Accurate moving cast shadow suppression based on local color constancy detection

Ariel Amato, Mikhail G. Mozerov, Andrew D. Bagdanov, and Jordi Gonzàlez

***Abstract*—This paper describes a novel framework for detection and suppression of properly shadowed regions for most possible scenarios ocurring in real video sequences. Our approach requires no prior knowledge about the scene, nor is it restricted to specific scene structures. Furthermore, the technique can detect both achromatic and chromatic shadows even in the presence of camouflage that occurs when foreground regions are very similar in color to shadowed regions. The method exploits local color constancy properties due to reflectance suppression over shadowed regions. To detect shadowed regions in a scene, the values of the background image are divided by values of the current frame in the RGB color space. We show how this luminance ratio can be used to identify segments with low gradient constancy, which in turn distinguish shadows from foreground. Experimental results on a collection of publicly available datasets illustrates the superior performance of our method compared with the most sophisticated, state-of-the-art shadow detection algorithms. These results show that our approach is robust and accurate over a broad range of shadow types and challenging video conditions.**

***Index Terms*—Shadow removal, motion detection, color constancy.**

I. Introduction

MOVING object detection plays an important role in Computer Vision [1], [2]. It is a necessary preprocessing step in applications such as tracking, video compression and video surveillance. One of the most common and effective approaches to moving object localization is background subtraction, in which a model of the static scene background is subtracted from each frame of a video sequence [3], [4], [5], [6]. The task of moving object detection is strongly hindered by shadows cast by moving objects in the scene. In practice, cast shadows cause problems such as shape distortion, object merging, and even total failure of object detection and segmentation. An effective shadow detection algorithm is highly desirable for a wide range of applications in computer vision, including all applications requiring robust object segmentation.

Cast shadows are the areas projected on a surface due to an object that partially, or totally blocks the direct light source. The region where the direct light source is totally blocked is called the *umbra*, while the region where it is partially blocked is known as the *penumbra*. Obviously, an area affected by cast shadow experiences a change of illumination. Often this illumination change is considered only as a decrease in brightness, without significant variation in chromaticity. However, the assumption that chromaticity is invariant to cast shadows is not always correct. It is correct, in fact, only when light sources are white and there is no color blending among objects. This type of shadow is often called an *achromatic shadow*, while those that are not achromatic are referred to as *chromatic shadows*.

Removing chromatic shadows is a particularly challenging task due to the fact that they are extremely difficult to distinguish from the foreground because they have no clearly defined pattern. Interactions between color and texture in the background and shadows are highly variable and difficult to characterize. This paper introduces a framework for detection and suppression of moving chromatic and achromatic shadows. It does not require *a priori* knowledge about or restriction of the scene and lighting conditions. First, the proposed algorithm generates a mask of moving objects using a standard background subtraction technique. Then, our proposed method is used to detect cast shadow areas inside each detected moving object segment. To do this, luminance values of the background image are divided by the corresponding luminance values of the current frame, thus suppressing the reflectance component in shadow areas. Next, each object area is partitioned into a set of segments using a simple graph-based method. Finally, these sub-segments are classified as foreground or shadow by analyzing the intrinsic parameters of sub-segments.

This selective, region-based analysis is a distinguishing feature our approach, as almost all previously developed algorithms first use pixel-by-pixel (or small image patches) to determine the presence of shadows and then extend these pixel-wise solutions to neighboring areas. In the next section we explain the drawbacks of such techniques. The other important contribution of our method is its low computational complexity. The complexity is proportional to $O(n)$, where n is the number of pixels in the ROI of the frame. We achieve this by applying graph-based segmentation, where the area expansion algorithm exploits local gradient values instead global region analysis. This paper is organized as follows. In Section II we discuss shadow handling and related work. In Section III the framework of the proposed method is introduced. Experimental results are discussed in Section IV and finally concluding remarks are given in Section V.

II. Shadow Handling

Many publications are devoted to the shadow suppression techniques [7]. Some of them are extremely accurate [8], [9],

A. Amato (aamato@cvc.uab.es), M. G. *Mozerov (mozerov@cvc.uab.es), A. D. Bagdanov (bagdanov@cvc.uab.es) and J. Gonzàlez (poal@cvc.uab.es) are all with the Computer Vision Center (CVC), Universitat Autonoma de Barcelona, Campus UAB Edifici O, 08193 Bellaterra, Spain. Phone: +34-935811828, Fax:+34-935811670.-

but are not very useful in the background subtraction context. In this section we consider papers that are directly related to our work.

A. Related Work

Several moving cast shadow detection methods have been reported in recent years. Many of them work at the pixel level; for example: Horprasert et. al. [10] propose a color model that compares intensity to the chromaticity component at each pixel. Each pixel is classified as background, shaded, highlighted or moving foreground through a combination of three threshold values which are defined over a single Gaussian distribution. Cucchiara et al. [11] used shadow properties in the HSV color space to distinguish shadows from moving objects. These properties show that cast shadows darken the background in the luminance component, while the hue and saturation components change within certain limits. McKenna et al. [3] assume that cast shadows result in significant change in intensity without much change in chromaticity. Pixel chromaticity is modeled using its mean and variance, and the first-order gradient of each background pixel modeled using gradient means and magnitude variance. Moving shadows are then classified as background if the chromaticity or gradient information supports their classification as such. The major limitation of these algorithms is that they are restricted essentially to achromatic shadows, and they often require explicit tuning of parameters for each scene.

Statistical learning-based approaches have been developed to learn and remove cast shadows [12], [13], [14], [15]. For example, in [14] was presented a nonparametric framework to model surface behavior when shadows are cast on them. Physical properties of light sources and surfaces are employed in order to identify a direction in RGB space at which background surface values under cast shadows are found. However, these approaches are particularly affected by the training phase. These methods require a longer training period.

Methods which exploit region-based and photometric invariant color spaces are described in [16], [17], [18]. Stauder et al. [16] use a physics-based luminance model to describe illumination changes. They assume a plain textured background and a cast shadow is determined by combining the results of change detection, static edge detection, shading change detection and penumbra detection. Toth et al. [17] proposed a shadow detection algorithm based on color and shading information. They segment an image in to several regions based on color information and the mean shift algorithm. They consider that the intensity values of a shadow pixel divided by the same pixel in the background image should be constant over a small neighborhood. Yang et. al [18] proposed a moving cast shadow detection algorithm that combines shading, color, texture, neighborhoods and temporal consistency in the scene. However, many of these kinds of approaches may suffer from the composition of the scene in term of the background-foreground texture, and these methods cannot distinguish between shadows and foreground in a flat and nontextured background when exist camouflage in shadows.

The method presented in this paper is a selective, region-based approach that exploits the properties of the color constancy field which exists over shadowed regions due to the effect of reflectance suppression. The proposed method does not need any prior knowledge about scene and does not imply any kind of restriction about the structure of the scene. Furthermore, the technique can detect both achromatic and chromatic shadows even in the presence of shadow camouflage.

Table I summarizes and compares key features of the shadow detection algorithms mentioned above. In the table, "Texture constraint" indicates that the technique is limited to detection of shadows moving across a specific type of background: either textured or texture-less. The "Umbra & Penumbra" column indicates whether the method is capable of detecting both umbra and penumbra.

B. Drawbacks of pixel-wise analysis

When an object casts a shadow on a surface, it partially or completely blocks the surface of direct illumination from a light source, producing a change in its appearance. The measurement of this change, between a pixel affected by shadow and the same pixel in the absence of cast shadows, is one of the main properties used to classify cast shadows in background-foreground segmentation algorithms.

The earliest approaches to shadow suppression were based on the assumption that the value of the surface under cast shadows is just a linear scaling of its brightness component, without significant variation in its chromaticity components. This classification process can be done directly in chromaticity space [3] or in RGB space by separating brightness from the chromaticity components and introducing a color distortion term to separate shadow-foreground pixels [10]. The most sophisticated algorithms assume that it is possible to distinguish shadowed pixels from foreground pixels in RGB space, measuring difference both in terms of the angular and the intensity measure between two color vectors belonging to the background and the current frame respectively. This follows from the assumption that cast shadows do not produce significant changes in the color vector direction [6]. This idea has also been explored in the HSV color space to distinguish between shadows and foreground pixels [11].

Unfortunately, these pixel-wise approaches are subject to severe degradation in accuracy in many cases. Two of the most problematic situations are:

- **Shadow camouflage**
  Shadow-foreground discrimination fails when there is no difference in chromaticity between foreground and background (e.g. black car is moving in highway), hence inducing a strong similarity between shadow-foreground regions.
- **Chromatic Shadows**
  All these methods fail in the presence of strong chromatic shadows where the dominant illumination is not white or there is color bleeding due to multiple light reflection.

These limitations are a direct result of the pixel-wise approach. These approaches are based on the assumption that foreground and shadow pixels can be effectively clustered

METHODS	Chromatic shadow	Texture constraint	Shadows camouflage	Umbra & Penumbra
Horpraset [10]	No	No	No	No
Cucchiara [11]	No	No	No	Yes
McKena [3]	No	No	No	No
Stauder [16]	Yes	Yes	No	Yes
Toth [17]	Yes	Yes	No	No
Yang [18]	Yes	No	No	Yes
Martel-Brisson [14]	Yes	No	Yes	Yes
Jia-Bin Huang [15]	Yes	No	Yes	Yes
Proposed method	Yes	No	Yes	Yes

TABLE I

COMPARISON OF DIFFERENT SHADOW DETECTION ALGORITHMS. EACH ROW REPRESENTS AN ALGORITHM FROM THE LITERATURE, AND THE COLUMNS REPRESENT A RANGE OF CHARACTERISTICS OF SHADOW DETECTION METHODS.

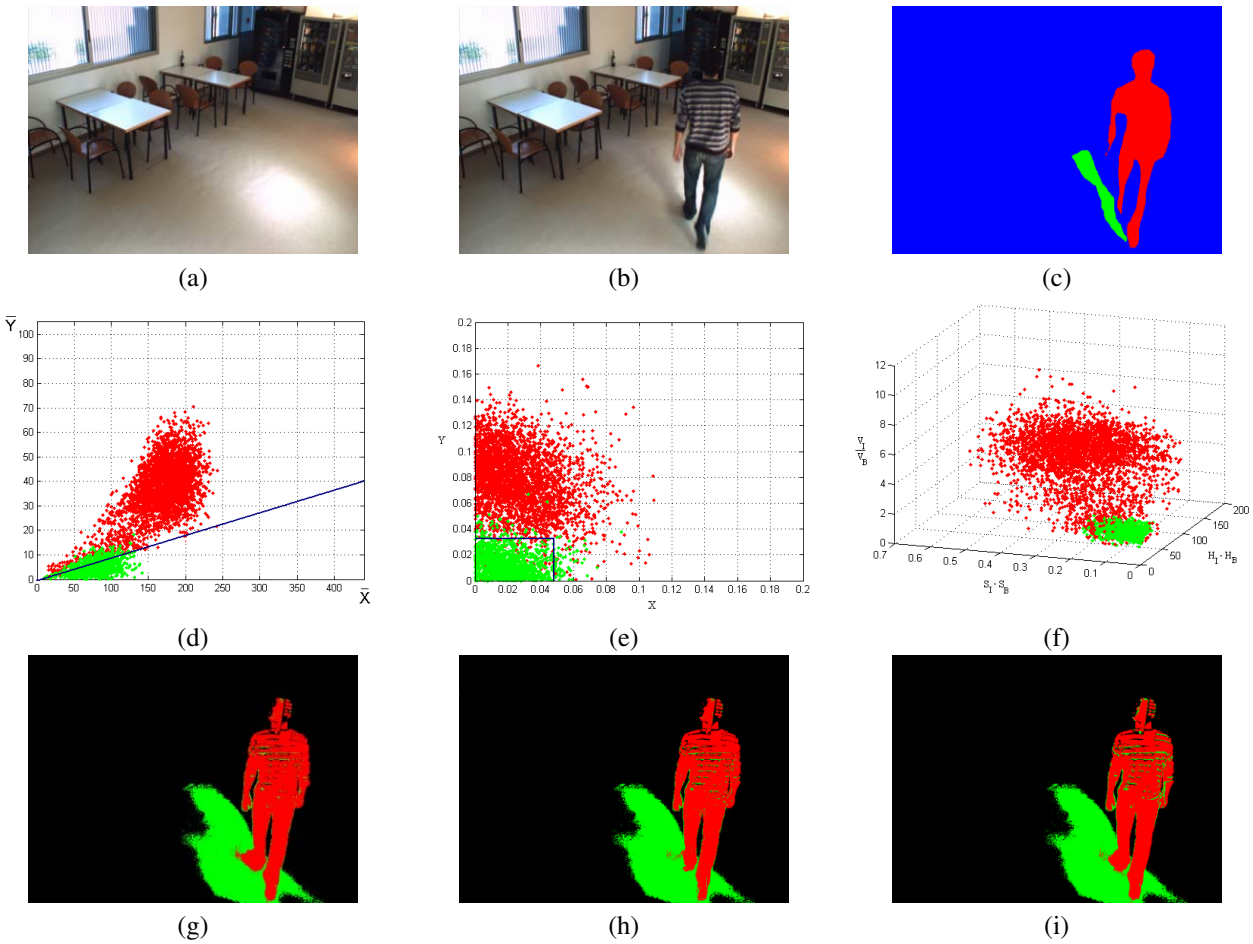


Fig. 1. (a), (b) and (c) are the background image, current image and hand-labeled segmentation image, respectively. (d) Histogram in polar RGB space, (e) Histogram in Delta Chromaticity Space, (f) Histogram in DHSV space. (g), (h) and (i) illustrate the result of the segmentation process based on an optimal threshold from different color spaces.

using a global distribution of pixels in one of various color spaces. In other words, that there exists a set of thresholds that minimizes detection error for foreground-shadow segmentation. In many specific cases these approaches do not work properly, and below we illustrate these phenomena with three examples. In Fig. 1, the achromatic shadow case is considered. Fig. 2 also illustrates the achromatic shadow case, but in this situation there are regions with similarity in chromaticity components between foreground and background or shadow camouflage. Finally, Fig. 3 illustrates a typical chromatic shadows case.

The real distribution of the foreground and shadow pixels based on a hand-labeled segmentation image is shown in Fig. 1(d-f). Here three different histograms are accumulated in the three color spaces. In Fig. 1(d) are shown the difference in angle and magnitude of the compared color in the RGB color space [6]; where the axis are computed as:

$$\bar{X} = \left|I^{bg}(\mathbf{x}) - I^{im}(\mathbf{x})\right| \cos\left(\cos^{-1}\left(\frac{I^{bg}(\mathbf{x}).I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})|\,|I^{im}(\mathbf{x})|}\right)\right)$$

$$\bar{Y} = \left|I^{bg}(\mathbf{x}) - I^{im}(\mathbf{x})\right| \sin\left(\cos^{-1}\left(\frac{I^{bg}(\mathbf{x}).I^{im}(\mathbf{x})}{|I^{bg}(\mathbf{x})|\,|I^{im}(\mathbf{x})|}\right)\right)$$

Note that the hand-labeled segmentations of Figs. 1(c), 2(c), and 3(c) are deliberately conservative in labeling pixels either

foreground or shadow. This is done in order to minimize the chance of falsely labeling pixels in the ambiguous region at the boundary between object and shadow. These hand-labeled segmentations should not be confused with the ground-truth segmentations used for quantitative evaluation in IV.

Fig. 1(e) shows differences in the chromaticity space [3], where the axis are:

$$X = \left| \frac{R^{bg}}{R^{bg}+G^{bg}+B^{bg}} - \frac{R^{im}}{R^{im}+G^{im}+B^{im}} \right|$$
$$Y = \left| \frac{G^{bg}}{R^{bg}+G^{bg}+B^{bg}} - \frac{G^{im}}{R^{im}+G^{im}+B^{im}} \right|$$

and in Fig. 1(f) differences in the modified HSV space (DHSV) [11].

Using this prior distribution one can estimate a set of optimal thresholds that minimize the global detection error. These thresholds for each distribution are drawn in the related plots in Fig. 1. The result of foreground-shadow classification for the entire frame using different sets of thresholds is shown in Fig. 1(g-i). We can see that for this achromatic shadow case, pixel-wise approaches can yield satisfactory results. Reconstruction in the case of shadows camouflage, however, as illustrated in Fig 2(g-i) and for the case of chromatic shadow in Fig 3(g-i)) is not satisfactory. To overcome the limitations of the pixel-wise approach,we propose a technique based on region analysis.

III. Moving Shadow Detection

In this section we present the framework of our method. Our approach is able to detect and suppress moving cast shadow regions for most possible scenarios that occur in real video sequences. The approach is based on the assumption that in the luminance ratio space, a low gradient constancy is present in all shadowed regions due to a local color constancy effect caused by reflectance suppression. In section III-B we give a formal proof of this property.

Note that our method aims to detect only moving cast shadows. In this case all pixels in the frame image must be previously segmented into background and motion regions. Our algorithm works only on the pixels in motion areas.

A. Motion Region Mask Formation

To obtain a binary mask of motion regions, we use a standard background subtraction algorithm, one based on differencing the current image frame and the background model [3], [4], [5], [6]. Note that the shadow detection method is not limited to any specific background model or algorithm.

The initial motion mask $M(\mathbf{x})$ is shown in Fig. 4(c). For our region-based approach this binary mask must be transformed into a set of independent regions $\Phi = \{\mathrm{o}_1, \mathrm{o}_2, ..., \mathrm{o}_k\}$, where K is the number of motion regions in the current frame. This can be achieved using strongly connected component analysis [19]. Relatively small components are suppressed by a simple morphological filter. Consequently, one component segment in the label mask Fig. 4(d) corresponds to one motion object in the set Φ.

B. Shadow Model and Reflectance Suppression

Assuming Lambertian reflectance, an image obtained from a scene by a standard RGB camera can be analyzed using a simple luminance model [16]:

$$\mathbf{L}(\mathbf{x}) = \mathbf{E}(\mathbf{x})\,\boldsymbol{\rho}(\mathbf{x}),$$

where $\mathbf{L}(\mathbf{x}) = [L_R(\mathbf{x}), L_G(\mathbf{x}), L_B(\mathbf{x})]^T$ is the luminance vector of the RGB color space in the image plane $\mathbf{x} \in \mathbf{X}$, $\mathbf{E}(\mathbf{x}) = [E_R(\mathbf{x}), E_G(\mathbf{x}), E_B(\mathbf{x})]^T$ is the irradiance vector of the input signal, and $\boldsymbol{\rho}(\mathbf{x}) = [\rho_R(\mathbf{x}), \rho_G(\mathbf{x}), \rho_B(\mathbf{x})]^T$ is the reflectance vector of the object surface reflected at the pixel $\mathbf{x}$. Though vectors $\mathbf{E}$ and $\boldsymbol{\rho}$ should properly be expressed by diagonal matrices, we simplify our notation throughout the text the product of two vectors indicates simple, component-wise multiplication $\mathbf{ab} = a_i b_i$. Division is similarly indicated.

The irradiance component of the input signal for one light source in shadows areas can now be expressed as:

$$\mathbf{E}(\mathbf{x}) \;=\; \mathbf{C}_a + \mathbf{C}_b \cos(\theta(\mathbf{x}))\varsigma(\mathbf{x}), \tag{1}$$

where $\mathbf{C}_a$, $\mathbf{C}_b$, $\theta(\mathbf{x})$, are the intensity of ambient light, the intensity of the light source, and the angle between the light source direction and a surface normal, respectively, $\varsigma(\mathbf{x}) \in [0,1]$ is a shadow parameter that represents the transition inside the penumbra, which depends on the light source and scene geometry. Generally it is characterized by slow spatial variation [20]. When $\varsigma(\mathbf{x})$ is equal to 0, the quantity of light reflected at pixel $\mathbf{x}$ is from ambient light alone and it belongs to the umbra region. For $0 < \varsigma(\mathbf{x}) < 1$, the pixel is located in the penumbra and when $\varsigma(\mathbf{x}) = 1$ the pixel is outside of the shadow region.

The appearance of an arbitrary pixel $\mathbf{x}$ in an image sequence will vary according to illumination conditions and the configuration of objects that may cast shadows. Letting $\mathbf{L}^{im}(\mathbf{x})$ denote the pixels belonging to cast shadows in the current frame and $\mathbf{L}^{bg}(\mathbf{x})$ those that do not. The luminance ratio of these pixels can be written as:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{L}^{bg}(\mathbf{x})}{\mathbf{L}^{im}(\mathbf{x})} = \frac{\mathbf{E}^{bg}(\mathbf{x})\boldsymbol{\rho}^{bg}(\mathbf{x})}{\mathbf{E}^{im}(\mathbf{x})\boldsymbol{\rho}^{im}(\mathbf{x})}. \tag{2}$$

If the point $\mathbf{x}$ belongs to the shadow region R^{sh} of the current image $\mathbf{L}^{im}(\mathbf{x})$ then the two reflectances in equation (2) are equal because the reflectance $\boldsymbol{\rho}(\mathbf{x})$ of the projected surface point $\mathbf{x}$ does not changes with time.

Therefore, the result of the luminance ratio $\mathbf{D}$ is reduced to:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{E}^{bg}(\mathbf{x})}{\mathbf{E}^{im}(\mathbf{x})}, \forall \mathbf{x} \in R^{sh}. \tag{3}$$

After substituting (1) in (3) we have

$$\mathbf{D}(\mathbf{x}) \;=\; \frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg}\cos(\theta(\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im}\cos(\theta(\mathbf{x}))\varsigma(\mathbf{x})}. \tag{4}$$

Let $\Delta\mathbf{x}$ represents the distance between two neighboring pixels in the image plane. The difference between these two luminance ratios can be written as:

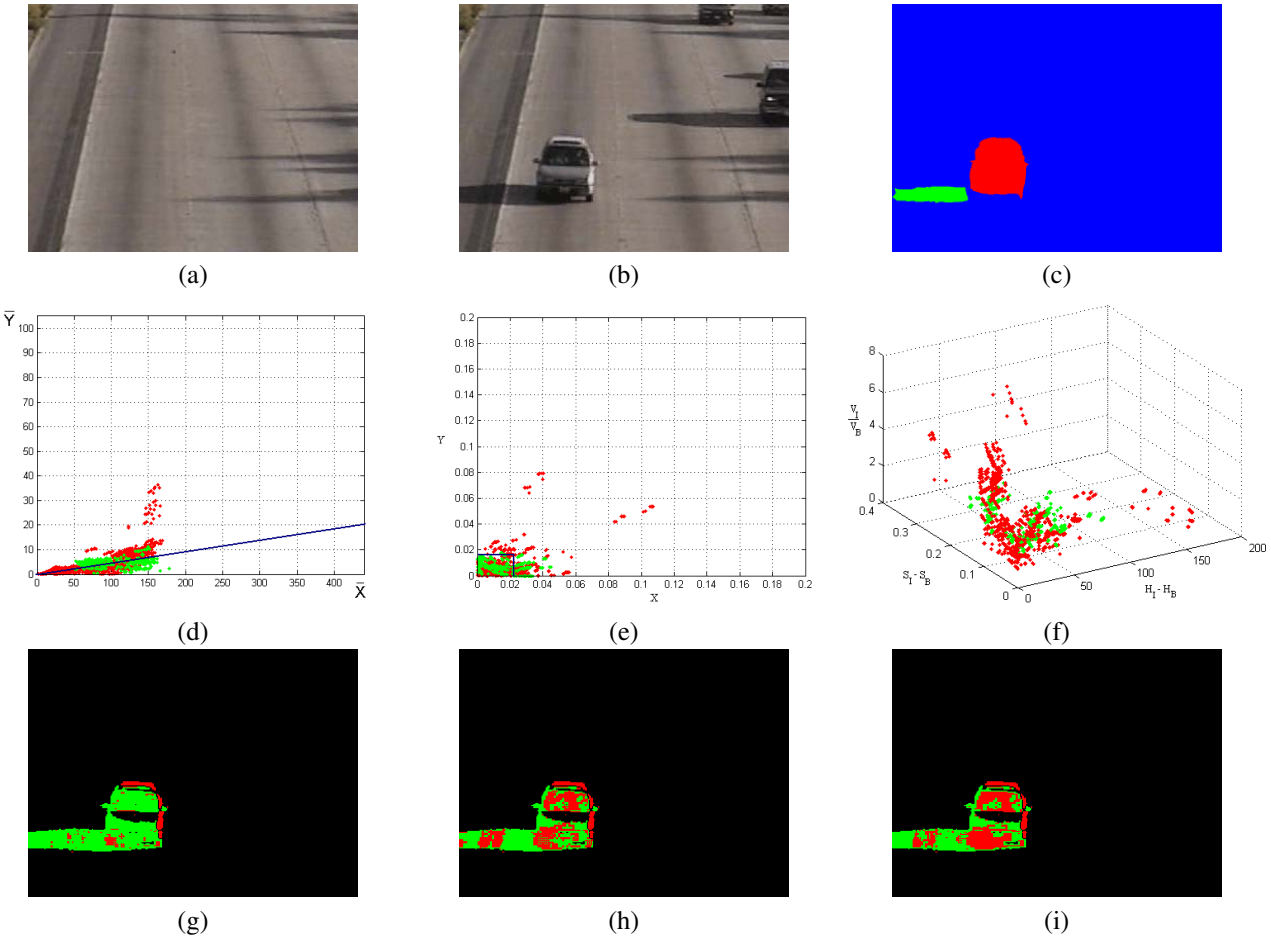


Fig. 2. (a), (b) and (c) are the background image, current image and hand-labeled segmentation, respectively. (d) Histogram in polar RGB space, (e) Histogram in Delta Chromaticity Space, (f) Histogram in DHSV space. (g), (h) and (i) illustrate the result of the segmentation process based on an optimal threshold from different color spaces.

$$\mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x}+\Delta\mathbf{x}) = \frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg}\cos(\theta(\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im}\cos(\theta(\mathbf{x}))\varsigma(\mathbf{x})} - \frac{\mathbf{C}_a^{bg} + \mathbf{C}_b^{bg}\cos(\theta(\mathbf{x}+\Delta\mathbf{x}))}{\mathbf{C}_a^{im} + \mathbf{C}_b^{im}\cos(\theta(\mathbf{x}+\Delta\mathbf{x}))\varsigma(\mathbf{x}+\Delta\mathbf{x})}. \quad (5)$$

Assuming that the scale factor ς and the angle θ are slowly varying functions: $\varsigma(\mathbf{x}) \approx \varsigma(\mathbf{x}+\Delta\mathbf{x})$ and $\theta(\mathbf{x}) \approx \theta(\mathbf{x}+\Delta\mathbf{x})$, we obtain

$$\mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x}+\Delta\mathbf{x}) \approx 0, \quad (6)$$

which means that a local constancy exists for any pair of pixels belonging to the shadow region. In contrast, local constancy in equation (6) does not hold for foreground pixels because of the inequality of the reflectance components in (2).

The local color constancy condition in equation (6) is derived using a single light source model. However, the assumption used will also hold when $\mathbf{E}(\mathbf{x})$ in equation (1) is formed by a linear combination of multiple light sources. The commonly used irradiance model of equation (1) from [16] assumes that the intensity of ambient light $\mathbf{C}_a$ is constant and that the intensity of the light source $\mathbf{C}_b$ is proportional to $(\frac{1}{r^2})$, where r is the distance between the object and light source [21]. These assumptions are valid for a broad range of imaging conditions, and for deriving our local color constancy criterion it is reasonable to use a model where both $\mathbf{C}_a$ and $\mathbf{C}_b$ are constants. This same, simplified lighting model has been used in the derivation of several other shadow suppression models [16], [17], [18].

Using the local constancy effect that exists in shadow regions, our algorithm distinguishes between shadows and foreground regions.

C. Regions with Local Color Constancy

To detect regions with local color constancy, we first calculate the luminance ratio image. The luminance ratio for a single pixel is written as:

$$\mathbf{D}(\mathbf{x}) = \frac{\mathbf{L}^{bg}(\mathbf{x}) + \nu}{\mathbf{L}^{im}(\mathbf{x}) + \nu}, \quad (7)$$

where ν is a quantization constant, which is chosen to be unity for the standard eight bit input signal. It is very important to note that for our shadow detection algorithm it is more essential to make fine segmentation in shadow-like pixels (i.e. those with higher luminance values in background pixels than the pixels in the current image). When the background image is divided by the current image, the luminance ratio image $\mathbf{D}$ is segmented into two types of regions: foreground regions

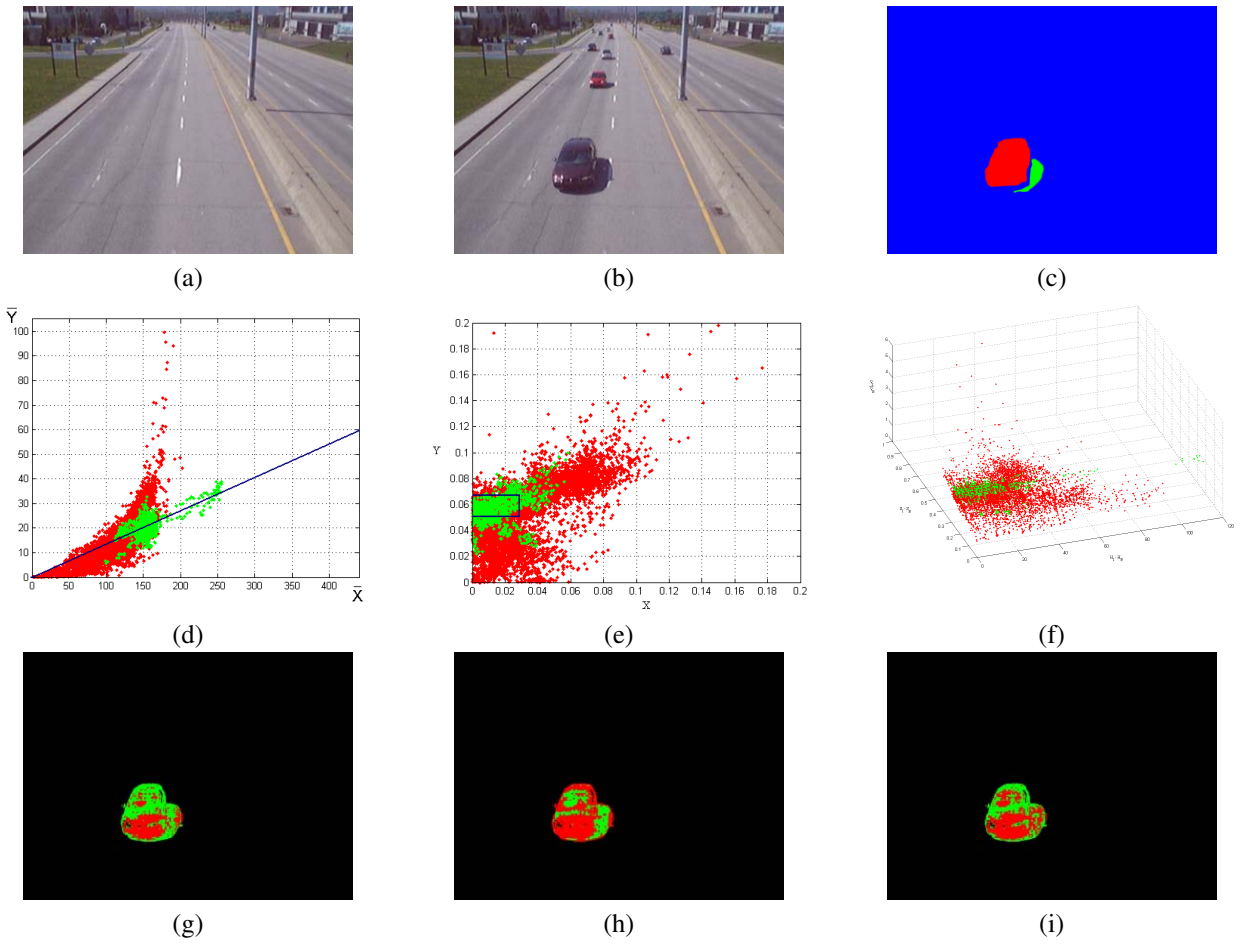


Fig. 3. (a), (b) and (c) are the background image, current image and hand-labeled segmentation, respectively. (d) Histogram in polar RGB space, (e) Histogram in Delta Chromaticity Space, (f) Histogram in DHSV space. (g), (h) and (i) illustrate the result of the segmentation process based on an optimal threshold from different color spaces.

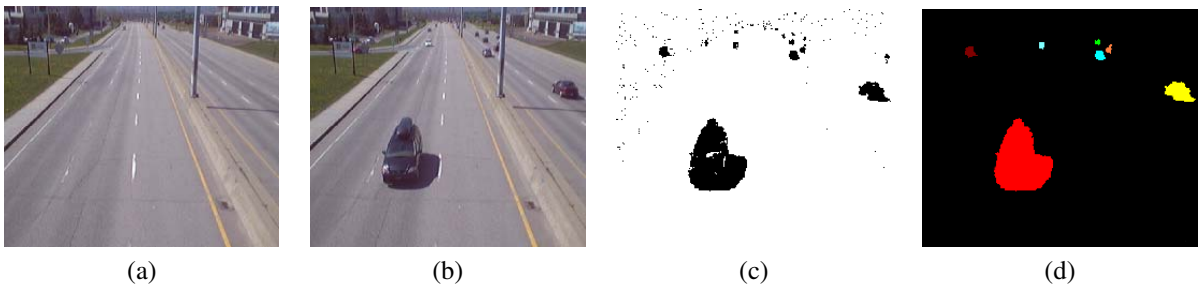


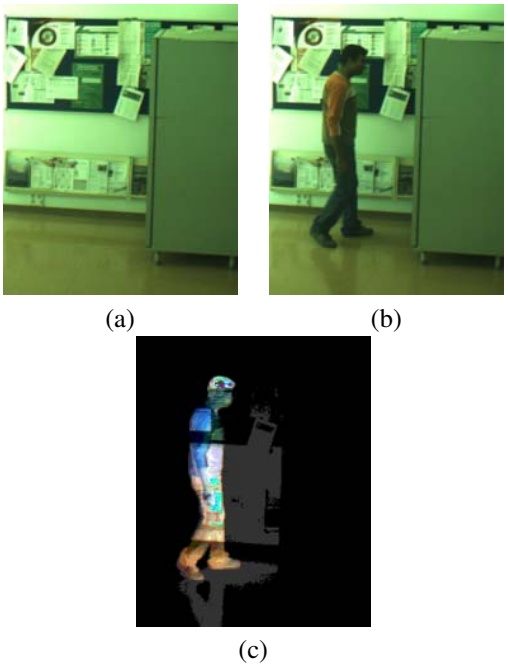
Fig. 4. Background image (a), current frame (b), binary motion pixels mask (c) and object mask (d).

where $2^{-8} \leq \mathbf{D}(\mathbf{x}) \leq 1$ and shadow like regions where $1 \leq \mathbf{D}(\mathbf{x}) \leq 2^{8}$. It is easy to see that in the shadow-like regions, measurement is more precise.

An example of the luminance ratio image in the RGB color space is illustrated in Fig. 5(c), where Fig. 5(a) shows the background model and Fig. 5(b) is a frame of the test sequence.

Let us analyze the values of the luminance ratio $\mathbf{D}$ inside each motion areas. As we explained before in (3), the value of the function $\mathbf{D}(\mathbf{x})$ in the shadow area depends only on the irradiance ratios between the background and current image, and form regions (usually one shadows region per motion segment) that are characterized by smooth spatial changes in (6), which we call local color constancy. We stress that local color constancy does not assume, in general, color constancy in a full shadow region and a value of the luminance ratio $\mathbf{D}(\mathbf{x}_1)$ can be considerably different from a value $\mathbf{D}(\mathbf{x}_2)$ if the distance between two pixels $|\mathbf{x}_1 - \mathbf{x}_2|$ is significant.

In contrast, in the foreground areas the value of the function $\mathbf{D}(\mathbf{x})$ depends both on the irradiance ratios and on reflectance ratios in (2) and, in general, do not form considerable regions with local color constancy. In rare cases, when this assumption in the foreground areas does not hold (i.e. poor texture), we use additional features (extrinsic terminal points, which will be explained later) to solve the problem. In this way, local color constancy detection allows us to distinguish background



(a) (b)

(c)

Fig. 5. (a) Background image, (b) current image, (c) luminance ratio image in the RGB color space.

from foreground.

Thus, the main goal of our algorithm is to detect local color constancy regions, which is classical segmentation task. Segmentation problems can usually be solved by local neighborhood analysis. Local neighborhood analysis is a widely used technique in image processing. Most of them utilize sliding windows with a fixed shape and size or with a locally adaptive shape.

Other approaches work with neighborhoods that are the result of an initial segmentation of the entire region of interest, for example watershed [19] or mean shift segmentation [20]. But in our case, to detect shadow segments we have to overcome the problem of gradual change. Therefore, we apply a gradient-based segmentation technique. More precisely, our algorithm forms a set of gradient space connected neighborhoods (GSCN) or a set of nonoverlapping segments. In the next section we explain this technique.

D. Gradient-space Connected Neighborhood Segmentation

A gradient-space connected neighborhood is defined as a set of pixels in an image in which any two pixels are gradient connected. That is, there exists a (four- or eight-connected) path between any pair of pixels in the neighborhood. All pixels of the path satisfy the following condition: $|D(\mathbf{p})-D(\mathbf{q})| \leq \partial$, where ∂ is a given threshold, $\{D(\mathbf{p}), D(\mathbf{q})\}$ are values in of any pair of adjacent pixels $\{\mathbf{p}, \mathbf{q}\} \in \mathbf{X}$ on the path.

Figure 6(a) illustrates two paths between pixels $(0,2) - (2,0)$ and $(0,5) - (7,0)$ of gradient-space connected neighborhoods. Figure 6(b) shows two formed gradient adaptive neighborhoods. The edges that separate these pixels, or cells, may be represented as partitions or thin dams (see Fig. 6(b)). The heights of these partitions are proportional to the difference between adjacent pixels. When any pixel of the image is flooded by letting water rise to a fixed level ∂ then several noncommunicating pools are formed, each with a unique levels of water.

To form GSCNs we use a standard graph-based technique. An undirected image cover graph is defined as $G = (V, E)$, where each pixel $\mathbf{x}$ of a segmented object $o_k \in \Phi$ has a corresponding node $v(\mathbf{x})$, and to each pair of neighboring pixels $\mathbf{x}_i, \mathbf{x}_j \in o_k$ corresponds one edge $e(\mathbf{x}_i, \mathbf{x}_j)$ with weight:

$$w(e(\mathbf{x}_i, \mathbf{x}_j)) = \prod_{c \in \{R,G,B\}} H(\partial - |D_c(\mathbf{x}_i) - D_c(\mathbf{x}_j)|), \quad (8)$$

where H is the Heaviside step function:

$$H(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{otherwise} \end{cases}.$$

Then a new graph is formed $\tilde{G} = (V, \tilde{E})$ such that $\tilde{G} \subset G, \forall e \in E \text{ s.t. } |e| \neq 0 \Rightarrow e \in \tilde{G}$. In other words, this new graph inherits all vertices of the graph G and only edges with nonzero weight $|e|$. To reach a new subdivision of an object o_k, a forest of graph trees is formed with the aid of depth first or breadth first search algorithms [19]. Finally, all vertices (pixels) of a tree form a sub-segment inside a considered object such that:

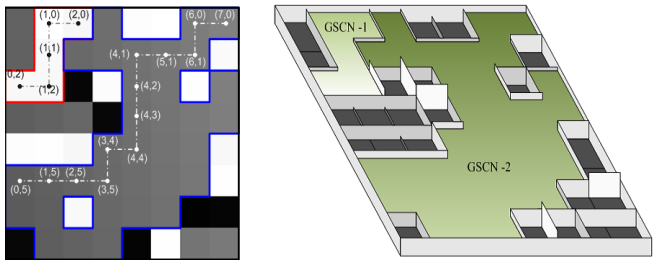
$$\bigcup_{l=1}^{L_k} s_l^k = o_k, \bigcap_{l=1}^{L_k} s_l^k = \emptyset, \bigcup_{k=1}^{K} o_k = \Phi, \quad (9)$$

where L_k is the number of trees, i.e. the number of sub-segments s_l^k of object o_k, of the graph $\tilde{G}$.

Such segmentation is unique and depends only on a given threshold ∂. Further, it permits us to classify as shadow or foreground all pixels inside a sub-segment s_l. In contrast, approaches based on sliding windows must form a neighborhood for each pixel separately, and consequently have much higher computational complexity. However, we choose gradient-space connected neighborhoods not only because of the low computational complexity relative to similar segmentation techniques, but also because for our method it is very important to obtain a unique shadow segment for every shadow region. The main reason for this requirement is that sometimes shadows are formed by a large penumbra, and therefore segmentation algorithms based on global pixel analysis, such as mean shift and watershed segmentations, can fail due to over- and/or under-segmentation. This is illustrated in Fig. 7. For example, if a considered region must include only the pixels with values in some fixed interval ∂ then the segmentation process usually splits shadows region or merges shadow and foreground pixels into one segment. This situation is shown in Fig. 7(b), where ∂ is a small interval value that results in three segments $\{s_1, s_2, s_3\}$, while $\tilde{\partial}$ is a big interval value that under-segments by merging $SR \subset \tilde{s}$. If there is no penumbra region Fig. 7(a)), over-segmentation does not occur.

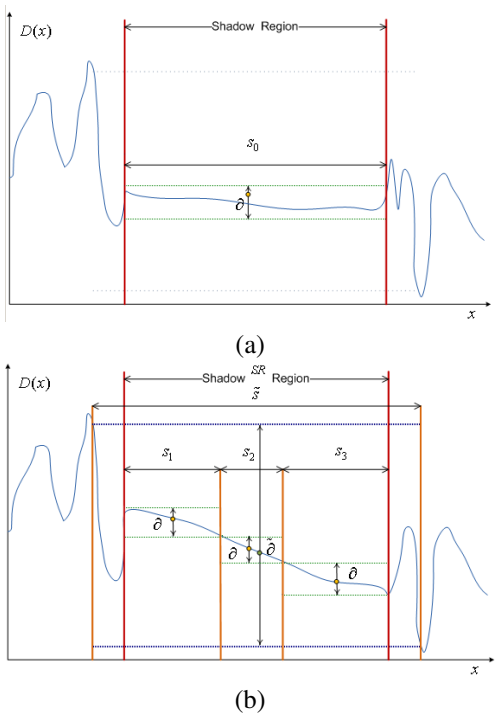
E. Classification Process

To classify each sub-segment s_l^k of motion region o_k as shadow, we exploit local features of the region and apply three classification criteria.



(a) (b)

Fig. 6. (a) Illustration of the GSCN concept, (b) formed GSCNs.



(a)

(b)

Fig. 7. Illustration of shadow splitting and over-segmentation effects. (a) Signal without penumbra effect. (b) Signal with penumbra that causes splitting and over-segmentation.

The features used to perform this local classification are: $\{\mu_l^k, |s_l^k|, \tau_l^k\}$, where:

- μ_l^k: Mean value in the region s_l^k.

$$\mu_l^k = \left|s_l^k\right|^{-1} \sum_{\mathbf{x} \in s_l^k} \mathbf{D}\left(\mathbf{x}\right).$$

- $|s_l^k|$: Number of pixels that belong to a segment s_l^k.

$$\left|s_l^k\right| = \sum_{\mathbf{x} \in o_k} M_{\mathbf{x} \in s_l^k}\left(\mathbf{x}\right).$$

- τ_l^k: Terminal pixel weight, where $\hat{t}_l^k\left(s_l^k\right)$ is the number of external terminal pixels of a sub-segment s_l^k. This means, the number of edge of sub-segment s_l^k that are neighbor to non-object parts of the image. Finally, $t_l^k\left(s_l^k\right)$ is the number of all terminal pixels of the sub-segment s_l^k. This includes all internal and external pixels.

$$\tau_l^k = \frac{\hat{t}_l^k\left(s_l^k\right)}{t_l^k\left(s_l^k\right)}.$$

Sub-segments are classified based on the combination of three decision rules:

1) *Luminance difference criterion.*
The first classification rule is based on the assumption that in shadow regions the luminance of each RGB component in the background image is greater than the current frame. Following this concept we introduce the shadow-like indicator function:

$$Sh_\mu(\mu_l^k) = \prod_{c \in \{R,G,B\}} H(\mu_c(s_l^k) - 1), \tag{10}$$

where $Sh(s_l^k) = 1$ when the sub-segment s_l^k can belong to the shadow class. When $Sh(s_l^k) = 0$, the sub-segment s_l^k can be directly classify as foreground.

2) *Segment size criterion.*
The size classification rule is based on the assumption that for each shadow region of an object o_k, there is just one shadow sub-segment (the ideal case), or that there are a few relatively large shadow sub-segments. In contrast, the object o_k is generally formed by many foreground sub-segments, but each of these sub-segments contains few pixels as a result of superposition of two topological structures: background and foreground. Thus, the segment size criterion is represented by the indicator function:

$$Sh_{|s|}\left(\left|s_l^k\right|\right) = \begin{cases} 1, \text{if } \left(\left|s_l^k\right| > |o_k|\,\lambda\right) \\ 0, \text{otherwise,} \end{cases} \tag{11}$$

where λ is the relative size of the smallest sub-segment in an object o_k that can be shadow.

3) *Extrinsic terminal point weight criterion.*
This rule is based on the spatial topology of shadows. Shadow regions are usually located around foreground regions. Therefore any shadow sub-segment of an object o_k contains a considerable amount of extrinsic terminal points of the region (see Fig. 8), relative to the total number of terminal points in the region. In contrast, foreground sub-segments typically have the weight of the extrinsic terminal points equal to zero or have an insignificant amount of such points. Therefore, the extrinsic terminal point weight criterion is:

$$Sh_\tau\left(\tau_l^k\right) = \begin{cases} 1, \text{if } \left(\tau_l^k > \tau_0\right) \\ 0, \text{otherwise,} \end{cases} \tag{12}$$

where τ_0 is an experimentally determined threshold.

Joint classification rule

The final shadow classification rule is based on the superposition of all previously described criteria (10), (11) and (12):

$$s_l^k = \begin{cases} \textbf{Shadow}, \text{if } \left(Sh_\mu\left(\mu_l^k\right) \cap Sh_{|s|}\left(\left|s_l^k\right|\right) \cap Sh_\tau\left(\tau_l^k\right)\right) \\ \textbf{Foreground}, \text{otherwise} \end{cases} \tag{13}$$

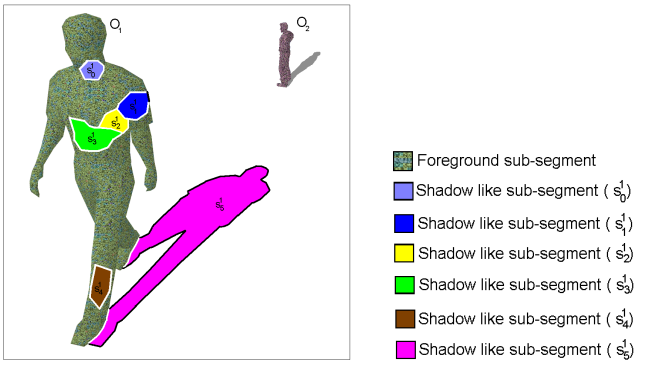


Fig. 8. Point-wise border representation: white borderlines represent end points of the segment spatially connected with another object's segments Black borderlines represent extrinsic terminal points.

F. Edge Noise Correction

When motion regions $o_k \in \{\Phi\}$ are detected using background subtraction techniques, the algorithm usually includes outlier pixels in the object mask that form a bright, narrow fringe between shadow and background regions in the image plane. This edge effect can result from JPEG or similar compression techniques and can also be the result of other signal transmission artifacts. This outlier fringe can completely spoil the result of classification based on the external terminal point weight criterion. To overcome the negative consequences of the "bright edge effect" we shrink each individual motion region o_k mask by a small, radius-5 morphological erosion. Only then do we start the classification process described in Section III-E. When the classification is finished we add all unclassified pixels that belong to the edge region to the nearest shadow or foreground sub-region of the motion region o_k.

IV. Experimental Results

In this section we explore the performance of our approach and discuss the parameters which are involved in the method and their estimation.

A. Parameter Analysis

Our algorithm relies on several parameters that must be set. In this section we describe each parameter and how they may each be estimated directly from data.

1) *Minimum gradient threshold ∂.*
For every motion object segment o_k a specific ∂_k must be computed:

$$\partial_k = \alpha \left|o_k\right|^{-2} \left(\sum_{\mathbf{x}\in o_k} \left|\mu^{bg}(\mathbf{x})\right|\right) \left(\sum_{\mathbf{x}\in o_k} \left|\sigma^{bg}(\mathbf{x})\right|\right), \quad (14)$$

where $\mu^{bg}(\mathbf{x})$ and $\sigma^{bg}(\mathbf{x})$ (mean value and standard deviation) are the pixel-wise parameters of the background image model described in Section III-A, and $|o_k|$ represents the number of pixels in region k. Because both parameters are vectors in the RGB color space, $\left|\mu^{bg}(\mathbf{x})\right|$ and $\left|\sigma^{bg}(\mathbf{x})\right|$ are the magnitude or grayscale value of a color vector. So, the threshold ∂_k is proportional to the mean values of the background parameters over an entire motion segment o_k that must be sub-segmented into a set $s_i^k \in \{o_k\}$. The experimental parameter α of (14) is fortunately robust and can be considered as an intrinsic constant of the method. At least for all tested sequences the value of $\alpha = 0.00621$ was optimal.

2) *Relative size threshold λ.*
The value of the relative size threshold λ has been calculated on the basis of the optimization of two criteria: true positive foreground (TPf) and false positive foreground (FPf). The value of this threshold could be optimized for each scene. However, a scene independent value can be used, since the final result does not show a big variation relative to optimal (less than 0.5%) with λ equal to 0.04. Note that this parameter can be decreased if image or scene conditions warrant. The segment size criteria uses λ to quickly discard sub-segments formed by a small number of pixels. In the case of extreme situations where shadows are formed by few pixels, the value of λ can be safely decreased – at the cost of increasing the number of sub-segments to be classified by the "Extrinsic terminal point wight criterion".

3) *Extrinsic terminal point weight threshold τ_0.*
The value of the threshold τ_0 defined in equation (12) is calculated using the same optimization process as in the previous for λ. The value of this threshold could also be optimized for each scene. However, a scene independent value has been found, and this value guarantees even more robustness than the value of λ (less than 0.1% of error change) with τ_0 equal 0.2.

B. Performance Evaluation

Here we provide qualitative and quantitative results of our approach on publicly available sequences.

Qualitative results

Rows (a-i) in Fig. 9 illustrate all steps of our algorithm:

- (a) The image being segmented.
- (b) The binary motion detection binary mask *(Section III-A)*.
- (c) Object masks. *(Section III-A)*.
- (d) Difference image plane *(Section III-B)*.
- (e) Result of GSCN segmentation *(Section III-D)*.
- (f) Edge noise correction *(Section III-F)*.
- (g) Classification based on the luminance difference and segment size criteria *(Section III-E (1)-(2))*.
- (h) Classification based on the terminal point weight criterion *(Section III-E (3))*.
- (i) Final segmentation *(Section III-E (Joint classification rule))*.

The three different columns of Fig. 9 represent three diverse image scenes taken from sequences: (I) - Grass field #184, (II) - Highway II #157, (III) Highway II #801. For scene (I) the classification was done completely based on the segment size criterion (I-g), in spite of that the scene has an irregular background. The classification could be done directly with segment size criterion because in this case the background is rich in term of texture and the foreground is darker than the background. In the scene (II) the luminance difference

criterion also plays an important role since the big part of the foreground is brighter than the background, but the final decision was done by terminal point weight criterion. The classification of the scene (III) can be done only with all three proposed criteria.

The results of our method in different scenarios is illustrated in Fig. 10. In the figure we include five different frames taken from five diverse scenes: (I) - Hallway frame #163, (II) - Auto frame #1143, (III) - Highway II frame #253, (IV) - Highway I frame #353 and (V) - CVC outdoor #509. The columns in Fig. 9 (a), (b) and (c) represent:

- (a) The image being segmented.
- (b) Motion object mask.
- (c) Final segmentation.

Scene (I) is an indoor scenario where shadows are projected on the floor and on the wall it being a two disconnected shadows patch. Scene (II) is also an indoor scenario, but the environment contains multiple overlapping light sources and a large penumbra region. Scene (III) shows an outdoor scene with flat gray background which it is affected by severe shadow camouflage and with chromatic shadow . Scene (IV) is another outdoor scenario where multiple real objects are combined in a single object mask. Finally scene (V) contains a long chromatic shadow cast on an irregular background surface.[1]

Quantitative results

The quantitative evaluation is based on two standard metrics for evaluating the performance of cast shadow detection algorithm introduced by Prati et al. [7]: shadow detection rate (η) and shadow discrimination rate (ξ). These two metrics are as follows:

$$\eta = \frac{TP_S}{TP_S + FN_S}; \qquad \xi = \frac{\overline{TP_F}}{TP_F + FN_F}, \tag{15}$$

where the subscript S stands for shadow and F for foreground. $\overline{TP_F}$ is the number of true positive foreground pixels detected minus the number of points detected as shadows but belonging to the foreground. The quantitative comparison was done with the Martel-Brisson et. al method [14], and with the Jia-Bin Huang et. al method [15] since these methods perform best among current state-of-the-art techniques and are also robust under many different scene conditions (see Table I). Note that results of the other approaches are taken from [15] and [14]. Table II shows the quantitative results. Comparative results are shown on the standard benchmark sequences: Highway I, Highway II and Hallway (the first three rows of table II).

To provide further evidence of the performance of our approach, we have ground-truthed an additional three sequences for the purpose of benchmarking shadow suppression algorithms. These sequences and ground truth are publicly available and were selected to contain a variety of imaging scenarios and challenging background conditions. The new sequences are Pets-2009 View 7, CVC-outdoor and Football Match. [2] In the last three rows of table II we show the performance of our approach on these three new sequences. There is no publicly available source code or executables for the two other reported methods, and for this reason we only provide performance evaluation for our approach on the three new sequences. All sequences and ground truth are publicly available.

Sequences		Methods		
		Our method	Huang [15]	M.Brisson [14]
Highway I	η	0.81	0.70	0.70
	ξ	0.85	0.82	0.84
Highway II	η	0.72	0.76	0.68
	ξ	0.75	0.74	0.71
Hallway	η	0.84	0.82	0.72
	ξ	0.91	0.90	0.86
Pets2009 V7	η	0.96	-	-
	ξ	0.95	-	-
CVC Outdoor	η	0.91	-	-
	ξ	0.96	-	-
Football Match	η	0.80	-	-
	ξ	0.95	-	-

TABLE II
QUANTITATIVE RESULTS FOR DIFFERENT SEQUENCES

This comparison shows that our proposed technique surpasses the performance of known methods. We also note that our approach is fast, needing just $O(n)$ operations, where n is the number of pixels in the frame.

V. Conclusion

This paper describes a novel approach to distinguishing moving objects from their shadows. Our method exploits the property of local color constancy for the shadow region formed by reflectance suppression. The approach is uses image ratios to coarsely segment a frame into motion regions, and we show how these ratios can then be used to segment foreground from shadows. All parameters in our algorithm can be estimated directly from the data, and it deals efficiently and accurately with the most difficult issues in the field, including umbra/penumbra detection, chromatic shadows and shadow camouflage.

The effectiveness of the proposed method was validated by the higher recognition rates achieved over a collection of publicly available sequences Our algorithm is fast and the computational complexity is linear in the number of pixels in the frame.

Acknowledgment

This work is supported by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDIvideo project, and by the Spanish MEC under projects TIN2006-14606 and CONSOLIDER-INGENIO 2010 MIPRCV CSD2007-00018.

References

[1] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, August 2002.

[1] In order to provide more qualitative information for our method, several segmented sequences are available at http://www.cvc.uab.es/~aamato/results/Shadows_Detection/

[2] Sequences and ground truth are available at: http://www.cvc.uab.es/~aamato/results/Shadows_Detection/

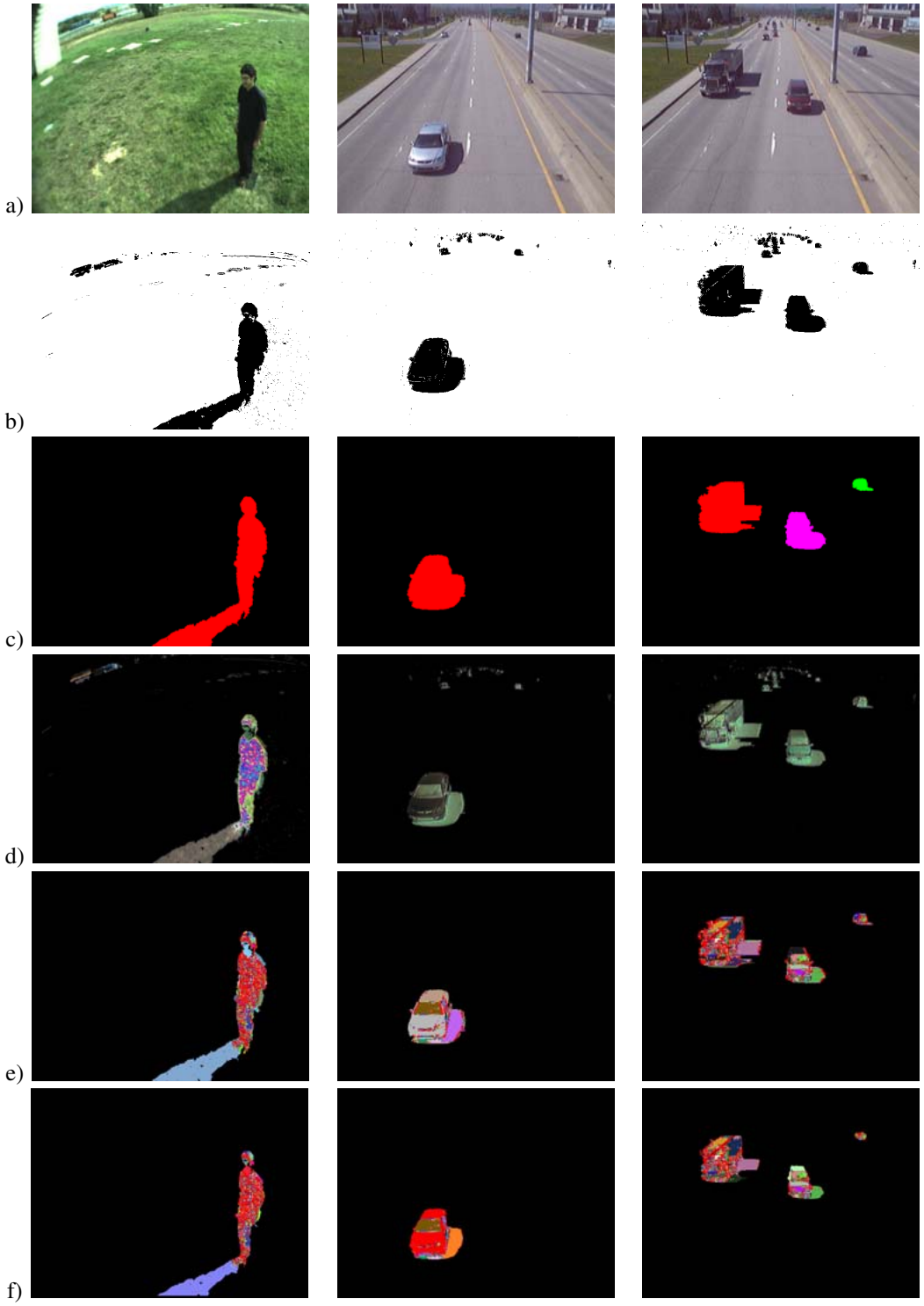


Fig. 9. The result of different steps of our method (red color means foreground, green color means shadow): (a) image being segmented; (b) motion detection mask; (c) object masks; (d) - image difference plane; (e) result of GSCN segmentation; (f) edge noise correction. *(figure continued on next page)*.

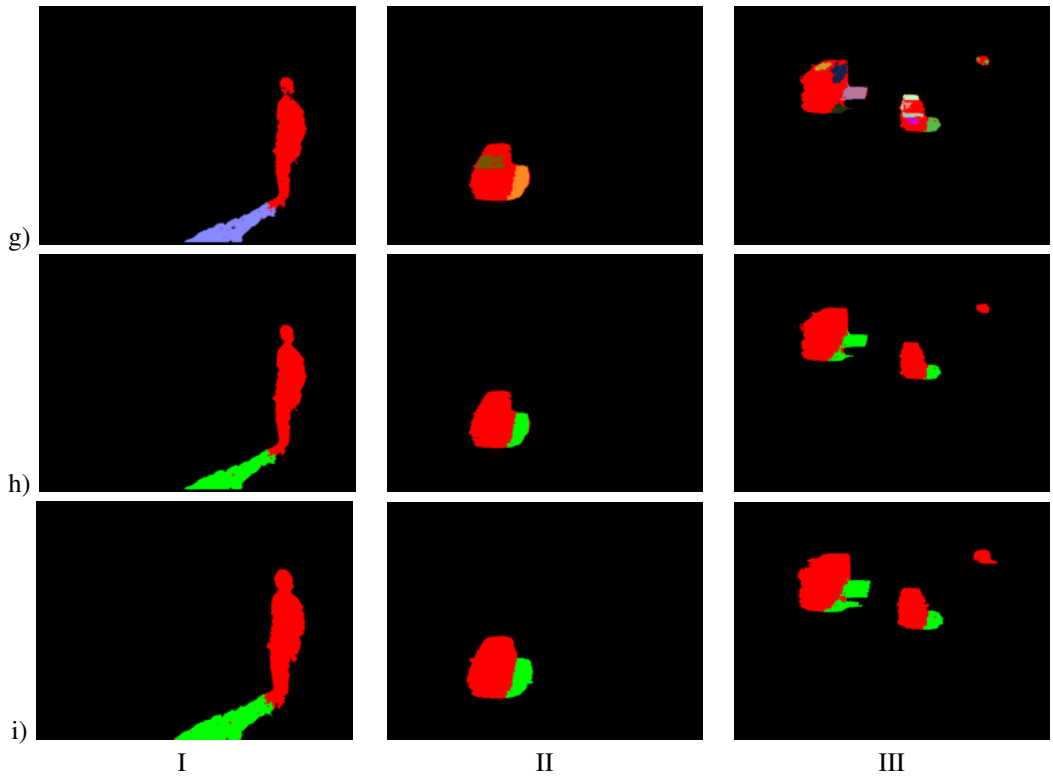


Fig. 9. *(continued)* (g) classification based on the luminance difference and segment size criteria; (h) classification based on the terminal point weight criterion; (i) final segmentation. The sequences used are: (I) Grass field #184, (II) Highway II #157, (III) Highway II #801.

[2] G. Obinata and A. Dutta, *Vision Systems: Segmentation and Pattern Recognition*. I-Tech Education and Publishing, 2007.

[3] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding: CVIU*, vol. 80, no. 1, pp. 42–56, 2000.

[4] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 809–830, 2000.

[6] A. Amato, M. Mozerov, I. H. Casado, J. Gonzàlez, and J. Villanueva, "Background subtraction technique based on chromaticity and intensity patterns," in *19th International Conference on Pattern Recognition (ICPR 2008), December 8-11, 2008, Tampa, Florida, USA*. IEEE, 2008, pp. 1–4.

[7] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 918–923, 2003.

[8] G. Finlayson, DHordley, S. D., and C. Lu, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, 2006.

[9] R. Ramamoorthi, M. L. Koudelka, and P. N. Belhumeur, "A fourier theory for cast shadows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 288–295, 2005.

[10] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *ICCV Frame-Rate WS*. IEEE, 1999.

[11] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with hsv color information," in *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, 2001, pp. 334–339.

[12] N. Martel-Brisson and A. Zaccarin, "Learning and removing cast shadows through a multidistribution approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 7, pp. 1133–1146, 2007.

[13] F. Porikli and J. Thornton, "Shadow flow: a recursive method to learn moving cast shadows," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, 2005, pp. 891–898 Vol. 1.

[14] N. Martel-Brisson and A. Zaccarin, "Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation," in *CVPR08*, 2008, pp. 1–8.

[15] J.-B. Huang and C.-S. Chen, "Moving cast shadow detection using physics-based features," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 2310–2317, 2009.

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

[17] D. Toth, I. Stuke, A. Wagner, and T. Aach, "Detection of moving shadows using mean shift clustering and a significance test," in *International Conference on Pattern Recognition (ICPR 2004)*, vol. 4, 2004, pp. 260–263.

[18] Yang, Lo, Chinag, and Tai, "Moving cast shadow detection by exploiting multiple cues," *Image Processing, IET*, vol. 2, no. 2, pp. 95–104, 2008.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition*. The MIT Press and McGraw-Hill Book Company, 2001.

[20] A. Watt, *3D Computer Graphics*. Addison-Wesley, 1993.

[21] J. Stauder, "Estimation of point light source parameters for object-based coding," *Signal Processing: Image Communication*, vol. 7, no. 4-6, pp. 355 – 379, 1995.

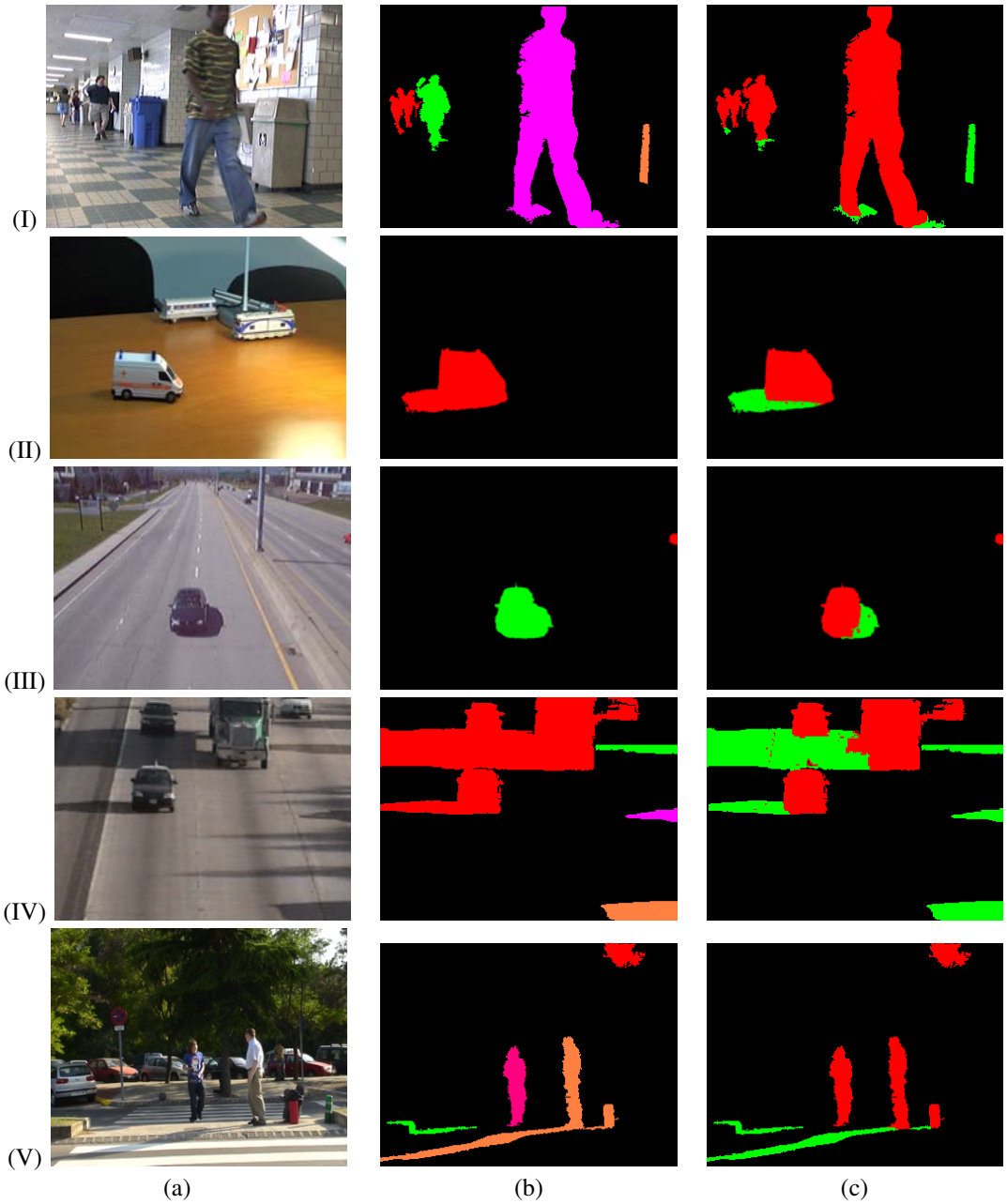


Fig. 10. The result of the implementation of our method in different scenarios: (I) Hallway #163, (II) Auto #1143, (III) Highway II #253, (IV) Highway I #353, (V) CVC outdoor #509. The meaning of each column here is: (a) current image; (b) motion object mask; (c) final classification