# Improved Shadow Removal for Robust Person Tracking in Surveillance Scenarios

Andres Sanin, Conrad Sanderson, Brian C. Lovell
*NICTA, PO Box 6020, St Lucia, QLD 4067, Australia* *
*The University of Queensland, School of ITEE, QLD 4072, Australia*

## Abstract

*Shadow detection and removal is an important step employed after foreground detection, in order to improve the segmentation of objects for tracking. Methods reported in the literature typically have a significant trade-off between the shadow detection rate (classifying true shadow areas as shadows) and the shadow discrimination rate (discrimination between shadows and foreground). We propose a method that is able to achieve good performance in both cases, leading to improved tracking in surveillance scenarios. Chromacity information is first used to create a mask of candidate shadow pixels, followed by employing gradient information to remove foreground pixels that were incorrectly included in the mask. Experiments on the CAVIAR dataset indicate that the proposed method leads to considerable improvements in multiple object tracking precision and accuracy.*

## 1. Introduction

Modern tracking algorithms typically require that the objects to be tracked are segmented prior to tracking (e.g. particle filters [9]) or while tracking (e.g. blob matching [8]). Existing foreground segmentation algorithms, such as background subtraction or detection of foreground pixels using Gaussian mixture models [2], can be used for this task. However, current techniques typically have one major disadvantage: shadows tend to be classified as part of the foreground. This is due to shadows in many situations having a similar magnitude of intensity change as that of the foreground objects [6].

Since cast shadows can be as big as the actual objects, their incorrect classification as foreground decreases tracking performance. Example scenarios in which tracking performance is affected include: **(i)** several people are merged together because of their cast shadows, **(ii)** the inclusion of shadow pixels decreases the reliability of the appearance model for each person, increasing the likelihood of tracking loss. As such, removing shadows has become an important step in the implementation of robust tracking systems [5].

*Acknowledgements:** NICTA is funded by the Australian Government as represented by the *Department of Broadband, Communications and the Digital Economy* as well as the Australian Research Council through the *ICT Centre of Excellence* program.

Prati *et al* [7] evaluated the usefulness of several shadow detection methods using the following two metrics, which indicate the shadow detection rate ($\eta$) and the shadow discrimination rate ($\xi$):

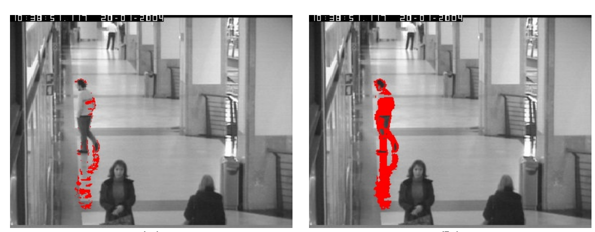
$$\eta = \frac{TP_S}{TP_S + FN_S} \quad \text{and} \quad \xi = \frac{TP_F}{TP_F + FN_F} \qquad (1)$$

Here $TP$ and $FN$ stand for true positive and false negative pixels with respect to either shadows ($S$) or foreground objects ($F$). The shadow detection rate is concerned with labelling the maximum number of cast shadow pixels as shadows. The shadow discrimination rate is concerned with maintaining the pixels that belong to the moving object as foreground. Current shadow detection methods present a compromise between the two rates [7].

In this paper we propose a shadow detection method that is able to achieve both good shadow detection and discrimination rates, leading to improved tracking in surveillance scenarios. The proposed method first uses chromacity information to create a mask of candidate shadow pixels, followed by employing gradient information to remove foreground pixels that were incorrectly included in the mask.

A further contribution of this work is the manually obtained shadow ground-truth data[1] for the CAVIAR dataset[2], allowing for the objective evaluation and comparison of shadow removal algorithms.

[1] Shadow ground-truth data: *http://itee.uq.edu.au/~uqasanin/*
[2] CAVIAR dataset: *http://homepages.inf.ed.ac.uk/rbf/CAVIAR/*



**(a)** **(b)**

**Figure 1.** Detection rate and discrimination rate trade-off in chromacity based shadow detection (the original scene is shown in greyscale rather than in colour for ease of interpretation only). Red areas indicate pixels classified as being part of the person's shadow. Higher thresholds used in **(a)** reduce the shadow detection rate. Lower thresholds used in **(b)** increase the detection rate but reduce the discrimination rate.

We continue the paper as follows. In Section 2 we overview pertinent existing shadow detection algorithms, leading to Section 3 where we detail our proposed method. Section 4 shows two experiments: **(i)** comparison of shadow detection and discrimination rates against existing techniques, **(ii)** evaluation of the proposed method's effect on tracking performance. In Section 5 we provide the main findings.

## 2. Related Work

Cucchiara *et al* [3] proposed a method based on the assumption that shadow pixels in the given frame are darker compared to the background reference image while having similar chromacity. The HSV colour space is used since it gives a better separation of chromacity and intensity. A pixel $p$ is considered to be part of a shadow if:

$$\alpha \leq \left(F_p^V / B_p^V\right) \leq \beta, \tag{2}$$

$$\left(F_p^S - B_p^S\right) \leq \tau_S \quad \text{and} \quad \left|F_p^H - F_p^H\right| \leq \tau_H \tag{3}$$

In the above equations, $F_p^C$ and $B_p^C$ represent the component values, $C$, of HSV for the pixel position $p$ in the frame ($F$) and in the background reference ($B$) image, respectively. $\alpha$, $\beta$, $\tau_S$ and $\tau_H$ represent thresholds that are optimised empirically.

The main limitation of this method is its sensitivity to the selection of thresholds. If the thresholds are too restrictive (high) it will result in a low detection rate (Fig. 1a). If the thresholds are made more permissive (low), in cases where regions of the object have similar chromacity to that of the background, they will be misclassified as shadows resulting in a low discrimination rate (Fig. 1b). By using only one feature (chromacity in this case), it is difficult to have both good shadow detection and discrimination rates at the same time.

Although chromacity based methods are the most common in the literature, other features have been successfully used in shadow detection. Javed and Shah [4] proposed a shadow detection mechanism based on a different assumption: shadow regions are darker compared to the background reference but they maintain the same texture. Their shadow detection method has four steps as follows. **(i)** Pixels that are darker in the frame than in the background reference are selected and their gradient direction is calculated using:

$$\theta = \arctan\left(\bigtriangledown_y / \bigtriangledown_x\right) \tag{4}$$

where the $\bigtriangledown_y$ is the vertical gradient (difference in intensity between the pixel and the pixel in the next row) while $\bigtriangledown_x$ is the horizontal gradient. The gradient direction is a simple way of determining the texture at the pixel level. **(ii)** All selected pixels are segmented by colour, resulting in several candidate shadow regions. **(iii)** The gradient direction of the pixels in each region in the foreground is correlated with the corresponding region in the background. **(iv)** Regions with high gradient direction correlation ($> 0.75$) are considered to be cast shadow regions.

The major drawback of the above method is that colour segmentation is used to produce the candidate shadow regions. In most cases this will break the textures into smooth regions that are more sensitive to noise, resulting in a lower detection rate.

## 3. Proposed Shadow Detection Method

We propose a new method that uses chromacity and gradient features to achieve high shadow detection and discrimination rates at the same time. The method has five steps: **(1)** pre-selection of shadow pixels based on chromacity invariance, **(2)** grouping of shadow pixels into candidate shadow regions, **(3)** selection of pixels with significant gradient magnitude in each region, **(4)** calculation of the gradient direction distance between the given frame and the background reference image for each selected pixel, **(5)** for each candidate shadow region, correlation of the gradient direction between the given frame and the background image to determine which regions are shadows. The five steps are described in more detail below, and the whole process is illustrated in Fig. 2.

1. The chromacity method proposed by Cucchiara *et al* [3] is used to create a mask of possible shadow pixels. The thresholds are deliberately set to low values in order to increase the likelihood of including all the shadow pixels in the shadow mask (i.e. high detection rate).
2. Connected components are extracted from the resulting mask, with each component corresponding to a candidate shadow region. Unlike the colour segmentation approach used in [4], this approach of producing shadow regions has the advantage of not breaking textures.
3. For each connected component, the gradient magnitude $|\bigtriangledown_p|$ and gradient direction $\theta_p$ at each pixel $p = (x, y)$ are calculated using:

$$|\bigtriangledown_p| = \sqrt{\bigtriangledown_x{}^2 + \bigtriangledown_y{}^2} \tag{5}$$

$$\theta_p = \operatorname{arctan2}\left(\bigtriangledown_y / \bigtriangledown_x\right) \tag{6}$$

   where $\bigtriangledown_x$ and $\bigtriangledown_y$ are calculated as described for Eqn. (4). Function $\operatorname{arctan2}(\cdot)$ returns an angle in the full angular range $[-\pi, \pi]$ allowing the gradient direction to be treated as a true circular variable. Only the pixels with $|\bigtriangledown_p|$ greater than a certain threshold $\tau_m$ are taken into account, in order to avoid the effects of noise (which is stronger in the smooth regions of the image) and to give weight to the pixels near the edges that have more robust information about the textures.
4. For each pixel $p = (x, y)$ that was selected due to significant magnitude, the difference in gradient direction between frame $F$ and background $B$ is calculated:

$$\Delta\theta_p = \arccos\left[\frac{\bigtriangledown_x^F \bigtriangledown_x^B + \bigtriangledown_y^F \bigtriangledown_y^B}{\left\{\left(\bigtriangledown_x^{F^2} + \bigtriangledown_y^{F^2}\right)\left(\bigtriangledown_x^{B^2} + \bigtriangledown_y^{B^2}\right)\right\}^{\frac{1}{2}}}\right] \tag{7}$$

   Since the gradient direction is a circular variable, the difference has to be calculated as an angular distance.

5. The gradient direction correlation between the frame and the background is estimated using:

$$c = \left\{ \sum_{p=1}^{n} \mathrm{H}(\tau_a - \Delta\theta_p) \right\} / n \tag{8}$$

where $n$ is the number of pixels selected in the candidate shadow region and $H(\cdot)$ is the unit step function which, in this case, evaluates to 1 if the angular difference is less than or equal to the threshold $\tau_a$, and 0 otherwise. In essence, $c$ is the fraction of pixels in the region whose gradient direction is similar in both the frame and the background. If $c$ is greater than threshold $\tau_c$, the candidate region is considered a shadow region and it is removed from the foreground mask.

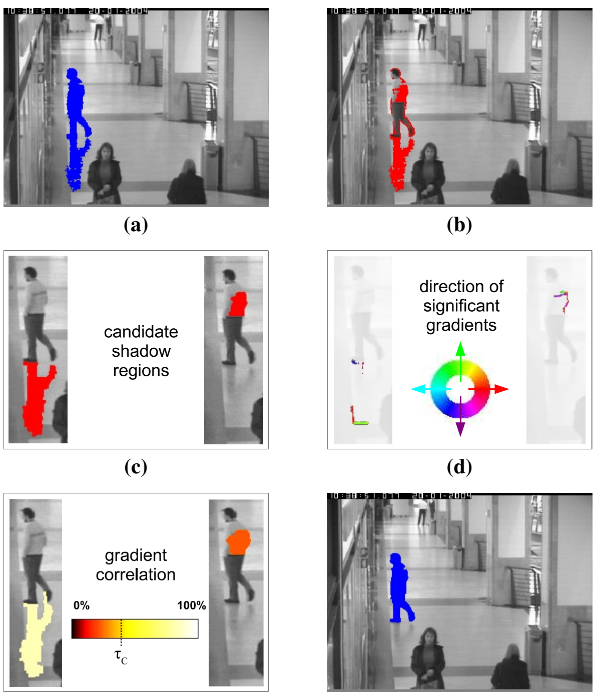



**Figure 2.** Processing in the proposed method: **(a)** foreground obtained after background subtraction, marked in blue; **(b)** pre-selection of shadow pixels based on chromacity; **(c)** candidate shadow regions; **(d)** gradient direction of pixels with significant gradient magnitude; **(e)** gradient direction correlation for each candidate region (only the region in the left pane is above the threshold $\tau_c$); **(f)** final foreground object pixels marked in blue, omitting the shadow.

| **Method** | $\eta$ (%) | $\xi$ (%) | **Time (ms)** |
|---|---|---|---|
| Chromacity | 95.99 | 78.15 | 7.38 |
| Gradients | 66.24 | 88.20 | 9.47 |
| Proposed | 92.05 | 97.85 | 13.12 |

**Table 1.** Shadow detection ($\eta$) and discrimination ($\xi$) rates as well as time per frame on CAVIAR sequences. C++ implementation on an Intel Dual Core 2.6 GHz machine.

# 4. Experiments

We performed two experiments to evaluate the proposed method: **(i)** comparison of shadow detection and discrimination rates against the two methods described in Sec. 2, **(ii)** evaluation of the improvement in tracking performance.

Since the publicly available surveillance datasets do not include shadow ground-truth data which was needed for the first experiment, we manually labelled the moving cast shadows in 25 sequences from the CAVIAR dataset (corridor view clips from the second set). The detection and discrimination rates reported were calculated by averaging the results obtained with all the test sequences. The thresholds for all methods were selected to obtain the best overall performance on the test sequences, using one setting for all sequences (i.e. no specific tuning for each sequence).

Table 1 compares the performance of the chromacity based shadow detection method [3], the gradient correlation based method [4], and our proposed method. The values of the shadow detection rate ($\eta$), shadow discrimination rate ($\xi$) and time per framerepresent the averages obtained on the 25 labelled sequences. The results show that the previous approaches exhibit a significant trade-off between the detection and discrimination rates (e.g. high detection rate at the expense of a reduced discrimination rate). The proposed method, on the other hand, is able to achieve both high detection and discrimination rates (i.e. no trade-off) with only a minor increase in the average amount of time required for processing each frame.

The gradient correlation method in [4] was designed to work in scenes that have rich texture information but uniform chromacity (e.g. grass, asphalt), which are more common in outdoor environments. Since indoor scenes tend to have smoother surfaces, the edges that are more useful for the gradient correlation are produced by colour changes on the background. In this case, using colour segmentation to generate the candidate shadow regions is not a good approach as the textures are broken into smooth segments that do not have enough pixels with significant magnitude. This scenario is solved in our proposed method by using the output of the chromacity based method to produce the candidate shadow regions, which avoids breaking the shadow regions into smooth segments.

In the second experiment we studied the effect of shadow removal on tracking performance. We used the tracking systems implemented in the video surveillance module of OpenCV v2.0 [2] and the tracking ground truth data that is available for the 50 sequences in the second set of the CAVIAR dataset. The tracking performance was measured with the two metrics proposed by Bernardin and Stiefelhagen [1], namely multiple object tracking accuracy (MOTA) and multiple object tracking precision (MOTP).

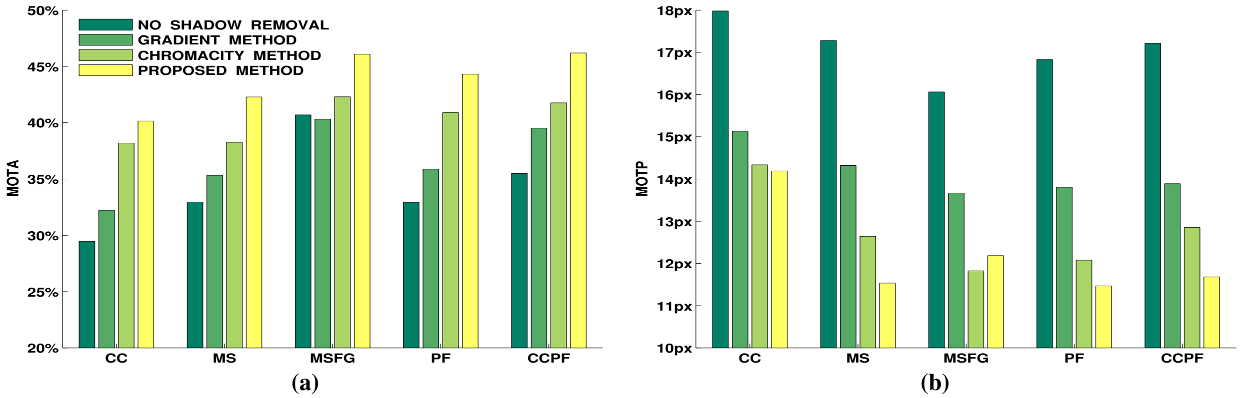



**Figure 3.** Effect of shadow removal methods on: **(a)** multiple object tracking accuracy (MOTA), where taller bars indicate better accuracy; **(b)** multiple object tracking precision (MOTP), where shorter bars indicate better precision. Results are grouped by tracking algorithm: blob matching (CC), two mean shift trackers (MS and MSFG), particle filter (PF) and hybrid tracking (CCPF).

Briefly, MOTP measures the average pixel distance between the ground-truth locations of objects and their locations according to a tracking algorithm. The lower the MOTP, the better. MOTA accounts for object configuration errors, false positives, misses as well as mismatches. The higher the MOTA, the better.

We performed 20 tracking evaluations by combining four shadow removal options (no shadow removal, chromacity based method, gradient based method and the proposed method) with five tracking algorithms (blob matching, mean shift, mean shift with foreground feedback, particle filter and blob matching with particle filter for occlusion handling). The performance result in each evaluation is the average performance of the 50 test sequences. The results, presented in Fig. 3, indicate that in almost all cases the proposed method obtains the best MOTA and MOTP, with considerable improvements over the next best method (chromacity based).

## 5. Discussion & Main Findings

The proposed shadow detection and removal method can be considered as a mixture of the chromacity based method [3] and an improved version of the gradient direction correlation method [4]. Experiments on the CAVIAR datatset suggest that proposed method is able to achieve both high detection and discrimination rates, unlike the chromacity or gradient based methods which exhibit a significant trade-off between the two rates.

Improved shadow removal results in more accurate delineation of foreground objects, which in turn has several important effects from a tracking point of view: **(i)** reduction of the probability of objects merging together, **(ii)** reduction of the amount of false positives caused by fragmented shadows, **(iii)** more accurate appearance models due to less noise. Using objective metrics we have shown that in almost all cases the proposed method results in considerable tracking improvements over the next best method (chromacity based), across a wide range of tracking algorithms.

The proposed method assumes that shadows and foreground objects will break into separate regions. In all our experiments this assumption held true, but under different conditions this could lead to errors in either detection or discrimination. However, the method can be expanded so that the candidate shadow regions are divided using significant edges that occur in the frame but are not present in the background estimation. This extra step can address the limitation of the algorithm at the cost of additional complexity.

## References


[1] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, 2008.

[2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008.

[3] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. In *Proc. Int. Conf. Image Analysis and Processing*, pages 360–365, 2001.

[4] O. Javed and M. Shah. Tracking and object classification for automated surveillance. In *Proc. European Conf. Computer Vision*, volume 4, pages 343–357, 2002.

[5] B. Mitra, R. Young, and C. Chatwin. On shadow elimination after moving region segmentation based on different threshold selection strategies. *Optics and Lasers in Engineering*, 45 (11 ):1088–1093, 2007.

[6] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26 (8 ):1079–1087, 2004.

[7] A. Prati, R. Cucchiara, I. Mikic, and M. Trivedi. Analysis and detection of shadows in video streams: a comparative evaluation. In *IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, pages 571–576, 2001.

[8] Q. Zhou and J. Aggarwal. Object tracking in an outdoor environment using fusion of features and cameras. *Image and Vision Computing*, 24 (11 ):1244–1255, 2006.

[9] S. K. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Trans. Image Processing*, 13 (11 ):1491–1506, 2004.