Illumination Invariant Motion Segmentation of Simply Connected Objects

Martin Bichsel *
University of Zurich
Department of Computer Science
MultiMedia Laboratory
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland
mbichsel@ifi.unizh.ch

Abstract

A new segmentation algorithm exploits local image quantities which are invariant to changing illumination. Local object-background probability estimates are obtained by comparing illumination invariant quantities in an actual image with the corresponding quantities in a reference image. The objects' simply connectedness is included directly into the probability estimates and leads to an iterative optimization procedure that is implemented efficiently. This new approach avoids early thresholding, explicit edge detection, motion analysis, and grouping.

1 Introduction

In many object recognition applications the objects of interest are moving whereas the background is static or can be stabilized [1, 2]. Motion segmentation can enormously simplify subsequent object recognition steps. Therefore, segmenting moving objects in a static scene is an important computer vision task.

In real-world applications the illumination may vary considerably, *e.g.* due to clouds moving in the sky. Furthermore, many video cameras have a built-in automatic gain control keeping the pixel values within a reasonable range while the illumination changes. Moving objects activate this gain control and lead to brightness variations of the whole scene.

In recent years a number of different approaches have been proposed for moving object segmentation (*e.g.* [3],[4],[5],[6],[7],[8]). Little attention, however, has been paid to the problems caused by varying illumination. Most approaches are based on absolute differences of subsequent frames or on the difference between the current frame and an estimate of the static background [5]. Using image derivatives (*e.g.* [9]) rather than absolute values reduces illumination effects but cannot completely

*This work was supported by the consortium VISAGE and KWF grant No. 2440.1

BMVC 1994 doi:10.5244/C.8.45

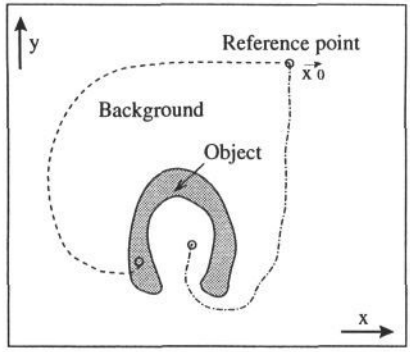

Figure 1: Any background point can be connected with reference background point $\vec{x}_0$ by a path remaining completely within the background. Any path from an object point to $\vec{x}_0$ at least touches the object contour once.

eliminate them. Therefore, this paper focuses on illumination invariant image quantities.

A second problem arises since local motion detectors do not lead to a dense object map, in general. Dense maps are therefore obtained by combining connectivity information with the local object information. Most approaches apply a thresholding operation in an early stage of the segmentation process and exploit the fact that objects consist of multiple connected pixels only after thresholding. With early thresholding, however, useful information is thrown away that is missing in the final segmentation step.

The approach presented in this paper includes connectivity information directly into the probability estimates and applies a thresholding operation only at the very end. Traditional heuristics are replaced by experimental properties of local image statistics.

2 Definitions

The following discussion assumes that the considered image sequence consists of moving *simply connected* objects in front of a *single, connected, and stationary* background. Additionally, at least one point $\vec{x}_0$ is known to belong to the background.

The following conditions uniquely define a binary object-background function for all image points (see Fig. 1):

1. Let the reference point $\vec{x}_0$ belong to the background.

2. For any background point $\vec{x}$ there exists at least one connected path from $\vec{x}_0$ to $\vec{x}$ that touches no object point.

3. Every path from $\vec{x}_0$ to any object point $\vec{x}$ touches at least one object point. The first object point being touched by such a path is a contour point, *i.e.* an object point which has at least one neighbouring background point.

3 Illumination invariant image quantities and their probability distribution

If a (single) light source varies in luminance from L_0 to L then the pixel response $I(x,y)$ of a linear camera varies from $I_0(x,y)$ to $\frac{L}{L_0} \cdot I_0(x,y)$ (*e.g.* [10]). Local illumination invariant quantities can be obtained by dividing $I(x,y)$ by a normalizing function $f(\mathbf{I},L)$ of the image such that $f(\mathbf{I},x,y,L) = \frac{L}{L_0} \cdot f(\mathbf{I},x,y,L_0)$, where $\mathbf{I}$ denotes the image. Thus, for any such function f, the quantity $I(x,y)/f(\mathbf{I},x,y,L)$ is invariant to illumination changes.

Requiring that the new image quantities should not only be invariant to a global change of luminance but also to smooth variations of the light distribution (*e.g.* due to smooth shadows) the function $f(\mathbf{I},x,y,L)$ should be as local as possible. The smallest symmetric neighbourhood around an image point is 3 by 3. Furthermore, $f(\mathbf{I},x,y,L)$ should be insensitive to Gaussian noise due to electronic noise in the static background. Therefore, the local mean

$$f(\mathbf{I},x,y,L) = m(x,y) = \frac{1}{9}\sum_{j'=-1}^{1}\sum_{k'=-1}^{1} I(x+j',y+k') \tag{1}$$

is ideally suited as a normalizing function and leads to illumination invariant local image quantities

$$r(x,y) = \frac{I(x,y)}{m(x,y)} \tag{2}$$

These quantities can be transformed into

$$r(x,y) = \frac{I(x,y)-m(x,y)+m(x,y)}{m(x,y)} = \frac{I(x,y)-m(x,y)}{m(x,y)} + 1 \tag{3}$$

so that, up to a constant, they become a Prewitt Laplacian [14] which is normalized by the local mean. It should be pointed out that this is very similar to the response of a retinal ganglion cell. Retinal ganglion cells show a typical ON-center or OFF-center response [11] combined with a local gain control mechanism which makes sure that the response is independent of luminance changes [12].

A straight forward application to motion segmentation consists in calculating the illumination invariant frame differences $r(x,y,t) - r(x,y,t\text{-}1)$ for change detection, where t is the frame index.

This choice, however, leads to illumination dependent noise in the static background. This is due to the fact that the noise of typical CCD cameras is dominated by amplifier noise and quantization noise so that noise is considerable at $L = 0$ and is only slowly increasing as a function of luminance L [13]. Due to averaging, the noise of $m(x,y)$ is considerably smaller than the noise of $I(x,y,t)$. As a consequence, the noise of the quantity $r(x,y,t)-r(x,y,t-1)$ is approximately proportional to $1/m(x,y)$ so that the local object-background decision has to depend on $m(x,y)$.

Approximatively constant noise, on the other hand, is obtained by calculating the illumination invariant quantity

$$\begin{aligned} D(x,y,t,t-1) &= m(x,y,t)\cdot(r(x,y,t)-r(x,y,t-1)) &(4)\\ &= I(x,y,t) - m(x,y,t)\cdot r(x,y,t-1) &(5)\end{aligned}$$

In stationary regions the expectation value of $D(x,y,t,t-1)$ is zero, independent of varying illumination, and the variance of $D(x,y,t,t-1)$ approximately equals the noise variance of the sensor multiplied by $\sqrt{2}$ since both $I(x,y,t)$ and $m(x,y,t) \cdot r(x,y,t-1)$ are approximately equally effected by the sensor noise. The noise is expected to be Gaussian.

Instead of segmenting based on interframe differences we can also segment the difference between a reference image and the actual frame $D(x,y,t,0) = I(x,y,t) - m(x,y,t) \cdot r(x,y,0)$. The reference image can be obtained by averaging over several frames and converting this image into the illumination invariant representation $r(x,y,0)$. In this case the noise of $m(x,y,t) \cdot r(x,y,0)$ can be neglected compared to the noise of $I(x,y,t)$ so that the variance of $D(x,y,t,0)$ approximately equals the noise variance of the sensor and is thus reduced by a factor of $\sqrt{2}$ compared to $D(x,y,t,t-1)$.

In moving areas which move with constant velocity (v_x, v_y) we will measure

$$\begin{aligned} D_m(x,y,t,\Delta x,\Delta y) &= D(x,y,t,t-\Delta t) && (6)\\ &= I(x,y,t) - m(x,y,t) \cdot r(x-\Delta x, y-\Delta y, t) && (7) \end{aligned}$$

where $(\Delta x, \Delta y) = \Delta t \cdot (v_x, v_y)$ and Δt is the time difference between the measured frames.

For small displacements, $D_m(x,y,t,\Delta x,\Delta y)$ is approximated by the Taylor series

$$\begin{aligned} D_m(x,y,t,\Delta x,\Delta y) \approx\ & -\Delta x\, I_x(x,y,t) - \Delta y\, I_y(x,y,t) \\ & + (\Delta x\, m_x(x,y,t) + \Delta y\, m_y(x,y,t)) \cdot r(x,y,t) \qquad (8) \end{aligned}$$

where I_x, I_y, m_x, and m_y are the spatial derivative of I and m in the directions x and y. Within textured image regions and along object contours m is smooth compared to I. Hence, m_x is small compared to I_x and m_y is small compared to I_y so that

$$D_m(x,y,t,\Delta x,\Delta y) \approx -\Delta x\, I_x(x,y,t) - \Delta y\, I_y(x,y,t). \qquad (9)$$

In order to calculate the local object-background probability, the statistics of the quantity $D_m(x,y,t,\Delta x,\Delta y) = I(x,y,t) - m(x,y,t) \cdot r(x-\Delta x, y-\Delta y, t)$ in moving object region must be compared to the Gaussian distribution of $D(x,y,t,t-1)$ within the static background. The shape of the distribution of D_m can be estimated by calculating $D_m(x,y,t,\Delta x,\Delta y)$ for a static scene. In this way, distributions were measured for a number of images, selecting $(\Delta x, \Delta y) = (1,0)$ in order to simulate a small displacement and $(\Delta x, \Delta y) = (50,0)$ in order to simulate a very large displacement as expected when $D(x,y,t,0)$ is calculated based on a reference image.

In both cases we found distributions that can be approximated by a two-sided exponential (Laplacian) distribution

$$P\{D_m(x,y,t,\Delta x,\Delta y)\} \approx \frac{1}{2\lambda} \exp\left(-\frac{|D_m(x,y,t,\Delta x,\Delta y)|}{\lambda}\right) \qquad (10)$$

where λ is an experimental constant. Examples of this qualitative law, which was confirmed for a number of pictures, are shown in Figures 2(a) and (b) for the

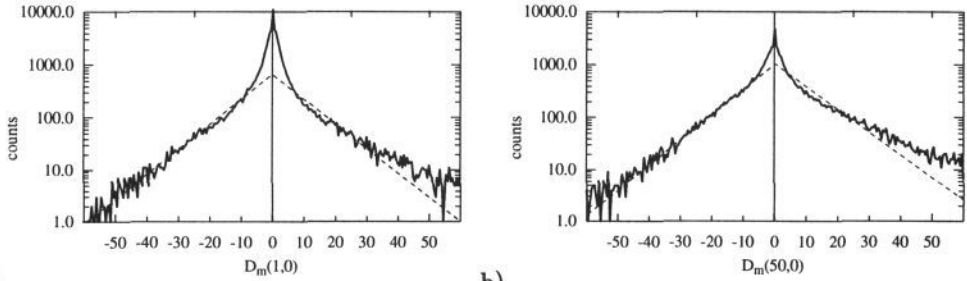

Figure 2: Histograms of $D_m(x, y, t, \Delta x, \Delta y)$ for the Lenna picture. In the linear-log plots an approximately linear law (triangular shape) can be seen.
(a) Log-histogram of $D_m(x, y, t, 1, 0)$.
(b) Log-histogram of $D_m(x, y, t, 50, 0)$.

picture Lenna. In linear-log plots the graphs show a linear law, corresponding to an exponential distribution. The noisy structures at both sides of the triangle are due to the Poisson statistics of histograms which become prominent for low count numbers [15]. Peaks in the central region of the plots correspond to untextured regions within objects. Note that, in the plots, a Gaussian distribution would show up as a parabola which is not consistent with the data. The slight asymmetry of the distribution seems to be due to a nonlinear camera response in the acquisition of the picture Lenna.

4 Local Background Probability Estimation

In natural image sequences we cannot expect a correct classification of each image point. Therefore, we can at best assign a probability value $P\{b(\vec{x})\}$ that a point $\vec{x} = (x, y)$ belongs to the background or, equivalently, a probability value $P\{o(\vec{x})\} = 1 - P\{b(\vec{x})\}$ that $\vec{x}$ belongs to an object. According to the definitions in Section 2, object-background probability estimation is especially important for object contour points.

As discussed in the previous section, the illumination invariant quantities $D(x, y, t_1, t_2)$ are expected to approximately show a Laplacian distribution in moving textured regions and, especially, along contours c of moving objects:

$$P\{D(x, y, t_1, t_2) \,|c\} \approx \frac{1}{2\lambda} \exp\left(-\frac{D(x, y, t_1, t_2)}{\lambda}\right) \tag{11}$$

Gaussian white noise, on the other hand, is expected in the stationary background.

Using notation $D_T(\vec{x})$=$D(x,y,t_1,t_2)$, let us estimate $P\{b(\vec{x})|D_T(\vec{x}),b(\text{neighbour})\}$, *i.e.* the conditional probability that the pixel at location $\vec{x}$ belongs to the background, given that at least one neighbouring pixel belongs to the background (denoted by b(neighbour)). It is important to know $P\{b(\vec{x})|D_T(\vec{x}), b(\text{neighbour})\}$ because, according to Section 2, a path from a reference background point x_0 to x can only remain completely within the background if it never touches a contour point. Let us subsequently omit $\vec{x}$ in order to shorten the formulas.

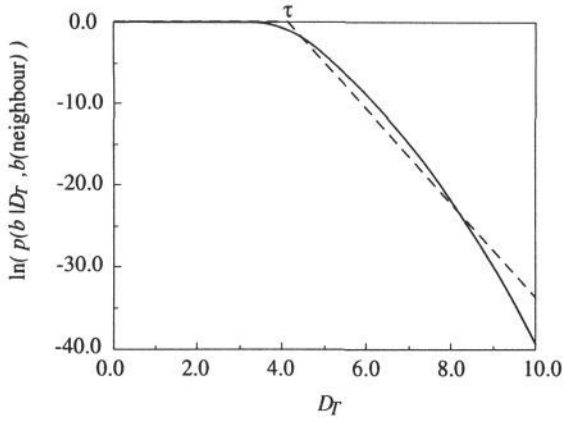

Figure 3: Logarithm of conditional background probability $\ln(P\{b|D_T, b(\text{neighbour})\})$ versus measured D_T. The exact curve (solid line) is approximated with two linear segments (dashed line).

We can determine $P\{b|D_T, b(\text{neighbour})\}$ by applying Bayes' theorem [15]:

$$P\{b|D_T, b(\text{neighbour})\} = \frac{1}{1+\frac{P\{c\}}{P\{b\}}\frac{P\{D_T|c\}}{P\{D_T|b\}}} \approx \frac{1}{1+\frac{P\{c\}}{P\{b\}}\frac{\sqrt{2\pi}\sigma}{2\lambda}\exp\left(\frac{D_T^2}{2\sigma^2}-\frac{|D_T|}{\lambda}\right)} \tag{12}$$

$$= \frac{1}{1+\exp\left(\left(\frac{|D_T|-\alpha}{\beta}\right)^2+\gamma\right)} \tag{13}$$

$$\text{with} \qquad \alpha = \frac{\sigma^2}{\lambda}, \quad \beta = \sqrt{2}\sigma, \quad \gamma = -\frac{\sigma^2}{2\lambda^2} + \ln\left(\frac{P\{c\}}{P\{b\}}\frac{\sqrt{2\pi}\sigma}{2\lambda}\right) \tag{14}$$

where $P\{b\}$ and $P\{c\}$ are the *a priori* probabilities for background and object contour.

In a plot of $P\{b|D_T, b(\text{neighbour})\}$ versus D_T, variable α describes a shift to the right, variable β describes a scaling of the D_T-axis, whereas γ determines the curve shape.

The ratio $P\{c\}/P\{b\}$ can be estimated by dividing the expected number of contour pixels by the expected number of background pixels. Using typical experimental values, σ=1, λ=10, and $P\{c\}/P\{b\}$=0.01 the shape determining parameter γ is on the order of -7.

The function $\ln(P\{b|D_T, b(\text{neighbour})\})$ versus D_T is plotted in Figure 3, taking parameters α=0.5, β=1.4, and γ=-6.0. Figure 3 also shows that the function $\ln(P\{b|D_T, b(\text{neighbour})\})$ can be approximated by two straight lines:

$$\ln(P\{b|D_T, b(\text{neighbour})\}) \approx \begin{cases} -|\mu| \cdot (|D_T| - \tau), & \text{if } |D_T| \geq \tau \\ 0, & \text{if } |D_T| \leq \tau \end{cases} \tag{15}$$

where τ is the transition threshold. The slope $|\mu|$ of the straight line above the transition threshold is not required to be known because different slope values only

correspond to different overall scalings of the final object-background function that can be included in the final segmentation threshold. Approximation (15) even becomes better if γ is more negative than in Figure 3. The piecewise linear approximation is advantageous for an efficient implementation and it reduces the three parameters α, β, and γ to a single parameter τ.

5 Including Connectivity Information

Given any path from $\vec{x}_0$ to $\vec{x}$ let us estimate the probability that the path remains completely within the background, *i.e.* no contour point lies within the path. The background probability for the point $\vec{x}$ is estimated following the definitions in Section 2: select the path with the highest probability that all points on the path belong to the background. Fortunately, this global optimization over all possible paths can be calculated locally.

Let $\mathbf{I}$ denote the whole picture information. For an estimation of the background probability $P\{b(\vec{x})|\mathbf{I}\}$ at point $\vec{x}$, including global connectivity information, let us assume that the background probability $P\{b(\vec{x}_{ni})|\mathbf{I}\}$ is already known for the neighbours $\vec{x}_{ni}$ of $\vec{x}$. The global optimum at $\vec{x}$ is then obtained by optimizing over all local paths from the neighbours to $\vec{x}$:

$$P\{b(\vec{x})|\mathbf{I}\} = \max_{i \in \text{neighbours of } \vec{x}} P\{b(\vec{x}_{ni})|\mathbf{I}\} \cdot P\{b|D_T, b(\text{neighbour})\} \tag{16}$$

$$\ln(P\{b(\vec{x})|\mathbf{I}\}) = \max_{i \in \text{neighbours of } \vec{x}} \ln(P\{b(\vec{x}_{ni})|\mathbf{I}\}) + \ln(P\{b|D_T, b(\text{neighbour})\}) \tag{17}$$

where $P\{b|D_T, b(\text{neighbour})\}$ is estimated according to the preceding section.

The probabilities $P\{b(\vec{x}_{ni})|\mathbf{I}\}$ and $P\{b|D_T, b(\text{neighbour})\}$ have to be multiplied in equation (16) because the path remains completely in the background only if the path to the neighbour remains in the background *and* the point $\vec{x}$ belongs to the background. It is further assumed that the D_T-values along the global path to a neighbour and the D_T-value at point $\vec{x}$ are approximately independent.

The global estimate of the background probability at $\vec{x}$ can be evaluated by applying equation (17) iteratively, whereas initially all pixels besides the known background pixels are set to a large negative logarithmic probability value. If, at any iteration, for some point a lower probability is calculated than in the previous iteration then the previous value is retained so that the logarithmic probability values increase monotonically for each iteration.

The logarithmic global estimate of the local background probability decreases monotonically as a function of path length because $\ln(P\{b|D_T, b(\text{neighbour})\}) \leq 0$. Combining this fact with the fact that the logarithmic probability values increase monotonically the convergence of the iterative approach can be proven and the resulting probability distribution corresponds to the global optimum over all paths [16]. After a few iterations, when the result is stable, the logarithmic background probability is known for each image point and thus denotes the desired foreground-background function on which a threshold operation is applied for the final segmentation.

6 Implementation

The computation for the global optimization over all paths consists of 3 steps:

1. Calculate D_T at each image point.
2. Estimate local conditional background probabilities based on equation (15).
3. Estimate local background probabilities based on global image information by applying equation (17) iteratively with a fixed number of iterations.

Pixels were defined to be 4-connected, *i.e.* each pixel has two horizontal and two vertical neighbours. The upper image boundary was defined to belong to the background in order to provide boundary values for the iterative process.

On a sequential computer the algorithm converges fastest with a Gauss-Seidel procedure where updated probability values become immediately available to the calculation of the following values. The algorithm typically converges in less than 5 iterations for simply shaped objects, where an iteration means one scan over all pixels. In the current implementation two subsequent 512 by 512 frames are segmented in less than 5 seconds on a SUN SPARCstation 2. The processing time is dominated by pyramid reduction and conversion of raw values into illumination invariant quantities so that the dominating part is independent of image content.

7 Experiments

The presented segmentation algorithm was tested on a number of image sequences. In order to reduce the random noise, the original 512 by 512 images were reduced to 128 by 128 images by applying a standard Burt reduction routine [17]. For the examples in Figure 4, a transition threshold of $\tau = 3.0$ was used and the logarithmic global probabilities $\ln(P\{b(\vec{x})|\mathbf{I}\})$ were thresholded at -4.

The algorithm was tested with a person walking in an office scene. Images were captured with a Sony77CE video camera (with gamma correction switched off) and a Silicon Graphics VisionLab digitizer board. This equipment shows an approximatively linear luminance response [18] as long as the pixel values are below saturation.

In the middle of the sequence a controlled change of illumination was achieved by quickly placing a filter in front of the camera lens. This filter reduced the luminance of the scene by a factor 1.5. Frame 26 (before luminance reduction) and frame 27 (after luminance reduction) as well as the reference frame are shown in Figures 4(a-c). The illumination invariant representation of the reference frame is shown in Figure 4(d). The thresholded logarithmic global probabilities are shown in Figures 4(e) and 4(f), resulting in a correct segmentation of the person in spite of the varying illumination. In the static background, only a few pixels are classified as object pixels. These pixels could easily be removed with standard image processing techniques.

The segmentation accuracy in the reduced images is close to pixel precision as long as motion smear is low. With high motion smear, errors may occur in front of homogeneous background since this situation leads to similar illumination invariant quantities at object contours and in the background.

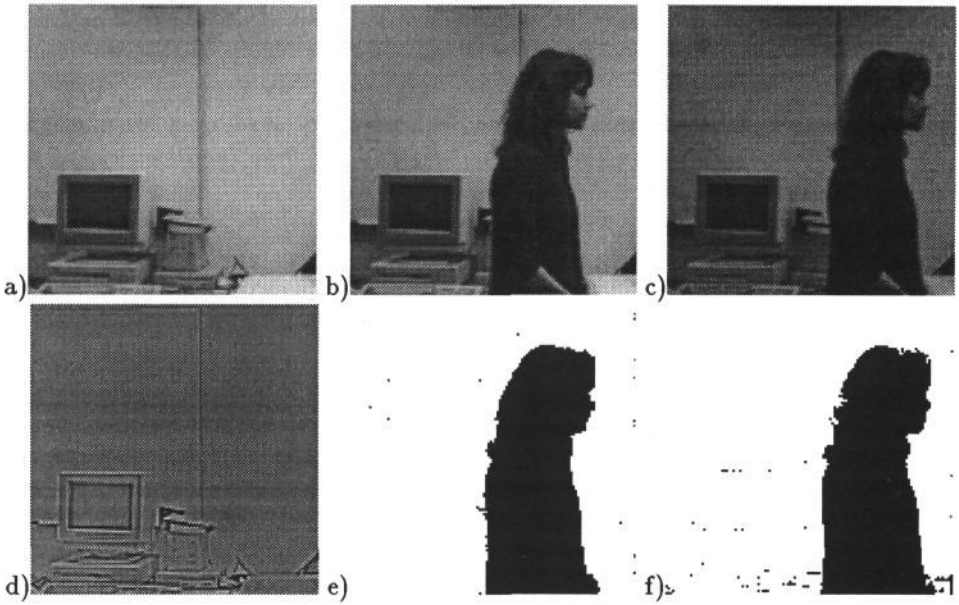

Figure 4: Illumination invariant motion segmentation.
(a) Reference background.
(b) Moving person in front of a stationary background, before illumination change.
(c) Moving person in front of a stationary background, after illumination change.
(d) Illumination invariant representation of picture 4(a).
(e) Segmentation of picture 4(b).
(f) Segmentation of picture 4(c) .

8 Conclusions

In this paper a method was derived for segmenting moving objects under varying illumination. This method was successfully tested on image sequences in a laboratory environment. It is ideally suited to be used in conjunction with an automatic gain control mechanism that keeps pixel values within a reasonable range.

9 Acknowledgements

I would like to thank K. Szabo, for carefully looking through this manuscript.

References

[1] P. J. Burt, R. Hingorani, and R. J. Kolczynski, "Mechanisms for Isolating Component Patterns in the Sequential Analysis of Multiple Motion" *Proc. of the IEEE Workshop on Visual Motion*, Nassau Inn, Princeton, New Jersey, pp. 187-193, 1991.

[2] J. R. Bergen et al., "Computing Two Motions from Three Frames", *IEEE Proc. of the Third Int. Conf. on Computer Vision*, Osaka, Japan, pp. 27-32, 1990.

[3] R. C. Jain, "Segmentation of Frame Sequences Obtained by a Moving Observer", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 5, pp. 624-629, 1984.

[4] M. K. Leung and Y. Yang, "Human Body Motion Segmentation in a Complex Scene" *Pattern Recognition*, Vol. 20, No. 1, pp. 55-64, 1987.

[5] A. Shio and J. Sklansky, "Segmentation of People in Motion", *IEEE Proceedings*, pp. 325-332, 1991.

[6] S. D. Blostein and T. S. Huang, "Detecting Small, Moving Objects in Image Sequences using Sequential Hypothesis Testing", *IEEE Trans. on Signal Processing*, Vol. 39, No. 7, pp. 1611-1929, 1991.

[7] G. W. Donohoe et al., "Change Detection for Target Detection and Classification in Video Sequences", *Proc. Int. Conf, Acoustics, Speech and Signal Processing*, New York, pp. 1084-1087, 1988.

[8] Y. Z. Hsu et al., "New Likelihood Test Methods for Change Detection in Image Sequences", *CVGIP*, Vol. 26, pp. 73-106, 1984.

[9] M. Bichsel, "Segmenting Simply Connected Moving Objects in a Static Scene", *Transactions on Pattern Recognition and Machine Intelligence*, to be published.

[10] M. Bichsel, "Analyzing a Scene's Picture Set under Varying Illumination", *CVGIP: Image Understanding*, submitted, 1993.

[11] D. Hubel, *Eye, Brain and Vision*, Scientific American Library, 1988.

[12] J. E. Dowling, *The Retina: An Approachable Part of the Brain*, Belknap Press, 1987.

[13] G. Healey and R. Kondepudy, "CCD Camera Calibration and Noise Estimation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Champaign, Illinois, 1992.

[14] W. K. Pratt, *Digital Image Processing, Second Edition*, John Wiley & Sons, Inc., 1991.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd Edition, McGraw-Hill, Inc., 1987.

[16] M. Bichsel and A. P. Pentland, "A Simple Algorithm for Shape from Shading", *Proc. of the IEEE CVPR Conference*, Champaign, Illinois, pp. 459-465, 1992.

[17] P. J. Burt, "Fast Filter Transforms for Image Processing", *Computer Vision, Graphics and Image Processing*, Vol. 16, pp. 20-51, 1981.

[18] M. Bichsel and K. W. Ohnesorge, "How to Measure a Camera's Response Curve from Scratch", *Pattern Recognition Letters*, submitted, 1993.