

Image difference threshold strategies and shadow detection

Paul L. Rosin
Inst. Remote Sensing Appl.
Joint Research Centre
I-21020 Ispra (VA), Italy
paul.rosin@jrc.it

Tim Ellis
Centre for Info. Eng.
City University
London, EC1V 0HB, UK
T.J.Ellis@city.ac.uk

Abstract

The paper considers two problems associated with the detection and classification of motion in image sequences obtained from a static camera. Motion is detected by differencing a reference and the “current” image frame, and therefore requires a suitable reference image and the selection of an appropriate detection threshold. Several threshold selection methods are investigated, and an algorithm based on hysteresis thresholding is shown to give acceptably good results over a number of test image sets. The second part of the paper examines the problem of detecting shadow regions within the image which are associated with the object motion. This is based on the notion of a shadow as a semi-transparent region in the image which retains a (reduced contrast) representation of the underlying surface pattern, texture or grey value. The method uses a region growing algorithm which uses a growing criterion based on a fixed attenuation of the photometric gain over the shadow region, in comparison to the reference image.

1 Introduction

Frame differencing is a particularly efficient and sensitive method for detecting grey level changes between images which are co-registered. It is widely used in motion detection, where a fixed camera is used to observe dynamic events in a scene.

The frame differencing algorithm may be sub-divided into three parts: firstly, the generation of a suitable reference or background; secondly, the arithmetic subtraction operation; and thirdly, the selection (and application) of a suitable threshold. Reference images can be generated by a variety of methods, e.g. on a background image acquired during a period of relative inactivity within the scene or from a temporally adjacent image from a dynamic sequence. In order to adapt to both global and local illumination changes (e.g. clouds, shadows), updating strategies can be applied to the reference image in order to keep it up-to-date.

Another problem in motion estimation occurs because of the detection of shadows, generated as the result of bright point-like illumination sources. These shadows may either be in contact with the detected object, or disconnected from it.

In the first case, the shadow distorts the object shape, making the use of subsequent shape recognition methods less reliable. In the second case, the shadow may be classified as a totally erroneous object in the scene. For analysing many natural world scenes (e.g. [5]) the disambiguation of these shadow regions would substantially benefit the object classification.

2 Change Detection

We assume a stationary camera; any movement (e.g. caused by wind shaking the camera) is corrected by first translating the images in the sequence (generally by a small amount) with respect to any image in the sequence or to some reference image such that their cross-correlation is minimised. Change detection can then be performed by simply taking image differences. The differencing can be performed between subsequent frames in the image sequence (e.g. [9]). This has the advantage that little spurious change should occur in the small time gap between frames. But the disadvantages are: 1/ that only the motion “wavefront” will produce any change, so that only part of the moving object is highlighted, and 2/ objects that become stationary for short periods of time will “disappear”. The alternative is to difference the image sequence against some reference image representing the background. If the background image is acquired some time previously (when it is known that no unwanted foreground objects were present) then there is a danger that changes in the ambient conditions (e.g. position of light source, light intensity), will cause the background image to become outdated. Therefore a potentially more robust approach is to dynamically generate the background image from some portion of the image sequence.

2.1 Background Generation

The task is as follows: the background image $B_{x,y}$ is to be generated from a sequence of images $I_{x,y}^t$ which may contain moving objects.

One approach takes an estimate of the background generated from the previous frames, and updates it using the current frame, which can be formulated as a Kalman filter [11]. However, various parameters are required which specify the degree of smoothing the previous estimates have on the current background prediction, and the model for background change (e.g. constant rate of change).

Alternatively, Long and Yang [13] analyse the temporal signature at each pixel for a stable section, i.e. a sequence of values which only changes by small amounts over time. The disadvantages are again the need for various parameters as well as the requirement of a continuously unoccluded view of the background.

Our approach is to perform background detection using L-filters, i.e. a linear combination of the ordered samples of the image sequence. This has several advantages: since the data sequence is (re)ordered it is not dependent on the background appearing unoccluded over a continuous sequence; L-filters are a class of robust statistics, and can tolerate large amounts of (e.g. non-Gaussian) noise, and; it does not require parameters.

Previously we have generated the background using a median filter at each pixel [16]: $B_{x,y} = \text{med}_t I_{x,y}^t$. Alternatively, Yang and Levine [21] have suggested

the least median of squares (LMedS) estimate: $B_{x,y} = \min_b \text{med}_t (I_{x,y}^t - b)^2$.

2.2 Automatic Thresholding of Difference Images

A popular approach to performing the automatic thresholding of difference images is to assume particular distribution models for the difference of image samples and the noise [1, 4, 9].

Instead our first method uses simple methods from robust statistics, and does not require any distribution assumptions. We analyse the difference image $D_{x,y} = |I_{x,y} - B_{x,y}|$ to determine the median $\text{MED} = \text{med}_{x,y \in I} D_{x,y}$ and the median absolute deviation $\text{MAD} = \text{med}_{x,y \in I} |D_{x,y} - \text{MED}|$. Assuming less than half the image is in motion the median should correspond to typical noise values, and a suitable threshold is at $T = \text{MED} + 3 \times 1.4826 \times \text{MAD}$, where 1.4826 is a normalisation factor wrt. a Gaussian distribution.

2.2.1 Connectivity preserving thresholding

In the context of document analysis O’Gorman [14] proposed a technique for image thresholding based on image connectivity. The image was thresholded at multiple intensities, and the connectivity value of each calculated. The threshold was selected from an intensity range that produced a stable set of connectivity values.

Rather than measuring connectivity, the number of regions may be more appropriate. However, the advantage of calculating connectivity over region counting is that the Euler number is locally countable [6], and can therefore be determined efficiently in a single raster scan of the image. We have experimented with calculating both the number of regions and the Euler number at all possible thresholds. The mode of the measures is calculated, and the threshold is selected as the lowest difference intensity that produces the mode value. We have found both the topology and connectivity methods give very similar results.

2.2.2 Thresholding with hysteresis

In his influential paper on edge detection Canny [2] popularised the application of connectivity-based hysteresis to thresholding. A bilevel edge magnitude threshold is applied, producing three classes of edges. All edges above the high threshold are retained (class H), and all edges below the low threshold are rejected (class L). The remaining edges (class M) are retained only if they are adjacent to class H edges or are connected to class H edges via other class M edges. The advantage of applying hysteresis is that it incorporates spatial context into the thresholding decision, and effectively enables isolated (noisy) medium strength edges to be eliminated without fragmenting long curves containing low strength sections.

We can apply the same technique of incorporating context to region thresholding as a method for eliminating small noisy regions without fragmenting larger regions. The difference image is thresholded at two levels, and regions in the intermediate range of intensities are rejected unless they are connected to regions generated by the lower threshold. Determining the connectivity is implemented by iteratively dilating the high threshold image, and performing a logical AND with the low threshold image. This has the advantage that it can be done relatively

efficiently. Also, if desired, the amount of expansion of the high threshold image can be controlled by limiting the number of dilation iterations.

Canny experimentally determined that a ratio of 2:1 between upper and lower threshold values produced good results. In [7] this was formulated as

$$R = \sqrt{\frac{\ln 2}{\ln \frac{1+2P}{1+P}}}$$

where P is the probability of an edge (and $1 - P$ is the probability of a non-edge). In this context Canny's ratio is obtained when $P = 0.23$ which may be a reasonable assumption for typical edge maps.

We can apply the same reasoning to determining the threshold ratio for applying hysteresis to the difference images. Our sequences tend to only have small areas of motion, normally in the range $P = [0.01, 0.05]$, which gives $R = [8.39, 3.86]$.

An alternative approach is to use a hybrid threshold selection scheme, where the upper and lower hysteresis thresholds are selected by different methods.

2.2.3 Local and Global Information

It should be noted that the hysteresis methodology attempts to combine local and global information: the two thresholds are calculated globally while the thresholding in the intermediate range uses local information. Local and global information have also been combined in different ways by other thresholding methods. Song *et al.* [19] use a single high threshold on the difference image and then grew the thresholded regions. This, however, assumes that both the moving objects and the background are homogeneous.

Yang and Levine [21] determine individual pixel thresholds by the following:

1. The background image $B_{x,y}$ is generated using the LMedS criterion as described in section 2.1.
2. A threshold image $T_{x,y}$ is generated from the median absolute deviation (MAD) at each pixel $T_{x,y} = B_{x,y} + 2.5 \times 1.4826 \times \text{MAD}_{x,y}$, where $\text{MAD}_{x,y} = \text{med}_t |I_{x,y}^t - B_{x,y}|$.
3. For the set of values in the difference image above their local threshold the global statistics (LMedS^g and MAD^g) are calculated. An additional threshold is applied to those previously retained pixels: pixels with difference values less than or equal to $\text{LMedS}^g + 2.5 \times 1.4826 \times \text{MAD}^g$ are removed.

In addition, local outliers are removed by non-maximal suppression, and erosion and dilation is performed. In our experiments these additional stages were not included – they were used by Yang and Levine [21] since they differenced edge maps, and wanted connected contours. The calculation of the MAD was modified according to Rousseeuw and Leroy [17] to take into account a finite sample correction factor which they determined as $1 + \frac{5}{n-p}$, where n is the number of data samples, and p is the data dimensionality. For our examples containing short image sequences, this factor is substantial (e.g. 1.7 for $n = 8$ and $p = 1$).

3 Computational Efficiency

Both the median and LMedS methods for background generation can be simply implemented based on sorting the F frames (each containing P pixels) in the sequence, and so their computational complexity is $O(PF \log F)$.

For determining the thresholds the three methods are:

- Calculating the global MED and MAD of the difference image can be calculated in $O(P)$ time using the histogram method [21].
- The Euler number only requires a single raster scan, and is applied at all G grey levels, and is therefore $O(GP)$.
- The per pixel MAD method suggested by Yang and Levine [21] requires $O(PF \log F)$ to generate the threshold image. Using the histogram method LMedS^g and MAD^g are calculated in $O(P)$ time.

We use a simple iterative raster-scanning method for performing the hysteresis. If I iterations are required then the complexity is $O(PI)$. However, if propagation is restricted to the blob boundaries then more efficient methods could be designed.

4 Examples of Thresholding

The alternative methods for the individual stages of processing (shown in fig. 1) produces a large number of possible combinations. Due to limitations of space we will describe results for only some of these combinations.

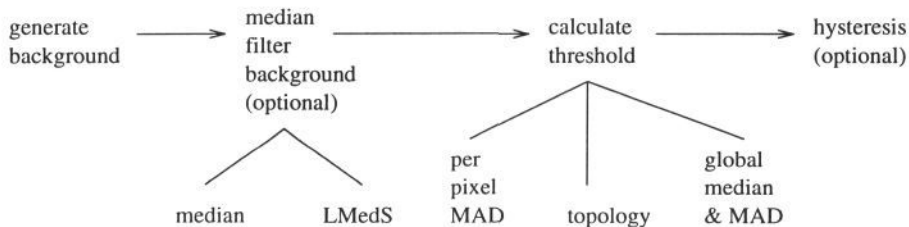


Figure 1: Processing steps

Figure 2a shows the first of eight frames from sequence SRDB018 in which a moving bird is located in the centre of the image. Note the low dynamic range, poor contrast between the bird and background, and the small size of the target. The following examples of thresholding show only the right half of the image. Detecting the background using the median method, and then thresholding based on the median and MAD (section 2.2) of the difference image gave very noisy results (fig. 2c). Median filtering the difference image first improved the results, but there are still many noisy blobs (fig. 2c). Using the LMedS method for background detection gave similar results as above. The local threshold approach of Yang and Levine [21] (without the non-maximal suppression and erosion/dilation stages) also gave noisy results (fig. 2d). The connectivity method applied directly to the

difference image failed to detect the moving object. Instead four tiny bright noise points were retained instead since they persisted over a large range of thresholds. However, when the difference image was median filtered, removing these points, a single blob was retained, corresponding to the bird (fig. 2e). It can be seen that a high threshold was necessary to eliminate all other blobs, resulting in the target blob being shrunk since its boundaries are blurred. Applying hysteresis thresholding ($R = 8$) produces a good result (fig. 2f). The bird is well thresholded whilst also avoiding spurious blobs. For comparison, some standard image thresholding techniques were also applied [12, 15, 20]. Without median filtering the difference image Otsu's method performed very poorly (fig. 2g), but with the addition of filtering it gave the best result of the three techniques (fig. 2h).

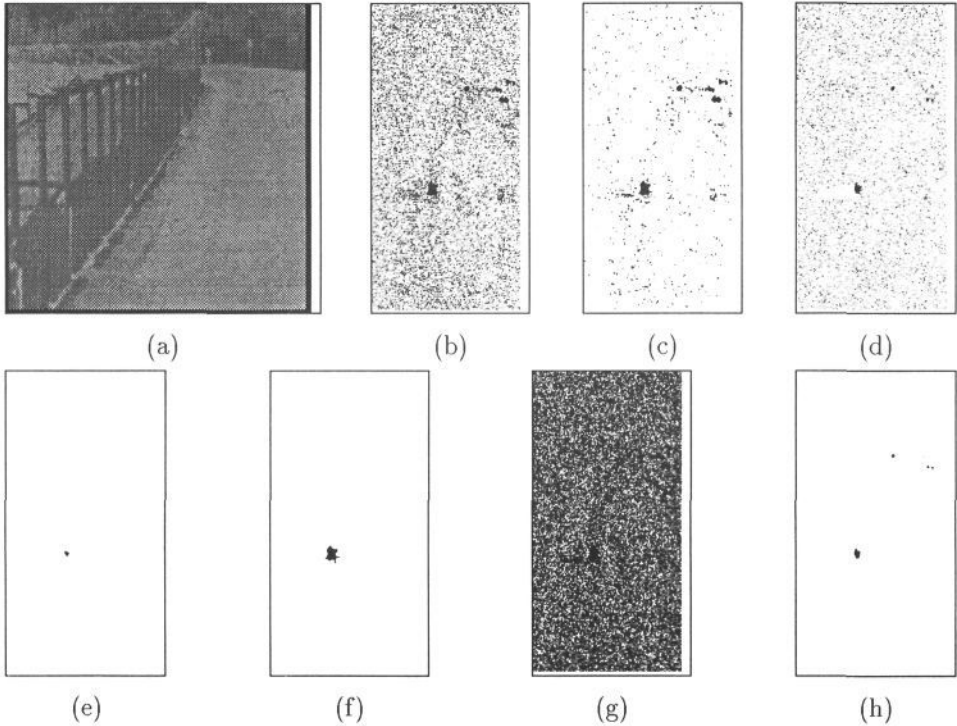


Figure 2: SRDB018 (a) Frame 1, (b) median, (c) median(difference) + median, (d) LMedS background + local thresholds, (e) median(difference) + Euler, (f) median(difference) + Euler + hysteresis, (g) Otsu, (h) median(difference) + Otsu

A second example is given in fig. 3a of the first of eight frames from sequence SRDB044 showing a man walking in the shadow at the rear of the scene. Again, the connectivity method with prior median filtering of the difference image and hysteresis performs well (fig. 3b) – a single blob is extracted corresponding to the man. The other methods give poor results (eg. the median method applied after median filtering of the difference image, fig. 3c). Otsu's method underthresholds, and the man is fragmented into four blobs (fig. 3d).

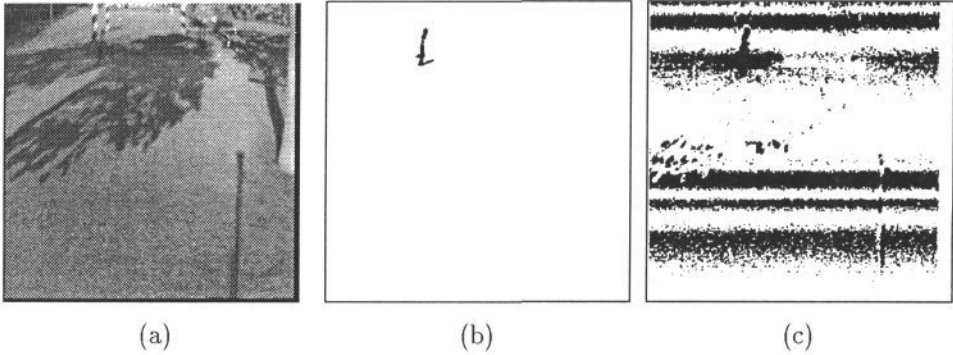


Figure 3: SRDB044 (a) Frame 1, (b) median(difference) + Euler + hysteresis, (c) median(difference) + median

5 Shadow Detection

Previous research on the detection of shadows [10, 3, 18] has focused on two main uses: disambiguation for object recognition and recovery of the underlying surface detail. Here we consider only the former problem.

We can interpret shadows in the image, and the effect they have on the pixels in the scene, as a semi-transparent region in which the scene reflectance undergoes a local attenuation. Under the constraint that the imaging sensor is not undergoing motion, it is feasible to identify those regions within shadow by analysis of their photometric properties: firstly, they will have a photometric gain with respect to the background image, which is less than unity; secondly, this gain will be reasonably constant over the shadow region, except at the edges, where the effects of a finite size illumination source will tend to reduce the attenuation (i.e. the penumbra). Although similar photometric characteristics may also be exhibited by actual objects in the scene (i.e. those that are darker than the background and have a uniform gain with respect to the surface they occlude), their occurrence is expected to be less likely, and hence they may be interpreted as rare “accidents”.

The shadows are modelled as a constant contrast change between the reference or background image and the current image, and are detected by performing region growing to locate areas of constant photometric gain in the difference image. Heuristic rules are then used to cue possible shadow regions.

5.1 Region Growing

The algorithm starts with a thresholded image resulting from a frame differencing operation, generated using one of the methods described earlier in the paper. The algorithm calculates for each pixel within the binary detected blobs the intensity ratio between the current and background image. A single pass neighbourhood connectivity algorithm [8] is used for region growing, which performs a raster scan through the image, propagating region labels based on local eight-neighbour connectivity using constant values of the intensity ratio (i.e. the gain). The gain is simply defined as the ratio of the reference pixel intensity to the image intensity,

$gain_{x,y} = \frac{R_{x,y}}{I_{x,y}}$, resulting in ratios of less than unity in regions where the image is brighter than the reference, and greater than unity where it is darker.

For each of the four previously examined neighbours in the raster scan (which will already have been assigned a region label), the minimum difference between the pixel gain and the mean gain of each of the regions is used to identify into which region the pixel might be merged. If the gain is less than some prescribed threshold, then the pixel is labelled as belonging to that region and its gain is used to update the region mean and variance; otherwise, a new region is initiated.

A second stage of the algorithm merges similar neighbouring regions by using a t-test to compare the mean and variance of each pair of neighbouring regions. A significance level of 0.05 was used.

5.2 Shadow Identification

Following region growing, several rules are applied to the analysis of local regions to discriminate the shadow regions from the object. In the first instance, the variation of region statistics within the shadow region should vary smoothly and the shadow region should contain relatively homogeneous intensity ratio regions. Secondly, the gain values within the shadow region should always be less than unity (i.e. the pixels in the shadow region will be darker than those in the reference image).

The homogeneity of the region is estimated by considering its neighbours. The proportion of a regions' boundary which is shared with other regions is computed, and the ratio of the boundary shared with the background, against the total boundary length is determined. Secondly, the area of all directly bordering regions is calculated, and expressed as a proportion of the regions own area. These two values are thresholded to select homogeneous regions that have no substantial border with other regions which have no significant similarity in the gain ratio.

5.3 Results

Fig. 4a shows a composite image of a person walking through a car park. A reference image frame of the background is from the first frame in the sequence, acquired several seconds before the person enters the field of view. The shadows obtained are fairly strong, though they contain some significant brightness variations within the shadow region (i.e. the white lines). Fig. 4b shows the result of binary thresholding the difference image. The results of the first stage of region-growing are shown in fig. 4c, where it can be seen that both shadow and object are divided into a number of regions. The regions resulting from the merging operation are shown in fig. 4d. The shadow has been detected as mainly a single region, whilst the person (which is composed of a number of regions of significantly different grey levels), remains fragmented. The final classification of the shadow regions is shown in fig. 4e. In this composite taken over all 5 images in the sequence, the shadow detection fails to find the shadow for the 4th frame, which violates one of the identification rules used above, and is found to contain an internal region which is classified as background, resulting from light passing between the legs.

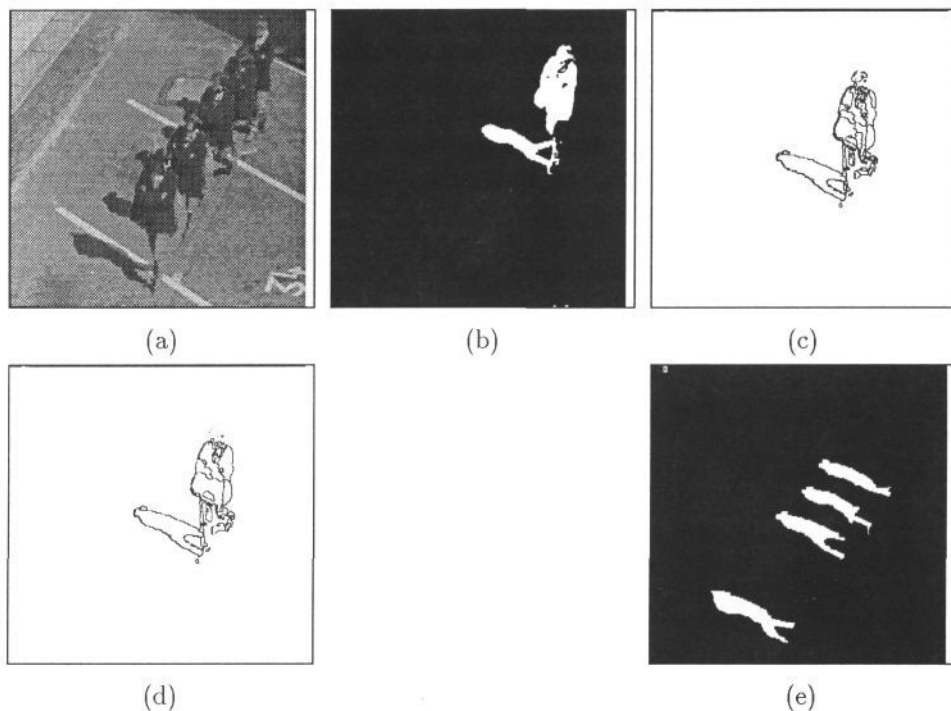


Figure 4: SHADOW (a) Grey-level composite (5 frames), (b) frame differenced and thresholded, (c) first region boundaries, (d) second stage regions, after merging, (e) composite of shadow classified regions

6 Discussion

Our initial results show that the most reliable method for thresholding a difference image to obtain the target blobs without spurious clutter is to first median filter the difference image, and then use the connectivity thresholding method followed by hysteresis. Comparisons were made with several other thresholding approaches, including those designed specifically for difference images as well as some more general standard image thresholding methods. More substantial verification is under progress using a larger test set. While it would be desirable to use performance measures to objectively rate the various approaches, our initial experience with several such measures has found little correlation between their ratings and subjective assessment.

The shadow detection algorithm seems to perform well on the image sequences that we have applied it to. However, several observations may be made on the experiments thus far. The shadow regions will be easier to find in images containing many (moving) objects that create shadows, since the photometric gain can be expected to be fairly constant over the image, and to exhibit a reasonable temporal constancy. The region growing algorithm can be affected by the shadow penumbra, though the application of a binary erosion operator, applied to the original differenced image can significantly minimise any deleterious effects. Some of

the initial observations on the characteristics of the shadows (especially associated with enclosed regions of significantly differing gain values) do not hold up well in practice. In particular, objects in the scene which are transparent (or semi-transparent) will contradict the assertion that the shadow should be homogenous. Further work is also in progress to investigate the potential for discriminating shadows from colour sequences, and at methods of identifying the shadows in data where the camera is not stationary.

7 REFERENCES

- [1] M. Bichsel. Segmenting simply connected moving objects in a static scene. *IEEE Trans. PAMI*, 16:1138–1142, 1994.
- [2] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8:679–698, 1986.
- [3] Wang Chengye, Huang Liuqing, and A. Rosenfeld. Detecting clouds and cloud shadows on aerial photographs. *Pattern Recognition Letters*, 12, no. 1:55–64, 1991.
- [4] W.S. Ching. A novel change detection algorithm using adaptive threshold. *Pattern Recognition Letters*, 12:459–463, 1994.
- [5] T.J. Ellis, P. Rosin, and P. Golton. Model-based vision for automatic alarm interpretation. *IEEE Aerospace and Electronic Systems Magazine*, 6(3):14–20, 1991.
- [6] S.B. Gray. Local properties of binary images in two dimensions. *IEEE Trans. Computers*, 20:551–561, 1971.
- [7] E.R. Hancock and J. Kittler. Adaptive estimation of hysteresis thresholds. In *Proc. CVPR*, pages 196–201, 1991.
- [8] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision 1*. Addison Wesley, 1992.
- [9] Y.Z. Hsu, H.H. Nagel, and G. Rekers. New likelihood test methods for change detection in image sequences. *CVGIP*, pages 73–106, 1984.
- [10] C. Jiang and M. O. Ward. Shadow indentification. pages 606–12, 1992.
- [11] K.P. Karmann and A. von Brandt. Moving object recognition using an adaptive background memory. In V. Cappellini, editor, *Time-Varying Image Processing and Moving Object Recognition 2*, pages 289–296. Elsevier, 1990.
- [12] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19:41–47, 1986.
- [13] W. Long and Y.H. Yang. Stationary background generation: An alternative to the difference of two images. *Pattern Recognition*, 23:1351–1359, 1990.
- [14] L. O’Gorman. Binarization and multi-thresholding of document images using connectivity. In *Symp. on Document Analysis and Info. Retrieval*, pages 237–252, 1994.
- [15] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man, and Cybernetics*, 9:62–66, 1979.
- [16] P.L. Rosin and T. Ellis. Detecting and classifying intruders in image sequences. In *British Machine Vision Conf.*, pages 293–300, 1991.
- [17] P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.
- [18] J. M. Scanlan, D. M. Chabries, and R. W. Christiansen. A shadow detection and removal algorithm for 2-d images. volume 4, pages 2057–60, 1990.
- [19] S. Song, M. Liao, and J. Qin. Multiresolution image motion detection and displacement estimation. *Machine Vision Applic.*, pages 17–20, 1990.
- [20] W.H. Tsai. Moment-preserving thresholding: a new approach. *CVGIP*, 29:377–393, 1985.
- [21] Y.H. Yang and M.D. Levine. The background primal sketch: An approach for tracking moving objects. *Machine Vision Applic.*, 5:17–34, 1992.