Dealing with shadows and highlights is essential in object detection and tracking applications such as automated video surveillance systems for outdoors environments where accurate operation is required even under variable weather conditions. In this paper, we present a novel scheme for effective shadows (highlights) detection using both colour and texture cues. Since in any shadow removal algorithm, misclassification errors often occur, resulting in distorted object shapes. The core of this scheme is the use of a technique that is capable of correcting these errors. The technique is based on morphological reconstruction of the shadow-removed blobs conditioned on the blobs prior to the shadow-removal process, in which the object shapes are still well defined along the most part of their contours. The experiments on variety real-world video data demonstrate the favourable performance and robustness.

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 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.

One of the fundamental challenges for accurate tracking is achieving invariance to illumination variations and more concretely to shadows and highlights. There are two types of shadows that should be treated differently:

- Cast-shadows are the area in the background projected by the object in the direction of light rays producing inaccurate silhouettes. (A typical scenario with very long cast shadows is seen in Figure 3a)
- 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 exceptional lightness in an image. Cluttered scenes in the background such as trees should not be detected as new objects when being illuminated by sun lights in cloudy days, for instance.

Usually, shadows and highlights detection algorithms form part of more general object tracking systems. These object tracking systems often first segment incoming images into foreground and background representations by means of different techniques. In a controllable indoor scene a background representation image can be created in advance. Then the foreground moving pixels can be detected by subtracting the reference image from an incoming image. In other more robust techniques, probabilistic adaptive models are created for every pixel to classify incoming image pixels into foreground or background. Afterwards, a connected component analysis (CCA) [5] is usually employed to isolate meaningful blobs from individual foreground pixels. For each blob some representative features can be extracted to describe its spatial-temporal properties. Finally, there is a blob-based feature matching process in order to find persistent blob correspondences between consecutive frames. An example of object tracking systems can be found in [6].

Shadow removal algorithms are usually incorporated in the background subtraction/modelling 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 is a cast shadowed/lightened pixel or 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 colour 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 sources of information is still an 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 to use images prior to the shadow-removal process 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 outline on Stauffer and Grimson’s approach [1], suppression of falsely detected foreground pixels technique, and the extraction procedure of a background image. Section 3 discusses issues concerning colour and texture-based shadows detection. Combination of both techniques is explained in Section 4 where a novel morphological foreground reconstruction technique is presented. Finally, Section 5 explains the experimental evaluations issues. 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.

SHADOW REMOVAL WITH BLOB-BASED MORPHOLOGICAL RECONSTRUCTION FOR ERROR CORRECTION

Li-Qun Xu†, José Luis Landabaso‡, Montse Pardàs‡

† BT Exact, Adastral Park, Ipswich, UK
‡ Technical University of Catalunya (UPC)

ABSTRACT

Dealing with shadows and highlights is essential in object detection and tracking applications such as automated video surveillance systems for outdoors environments where accurate operation is required even under variable weather conditions. In this paper, we present a novel scheme for effective shadows (highlights) detection using both colour and texture cues. Since in any shadow removal algorithm, misclassification errors often occur, resulting in distorted object shapes. The core of this scheme is the use of a technique that is capable of correcting these errors. The technique is based on morphological reconstruction of the shadow-removed blobs conditioned on the blobs prior to the shadow-removal process, in which the object shapes are still well defined along the most part of their contours. The experiments on variety real-world video data demonstrate the favourable performance and robustness.

One of the fundamental challenges for accurate tracking is achieving invariance to illumination variations and more concretely to shadows and highlights. There are two main set of works which incorporate these extracted clues. A first set uses colour 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 sources of information is still an 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 to use images prior to the shadow-removal process 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 outline on Stauffer and Grimson’s approach [1], suppression of falsely detected foreground pixels technique, and the extraction procedure of a background image. Section 3 discusses issues concerning colour and texture-based shadows detection. Combination of both techniques is explained in Section 4 where a novel morphological foreground reconstruction technique is presented. Finally, Section 5 explains the experimental evaluations issues. 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 the necessary steps to take, prior to handling cast shadows and highlights removal, including the suppression of falsely detected foreground pixels and extraction of a reference background image. Note that the method we will introduce does not depend on this particular background learning technique. It is equally applicable to any other background learning algorithm where a background reference image can be obtained.

### 2.1 The Stauffer and Grimson 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. Different Gaussians are assumed to characterise different colour appearances in each pixel, and each Gaussian is weighted ($w$) depending on how often the Gaussian has explained the same appearance. Using multiple Gaussians ensures 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 colour 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 Gaussian explains an incoming pixel, its variance ($\sigma^2$) and mean ($\mu$) are updated as in (1).

$$
\begin{align*}
\mu_i &= (1 - \rho) \mu_{i-1} + \rho X_t \\
\sigma^2_i &= (1 - \rho) \sigma^2_{i-1} + \rho (X_t - \mu_{i-1}) (X_t - \mu_{i-1})
\end{align*}
$$

where $\rho$ is the Gaussian adaptation learning rate.

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

$$
\begin{align*}
w_t &= w_{t-1} + \alpha [1 - w_{t-1}] & \text{matched} \\
w_t &= (1 - \alpha) w_{t-1} & \text{non-matched}
\end{align*}
$$

$\alpha$ being the weight learning rate.

Thus, the more often a Gaussian explains an incoming pixel, the higher is 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 $w/\sigma$ in 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 is seen most of the time (high weight $w$). Analogously, the incoming foreground pixels correspond to the last Gaussians in the list.

This can be formulated as follows: When a pixel matches any of the first $B$ distributions decided by (3), it will be classified as a background pixel, otherwise, a foreground pixel.

$$
B = \arg \min \left\{ \sum_{i=1}^{k} w_i > T \right\}
$$

### 2.2 Suppression of falsely detected foreground pixels

The S&G background learning is very robust, though 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 neighbourhood information is introduced here. The proposal is that, when a pixel is classified as foreground, it is again examined by its 3x3 spatial neighbouring 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

Since the classification of foreground pixels in the scene is directly performed on incoming images, so far an explicit background reference image is not required. However, the needs arise in shadow removal techniques where the properties of 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 colours 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 colour in the pixel) is chosen as the background pixel colour.

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 normalised chromatic colour 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 analyse 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 Colour-based detection

Based on the fact that both brightness and chromaticity are very important, a good distortion measure between foreground and background pixels should account for the discrepancies in both their 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, colour distortion (CD) can be defined as the orthogonal distance between the expected colour and the observed chromaticity line. Both measures are shown in Figure 1 and formulated in (4).

\[ BD = \arg \min \left( \text{Fore} - a \cdot \text{Back} \right) \]

\[ CD = \left| \text{Fore} - a \cdot \text{Back} \right| \] (4)

Figure 1. Distortion measurements in the RGB colour space: \( \text{Fore} \) denotes the RGB value of a pixel in the incoming frame that has been classified as foreground. \( \text{Back} \) is that of its counterpart in the background.

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

\[ CD = \left| \text{Fore} - a \cdot \text{Back} \right| \]

The brightness distortion can be easily obtained by computing the derivative of the first expression, i.e. \( BD = \text{Fore} \cdot \text{Back} / \text{Back}^2 \).

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

<table>
<thead>
<tr>
<th>Condition</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CD &lt; 10.0 )</td>
<td>END detection</td>
</tr>
<tr>
<td>( 0.5 &lt; BD &lt; 1.0 )</td>
<td>SHADOW</td>
</tr>
<tr>
<td>( 1.0 &lt; BD &lt; 1.25 )</td>
<td>HIGHLIGHT</td>
</tr>
<tr>
<td>Otherwise</td>
<td>FOREGROUND</td>
</tr>
</tbody>
</table>

Table 1. Thresholds for shadow and highlight detection.

Note that it is still possible to achieve more precise results by normalising variations in colour bands at the expense of increased computational cost. Also, many other approaches as [2] are based on the same underlying idea of decomposing colour and brightness. Our reconstruction process to be described in Section 4 does not 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 colour based shadow removal procedure, a texture distortion measure can be defined to detect possible foreground shadow pixels as well.

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 colour- and texture-based shadow removal techniques suffer from weaknesses of their own. The colour based algorithm generates errors when the underlying assumptions are violated, meaning that foreground objects having similar colours 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 colour 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 (Figure 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 (Figure 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, \( M \), 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, \( \hat{M} \).

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

\[ \hat{M} = M \wedge (M \oplus N) \] (5)

where \( M \) is the binary image generated after shadow removal and assertion process, \( N \) denotes the structuring element in Figure 2 with the origin at the centre:

\[ \begin{array}{ccc}
\text{O} & + & \text{O} \\
\text{O} & + & \text{O}
\end{array} \]

Figure 2. The 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, $M$ identifies all the pixels that are in $M$ and also have a four-connected neighbour, eliminating the isolated points in $M$.

![Figure 3](image_url)

**Figure 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 from shadows/highlights (Figure 3c) are subject to the shape reconstruction process shown in (6):

$$R = M_1 \cap (M \ominus SE)$$

(6)

where $M_1$ is the mask, $M$ the marker and $SE$ the structuring element whose size usually depends on the size of the objects of interest, although a $9 \times 9$ 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 (prior to the shadow removal as in Figure 3b). The full reconstruction blobs are showed in Figure 3d.

5. EXPERIMENTAL RESULTS

The algorithm performs well in our experiments on various outdoor scenarios and recordings except for 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. An example of the algorithm in action is shown in Figure 3 on a real world scenario. First one corresponds to the incoming image. The second one (the mask image) shows the initial foreground blobs extracted before any shadow removal attempt. Following, the “marker image” obtained after applying the colour based shadow removal and texture based assertion is shown. The last one gives the reconstructed image calculated, using the “marker image” as the mark and the image with original extracted blobs as the “mask image”.

A small defect of the algorithm is that the reconstructed image contains a wrongly reconstructed segment of shadows in the objects exteriors where the cast shadow starts (see the feet of the persons in Figure 3d). This segment has 1/2 size of the structuring element used, 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.

Finally, this novel scheme has been incorporated in our object tracking system, which has been evaluated broadly using the publicly available benchmarking video sequences PETS 2001 and our own recordings. The sequences contain persons groups of people and vehicles. Some results can be seen from URL: http://gps-tsc.upc.es/imatge/_jl/Tracking.html.

6. CONCLUSION

We have presented in this paper a novel scheme for effective shadows and highlights detection, which has been successfully incorporated in an object tracking system. The scheme exploits information from both colour and texture cues between an incoming image and an adaptive background reference, and performs an error correction procedure to recover original object shapes using conditional morphological reconstruction process. Experiments have demonstrated favourable results on various real-world scenes on both raw and compressed image sequences. Some of the future works to improve further the current results include using region-based instead of pixel-based domain processing in both the texture and the colour-based shadow detection. Also, different heuristics can be examined to prohibit the minor reconstructions of shadows in objects exteriors.

ACKNOWLEDGMENTS

This work was performed at the Content and Coding Lab, BT Exact, UK, where JLL was supported by a BT Student Internship, in connection with the EU Framework V SCHEMA project (IST-2000-32795). JLL also acknowledges the bursary from Spanish national project grant number TIC2001-0996.

REFERENCES


