\text{m} \cdot \text{K}).$$

Taking $T_{\text{day}} = 6500$ K (standard daylight illuminant D_{65} [29]) for daylight, $T_{\text{sun}} = 5500$ K for direct sunlight, and setting the wavelengths at red, green, blue equal to 650, 550, and 440 nm, respectively, it is easy to get

$$\frac{E(\lambda_B, T_{\text{day}})}{E(\lambda_R, T_{\text{day}})} = \left(\frac{\lambda_B}{\lambda_R} \right)^{-5} \cdot \frac{e^{\frac{c_2}{T_{\text{day}}\lambda_R}} - 1}{e^{\frac{c_2}{T_{\text{day}}\lambda_B}} - 1} = 1.36 \tag{15}$$

similarly

$$\frac{E(\lambda_B, T_{\text{day}})}{E(\lambda_G, T_{\text{day}})} = 1.11;$$
$$\frac{E(\lambda_R, T_{\text{sun}})}{E(\lambda_B, T_{\text{sun}})} = 0.97; \quad \frac{E(\lambda_G, T_{\text{sun}})}{E(\lambda_B, T_{\text{sun}})} = 1.07.$$

Substituting them into (13) and setting $\alpha = 0$ for simplicity, it is easy to get: $m = 1.31$ and $n = 1.19$. The parametric values of m and n are just approximations because they are calculated from the standard CCTs that may not be same as the actual CCTs under which the experimental images are taken. In fact, it is impossible to calculate the accurate values because the SPDs of daylight and skylight are extremely variable. They vary with latitude, season, and weather conditions etc. However, The SPDs we used here are under the situation that shadows most likely take place. Hence, the estimated parameters will be not off the real ones much.

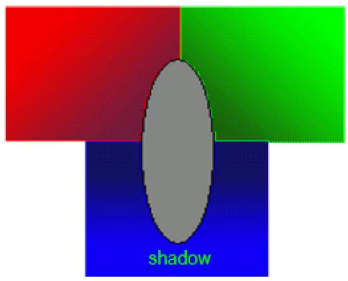


Fig. 3. Same shadow falls on different color surface.

IV. Shadow Detection Algorithm Based on the Model

Based on the theory discussed above, we introduce our multistep algorithm for shadow detection. Before detection, the calculating regions of the algorithm should be determined. It is meaningless to calculate $[m \cdot F_{\text{NS}_R}/F_{\text{NS}_B} n \cdot F_{\text{NS}_G}/F_{\text{NS}_B} 1]$ pixel-wisely because shadow and nonshadow will certainly not be a same pixel. Calculating it on the whole image is also not proper. As shown in Fig. 2(a), region ① is nonshadow background and region ② is shadow region. Red color is the domain color (in region ①R > G > B). If we subtract B channel from R channel, as shown in Fig. 2(b), not only the shadow region but also the blue region will be apparent from domain color region, that is, region ② and region ③ are both darker than region ①. Then the blue region may be falsely detected as shadows.

If a same shadow falls on different color surface, as shown in Fig. 3, calculating it on the whole image may also cause false shadow detection.

Therefore, we introduce a preprocessing to segment the original image into sub-regions with similar color. The algorithm proposed here just needs a rough segmentation but does not need an accurate one. Because segmentation methods are

easily affected by shadows, it is better to segment on a shadow invariant image instead of the original one. Although there are many researches on illumination invariance, such as Retinex [31] and intrinsic image [32], they are not easy to be used due to the limitation of preconditions and the complexity of the algorithms. Another well-known illumination invariant color space HSI has disadvantages of nonremovable singularity numerical instability at low saturation due to nonlinear transformation [33]. In the paper, we adopt formula (16) to transform a RGB image to a gray one that is not sensitive to shadows so much compared with the original one

$$Y = \log\left(\frac{\max[F_R\ F_G\ F_B]}{\min[F_R\ F_G\ F_B]+1}\right). \tag{16}$$

Here, adding 1 is to avoid zero value of denominator.

The reasons to use the transformation for computing the shadow invariant image are listed below.

1) The color ratios (chromaticity) do not vary so obviously as pixel intensity when illumination changes.
2) Constant color ratios such as F_R/F_B is not appropriate. Taking $F_{\mathrm{NS}_R} > F_{\mathrm{NS}_G} > F_{\mathrm{NS}_B}$ for an example, we get $\Delta R > \Delta G > \Delta B$ from formula (11). However, the inequality $F_{S_R} < F_{S_G} < F_{S_B}$ may happen according to (1). The difference between shadow and nonshadow obtained by color ratios is large due to $F_{\mathrm{NS}_R}/F_{\mathrm{NS}_B} > 1$ and $F_{S_R}/F_{S_B} < 1$. The formula (16) using the ratio of maximum channel and minimum channel instead of fixed channel ratio, so, $F_{\mathrm{NS}_R}/F_{\mathrm{NS}_B} > 1$ and $F_{S_B}/F_{S_R} > 1$. In most case, the difference between shadow and nonshadow obtained by formula (16) is smaller than that obtained by color ratios.
3) The logarithm operation is used for compressing the dynamic range to overcome the over segmentation of the watershed segment algorithm that will be employed by our algorithm.

Our multistep algorithm for shadow detection is as follows.

Step 1: To transform the original color image F into the gray image Y by formula (16).

Step 2: To segment Y into sub-regions with similar color by the well-known watershed algorithm, that is, $Y = \cup Y^i$, where $Y^i \cap Y^j = \Phi$ if $i \neq j$, and i is the segmented region number.

Step 3: For each region, to calculate the mean value of $[F_R^i\ F_G^i\ F_B^i]$ by

$$\left[\overline{F_R^i}\ \overline{F_G^i}\ \overline{F_B^i}\right] = \frac{1}{M}\left(\sum_{k\in F^i}^{k=1\ldots M.}\left[F_R^{i,k}F_G^{i,k}F_B^{i,k}\right]\right) \tag{17}$$

where $F_R^{i,k}$ denotes the kth pixel in the ith region of F in R channel, and M is the number of pixels in region i.

Step 4: To calculate the mean value of $[F^i_{\mathrm{NS}_R}\ F^i_{\mathrm{NS}_G}\ F^i_{\mathrm{NS}_B}]$ by (17). As shown in formula (11), $[\Delta R\ \Delta G\ \Delta B]$ is determined by $F_{\mathrm{NS}_R}/F_{\mathrm{NS}_B}$ and $F_{\mathrm{NS}_G}/F_{\mathrm{NS}_B}$. However, we do not know where is nonshadow and where is shadow before detection. We take pixels whose values are larger than the mean value of region i as the nonshadow background in this step. Then $[m\cdot\overline{F^i_{\mathrm{NS}_R}}/\overline{F^i_{\mathrm{NS}_B}}\,n\cdot\overline{F^i_{\mathrm{NS}_G}}/\overline{F^i_{\mathrm{NS}_B}}\,1]$ is calculated.

Step 5: To subtract the minimum channel from the maximum one. Namely, if $m\cdot\overline{F^i_{\mathrm{NS}_R}}/\overline{F^i_{\mathrm{NS}_B}} > n\cdot\overline{F^i_{\mathrm{NS}_G}}/\overline{F^i_{\mathrm{NS}_B}}>1$ in F^i, we get $X^i = F^i_R - F^i_B$, and vise versa.

Step 6: To binarizate X^i by the following threshold based on the observation that shadows are often darker than the mean value of X^i

$$T = \frac{1}{M}\left(\sum_{k\in X^i}^{k=1\ldots M.} X_k^i\right) \tag{18}$$

where M is the number of pixels in region i. An initial shadow result in F^i can be obtained by

$$S^i = \{(x,y)|(x,y)\subset F^i, X^i(x,y) < T\}. \tag{19}$$

Step 7: To verify the shadow. Shadows gotten in each sub-region may not be real ones due to false detection. Denoting S as the shadow region in F, S is initialized by the first S^i gotten by formula (19)

$$S = \begin{cases} S\cup S^i & \text{if } \overline{F^{NS^i}_{[RGB]}} - \overline{F^{S^i}_{[RGB]}} \subset [k_1\cdot Lk_2\cdot L] \\ S & \text{if } \overline{F^{NS^i}_{[RGB]}} - \overline{F^{S^i}_{[RGB]}} \not\subset [k_1\cdot Lk_2\cdot L] \end{cases} \tag{20}$$

where $L = [m\cdot\overline{F_R^{NS^i}}/\overline{F_B^{NS^i}}\,n\cdot\overline{F_G^{NS^i}}/\overline{F_B^{NS^i}}\,1]\cdot\Delta B$, the coefficients k_1 and k_2 are empirically set to 0.8 and 1.2, respectively. The two coefficients are adopted because m and n are not accurate values.

Step 8: To obtain accurate boundaries of shadows. Shadows detected by previous steps are based on the subtractive image X. However, subtractive operation blurred image X because of high correlation among R, G, and B components. This may cause inaccurate boundaries of shadows. The blurred information of X can be regained from the original image. Therefore, another constraint is imposed: the shadow regions are often darker than the mean values of the original image in each channel. The final result of detected shadows is denoted as (21). $F^i_{[RGB]}(x,y)$ denotes the tricolor vector at location (x,y) in i th region of original image F

$$\text{shadow} = \left\{(x,y)|S(x,y)\cap F^i_{[RGB]}(x,y) < \left[\overline{F_R^i}\ \overline{F_G^i}\ \overline{F_B^i}\right]\right\}. \tag{21}$$

V. Experimental Results

The multistep algorithm based on the TAM is evaluated with different images including four simple images, one aerial image and three complex images. Fig. 4 shows the shadow detection in simple images by our method and by others as comparisons. Image (a) has a shadow of man half on the grass and half on the road. It is detected more accurately by our method [shown in image (b)] than by the method in [26] [shown in image (c)]. Image (d) has a shadow of a bird standing on the ground. Though

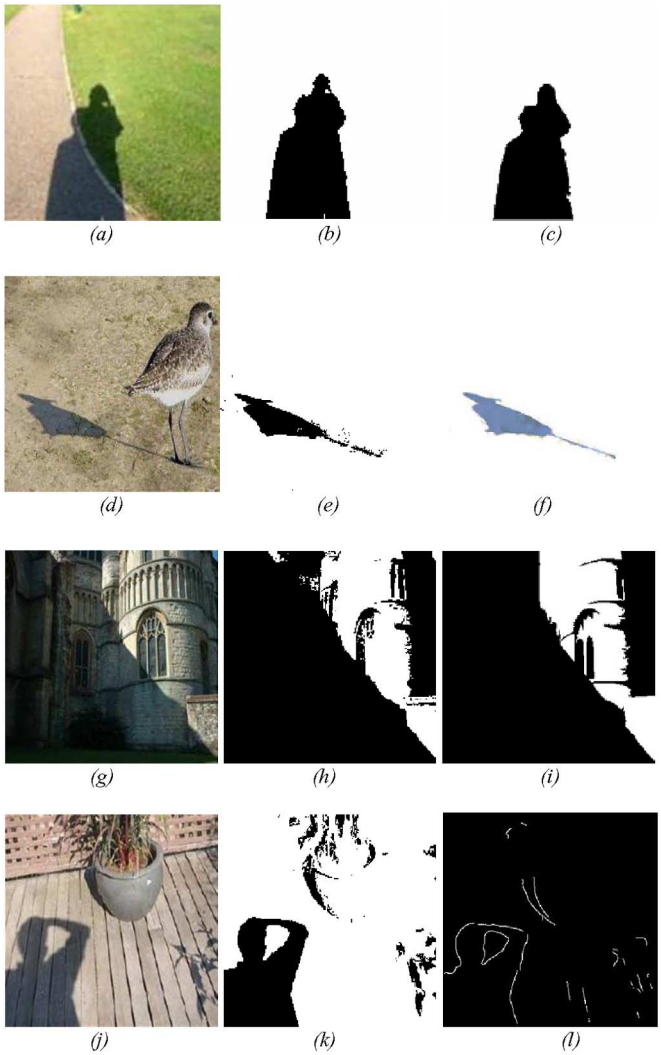


Fig. 4. Results of simple scenes. Images (a), (d), (g), (j): original images; images (b), (e), (h), (k): the output of our algorithm; images (c), (f), (i), (l): the results given in [26], [22], [19], and [33], respectively. Images (c), (f), (i) are derived from the corresponding articles and image (l) is derived from the author's website.

the method in [22] processes it perfectly [shown in image (f)], it requires users to identify the rough shadow and nonshadow regions manually, whereas our method is automatic. Image (e) is the output of our method. Both our method and the method in [19] obtain similar results on detecting a big slanting cast shadow on the cathedral and a self shadow in the right of image (g), as shown in image (h) and image (i) respectively. However, the method in [19] requires a chromagenic camera to take two pictures of each scene. For the shadows on the floor in image (j), the result of method in [33] is obtained from the author's website, shown in image (l). The shadow edge in image (l) is found by comparing edges in the intrinsic image and the original one. Our method [shown in image (k)] can not only detect the shadow regions, but also detect the shadow edges more accurately. Some details such as shadows of leaves and potted plant are detected by ours, but not detected by the method in [33] as shown in image (l).

Our algorithm is not designed for specific applications. In addition to four simple images, an aerial image and three images with complex content are chosen as the inputs. In Fig. 5, the aerial image processed by our algorithm (image(b)) shows smooth shadow regions, while image (c) given in [8] has a little noisy. The shadow on the side of the left building is detached into pieces by the method in [8], which is an whole part in ours. The result given in [6] is better than ours: the two long, slender shadows, which are missed in ours, are detected by method in [6]. But their method is designed for aerial images, and is time-consuming because it is a two-level hierarchical algorithm with two optimizations.

As shown in Fig. 6, our algorithm is used for dealing with three images obtained from real complex scenes. Image (a) is a forest picture that was taken from 100 meter high, with a big cast shadow in the middle and a small one in the bottom. Both of them can be detected out, as shown in Image (b), and are free of affection of the complicate texture in the background. Image (c) is a scene that contains shadows of a man and of a tree. Image (d) shows the extracted shadows. Though there is a little noise on the grass, the shadows on the ground are detected out from the complex image contents. Because the shadow on the upper portion of the trunk is encompassed by leaves and is lighted by only part of skylight, which violates the light environments of our model (shadow lighted by clear skylight) and further causes the changes of parameter values of m and n, the shadow is missed in our detecting results. To be exact, the shadow region is removed by step 7 in our method. Image (e) has shadows on the road and on the soil. As shown in Image (f), the weak penumbra shadows on the road, the self shadow on the right side of the road, and the shadows on the soil are all detected, except that there are some pixels misclassified as shadows.

Because shadow detection usually is a preprocessing step of practical applications, mean values are simply chosen as the thresholds for fast computing in step 6 and step 8. Experiments show that our algorithm with these simple thresholds works well in most cases. However, it unavoidably causes problems in parts of some images. In our experiments, in Fig. 6(c), the shadow on the pavement behind the tree, and in Fig. 5(a), the shadow on the upper left corner of the roof beneath the tall tree are not detected by our algorithm. These slight shadows are not clearly darker than mean values of their sub regions, so they are omitted by step 6. If iterative thresholds method like OTSU are employed, better results can be expected.

VI. Discussion

Generally, shadows occur in multilight sources environments (in which, at least, a point light source). For example, in outdoor scenes, the light sources are daylight and skylight; in indoor scenes, the light sources are lamplight and the reflected light. When light sources change, the proposed method can easily be extended to new light environments. What we need to do is just using the CCTs of the new light sources to reestimate m and n. Many methods assume that the outdoor light source is a white one. It is not correct. The white light emitted from the sun is generally composed of a mixture of energy at different wavelengths. When passing through the earth's atmosphere, the atmospheric effects, such as reflection, scattering, and absorption will largely

Fig. 5. Result of aerial image. (a) original image; (b) result of our method; (c) result given in [8]; (d) result given in [6].

change the spectral of the white light of solar irradiance. The spectral of the light that ultimately reaches the earth's surface changes, i.e., the light is not a white one anymore. The more reasonable approach adopted in this paper is to use blackbody irradiance to approximate the SPDs of daylight and skylight. Sky plays an important role in shadow generation. Although some researchers like [34], [35] have noticed that shadow has relation to the blue sky: they simply assume shadows are bluer than nonshadows, which is not totally true. The pixel values of a shadow also have relation to surface reflection and camera sensor responsibility, but the product of that can be estimated from pixels of nonshadow backgrounds [see (8)].

A. Approximations on TAM

For deducing the tricolor attenuation model, we use the three approximations.

1) Using blackbody irradiance to approximate the daylight and sunlight. Some experiments as in [36] have validated the reasonability of the approximation.
2) Using impulse function to approximate the camera response properties. However, real cameras may not have infinitely narrow sensors. We find that our algorithm still holds when the assumption is violated in practical applications. This is because the violation of infinitely narrow assumption may not change the maximum and minimum in $[\Delta R\ \Delta G\ \Delta B]$. Even if it changes them, definitely does not change them on the whole image due to our algorithm is working on the segmented small regions. Furthermore, if the assumption does have effect when used in wide-band cameras, the caused error can be solved by using the spectral-sharpening technique [37] to narrow the response curve of the camera's sensor.
3) Ignoring the nonlinear postprocessing functions. In fact, this nonlinear processing is common (like JPEG, BMP format image). If our method is used on RAW data, the results are expected to be better.

B. Future Work

As shown in Fig. 7, assuming [r g b] is the shadow we have detected, and [R G B] is the corresponding nonshadow background. Assuming $\Delta B \neq 0$, formula (1) always holds

$$(R-r)+(G-g)=(B-b)\cdot\frac{(R-r)+(G-g)}{(B-b)}$$
$$\Rightarrow (R-r)+(G-g)=(B-b)\cdot\left(\frac{\Delta R}{\Delta B}+\frac{\Delta G}{\Delta B}\right). \quad (1)$$

So

$$R+G-\left(\frac{\Delta R}{\Delta B}+\frac{\Delta G}{\Delta B}\right)\cdot B=r+g-\left(\frac{\Delta R}{\Delta B}+\frac{\Delta G}{\Delta B}\right)\cdot b. \quad (2)$$

Denoting T operator as $T = [1\ 1 - ((\Delta R/\Delta B) + (\Delta G)/(\Delta B))]$, (2) can be rewritten as $T[R\,G\,B] = T[r\,g\,b]$. We find that after T operation, the shadow pixels and the nonshadow pixels are strictly equal. Apparently, we obtain a shadow invariant image. Two results are given in Fig. 8. Shadow invariant is useful for shadow detection (see step 1, in Section IV), and shadow detection is useful to get an invariant image (T operator). Our future research might focus on using shadow detection and shadow invariant together to develop an iterative algorithm which will output two results: one is shadow image and another is shadow invariant image.

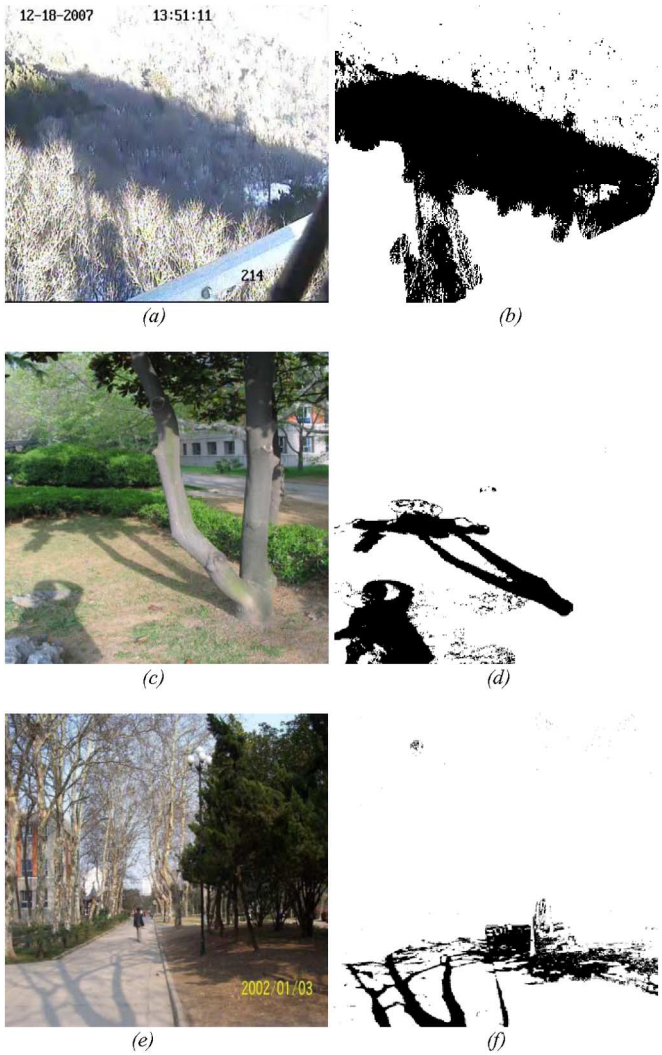


Fig. 6. Results of complex scenes. Images (a)–(c): original images; images (d)–(f): detected shadows.

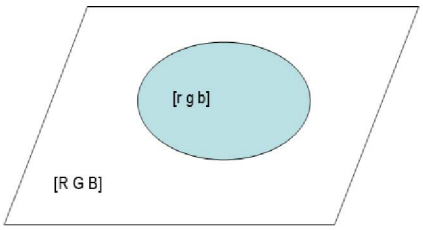


Fig. 7. Shadow on background, in which [r g b] is shadow region; [R G B] is the corresponding nonshadow background.

VII. Conclusion

Shadow identification in single image is difficult but has wide applications. In the paper, we use image formation theory to deduce the tricolor attenuation model, and employ blackbody irradiance to estimate its parameters. Based on the new model, we present an algorithm to detect shadows. Unlike most previous methods are suitable for image sequences, our method can extract shadows from only a single image, even for image with complex outdoor scenes. Definitely, the method proposed here also can be used in video sequence (in each frame).

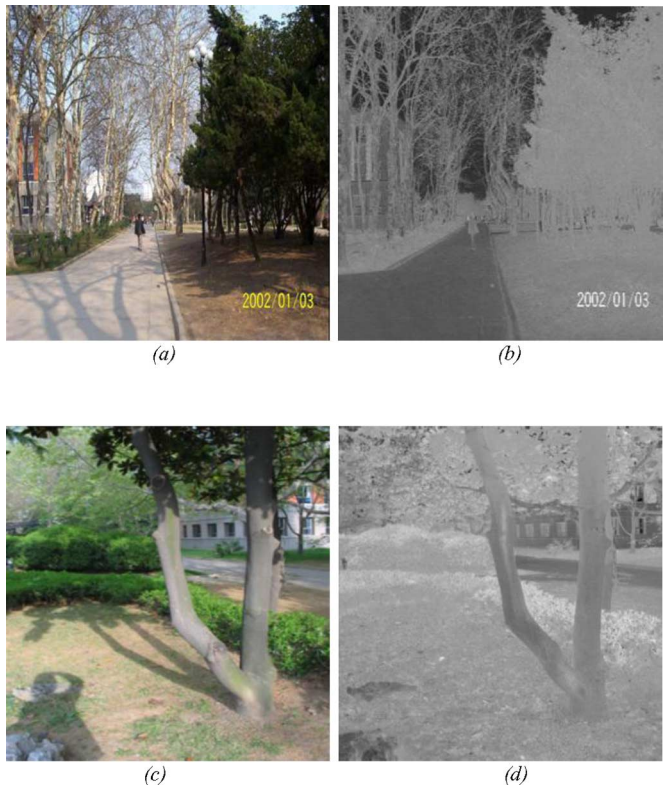


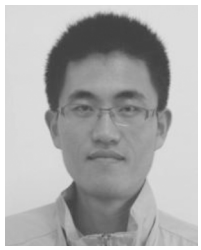
Fig. 8. Using T operator to obtain a shadowless image. (a), (c) original image; (b), (d): shadow invariant image after T operator.

In the theoretical analysis, for simplicity, we use the three above-mentioned approximations (Section VI-A). However, the shadow detection algorithm needs not keep those preconditions. Furthermore, the algorithm proposed in this paper is not designed for specific applictions. As shown in experimental section, it can extract both self shadows and cast shadows. Self shadows and cast shadows share common property: lighted by skylight but sunlight is occluded, which is exactly what the TAM used for estimating m and n. The weakness of our algorithm is that it will fail on detecting shadows in sunrise and sunset because in this time the CCTs of sunlight and skylight are very different from the CCTs we adopted in the paper. This will affect the SPDs and further affect m and n. In this situation, the sunlight and skylight should be treated as new light source, i.e., m and n should be reestimated.

References

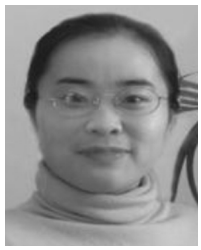
[1] D. C. Knill, D. Kersten, and P. Mamassian, "Geometry of shadows," *J. Opt. Soc. Amer. A*, vol. 14, no. 12, pp. 3216–3232, 1997.
[2] H. Jiang and M. Drew, "Tracking objects with shadows," in *Proc. ICME03: Int. Conf. Multimedia and Expo.*, 2003, pp. 100–105.
[3] Y. Wang, K. F. Loe, and J. K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2006.
[4] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," *Comput. Vis. Pattern Recognit.*, pp. 643–648, 2005.
[5] S. Nadimi and B. Bhanu, "Physical models for moving shadow and object detection in video," *Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1079–1087, Aug. 2004.
[6] J. Yao and Z. F. Zhang, "Hierarchical shadow detection for color aerial images," *Comput. Vis. Image Understand.*, vol. 102, pp. 60–69, 2006.

[7] L. Xu, F. H. Qi, and R. J. Jiang, "Shadow removal from a single image," in *Proc. 6th Int. Conf. Intelligent Systems Design and Applications*, 2006, pp. 1049–1054.
[8] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Shadow identification and classification using invariant color models," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2001, pp. 1545–1548.
[9] G. Finlayson, S. Hordley, G. Schaefe, and G. Y. Tian, "Illuminant and device invariant colour using histogram equalization," *Pattern Recognit.*, vol. 38, pp. 179–190, 2005.
[10] A. Leone and C. Distante, "Shadow detection for moving objects based on texture analysis," *Pattern Recognit.*, vol. 40, pp. 1222–1233, 2007.
[11] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Comput. Vis. Image Understand.*, vol. 95, no. 2, pp. 238–259, 2004.
[12] K. Barnard and G. Finlayson, "Shadow identification using colour ratios," in *Proc. IS&T/SID 8th Color Imaging Conf. Color Science, Systems and Applications*, 2000, pp. 97–101.
[13] R. Ramamoorthi, D. Mahajan, and P. Belhumeur, "A first-order analysis of lighting, shading, and shadows," *ACM Trans. Graph.*, vol. 26, no. 1, pp. 1–21, Jan. 2007, Article 2.
[14] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2003, pp. 1343–1350.
[15] G. Funka-Lea and R. Bajcsy, "Combining color and geometry for the active, visual recognition of shadows," in *Proc. ICCV*, 1995, pp. 203–209.
[16] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithm and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 918–923, Mar. 2003.
[17] F. Porikli and J. Thornton, "Shadow flow: A recursive method to learn moving cast shadows," in *Proc. ICCV*, 2005, pp. 891–898.
[18] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 65–76, Mar. 1999.
[19] G. Finlayson, C. Fredembach, and M. S. Drew, "Detecting illumination in images," in *Proc. ICCV*, 2007, pp. 1–8.
[20] G. Finlayson, S. Hordley, and M. S. Drew, "Removing shadows from images," in *Proc. ECCV*, 2002, pp. 823–836.
[21] A. Prati, R. Cucchiara, I. Mikic, and M. M. Trivedi, "Analysis and detection of shadows in video streams: A comparative evaluation," *Comput. Vis. Pattern Recognit.*, 2001.
[22] T. P. Wu and C. K. Tang, "A Bayesian approach for shadow extraction from a single image," in *Proc. ICCV*, 2005, pp. 480–487.
[23] T. P. Wu, C. K. Tang, M. S. Brown, and H. Y. Shum, "Natural shadow matting," *ACM Trans. Graph*, vol. 26, no. 2, pp. 8:1–8:21, June 2007.
[24] M. Nielsen and C. B. Madsen, "egmentation of soft shadows based on a daylight- and penumbra model," in *Proc. Mirage*, 2007, pp. 341–352.
[25] Z. Du, X. Qin, W. Hua, and H. Bao, "Shadow removal in sole outdoor image," in *Proc. PCM*, 2006, pp. 999–1007.
[26] C. Lu and M. S. Drew, "A Markov random field framework for finding shadows in a single colour image," in *Proc. 10th Congr. Int. Colour Association*, Granada, Spain, May 8–13, 2005.
[27] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting objects, shadows and ghosts in video streams by exploiting color and motion information," in *Proc. 11th Int. Conf. Image Analysis and Processing*, 2001, pp. 360–365.
[28] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognit.*, no. 32, pp. 453–464, 1999.
[29] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 1982.
[30] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 34–43, Jan. 2005.
[31] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for Retinex," *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, 2003.
[32] G. Finlayson, M. Drew, and C. Lu, "Intrinsic images by entropy minimization," in *Proc. Eur. Conf. Computer Vision*, 2004, vol. 3, pp. 582–595.
[33] H. D. Cheng, X. H. Jiang, Y. Sun, and J. L. Wang, "Color image segmentation: Advances and prospects," *Pattern Recognit.*, vol. 34, pp. 2259–2281, 2001.
[34] S. Nadimi and B. Bhanu, "Moving shadow detection using a physics-based approach," in *Proc. 16th IEEE Int. Conf. Pattern Recognition*, Quebec City, QC, Canada, Aug. 2002, pp. 701–704.
[35] A. M. Polodorio, F. C. Flores, N. N. Imai, A. M. G. Tommaselli, and C. Franco, "Automatic shadow segmentation in aerial color images," in *Proc. IEEE 16th Brazilian Symp. Computer Graphics and Image Processing*, 2003, pp. 270–277.
[36] J. A. Marchant and C. M. Onyango, "Shadow-invariant classification for scenes illuminated by daylight," *J. Opt. Soc. Amer. A*, vol. 17, pp. 1952–1961, 2000.
[37] M. S. Drew and G. D. Finlayson, "Spectral sharpening with positivity," *J. Opt. Soc. Amer. A*, vol. 17, pp. 1361–1370, 2000.



Jiandong Tian received the B.S. Tech. degree in automation from HeiLongjiang University, China, in 2005. He is currently pursuing the Ph.D. degree in pattern recognition and intelligent systems at the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences.

His research interests include image processing, pattern recognition, and robotic navigation.



Jing Sun graduated from Northeast University, China, in 2003, with a major in English and the M.S. degree in computers in 2006. She is currently pursuing the Ph.D. degree in pattern recognition and intelligent systems at the State Key laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences.

Her research interests include computer vision and image processing.

Yandong Tang received the B.S. and M.S. degrees in mathematics from Shandong University, China, in 1984 and 1987, respectively. He received the Ph.D. degree in applied mathematics from the University of Bremen, Germany, in 2002. He is a Professor at the Shenyang Institute of Automation, Chinese Academy of Sciences.

His research interests include numerical computation, image processing, and computer vision.