DETECTING SHADOWS IN IMAGE SEQUENCES

Andrea Cavallaro (+), Elena Salvador (*), Touradj Ebrahimi (*)
(+) Queen Mary, University of London (United Kingdom)
(*) Swiss Federal Institute of Technology (Switzerland)

Abstract

In this paper, we present an algorithm for the detection of local illumination changes due to shadows in real world sequences. The algorithm is designed to be able to work when camera, illumination and scene's characteristics are unknown. This feature is highly desirable for a wide range of applications, such as video production, immersive gaming, and visual surveillance. The algorithm operates as follows. First colour information is exploited, then multiple constraints from physical knowledge are embedded to define the shadow detection algorithm. Colour information is exploited by means of the RGB colour space and by means of photometric invariant features. After colour analysis, a spatio-temporal verification stage is introduced to refine the results. Experimental results show that the proposed algorithm outperforms state-of-the-art methods and can be applied on both indoor and outdoor image sequences.

1 INTRODUCTION

The diffusion of digital video cameras and powerful personal computers is favoring the introduction of authoring tools to the home and corporate markets. Immersive gaming, realistic video conferencing, natural human-computer interfaces, home video and corporate communications are examples of applications that benefit from the advances in digital video analysis and editing technologies. Authoring techniques, such as the use of chroma keying for separation of an anchorperson from background and the subsequent creation of an augmented scene in television production and movie creation, are becoming affordable for the home market. Video editing itself is a potential killer application for the near future [1].

One way to create new and richer content is by extracting natural objects from a scene. The rich content is created by composing a new scene with different objects captured by different sensors (e.g., immersive video conferencing) and mixed with artificial objects (e.g., immersive gaming and rich media presentations). An important shift between the professional and home production is the simplification of the set-up of the scene from which the objects are extracted. Studio production can afford both controlled lighting to avoid shadows that would add noise to the final composition and special cameras with ring of leds, [2], or depth cameras coupling a depth sensor with a traditional camera. The goal is to make the process of extracting natural objects affordable for media production not only for the professional market but also for home and corporate markets. The ideal solution would be using a digital camera to extract characters without the need of ad-hoc scenes or ad-hoc cameras. This simplification leads to the problem of segmenting video objects without using a blue screen. An important problem related to this approach is shadow segmentation [3].

The shadow segmentation problem has been increasingly studied in the past years ([4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25]). However, there is not a generally accepted method to detect shadows in image sequences. As concluded by the review presented in [25], different approaches to shadow detection should be taken when addressing different kind of scenes. Furthermore, many methods need the selection ad-hoc thresholds, thus making the methods unpractical to use. In addition to this, we observed that some methods might even use the same information in a contrasting way as cue to detect shadow. An example is given by the use of Saturation. In [10] shadows are detected when the value of the Saturation decreases, whereas in [17] shadows are detected when the value of the Saturation increases. This contradiction demonstrates the difficulty of defining a general model describing the effect of a shadow. Shadows are in fact a difficult phenomena to model. A shadow does not have its own appearance, but that of the material it is cast upon. Furthermore, a number of factors influence the appearance of a shadow. The shadow segmentation problem is generally faced by embedding multiple constraints when processing a video. The problem is to define the most appropriate constraints and how to embed them in the shadow detection algorithm. In this paper, we propose a shadow segmentation method which is based on colour information and spatio-temporal verification. Colour information is exploited by means of the RGB colour space and by means of photometric invariant features. A spatio-temporal verification stage is then introduced to re-

Printed and published by the IEE, Michael Faraday House, Six Hills Way, Stevenage, Herts SG1 2AY, UK

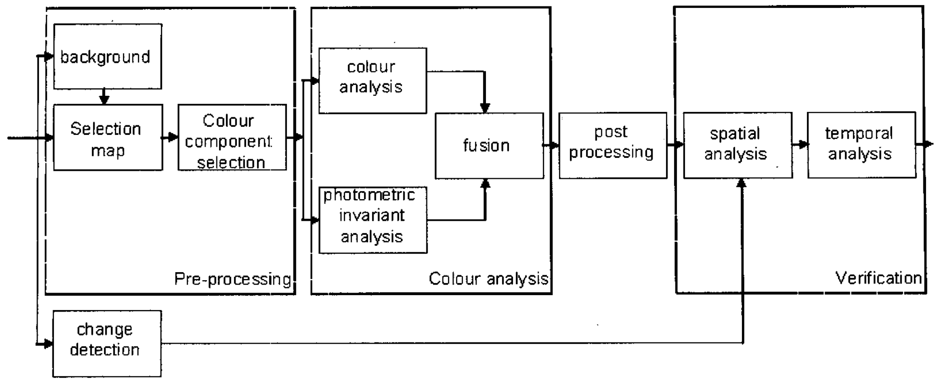

Figure 1: Block diagram of the proposed shadow detection system

fine the results. The proposed algorithm is described in Section 2. Experimental results are discussed in Section 3. Finally, in Section 4 we draw the conclusions.

2 PROPOSED METHOD

Given the conclusions from the previous section, we address the shadow detection problem by exploiting three sources of information, namely colour, spatial, and temporal information. The block diagram of the proposed system is shown in Fig. 1.

Colour information is exploited in two forms, by means of the *RGB* colour space and by means of photometric invariant features. Photometric invariant features are functions describing the colour configuration of each image coordinate discounting local illumination variations, such as shadings and shadows. Examples of photometric invariant features are Hue and Saturation in the HSV colour space, the c_1, c_2, c_3 colour model [22], and the normalised-*RGB* colour space. The normalised-*RGB* colour space, the *rgb* colour system, was chosen for its fast computation since it can be obtained by dividing the R, G and B coordinates by their total sum. The transformation from the RGB coordinates to the normalized colour space is given by

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \tag{1}$$

This transformation projects a colour vector in the RGB cube into a point on the unit plane $r+g+b = 1$. Only two among the three *rgb* values suffice to define the coordinates of the colour point in this plane.

Given the *RGB* and the *rgb* colour spaces, colour information is exploited in a selective way. First, the relevant parts to analyse are identified in each image. Then, the colour components that carry most of the needed information are selected. Finally, spatial and temporal constraints are embedded in the algorithm to verify the results of the colour analysis. The details of the proposed method are described in the following sections.

2.1 PREPROCESSING

The preprocessing stage prepares the input image to be analysed. This stage is composed of two steps, namely the selection map and the colour component selection. The first step selects the areas to analyse and results in a selection map. The selection map identifies the pixels in the image that are suitable for the colour analysis. The computation of the map aims at eliminating the information which is useless or might mislead the detection process. In particular, the selection map identifies the pixels in achromatic parts of the scene. Typically a lower threshold is used to eliminate from the analysis areas with low luminance values [10, 13]. Moreover, when considering *rgb* colour components (see Eq(1)), the invariance is obtained at the cost of singularities and instabilities near the black vertex of the *RGB* cube. To avoid these instabilities, the volume close to the black vertex is eliminated from the analysis process. The shadow selection map eliminates from the subsequent analysis all the pixels whose colour components are smaller that the 20% of the colour dynamic range. The shadow selection map is a binary map indicating which part of a frame should be further processed by the shadow detection process. An example of selection map is shown in Fig. 2. The points identified by the shadow selection map are eliminated from the colour processing and are then treated separately in the post processing.

The computation of the shadow selection map is followed by a second stage that reduces the set of colour components to be considered in the subsequent colour analysis. This module aims at choosing only those components that carry most of the information for the shadow detection task. We observed that a shadow can be better characterised

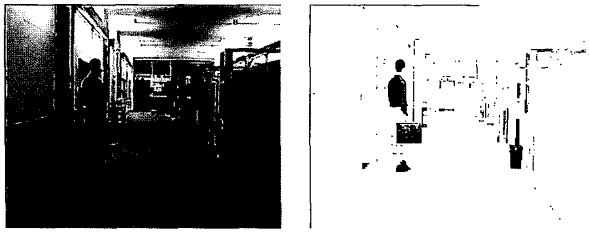

Figure 2: Example of shadow selection map. (a) Frame 37 of the test sequence *Hall Monitor* and (b) corresponding shadow selection map. The black labels identify pixels that are not considered in the colour analysis stage

by analysing the behavior of the colour components with the larger values. This conclusion can guide the algorithm in eliminating from the analysis those colour features that do not add relevant information for the detection process. The smallest component is therefore eliminated from the analysis. The rationale behind this stage is that a colour component with a small value is highly influenced by the effects of reflections and ambient lighting and it is less discriminative for detecting shadows.

To conclude, the pre-processing stage is based on the following observations: (1) not all areas of the image are suitable for shadow detection (e.g., dark areas are likely to mislead the detection process), (2) the photometric invariants are unreliable for points near the black vertex of the RGB cube, (3) the selection of the colour on which to perform/not perform the analysis increases the reliability of the detection. Each pixel (x, y) belonging to the selection map is finally described by a feature vector $\mathbf{f}(x, y)$ representing the value of the colour features $c_i(x, y)$ that have been selected through the colour selection stage. The feature vector can be represented as

$$\mathbf{f}(x,y) = \left(c_1(x,y), c_2(x,y), c_3(x,y), c_4(x,y)\right), \quad (2)$$

where $c_1(x, y)$ and $c_2(x, y)$ are the selected components from the RGB space, and $c_3(x, y)$ and $c_4(x, y)$ are the selected components from the rgb space. The feature vector $\mathbf{f}(x, y)$ is analysed as described in the following stage.

2.2 COLOUR ANALYSIS

Shadows cannot be defined by a specific colour appearance. However, it is possible to characterise a shadow by considering its effect on the colour appearance of the region on which it is cast. To exploit this property of shadows, each frame of a video sequence is compared to a reference frame in order to verify the presence or absence of a shadow. The comparison between the current and the reference frame is based on colour information. Colour analysis is performed in order to identify those pixels in the image that respect the chromatic property of a shadow. Comparing the incoming video frame with a reference frame is a widely used approach in the related literature. For the applications addressed in this paper, it is reasonable to assume that a reference image is available, either as a snapshot of the scene or as a model resulting from a learning process [24]. The advantage of using a reference image representing the background compared to the alternative of using the previous frame is that it is possible to avoid the dependence on objects speed.

Colour information is exploited at this stage by means of the traditional colour components of the RGB space and by means of the photometric invariant features of the rgb space. The effect of a shadow is the darkening of each point on which it is cast (Fig. 3). Let (x', y') be a background pixel and (x, y) the corresponding pixel in the image under test. The test in the RGB space is defined by

$$c_1(x',y') > c_1(x,y) \quad \wedge \quad c_2(x',y') > c_2(x,y). \quad (3)$$

The test (3) is satisfied by a shadow, but can also be satisfied by an object. The colour information in the RGB space needs to be complemented by additional constraints. These constraints should help in identifying a shadow from an object which is darker than the corresponding background. The first constraint stems from the empirical observation that the colour components do not change their order when a shadow occurs (Fig. 4). The second constraint is based on exploiting the property of photometric invariant features. Photometric invariant features do not change their value when an illumination change, such as a shadow, occurs; whereas they are likely to change their value in case of a material change (Fig. 4). This constraint can be represented by the following condition

$$\left(c_3(x,y), c_4(x,y)\right) \cong \left(c_3(x',y'), c_4(x',y')\right). \quad (4)$$

The results of the colour analyses on the RGB and rgb spaces are then fused to produce the shadow map. The fusion is the logical AND between the two partial results. The shadow map is then post processed as described in the following section.

2.3 POST PROCESSING AND VERIFICATION

Colour information alone is not discriminative enough to allow for reliable shadow detection. Other constraints can be embedded in the algorithm based on contextual information and the spatial nature of shadow (cast by an object). For this reason, after colour analysis, the shadow map is first post processed based on morphology and then it undergoes a spatio-temporal verification process as described in the following.

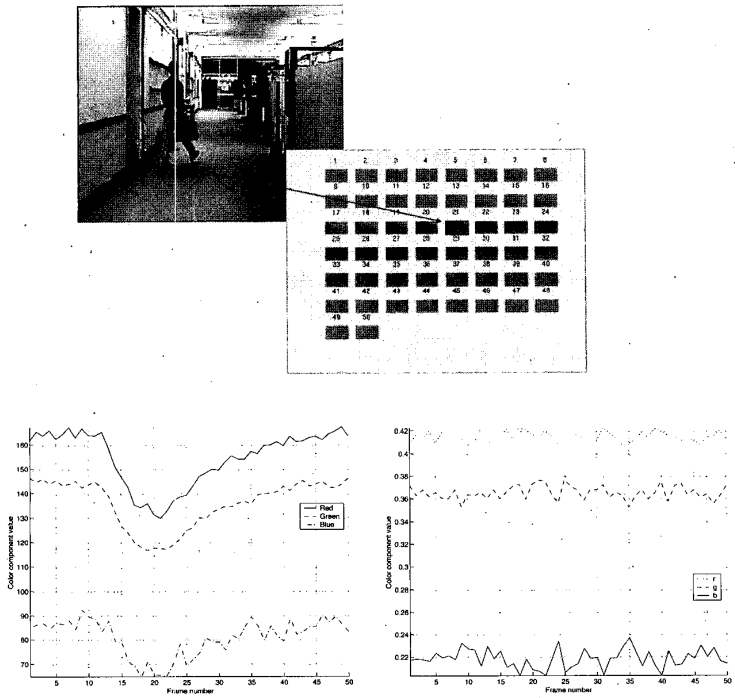

Figure 3: Effect of a moving shadow on the colour components of the *RGB* and *rgb* colour spaces. (Top) Sample frame from the test sequence *Hall Monitor* and Region of Interest (ROI) with a cast shadow. The passage of the cast shadow is highlighted for 50 frames. (Bottom, left) Profile of the R, G, and B colour components in the selected ROI. (Bottom, right) Profile of the r, g, and b components in the same ROI. It is possible to notice the invariance properties of the *rgb* colour features

- *Spatial verification* Although the colour appearance of a shadow is undefined by nature, it is possible to define some constraints for its geometric appearance. The geometrical characteristics of shadows can be defined without any knowledge of the structure of the object or of the scene. In particular, the existence a line separating the shadow from the background is a necessary condition for the presence of a shadow. In order to check this condition, the moving objects are first detected. Then the existence of the shadow separating the detected shadows from the corresponding object is verified. This allows the algorithm to eliminate those shadows that are erroneously detected *inside* an object. In the specific implementation, we use the change detection method presented in [24]. This method is embedded in the shadow detection process as shown in Fig. 1.

- *Temporal verification* The final verification is based on the temporal consistency of shadows. To this end, we employ a temporal filter based on Nearest Neighbor. The segmented shadow regions are tracked over time. Tracking provides a reliability estimation for the moving cast shadow extraction stage ([23]). This allows us to remove shadows that have a low temporal reliability. At each time instant, each extracted moving cast shadow is put in correspondence with previously extracted shadows. A correspondence between two shadows is established based on the Nearest Neighbor filter. This simple tracking mechanism allows one to eliminate undesiderable detections as shown in Fig. 5.

3 RESULTS

This section presents the results of the shadow detector described in this paper. Subjective and objective evaluation and comparison with state-of-the-art techniques are introduced in order to evaluate the performance by comparison with alternative methods. The results are evaluated subjectively by showing the detected shadows superimposed over the original image and colour-coded in white. Furthermore, objective evaluation is performed with respect to a ground-truth segmentation by comparing the results of shadow detection combined with an object detector.

Test sequences from the MPEG-4 and MPEG-7 data set are used, as well as test sequences from

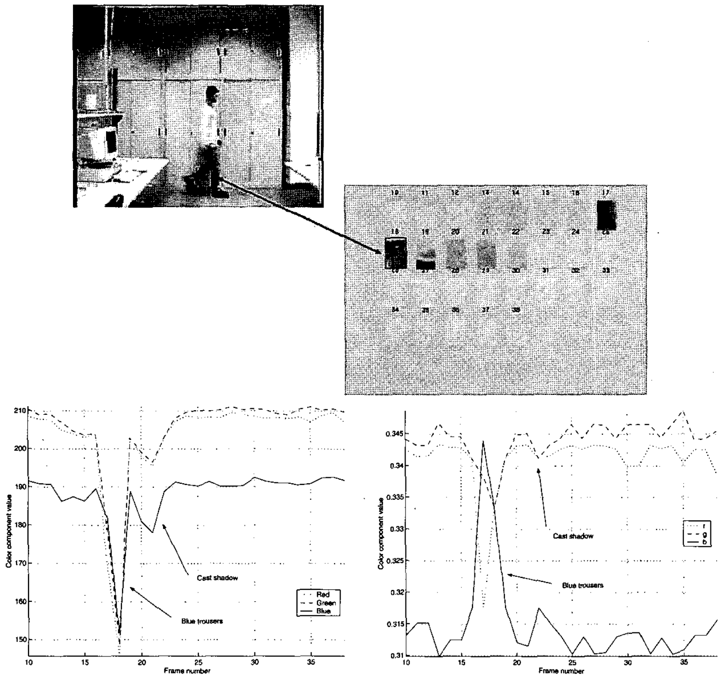

Figure 4: Comparison between the effect of a moving shadow and of an object on the colour components of the *RGB* and *rgb* colour spaces. (Top) Sample frame from the test sequence *Laboratory* and Region of Interest (ROI) with the passage of an object and a cast shadow (highlighted for 29 frames). (Bottom, left) Profile of the R, G, and B colour components in the selected ROI. The colour components decrease their value but do not change their order when a shadow occurs, whereas the colour components may change their order when an object passes. (Bottom, right) Profile of the r, g, and b components in the same ROI. It is possible to notice the different behavior of the colour components in the presence of an object and in the presence of a shadow

(a) (b) (c)

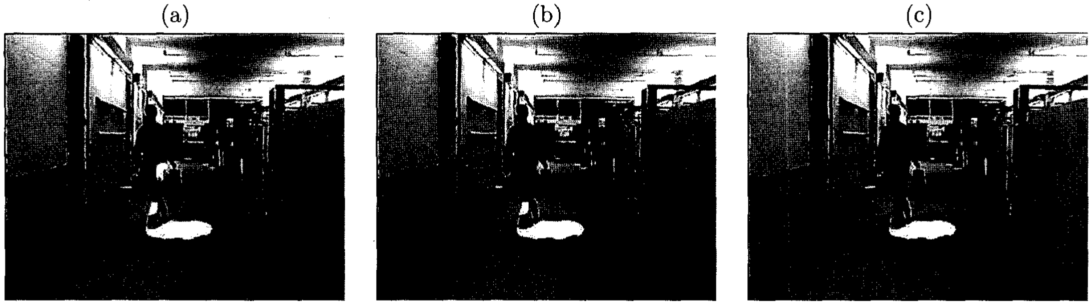

Figure 5: Shadow detection results with (a) colour analysis only; (b) colour analysis and spatial verification; (c) colour analysis and spatio-temporal verification

the test set of the ATON project and the European project *art.live*. The sequences are in CIF format (288×352 pixels) and the frame rate is 25 Hz, unless otherwise stated in the remainder of the section.

In Figure 6 the sequence *Improvisation* from the IST European project *art.live* is considered. This is a scene representing a typical situation for video production. An actor moves in the scene producing multiple shadows. It is possible to notice that the shadow detector correctly identifies the shadows and the segmentation results are stable over time. We would like to highlight here that the detection is performed without any model of the scene, the illumination, or the captured object. The advantage of this approach is also demonstrated in a more complex scene with many interactions among objects, the sequence *Group* (Figure 7). The people walking in the room cast several shadows which are caused by their interaction with multiple light sources. In this scene,

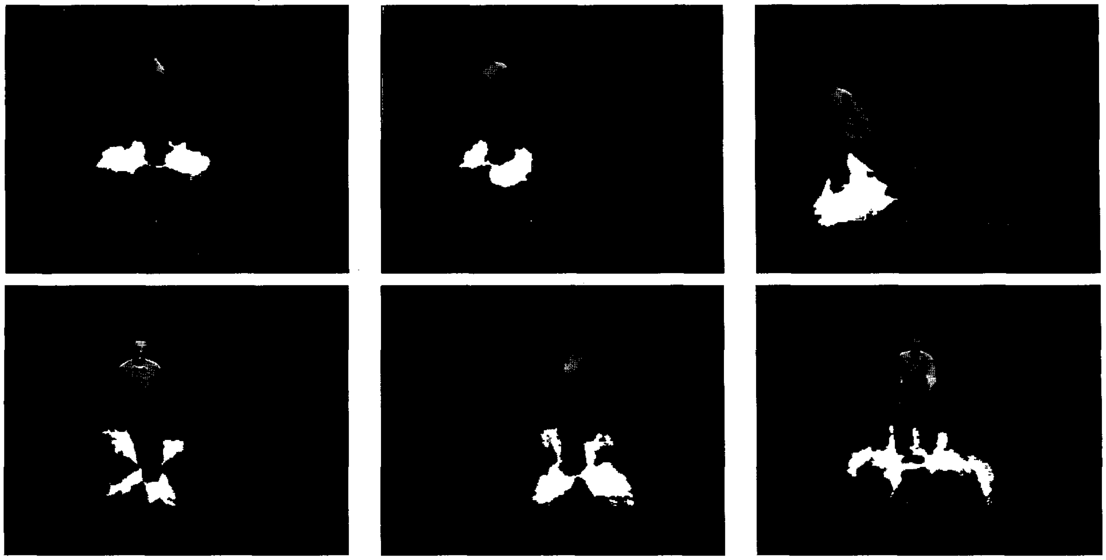

Figure 6: Shadow detection results for sample frames of the test sequence *Improvisation*

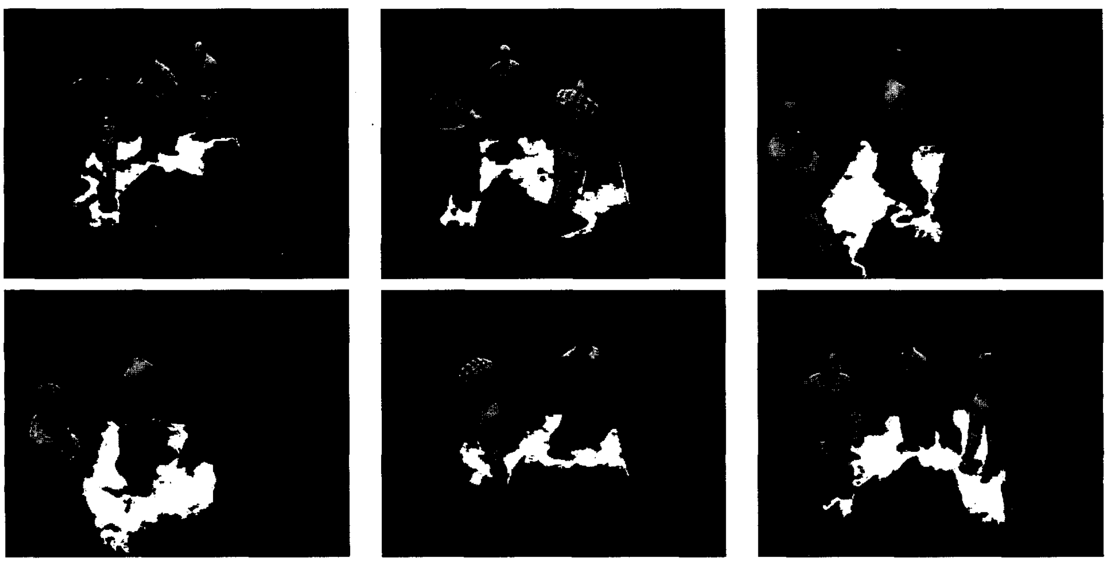

Figure 7: Shadow detection results for sample frames of the test sequence *Group*

a model based method for shadow recognition would fail due to the complexity of the scene. The proposed method is based on shadow properties, and it can be therefore applied to complex scenes, when shadows and object occlude each other. The results show the extraction of the shadow also for this indoor scene with large objects close to the camera.

A different scene set-up is shown for the test sequence *Intelligent Room* in Fig. 8. The format of this sequence is 320×240. Here the scene is more complex compared to the previous sequences and the object casting shadows is smaller. However, shadows cast both on the floor and on the walls are correctly detected. Similarly, Fig. 9 shows the shadow detection results for the MPEG-4 test sequence *Hall Monitor*. To demonstrate the performance of the proposed method in outdoor scenes, the test sequences *Surveillance* (Fig. 10 , format 352×240) and *Highway* (Fig. 11) are considered. Fig. 10 shows the shadow detection results in case of a deformable object, which can illustrate, for instance, a situation of outdoor video production. Finally, the MPEG-7 test sequence *Highway*, shown in Fig. 11, illustrate how the proposed method can work for fast moving objects.

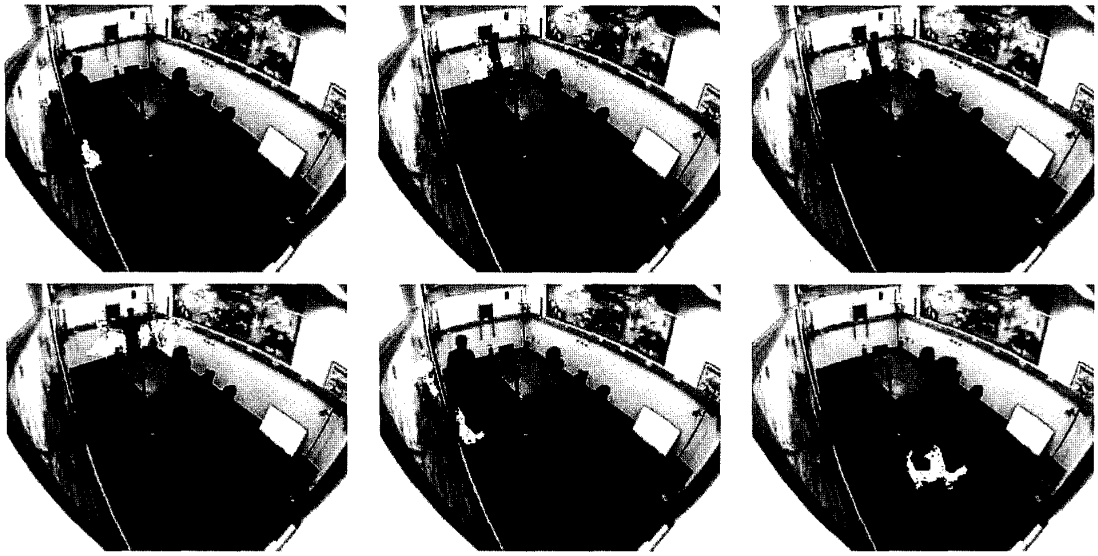

Figure 8: Shadow detection results for sample frames of the test sequence *Intelligent Room*

Figure 9: Shadow detection results for sample frames of the test sequence *Hall Monitor*

Finally, we present the results of objective evaluation and comparison with state-of-the-art methods. The evaluation of shadow segmentation is done through the evaluation of video object segmentation. The detection of shadows is exploited to improve the performances of algorithm extracting video objects which are based on change detection. The objective evaluation is performed with respect to a ground-truth segmentation. The ground-truth segmentation allows to evaluate the results of the shadow detector combined with an object detector. In Fig. 12, examples of change detection results before and after shadow detection are shown. In order to obtain an objective evaluation, the deviation of segmentation mask with respect to a ground-truth segmentation is considered. Two types of errors can be defined in each frame of the sequence n, namely false positives $\epsilon_p(n)$, and false negatives $\epsilon_n(n)$. False positives are pixels uncorrectly detected as belonging to the object mask, while false negatives are pixels belonging to the object but not detected. If $card(C(n))$ represents the number of pixels detected as object pixels at frame n, and $card(C_g(n))$ the number of pixels belonging to the ground-truth, then we compute the

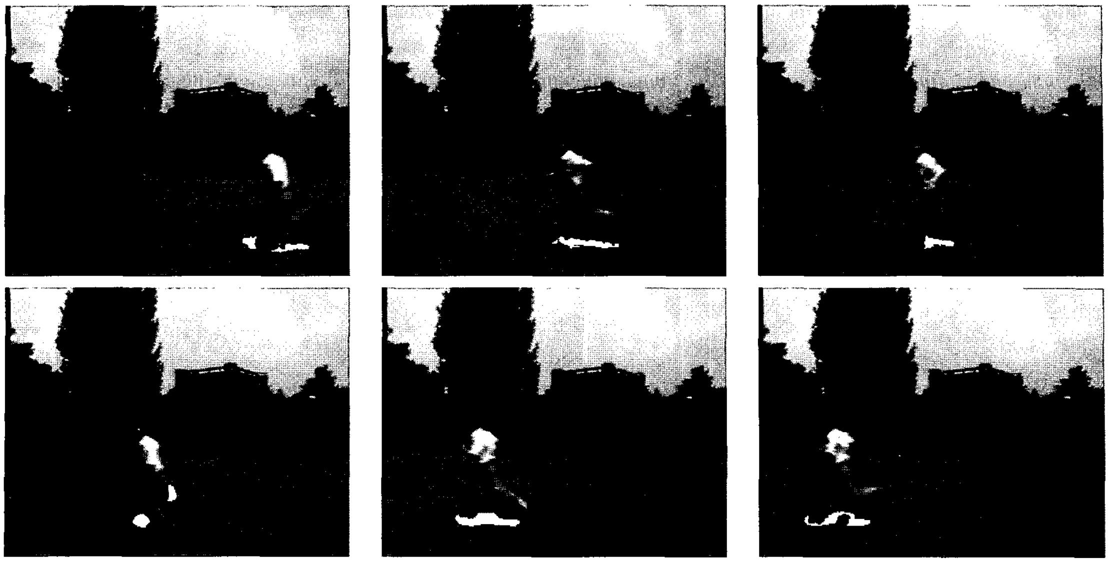

Figure 10: Shadow detection results for sample frames of the test sequence *Surveillance*

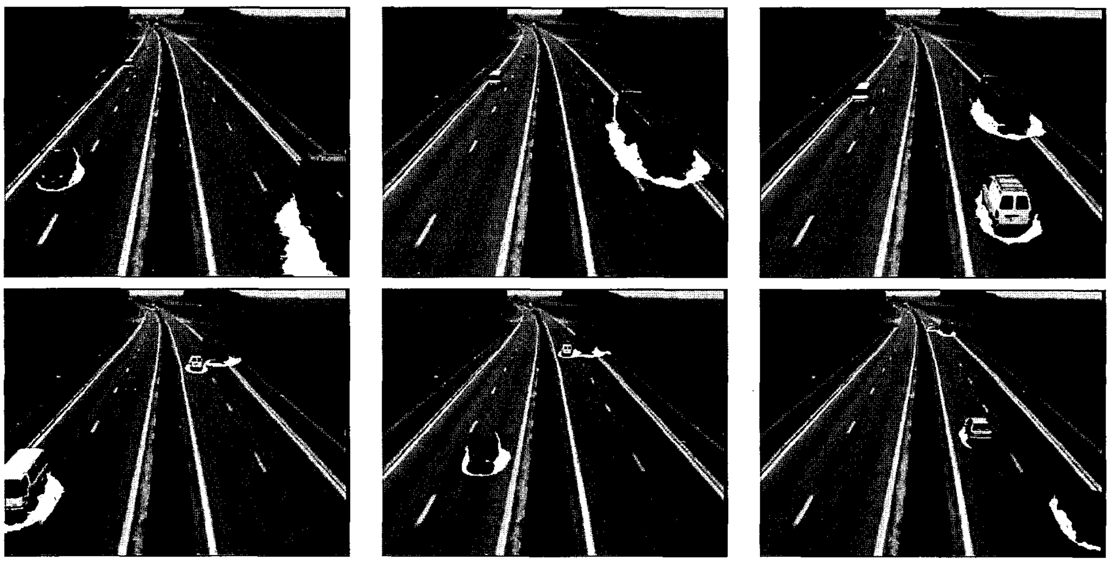

Figure 11: Shadow detection results for sample frames of the test sequence *Highway*

deviation from the reference segmentation as:

$$\epsilon(n) = \begin{cases} 0 & \text{if } Card(C(n)) = 0 \wedge Card(C_g(n)) = 0 \\ \frac{\epsilon_n(n)+\epsilon_p(n)}{Card(C(n))+Card(C_g(n))} & \text{otherwise} \end{cases} \quad (5)$$

where $\epsilon(n)$ is in $[0,1]$. The spatial accuracy of the segmentation result is then quantified by:

$$\nu(n) = 1 - \epsilon(n) \quad (6)$$

that takes again values in $[0,1]$. If $\nu(n) = 1$ then there is a perfect match between segmentation results and ground-truth. The results of the objective comparison for the test sequence *Hall Monitor* and *Intelligent room* are presented in Fig. 13. The symbols in the legend refer to the shadow detection technique used in the object extraction process: DNM1 [10], DNM2 [8], SP [12], SNP [13]. The mean values of accuracy corresponding to the plots in Fig. 13 are the following. For the test sequence *Hall Monitor*, DNM1: 0.78, DNM2: 0.60, SP: 0.59, SNP: 0.63, Proposed: 0.86. For the test sequence *Intelligent room*, DNM1: 0.86, DNM2: 0.77, SP: 0.89, SNP: 0.89, Proposed: 0.90. The combination of the proposed shadow recognition method with [24] provides a more accurate segmentation than state-of-

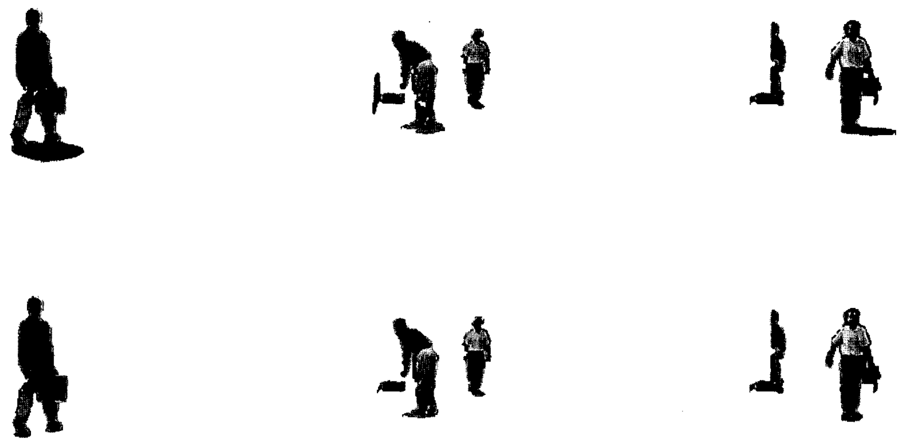

Figure 12: Comparative results of change detection for the sequence *Hall Monitor* in case of change detection only (top) and change detection followed by shadow detection (bottom)

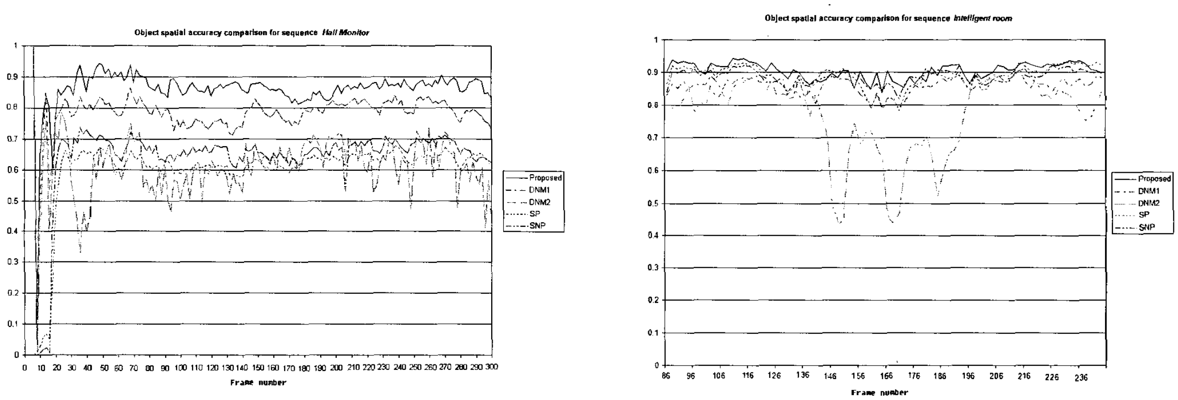

Figure 13: Objective evaluation. Comparison of the proposed algorithm with state-of-the-art methods. The plots represent the spatial accuracy in the object segmentation task. Larger values of accuracy are associated with higher accuracy in the shadow segmentation. (Left) Test sequence *Hall Monitor*. (Right) Test sequence *Intelligent room*

the-art methods.

4 CONCLUSIONS

We presented an algorithm that uses colour and colour invariants combined with spatio-temporal constraints for segmenting shadows. The proposed algorithm is designed to work when the imaging conditions and the scene set-up are not under control. This method does not require the knowledge of objects or scene models, nor requires external intervention and therefore is suitable for a wide range of applications. The selective use of colour, the use of photometric invariants, and the integration of spatial and temporal information allowed to improve the performance of state-of-the-art methods. Given the modularity of the proposed method, the different stages can be modified according to speed or to accuracy requirements. For example, the colour analysis alone could be enough for applications that do not require high accuracy (i.e. for identifying the direction of the light source). Future work includes the extension of the method to cope with a multi camera environment and the optimisation of the code to reach real-time performance.

5 ACKNOWLEDGMENTS

We would like to acknowledge Andrea Prati for providing us with the segmentation results of the state-of-the-art methods ([25]) that have been used in the performance comparison.

References

[1] T. Glover, "Cut! Edit! Why digital video editing is the next big thing," *The Business,* **280**, pp. 1–6, July 2003.

[2] O. Grau, "A Studio Production System For Dynamic 3D Content," in *Proc. of Visual Communications and Image Processing*, 2003.

[3] X. Marichal and T. Umeda "Real-Time Segmentation of Video Objects for Mixed-Reality Interactive Applications," in *Proc. of Visual Communications and Image Processing*, 2003.

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

[5] K. Onoguchi, "Shadow elimination method for moving object detection," in *Proc. of IEEE Int. Conf. on Pattern Recognition (ICPR)*, pp. 583–587, 1998.

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

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

[8] J. Stauder, R. Melch, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Transactions on Multimedia,* **1**(1), pp. 65–77, 1999.

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

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

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

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

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

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

[15] G. Finlayson, S. Hordley, and M. S. Drew, "Removing shadows from images," in *European Conf. on Computer Vision*, pp. 823–836, 2002.

[16] S. Nadimi and B. Bhanu, "Moving shadow detection using a physics-based approach," in *Proc. of IEEE Int. Conf. on Pattern Recognition*, **2**, pp. 701–704, 2002.

[17] V. Risson, *Application de la Morphologie Mathématique à l'Analyse des conditions d'Éclairage des Images Couleur.* PhD thesis, Paris School of Mines, December, 2001.

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

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

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

[21] S. A. Shafer, "Using colour to separate reflection components," *Colour Research Applications,* **10**(4), pp. 210–218, 1985.

[22] T. Gevers and A. W. M. Smeulders, "Colour-based object recognition," *Pattern Recognition,* **32**, pp. 453–464, 1999.

[23] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Spatio-temporal Shadow Segmentation and Tracking," in *Proc. of Visual Communications and Image Processing*, 2003.

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

[25] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, pp. 918–923, 2003.