Multiresolution Statistical Object Recognition

Jonathan D. Baker *
Massachusetts Institute of Technology
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02319
jonbaker@ai.mit.edu

William M. Wells, III †
Harvard Medical School and
Brigham and Women's Hospital
Department of Radiology
75 Francis Street, Boston, MA 02115
sw@ai.mit.edu

Abstract

A multiresolution statistical estimation approach to object recognition and pose estimation has been developed that incorporates models for the effects of background clutter, occlusion, and sensor noise. The resulting formulation is essentially an alignment approach in which separate methods are used to align, refine, and verify object poses. A new model for background features is proposed along with a method of applying hypothesis testing techniques to determine whether a model object appears in an image. In addition, the use of multiple resolutions to improve computational efficiency and robustness is described. A number of experiments are presented which demonstrate the multiresolution statistical object recognition system.

1 Introduction

The ability to recognize objects quickly, determine their pose, and provide a measure of confidence in the solution is crucial to a wide range of applications including automatic target detection and recognition, robotics, medical imaging, manufacturing, and inspection. The goal of this research is to enable a computer to recognize an object and compute its position and orientation in a cluttered and noisy image using a multiresolution statistical formulation.

A number of statistical estimation techniques for object recognition in the visual domain have been developed that provide explicit models for the uncertainties in object recognition and allow the application of methods of statistical estimation to the recognition of objects in images [Wells, 1990; Wells, 1991; Wells, 1992; Wells, 1993]. This paper expands upon these statistical estimation algorithms by incorporating an improved noise model for background features, hypothesis testing for detecting the presence of the model object, and multiresolution processing for pose estimation.

*Research supported by an Office of Naval Research Fellowship.

†Also affiliated with the MIT Artificial Intelligence Laboratory.

[1]This paper describes research supported in part by ARPA under ONR contract N00014-91-J-4038.

1.1 Modeling Objects

In this research, we utilize an image-based approach to object modeling in which images of the object are used to construct the model. Ullman and Basri [Ullman and Basri, 1989] describe a method known as linear combination of views for building a model of an object using a number of 2D views and then using an interpolation of the views to predict what the object will look like at a given pose. This is the approach to 3D recognition used in this research, although we focus on the 2D recognition problem assuming that the interpolated view has already been generated.

1.2 The Statistical Approach

In this paper, statistical methods are used to refine poses and evaluate hypotheses about the presence of an object in a scene. This approach converts the recognition problem into a well-defined optimization problem and also allows the assumptions about the problem to become explicit.

The resulting statistical recognition algorithm falls within the general category of an alignment approach with indexing. The three primary steps of the algorithm are align, refine, and verify. In brief summary, the algorithm operates by generating a number of hypotheses of minimal correspondences at low resolution using an indexing table. The initial hypotheses for pose are then refined using statistical estimation and the Expectation-Maximization (EM) Algorithm which is an iterative optimization algorithm. The best scoring poses are propagated on to higher resolution levels where the pose is reestimated after the addition of the finer detail. After convergence at maximum resolution, the highest scoring pose is evaluated using a binary hypothesis testing procedure to determine whether the evidence indicates that the object is present in the image or not.

2 Statistical Models

In order to apply statistical tools to the problem of object recognition, we must first formulate explicit models of the statistical aspects of the recognition problem. The statistical models for recognition are based on those developed by Wells [Wells, 1992] and are discussed more fully in [Baker, 1994].

2.1 Image Features

In this research, we use edge-based features since the edge contours in an image contain a great deal of information about the objects in a scene.

There are three sources of image features that we will be working with: model features, scene features, and background features. Model features are the image features we extract from a model image and use to build the representation of the model object. Scene features are the image features that we extract from an input image in which we try to locate the model object. And background features are the source of all the scene features that do not come from the model object.

Let us now present a notation to describe these different categories of features. Suppose we extract k features from the model image and label them $m_1, m_2, \ldots, m_k$. The collection of these model features is the set

$$M = \{m_1, m_2, \ldots, m_k\}.$$

Similarly, suppose we extract n features from a scene image and label them $y_1, y_2, \ldots, y_n$. The collection of these scene features is the set

$$Y = \{y_1, y_2, \ldots, y_n\}.$$

Furthermore, in any scene most of the features will not correspond to the model, but to other objects in the background of the scene. We can define a set of n background features

$$B = \{b_1, b_2, \ldots, b_n\}$$

to represent the true location of scene features that come from the background of the image. Of course, we are not able to directly measure the true location of these background features since all of our measurements involve noise, and their values therefore have to be estimated. The concept of multiple background features is an extension to Wells' single background feature '$\perp$' [Wells, 1992] and to Grimson's null character '$*$' [Grimson, 1990]. These single background feature formulations treated all background features interchangeably. In this work, however, we introduce the notion of individual background features which arise from specific localized features in the background of the image. This makes intuitive sense since each observed scene feature corresponds to some distinct physical aspect of the scene image.

The label of a feature will be written in an italic typeface such as y_i and the measurement associated with that feature will be written in boldface as in $\mathbf{y}_i$.

2.2 Feature Correspondence Model

The interpretation of a scene image is described by a *correspondence function* $\Gamma(y_i)$ or *correspondence vector* $\mathbf{\Gamma}$ where $\Gamma(y_i)$ indicates the model or background feature to which scene feature y_i corresponds and $\mathbf{\Gamma}_i \equiv \Gamma(y_i)$.

2.3 Prior Densities for Correspondences

The prior density model for feature correspondences is based on the clutter level in the image. If the ratio of background features to scene features is denoted by P_B, then the probability that an arbitrary image feature corresponds to the background is

$$\Pr[\Gamma_i \in B] = P_B. \tag{1}$$

Assuming independence of all of the correspondences before observation of the image, we can write the joint probability density of the correspondence vector as the product of the marginal probabilities

$$p_{\mathbf{\Gamma}}(\mathbf{\Gamma}) = \prod_{i=1}^{n} p_{\mathbf{\Gamma}_i}(\Gamma_i) = \prod_{ij:\Gamma_i = m_j} \frac{P_B}{n} \prod_{ij:\Gamma_i = b_j} \frac{1-P_B}{k}. \tag{2}$$

2.4 Projection Model

In the absence of noise, the positions of image features are completely determined by their correspondence and pose. The projection model expresses the mathematics of this deterministic transformation from model features to scene features. In its most general form, this projection relation can be written as

$$\mathbf{y}_i = \mathcal{P}(\mathbf{M}_j, \boldsymbol{\beta}) \quad \text{for } \Gamma(y_i) = m_j$$

where $\mathbf{y}_i$ represents the feature measurement associated with scene feature y_i, $\mathbf{M}_j$ represents the feature measurement associated with model feature m_j, $\boldsymbol{\beta}$ represents the pose of the object, and $\mathcal{P}(\cdot, \cdot)$ represents an arbitrary projection function.

The pose estimation problem is significantly simplified when the feature projection model is linear with respect to pose so that

$$\mathbf{y}_i = M_j \boldsymbol{\beta}.$$

In this paper, we use a 2D point feature projection model which incorporates a pose vector with four degrees of freedom. A pose vector of this type corresponds to a 2D similarity transform with a two-dimensional translation t_x and t_y, a scaling factor s, and a rotation θ in the plane.

2.5 Noise Model

The measurement of image features as expressed in the projection model of the previous section only captures part of the measurement process because in a real image the projection of features is not exact and our recognition system must be able to deal with uncertainty in the measurement of features. We model the sum effect of the various types of noise as independent white Gaussian noise. The appropriateness of this Gaussian model was demonstrated in [Wells, 1992].

Previously, we used a Gaussian model for the error in features matched to the model and a uniform density for features matched to the background [Wells, 1992]. In this research, however, we utilize Gaussian models for both cases.

2.5.1 Measurement Noise for Model Features

For scene features that correspond to model features, we will approximate the measurement noise as independent identically distributed white Gaussian noise with zero mean and covariance $\mathbf{\Lambda}_1$. If we let $\mathbf{w}_1$ represent the noise in the measurement of features corresponding to the model, we can write

$$\mathbf{w}_1 \sim N(\mathbf{0}, \mathbf{\Lambda}_1).$$

The actual values for the variance of the noise can be estimated for any image domain from sample images in which the correct pose of the model object is known.

2.5.2 Measurement Noise for Background Features

Modeling the noise for features corresponding to the background is more complicated because we have no useful *a priori* information about the distribution of the background features. Previously, a uniform density bounded by the limits of the scene feature coordinates was used to capture the maximum entropy nature of the problem.

Experiments conducted in the course of this research with the uniform density model for background features have indicated that this model significantly underestimates the probability of background features. As a prior density on background features, the uniform model tells us the expected density of features before observation of the image. However, we can use the observed image itself in estimating the background density at recognition time. The key point is that when presented with an actual image, we are not looking at any arbitrary background, but rather a particular background, and our observation of the image gives us significant information about that particular background density.

The inadequacy of the uniform distribution of image features is also evident from the observation that the clutter is not uniformly distributed, but follows a certain local probabilistic distribution.

We therefore model each scene feature corresponding to the background as the measurement of some true background feature corrupted by some measurement noise.

Such a model makes intuitive sense since an observed image feature must correspond to some physical aspect of the image and as we have discussed previously the features we are working with are physically localized features. We can therefore use the observed scene image to make an estimate of the true locations of background features.

In summary, the noise for background features will be modeled as zero-mean Gaussian noise with covariance $\mathbf{\Lambda}_0$

$$\mathbf{w}_0 \sim N(\mathbf{0}, \mathbf{\Lambda}_0).$$

One final point is that we require $\mathbf{\Lambda}_0 > \mathbf{\Lambda}_1$ to enforce the desired metrical consistency constraint on our solution. Furthermore, it seems reasonable that the variance of background features should be greater than model features since we have no knowledge about their true location. In this research, we have used a value of $\mathbf{\Lambda}_0 = c\mathbf{\Lambda}_1$ where c is a parameter that must be set. Determination of this parameter from the known statistical characteristics of the problem is still being investigated. Empirically, we have determined that a value of $c \approx 1.2$ works reasonably well and this is the value that we have used in the experiments in Section 5.

2.6 Measurement Model

The correspondence, projection, and noise models can be combined in a measurement model that expresses the relationship between the pose, the correspondences, and the observed image features.

Using the linear projection model and the Gaussian noise models discussed above, our observation model becomes

$$\mathbf{y}_i = \begin{cases} \mathbf{M}_j\boldsymbol{\beta} + \mathbf{w}_1 & \text{if } \Gamma(y_i) = m_j \\ \mathbf{b}_j + \mathbf{w}_0 & \text{if } \Gamma(y_i) = b_j \end{cases}$$

where $\mathbf{w}_1 \sim N(\mathbf{0}, \mathbf{\Lambda}_1)$ and $\mathbf{w}_0 \sim N(\mathbf{0}, \mathbf{\Lambda}_0)$. With this observation model, the probability distribution of an image feature $\mathbf{y}_i$ depends on its correspondence and on the pose $\boldsymbol{\beta}$ and we can write the conditional density of y_i as

$$p_{\mathbf{y}_i|\Gamma_i,\beta}(\mathbf{Y}_i \mid \Gamma_i, \boldsymbol{\beta}) = \begin{cases} N(\mathbf{M}_j\boldsymbol{\beta}, \mathbf{\Lambda}_1) & \text{if } \Gamma_i = m_j \\ N(\mathbf{b}_j, \mathbf{\Lambda}_0) & \text{if } \Gamma_i = b_j \end{cases}. \quad (3)$$

The interpretation of this conditional density is that the scene features corresponding to the model are normally distributed about their projected location with covariance $\mathbf{\Lambda}_1$ and scene features corresponding to the background are normally distributed about their true location with covariance $\mathbf{\Lambda}_0$.

We can collect the image features into an observation vector $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T$. Since we have assumed that the noise is independent, the joint conditional probability density is the product of the marginal conditional probability densities

$$p_{\mathbf{y}|\mathbf{\Gamma},\beta}(\mathbf{Y} \mid \mathbf{\Gamma}, \boldsymbol{\beta}) = \prod_{i=1}^{n} p_{\mathbf{y}_i|\Gamma_i,\beta}(\mathbf{Y}_i \mid \Gamma_i, \boldsymbol{\beta}). \quad (4)$$

Substituting in the appropriate probability densities, we can also write this as

$$p_{\mathbf{y}|\mathbf{\Gamma},\beta}(\mathbf{Y} \mid \mathbf{\Gamma}, \boldsymbol{\beta}) = \prod_{ij:\Gamma_i=m_j} N(\mathbf{M}_j\boldsymbol{\beta}, \mathbf{\Lambda}_1) \prod_{ij:\Gamma_i=b_j} N(\mathbf{b}_j, \mathbf{\Lambda}_0). \quad (5)$$

3 Hypothesis Testing

In this section, we discuss how we can use hypothesis testing to make a determination about whether the model object appears in a scene image.

We can define two hypotheses

H_0 : The object is not present in the image

H_1 : The object is present in the image and derive the decision rule using a binary hypothesis testing formulation. Hypothesis H_0 is equivalent to stating that all of the scene features are matched to background features, and hypothesis H_1 is equivalent to stating that at least one of the scene features is matched to the model.

3.1 Likelihood Ratio Test

Prior densities for each hypothesis are difficult to estimate in the case of object recognition since we do not know *a priori* the probability of the object appearing in the image. We therefore use the Neyman-Pearson criterion for hypothesis testing which is appropriate for problems without prior densities. The Neyman-Pearson criterion sets one of the conditional error probabilities to a fixed value α and then minimizes the other conditional error probability given this constraint. It can be shown that the optimal decision rule given these constraints is a likelihood ratio test of the form

$$\lambda(\mathbf{Y}) \triangleq \frac{p_{\mathbf{y}|H_0}(\mathbf{Y} \mid H_0)}{p_{\mathbf{y}|H_1}(\mathbf{Y} \mid H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \eta \qquad (6)$$

where η is the threshold of the decision rule [Trees, 1968].

3.2 Generalized Likelihood Ratio Test

The likelihood ratio test defined in Equation 6 can be used when the conditional density $p_{\mathbf{y}|H_i}(\mathbf{y} \mid H_i)$ for each hypothesis is completely known. Often however, the conditional density for a hypothesis contains several unknown random or nonrandom parameters. A hypothesis whose conditional density contains one or more unknown parameters is referred to as a *composite hypothesis*.

An extension to the likelihood ratio test called the *generalized likelihood ratio test* can be used to make decisions involving composite hypotheses. To derive the generalized likelihood ratio test, suppose that $\mathbf{y}$ is an n-dimensional random variable with conditional probability density $p_{\mathbf{y}|\theta}(\mathbf{Y} \mid \boldsymbol{\theta})$ conditioned on some r-dimensional parameter $\boldsymbol{\theta}$ whose values range over a parameter subspace $\Omega \subset R^r$. We can then partition the parameter space Ω into two subspaces Ω_0 and Ω_1 so that $\boldsymbol{\theta} \in \Omega_0$ for hypothesis H_0 and $\boldsymbol{\theta} \in \Omega_1$ for hypothesis H_1.

Ideally, if we knew the value of $\boldsymbol{\theta}$ then we could use this value directly in a likelihood ratio test. The error probability achieved by using the correct value of $\boldsymbol{\theta}$ is called the perfect measurement bound and represents a limit on the performance of any decision rule for composite hypotheses. Since we do not know the actual value of $\boldsymbol{\theta}$, we can approach the perfect measurement bound by estimating $\boldsymbol{\theta}$ given H_1 and given H_0 and then use these estimates in a likelihood ratio test as if they were correct. This procedure defines the generalized likelihood ratio test which we can write as

$$\lambda_g(\mathbf{Y}) \triangleq \frac{\max_{\theta_0 \in \Omega_0} p_{\mathbf{y}|\theta_0}(\mathbf{Y} \mid \theta_0)}{\max_{\theta_1 \in \Omega_1} p_{\mathbf{y}|\theta_1}(\mathbf{Y} \mid \theta_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \eta. \qquad (7)$$

If we let $\hat{\boldsymbol{\theta}}_0$ denote the maximum likelihood estimate of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_1$ denote the maximum likelihood estimate of $\boldsymbol{\theta}_1$ then we can write Equation 7 as

$$\lambda_g(\mathbf{Y}) \triangleq \frac{p_{\mathbf{y}|\theta_0}(\mathbf{Y} \mid \hat{\boldsymbol{\theta}}_0)}{p_{\mathbf{y}|\theta_1}(\mathbf{Y} \mid \hat{\boldsymbol{\theta}}_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \eta. \qquad (8)$$

We now have a procedure for testing composite hypotheses, however the issue of computing the threshold η remains a problem since we still need to evaluate

$$\alpha = \int_{-\infty}^{\eta} p_{\lambda_g|H_0}(\lambda_g \mid H_0) d\lambda_g. \qquad (9)$$

In practice, solving this integral equation for η is often nontrivial and we are currently investigating the use of an asymptotic approximation to set the threshold without explicitly solving Equation 9.

3.3 Application to Object Recognition

The conditional density of our observation vector given hypothesis H_0 is

$$\begin{aligned} p_{\mathbf{y}|H_0}(\mathbf{Y} \mid H_0) &= p_{\mathbf{y}|\boldsymbol{\Gamma}}(\mathbf{Y} \mid \boldsymbol{\Gamma} \in B^n) \\ &= \prod_{\Gamma_i \in B} N_0 \\ &= N_0^n \end{aligned}$$

where $\boldsymbol{\Gamma} \in B^n$ indicates that all of the n scene features are matched to background features and $N_0 = p_{\mathbf{y}_i|\Gamma_i}(\mathbf{Y}_i \mid \Gamma_i \in B)$. The conditional density of the observation vector given hypothesis H_1 is

$$\begin{aligned} p_{\mathbf{y}|H_1}(\mathbf{Y} \mid H_1) &= p_{\mathbf{y}|\boldsymbol{\Gamma},\beta}(\mathbf{Y} \mid \boldsymbol{\Gamma} \notin B^n, \boldsymbol{\beta}) \\ &= \prod_{ij:\Gamma_i = m_j} N(\mathbf{M}_j\boldsymbol{\beta}, \boldsymbol{\Lambda}_\mathbf{1}) \prod_{i:\Gamma_i \in B} N_0 \end{aligned}$$

where $\boldsymbol{\Gamma} \notin B^n$ indicates that not all of the n scene features are matched to the background. Hypothesis H_1 is clearly a composite hypothesis since the conditional probability contains unknown parameters for correspondence $\boldsymbol{\Gamma}$ and pose $\boldsymbol{\beta}$.

The generalized likelihood ratio test requires an estimate of pose and correspondences under hypothesis H_1. The estimation of pose and correspondence has already been described extensively in [Wells, 1992] and [Baker, 1994].

4 Multiresolution Processing

In this section, we discuss how we can combine hypothesis testing and pose estimation in a multiresolution formulation. A number of advantages can be gained by multiresolution processing including improvements in computational efficiency and robustness.

4.1 Multiresolution Approaches

Multiresolution techniques have been used for a wide range of computer vision and image processing applications including edge detection [Marr and Hildreth, 1980], image compression [Burt and Adelson, 1983], stereo matching [Grimson, 1985], optical flow computation [Luettgen *et al.*, 1993], and object recognition [Cantoni *et al.*, 1988; Marsic and Micheli-Tzanakou, 1992; Neveu *et al.*, 1986; Rosenfeld and Vanderbrug, 1977; Tretter and Bouman, To appear]. There is also significant evidence that human vision utilizes multiple resolutions in analyzing images. The reason for the widespread use of multiresolution processing is the fact that different scales or resolution levels in an image contain different amounts of information. We utilize this observation by modeling objects at low resolution by their coarse features and at higher resolution by the addition of finer detailed features.

4.2 Multiresolution Object Recognition Algorithm

The multiresolution statistical object recognition algorithm operates in a coarse-to-fine manner by computing the optimal pose at each resolution level and using that estimate as the starting value for the pose estimation at the next finer resolution level. Initial hypotheses are generated using an indexing table at the lowest resolution level and the pose is computed for each hypothesis. Each hypothesis is then evaluated by the weight computation step or E-step of the EM algorithm which also provides an estimate of correspondences. The pose and correspondence estimate is then used to compute the score of the hypothesis and the hypotheses with the N highest scores are propagated through each resolution level. At the maximum resolution level, the highest scoring hypothesis is evaluated with the generalized likelihood ratio test to determine whether the data indicates that the model actually appears at this pose. If the model is located in the image, additional occurrences of the model can be found by eliminating those scene features which were matched to the model and running the algorithm again. We can continue this procedure until the hypothesis test indicates that there is no additional occurrence of the model object in the image.

5 Recognition Experiments

This section describes two of the experiments that we have conducted with the multiresolution statistical object recognition algorithm. The algorithm was tested on two images: one in which the model object appears and one in which the model object does not appear.

5.1 Model Object Present

An experiment was conducted to evaluate the performance of the multiresolution statistical object recognition algorithm on a scene image containing the model object. The scene image which was used for this experiment is shown in Figure 1. The image is 640 by 480 pixels. The true pose of the model object in the scene image is the null pose $t_x = 0$, $t_y = 0$, $s = 1$, and $\theta = 0$. We used three levels of smoothing to generate the multiresolution object model. The resulting model consisted of 269 features at the highest resolution level, 172 features at medium resolution, and 102 features at low resolution. The multiresolution scene image contained 1157 features at the highest resolution level, 801 features at medium resolution, and 484 features at low resolution.

Figure 1: Scene Image

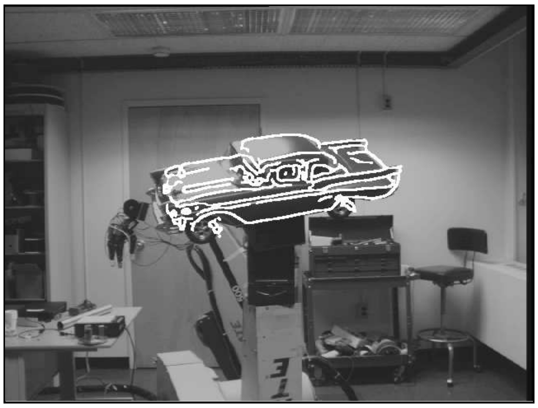

Figure 2: Model object present. Estimated pose: $t_x = 27.5$, $t_y = 4.55$, $s = 0.959$, $\theta = 0.009$. Score: $\lambda_g = 0.0131$.

The low resolution model features were used to construct an angle pair indexing table. The scene features at the lowest resolution level were used to index into the table and 2,199 hypotheses were generated. Of these hypotheses, twelve corresponded to correct poses. The top 100 scores were propagated through the next two levels of the resolution hierarchy. Of the twelve correct initial hypotheses, four of them were among the top 100 scores. At maximum resolution, the hypothesis with the highest score was the correct hypothesis shown in Figure 2.

The results of this experiment indicate that we can successfully discard many incorrect solutions at low resolution level and spend our computational effort verifying the most likely solutions at higher resolution levels.

Figure 3: Model object not present. Estimated pose: $t_x = 281.451$, $ty = 173.077$, $s = 0.381$, $\theta = 1.187$. Score: $\lambda_g = 0.0303$.

5.2 Model Object Not Present

This experiment was conducted on the scene image shown in Figure 3. The parameters for the multiresolution images are the same as those in the previous experiment The top 100 hypotheses were propagated to the maximum resolution level. Figure 3 shows the hypothesis with the highest score at maximum resolution level, however the magnitude of the score indicates that the model does not in fact appear in the image.

6 Future Work

Future areas of research include investigation of the use of an asymptotic approximation to the generalized likelihood ratio test for setting decision thresholds. Another topic concerns using the known statistical characteristics of the problem to determine the background feature variance.

References

[Baker, 1994] Jonathan D. Baker. Multiresolution statistical object recognition. Master's thesis, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, May 1994.

[Burt and Adelson, 1983] Peter J. Burt and Edward H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983.

[Cantoni *et al.*, 1988] V. Cantoni, L. Carrioli, M. Ferretti L. Lombardi, and K. Matthews. Object recognition in multi-resolution systems. In Anil K. Jain, editor, *Real-Time Object Measurement and Classification*, pages 359–374. Springer-Verlag, Berlin, 1988. Proceedings of the NATO Advanced Research Workshop on Real-Time Object and Environment Measurement and Classification, held in Maratea, Italy, August 31 – September 3, 1987.

[Grimson, 1985] W. Eric L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, 1985.

[Grimson, 1990] W. Eric L. Grimson. *Object Recognition by Computer*. The MIT Press, Cambridge, MA, 1990.

[Luettgen *et al.*, 1993] Mark R. Luettgen, W. Clem Karl, and Alan S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. LIDS 2175, Massachusetts Institute of Technology, Cambridge, April 1993.

[Marr and Hildreth, 1980] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London*, B207:187–217, 1980.

[Marsic and Micheli-Tzanakou, 1992] Ivan Marsic and Evangelia Micheli-Tzanakou. A framework for object representation and recognition. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 3, pages 272–277, Baltimore, MD, June 1992.

[Neveu *et al.*, 1986] Charles F. Neveu, Charles R. Dyer, and Roland T. Chin. Two-dimensional object recognition using multiresolution models. *Computer Vision, Graphics, and Image Processing*, 34(1):52–65, April 1986.

[Rosenfeld and Vanderbrug, 1977] Azriel Rosenfeld and Gordon J. Vanderbrug. Coarse-fine template matching. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 104–107, February 1977.

[Trees, 1968] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory*. John Wiley and Sons, Inc., New York, 1968.

[Tretter and Bouman, To appear] Daniel Tretter and Charles A. Bouman. Multiscale stochastic approach to object detection. In *Proceedings of the SPIE Visual Communications and Image Processing*, To appear.

[Ullman and Basri, 1989] Shimon Ullman and Ronen Basri. Recognition by linear combinations of models. A.I. Memo 1152, Massachusetts Institute of Technology, August 1989.

[Wells, 1990] William M. Wells, III. A statistical approach to model matching. In *Proceedings of SPIE – Intelligent Robots and Computer Vision IX: Algorithms and Techniques*, pages 22–29, Boston, MA, November 1990.

[Wells, 1991] William M. Wells, III. MAP model matching. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 486–492, Maui, Hawaii, June 1991.

[Wells, 1992] William M. Wells, III. *Statistical Object Recognition*. Ph.D. dissertation, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, November 1992.

[Wells, 1993] William M. Wells, III. Statistical object recognition with the Expectation-Maximization Algorithm in range-derived features. In *Proceedings of the DARPA Image Understanding Workshop*, pages 839–850, Washington, DC, April 1993.