Detection of Moving Cast Shadows for Object Segmentation

Jürgen Stauder, Roland Mech, and Jörn Ostermann

***Abstract*— To prevent moving shadows being misclassified as moving objects or parts of moving objects, this paper presents an explicit method for detection of moving cast shadows on a dominating scene background. Those shadows are generated by objects moving between a light source and the background. Moving cast shadows cause a frame difference between two succeeding images of a monocular video image sequence. For shadow detection, these frame differences are detected and classified into regions covered and regions uncovered by a moving shadow. The detection and classification assume plane background and a nonnegligible size and intensity of the light sources. A cast shadow is detected by temporal integration of the covered background regions while subtracting the uncovered background regions. The shadow detection method is integrated into an algorithm for two-dimensional (2-D) shape estimation of moving objects from the informative part of the description of the international standard ISO/MPEG-4. The extended segmentation algorithm compensates first apparent camera motion. Then, a spatially adaptive relaxation scheme estimates a change detection mask for two consecutive images. An object mask is derived from the change detection mask by elimination of changes due to background uncovered by moving objects and by elimination of changes due to background covered or uncovered by moving cast shadows. Results obtained with MPEG-4 test sequences and additional sequences show that the accuracy of object segmentation is substantially improved in presence of moving cast shadows. Objects and shadows are detected and tracked separately.**

***Index Terms*— Illumination, MPEG-4, object segmentation, shadow detection, VOP generation.**

I. Introduction

THE ISO/MPEG-4 standard [6], [24], [31], [36], [38] requires the segmentation of moving objects for so-called content-based functionalities. The video sequence to be transmitted is segmented into two-dimensional (2-D) video objects, and each video object is transmitted by a specific coding technique at a specific quality level. At the decoder side, a compositor reconstructs the scene based on the transmitted video objects. Applications are for example automatic surveillance [18], content-based [46], and object-based [12], [16], [32], [34] coding. Since many applications operate in real time, good automatic segmentation algorithms are required.

Current approaches to object segmentation are mainly based on motion information. In [23] and [37], an estimated displacement vector field is segmented into homogeneous regions. Due to the corona effect, such a segmentation is inaccurate. More accurate results can be achieved considering additionally texture information. In [11], in a first step an image is segmented in to regions with homogeneous texture. In a second step, regions with similar motion are merged. Due to the successive application of texture and motion analysis, this technique is suboptimal. In [43] and [44], an image is segmented into regions with homogeneous displacement vectors considering *concurrently* texture edges. This technique suffers from its high computational load. Furthermore, it tends to oversegmentation with respect to the number of moving objects in the scene.

In [22], an automatic, noise robust segmentation technique for segmentation of moving objects in video sequences considering a moving camera is presented, which overcomes the problems of oversegmentation by the assumption of a dominating background. First, the algorithm estimates and compensates apparent camera motion. Then, a change detection marks changed regions comparing the current image with the previous camera motion compensated image. A relaxation technique [1], [2] using a spatially adaptive threshold is applied. Removing regions of uncovered background from the binary change detection mask using motion information [15] results in a 2-D object mask. The algorithm from [22] has been under investigation within the framework of an MPEG-4 core experiment on automatic segmentation [25]–[31]. Furthermore, parts of the algorithm had been included into the European COST 211^{quat} Analysis Model [7], [8].

This paper addresses the problem of segmentation of moving objects in presence of moving cast shadows on the background. A cast shadow on the background is generated by an object moving between a light source and the background. Moving cast shadows cause a frame difference between two succeeding images of a monocular video sequence. In this case, neither motion segmentation nor change detection based methods can distinguish between moving objects and moving cast shadows. To solve this problem, moving cast shadows have to be detected explicitly to prevent them being misunderstood as moving objects or parts of moving objects.

In [3] and [18], cast shadows are identified using known three-dimensional (3-D) object geometry, known direction of the shadow causing light source and known 3-D geometry of

Manuscript received September 1, 1998; revised November 20, 1998. The associate editor coordinating the review of this paper and approving it for publication was Prof. Yao Wang.

J. Stauder is with IRISA/INRIA, Project TEMICS, Campus Universitaire de Beaulieu, F-35042 Rennes Cedex, France (e-mail: jurgen.stauder@irisa.fr, http://www.irisa.fr/prive/Jurgen.Stauder).

R. Mech is with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover, Germany (e-mail: mech@tnt.uni-hannover.de, http://www.tnt.uni-hannover.de/nmech).

J. Ostermann is with AT&T Laboratories, Speech and Image Processing Services Research Laboratory, Red Bank, NJ 07701-7033 USA (e-mail: osterman@research.att.com).

Publisher Item Identifier S 1520-9210(99)01582-5.

the background for constraint environments as a traffic scene [18] or simple buildings [3]. Shadows of persons are analyzed in [5].

Without knowing any 3-D geometry, image regions changed by moving cast shadows can be identified using assumptions on the unknown 3-D geometry. The approaches in the literature can be classified according to the assumptions they make. Four assumptions can be found. First, the light source is assumed to be strong [39]. Then, illumination changes due to a moving cast shadow are large in amplitude. Thus, a first criterion for image regions changed by moving cast shadows is a large frame difference between consecutive frames [35]. Second, the camera is assumed to be static and the background is assumed to be static and textured [33], [35], [42]. Then, image regions changed by shadows can be separated from image regions changes by objects by a search for static edges of the background texture. Third, the background is assumed to be plane [39]. Then, illumination changes due to a moving cast shadow are smooth. In [39], the illumination changes are measured by comparing an image with a reference image showing the same scene without cast shadows. Image regions changed by a moving cast shadow are identified by a large and smooth frame difference. The fourth assumption is that the light source causing a cast shadow has a certain extent [17], [47]. Then, cast shadows have a penumbra. The penumbra is a soft luminance transition from shadowed to non-shadowed background. The penumbra is modeled by a Gaussian step function [47].

In this paper, all four assumptions are used at the same time to detect image regions changes by moving cast shadows. As in [35], [39], and [42], the light source is assumed to be strong and image regions changed by moving cast shadows will be detected by a large frame difference. As in [33], [35], and [42], the camera is assumed to be static and the background is assumed to be static and textured. Candidates for image regions changed by moving cast shadows on static background will be detected by static edges. As in [39], the background is assumed to be plane. On the contrary to [39], no reference image will be used that shows the same scene without shadows. The illumination changes will be measured directly from two frames of an image sequence using a physics-based signal model. As in [17] and [47], the light source is assumed to have a certain extent, and image regions changed by moving cast shadows will be detected by the penumbra of the shadows. On the contrary to [47], the luminance transition will be assumed to follow linear signal model to simplify the detection algorithm.

The results of the four criteria emerging from the four assumptions will be combined to a binary mask for image regions changed by moving shadows. We propose the first algorithm that temporally integrates these image regions resulting in a detection and tracking of entire cast shadows in an image sequence.

The method for detection of moving cast shadows will be integrated in the object segmentation algorithm from [22]. Therefore, detected image regions changed by moving cast shadows will be deleted from the change detection mask before further processing as in [22].

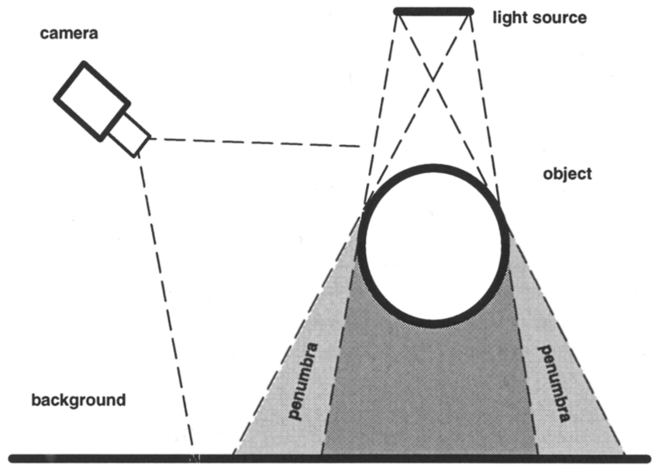


Fig. 1. Cast shadow generation: the scene grabbed by a camera consists of a moving object and a moving cast shadow on the background. The shadow is caused by a light source of certain extent and exhibits a penumbra.

The paper is organized as follows. In Section II, the appearance of cast shadows in video images is discussed. Based on these insights, a shadow detector is introduced in Section III. In Section IV, the integration of the shadow detector in the object segmentation algorithm from [22] is described. Section V presents experimental results for MPEG-4 and additional test sequences. Section VI gives conclusions.

II. Appearance of Cast Shadows in Video Images

Effects in the real world that are visible by a human eye are described in the field of optics. If the size of obstacles is large compared to the wave length of the light, the geometric optic is sufficient. The geometric optic describes the distribution of light by light rays of infinitesimal radius.

Cast shadows in the real world belong to so-called global illumination effects, because the light ray on its way from the light source to the eye or camera is affected by more than only one reflection on an object surface [10], [45]. In Fig. 1, the formation of a cast shadow is shown. The light coming from a single light source reaches the background only partially due to a moving object. The darkened region on the background is called cast shadow. It is illuminated by some ambient diffuse light or by other light sources only.

A cast shadow consists of a center part without any light from the light source, called the umbra, and a soft transition from dark to bright, called the penumbra, where some light from the light source reaches the background [45].

The appearance of a cast shadow in an image of a video camera can be described by an image signal model. It describes the image luminance

$$s_k(x, y) = E_k(x, y)\rho_k(x, y) \quad (1)$$

at time instant k at the 2-D image position x, y by the product of the irradiance $E_k(x, y)$ and reflectance $\rho_k(x, y)$ of the object surface.

The irradiance $E_k(x, y)$ is the amount of light power per receiving object surface area; see Fig. 2. It is a function of the direction $\mathbf{L}$ of the light source, the intensities c_P and c_A

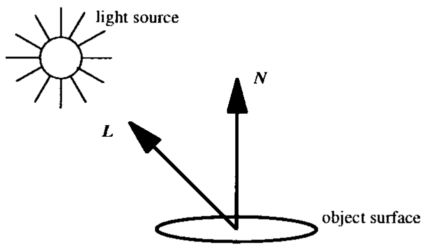


Fig. 2. Illumination of an object surface with surface normal $\mathbf{N}$ by a light source from direction $\mathbf{L}$.

of light source and ambient light, respectively, and the object surface normal $\mathbf{N}$ according to

$$E_k(x, y) = \begin{cases} c_A + c_P \cos \angle(\mathbf{N}(x, y), \mathbf{L}), & \text{if illuminated} \\ c_A + k(x, y) c_P \cos \angle(\mathbf{N}(x, y), \mathbf{L}), & \text{if penumbra} \\ c_A, & \text{if umbra.} \end{cases} \quad (2)$$

In (2), the term $0 \leq k(x, y) \leq 1$ describes the transition inside the penumbra and depends on the light source and scene geometry [45]. The intensity c_P of the light source is proportional to $1/r^2$ with r being the distance between object and light source [41]. Equation (2) is based on Lamberts cosine law [19] and is partially contained in Phongs illumination model [20].

In the image model (1), radial photometric distortions of perspective projection [14] are neglected. Also, the gamma nonlinearity and gain control of a video camera are not considered. Finally, the ambient light and all light sources are assumed to have the same color.

III. Shadow Detection

In this section, moving cast shadows on a dominating background will be detected. The shadows are generated by objects moving between a light source and the background. For shadow detection, the 3-D geometry of shadow generation as shown in Fig. 1 may be evaluated. Such algorithms [3], [18] need to know the following information:

- shape and position of the moving object;
- shape and position of the background;
- shape, position and intensity of the light source;
- intensity of the ambient illumination;
- camera geometry.

If such information is unknown, assumptions on it can help to identify shadows and distinguish them from moving objects. In this paper, four assumptions are made to detect image regions changed by moving cast shadows as a first step to shadow detection. These image regions are background regions covered or uncovered by a moving cast shadow from image to image. The four assumptions lead to four criteria that are evaluated on two succeeding images of an image sequence. The first assumption:

1) light source intensity c_P is high;

this implies that the frame difference at a pel changed by a moving cast shadow will be large. To see why this is the case, consider a picture element (pel) at position x, y that shows a part of the background. Let us assume that the pel is inside the umbra of a cast shadow at time instant k and outside the shadow at time instant $k+1$. The reflectance of a static background does not change with time, thus $\rho_k(x, y) = \rho_{k+1}(x, y)$ holds. According to (1), the frame difference will be then

$$\begin{aligned} &s_{k+1}(x, y) - s_k(x, y) \\ &\quad = \rho_k(x, y) c_P \cos \angle(\mathbf{N}(x, y), \mathbf{L}) \geq 0. \end{aligned} \quad (3)$$

The frame difference will be large if c_P is large. It can be concluded that if assumption 1 holds, this pel that belongs to a cast shadow will be part of the change detection mask. A change detection mask indicates those image regions having a large frame difference between the previous and the current image. The change detection mask is assumed to be available, see Section IV.

Three more assumptions:

2) camera and background are static,
3) background is plane, and the light source position is distant from background,
4) light source size and distance between moving object and background are not negligible compared to the distance between light source and object;

these lead to three criteria and will be discussed in Sections III-A, B, and C. In Section III-D, the results of the criteria are combined to a binary mask indicating regions being changed by moving cast shadows from one image to another. Section III-E describes how to temporally integrate these changed regions along an image sequence to detect and track entire cast shadows. Finally, Section III-F discusses the choice of algorithmic parameters that are used for shadow detection.

A. Detection of Static Background Edges

In image regions changed either by moving textured objects or moving cast shadows on static textured background, assumption 2 can be used to distinguish between shadow and object. In nontextured image regions, assumption 2 does not give any information. To use assumption 2 to distinguish between shadow and object in changed image regions, edges in the current and in the previous image are detected and classified.

For edge detection, each image is searched for pels with a zero crossing *and* high locally calculated variance of the second derivative of image luminance (see Fig. 3). The detection method is inspired by the sensitivity of a human eye to edges [13]. Then, the edges in the previous and current image are classified into moving and static edges; see Fig. 4. A pel belonging to an edge will be classified as static, if in a local neighborhood of this pel the energy in high frequencies of the frame difference between current and previous image is low [35]. Other edges are classified as moving edges. The threshold for the high-frequency energy is adaptively calculated from the local high-frequency energies of the frame difference outside the change detection mask. The threshold is calculated such that 98% of the pels outside the change detection mask have an energy below the threshold (see Section III-F).

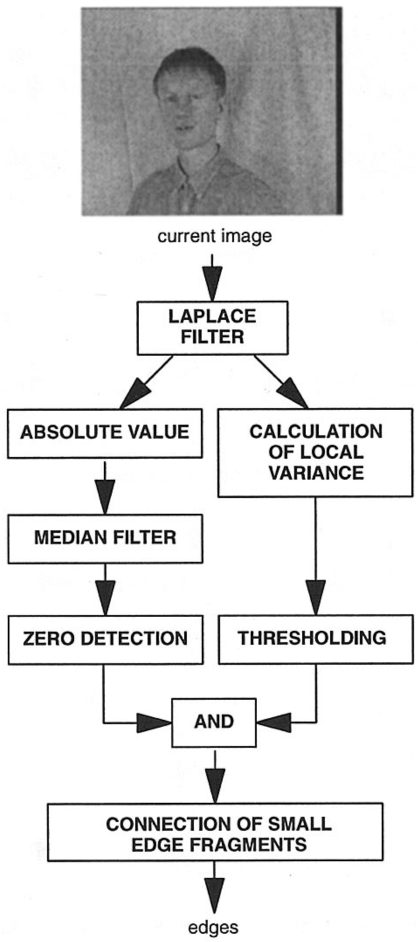


Fig. 3. Detection of edges by looking for pels with zero crossing *and* high local variance in second derivative of image luminance derived by a Laplacian filter.

The static edges are used to detect nonmoving regions inside the change detection mask. Equation (1) shows that static edges in an image $s_k(x, y)$ are caused by static, spatial discontinuities either of the reflectance $\rho_k(x, y)$ or of the irradiance $E_k(x, y)$. Static edges caused by discontinuities in the reflectance hint at texture of a static background. Static edges caused by discontinuities in the irradiance hint at discontinuous shading at 3-D shape edges of a static background. Thus, static edges hint at static background and define therefore possible image regions of a moving cast shadow on a static background.

In case of a moving camera, assumption 2 may still be applied if the previous image s_k is motion compensated with respect to the camera motion. Such a motion compensation is introduced in Section IV.

B. Detection of Uniform Changes of Shading

Assumption 3 says that the background is plane and the light source is distant from the background. In this case, the irradiance according to (2) is spatially constant, because the surface normal $\mathbf{N}(x, y)$ is spatially constant. Note that in this section, the shadow penumbra is neglected. To make use of

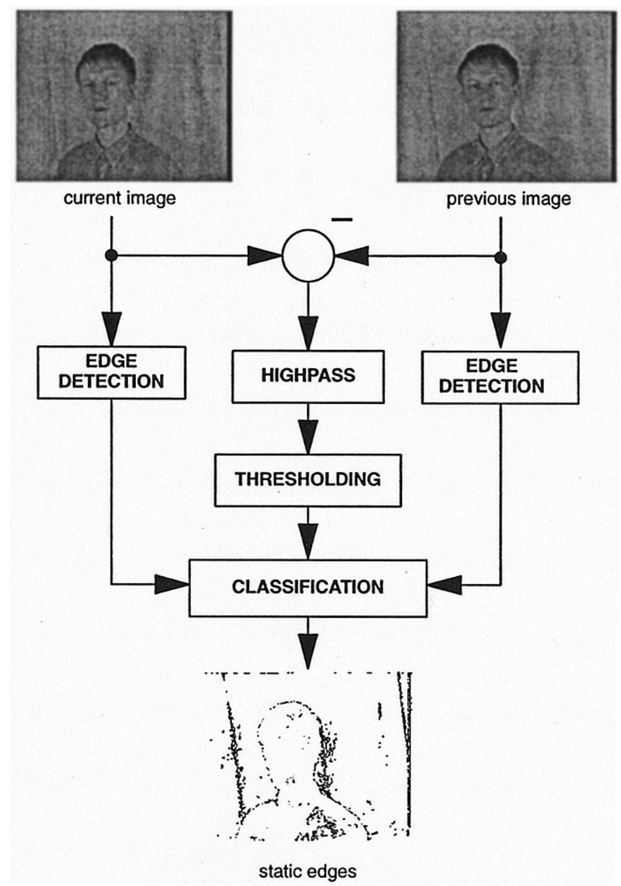


Fig. 4. Detection of static edges: edges either in the current or the previous image are classified as static edges if the high frequency part of the frame difference is low in a local neighborhood.

the constant irradiance for shadow detection, the frame ratio

$$FR(x, y) = \frac{s_{k+1}(x,y)}{s_k(x,y)} = \frac{E_{k+1}(x,y)}{E_k(x,y)} \frac{\rho_{k+1}(x,y)}{\rho_k(x,y)} \tag{4}$$

is evaluated inside the change detection mask. For each pel, the hypothesis is tested, that the luminance at position x, y has changed due to a moving cast shadow. If the hypothesis is valid, the background reflectance does not change and $\rho_k(x, y) = \rho_{k+1}(x, y)$ holds. Then, neglecting any camera noise, the frame ratio can be simplified to

$$FR(x, y) = \frac{E_{k+1}(x, y)}{E_k(x, y)}. \tag{5}$$

According to (5), the frame ratio will be spatially constant in a neighborhood of x, y, if the hypothesis holds. This is because the irradiance is constant as discussed previously. For shadow detection, this conclusion is used vice versa: if the frame ratio is locally spatially constant, a moving cast shadow is assumed at position x, y.

The frame ratio is tested for spatial constancy by evaluating its variance that is calculated for each pel in a local neighborhood, see Fig. 5. A pel with a variance smaller than a threshold indicates a uniform change of shading. We assume that these pels are changed by moving cast shadows. The threshold is adaptively set to 2% of the variance of the frame ratio of all pels inside the change detection mask (see Section III-F).

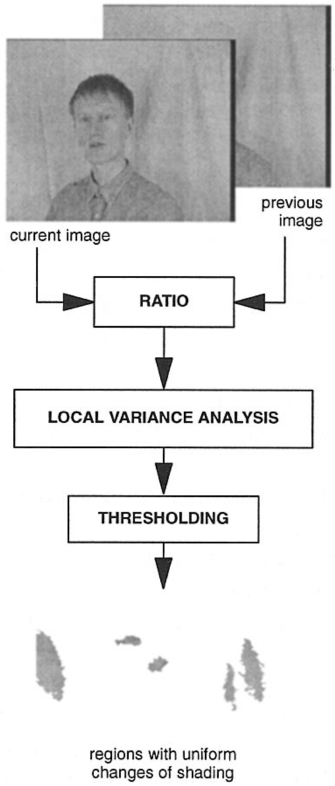


Fig. 5. Detection of uniform changes of shading: the luminances of the current and the previous image are divided pel-wise resulting in the frame ratio. Then, pels with a small variance in the frame ratio measured in a local spatial neighborhood are selected.

For each pel of the frame ratio, its local variance is calculated in a local neighborhood. In the neighborhood, only those pels will be evaluated that have a ratio larger or equal one if the center pel has a ratio larger than one and vice versa. If a majority of pels in the local neighborhood do not match this constraint, the criterion is not evaluated for this pel.

There is a case where the criterion will fail. The criterion will detect erroneously a pel as being changed by a moving cast shadow, if the pel shows a uniformly colored, rotating object. In this case, the simplification from (4) to (5) holds and the frame ratio will be locally spatially constant [40]. Such an error can be seen in Fig. 5, where two regions are detected in the facial area.

C. Penumbra Detection

Assumption 4 says that the extent of the light source and the distance between object and background are not negligible compared to the distance between light source and moving object. Then, a cast shadow has a penumbra [45]. The idea of the third criterion is to detect shadows by their penumbra.

The penumbra causes a soft luminance step at the contour of a shadow. The luminance step in an image perpendicular to a shadow contour is modeled by the luminance step model shown in Fig. 6. The luminance is assumed to rise linearly from a low luminance inside a shadow to a high luminance outside the shadow. The luminance step is characterized by its step height h, step width w, and its gradient g. If the width of a luminance step caused by a penumbra is much larger than that of edges caused by the camera aperture for object surface texture edges or object edges, assumption 4 can be used for shadow detection.

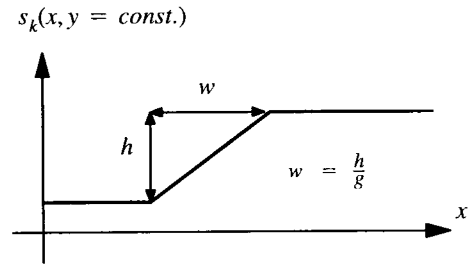


Fig. 6. Model of an image luminance step in an image s_k at time instant k in direction perpendicular to a shadow contour: the luminance step is defined by the step height h and the norm g of the gradient. From h and g, the width w can be calculated. In this figure, the shadow contour is assumed to be in y direction.

In Table I, the luminance step height, gradient, and width of different kinds of edges in an image are characterized. It can be seen, that shadow edges can be distinguished best from other edges by their luminance step width. The luminance step height is not appropriate, because either a shadow edge caused by a bright light source or a texture edge with high contrast may cause a high luminance step height and vice versa. The luminance step gradient is not appropriate because the gradient of a shadow edge caused by a bright light source (with a certain extent) may be comparable to that of a texture edge with less contrast (and small camera aperture).

For penumbra detection, edges are evaluated in the frame difference between the previous and the current image. The frame difference is considered because the relevant edges of the current and previous frame are included. The luminance edge model from Fig. 6 is therefore applied to the frame difference.

The pels at the border of the change detection mask are selected as penumbra candidate pels (see Fig. 7). The penumbra candidate pels may be object or shadow edges because the change detection mask contains image regions changed by moving objects or moving cast shadows. The candidate pel selection has two advantages. First, the number of candidate pels is low compared to the number of edges indicated by a standard edge detection algorithm as in [47]. Second, standard edge detection algorithms have difficulties in finding soft edges of a shadow [9]. The candidate pel selection is enhanced by two steps. First, the number of candidate pels is further reduced. Therefore, an object mask (see Section IV) for the moving objects in the previous image is, if available, or-connected with the change detection mask before candidate pel selection to close wholes. Or-connection means that a pel will be set in the result mask if it is set in at least one of the input masks. Second, to enhance the precision, refined penumbra candidate pels are searched perpendicular to the border of the change detection mask at a position of highest gradient in the frame difference. The gradient is measured perpendicular to the border of the change detection mask using a Sobel operator aligned perpendicular to the edge.

To decide whether a candidate pel belongs to a penumbra, the height and gradient of its frame difference step perpendicu-

TABLE I
DIFFERENT KINDS OF EDGES IN AN IMAGE AND THEIR CHARACTERIZING PARAMETERS: THE LUMINANCE STEP HEIGHT, GRADIENT AND WIDTH; IT CAN BE SEEN THAT SHADOW AND TEXTURE/OBJECT EDGES CAN BE DISTINGUISHED ONLY BY THEIR LUMINANCE STEP WIDTH

Edge Type	Edge Height	Edge Gradient	Edge Width
shadow contour, bright light source	large	medium	large
shadow contour, weak light source	small	small	large
texture/object edge, high contrast	large	large	small
texture/object edge, less contrast	small	medium	small

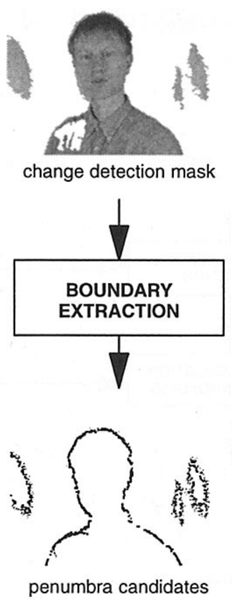


Fig. 7. Detection of penumbra candidates: penumbra candidates are searched near the boundary of the change detection mask.

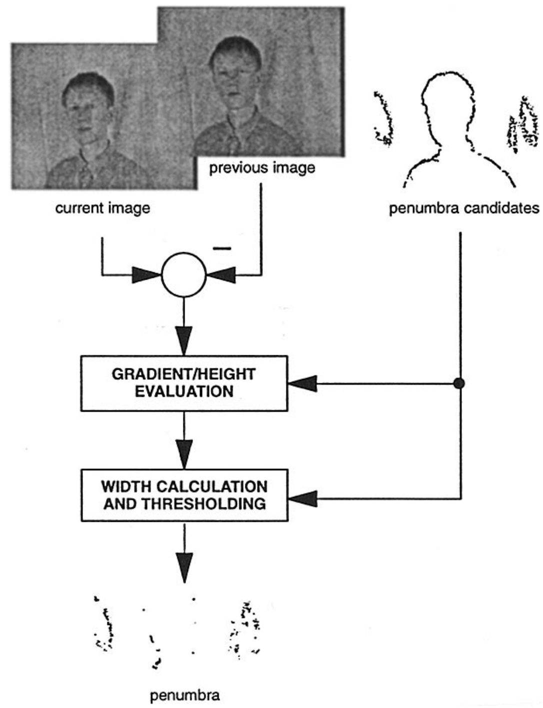


Fig. 8. Detection of penumbra: for pels being penumbra candidates, the gradient and height of the frame difference signal step is measured. Then, the step width is calculated and thresholded. Pels with a step width exceeding the threshold belong to penumbra.

lar to the edge is measured; see Fig. 8. The height is measured by the difference of averaged frame differences from both sides of the edge. Therefore, averaging windows are placed at both sides of the edge. The gradient is measured using a Sobel operator aligned perpendicular to the edge. The direction of the edge is measured by a regression line evaluating penumbra candidate pels in a local neighborhood. For each penumbra candidate pel, from height h and gradient g, the width $w = h/g$ of the signal step is calculated. The width w is thresholded. Each penumbra candidate pel having a width greater than a threshold is detected as penumbra. The threshold is set to 2.5 pels (see Section III-F). Other penumbra candidate pels are said to be object edges.

D. Detection of Image Regions Changed by Moving Cast Shadows

To detect image regions changed by moving cast shadows, the results of the three criteria from Sections III-A–C are evaluated by heuristic rules. First, the current image is divided into image regions of connected pels having the same result of change detection, static edge detection (Section III-A) and shading change detection (Section III-B). Then, the rules shown in Table II are applied to each image region to decide whether it inherits changes that are caused by moving cast shadows. The rules require changed image regions containing either no edges or static edges. Additionally, a uniform change of shading is required, at least in a local neighborhood.

Finally, the penumbra criterion from Section III-C is evaluated in a local neighborhood of each region: if a majority of object edges is observed, the shadow hypothesis from Table II is rejected.

E. Detection and Tracking of Moving Cast Shadows

For detection of entire regions of moving cast shadows, the detected image regions changed by moving cast shadows are classified and temporally integrated. The classification identifies the detected image regions pel-wise to be background that is newly covered or uncovered by a moving cast shadow. A pel is decided to belong to background that has been newly covered or uncovered by a moving cast shadow in the current image if the luminance in the current image has decreased or increased, respectively.

The final binary mask of moving cast shadows for a current image is derived from temporal integration. At the beginning of an image sequence or after a scene cut, the final shadow mask is cleared. For a current image, the regions newly covered by a moving cast shadow are added to the final shadow mask. Image regions newly uncovered by a moving cast shadow are deleted from the shadow mask if they were

TABLE II
SHADOW DETECTION; HEURISTIC EVALUATION RULES OF RESULTS OF CHANGE DETECTION, STATIC EDGE DETECTION (SECTION III-A) AND SHADING CHANGE DETECTION (SECTION III-B); THE RULES ARE APPLIED TO EACH REGION OF CONNECTED PELS OF SAME TYPE. THE NEIGHBORHOOD (NH) ARE ALL IMAGE REGIONS NEIGHBORING THE REGION UNDER CONSIDERATION

Image region type	Change detection result	Static edge detection result	Shading change detection result	Shadow detection result
0	unchanged	no edge	no result	NO SHADOW
1	unchanged	no edge	uniform	NO SHADOW
2	unchanged	no edge	non-uniform	NO SHADOW
3	unchanged	static edge	no result	NO SHADOW
4	unchanged	static edge	uniform	NO SHADOW
5	unchanged	static edge	non-uniform	NO SHADOW
6	unchanged	moving edge	no result	NO SHADOW
7	unchanged	moving edge	uniform	SHADOW, if at least one region of type 13 in NH NO SHADOW, else
8	unchanged	moving edge	non-uniform	SHADOW, if at least one region of type 13 in NH NO SHADOW, else
9	changed	no edge	no result	NO SHADOW
10	changed	no edge	uniform	SHADOW
11	changed	no edge	non-uniform	SHADOW, if at least one 10-region, but any 15-region in NH SHADOW, if at least one region of type 0-6, but any region of type 12 or 17 in NH NO SHADOW, else
12	changed	static edge	no result	SHADOW, if at least one region of type 13, but no regions of type 11 or 17 in NH NO SHADOW, else
13	changed	static edge	uniform	SHADOW
14	changed	static edge	non-uniform	SHADOW
15	changed	moving edge	no result	NO SHADOW
16	changed	moving edge	uniform	SHADOW, if at least one region of type 13 in NH NO SHADOW, else
17	changed	moving edge	non-uniform	NO SHADOW

marked already as shadow. After this update, the mask is median filtered, simplified and small regions are eliminated. Also, image regions that are covered by a moving object are deleted from the shadow mask to prevent artifacts of nondetected background uncovered by a shadow while being covered by an object.

A cast shadow will be completely detected as soon as it has completely uncovered its location in the first frame. Pels that are always covered by a cast shadow can not be detected by this algorithm.

F. Selection of Algorithmic Parameters

The main three algorithmic parameters of the shadow detection algorithm define the thresholds used by the three criteria presented in Sections III-A–C.

For the static edge criterion in Section III-A, the high-frequency energy threshold is automatically adapted to the current image assuming that erroneously 2% of pels outside the change detection mask belong to moving objects. The figure of 2% is fixed for all experiments and can be justified by a simplification step during change detection that eliminates small changed image regions from the change detection mask. Thus, this threshold calculation has been adapted to the chosen change detection algorithm (see Section IV).

For the illumination criterion in Section III-B, the threshold for the local variance of the frame ratio is set to 2% of the variance of the frame ratio of all pels inside the change detection mask. The figure of 2% is fixed for all experiments. The threshold calculation assumes that mainly the moving objects determine the variance inside the change detection mask. Then, only few regions corresponding to moving objects will be erroneously detected as shadows. If the objects are highly textured, the threshold will automatically rise, and even shadows on slightly nonplane background (with slightly nonconstant frame ratio) will be detected successfully by this criterion. If the objects are less textured, the threshold automatically decreases to prevent a false detection of shaded object regions as mentioned in Section III-B. Thus, the threshold calculation of this criterion automatically adapts to the distinctiveness of moving objects and moving cast shadows.

For the penumbra criterion in Section III-C, the threshold for the edge width should be chosen evaluating both camera aperture and 3-D scene geometry. Theoretically, it should be larger than the width of any edge caused by the aperture and smaller than the width of the sharpest shadow edge. For the experiments in this paper, the threshold is set to 2.5 pels for CIF image format. This threshold has been found to be sensible, because there are shadows in real image sequences with edges as sharp as object texture edges smoothed by the camera aperture. In this case, this criteria will fail. Generally, it should be larger for high resolution images and for cameras with larger apertures.

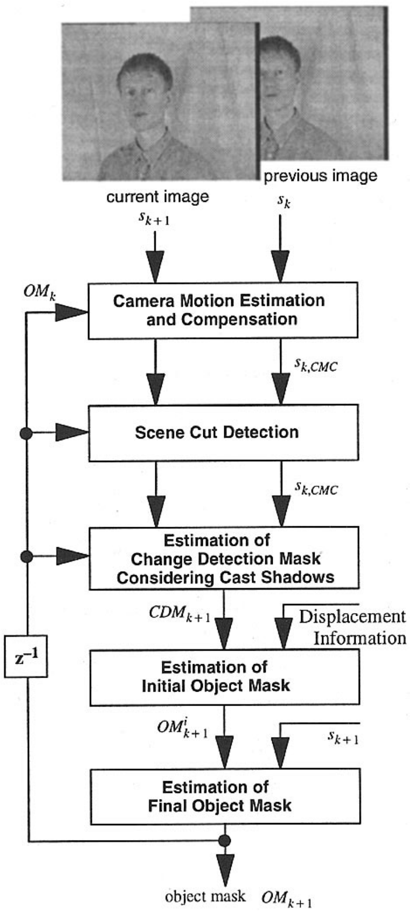


Fig. 9. Block diagram of the proposed segmentation algorithm.

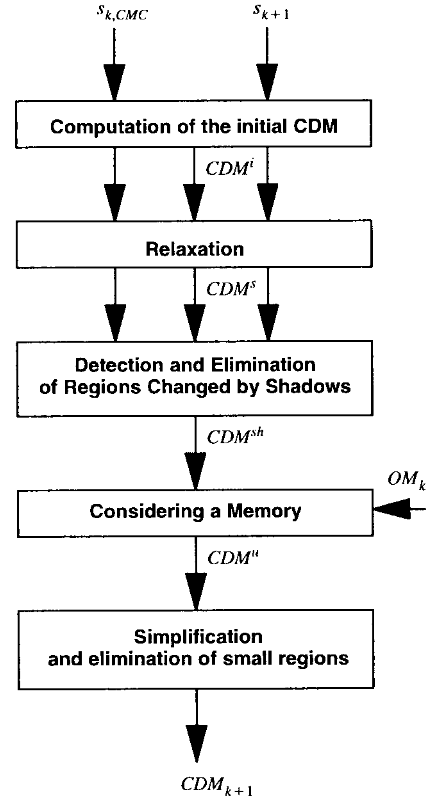


Fig. 10. Block diagram of the algorithm for estimating the change detection mask CDM_{k+1} from two images $s_{k,CMC}$ and s_{k+1} using the previous object mask OM_k as memory.

IV. Segmentation of Moving Objects Considering Moving Shadows

In this section, the shadow detection method from Section III is integrated into the method for 2-D shape estimation of moving objects in an image sequence from [22]. Thus, the resulting algorithm will be able to deal with image sequences captured by a static or moving camera, where moving cast shadows do appear or not. Fig. 9 gives an overview of the proposed segmentation algorithm. It can be subdivided into the following five steps: in the first step, an apparent camera/background motion is estimated and compensated using an eight parameter motion model, assuming that the background of the scene is a rigid plane [15]. Its eight parameters can reflect any kind of camera motion, especially zoom and pan. The algorithm is robust against small model failures, i.e., violations of the assumption of a rigid background plane.

In the second step, a scene cut detector evaluates whether the mean square error between the current original frame s_{k+1} at the current time instant $k+1$ and the camera motion compensated previous frame $s_{k,CMC}$ exceeds a certain threshold [22]. It causes a reset of the segmentation algorithm in these situations, i.e., all parameters are set to their initial value. The evaluation is only performed in background regions of the previous frame which are indicated by the previous object mask OM_k, if available. In OM_k, all pels are set to foreground which belong to a moving object in the previous frame.

By the third step, a change detection mask between current and the previous frame is estimated considering moving cast shadows (see Fig. 10). For that, first an initial change detection mask CDM^i between the current and the previous frame is generated by thresholding the frame difference using a global threshold [21]. Boundaries of changed image regions are smoothed by a relaxation technique using locally adaptive thresholds [1], [2], resulting in a mask denoted as CDM^s. Thereby, the algorithm adapts frame-wise automatically to camera noise [21]. The detected changes in CDM^s are due to both moving objects and moving cast shadows. Thus, the next step is to eliminate pels changed by moving cast shadows from the mask CDM^s using the algorithm described in Section III-D. The resulting mask is denoted as CDM^{sh}. To temporally stabilize the mask CDM^{sh}, it is or-connected with the previous object mask OM_k, resulting in a mask CDM^u [21]. Or-connection means that a pel will be set in the result mask if it is set in at least one of the input masks. This is based on the assumption that all pels which belonged to OM_k should belong to the current change detection mask. However, in order to avoid infinite error propagation, a pel from OM_k is or-connected only if it was also labeled as changed in one of the masks CDM^{sh} of the last L frames. The value L denotes the depth of memory and adapts automatically to the image by evaluating the size and motion amplitudes of the moving objects in the previous frame. At last, the mask CDM^u is simplified and small regions are eliminated, resulting in the final change detection mask CDM_{k+1}.

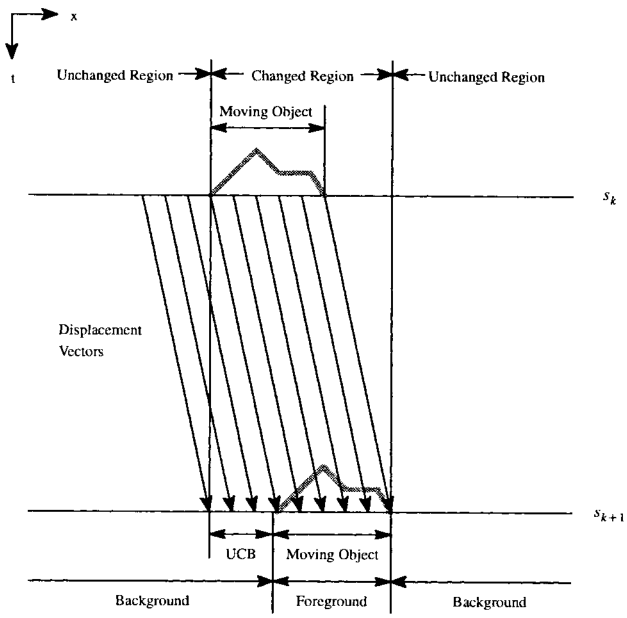


Fig. 11. Example for the separation of changed areas between two images s_k and s_{k+1} into moving object and uncovered background (UCB).

In the fourth step, an initial object mask OM^i is calculated by eliminating background regions from CDM_{k+1} that are uncovered by moving objects [15]. Therefore, displacement information for pels within the changed regions is used. The displacement is estimated by a hierarchical block matcher (HBM) [4]. For a higher accuracy of the calculated displacement vector field, the change detection mask from the first step is considered by the HBM. Uncovered background is detected by pels with foot- or top-point of the corresponding displacement vector being outside the changed regions in CDM_{k+1}. The example in Fig. 11 shows an object moving from the left to the right while uncovering background. The initial object mask OM^i is the mask CDM_{k+1} without uncovered background regions.

Finally, the boundaries of OM^i are adapted to luminance edges in the current image in order to improve the accuracy. The result is the final object mask OM_{k+1}, indicating moving objects in the current frame s_{k+1}.

V. Experimental Results

Simulation results with various test sequences have been obtained. The proposed algorithm was used to segment moving objects while detecting moving cast shadows. For all results, the same algorithmic parameters as indicated in Sections III and IV are used. The segmentation results for the moving objects of the proposed algorithm have been compared to that of the same algorithm without shadow detection as reference. The reference algorithm is identical to [22] and is included in the informative part of the forthcoming international standard ISO/MPEG-4 [25]–[31]. Furthermore, parts of the reference algorithm are included in the COST 211^{quat} Analysis Model [7], [8].

In Fig. 12, a typical video telephone sequence named "Erik" (CIF, 10 Hz) can be seen. The scene is illuminated by diffuse

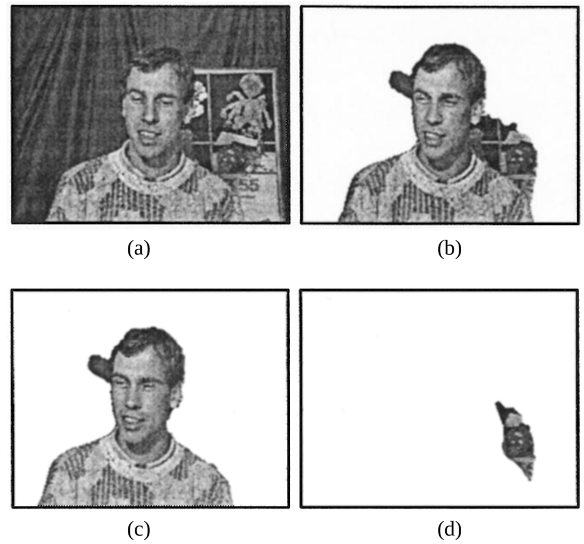


Fig. 12. Segmentation results for image sequence "Erik" (CIF, 10 Hz): (a) original frame 42, (b) detected objects by the reference algorithm, and (c) by the proposed algorithm. In (d), the detected cast shadows are shown.

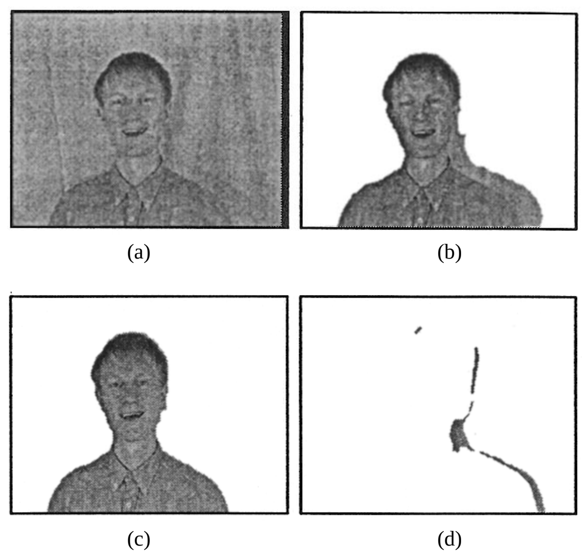
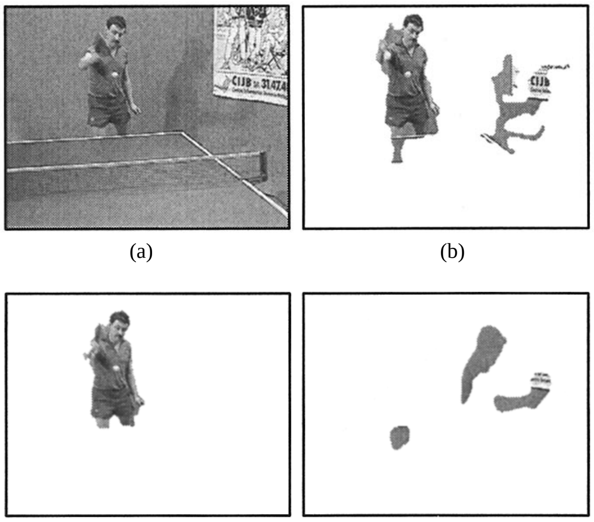


Fig. 13. Segmentation results for image sequence "Jürgen" (CIF, 10 Hz): (a) original frame 24, (b) detected objects by the reference algorithm, and (c) by the proposed algorithm. In (d) , the detected cast shadows are shown.

light and a spot light generating a cast shadow on the left side of the person's neck. The results sho that the cast shadow is detected by the proposed algorithm. The proposed algorithm does not include the shadow into the object mask as the reference algorithm does. A center part of the cast shadow is not detected, because it has been always part of the shadow.

In Fig. 13, the video telephone sequence "Jürgen" (CIF, 10 Hz) is shown. Here, the scene is illuminated from three directions causing several overlapping cast shadows on both sides of the person. As can be seen, the weak and diffuse shadows isolated from the person have not been detected. Due to small object motion, the temporal image signal changes are so weak, that they are not detected by the change detection. Here, assumption 1 from Section III is violated. The mentioned shadows are weak because the scene is illuminated by more than one light source. The shadows with respect to a



(a) (b)

(c) (d)

Fig. 14. Segmentation results for MPEG-4 test sequence “Table Tennis” (CIF, 30 Hz): (a) original frame 151, (b) detected objects by the reference algorithm, and (c) by the proposed algorithm. In (d), the detected cast shadows are shown.

certain light source are illuminated by the other two sources. On the other hand, the shadow directly neighboring the left side of the person is correctly detected as shadow. This shadow is less diffuse because the person is closer to the background.

Results for the test sequence “Table Tennis” (CIF, 10 Hz) shown in Fig. 14 demonstrate the performance of the proposed algorithm in case of camera motion and moving cast shadows. Compared to the reference algorithm, the proposed segmentation algorithm is able to detect shadows and objects separately. The object segmentation in Fig. 14(c) has still some artifacts but is much more accurate than the reference method result in Fig. 14(b). Some regions from the Fig. 14(b) are removed in Fig. 14(c) even if they are not detected as moving cast shadows in Fig. 14(d). This is because small regions contained in the mask of regions changed by moving shadows have been eliminated from the final shadow mask. Further, the shown results depend on results for preceding images. However, these results are different for the reference and the proposed algorithm.

In Fig. 15, the temporal evolution of the shadow integration for the sequence “Table Tennis” is shown. In can be seen that the image regions covered by the cast shadows at the wall are integrated. The lower shadow of the arm is more diffuse and is not detected. There is also a diffuse mirror image of the player on the table. This mirror image moves with the player and is mostly darker than the rest of the table that is illuminated by the brighter wall. Thus, the players mirror image is partially detected as shadow. Regions missing in the integrated shadow have been either part of the shadow since the scene cut and can thus not be detected by an integration of temporal changes, or, they were covered by the moving person and were thus deleted from the mask (see Section III-E).

Some problems of shadow detection can be seen in Fig. 12. Here, the moving object partly covers the cast shadow. Thus, cast shadow and moving object are contained in one region of the change detection mask. Due to the evaluation of neighboring pels for shadow detection, as explained in Section III-D, some shadow artifacts can be seen in the resulting object mask. Another problem exists for the temporal integration explained in Section III-E. The shadow of the head in Fig. 12 or the shadow of the person’s body in Fig. 14 are still not complete for the shown sample images. A cast shadow is detected completely as soon as he has totally moved away from its position from the first frame. Also, regions of background being covered by a moving cast shadow have to be visible to be able to integrate them. If the persons head in Fig. 12 moves further on to the right, the background being covered by the following shadow is behind the head an thus not visible. A successful example is the shadow of the person’s arm in Fig. 14. This shadow is completely visible at all time. Further, its motion is sufficient large to make the shadow leaving quickly its position from the first image after the scene cut. In this case, the tracking works well.

VI. Conclusions

In this paper, a method for detection of moving cast shadows is developed and integrated into an algorithm for segmentation of moving objects from the literature [22]. It is assumed that moving shadows are casted on the dominant background of the scene. They are caused by objects moving between a light source and the background. Moving cast shadows cause a marking frame difference between two consecutive images of a monocular video sequence. Image regions that are changed from frame to frame by moving cast shadows are detected and classified into being covered or uncovered by moving cast shadows. The covered regions are temporally integrated to detect and track entire cast shadows.

The detection of image regions that are changed from frame to frame by moving cast shadows is based on four assumptions on the unknown 3-D scene visible in the image sequence. The assumptions lead to four criteria. First, it is assumed that the shadow causing light source is strong compared to other light sources or camera noise. Then, image regions changed by moving cast shadows have an image difference larger than camera noise. Second, the camera and the background motion is assumed to be zero or compensated in advance. Then, image regions changed by moving cast shadows should not contain any moving edges. The third assumption is a plane scene background. Then, image regions changed by moving cast shadows on the background have a spatially homogeneous luminance ratio between two succeeding frames. Finally, the size of the shadow causing light source and the distance between object and background are assumed to be non negligible compared to the distance between light source and moving object. In this case, shadows have a penumbra causing a slow transition of image luminance from shadowed to nonshadowed regions. These four criteria are combined by heuristic rules resulting in a binary mask indicating image regions changed by moving cast shadows. The rules require that the majority of the criteria decide for a shadow. Also, a local neighborhood is evaluated for each pel.

The detected image regions changed by moving cast shadows are then pel-wise classified into uncovered or covered background if the luminance temporally increases or decreases,

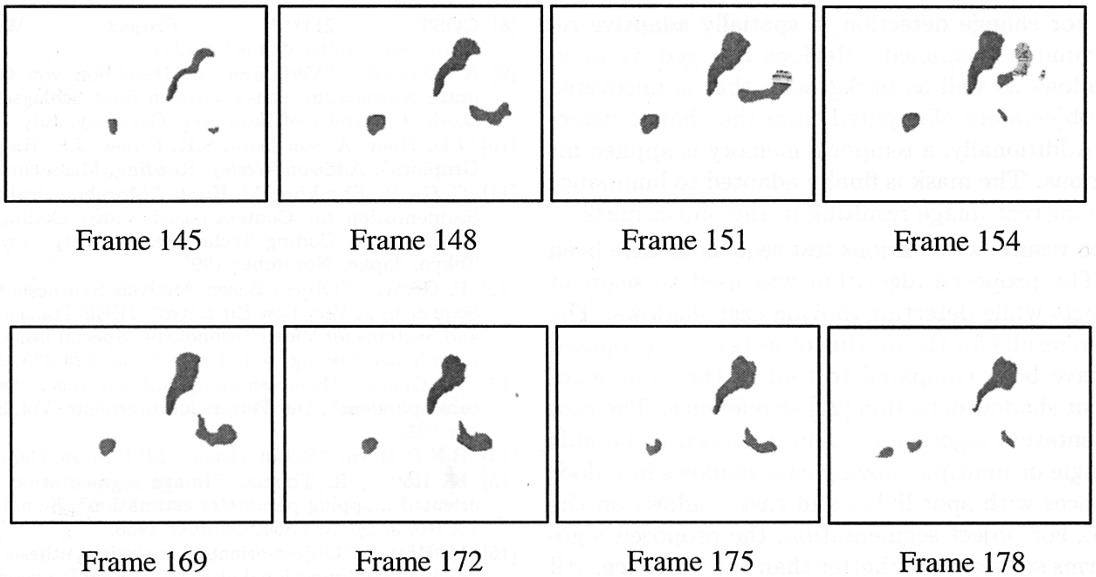


Fig. 15. Examples for temporal evolution of moving cast shadow integration for the test sequence "Table Tennis": Image regions that are newly covered by moving cast shadows are temporally integrated. For the shadow of the players arm, the integration is complete after few images and the arm is well tracked.

respectively. Entire moving cast shadows are detected and tracked by temporal integration of background regions being covered while subtracting background regions being uncovered by shadows.

The method for detection of moving cast shadows has been integrated into a complete method for segmentation of moving objects from [22]. For object segmentation, an apparent camera motion is first estimated and compensated. Then, a possibly apparent scene cut is detected to reset the algorithm in that case. From the current and the motion-compensated previous image, a change detection mask is estimated. For change detection, a spatially adaptive relaxation technique is applied. Regions changed by moving cast shadows as well as background that is uncovered by moving objects are eliminated from the change detection mask. Additionally, a temporal memory is applied for changed regions. The mask is finally adapted to luminance edges in the current image, resulting in the object mask.

Simulation results with various test sequences have been obtained. The proposed algorithm was used to segment moving objects while detecting moving cast shadows. The segmentation results for the moving objects of the proposed algorithm have been compared to that of the same algorithm without shadow detection [22] as a reference. The proposed segmentation algorithm has been shown to be able to detect single or multiple moving cast shadows in indoor video sequences with spot lights and cast shadows on the background. For object segmentation, the proposed algorithm performs substantially better than the reference. All cast shadows that have been detected as objects by the reference algorithm are now excluded from the object mask. Also, the MPEG-4 test sequence "Table Tennis" can be segmented successfully. It is captured by a moving camera and contains moving cast shadows on a wall. Moving objects and moving cast shadows are detected separately.

For shadows that are weak or diffuse, for shadows on highly structured background and for shadows with contours as sharp as object edges, the applied assumptions are violated and shadows may not be detected. Further, the entire shadow regions have to be visible to be tracked successfully. Finally, the presented shadow detection method assumes that the cast shadows move. Shadows can be entirely detected only, if they entirely cover new background along the image sequence. Image regions always shadowed cannot be detected by the proposed algorithm. To solve these problems, future work may address the integration of an intra-frame shadow detection algorithm and the consideration of moving shadows that are occluded temporally by moving objects.

REFERENCES

[1] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Process.*, vol. 31, no. 2, pp. 165–180, Mar. 1993.

[2] ——, "Change detection in image sequences using Gibbs random fields: A Bayesian approach," in *Proc. Int. Workshop on Intelligent Signal Processing and Communication Systems,* Sendai, Japan, Oct. 1993, pp. 56–61.

[3] M. Bejanin, A. Huertas, G. Mediani, and R. Nevatia, "Model validation for change detection," in *Proc. 2nd Int. IEEE Workshop on Applications of Computer Vision,* Los Alamitos, CA, Dec. 5–7, 1994, pp. 160–167.

[4] M. Bierling, "Displacement estimation by hierarchical blockmatching," in *Proc. 3rd SPIE Symp. Visual Communications and Image Processing,* Cambridge, MA, Nov. 1988, pp. 942–951.

[5] A. M. Bruckstein, R. J. Holt, Y. D. Jean, and A. N. Netravali, "On the use of shadows in stance recovery," Tech. Memo. BL0113380-971222-04TM, Bell Labs., Lucent Technol., Dec. 1997.

[6] L. Chiariglione, "MPEG and multimedia communications," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 5–18, Feb. 1997.

[7] COST 211^{ter} Simulation Subgroup: "Description of tools for automatic segmentation of moving objects in video sequences," Doc. SIM(97)13, June 1997.

[8] COST 211^{quat} Project Web Pages: http://www.teltec.dcu.ie/cost211.

[9] N. Elouardi, "Verfahren zur Detektion von Objektsilhouetten unter Ausnutzung von Vorwissen über Schlagschatten," student thesis, Univ. Hannover, Germany, July 1997.

[10] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics.* Reading, MA: Addison-Wesley, 1992.

[11] C. Gu, T. Ebrahimi, and M. Kunt, "Morphological spatio-temporal segmentation for content-based video coding," in *Int. Workshop on Coding Techniques for Very Low Bit-rate Video,* Tokyo, Japan, Nov. 1995.

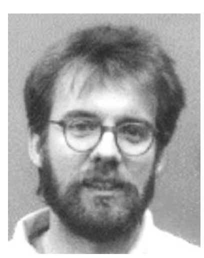
[12] P. Gerken, "Object-based analysis-synthesis coding of image sequences at very low bit rates," *IEEE Trans. Circuits Syst. Video Techn.: Special Issue on Very Low Bit Rate Video Coding*, vol. 4, pp. 228–235, June 1994.
[13] W. Geuen, "Konturfindung auf der Basis des visuellen Konturempfindens," *Der Fernmelde-Ingenieur*, vol. 37, no. 10, Oct. 1983.
[14] B. K. P. Horn, *Robot Vision.* Cambridge: MIT Press, 1987.
[15] M. Hötter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation," *Signal Process.*, vol. 15, no. 3, pp. 315–334, Oct. 1988.
[16] M. Hötter, "Object-oriented analysis-synthesis coding based on moving two-dimensional objects," *Signal Process.: Image Commun.*, vol. 2, no. 4, pp. 409–428, Dec. 1990.
[17] C. Jiang and M. O. Ward, "Shadow segmentation and classification in a constrained environment," *CVGIP: Image Understanding*, band 59, no. 2, pp. 213–225, Mar. 1994.
[18] D. Koller, K. Danilidis, and H.-H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *Int. J. Comput. Vis.*, vol. 10, no. 3, pp. 257–281, 1993.
[19] J. H. Lambert, *Lambert's Photometrie (Photometria sive de mensura de gratibus luminis, colorum et umbrae).* Leipzig, Germany: Wilhelm Engelmann, 1892.
[20] B. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, June 1975.
[21] R. Mech and M. Wollborn, "A noise robust method for segmentation of moving objects in video sequences," in *Int. Conf. Acoustics, Speech, and Signal Processing*, Münich, Germany, Apr. 1997.
[22] ——, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera," in *Workshop on Image Analysis for Multimedia Interactive Services*, Louvain-la-Neuve, Belgium, June 1997.
[23] MPEG-4 Video Group: Technical Description of the Video Encoder (Proposel for MPEG-4 Tests), Doc. ISO/IEC JTC1/SC29/WG11 MPEG95/504, Dec. 1995.
[24] MPEG-4 Video Group: MPEG-4 Video Verification Model Version 7.0, Doc. ISO/IEC JTC1/SC29/WG11 N1642, Apr. 1997.
[25] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG96/841, Mar. 1996.
[26] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG96/989, June 1996.
[27] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG96/1188, Sept. 1996.
[28] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG96/1549, Nov. 1996.
[29] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/1831, Feb. 1997.
[30] MPEG-4 Video Group: Automatic Segmentation of moving objects (core experiment N2), Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/1949, Apr. 1997.
[31] MPEG-4 Video Group: "Information Technology—Coding of Audio-visual Objects: Visual, ISO/IEC 14496-2, Committee Draft," Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/1902, Oct. 1997.
[32] H. G. Musmann, M. Hötter, and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images," *Signal Process.: Image Commun.*, vol. 1, no. 2, pp. 117–138, Oct. 1989.
[33] J. Ostermann, "Modeling of 3D moving objects for an analysis-synthesis coder," in *Proc. Conf. Sensing and Reconstruction of 3D-Objects and Scenes*, Atlanta, GA, Feb. 15–16, 1990, SPIE vol. 1260, pp. 240–249.
[34] ——, "Object-based analysis-synthesis coding (OBASC) based on the source model of moving rigid 3D objects," *Signal Process.: Image Commun.*, vol. 6, no. 2, pp. 143–161, May 1994.
[35] ——, "Segmentation of image areas changed due to object motion considering shadows," *Multimedia Communications and Video Coding*, Y. Wang, Ed. New York: Plenum, 1996.
[36] J. Ostermann and A. Puri, "Natural and synthetic video in MPEG-4," in *Proc. ICASSP'98*, May 1998.
[37] F. Pedersini, A. Sarti, and S. Tubaro, "Combined motion and edge analysis for a layer-based representation of image sequences," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 1996.
[38] F. Pereira, "MPEG-4: A new challenge for the representation of audio-visual information," keynote speech at Picture Coding Symp. PCS'96, Melbourne, Mar. 1996.
[39] G. G. Sexton and X. Zhang, "Supression of shadows for improved object discrimination," in *Proc. IEEE Colloq. Image Processing for Transport Applications*, London, U.K., Dec. 1993, pp. 9/1–9/6.
[40] J. Stauder, "Detection of shading for object-based coding," in *Proc. Picture Coding Symp. (PCS)*, Melbourne, Australia, Mar. 13–15, 1996, pp. 325–328.
[41] ——, "Estimation of point light source parameters for object-based coding," *Signal Process.: Image Commun.*, vol. 7, no. 4–6, pp. 355–379, Nov. 1995.
[42] ——, "Segmenation of moving objects in presence of moving shadows," in *Proc. Int. Workshop on Coding Techniques for Very Low Bit-rate Video*, Linköping, Sweden, July 28–30, 1997, pp. 41–44.
[43] C. Stiller, "Object-oriented video coding employing dense motion fields," in *Proc. Int. Conf. Acoustic, Speech and Signal Processing*, Adelaide, South Australia, Apr. 1994.
[44] ——, "Object-based estimation of dense motion fields," *IEEE Trans. Image Processing*, vol. 6, pp. 234–250, Feb. 1997.
[45] A. Watt, *3D Computer Graphics.* Reading, MA: Addison-Wesley, 1993.
[46] M. Wollborn, M. Kampmann, and R. Mech, "Content-based coding of videophone sequences using automatic face detection," in *Proc. Picture Coding Symp.*, Berlin, Germany, Sept. 1997.
[47] W. Zhang and F. Bergholm, "An extension of Marr's signature based edge classification and other methods determining diffuseness and height of edges, and bar edge weight," in *Int. Conf. Computer Vision*, Berlin, May 11–14, 1993, pp. 183–191.

Jürgen Stauder received the Dipl.-Ing. degree in electrical engineering from the University of Darmstadt, Darmstadt, Germany, in 1990. From 1990 to 1998, he was with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung at the University of Hannover, Germany as a Research Assistant, where he will defend his Ph.D. thesis on illumination estimation.

Since 1998, he has been with the IRISA Laboratory of INRIA, Rennes, France, as a Visiting Researcher. He contributes to the European research project COST 211 and participates in the European TMR program. His research and education interests are in the field of image sequence processing, where he works on illumination estimation, object segmentation, shadow detection, augmented reality, and 3-D object-based video coding.

Roland Mech received the Dipl.-Inform. degree in computer science from the University of Dortmund, Dortmund, Germany, in 1995.

Since 1995, he has been with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, Germany, where he works in the areas of image sequence analysis and image sequence coding. He is a member of the European projects COST 211 and ACTS-MoMuSys. He contributes actively to the ISO/MPEG-4 standardization activities. His present research interests and activities cover image sequence analysis, especially 2-D shape estimation of moving objects and the application to object-based image coding.



Jörn Ostermann studied electrical engineering and communications engineering at the University of Hannover, Hannover, Germany, and Imperial College London, London, U.K., respectively. He received the Dipl.-Ing. and Dr.-Ing. degrees from the University of Hannover in 1988 and 1994, respectively.

From 1988 to 1994, he worked as a Research Assistant at the Institut für Theoretische Nachrichtentechnik conducting research in low bit-rate and object-based analysis-synthesis video coding. In 1994 and 1995, he worked in the Visual Communications Research Department at AT&T Bell Laboratories, Red Bank, NJ. He has been working with Image Processing and Technology Research within AT&T Labs–Research since 1996. From 1993 to 1994, he chaired the European COST 211 simulation group coordinating research in low bit rate video coding. Within MPEG-4, he organized the evaluation of video tools to start defining the standard. Currently, he chairs the Adhoc Group on Coding of Arbitrarily-shaped Objects in MPEG-4 Video. His current research interests are video coding, computer vision, 3-D modeling, face animation, coarticulation of acoustic and visual speech, computer–human interfaces, and speech synthesis.