Title
Detecting moving shadows: Algorithms and evaluation

Journal
IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7)

ISSN
0162-8828

Authors
Prati, A
Mikic, I
Trivedi, Mohan Manubhai
et al.

Publication Date
2003-07-01

DOI
10.1109/TPAMI.2003.1206520

Peer reviewed

Detecting Moving Shadows: *Algorithms and Evaluation*

Andrea Prati, *Member*, *IEEE*,
Ivana Mikic, *Member*, *IEEE*,
Mohan M. Trivedi, *Member*, *IEEE*, and
Rita Cucchiara, *Member*, *IEEE*

Abstract—Moving shadows need careful consideration in the development of robust dynamic scene analysis systems. Moving shadow detection is critical for accurate object detection in video streams since shadow points are often misclassified as object points, causing errors in segmentation and tracking. Many algorithms have been proposed in the literature that deal with shadows. However, a comparative evaluation of the existing approaches is still lacking. In this paper, we present a comprehensive survey of moving shadow detection approaches. We organize contributions reported in the literature in four classes two of them are statistical and two are deterministic. We also present a comparative empirical evaluation of representative algorithms selected from these four classes. Novel quantitative (detection and discrimination rate) and qualitative metrics (scene and object independence, flexibility to shadow situations, and robustness to noise) are proposed to evaluate these classes of algorithms on a benchmark suite of indoor and outdoor video sequences. These video sequences and associated "ground-truth" data are made available at http://cvrr.ucsd.edu/aton/shadow to allow for others in the community to experiment with new algorithms and metrics.

Index Terms—Shadow detection, performance evaluation, object detection, segmentation, traffic scene analysis, visual surveillance.

1 Introduction

Detection and tracking of moving objects is at the core of many applications dealing with image sequences. One of the main challenges in these applications is identifying shadows which objects cast and which move along with them in the scene. Shadows cause serious problems while segmenting and extracting moving objects due to the misclassification of shadow points as foreground. Shadows can cause object merging, object shape distortion, and even object losses (due to the shadow cast over another object). The difficulties associated with shadow detection arise since shadows and objects share two important visual features. First, shadow points are detectable as foreground points since they typically differ significantly from the background. Second, shadows have the same motion as the objects casting them. For this reason, the shadow identification is critical both for still images and for image sequences (video) and has become an active research area, especially in the recent past. It should be noted that, while the main concepts utilized for shadow analysis in still and video images are similar, typically, the purpose behind shadow extraction is somewhat different. In the case of still images, shadows are often analyzed and exploited to infer geometric properties of the objects causing the shadow ("shape from shadow" approaches) as well as to enhance object localization and measurements. Examples of this can be found in aerial image analysis for recognizing buildings [1], [2], for obtaining 3D reconstruction of the scene [3], or even for detecting clouds and their shadows [4]. Another important application domain for shadow detection in still images is for the 3D analysis of objects to extract surface orientations [5] and light source direction [6].

Shadow analysis, considered in the context of video data, is typically performed for enhancing the quality of segmentation results instead of deducing some imaging or object parameters. In the literature, shadow detection algorithms are normally associated with techniques for moving object segmentation. In this paper, we present a comprehensive survey of moving shadow detection approaches. We organize contributions reported in the literature in four classes and present a comparative empirical evaluation of representative algorithms selected from these four classes. This comparison takes into account both the advantages and the drawbacks of each proposal and provides a quantitative and qualitative evaluation of them. Novel quantitative (detection and discrimination rate) and qualitative metrics (scene and object independence, flexibility to shadow situations, and robustness to noise) are proposed to evaluate these classes of algorithms on a benchmark suite of indoor and outdoor video sequences. These video sequences and associated "ground-truth" data are made available at http://cvrr.ucsd.edu/aton/shadow to allow for others in the community to experiment with new algorithms and metrics. This availability follows the idea of data-sharing embodied in Call for Comparison, like the project of European COST 211 Group (see http://www.iva.cs.tut.fi/COST211/ for further details).

In the next section, we develop a two layer taxonomy for surveying various algorithms presented in the literature. Each approach class is detailed and discussed to emphasize its strengths and its limitations. In Section 3, we develop a set of evaluation metrics to compare the shadow detection algorithms. This is followed by Section 4, where we present a results of empirical evaluation of four selected algorithms on a set of five video sequences. The final section presents concluding remarks.

2 Taxonomy of Shadow Detection Algorithms

Most of the proposed approaches take into account the shadow model described in [7]. To account for their differences, we have organized the various algorithms in a two-layer taxonomy. The first layer classification considers whether the decision process introduces and exploits uncertainty. *Deterministic approaches* use an on/off decision process, whereas *statistical approaches* use probabilistic functions to describe the class membership. Introducing uncertainty to the class membership assignment can reduce noise sensitivity. In the statistical methods (see [8], [9], [10], [11], [12]), the parameter selection is a critical issue. Thus, we further divide the statistical approaches in *parametric* and *nonparametric* methods. The study reported in [8] is an example of the parametric approach, whereas [10], [11] are examples of the nonparametric approach. The deterministic class (see [6], [7], [13], [14]) can be further subdivided. Subclassification can be based on whether the on/off decision can be supported by model-based knowledge or not. Choosing a *model-based* approach undoubtedly achieves the best results, but is, most of the time, too complex and time consuming compared to the *nonmodel-based*. Moreover, the number and the complexity of the models increase rapidly if the aim is to deal with complex and cluttered environments with different lighting conditions, object classes, and perspective views.

It is also important to recognize the types of "features" utilized for shadow detection. Basically, these features are extracted from three domains: *spectral*, *spatial*, and *temporal*. Approaches can exploit differently spectral features, i.e., using gray level or color

- *A. Prati and R. Cucchiara are with the Dipartimento di Ingegneria dell'Informazione, Università di Modena e Reggio Emilia, Via Vignolese, 905/b, Modena, Italy. E-mail: {prati.andrea, cucchiara.rita}@unimore.it.*
- *I. Mikic is with Q3DM Inc., 10110 Sorrento Valley Road, Suite B, San Diego, CA 92121. E-mail: imikic@q3dm.com.*
- *M.M. Trivedi is with the Computer Vision and Robotics Research Laboratory, Department of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037. E-mail: trivedi@ece.ucsd.edu.*

Manuscript received 7 June 2001; revised 19 Aug. 2002; accepted 18 Jan. 2003.
Recommended for acceptance by P. Anandan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114319.

TABLE 1
Classification of the Literature on Shadow Detection

Statistical parametric				Statistical nonparametric			
Paper	Spectral	Spatial	Temporal	Paper	Spectral	Spatial	Temporal
Friedman and Russell 1997 [12]	C	L	D	**Horprasert et al. 1999 [11]**	C	L	S
Mikić et al. 2000 [8][9]	C	R	D	Tao et al.[4] 2000 [16]	C	F	D
				McKenna et al. 2000 [17]	C	L	S
Deterministic model-based				**Deterministic nonmodel-based**			
Paper	Spectral	Spatial	Temporal	Paper	Spectral	Spatial	Temporal
Irvin and McKeown Jr.[1] 1989 [1]	G	L	S	Scanlan et al.[1] 1990 [18]	G	L	S
Wang et al. 1991 [4]	G	R	S	Jiang and Ward[1] 1992 [6]	G	F	S
Kilger 1992 [19]	G	R	S	Charkari and Mori 1993 [20]	G	R	S
Koller et al. 1993 [14]	G	L	S	Sexton and Zhang 1993 [21]	G	L	S
Onoguchi[2] 1998 [22]	G	L	S	Funka-Lea and Bajcsy[1] 1995 [23]	G	F	D
				Sonoda and Ogata 1998 [24]	G	F	S
				Tzomakas and von Seelen 1998 [25]	G	F	S
				Amamoto and Fujii 1999 [26]	G	N/A[3]	D
				Stauder et al. 1999 [7]	G	F	D
				Cucchiara et al. 2001 [13]	C	L	S

(G = Gray-Level, C = Color, L = Local/Pixel-Level, R = Region-Level, F = Frame-Level, S = Static, and D = Dynamic.)

[1] *This paper considers only still images.*

[2] *This paper is not properly a deterministic model approach. It uses an innovative approach based on inverse perspective mapping in which the assumption is that the shadow and the object that casts it are overlapped if projected on the ground plane. Since a model of the scene is necessary, we classify this paper in this class.*

[3] *This paper has the unique characteristic of using the DCT to remove shadow. For this reason, we can say that this paper works on frequency-level. The rationale used by the authors is that a shadow has, in the frequency domain, a large DC component, whereas the moving object has a large AC component.*

[4] *Since this paper uses a fuzzy neural network to classify points as belonging or not to a shadow, it can be considered a statistical approach. However, how much the parameter setting is automated is not clear in this paper.*

information. Some approaches improve results by using spatial information working at a region level or at a frame level instead of pixel level. This is a classification similar to that used in [15] for the background maintenance algorithms. Finally, some methods exploit temporal redundancy to integrate and improve results.

In Table 1, we have classified 21 papers dealing with shadow detection in four classes. We highlight spectral, spatial, and temporal features used by these algorithms. In this paper, we focus our attention on four algorithms (reported in bold in Table 1) representative of three of the above-mentioned classes. For the statistical parametric class, we choose the algorithm proposed in [8] since this utilizes features from all three domains. The approach reported in [11] can be considered to be a very good representative of the statistical nonparametric class and is also cited and used in [17]. Within the deterministic nonmodel-based class, we choose to compare the algorithm described in [13] because it is the only one that uses HSV color space for shadow detection. Finally, the algorithm reported in [7] has been selected for its unique capability of coping with penumbra. The deterministic model-based class has not been considered due to its complexity and due to its reliance on very specific task domain assumptions. For instance, the approach used in [14] models shadows using a simple illumination model: Assuming parallel incoming light, they compute the projection of the 3D object model onto the ground, exploiting two parameters for the illumination direction set offline and assumed to be constant during the entire sequence. However, in an outdoor scene, the projection of the shadow is unlikely to be perspective since the light source cannot be assumed to be a point light source. Therefore, the need for object models and the illumination position's manual setting make this approach difficult to implement in a general-purpose framework.

In the next sections, we describe briefly the selected approaches. For more details, refer to the corresponding papers or see the detailed description that we reported in [27].

2.1 Statistical Nonparametric (SNP) Approach

As an example of the statistical nonparametric (SNP) approach, we choose the one described in [28], and detailed in [11]. This work considers the *color constancy* ability of human eyes and exploits the Lambertian hypothesis to consider color as a product of irradiance and reflectance. The distortion of the brightness α_i and the distortion of the chrominance CD_i of the difference between the expected color of a pixel and its value in the current image are computed and normalized with regard to their root mean square of pixel i. The values $\widehat{\alpha}_i$ and $\widehat{CD_i}$ obtained are used to classify a pixel in four categories:

$$C(i) = \begin{cases} Foreground: & \widehat{CD_i} > \tau_{CD} \quad or \quad \widehat{\alpha}_i < \tau_{\alpha lo}, \quad else \\ Background: & \widehat{\alpha}_i < \tau_{\alpha 1} \quad and \quad \widehat{\alpha}_i > \tau_{\alpha 2}, \quad else \\ Shadowed\ backg.: & \widehat{\alpha}_i < 0, \quad else \\ Highlighted\ backg.: & otherwise \end{cases} \tag{1}$$

The rationale used is that shadows have similar chromaticity, but lower brightness than the background model. A statistical learning procedure is used to automatically determine the appropriate thresholds.

2.2 Statistical Parametric (SP) Approach

The algorithm described in [8] for traffic scene shadow detection is an example of statistical parametric (SP) approach. This algorithm claims to use two sources of information: *local* (based on the appearance of the pixel) and *spatial* (based on the assumption that the objects and the shadows are compact regions). The a posteriori probabilities of belonging to background, foreground, and shadow classes are maximized. The a priori probabilities of a pixel belonging to shadow are computed by assuming that $\mathbf{v} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]^{\mathbf{T}}$ is the value of the pixel not shadowed and by using an approximated linear transformation $\bar{\mathbf{v}} = \mathbf{D}\mathbf{v}$ (where $\mathbf{D} = diag(\mathbf{d_R}, \mathbf{d_G}, \mathbf{d_B})$ is a diagonal matrix obtained by experimental evaluation) to estimate the color of the point covered by a shadow. The $\mathbf{D}$ matrix is assumed approximately constant over flat surfaces. If the background is not flat over the entire image, different $\mathbf{D}$ matrices must be computed for each flat subregion. The spatial information is exploited by performing an iterative probabilistic relaxation to propagate neighborhood information. In this statistical *parametric* approach, the main drawback is the difficult process necessary to select the parameters. Manual segmentation of a certain number of frames has to be done to collect statistics and to compute the values of matrix $\mathbf{D}$. An expectation maximization (EM) approach could be used to automate this process, as in [12].

2.3 Deterministic Nonmodel-Based (DNM1) Approach

The system described in [13] is an example of the deterministic nonmodel-based approach (and we call it DNM1). This algorithm works in the HSV color space. The main reasons are that the HSV color space corresponds closely to the human perception of color [29] and it has revealed more accuracy in distinguishing shadows. In fact, a shadow cast on a background does not change its hue significantly [30]. Moreover, the authors exploit saturation information since they note that shadows often lower the saturation of the points. The resulting decision process is reported in the following equation:

$$SP_k(x,y) = \begin{cases} 1 & if \quad \alpha \leq \frac{I_k^V(x,y)}{B_k^V(x,y)} \leq \beta \ \wedge \ (I_k^S(x,y) - B_k^S(x,y)) \leq \\ & \tau_S \ \wedge \ |I_k^H(x,y) - B_k^H(x,y)| \leq \tau_H \\ 0 & otherwise, \end{cases} \tag{2}$$

where $I_k(x,y)$ and $B_k(x,y)$ are the pixel values at coordinate (x,y) in the input image (frame k) and in the background model (computed at frame k), respectively. The use of β prevents the identification as shadows those points where the background was slightly changed by noise, whereas α takes into account the "power" of the shadow, i.e., how strong the light source is with regard to the reflectance and irradiance of the objects. Thus, the stronger and higher the sun (in the outdoor scenes), the lower α should be chosen.

2.4 Deterministic Nonmodel-Based (DNM2) Approach

Finally, we compare the approach presented in [7]. This is also a deterministic nonmodel-based approach, but we have included it because of its completeness (it is the only work in the literature that deals with penumbra in moving cast shadows). The shadow detection is provided by verifying three criteria: the presence of a "darker" uniform region, by assuming that the ratio between the actual value and reference value of a pixel is locally constant in presence of cast shadows, the presence of a high difference in luminance with regard to reference frame, and the presence of static and moving edges. Static edges hint at a static background and can be exploited to detect nonmoving regions inside the frame difference. Moreover, to detect penumbra, the authors propose computing the width of each edge in the difference image. Since penumbra cause a soft luminance step at the contour of a shadow, they claim that the edge width is the more reliable way to distinguish between objects contours and shadows contours (characterized by a width greater than a threshold).

This approach is one of the most complete and robust proposed in the literature. Nevertheless, in this case, the assumptions and the corresponding approximations introduced are strong and they could lack in generality. Also, the penumbra criterion is not explicitly exploited to *add* penumbra points as shadow points, but it is only used to *remove* the points that do not fit this criterion. Moreover, the proposed algorithm uses the previous frame (instead of the background) as a reference frame. This choice exhibits some limitations in moving region detection since it is influenced by object speed and it is too noise sensitive. Thus, to make the comparison of these approaches as fair as possible, limited to the shadow detection part of the system, we implemented the DNM2 approach using a background image as a reference, as the other three approaches do.

3 Performance Evaluation Metrics

In this section, the methodology used to compare the four approaches is presented. In order to systematically evaluate various shadow detectors, it is useful to identify the following two important quality measures: *good detection* (low probability of misclassifying a shadow point) and *good discrimination* (the probability of classifying nonshadow points as shadow should be low, i.e., low false alarm rate). The first one corresponds to minimizing the *false negatives (FN)*, i.e., the shadow points classified as background/foreground, while, for good discrimination, the *false positives (FP)*, i.e., the foreground/background points detected as shadows, are minimized.

A reliable and objective way to evaluate this type of visual-based detection is still lacking in the literature. A very good work on how to evaluate objectively the segmentation masks in video sequences is presented in [31]. The authors proposed a metric based on *spatial accuracy* and *temporal stability* that aims at evaluating information differently than the FPs and FNs, depending on their distance from the borders of the mask, and at taking into account the shifting (instability) of the mask along the time. In [22], the authors proposed two metrics for moving object detection evaluation: the *Detection Rate (DR)* and the *False Alarm Rate (FAR)*. Assuming TP as the number of *true positives* (i.e., the shadow points correctly identified), these two metrics are defined as follows:

$$DR = \frac{TP}{TP + FN} \quad ; \quad FAR = \frac{FP}{TP + FP}. \tag{3}$$

The Detection Rate is often called *true positive rate* or also *recall* in the classification literature and the FAR corresponds to $1 - p$, where p is the so-called *precision* in the classification theory. These figures are not selective enough for shadow detection evaluation since they do not take into account whether a point detected as shadow belongs to a foreground object or to the background. If shadow detection is used to improve moving object detection, only the first case is problematic since false positives belonging to the background do not affect neither the object detection nor the object shape.

To account for this, we have modified the metrics of (3), defining the *shadow detection rate* η and the *shadow discrimination rate* ξ as follows:

$$\eta = \frac{TP_S}{TP_S + FN_S} \ ; \ \xi = \frac{\overline{TP_F}}{TP_F + FN_F}, \tag{4}$$

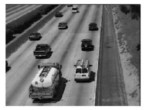
where the subscript S stands for shadow and F for foreground. The $\overline{TP_F}$ is the number of ground-truth points of the foreground objects minus the number of points detected as shadows, but belonging to foreground objects.

In addition to the above quantitative metrics, we also consider the following qualitative measures in our evaluation: *robustness to noise*, *flexibility to shadow strength, width and shape*, *object independence*, *scene independence*, *computational load*, and *detection of indirect cast shadows and penumbra*. Indirect cast shadows are the shadows cast by a moving object over another moving object and their effect is to decrease the intensity of the moving object covered, probably affecting the object detection, but not the shadow detection.

4 Empirical Comparative Evaluation

In this section, the experimental results and the quantitative and qualitative comparison of the four approaches are presented. First, a set of sequences to test the algorithms was chosen to form a complete and nontrivial benchmark suite. We select the sequences reported in Table 2, where both indoor and outdoor sequences are present, where shadows range from dark and small to light and large, and where the object type, size, and speed vary considerably.

TABLE 2
The Sequence Benchmark Used

	Highway I	**Highway II**	**Campus**	**Laboratory**	**Intelligent room**
					
Sequence type	outdoor	outdoor	outdoor	indoor	indoor
Sequence length	1074	1134	1179	987	900
Image size	320x240	320x240	352x288	320x240	320x240
Shadow strength	medium	high	low	very low	low
Shadow size	large	small	very large	medium	large
Object class	vehicles	vehicles	vehicle/people	people/other	people
Object size	large	small	medium	medium	medium
Object speed (pixels)	30-35	8-15	5-10	10-15	2-5
Noise level	medium	medium	high	low	medium

The *Highway I* and the *Highway II* sequences show a traffic environment (at two different lighting conditions), where the shadow suppression is very important to avoid misclassification and erroneous counting of vehicles on the road. The *Campus* sequence is a noisy sequence from an outdoor campus site where cars approach an entrance barrier and students are walking around. The two indoor sequences report two laboratory rooms in two different perspectives and lighting conditions. In the *Laboratory* sequence, besides walking people, a chair is moved in order to detect its shadow.

4.1 Quantitative Comparison

To compute the evaluation metrics described in Section 3, the ground truth for each frame is necessary. We obtained it by segmenting the images with an accurate manual classification of points in foreground, background, and shadow regions. We prepared ground truth on tens of frames for each video sequence representative of different situations (dark/light objects, multiple objects or single object, occlusions or not).

All four approaches, but the DNM2, have been faithfully and completely implemented. In the case of DNM2, some simplifications have been introduced: The memory MEM used in [7] to avoid infinite error propagation in the change detection masks (CDMs) has not been implemented since it is computationally very heavy and not necessary (in the sequences considered there is no error propagation); some minor tricks (like that of the closure of small edge fragments) have not been included due to the lack of details in the paper. However, these missing parts of the algorithm do not influence shadow detection at all. In conclusion, the comparison has been set up as fairly as possible.

Results are reported in Table 3. To establish a fair comparison, algorithms do not implement any background updating process (since each tested algorithm proposes a different approach). Instead, we compute the reference image and other parameters from the first N frames (with N varying with the sequence considered). The first N frames can be considered as the training set and the remaining frames as the testing set for our experimental framework. Note that the calculated parameters remain constant for the whole sequence. The visual results on a subset of the *Intelligent Room* and of the *Highway I* sequences are available at http://cvrr.ucsd.edu/aton/shadow. Fig. 1 shows an example of visual results from the indoor sequence *Intelligent Room*.

The SNP algorithm is very effective in most of the cases, but with very variable performances. It achieves the best detection performance η and high discrimination rate ξ in the indoor sequence *Laboratory*, with percentages up to 92 percent. However, the discrimination rate is quite low in the *Highway I* and *Campus* sequences. This can be explained by the dark (similar to shadows) appearance of objects in the *Highway I* sequence and by the strong noise component in the *Campus* sequence.

The SP approach achieves good discrimination rate in most of the cases. Nevertheless, its detection rate is relatively poor in all the cases, but the *Intelligent room* sequence. This is mainly due to the approximation of constant **D** matrix on the entire image. Since the background can be rarely assumed as flat on the entire image, this approach lacks in generality. Nevertheless, good accuracy in the case of the *Intelligent room* test shows how this approach can deal with indoor sequences where the assumption of constant **D** matrix is valid.

The DNM1 algorithm is the one with the most stable performance, even with totally different video sequences. It achieves good accuracy in almost all the sequences. It outperforms the other algorithms in the *Campus* and in the *Intelligent room* sequences.

The DNM2 algorithm suffers mainly due to the assumption of planar background. This assumption fails in the case of the

TABLE 3
Experimental Results

	Highway I		**Highway II**		**Campus**		**Laboratory**		**Intelligent Room**	
	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
SNP	81.59%	63.76%	51.20%	78.92%	80.58%	69.37%	84.03%	92.35%	72.82%	88.90%
SP	59.59%	84.70%	46.93%	91.49%	72.43%	74.08%	64.85%	95.39%	76.27%	90.74%
DNM1	69.72%	76.93%	54.07%	78.93%	82.87%	86.65%	76.26%	89.87%	78.61%	90.29%
DNM2	75.49%	62.38%	60.24%	72.50%	69.10%	62.96%	60.34%	81.57%	62.00%	93.89%

(a) (b) (c) (d) (e)

Fig. 1. Results of in the *Intelligent room* sequence. Gray pixels identify foreground points and dark pixels indicate shadow points. (a) Raw image, (b) SNP result, (c) SP result, (d) DNM1 result, and (e) DNM2 result.

Laboratory sequence where the shadows are cast both on the floor and on the cabinet. The low detection performance in the *Campus* sequence is mainly due to noise and this algorithm has proven low robustness to strong noise. Finally, this algorithm achieves the worst discrimination result in all the cases but the *Intelligent room* sequence. This is due to its assumption of textured objects: If the object appearance is not textured (or seems not textured due to the distance and the quality of the acquisition system), the probability that parts of the object are classified as shadow rises. In fact, in the *Intelligent room* sequence, the clothes of the person in the scene are textured and the discrimination rate is higher. This approach outperforms the others in the more difficult sequence (*Highway II*).

The statistical approaches perform robustly in noisy data due to statistical modeling of noise. On the other hand, deterministic approaches (in particular, if pixel-based and almost unconstrained as DNM1) exhibit a good flexibility to different situations. Difficult sequences like *Highway II*, require, however, a more specialized and complete approach to achieve good accuracy. To help evaluating the approaches, the results on the *Highway I* outdoor sequence and on the *Intelligent room* indoor sequence are available at http://cvrr.ucsd.edu/aton/shadow.

4.2 Qualitative Comparison

To evaluate the behavior of the four algorithms with respect to the qualitative issues presented in Section 3, we vote them ranging from "very low" to "very high" (see Table 4). The DNM1 method is the most robust to noise, thanks to its pre and postprocessing algorithms [13]. The capacity to deal with different shadow size and strength is high in both the SNP and the DNM1. However, the higher flexibility is achieved by the DNM2 algorithm, which is able to detect even the penumbra in an effective way. Nevertheless, this algorithm is very object-dependent in the sense that, as already stated, the assumption on textured objects strongly affects the results. Also, the two frame difference approach proposed in [7] is weak as soon as the object speeds increase. The hypothesis of a planar background makes the DNM2 and, especially, the SP approaches more scene-dependent than the other two. Although we cannot claim to have implemented these algorithms in the most efficient way, the DNM2 seems the more time consuming due to the amount of processing necessary. On the other hand, the SNP is very fast.

Finally, we evaluated the behavior of the algorithms in the presence of indirect cast shadows (see Section 3). The DNM2 approach is able to detect both the penumbra and the indirect cast shadow in a very effective way. The SP and the DNM1 methods failed in detecting indirect cast shadows. The pixel-based decision cannot distinguish correctly between this type of moving shadows and those shadows cast on the background. However, the SP approach is able to detect relatively narrow penumbra.

5 Concluding Remarks

Development of practical dynamic scene analysis systems for real-world applications needs careful consideration of the moving shadows. The research community has recognized this and serious, substantive efforts in this area are being reported. The main motivator for this paper is to provide a general framework to discuss such contributions in the field and also to provide a systematic empirical evaluation of a selected representative class of shadow detection algorithms. Papers dealing with shadows are classified in a two-layer taxonomy and four representative algorithms are described in detail. A set of novel quantitative and qualitative metrics has been adopted to evaluate the approaches.

The main conclusion of the empirical study can be described as follows: For a general-purpose shadow detection system with minimal assumptions, a pixel-based deterministic nonmodel-based approach (DNM1) assures best results. On the other hand, to detect shadows efficiently *in one specific environment*, more assumptions yield better results and the *deterministic model-based approach* should be applied. In this situation, if *the object classes are numerous* to allow modeling of every class, a *complete deterministic approach*, like the DNM2, should be selected. If *the environment is indoor*, the *statistical approaches* are the more reliable since the scene is constant and a statistical description is very effective. If there are different planes onto which the shadows can be cast, an approach like SNP is the best choice. *If the shadows are scattered, narrow, or particularly "blended" to the environment*, a region-based dynamic approach, typically deterministic, is the best choice (as DNM2 in the *Highway II* scene reported in this paper). Finally, if *the scene is*

TABLE 4
Qualitative Evaluation

	Robustness to noise	Flexibility to shadow	Object independence	Scene independence	Computational load	Indirect shadow & penumbra detection
SNP	high	high	high	high	very low	high
SP	high	medium	high	low	low	low
DNM1	very high	high	high	high	low	very low
DNM2	low	very high	low	medium	high	very high

noisy, a *statistical approach or a deterministic approach with effective pre and postprocessing steps* should be used. Finally, we want to remark that all the evaluated approaches exploit a large set of assumptions to limit complexity, and to avoid being unduly constrained to a specific scene model. This limits their shadow detection accuracies. This, in fact, points to the limitations of using only image-derived information in shadow detection. Further improvements would require feedback of specific task/scene domain knowledge.

A very interesting future direction has been suggested by an unknown reviewer. He/she suggested considering the *physically important* independent variables to evaluate the algorithms. If we can consider as parameters of the scene, for example, the type of illumination for indoor scene or the surface type upon which the shadows are cast in outdoor environments, we can build up a benchmark on which to test the different approaches. Results on accuracy on this benchmark would be more useful to future reserachers/developers of shadow detection (and motion detection) algorithms since they are more *physically linked* to the considered scene.

ACKNOWLEDGMENTS

This research was supported in part by the California Digital Media Innovation Program (DiMI) in partnership with the California Department of Transportation (Caltrans), Sony Electronics, and Compaq Computers. The authors wish to thank the collaborators from the Caltrans (TCFI) in Santa Barbara for their support and interactions. They extend special thanks to their colleagues in the CVRR Laboratory for their efforts in acquiring the video data sets utilized in their studies. They also wish to thank all the reviewers for their thoughtful comments that help to improve the paper. Ivana Mikic was with the Computer Vision and Robotics Department of Electrical and Computer Engineering, University of California, San Diego.

REFERENCES

[1] R.B. Irvin and D.M. McKeown Jr., "Methods for Exploiting the Relationship between Buildings and Their Shadows in Aerial Imagery," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 19, pp. 1564-1575, 1989.

[2] Y. Liow and T. Pavlidis, "Use of Shadows for Extracting Buildings in Aerial Images," *Computer Vision, Graphics, and Image Processing,* vol. 49, no. 2, pp. 242-277, Feb. 1990.

[3] G. Medioni, "Obtaining 3-D from Shadows in Aerial Images," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 73-76, 1983.

[4] C. Wang, L. Huang, and A. Rosenfeld, "Detecting Clouds and Cloud Shadows on Aerial Photographs," *Pattern Recognition Letters,* vol. 12, no. 1, pp. 55-64, Jan. 1991.

[5] S.A. Shafer and T. Kanade, "Using Shadows in Finding Surface Orientations," *Computer Vision, Graphics, and Image Processing,* vol. 22, no. 1, pp. 145-176, Apr. 1983.

[6] C. Jiang and M.O. Ward, "Shadow Identification," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 606-612, 1992.

[7] J. Stauder, R. Mech, and J. Ostermann, "Detection of Moving Cast Shadows for Object Segmentation," *IEEE Trans. Multimedia,* vol. 1, no. 1, pp. 65-76, Mar. 1999.

[8] I. Mikic, P. Cosman, G. Kogut, and M.M. Trivedi, "Moving Shadow and Object Detection in Traffic Scenes," *Proc. Int'l Conf. Pattern Recognition,* vol. 1, pp. 321-324, Sept. 2000.

[9] M.M. Trivedi, I. Mikic, and G. Kogut, "Distributed Video Networks for Incident Detection and Management," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 155-160, Oct. 2000.

[10] A. Elgammal, D. Harwood, and L.S. Davis, "Non-Parametric Model for Background Subtraction," *Proc. IEEE Int'l Conf. Computer Vision '99 FRAME-RATE Workshop,* 1999.

[11] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection," *Proc. IEEE Int'l Conf. Computer Vision '99 FRAME-RATE Workshop,* 1999.

[12] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," *Proc. 13th Conf. Uncertainty in Artificial Intelligence,* 1997.

[13] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati, "The Sakbot System for Moving Object Detection and Tracking," *Video-Based Surveillance Systems—Computer Vision and Distributed Processing,* pp. 145-157, 2001.

[14] D. Koller, K. Daniilidis, and H.H. Nagel, "Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes," *Int'l J. Computer Vision,* vol. 10, pp. 257-281, 1993.

[15] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and Practice of Background Maintenance," *Proc. IEEE Int'l Conf. Computer Vision,* vol. 1, pp. 255-261, 1999.

[16] X. Tao, M. Guo, and B. Zhang, "A Neural Network Approach to the Elimination of Road Shadow for Outdoor Mobile Robot," *Proc. IEEE Int'l Conf. Intelligent Processing Systems,* vol. 2, pp. 1302-1306, 1997.

[17] S.J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking Groups of People," *Computer Vision and Image Understanding,* vol. 80, no. 1, pp. 42-56, Oct. 2000.

[18] J.M. Scanlan, D.M. Chabries, and R.W. Christiansen, "A Shadow Detection and Removal Algorithm for 2-D Images," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing,* vol. 4, pp. 2057-2060, 1990.

[19] M. Kilger, "A Shadow Handler in a Video-Based Real-Time Traffic Monitoring System," *Proc. IEEE Workshop Applications of Computer Vision,* pp. 11-18, 1992.

[20] N.M. Charkari and H. Mori, "A New Approach for Real Time Moving Vehicle Detection," *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems,* pp. 273-278, 1993.

[21] G.G. Sexton and X. Zhang, "Suppression of Shadows for Improved Object Discrimination," *Proc. IEE Colloquium Image Processing for Transport Applications,* pp. 1-6, Dec. 1993.

[22] K. Onoguchi, "Shadow Elimination Method for Moving Object Detection," *Proc. Int'l Conf. Pattern Recognition,* vol. 1, pp. 583-587, 1998.

[23] G. Funka-Lea and R. Bajcsy, "Combining Color and Geometry for the Active, Visual Recognition of Shadows," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 203-209, 1995.

[24] Y. Sonoda and T. Ogata, "Separation of Moving Objects and Their Shadows, and Application to Tracking of Loci in the Monitoring Images," *Proc. Int'l Conf. Signal Processing,* pp. 1261-1264, 1998.

[25] C. Tzomakas and W. von Seelen, "Vehicle Detection in Traffic Scenes Using Shadows," Technical Report 98-06, IR-INI, Institut fur Nueroinformatik, Ruhr-Universitat Bochum, FRG, Germany, Aug. 1998.

[26] N. Amamoto and A. Fujii, "Detecting Obstructions and Tracking Moving Objects by Image Processing Technique," *Electronics and Comm. in Japan, Part 3,* vol. 82, no. 11, pp. 28-37, 1999.

[27] A. Prati, R. Cucchiara, I. Mikic, and M.M. Trivedi, "Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation," *Proc. Third Workshop Empirical Evaluation Methods in Computer Vision—IEEE Int'l Conf. Computer Vision and Pattern Recognition,* 2001.

[28] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-Time Surveillance of People and Their Activities," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 809-830, Aug. 2000.

[29] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, "A Color Segmentation Scheme for Object-Based Video Coding," *Proc. IEEE Symp. Advances in Digital Filtering and Signal Processing,* pp. 25-29, 1998.

[30] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 334-339, Aug. 2001.

[31] P. Villegas, X. Marichal, and A. Salcedo, "Objective Evaluation of Segmentation Masks in Video Sequences," *Proc. Workshop Image Analysis for Multimedia Interactive Services,* pp. 85-88, May 1999.

[32] G. Medioni, "Detecting and Tracking Moving Objects for Video Surveillance," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 319-325, 1999.