Shadow Removal with Morphological Reconstruction

José-Luis Landabaso
Technical University of Catalunya
Barcelona, Spain, e-mail:
jl@gps.tsc.upc.es

Montse Pardàs
Technical University of Catalunya
Barcelona, Spain, e-mail:
montse@gps.tsc.upc.es

Li-Qun Xu
BT Exact Technologies
Adastral Park,
Ipswich, UK

Abstract—In this paper we present the technical advances and system development for shadows and highlights detection in automated video surveillance applications using a hybrid color and texture information system. The paper also includes a novel technique which corrects errors in the image after shadow removal using a reconstruction process.

Coping with shadows and highlights is a crucial challenge in object detection and tracking applications. It is specially relevant in automated surveillance applications or in smart rooms where accurate tracking is very important even under sunset, sunrise and artificial light change situations. Robust approaches make use of background representation to conjecture whether points have similar chrominance in the image as in the background. Other approaches make use of texture information to detect similarities between regions in the background and the image. We present a scheme to combine both of them.

Furthermore, none of the previous studies have considered correcting misclassifications of the shadow removal algorithm using information of the images not shadow-removed where shapes are still well defined. We have developed an algorithm to correct these errors by morphological reconstruction of shadow-removed blobs conditioned to the not shadow-removed ones.

Index Terms—Shadow, highlight, morphology, cast, tracking.

1. INTRODUCTION

Accurate and robust segmentation and tracking of multiple moving objects in dynamic video sequences is one of the major challenges in computer vision. It is particularly relevant in the **video surveillance** (see Fig. 3) field where an automated system allows fast and efficient access to unforeseen events that need to be attended by security guards and law enforcement officers. It is also important for cataloguing and streaming useful information in a video database.

Tracking in **smart rooms** (see Fig. 4) is another situation where shadow removal takes an enormous relevance. This situation is part of a very challenging framework with the goal of freeing people to interact with people and reposition machines to hover in the background, observing the humans –like electronic butlers– attempt to anticipate and serve their needs.[1] And solving low-level problems such as robust tracking is a critical step towards accomplishing this fantastic objective.

1.1 Shadow and highlight invariant tracking

One of the fundamental challenges for accurate tracking is achieving invariance to illumination and more concretely to shadows and highlights.

Regarding shadows, there are two different types of them and they have to be considered differently.

- Cast-shadows are the area in the background projected by the object in the direction of light rays producing inaccurate silhouettes. See a typical scenario with very long cast shadows in the 1st image in Fig. 3.
- Self-shadows are part of the object not illuminated. A good shadow removal scheme must not remove them, as they are part of the silhouette.

With respect to highlights, formally they are areas of lightness in a picture. Objects in the background such as trees should not be detected as new objects when being illuminated by sunrays in cloudy days for instance.

1.2. Related work

Usually, shadows and highlights detection algorithms form part of more general object tracking systems. These object tracking systems divide incoming images into foreground and background representations by means of different techniques (we call incoming images to new images arriving from video stream or storage that are going to be processed).

Sometimes they create a background representation image and subtract the incoming image to detect foreground pixels. In other more robust techniques probabilistic adaptive models are created for every pixel to classify incoming image pixels into foreground or background. Afterwards, connected component analysis (CCA) [5] is usually employed to isolate meaningful blobs from individual foreground pixels. Blobs are then used to extract some representative features and, finally, there is a blob-based matching process that attempts to find blob persistent correspondences between consecutive frames. Further details and a novel approach on this topic will appear in a separate publication.

Shadow removal algorithms are usually incorporated in the background subtraction/modeling step. Several studies have been carried out to extract hints from the background reference images/models to use them later to identify whether a pixel may be a cast shadowed/lightened pixel or

[1] EU Project CHIL (IST-2002-2.3.1.6) where the U.PC. is involved as a partner, is committed to this fundamental shift in the way we use computers, and this work pretends to solve some of the problems involved.

not. Prati et al. have presented an in-depth survey of these algorithms [4].
There are two main set of works which incorporate these extracted clues. A first set uses color information to find chrominance similarities between the background representation and the incoming frame. In the second set of studies, texture similarities are used. Combination of both information is still and open issue.
But even combining these two approaches, shadow removal algorithms tend to be somewhat noisy and often misclassify foreground pixels. In order to correct these situations we propose using images not shadow-removed where shapes are still well defined to assist blob reconstruction. Up-to-date, none of these shadow and highlights removal algorithms have made use of a similar idea to correct errors derived from these pixel based operations.

The paper is structured as follows. In the next section the techniques for pixel-domain analysis leading to the segmented foreground object blobs are described including an overview to Stauffer and Grimson's approach [1], suppression of falsely detected foreground pixels technique, and the extraction procedure of a background image. Section 3 is devoted to discussion on issues concerning colour and texture-based shadow detection. Combination of both techniques is explained in Section 4 where a novel morphological foreground reconstruction technique is also presented. Finally, Section 5 illustrates the experimental evaluations of the system. The paper concludes in Section 6.

2. LEARNING THE BACKGROUND

Background learning techniques are very useful to achieve accurate and robust foreground objects segmentation in a dynamic scene. There are techniques in which an explicit reference image is first generated to be used in the "background subtraction" process. New approaches perform a classification of every pixel based on a pixel-wise probabilistic model so that the explicit subtraction step is skipped.
The Stauffer and Grimson (S&G) [1] algorithm has become a reference in the area of probabilistic classification of background and foreground. In this section, we first outline this technique, and then explain how to adapt it to handle cast shadows and highlights. It is noted, however, that the method we propose doesn't depend on this particular technique and could be applied to any other background learning algorithm where a background reference image can be obtained. We will describe a simple procedure to extract this background image from the S&G models to illustrate how this could be adapted to the other modeling schemes.

2.1 The Stauffer and Grimson (S&G) algorithm

The main idea of S&G algorithm is to model the photometric variations of each pixel along the time course by a mixture of K Gaussian distributions per pixel. Different Gaussians are assumed to characterize different color appearances in every pixel, and each Gaussian is weighted (w) depending on how often the Gaussian has explained the same appearance. Using multiple Gaussians guarantees that repetitive moving background as in tree leaves can be represented by different probabilistic functions.
An incoming pixel is considered to be explained by a Gaussian distribution if its color value is within say 2.5 standard deviations of the distribution mean. Basically, this is the same as in any clustering process.
Then, every time a distribution explains an incoming pixel, the variance (σ^2) and mean (μ) of the Gaussian are updated as in (1).

$$\begin{aligned} \mu_t &= (1-\rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 &= (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_{t-1})^T(X_t - \mu_{t-1}) \end{aligned} \quad (1)$$

,where ρ is the Gaussian adaptation learning rate.

By updating the mean and the variance we allow the system to adapt to slow illumination changes. The weight w_t associated to each Gaussian component is also updated depending on if the Gaussian explains the incoming pixel or not as in (2).

$$\begin{aligned} w_t &= w_{t-1} + \alpha(1 - w_{t-1}) && \textit{matched} \\ w_t &= (1-\alpha)w_{t-1} && \textit{non-matched} \end{aligned} \quad (2)$$

,being α the weight learning rate.

Thus, the more a Gaussian explains an incoming pixel, the highest its associated weight.
In order to classify an incoming pixel as being part of the foreground or background, the Gaussians of each pixel are reordered according to w/σ into descending order. The first few Gaussians in this list correspond to the ones with more supporting evidence (more times explaining incoming pixels) at the lowest variance (explained incoming pixels are always very similar). In other words, these first few most likely represent the background as the background is often very static (low variance) and it's seen during most of the time (high weight w). Analogously, the foreground incoming pixels correspond to the last Gaussians in the list. This can be formulated as following: When a pixel matches any of the first B distributions decided by (3), it will be classified as background, otherwise, foreground.

$$B = \arg\min\left(\sum_{k=1}^{b} w_k > T\right) \quad (3)$$

2.2 Suppression of falsely detected foreground pixels

Although the S&G background learning is very robust there remain classification errors due to the noise manifested in the images. On certain occasions, some background points fail to match their Gaussian and are classified as foreground. Research has been carried out to overcome this well-known problem [2]. Although typical post-processing techniques often depend on the background learning technique employed, a more general approach using local neighborhood information is introduced here. The proposal is that, when a pixel is classified as foreground, it is again examined by its 3x3 spatial neighboring pixel models. If 5 or more models agree on that it's a background pixel, then it's considered as a false detection. By means of this simple rule many small errors are automatically corrected and system operation is more robust.

2.3 Extracting a background reference image

Up till the discussion so far, a background reference image is never explicitly required as the classification of foreground pixels in the scene is directly performed in the incoming image. However, the shadow removal techniques often require a background reference image as the properties between the shadowed regions and the corresponding background are to be examined in conjunction.

For such purpose, a simple procedure is used to extract an adaptive background image using the S&G algorithm. The background pixels are obtained as follows: The pixel colors in the background image assume those of the incoming image if they are classified as background. In the case that the incoming pixels have been classified as foreground, then the mean of the Gaussian distribution with the largest weight at the lowest variance (the most probable background color in the pixel) is chosen as the background pixel color.

In summary, the background learning algorithm described is very robust but it doesn't handle local illumination problems such as shadows and highlights, leading to inaccurate foreground object segmentation. How to effectively deal with these problems is the subject of the following discussions.

3. COLOR- & TEXTURE-BASED SHADOW DETECTION

A shadow is normally an area that is not or only partially irradiated or illuminated because of the interception of radiation by an opaque object between the area and the source of radiation. Assuming that the irradiation consists only of white light, the chromaticity in a shadowed region should be the same as when it is directly illuminated. The same also applies to lightened areas in the image.

Based on the same assumption, a normalized chromatic color space, $r = R/(R+G+B)$, $g = G/(R+G+B)$, for instance, is immune to shadows, but the lightness information is unfortunately lost. Keeping lightness information is important in order to avoid some simple errors such as confusing a white car with a grey road.

Another important issue is that we are only interested in detecting shadows that form part of the foreground objects. Shadows that form part of the background are not a problem as they don't have to be tracked. Specifically, a shadow removal algorithm needs to analyze foreground pixels and detect those that have similar chromaticity but lower brightness to the corresponding region when it is directly illuminated. The adaptive background reference image provides the needed information.

3.1 Color-based detection

Based on the fact that both brightness and chromaticity are very important, a good distortion measure between foreground and background pixels has to be decomposed into its brightness and chromaticity components as in [3]. Brightness distortion (BD) can be defined as a scalar value that brings expected background close to the observed chromaticity line. Similarly, color distortion (CD) can be defined as the orthogonal distance between the expected color and the observed chromaticity line. Both measures are shown in Fig. 1 and formulated in (4).

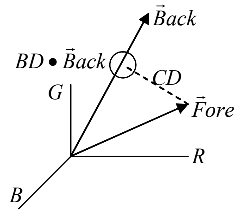

Fig. 1. Distortion measurements in the RGB color space. $\vec{F}ore$ denotes to the RGB value of a foreground pixel in the incoming frame which has been classified as foreground. $\vec{B}ack$ is that of its background counterpart.

Brightness distortion values over 1.0 correspond to lighter foreground. On the other hand, the foreground is darker when BD is below 1.0.

$$\begin{aligned} BD &= \arg\min_{\alpha}\left(\vec{F}ore - \alpha\vec{B}ack\right)^2 \\ CD &= \left\|\vec{F}ore - \alpha\vec{B}ack\right\| \end{aligned} \tag{4}$$

The brightness distortion can be easily obtained by computing the derivative of the first expression, *i.e.* $BD = \vec{F}ore \bullet \vec{B}ack / \vec{B}ack^2$.

Finally, a set of thresholds can be defined to assist the classification into foreground, highlighted or shadowed pixel.

If CD < 10.0 then: *If 0.5 < BD < 1.0 then SHADOW*   *Else if 1.0 < BD < 1.25 then HIGHLIGHT* *Else FOREGROUND*

Table 1. Thresholds for shadow and highlight detection.

It is still possible to achieve more precise results by normalizing variations in color bands increasing computational cost.

Many other approaches as [2] are also based on the same underlying idea of decomposing color and brightness. Our reconstruction process doesn't rely on any particular implementation so any approach can be used.

The last thing to mention is that the technique fulfils its objective not to remove self shadowed regions as they do not share similar brightness and chromaticity with the background reference image.

3.2 Texture-based detection

The same regions with or without cast shadows should have the same texture properties. Similar to the color based shadow removal; a texture distortion measure can be defined to detect possible foreground shadow pixels.

A simple way of computing the texture is to use the first-order spatial derivatives, though other more sophisticated measures can also be employed. We apply X and Y Sobel filters to both the background and incoming frame and then compute the Euclidean distance between them. If this distance is lower than a certain threshold, *i.e.* very similar texture, then the pixels are probably part of a shadowed region.

4. HYBRID SHADOW REMOVAL

The color- and texture-based shadow removal techniques suffer from weaknesses of their own. The color based algorithm generates errors when the underlying assumptions are violated, meaning that foreground objects having similar colors to that of the shadowed background regions may be wrongly diagnosed and removed. Similarly with the texture based approach, the foreground regions having similar textures to that of their corresponding background may also be deleted by mistake.

In our approach, both the color and texture-based procedures discussed above are used in parallel, followed by an assertion process that combines the results of the two, *i.e.*, the pixels are confirmed as shadows if and only if the result of both the two approaches corroborates. This process paves the way for the proposed foreground object shape reconstruction process.

4.1 Foreground reconstruction

The cast shadow/highlights removal algorithm is a destructive process in the sense that, despite the assertion process described above, original object shapes are likely distorted and some pixels will remain misclassified. Mathematical morphology theory can be employed in order to reconstruct the original image without cast shadow or highlights.

Mathematical morphology reconstruction filter uses an image called "marker" image as a mark to rebuild an object inside in an original image called "mask" image. In our case the "marker" image (Fig. 3c) is a binary image where a pixel is set at "1" when it corresponds to a foreground, not cast shadow/highlight pixel. On the other hand, the "mask" image (Fig. 3b) is also a binary image where a "1" pixel can correspond to a foreground pixel, or cast shadow/highlight pixel, or speckle noise.

It is highly desirable that the "marker" image, $\widetilde{M}$, contains only real foreground object pixels, *i.e.*, not any shadow/highlight pixels so that those regions will not be reconstructed. Therefore, the use of very aggressive thresholds is necessary in the foregoing color-based removal process to assure that all the shadow/highlight pixels are removed. A speckle noise removal filter is also applied to suppress isolated noisy foreground pixels that remain and obtain a good quality "marker" image, $\widetilde{M}$.

The speckle removal filter is also implemented using mathematical morphology operation as shown in (5)

$$\widetilde{M} = M \cap (M \oplus N) \qquad (5)$$

,where M is the binary image generated after shadow removal and assertion process; N denotes the structuring element in Fig. 2 with the origin at the centre:

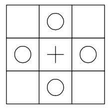

Fig. 2. 3x3 morphological structuring element used for speckles filtering. Note that the origin is not included.

The dilation operation $M \oplus N$ in (5) identifies all the pixels that are four-connected to (*i.e.* next to) a pixel of M. Hence, $\widetilde{M}$ identifies all the pixels that are in M and also have a four-connected neighbor, eliminating the isolated points in M.

(a)

(b)

(c)

(d)

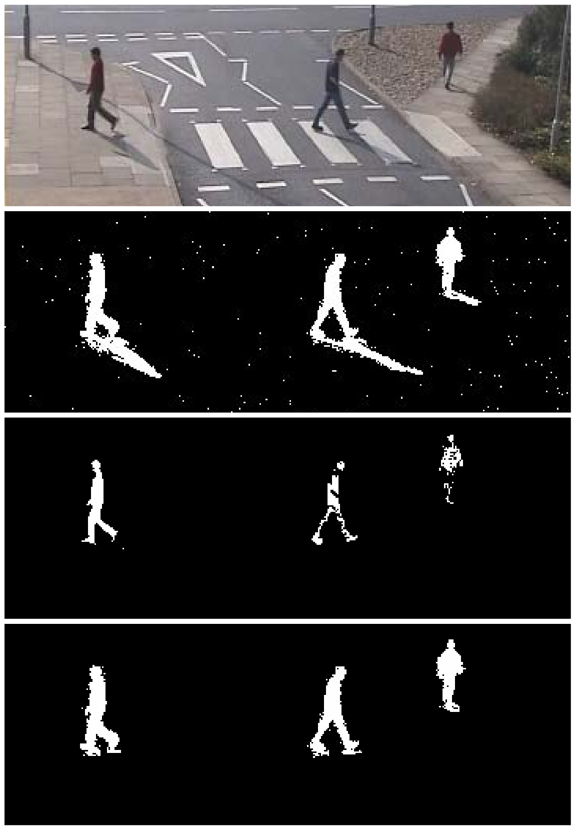

Fig. 3. Illustration of the foreground regions shape reconstruction process after shadows/highlights removal. (a) the incoming image; (b) the "mask" image from foreground segmentation; (c) the "marker" image after shadows/highlights removal; and (d) the final reconstructed objects shapes.

As a result, only the regions not affected by noise which are clearly free of shadows/highlights (Fig. 3c) are subject to the shape reconstruction process shown in (6).

$$R = M_s \cap (\widetilde{M} \oplus SE) \qquad (6)$$

where M_s is the mask, $\widetilde{M}$ the marker and SE the structuring element whose size usually depends on the size of the objects of interest, although a 9 x 9 square element proved to work in all our tests.

Basically this process consists of a dilation of the "marker" image, followed by the intersection with the "mask" image.

The underlying idea is that the shadow removed blobs keep at least a number of points that have been robust to erroneous shadow removal. These robust points are appropriate for leading the reconstruction of neighbouring points as long as they form part of the silhouette in the original blob (previous to the shadow removal as in Fig. 3b). The full reconstruction blobs are showed in Fig. 3d.

5. EXPERIMENTAL RESULTS

The system has been evaluated using the publicly available benchmarking video sequences PETS 2001 and our own recording at BT Adastral Park site. The sequences contain persons, groups of people and vehicles. Results are available at: http://gps-tsc.upc.es/imatge/_jl/Tracking.html

The algorithm performs well except on very large cast shadows where sometimes they are not completely removed. This is mainly due to the fact that brightness decreases below the BD threshold. The problem can be corrected using lower thresholds in the BD with the drawback of introducing false shadow pixel detection.

See for instance in Fig. 3 a real world scenario. These images show the process of shadow removal and reconstruction. First one corresponds to the incoming image. The second one (the mask image) shows the original blobs extracted before any shadow removal attempt. Following, the “marker image” obtained after applying the color based shadow removal and texture based assertion is shown. See in the last picture the reconstructed image calculated by using the “marker image” as the mark and the image with original extracted blobs as the “mask image”.

A small inconvenience of the algorithm is that the reconstructed image presents a wrongly reconstructed segment of shadow in the extremes where the cast shadow starts (see the feet of the persons in Fig. 3). This segment has 1/2 size of the structuring element and is produced during the dilation. Intersection with the mask image cannot suppress the segment as all the shadowed regions form part of the mask.

A similar situation in an indoors scenario is shown in Fig. 4.

6. CONCLUSION

In this paper, we have presented a system able to detect and suppress shadows and highlights. The system combines colour and texture information and performs a reconstruction process for superior results.

Some of the directions to take to improve results include using regions instead of isolate pixels both during the texture and colour shadow detection. Also, different heuristics can be examined to not allow wrong reconstructions in the margins of the shadows.

ACKNOWLEDGMENTS

Part of this work was carried out at the Content and Coding Lab, BT Exact, England, based upon work supported by the IST programme of the EU in the project IST-2000-32795 SCHEMA.

This work has also been supported by the project TIC2001-0996 of the Spanish Government.

(a)

(b)

(c)

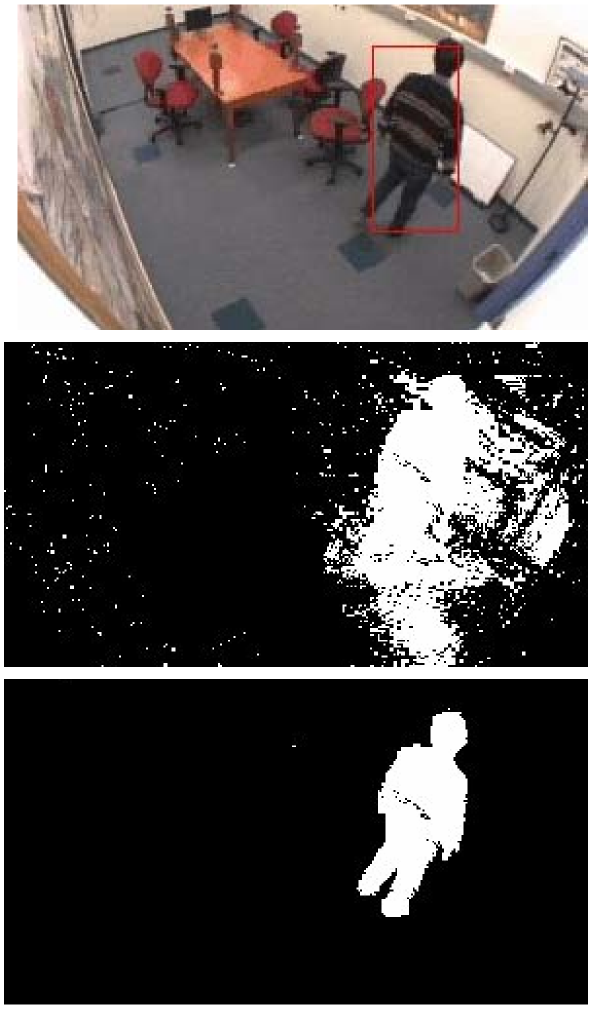

Fig. 4. Illustration of an indoors smart room scenario. (a) the incoming image; (b) the “mask” image from foreground segmentation; and (c) the final reconstructed objects shapes.

REFERENCES

[1] C. Stauffer, W.E.L. Grimson, “Learning Patterns of Activity Using Real-Time Tracking”, IEEE transactions on pattern analysis and machine intelligence, Vol. 22, No. 8, 2000

[2] A Elgamal, R Duraiswami, D. Harwood and L. Davis, “Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance”, Proceedings of the IEEE, Vol. 90, No. 7, 2002

[3] T. Horpraset, D. Harwood and L. Davis, “A Statistical approach for real-time robust background subtraction and shadow detection”, IEEE ICCV’99 FRAME-RATE workshop

[4] A. Prati, I. Mikic, M.M. Trivedi and R. Cucchiara, “Detecting Moving Shadows: Algorithms and Evaluation”, IEEE trans. on pattern recogn. and machine intelligence, Vol. 25, No. 7, 2004

[5] K.P. Horn, “Robot Vision”, pp. 299-333.MIT Press, 1986