Cast shadow segmentation using invariant color features

Elena Salvador,[a] Andrea Cavallaro,[b,*] and Touradj Ebrahimi[a]

[a] *Signal Processing Institute, Swiss Federal Institute of Technology, Lausanne, Switzerland*
[b] *Multimedia and Vision Laboratory, Queen Mary, University of London, London, UK*

Abstract

Shadows are integral parts of natural scenes and one of the elements contributing to naturalness of synthetic scenes. In many image analysis and interpretation applications, shadows interfere with fundamental tasks such as object extraction and description. For this reason, shadow segmentation is an important step in image analysis. In this paper, we propose a new cast shadow segmentation algorithm for both still and moving images. The proposed technique exploits spectral and geometrical properties of shadows in a scene to perform this task. The presence of a shadow is first hypothesized with an initial and simple evidence based on the fact that shadows darken the surface which they are cast upon. The validity of detected regions as shadows is further verified by making use of more complex hypotheses on color invariance and geometric properties of shadows. Finally, an information integration stage confirms or rejects the initial hypothesis for every detected region. Simulation results show that the proposed algorithm is robust and efficient in detecting shadows for a large class of scenes.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Color image analysis; Segmentation; Shadow detection; Video object extraction

1. Introduction

Shadows provide relevant information about the scene represented in an image or a video sequence. They contain cues about the shape and the relative position of

---

* Corresponding author.
E-mail addresses: elena.salvador@epfl.ch (E. Salvador), andrea.cavallaro@elec.qmul.ac.uk (A. Cavallaro), touradj.ebrahimi@epfl.ch (T. Ebrahimi).

objects, as well as about the characteristics of surfaces and light sources. Despite this, in applications requiring the identification of objects, shadows modify the perceived shape and color, thus introducing a distortion in the object detection process. For this reason, the problem of shadow detection has been increasingly addressed over the past years.

Shadow detection techniques can be classified into two groups: model-based, and property-based techniques. *Model-based* techniques rely on models representing the a priori knowledge of the geometry of the scene, the objects, and the illumination. *Property-based* techniques identify shadows by using features such as geometry, brightness or color of shadows.

Model-based techniques are designed for specific applications, such as aerial image understanding [1–5] and video surveillance [6–8]. They are based on matching sets of geometric features such as edges, lines or corners to 3D object models. Model-based schemes generally handle simple objects and are only applicable to the specific application they are designed for.

The above-mentioned limitations are overcome by using spectral and geometric features of shadows in property-based techniques. Luminance information alone is exploited in early techniques by analyzing edges [9,10], histograms [11] or texture information [12]. Edge and texture information is exploited more recently in [13] to detect background regions which are covered or uncovered by a moving cast shadow.[1] Static edges and uniform changes of shading in the background texture are searched for over time. An additional cue is given by the penumbra of shadows. The penumbra of a cast shadow is the part of the shadow where direct light from an extended light source is only partially occluded. The penumbra in outdoor scenes, when shadows present sharp edges due to the illumination source that is far from the objects, could be difficult to detect.

Luminance, chrominance, and gradient density information is used in [14]. A combined shadow confidence score is derived for extracted foreground regions in video sequences that allows for separation of a cast shadow from the corresponding object. This method is based on the hypothesis that shadow areas are not textured. This assumption is valid for applications such as traffic surveillance when shadows are cast on the road. Moreover, the method makes use of the convex hull of objects which is appropriate for vehicles, but less appropriate for non-rigid objects such as pedestrians. Color information is used also in [15], where the Dichromatic Reflection Model is exploited for learning background color during an off-line training phase and separating shadowed background from foreground regions in outdoor image sequences.

A physics-based approach to distinguish material changes from shadow boundaries in chromatic still images is presented in [16]. This method takes into account the effect of ambient illumination which is usually neglected by other approaches. An a priori estimate for the strength of the ambient illumination has to be provided to the method. Color ratios are used in [17] for the same purpose. Here, an estimate of the illumination field from the integration of the edge information is also derived.

---

[1] The definition of cast shadows will be given in Section 2.

In [18], a camera calibration procedure requiring a sequence of images is used to generate a 1-D illumination invariant shadow-free image. This image is used together with the original image to locate shadow edges and to reconstruct a shadow-free full color image. A classification of color edges by means of photometric invariant features into shadow-geometry edges, highlight edges, and material changes is proposed in [19].

A method for real-time cast shadow detection for video conference applications is proposed in [20]. The algorithm uses color information in the YUV color space in order to avoid time consuming color transformations. The authors observed that a shadow reduces the YUV values of a point linearly with respect to the same point in light. This observation is used in the method to detect shadows and remove them from segmented objects.

Intensity, hue and saturation are exploited in [21] to detect moving cast shadows. The detection is based on the observation that shadows change significantly the brightness of an area without significantly modifying the color information. On the basis of the same observation, a statistical background subtraction algorithm which exploits a computational color model that separates the brightness from the chromaticity components of a pixel is presented in [22]. In [23], the change in appearance of a pixel when shadowed and when illuminated is statistically modeled. The parameters of the method, which exploits a diagonal model of illumination change, require a time-consuming setting and are optimized for traffic monitoring.

Geometric properties of shadows are generally less exploited for shadow segmentation than spectral properties, but they can provide valuable information. In [24], both intensity and geometry constraints are used to detect and to classify shadows in images of a constrained, simple environment. In [25], geometry information is combined with color information to detect cast shadows. A limitation of the method is the active process that is required to determine the location of the light source. The approach proposed in [26] overcomes this limitation by presenting a method for the estimation of the projection of the light source direction in the 2D image plane. The estimated direction is used to guide the cast shadow detection process. However, the method requires the manual segmentation of the shadow in the first frame of the sequence and segmented shadow-casting objects without shadows. The estimation of the 3D illuminant direction in a scene from shading and shadow information has a long history and dates back to Pentland's original work [27]. Classical methods for illumination direction estimation make strong assumptions on the scene making them not applicable to complex real world images. Recent methods exploit brightness distribution in shadows cast by objects of known 3D shapes [28–30] or insert in the scene objects of known shape, reflectance, and position [31,32] to compute an illumination distribution in the scene. These methods mainly address augmented reality applications. Although they achieve good results in real images with complex illumination distributions, the need for specific equipment, multiple images or knowledge of the 3D shape, pose and location of objects prevents them to be used in other applications where no control on the scene is available.

In this paper, we propose an image understanding system for the detection of cast shadows which exploits shadows' spectral and geometric properties. The analysis is

organized in three stages. A hypothesis about the presence of a shadow is first generated on the basis of an initial and simple evidence, i.e., shadows normally darken the surface upon which they are cast. The validity of this hypothesis is further verified on each detected region by making use of hypotheses on color invariance and geometric properties of shadows. Finally, an information integration stage confirms or rejects the initial hypothesis for every detected region. The system is conceived to work with uncalibrated images from different sources and with objects of different nature.

We demonstrate that the proposed methodology can be applied to both video sequences and still images. In the case of video, the proposed algorithm's performance is compared to [21–23], which also exploit shadows' spectral properties but use different color models than those used here, and additionally to [13], which represents, to the best of our knowledge, one of the most complete methods in the literature. In the case of still images, the method extends and simplifies that in [24], which also exploits shadows' geometry for shadow extraction in single uncalibrated images.

An accurate characterization of shadows in terms of spectral and geometric properties and a study of the solutions proposed in the literature to the problem of shadow segmentation according to this characterization allows here to propose an analysis method for detecting shadows in both still and moving images which improves on the existing techniques.

The paper is organized as follows. In Section 2, the physical generation of shadows in a scene is described. The characterization of shadows in terms of spectral and geometrical properties leads to the definition of explicit criteria for their identification, as discussed in Section 3. Experimental results are presented in Section 4, and in Section 5 we draw the conclusions.

2. What is a shadow?

A shadow occurs when an object partially or totally occludes direct light from a source of illumination. Shadows can be divided into two classes: self and cast shadows. A *self shadow* occurs in the portion of an object which is not illuminated by direct light. A *cast shadow* is the area projected by the object in the direction of direct light. In the following, the relationship between shadows and lit regions is formalized in order to derive relevant shadow properties.

2.1. Spectral properties of shadows

To describe the spectral appearance of a surface in shadow, let us consider the physics of color generation. The appearance of a surface is the result of the interaction among illumination, surface reflectance properties, and the responses of a chromatic mechanism. This chromatic mechanism is composed of three color filters in a color camera.

To model the physical interaction between illumination and object's surface, let us consider the Dichromatic Reflection Model [33]. The radiance [34] of the light,

$L_r(\lambda, \vec{p})$, reflected at a given point $\vec{p}$ on a surface in the 3D space, given some illumination and viewing geometry, is formulated as

$$L_r(\lambda, \vec{p}) = L_a(\lambda) + L_b(\lambda, \vec{p}) + L_s(\lambda, \vec{p}), \qquad (1)$$

where $L_a(\lambda)$, $L_b(\lambda, \vec{p})$, $L_s(\lambda, \vec{p})$ are the ambient reflection term, the body reflection term, and the surface reflection term, respectively, and λ is the wavelength. The ambient illumination term is assumed to account for all the light indirectly reflected among surfaces in the environment and does not vary with geometry. If there is no direct illumination because an object is obstructing the direct light, then the radiance of the reflected light is

$$L_{r_{shadow}}(\lambda, \vec{p}) = L_a(\lambda), \qquad (2)$$

which represents the intensity of the reflected light at a point in a shadow region.

Let $S_R(\lambda)$, $S_G(\lambda)$, and $S_B(\lambda)$ be the spectral sensitivities of the red, green, and blue sensors of a color camera, respectively. The color components of the reflected intensity reaching the sensors at a point (x, y) in the 2D image plane are

$$C_i(x, y) = \int_\Lambda E(\lambda, x, y) S_{C_i}(\lambda) \, d\lambda, \qquad (3)$$

where $C_i \in \{R, G, B\}$ are the sensor responses, $E(\lambda, x, y)$ is the image irradiance [34] at (x, y), and $S_{C_i}(\lambda) \in \{S_R(\lambda), S_G(\lambda), S_B(\lambda)\}$. The interval of summation is determined by $S_{C_i}(\lambda)$, which is non-zero over a bounded interval of wavelengths Λ. Since image irradiance is proportional to scene radiance [34], for a pixel position (x, y) representing a point $\vec{p}$ in direct light, the sensor measurements are

$$C_i(x, y)_{lit} = \int_\Lambda \alpha \Big(L_a(\lambda) + L_b(\lambda, \vec{p}) + L_s(\lambda, \vec{p}) \Big) S_{C_i}(\lambda) \, d\lambda \qquad (4)$$

giving a color vector $\vec{C}(x, y)_{lit} = (R_{lit}, G_{lit}, B_{lit})$. α is the proportionality factor between radiance and irradiance. For a point in shadow the measurements are

$$C_i(x, y)_{shadow} = \int_\Lambda \alpha L_a(\lambda) S_{C_i}(\lambda) \, d\lambda \qquad (5)$$

giving a color vector $\vec{C}(x, y)_{shadow} = (R_{shadow}, G_{shadow}, B_{shadow})$. It follows that each of the three RGB color components, if positive and not zero, decrease when passing from a lit region to a shadowed one, that is

$$\begin{aligned} R_{shadow} &< R_{lit}, \\ G_{shadow} &< G_{lit}, \\ B_{shadow} &< B_{lit}. \end{aligned} \qquad (6)$$

Ambient light can have different spectral characteristics with respect to direct light [16]. The case of outdoor scenes, where the diffuse light from the sky differs in spectral composition with respect to the direct light from the sun, provides an example [15,18,35]. Since we aim in this work at avoiding calibration procedures and camera-dependent computations as in [18,35], so as to propose a segmentation algorithm that can be applied even when no control on the imaging conditions and the scene is

possible, we assume that the ambient light is proportional to the direct occluded light. In this case, the assumption is a commonly used one [14,20–22,26]. Another situation is when inter-reflections among different surfaces in the scene are present. Local effects due to inter-object reflection are extremely hard to analyze [34] in real world complex scenes as those addressed in this paper.[2] Experiments will however show that the proposed method is robust to the above mentioned assumption even in scenes where the effect of mutual illumination could be non-negligible.

A second spectral property of shadows can be derived in the commented hypothesis by considering photometric color invariants. Photometric color invariants are functions which describe the color configuration of each image point discounting shading, shadows, and highlights. These functions are demonstrated to be invariant to a change in the imaging conditions, such as viewing direction, object's surface orientation and illumination conditions [38]. Let us define F as one of the above mentioned photometric color invariants. F_l is the value assumed in a point in light, and F_s is the value in the same point in shadow. Then,

$$F_l = F_s. \tag{7}$$

Examples of photometric color invariants are normalized rgb, hue (H), saturation (S), $c_1 c_2 c_3$ and $l_1 l_2 l_3$ [38]. In particular, among the different photometric invariant color features, as stated in [39], we adopted the $c_1 c_2 c_3$ model. The $c_1 c_2 c_3$ invariant color features are defined as follows:

$$c_1(x,y) = \arctan \frac{R(x,y)}{\max(G(x,y), B(x,y))}, \tag{8}$$

$$c_2(x,y) = \arctan \frac{G(x,y)}{\max(R(x,y), B(x,y))}, \tag{9}$$

$$c_3(x,y) = \arctan \frac{B(x,y)}{\max(R(x,y), G(x,y))}, \tag{10}$$

for $R(x,y)$, $G(x,y)$, and $B(x,y)$ representing the red, green, and blue color components of a pixel in the image.

It is known from Kender [40] that normalized color rgb is unstable near the black vertex of the RGB space, where it is undefined, while hue is unstable near its singularities at the entire achromatic axis. The analysis of the noise propagation through the invariant transformations in Eqs. (8)–(10) is addressed in detail in [19]. The instability of photometric invariant transformations will be taken into account by the algorithm proposed in Section 3.

2.2. Geometrical properties of shadows

The geometric appearance of a shadow depends on objects and scene layout. However, it is possible to identify some geometrical characteristics of shadows, the

---

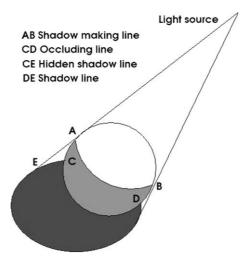[2] A detailed analysis of mutual illumination for simple scene geometries can be found in [36,37].

Fig. 1. Shadow lines definition.

shadow boundaries, without any knowledge of the structure of the object or of the scene. Shadow boundaries can be classified into four classes: shadow-making lines, shadow lines, occluding lines, and hidden shadow lines. These lines are depicted in Fig. 1. *Shadow-making lines*, *AB*, separate the illuminated surface and the non-illuminated surface of an object. They appear to be the outlines of an object if the position of the observer is aligned with the direction of the light source. The projections of the shadow-making lines in the direction of light rays are called *shadow lines*, *DE*. *Occluding lines*, *CD*, separate an object from its cast shadow. A *hidden shadow line*, *CE*, is a shadow line corresponding to a non-visible shadow-making line.

3. Proposed method

We exploit the spectral and geometrical properties described in Section 2 to automatically recognize shadows in both video sequences and still images. A bottom-up analysis organized in three levels is performed. This hierarchical control structure uses the hypothesize-and-test scheme. The presence of a shadow is first hypothesized based on some initial evidence. The hypothesized shadow region is then verified by checking its consistency with other additional hypotheses. Finally, an information integration confirms or rejects the initial hypothesis. A detailed description of the three levels is given in the following sections.

3.1. Hypothesis generation

The first level of the proposed strategy makes use of the property that shadows darken the surface upon which they are cast. This results in the identification of

potential shadows. Let us refer to the image under analysis[3] with $\mathbf{I}(x, y) = (R(x, y), G(x, y), B(x, y))$, where R, G, B represent the three color channels and (x, y) indicates a generic pixel position. The intensity of each pixel $\mathbf{I}(x, y)$ is compared to the intensity of a *reference pixel*, $\mathbf{I}(x_r, y_r)$. The *reference pixel* (x_r, y_r) is defined differently for still images and video, and it is described in the following.

Equation (6) states that each camera sensor must have a lower response for a point in shade. The pixel (x, y) becomes a candidate shadow if its intensity is smaller than that of the *reference pixel* for all three channels. This results in the identification of a set of pixels, $\mathcal{S}_c$, which are candidate to be shadow pixels

$$\mathcal{S}_c = \{(x, y) : R(x_r, y_r) > R(x, y), G(x_r, y_r) > G(x, y), B(x_r, y_r) > B(x, y)\}. \tag{11}$$

In the following, the details of the implementation of the first level of analysis are given for still images and video sequences, respectively.

3.1.1. Still images

For still images, shadow points are identified through the detection of shadow contours. In the first level of analysis, the candidate shadow contour $\mathcal{S}_c'$, with $\mathcal{S}_c' \subset \mathcal{S}_c$, is first extracted. The *reference pixel* (x_r, y_r) is a neighbor of the pixel under analysis defined as $(x_r, y_r) = (x + \delta, y + \gamma)$, with δ and $\gamma \in \{0, 1, -1\}$, where δ and γ are not simultaneously equal to zero, i.e., $(x_r, y_r) \neq (x, y)$. Edges are first extracted from the image, then the property described in Eq. (11) is tested.

The edge map is obtained by applying the Sobel operator [41], separately on the three color channels. The final edge map results from a logical OR-connection operation on the three edge maps corresponding to the three color channels. The Sobel operator has been chosen for its simplicity having verified that it provided satisfactory results in our experiments.

The property described in Eq. (11) is tested by analyzing the gradient image on the edges. A contour point (x, y) becomes a candidate shadow contour point, that is $(x, y) \in \mathcal{S}_c'$, if the gradient has the same orientation in all the three components (Fig. 2B). This is verified by analyzing the coherence of the signs of the horizontal and vertical gradients for the three color channels.

3.1.2. Video

For video sequences, the *reference pixel* (x_r, y_r) belongs to a reference image which represents the background of the scene. The reference image can be either a frame in the sequence or a reconstructed one [42]. The *reference pixel* (x_r, y_r) is at the same location as (x, y) in the image under analysis. The analysis is performed only in the areas of the image which have been identified as changed by a motion detector [42]. The identified areas correspond to both moving objects and their shadows.

Candidate shadow points are detected by analyzing the image difference $\mathbf{D}(x, y)$ computed as

$$\mathbf{D}(x, y) = \mathbf{I}(x_r, y_r) - \mathbf{I}(x, y). \tag{12}$$

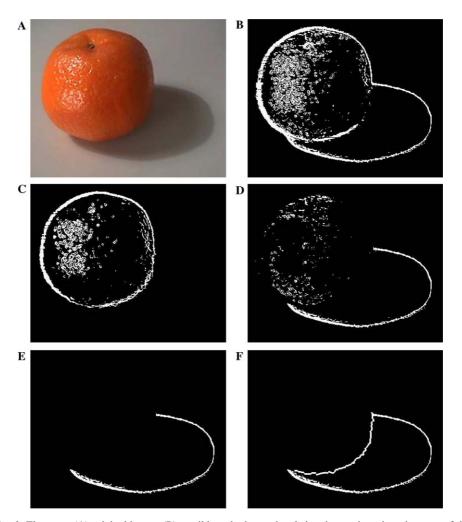[3] This image is the current image of the sequence in case of video.

Fig. 2. First row: (A) original image; (B) candidate shadow points belonging to the color edge map of the RGB image and verifying property in Eq. (11). Second row: (C) color edge map of the invariant features containing material boundaries for which the shadow hypothesis is weakened; (D) integration of the shadow evidence from the spectral analyzes of (B) and (C). Third row: Refinement by means of geometric analysis providing the shadow line and hidden shadow line (E), and complete shadow contours (F).

In a noise-free case, the conditions $R(x_r, y_r) - R(x, y) > 0$, $G(x_r, y_r) - G(x, y) > 0$, $B(x_r, y_r) - B(x, y) > 0$ would suffice to state that the pixel (x, y) belongs to $\mathcal{S}_c$. In real situations, the noise introduced by the acquisition process alters the above test, so that it becomes $R(x_r, y_r) - R(x, y) > b_1$, $G(x_r, y_r) - G(x, y) > b_2$, $B(x_r, y_r) - B(x, y) > b_3$. The vector $\mathbf{b} = (b_1, b_2, b_3)$ takes care of the distortions introduced by the noise. In addition to this, to obtain a more robust result for each pixel position (x, y), we extend the analysis to an observation window, $\mathcal{W}_{C_{(x,y)}}$, of $(2N + 1)$

$(2M + 1)$ pixels, centered in (x, y). In $\mathcal{W}_{C_{(x,y)}}$, we analyze the sum of differences $\mathbf{D}_w(x, y)$, given by

$$\mathbf{D}_w(x, y) = \frac{1}{(2N + 1)(2M + 1)} \sum_{i=-N}^{N} \sum_{j=-M}^{M} \mathbf{D}(x + i, y + j). \tag{13}$$

If each component of $\mathbf{D}_w(x, y)$ is larger than the corresponding component of $\mathbf{b}$, then (x, y) belongs to $\mathcal{S}_c$. The threshold $\mathbf{b}$ is content dependent and should be tuned for each sequence. To avoid the tuning of the threshold we employ a statistical approach. This approach is based on the assumption that the noise in the signal $\mathbf{D}_w(x, y)$ respects a certain distribution. We analyzed the difference $\mathbf{D}_w(x, y)$ in several sequences for indoor and outdoor scenes and we derived that it follows a Gaussian distribution. The statistical approach is based on a significance test [42]. The goal of the statistical test is to check the validity of the hypothesis that a sample $\mathbf{D}_w(x, y)$ comes from a given probability distribution, the Gaussian distribution. The pixel (x, y) is defined as candidate shadow pixel if the significance test is satisfied for all color channels.

3.2. Accumulation of evidence

The result of the first level of analysis is the identification of a set of candidate shadow pixels. This analysis leads to the detection of shadow pixels but also of object pixels. These spurious object pixels are darker than the corresponding reference pixels. A further analysis is required to confirm or to reject this initial hypothesis. Photometric invariant color features and spatial constraints are exploited at this level of the shadow segmentation process to provide additional evidences to the hypothesized shadows.

As we saw in Section 2.1, the presence of a shadow does not alter the value of the invariant color features. Let us define the set of pixels $\mathcal{S}_e$ as

$$\mathcal{S}_e = \{(x, y) : \mathbf{Inv}(x_r, y_r) = \mathbf{Inv}(x, y)\}. \tag{14}$$

The set $\mathcal{S}_e$ is defined by comparing the invariant color features of every pixel with the features of the *reference pixel*. If the value of the invariant color features has not changed with respect to the reference, the hypothesized shadow is strengthened. In the specific implementation of this paper

$$\mathbf{Inv}(x, y) = (c_1(x, y), c_2(x, y), c_3(x, y)), \tag{15}$$

as commented in Section 2.1. A precaution is required when analyzing the photometric invariant features. Due to the considerations discussed in Section 2.1, pixels that fall near the black vertex of the RGB space may give rise to unstable invariant features values in presence of noise. As suggested in [43], we extracted those pixel whose RGB values are below a value of 30 (on a range of 256 levels). We verified in a wide range of test images that these regions did not correspond to cast shadow pixels. It is therefore possible to discard these pixels from our shadow analysis. The works of [21,22] arrive at the same conclusions.

The last evidence about the existence of a shadow is derived from geometrical properties. We want to analyze the position of the shadow with respect to the object. This verification is based on checking the existence of the shadow line, *DE*, and hidden shadow line, *CE* (Fig. 1). These lines separate the shadow pixels from the background pixels, and constitute a necessary condition for the existence of a shadow.

The following sections give the details of the implementation of the two tests for still images and video sequences, respectively.

3.2.1. Still images

According to Eq. (7), contours in the invariant color features will correspond to surface boundaries and not to shadow contours. Color edge detection is therefore performed in the invariant space. A morphological dilation operation is applied on the invariant feature edge map to close the contours. Then isolated pixels are removed so that to obtain the final map, $\mathcal{E}(\mathbf{Inv}(x,y))$ (Fig. 2C). If we define $\mathcal{S}'_e$ as

$$\mathcal{S}'_e = \{(x,y) : \mathcal{E}(\mathbf{Inv}(x,y)) = 0\}, \tag{16}$$

then the shadow hypothesis is strengthened for the set of pixels (Fig. 2D)

$$\mathcal{S}' = \mathcal{S}'_c \cap \mathcal{S}'_e, \tag{17}$$

where object edges belonging to $\mathcal{S}'_c$ have been discarded. The shadow points which form the border between shadowed background and object cannot however be found by means of this analysis. This is clear from Fig. 2D. These points form the occluding line, *CD* (Fig. 1). The occluding line does not indeed belong to $\mathcal{S}'_e$ since it represents a material change. In real images, moreover, $\mathcal{S}'$ contains misclassified pixels due to sensor noise and approximations in the model underlying Eq. (7). Geometrical information is used therefore at this point to reduce the misclassification in $\mathcal{S}'$ and to extract the missing parts of the shadow contour.

Geometrical evidence is verified by checking the existence of the shadow line, *DE*, and the hidden shadow line, *CE*. This is done by extracting segments in $\mathcal{S}'$ and rejecting isolated and disconnected pixels, thus obtaining the subset $\mathcal{S}''$ (Fig. 2E). To this end, isolated groups of pixels are eliminated after connected component analysis. This decision is based on a threshold whose value is set to 30% of the number of pixels of the largest connected component in $\mathcal{S}'$. This value has been determined by means of extensive tests. Since it is relative to the largest component, it is adapted to the image content and does not require content-dependent setting. To extract the missing part of the shadow contour, the definition of occluding line, *CD*, is finally exploited. First, the contact points between shadow contour and object contour $\mathcal{E}(\mathbf{Inv}(x,y))$ are detected. A contour dilation operation is applied to the shadow contour in order to more effectively extract contact points. Then, the position of the shadow with respect to the line that connects the two points is computed and the occluding line is extracted from $\mathcal{E}(\mathbf{Inv}(x,y))$ giving the complete shadow contour (Fig. 2F). Finally, the shadow area is obtained by filling each closed shadow contour.

3.2.2. Video

The identification of the pixels satisfying the first evidence is achieved by analyzing the difference in the invariant feature values, $\mathbf{d}(x, y)$, computed as

$$\mathbf{d}(x, y) = (|c_1(x_r, y_r) - c_1(x, y)|, |c_2(x_r, y_r) - c_2(x, y)|, |c_3(x_r, y_r) - c_3(x, y)|). \quad (18)$$

As for Eq. (12), the test becomes $d_i(x, y) < f_i$ for $i = 1, 2, 3$, where $\mathbf{f} = (f_1, f_2, f_3)$ takes care of the distortions introduced by noise and approximations in the model. As for the first level, we consider a window, $\mathcal{W}_{\mathcal{I}(x,y)}$, centered in (x, y), and we analyze the sum of differences $\mathbf{d}_w(x, y)$, given as in Eq. (13). The setting of the threshold $\mathbf{f}$ is driven by experiments on different sequences. The statistical approach is not used in this step of the algorithm since we observed that the difference of invariant features $\mathbf{d}_w(x, y)$ could not be approximated by a known distribution.

Once the set of pixels $\mathcal{S}_e$ is obtained, the position of shadows with respect to objects is tested (geometric property). In case a hypothesized shadow is fully included in an object, the shadow line is not present, and the shadow hypothesis is then weakened.

3.3. Information integration

Once the additional evidences have been extracted, a decision making step is performed. This final step allows the fusion of the different pieces of information. The result is a rejection of the initial hypothesis in case the rules are not respected. Otherwise the hypothesis is confirmed.

If the analysis of the photometric invariant color features on the candidate shadow is not successful, the pixel is labeled as material change. If the analysis is successful, the candidate shadow undergoes further analysis by means of the geometrical constraints. This final verification is required to eliminate the last ambiguities.

4. Experimental results

4.1. Test set

The performance of the proposed algorithm and a comparison with state-of-the-art methods are presented in this section. Extensive experiments have been carried out on different test images and test sequences. A selection of test images is shown in Fig. 3A. The image size is 288×352 pixels. Among the test sequences, two indoor and one outdoor scenes with different complexity are shown in Fig. 4A, Fig. 5A, and Fig. 6A: the MPEG-4 test sequence *Hall Monitor*, the *art.live*[4] test sequence *Group*,

---

[4] European project IST 10942 *art.live* (Architecture and authoring tools for prototype for Living Images and new Video Experiments), http://www.tele.ucl.ac.be/PROJECTS/art.live/
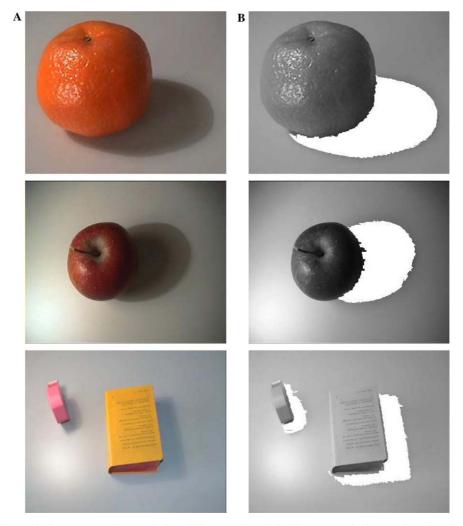
Fig. 3. Shadow segmentation results for still images. (A) Original image; (B) shadow mask (white pixels) superimposed on the original image.

and the MPEG-7 test sequence *Highway*. The image size is 288 × 352 pixels and the temporal resolution is 25 images per second.

4.2. Shadow segmentation results

4.2.1. Still images

Figure 3 shows the results of the proposed algorithm for three test images. The original image (Fig. 3A) and the superimposition of shadow masks on the original image (Fig. 3B) are displayed.

Fig. 4. Shadow segmentation results for the test sequence *Hall Monitor*. (A) Original image; (B) shadow mask (white pixels) superimposed on the original image.

The parameters of the proposed algorithm for still images are the edge detection thresholds (Sections 3.1.1 and 3.2.1), whose values have been determined empirically based on the following reasoning. The threshold value for the invariant features analysis must be large enough to minimize the occurrence of false positives detected due to noise far outside the object contours. The threshold for the RGB color space analysis should be small enough to minimize the occurrence of false negatives and to
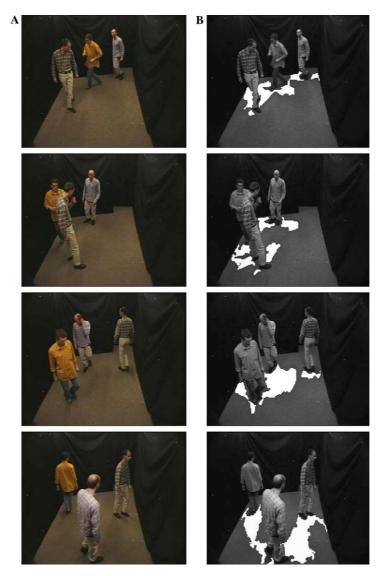
Fig. 5. Shadow segmentation results for the test sequence *Group*. (A) original image; (B) shadow mask (white pixels) superimposed on the original image.

obtain closed contours. In Fig. 3, for the images from top to bottom, the values are respectively 0.06, 0.05, 0.02 for the edge map on the RGB color space; and 0.12, 0.14, 0.14 for the edge map on the invariant features.

Shadows are correctly detected by the proposed algorithm. Smeared edge markings can be observed in the extracted shadows, particularly for Fig. 3, bottom. This type of error is caused by the use of a small threshold for edge detection in the RGB
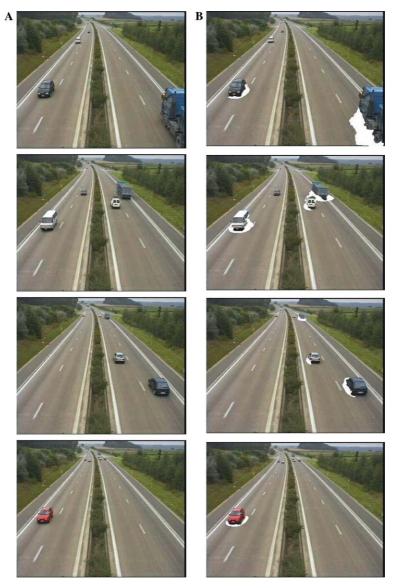
Fig. 6. Shadow segmentation results for the test sequence *Highway*. (A) Original image; (B) shadow mask (white pixels) superimposed on the original image.

space. To overcome this problem, a morphological post-processing depending on the application at hand may be used to improve the final segmentation results.

4.2.2. Video

Figure 4B, Fig. 5B, and Fig. 6B show the results of the proposed shadow segmentation algorithm in video sequences. The reference image is the first frame of the

sequence, acquired before the objects enter in the field of view. The parameters to choose are the size of the observation windows for the generation of the hypothesis, $\mathcal{W}_C$ (Section 3.1.2), and for the first analysis of evidence, $\mathcal{W}_I$ (Section 3.2.2) and the value of the threshold f_i for the photometric invariant color features test (Section 3.2.2). The values of the above-mentioned parameters are the same in all the tests and they are the result of an extensive analysis: $\mathcal{W}_C$ is 5×5 pixels, $\mathcal{W}_I$ is 7×7 pixels, and $f_i = 7$. An analysis of the method's performance with different parameter sets is presented in Section 4.3.

The segmentation results for four sample frames of the test sequence *Hall Monitor* are shown in Figure 4B. The method correctly identifies the shadows which moving objects cast on the floor and on walls. In the second image it is possible to notice that an error occurred: a part of the trousers of the man are detected as shadow region. This is due to the fact that the color of the trousers and the color of the corresponding background region are similar. In addition, the trousers are slightly darker than the background. This portion of the trousers has therefore the same characteristics as a shadow cast on the background. The additional test on geometrical evidence does not succeed in eliminating the candidate shadow, because of the existence of the shadow line. To overcome this problem, our current research direction [44] is to introduce a reliability estimation of the shadow over time. This will allow to discard shadows which do not present time coherence, as in the case of part of background and moving objects presenting similar color characteristics.

A different scenario is depicted in Fig. 5. People walking in a room cast several shadows which are caused by their interaction with multiple light sources. In this scene, a model-based method for shadow recognition would fail due to the complexity of the scene. The proposed method is based on shadow properties and therefore it can be applied to complex scenes, when shadows and objects occlude each other.

Finally, an outdoor scene is depicted in Fig. 6. Vehicles of different dimensions are running on a highway. Objects are smaller and lighting conditions are different when compared to the previous indoor sequences. These results demonstrate that the proposed method can be applied on a large class of scenes, without changing the values of the parameters.

4.3. Objective performance evaluation

To quantitatively analyze the performance of the method with different parameter sets, a ground-truth segmentation should be obtained. However, the generation of a ground-truth for shadow regions in real world scenes is a difficult task. In fact, the outer boundary of a shadow occurs at points of infinitesimal decrease in the amount of illumination. As a result, the exact boundary of a shadow cannot be manually determined in a reliable way. As a solution, a more significant analysis can be obtained by combining a shadow detection method with an object extraction method and by evaluating the object detection accuracy. To this end, the proposed shadow segmentation is therefore combined with the video object extraction method described in [42]. This combination will also demonstrate an important application of the shadow segmentation method (Fig. 7).

Fig. 7. Video object extraction for the test sequence *Hall Monitor* without (left) and with (right) shadow detection.

The ground-truth object segmentation is obtained manually. The comparison is based on computing the pixel-wise deviation of a segmentation result from the corresponding ground-truth segmentation. The deviation is computed by taking into account two types of errors, namely false positives and false negatives. False positives, ϵ_p, are pixels incorrectly detected as belonging to an object. False negatives, ϵ_n, are pixels belonging to an object that have not been detected. Let $card(C)$ represent the number of pixels detected as object pixels, and $card(C_g)$ the number of pixels belonging to the ground-truth segmentation. The deviation from the reference segmentation can be computed as

$$\epsilon = \begin{cases} 0 & \text{if } card(C) = card(C_g) = 0, \\ \frac{\epsilon_n + \epsilon_p}{card(C) + card(C_g)} & \text{otherwise,} \end{cases} \tag{19}$$

where $\epsilon \in [0,1]$. The accuracy of the segmentation is quantified by $v = 1 - \epsilon$, with $v \in [0,1]$. The larger v, the higher the accuracy. When $v = 1$, then there is a perfect match between segmentation results and ground-truth segmentation.

In Table 1, the false negatives and false positives of object segmentation are reported as percentage of the corresponding area in the ground-truth for different sets of parameters for 300 frames of the test sequence *Hall Monitor*. The obtained results show that the method's performance remains stable for different parameter configurations.

In order to further evaluate the performance of the proposed algorithm, the video object extraction results obtained by combining the method described in [42] with the proposed shadow segmentation have been compared to four state-of-the-art object extraction methods which are robust to shadows [45]. The object detec-

Table 1
System performance with different parameter sets

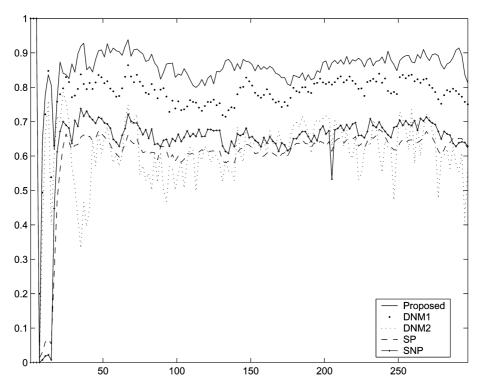| $\mathcal{W}_C$	$\mathcal{W}_I$	f_i	$\%\epsilon_p$	$\%\epsilon_n$	v
3×3	7×7	7	22.22	6.94	0.865
5×5	7×7	7	23.84	5.71	0.865
5×5	5×5	7	21.63	7.59	0.863
5×5	7×7	4	28.55	4.97	0.850
5×5	7×7	5	25.35	5.10	0.862
5×5	7×7	6	23.34	5.81	0.866

Fig. 8. Comparison of video object segmentation accuracy for the test sequence *Hall Monitor*. The large error in the first frames of the sequence is due to the fact that these frames correspond to the entrance of the man in the scene. The first part of the man entering the scene is his shoe. The shoe has a color that is very similar to that of the background. For this reason, the detection algorithms may be misled and do not detect the shoe that is instead present in the ground-truth segmentation.

tion accuracy v (Eq. 19) is computed to perform the comparison. The results for the test sequence *Hall Monitor* are presented in Fig. 8. The symbols in the legend of Fig. 8 refer to the shadow detection techniques used in the object extraction process: DNM1 [21], DNM2 [13], SP [23], and SNP [22]. The mean values of accuracy corresponding to the plots of Fig. 8 are the following: DNM1 0.78; DNM2 0.60; SP 0.59; and SNP 0.63; proposed 0.86. The combination of the proposed shadow recognition method with [42] results in a more accurate object detection over time when compared to state-of-the-art shadow-invariant object detection algorithms.

5. Conclusions

In this paper, we described an efficient method for segmenting cast shadows in both still images and video sequences. The proposed method is based on a

bottom-up analysis approach. An initial hypothesis is tested to identify candidate shadow regions. This initial hypothesis is then verified by exploiting photometric and geometric properties of shadows. The proposed approach was demonstrated through application to a number of test images and video sequences. In addition to this, the proposed method was combined with a video object segmentation algorithm in order to improve its segmentation accuracy. This improvement has been quantified by means of an objective evaluation metric and compared to state-of-the-art algorithms.

The method presented in this paper can be extended as outlined in the following. The proposed segmentation algorithm is designed to detect a particular type of shadows, namely cast shadows. Other types of shadows, such as self shadows, could be detected by adding or replacing appropriate parts of the operators used in this algorithm by others taking into account the specific photometric and geometric properties of such shadows. Moreover, the proposed approach addresses applications that use a monocular camera, such as those used in video surveillance. For applications that employ stereo or multiple cameras, the segmentation algorithm could exploit further hypotheses based on homography and 3D photometric and geometric analyzes. Finally, due to the uncertainty in defining the shadow lines when shadows present a very diffuse penumbra, more sophisticated operators such as that proposed by [46] could be used to extend the proposed method.

Acknowledgment

References

[1] A. Huertas, R. Nevatia, Detecting buildings in aerial images, Comput. Vis. Graph. Image Process. 41 (1988) 131–152.

[2] R. Irvin, D. Mckeown, Methods for exploiting the relationship between buildings and their shadows in aerial imagery, IEEE Trans. Syst., Man, Cybernet. 19 (1989) 1564–1575.

[3] C. Wang, L. Huang, A. Rosenfeld, Detecting clouds and cloud shadows on aerial photographs, Pattern Recogn. Lett. 12 (1991) 55–64.

[4] Y. Liow, T. Pavlidis, Use of shadows for extracting buildings in aerial images, Comput. Vis. Graph. Image Process. 49 (1991) 242–277.

[5] M. Bejanin, A. Huertas, G. Medioni, R. Nevatia, Model validation for change detection, in: 2nd Int. IEEE Workshop on Applications of Computer Vision, 1994, pp. 160–167.

[6] D. Koller, K. Danilidis, H.-H. Nagel, Model-based object tracking in monocular image sequences of road traffic scenes, Int. J. Comput. Vis. 10 (3) (1993) 257–281.

[7] Y. Sonoda, T. Ogata, Separation of moving objects and their shadows, and application to tracking of loci in the monitoring images, in: Proc. IEEE Int. Conf. on Signal Processing, 1998, pp. 1216–1264.

[8] A. Yoneyama, C.H. Yeh, C.-C.J. Kuo, Moving cast shadow elimination for robust vehicle extraction based on 2d joint vehicle/shadow models, in: Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS'03), 2003.

[9] A.P. Witkin, Intensity-based edge classification, in: Proc. Nat. Conf. on Artificial Intelligence, 1982, pp. 46–51.

[10] J.M. Scanlan, D.M. Chabries, R. Christiansen, A shadow detection and removal algorithm for 2-d images, in: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 1990, pp. 2057–2060.

[11] M. Nagao, T. Matsutyama, Y. Ikeda, Region extraction and shape analysis in aerial photographs, Comput. Vis. Graph. Image Process. 10 (1979) 195–223.

[12] M. Adjouadj, Image analysis of shadows, depressions, and upright objects in the interpretation of real world scenes, in: Proc. IEEE Int. Conf. on Pattern Recognition (ICPR), 1986, pp. 834–838.

[13] J. Stauder, R. Melch, J. Ostermann, Detection of moving cast shadows for object segmentation, IEEE Trans. Multimedia 1 (1) (1999) 65–77.

[14] G.S.K. Fung, N.H.C. Yung, G.K.H. Pang, A.H.S. Lai, Effective moving cast shadows detection for monocular color image sequences, in: Proc. 11th Int. Conf. on Image Analysis and Processing (ICIAP), 2001, pp. 404–409.

[15] S. Nadimi, B. Bhanu, Moving shadow detection using a physics-based approach, in: Proc. IEEE Int. Conf. Pattern Recognition, vol. 2, 2002, pp. 701–704.

[16] R. Gershon, A.D. Jepson, J.K. Tsotsos, Ambient illumination and the determination of material changes, J. Opt. Soc. Am. A 3 (10) (1986) 1700–1707.

[17] K. Barnard, G. Finlayson, Shadow identification using colour ratios in: IS&T/SID 8th Color Imaging Conf.: Color Science, Science, Systems and Appl., 2000 pp. 97–101.

[18] G. Finlayson, S. Hordley, M.S. Drew, Removing shadows from images, in: Eur. Conf. Computer Vision, 2002, pp. 823–836.

[19] T. Gevers, H. Stokman, Classifying color edges in video into shadow-geometry, highlight, or material transitions, IEEE Trans. Multimedia 5 (2003) 237–243.

[20] O. Schreer, I. Feldmann, U. Goelz, P. Kauff, Fast and robust shadow detection in videoconference applications, in: Proc. VIPromCom 2002, 4th EURASIP IEEE International Symposium on Video Processing and Multimedia Communications, 2002, pp. 371–375.

[21] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, Detecting objects, shadows and ghosts in video streams by exploiting color and motion information, in: Proc. of 11th Int. Conf. on Image Analysis and Processing (ICIAP), 2001, pp. 360–365.

[22] T. Horprasert, D. Harwood, L.S. Davis, Statistical approach for real-time robust background subtraction and shadow detection, in: Proc. IEEE Int. Conf. on Computer Vision, FRAME-RATE Workshop, 1999.

[23] I. Mikic, P.C. Cosman, G.T. Kogut, M.M. Trivedi, Moving shadow and object detection in traffic scenes, in: Proc. IEEE Int. Conf. on Pattern Recognition (ICPR), 2000, pp. 321–324.

[24] C. Jiang, M.O. Ward, Shadow segmentation and classification in a constrained environment, CVGIP: Image Understanding 59 (2) (1994) 213–225.

[25] G. Funka-Lea, R. Bajcsy, Combining color and geometry for the active, visual recognition of shadows, in: Proc. IEEE Int. Conf. on Computer Vision, 1995, pp. 203–209.

[26] J. Pinel, H. Nicolas, Estimation 2d illuminant direction and shadow segmentation in natural video sequences, in: Proceedings of VLBV, 2001, pp. 197–202.

[27] A.P. Pentland, Finding the illuminant direction, J. Opt. Soc. Am. 72 (1982) 448–455.

[28] I. Sato, Y. Sato, K. Ikeuchi, Illumination distribution from shadows, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1999, pp. 306–312.

[29] I. Sato, Y. Sato, K. Ikeuchi, Illumination distribution from brightness in shadows: adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions, in: Proc. IEEE Conf. on Computer Vision, 1999, pp. 875–882.

[30] I. Sato, Y. Sato, K. Ikeuchi, Stability issues in recovering illumination distribution from brightness in shadows, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, 2001, pp. 4000–4007.

[31] Y. Zhang, Y. Yang, Illuminant direction determination for multiple light sources, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2000, pp. 269–276.

[32] M.W. Powell, S. Sarkar, D. Goldgof, A simple strategy for calibrating the geometry of light sources, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 1022–1027.

[33] S.A. Shafer, Using color to separate reflection components, COLOR Res. Appl. 10 (4) (1985) 210–218.

[34] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall, New York, 2003.

[35] J.A. Marchant, C.M. Onyango, Shadow-invariant classification for scenes illuminated by daylight, J. Opt. Soc. Am. A 17 (2000) 1952–1961.

[36] B.V. Funt, M.S. Drew, Color space analysis of mutual illumination, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 1319–1326.

[37] D. Forsyth, A. Zisserman, Mutual illumination, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1989, pp. 466–473.

[38] T. Gevers, A.W.M. Smeulders, Color-based object recognition, Pattern Recogn. 32 (1999) 453–464.

[39] E. Salvador, A. Cavallaro, T. Ebrahimi, Shadow identification and classification using invariant color models, in: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2001, pp. 1545–1548.

[40] J. Kender, Saturation, Hue, and Normalized colors: Calculation, Digitization Effects, and Use, Tech. Rep., Carnegie-Mellon University, 1976.

[41] W. Pratt, Digital Image Processing, Wiley, New York, 1991.

[42] A. Cavallaro, T. Ebrahimi, Video object extraction based on adaptive background and statistical change detection, in: Proc. Visual Communications and Image Processing, 2001, pp. 465–475.

[43] Y. Otha, T. Kanade, T. Sakai, Color information for region segmentation, Comput. Graph. Image Process. 13 (1980) 222–241.

[44] E. Salvador, A. Cavallaro, T. Ebrahimi, Spatio-temporal shadow segmentation and tracking, in: Proc, SPIE's Image and Video Communications and Processing, vol. 5022, 2003, pp. 389–400.

[45] A. Prati, I. Mikic, M. Trivedi, R. Cucchiara, Detecting moving shadows: algorithms and evaluation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 918–923.

[46] W. Zhang, F. Bergholm, Multi-scale blur estimation and edge type classification for scene analysis, Int. J. Comput. Vis. 24 (1997) 219–250.