
Xu, Dong; Liu, Jianzhuang; Li, Xuelong; Liu, Zhengkai and Tang, Xiaoou (2005). Insignificant shadow detection for video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology* **15** (8) 1058-1064.

Insignificant Shadow Detection for Video Segmentation

Dong Xu, Jianzhuang Liu, *Senior Member, IEEE*, Xuelong Li, *Member, IEEE*, Zhengkai Liu, and Xiaoou Tang, *Senior Member, IEEE*

***Abstract*—To prevent moving cast shadows from being misunderstood as part of moving objects in change detection based video segmentation, this paper proposes a novel approach to the cast shadow detection based on the edge and region information in multiple frames. First, an initial change detection mask containing moving objects and cast shadows is obtained. Then a Canny edge map is generated. After that, the shadow region is detected and removed through multiframe integration, edge matching, and region growing. Finally, a post processing procedure is used to eliminate noise and tune the boundaries of the objects. Our approach can be used for video segmentation in indoor environment. The experimental results demonstrate its good performance.**

***Index Terms*—Insignificant shadow detection, multiframe integration, region growing, video segmentation.**

I. Introduction

VIDEO segmentation is a key technique for semantic object extraction and plays an important role in content-based video coding in MPEG4 [4], [5], [11]–[13], [15], [16], [18], [22]. It is often necessary to separate moving objects from their shadows in video segmentation for many applications such as traffic monitoring, surveillance and video conferencing [17]. There are some different shadow classifications in the literature. According to the scene, there are indoor shadow and outdoor shadow. Shadows can also be classified as ego shadow and cast shadow [9]. The former is part of the object that is not illuminated by direct light, while the latter is part of the background where direct light is blocked by the object. When a light source that causes shadows cannot be treated as a point, a penumbra (a partial shadow) and an umbra (a shadow where light is completely blocked) will appear simultaneously [20].

According to shadow extent, it can be categorized into significant shadow and insignificant shadow. Shadow significance is a relative concept. A shadow is called significant when the edge of the shadow region is as sharp as the edge of its corresponding object. Otherwise, it is considered as insignificant. Shadow significance depends on several factors such as light source intensity, the width of a penumbra, and the background. In sunny days and on cement roads, shadows cast by pedestrians or vehicles can be considered as significant shadows, where the light is strong with no penumbra area, and the contrast is large between the cement roads and the dark shadows. On the other hand, in normal indoor environment, there is often more than one light in a scene and the size of a light cannot be ignored when compared with the distance between the light and a moving object. Thus, penumbras usually appear and they may be quite large, forming insignificant shadows.

In the literature, much work has been done on significant shadow detection [3], [6], [7], [10], [14], [17], [19], [21]. However, less work [4], [6], [20] can be found on insignificant shadow detection, which is more challenging than the former. In [6], Cucchiara *et al.* used some shadow properties in HSV color space to distinguish shadows from moving objects. These properties show that the cast shadow darkens the background in luminance component, and the hue and saturation components change within certain limits. The main drawback of this method is that, when a video sequence is noisy such as one acquired by a low-end camera, pixel-wise classification of moving objects and shadows is not reliable. Other efforts like pre-processing and post-processing are needed [17]. Moreover, pixel-wise processing makes the parameters used in the algorithm sensitive with respect to the image luminance and gradient [6]. We believe that edge and region information should be employed for more robust shadow detection.

Stauder *et al.* [20] analyzed the moving cast shadow behavior extensively and based their approach on four assumptions: 1) a strong light source causing a cast shadow; 2) a static camera, and static and textured background; 3) planar background; and 4) the light source with a certain extent. A cast shadow is found mainly according to the results of change detection, static edge detection, shading change detection and penumbra detection. There are some problems in this approach. Some regions of a moving object, such as the facial part of a human, are easy to be misclassified as shadow regions because the uniform colors there present the same characteristics as the shadow regions. The regions that are always shadowed along the sequence cannot be detected by their algorithm, as pointed out by the authors in [20]. If the assumptions are violated, the shadow will not be detected correctly [17], [20]. Moreover, the computation is quite complex [4], [17].

Chien *et al.* [4] applied a gradient filter to remove moving cast shadows for video segmentation. The reason to use the gradient filter is that in general, a shadow region tends to have a gradual change in luminance value. Moreover, if the illumination or the camera gain changes within the sequence, the shadow effect is small in the gradient domain. As they pointed out, however, this approach has some limitations. The removal of shadow effect relies on smooth change in the shadow region. If a shadow appears in a region with strong texture, the benefit of the gradient

Manuscript received January 4, 2004; revised November 20, 2004. The work was supported by the Natural Science Fund, Anhui, China under Grant 03042307, and in part by the Research Grants Council of the Hong Kong SAR. This paper was recommended by Associate Editor H. Shum.

D. Xu and Z. Liu are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: dongxu@mail.ustc.edu.cn; zhengkai@ustc.edu.cn).

J. Liu and X. Tang is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: jzliu@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

X. Li is with the School of Computer Science and Information Systems, University of London, Birkbeck College, London WC1E 7HX, U.K. (e-mail: xuelong@dcs.bbk.ac.uk).

Digital Object Identifier 10.1109/TCSVT.2005.852402

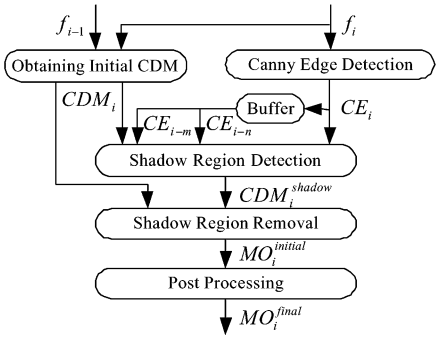

Fig. 1. Block diagram of the cast shadow detection algorithm.

filter will be lost. In addition, when the moving object has weak edges and is less textured, the segmentation result may degrade since the gradient filter also removes some useful information from the original image.

In this paper, we propose an effective approach to the detection and removal of moving cast shadows in normal indoor scenes where the camera is stationary. It is especially appropriate for the applications of indoor video surveillance and conferencing. The approach includes: 1) the generations of initial change detection masks and Canny edge maps; 2) shadow region detection by multiframe integration, edge matching, and region growing; and 3) shadow region removal and post-processing for eliminating noise and tuning object boundaries. The main contribution of this paper is that we successfully separate a moving object from its cast shadow by a region growing scheme, where the selection of seeds in shadow regions and the growing stop criterion are decided by the edge information in multiple frames. We have compared our approach with the gradient filter in [4], which is the most recent and related to our work. The experimental results show that our approach performs better.

II. Cast Shadow Detection Algorithm

A. Outline of the Algorithm

The block diagram of the algorithm is shown in Fig. 1. Given a sequence $f_1, f_2, \ldots, f_i, \ldots$, first an initial change detection mask (CDM) at frame i, CDM_i, is obtained. The Canny edge map of this frame CE_i is also generated. Then CE_i, CE_{i-n} and CE_{i-m} are fed into the step of the shadow region detection, where CE_{i-n} and CE_{i-m} are the Canny edge maps of two previous frames $i-n$ and $i-m$, respectively. The output of this step, denoted by $\mathrm{CDM}_i^{\mathrm{shadow}}$, is the detected shadow region in CDM_i. With CDM_i and $\mathrm{CDM}_i^{\mathrm{shadow}}$, the initial moving object $\mathrm{MO}_i^{\mathrm{initial}}$ can be created directly. The last step is the post-processing for eliminating noise and tuning the boundary of the moving object, from which we have the final detected moving object $\mathrm{MO}_i^{\mathrm{final}}$.

B. Creating Initial CDMs and Canny Edge Maps

To obtain the initial CDM for a frame is an important step for moving object segmentation. This mask usually contains moving objects and cast shadows. We adopt the background registration algorithm proposed by Chien *et al.* [4] to create it. This scheme has been verified, both in [4] and in our experiments, to have better performance than the commonly used frame difference technique [12]. Figs. 2(b) and 3(b) shows an example of CDM_i.

We use the popular Canny edge detection algorithm [2] to generate an edge map because it is excellent for obtaining precise edges. The result CE_i is of values 0 and 255, as shown in Figs. 2(c) and 3(c).

From Figs. 2(b) and 3(b), we see that the moving cast shadows appear in the initial CDMs. However, the fact that the transition from the backgrounds to the shadow regions is gradual makes the edges caused by the shadow boundaries almost invisible in Figs. 2(c) and 3(c). This phenomenon exists in indoor environment, and is very useful for the detection of indoor shadows.

C. Shadow Region Detection

The block diagram of the shadow region detection is shown in Fig. 4, which is the kernel of our approach. The detailed explanations of the steps are given in the following.

1) Multiframe Integration: In normal indoor environment, an insignificant shadow often appears in an area where the gray levels change gradually from the background to the shadowed region. An obvious example is shown in Fig. 3(a). Therefore, we can use the Canny edge operation to suppress the weak edges. From Fig. 3(c), we see that few edges exist in the shadow region. The effect of the shadow has been excluded from the edge map. For this example, it is not difficult to remove the shadow from the moving person with the information in Fig. 3(b) and (c). However, in some video sequences, shadows appear in highly textured background, as shown in Fig. 2, where the textured backgrounds bring the edges and affect moving object detection. Therefore, how to remove those edges in shadow regions becomes an important and challenging task.

Edges can be classified as *static* edges and *moving* edges. The former belong to the background, and the latter to the moving object. Moving Canny edges are the most important cue for the region growing algorithm discussed later. Thus, it is necessary to distinguish the moving edges from static edges. In [20], Stauder *et al.* used the information in two consecutive frames to find static and moving edges. But this method is easy to generate many *false negative edges*, which denotes the moving edges considered as static edges. Fig. 5 shows an example. The original interior Canny edges in Fig. 5(a) are the Canny edges lying inside CDM_i. (In what follows, the term *interior* is used to denote *inside* CDM_i.) Fig. 5(b) shows the interior moving Canny edges obtained by the two-consecutive-frame-based method. Comparing Fig. 5(a) and (b), we see that many moving edges are lost (classified as static edges). In our experiments, we found that these false negative edges may result in seriously bad outcome. For example, there are many gaps in the person's boundary in Fig. 5(b), indicating that some moving edges have been considered as static edges. Through these gaps, the region growing algorithm may grow into the head and shoulder of the person and thus misclassify part of the moving person into the shadow region.

Now, we propose a scheme called multiframe integration to remove the edges in the background and to preserve the edges of the moving object as much as possible. From our experiments, we have found that three frames are suitable to achieve this goal.

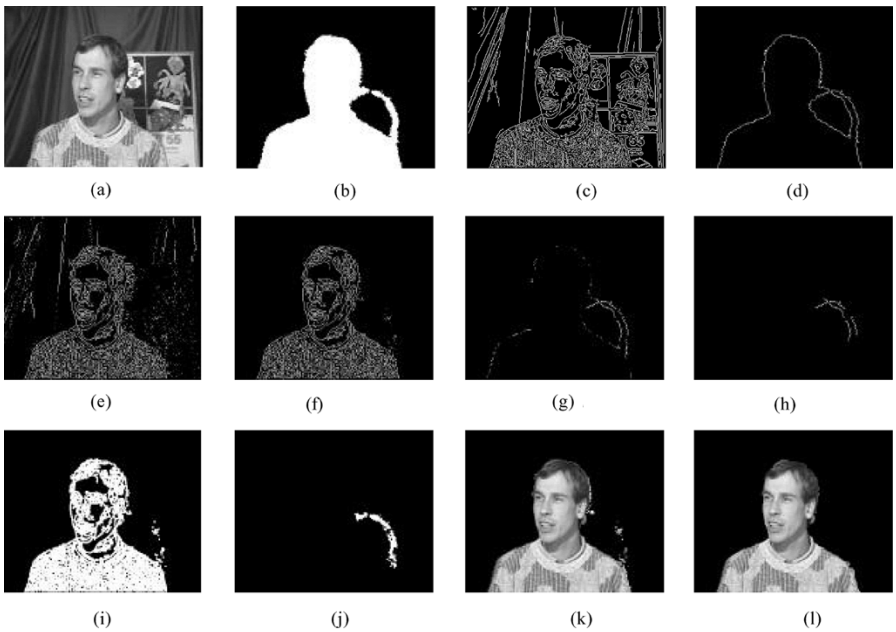

Fig. 2. Illustration of the results of the main steps in our algorithm with frame 46 of the video sequence Erik. (a) f_{46} of Erik. (b) CMD_{46}. (c) CE_{46}. (d) ECMD_{46}. (e) ICE_{46}. (f) $\mathrm{MCE}_{46}^{\mathrm{interior}}$. (g) SRE_{46}. (h) $\mathrm{SRE}_{46}^{\mathrm{final}}$. (i) $\mathrm{DMCE}_{46}^{\mathrm{interior}}$. (j) $\mathrm{CMD}_{46}^{\mathrm{shadow}}$. (k) $\mathrm{MO}_{46}^{\mathrm{intial}}$. (l) $\mathrm{MO}_{46}^{\mathrm{final}}$.

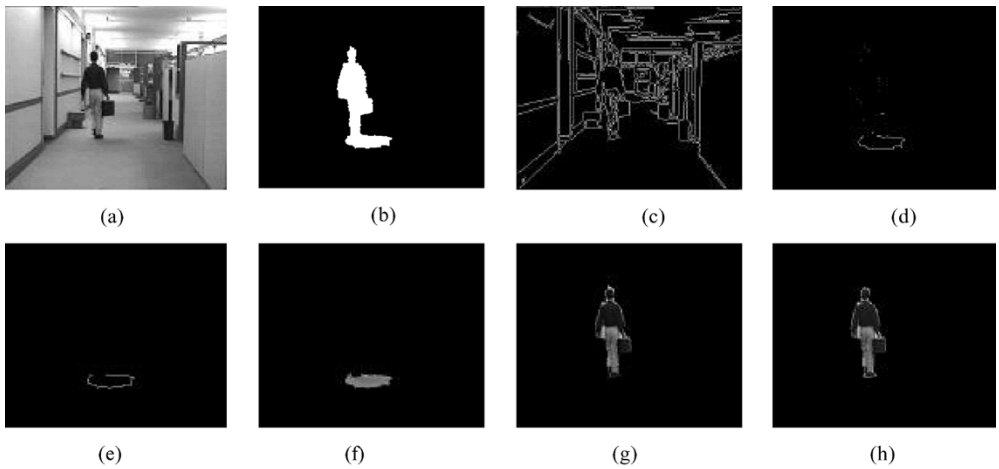

Fig. 3. Main results for the frame 53 of the video sequence Hall. (a) f_{53} of Hall. (b) CMD_{53}. (c) CE_{53}. (d) SRE_{53}. (e) $\mathrm{SRE}_{53}^{\mathrm{final}}$. (f) $\mathrm{CDM}_{53}^{\mathrm{shadow}}$. (g) $\mathrm{MO}_{53}^{\mathrm{initial}}$. (h) $\mathrm{MO}_{53}^{\mathrm{final}}$.

The integrated edge information is kept in ICE_i, which is defined in (1), shown at the bottom of the page, where p denotes a pixel. In this scheme, the edge points in the Canny edge map of the current frame i are considered as static edges only when they are also the Canny edges in the past frames $i-n$ and $i-m$. Thus, it is not easy to misclassify the moving object edges as the static edges, preserving the object edges well, and being able to remove the edges in the background effectively (see (4) below where the static edges in ICE_i are discarded).

In our experiments, n and m are chosen as 3 and 5, respectively; they are not set to 1 and 2, because the former choice can better preserve edges when the moving object moves very slowly. Fig. 5(c) gives an example obtained by our three-frame-based method where the moving Canny edges in CDM_i are shown. It is not difficult to see that compared with Fig. 5(b), there are much fewer false negative edges, and the boundary of the moving object is preserved better. Fig. 2(e) shows another example with all the moving Canny edges given. Similar results obtained in our experiments reveal that our multiframe Canny edge integration technique can be used to detect shadow regions in both highly textured and smooth backgrounds. The following analysis shows the relations among the two-frame-based

$$\mathrm{ICE}_i(p) = \begin{cases} \text{not edge}, & \text{if } \mathrm{CE}_i(p) = 0 \\ \text{static edge}, & \text{if } \mathrm{CE}_i(p) = \mathrm{CE}_{i-n}(p) = \mathrm{CE}_{i-m}(p) = 255 \\ \text{moving edge}, & \text{otherwise} \end{cases} \tag{1}$$

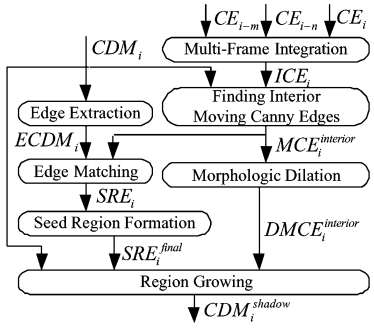

Fig. 4. Block diagram of the shadow region detection.

Fig. 5. Generating interior moving Canny edges by the two methods. (a) Original interior Canny edges. (b) Interior moving Canny edges obtained by the two-frame-based method. (c) Interior moving Canny edges obtained by our multiframe-based method.

method, the multiframe-based method, and the false negative edges.

An edge map can be treated as a grey-level image before converted to a binary one in the final step of the Canny edge detection. Our analysis is based on these grey-level edge maps. Let FD be the frame difference of two edge maps. Let β_t and β_m be the false negative rates generated by the two-frame-based and multiframe-based methods, respectively. We define the *false negative rate* as the probability of a moving edge point classified as a static one. Let H be the hypothesis that a pixel is a moving edge point. Then

$$\beta_t = \mathrm{Prob}(|\mathrm{FD}| < \mathrm{TH}_t \,|\, H) \tag{2}$$

$$\beta_m = [\mathrm{Prob}(|\mathrm{FD}| < \mathrm{TH}_m \,|\, H)]^{r-1} \tag{3}$$

where r is the number of edge maps in the multiframe-based method, all the edge point in FD's are assumed to be independentaly identically distributed (i.i.d.), and TH_t and TH_m are the decision thresholds for the two methods. With these analysis and assumptions, we have $\beta_t > \beta_m$ if $\mathrm{TH}_t = \mathrm{TH}_m$. Therefore, our multiframe-based method can produce less false negative edges.

The multiframe integration may produce more *false positive edges* than the two-frame-based method, which can be found in the shadow regions in Fig. 5(b) and (c). Here false positive edges denote the static edges that are misclassified as moving edges. However, this drawback is not serious in our approach. These edges have little effect on the region growing described later and thus little effect on the shadow detection. Moreover, they can be removed in the postprocessing step.

2) Finding Interior Moving Canny Edges and the Edges of CDM_i*:* With ICE_i and CDM_i, it is not difficult to find the interior moving Canny edges, $\mathrm{MCE}_i^{\mathrm{interior}}$, and the edges of CDM_i and ECDM_i. Two sets are defined for obtaining them

$$\mathrm{MCE}_i^{\mathrm{interior}} = \{p \,|\, \mathrm{ICE}_i(p) = \text{moving Canny edge}, p \in \mathrm{CDM}_i\} \tag{4}$$

$$\mathrm{ECDM}_i = \{p \,|\, p \in \mathrm{CDM}_i, p' \notin \mathrm{CDM}_i, \mathrm{NG}(p,p') = \text{true}\} \tag{5}$$

where p and p' are pixels, and $\mathrm{NG}(p,p')$ is a function that returns *true* when p and p' are in 4-neighborhood and *false* otherwise. Fig. 2(d) shows an example of ECDM_i, and Figs. 2(f) and 5(c) give two $\mathrm{MCE}_i^{\mathrm{interior}}$. The interior moving Canny edges $\mathrm{MCE}_i^{\mathrm{interior}}$ provide important information for the seed point selection and region growing that are discussed later.

3) Edge Matching, Seed Region Formation, and Morphologic Dilation: Because we use a region growing algorithm to detect the cast shadow, we have to generate some seed points at first. The edges of the shadow region in CDM_i are good seeds. Therefore, we hope to separate the shadow region edges (SRE_i) in ECDM_i from the moving object edges (MOE_i). This can be achieved with ECDM_i and $\mathrm{MCE}_i^{\mathrm{interior}}$ because most edge points in MOE_i have *matching* (i.e., corresponding) moving edges in $\mathrm{MCE}_i^{\mathrm{interior}}$ when a small neighbor of each point is considered, while most edge points in SRE_i fail to find such matching edges [see Fig. 2(d) and (f)]. Thus, we use the following equation to obtain SRE_i:

$$\mathrm{SRE}_i = \{p \,|\, p \in \mathrm{ECDM}_i, \min \|p - p'\| > T_{\mathrm{search}}, p' \in \mathrm{MCE}_i^{\mathrm{interior}}\} \tag{6}$$

where T_{search} (chosen as 2) is a threshold that confines the searching neighbor, and $\|\cdot\|$ denotes the chosen chessboard distance. The set SRE_i serves as the initial seed points for the region growing algorithm.

Fig. 2(g) shows an example of SRE_i. Comparing Fig. 2(b) and (g), it can be found that the connected white pixels correspond to the edges of the shadow region very well. However, some sporadic pixels, which belong to the edges of the moving object, also appear in SRE_{46}. In order to remove these points, the component connection algorithm in [8] is implemented here to connect the initial seed points. Among all the connected regions, only those with points more than a threshold $T_{\mathrm{minseedpoint}}$ (chosen as 30) are used to be the final seed region of edges $(\mathrm{SRE}_i^{\mathrm{final}})$ in the shadow region. Fig. 2(h) gives an example of $\mathrm{SRE}_i^{\mathrm{final}}$ where the moving edge points in Fig. 2(g) have been removed and the remaining seed points all belong to the shadow region. Although we have focused on the discussion for textured background, our approach works also for smooth background. An example is given in Figs. 3(d) and (e).

The morphologic dilation step in Fig. 4 dilates $\mathrm{MCE}_i^{\mathrm{interior}}$, the result of which, $\mathrm{DMCE}_i^{\mathrm{interior}}$, is part of the input to the region growing algorithm. This operation is used to fill small gaps possibly existing on the common boundary between the moving object and the shadow region. These gaps are caused by the wrong connection in the heuristic searching step of the Canny edge detection algorithm [2]. The structure element is a 3×3 square. Fig. 2(i) shows an example of this operation.

4) Region Growing: The region growing algorithm in [1] is modified for the shadow detection. The algorithm begins with the seed points in the shadow region (i.e., those in $\mathrm{SRE}_i^{\mathrm{final}}$), expands them pixel by pixel in CDM_i, and puts the detected shadow points in a set denoted by $\mathrm{CDM}_i^{\mathrm{shadow}}$. It will stop when the region cannot grow anymore. Note that the points in $\mathrm{DMCE}_i^{\mathrm{interior}}$, which mainly belong to the moving object, should not be put in $\mathrm{CDM}_i^{\mathrm{shadow}}$. The pseudocode of the modified region growing algorithm is described as follows.

1. Initialization: Put points $\in \mathrm{SRE}_i^{\mathrm{final}}$ into a circle queue cq **and** set $\mathrm{CDM}_i^{\mathrm{shadow}} = \emptyset$
2. **while** cq $\neq \emptyset$ **do**
3. **begin**
4. Remove a point x from cq
5. Label x as visited **and** put x into $\mathrm{CDM}_i^{\mathrm{shadow}}$
6. **for** every neighbor y of x in 4-neighborhood **do**
7. **if** $y \in \mathrm{CDM}_i$ **and** $y \notin \mathrm{DMCE}_i^{\mathrm{interior}}$ **and** y has not been visited **then** put y into cq
8. **end**

Figs. 2(j) and 3(f) show two examples of the region growing. From the results, it can be seen that the shadow regions in CDM_i are detected very well.

D. Removing the Shadow Region and Post-Processing

With CDM_i and $\mathrm{CDM}_i^{\mathrm{shadow}}$, the initial moving object mask, $\mathrm{MOM}_i^{\mathrm{initial}}$, can be obtained by $\mathrm{MOM}_i^{\mathrm{initial}} = \mathrm{CDM}_i - \mathrm{CDM}_i^{\mathrm{shadow}}$. The detected initial moving object $\mathrm{MO}_i^{\mathrm{initial}}$ is simply the region of the original frame in $\mathrm{MOM}_i^{\mathrm{initial}}$. Figs. 2(k) and 3(g) give two examples of $\mathrm{MO}_i^{\mathrm{initial}}$, in which there exist some noise points that do not belong to the moving person. There noise points can be eliminated by the post-processing described below.

Several post-processing operations on $\mathrm{MO}_i^{\mathrm{initial}}$ are applied to remove the noise and tune the boundary of the moving object. First, a morphologic erosion is applied with a 5×5 square structure element. Second, the component connection algorithm [8] is employed again to remove the small connected regions, which are noise in the background. Finally, a morphologic dilation operation with a 3×3 square structure element is used to compensate the effect of the morphologic erosion. The final detected moving object in frame i is denoted by $\mathrm{MO}_i^{\mathrm{final}}$, two examples of which are given in Figs. 2(l) and 3(h), where satisfactory results are obtained.

III. Experimental Results

Standard MPEG-4 indoor testing sequences have been used to test our algorithm. Both objective and subjective evaluations are carried out in the experiments. We also compare our approach with the most recent and related one, namely the gradient filter, proposed in [4].

In our algorithm, several parameters need to be pre-defined. The low and the high thresholds in the Canny edge detection algorithm are set to 0.2 and 0.8. The frame intervals n and m in Section II-C.1, and the two thresholds T_{search} and $T_{\mathrm{minseedpoint}}$ in Section II-C.3, are chosen as 3, 5, 2, and 30, respectively. The three structure elements for the morphologic operations in Section II-C.3 and Section II-D are three squares of 3×3, 5×5, and 3×3 pixels. These parameters are chosen by some initial experiments, and then they are fixed for all the experiments described in the following.

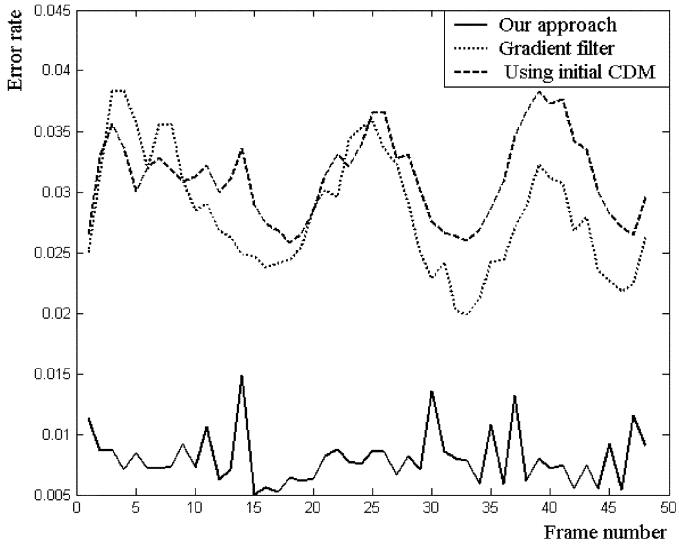

Fig. 6. Error rates in each frame of the Erik sequence for three approaches.

A. Objective Evaluation

Similarly to [4], we use the error rate ε of the final detected object mask $\mathrm{MO}_i^{\mathrm{final}}$ to show the effectiveness of our approach. It is defined as $\varepsilon = N_e/N_F$, where N_F is the frame size, and N_e is the number of pixels in $\mathrm{MO}_i^{\mathrm{final}}$ that are different from the reference alpha plane.

The test sequence is Erik, a typical video telephone sequence in format of CIF 352×288 at 10 f/s, where diffuse light and a spot light illuminate the scene. The cast shadow is formed on the left side of the person's neck. Fig. 6 shows the error rates in every frame of the Erik sequence for three approaches: ours, the gradient filter, and using CDM_i as the detected object mask. For the third approach, to obtain CDM_i, the close and open morphologic operations are applied for tuning the boundary of the object, as in [4]. Not surprisingly, using CDM_i as the object mask presents the largest average error rate 3.1% since no any technique is applied to remove the shadows. The gradient filter cannot well handle the case where shadows appear in textured background. Its average error rate is 2.8%. However, our approach performs best with a much lower average error rate 0.8%. Moreover, even the largest error rate of our approach is less than the smallest error rate of the gradient filter for this sequence.

Table I gives the average processing time in each step of our algorithm (in Visual C++) for Erik on a Pentium IV-1.4 GHz PC with 256 MB RAM. The average time needed to deal with one frame is 124.1 ms. In this first version of our algorithm, no much effort was paid on the optimization of the source code. We believe that a speed of 30 f/s can be reached with a faster PC and/or after some algorithmic optimization (e.g., programming some procedures in Assemble).

TABLE I
AVERAGE PROCESSING TIME IN EACH STEP OF OUR ALGORITHM

		Average Time (ms)	
Canny Edge Detection			35.0
Obtaining Initial CDM			34.6
Shadow Region Detection	Multi-Frame Integration	1.3	22.1
	Edge Extraction	1.2	
	Finding Interior Moving Canny Edges	1.0	
	Edge Matching	0.4	
	Seed Region Formation	6.0	
	Morphologic Dilation	12.0	
	Region Growing	0.2	
Shadow Region Removal			1.4
Post Processing			31
Total			124.1

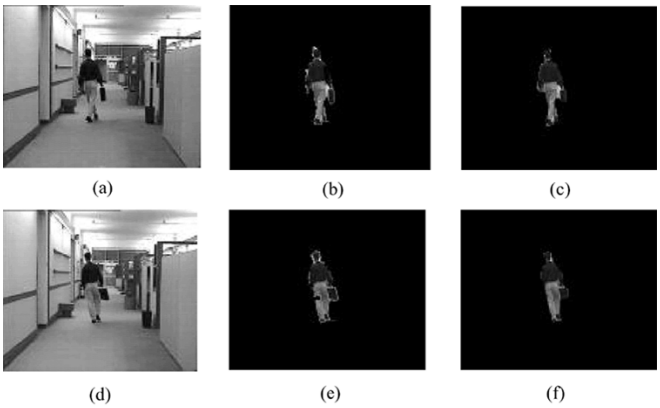

Fig. 7. Comparison of the gradient filter and our approach for Hall frames 60 and 70. (a) Hall frame 60. (b) Result by GF. (c) Result by ours. (d) Hall frame 70. (e) Result by GF. (f) Result by ours.

B. Subjective Evaluation

Three most commonly used MPEG-4 indoor testing sequences, Hall (CIF 352×288 at 30 f/s), Erik (CIF 352×288 at 10 f/s), and Silent (CIF 352×288 at 30 f/s), have been used to test our algorithm. For all the frames in these sequences, our algorithm can obtain satisfactory results of moving object segmentation. We have also compared our approach with the gradient filter (GF) in [4] and found that our algorithm performs better.

Fig. 7 shows the results obtained by our algorithm and the gradient filter for Hall frames 60 and 70. We see that some parts of the moving person are misclassified by the gradient filter due to the weak edges and less texture on the walking person. There is a gap on the right arm in Fig. 7(b) and a gap on the left leg in Fig. 7(e) of the person. On the other hand, our algorithm presents very good results [Fig. 7(c) and (f)] Fig. 8 shows the comparisons between the two approaches for Erik frames 18 and 38. In this sequence, the shadows appear in the highly textured background. We can see that the gradient filter cannot well distinguish the shadow in the textured background from the moving object [Fig. 8(b) and (e)] However, our algorithm obtains much better results [Fig. 8(c) and (f)] More comparisons are given in

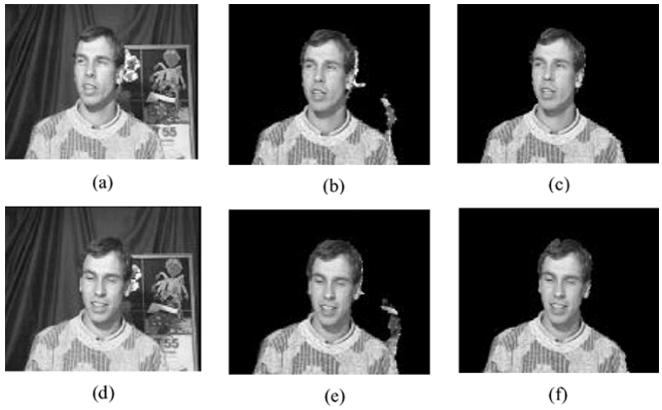

Fig. 8. Comparison of the gradient filter and our approach for Erik frames 18 and 38. (a) Erik frame 18. (b) Result by GF. (c) Result by ours. (d) Erik frame 38. (e) Result by GF. (f) Result by ours.

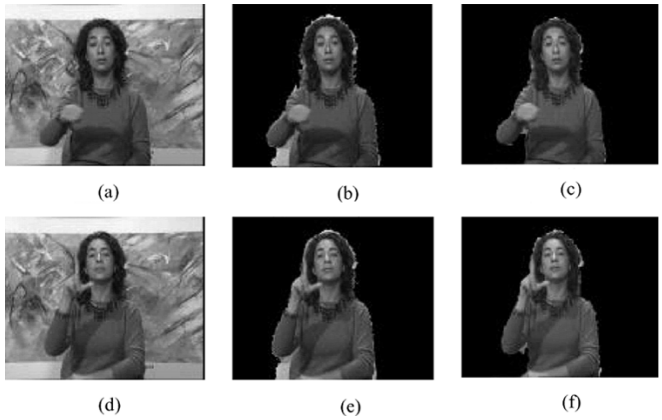

Fig. 9. Comparison of the gradient filter and our approach for Silent frames 112 and 268. (a) Silent frame 112. (b) Result by GF. (c) Result by ours. (d) Silent frame 268. (e) Result by GF. (f) Result by ours.

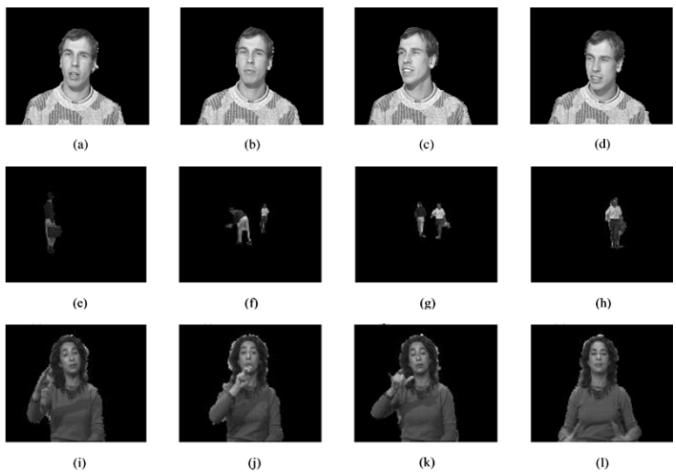

Fig. 10. Results obtained by our algorithm for some frames of the three sequences. (a) Erik frame 8. (b) Erik frame 16. (c) Erik frame 24. (d) Erik frame 42. (e) Hall frame 35. (f) Hall frame 105. (g) Hall frame 199. (h) Hall frame 257. (i) Silent frame 34. (j) Silent frame 96. (k) Silent frame 153. (l) Silent frame 217.

Fig. 9 for two Silent frames. Although the gradient filter can remove the shadows on the left side of the lady, it fails to detect the shadows on her right side [Fig. 9(b) and (e)] but our approach

can remove all the shadows [Fig. 9(c) and (f)] Again, these results show that our approach outperforms the gradient filter.

More segmentation results obtained by our algorithm are given in Fig. 10 for the three sequences. The frames are picked randomly. These results also demonstrate the effectiveness of our approach to the removal of moving cast shadows.

In our experiments, we found that, similarly to other shadow removal techniques, it is possible for our algorithm to misclassify part of a moving object as the shadow or background region, and vice versa, when the moving object (or part of it) moves very slowly and there is little change at the boundary between the object and the background. To overcome this problem, more complicated techniques and further study are required.

IV. Conclusion

We have presented a new approach to the detection of moving cast shadows in normal indoor environment. Compared with outdoor shadows at daytime, these shadows are generally insignificant and more difficult to detect. In our approach, a number of techniques have been developed to achieve the goal. They include the generations of initial change detection masks and Canny edge maps, shadow region detection by multiframe integration, edge matching and region growing, and post-processing.

We have tested our algorithm with the most commonly used MPEG-4 testing sequences and compared it with a recently published and most related method called the gradient filter [4]. Both objective and subjective evaluations are carried out in the experiments. The results show that our approach performs better. It can produce satisfactory segmentation when shadows appear in both smooth and highly textured backgrounds. Our algorithm is also efficient; it is not difficult to make it applicable in real time (30 f/s) after some algorithmic optimization. Further work includes doing more experiments to see whether there are better choices of parameters and combining our approach with others to achieve better performance.

References

[1] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, Jun. 1994.

[2] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 11, pp. 679–698, Nov. 1986.

[3] C. J. Chang, W. F. Hu, J. W. Hsieh, and Y. S. Chen, "Shadow elimination for effective moving object detection with Gaussian models," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2002, pp. 540–543.

[4] S. Y. Chien, S. Y. Ma, and L. G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 577–586, Jul. 2002.

[5] J. G. Choi, S. W. Lee, and S. D. Kim, "Spatio-temporal video segmentation using a joint similarity measure," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 279–286, Apr. 1997.

[6] R. Cucchiara, C. Granan, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[7] G. S. K. Fung, N. H. C. Yung, G. K. H. Pang, and A. H. S. Lai, "Toward detection of moving cast shadows for visual traffic surveillance," in *Proc. IEEE Int. Conf. System Man and Cybernetics*, 2001, pp. 2505–2510.

[8] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Reading, MA: Addision-Wesley, 1992, pp. 28–48.

[9] C. Jiang and M. O. Ward, "Shadow identification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1992, pp. 606–612.

[10] M. Kilgar, "A shadow handler in a video-based real-time traffic monitoring system," in *Proc. IEEE Workshop Applications of Computer Vision*, 1992, pp. 11–18.

[11] C. Kim and J. N. Hwang, "Fast and automatic video object segmentation and tracking for content-based application," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 2, pp. 122–129, Feb. 2002.

[12] R. Mech and M. Wollborn, "A noise robust method for 2-D shape estimation of moving objects in video sequences considering a moving camera," *Signal Process.*, vol. 66, pp. 203–217, Apr. 1998.

[13] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1190–1203, Dec. 1999.

[14] I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, "Moving shadow and object detection in traffic scenes," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2000, pp. 321–324.

[15] F. Moscheni, S. Bhattacharjee, and M. Junt, "Spatiotemporal segmentation based on region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 897–915, Sep. 1998.

[16] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Process.*, vol. 66, pp. 219–232, Apr. 1998.

[17] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara, "Detecting moving shadows, algorithms, and evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, pp. 918–923, July 2003.

[18] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[19] Y. Sonoda and T. Ogata, "Separation of moving objects and their shadows, and application to tracking of loci in the monitoring images," in *Proc. IEEE Int. Conf. Signal Processing*, 1998, pp. 1261–1264.

[20] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Trans. Multimedia*, vol. 1, pp. 65–77, Mar. 1999.

[21] X. Tao, M. Guo, and B. Zhang, "A neural network approach to the elimination of road shadow for outdoor mobile robot," in *Proc. IEEE Int. Conf. Intelligent Processing Systems*, Oct. 1997, pp. 28–31.

[22] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: A region labeling approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 597–612, Jul. 2002.