STATISTICAL MODELING OF RELATIONS FOR 3–D OBJECT RECOGNITION

Joachim Hornegger

Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen–Nürnberg, Martensstr. 3, D–91058 Erlangen, Germany
email: hornegger@informatik.uni-erlangen.de, www: http://www5.informatik.uni-erlangen.de

ABSTRACT

A new Bayesian framework for 3–D object classification and localization is introduced. Objects are represented as probability density functions, and observed features are treated as random variables. These probability density functions turn out a non geometric nature of models and characterize the statistical behavior of local object features like points or lines. The parameterization of model densities covers several terms of object recognition: locations and instabilities of features, rotation and translation, projection, the assignment of image and model features, as well as relations. This paper treats especially the probabilistic modeling of relational dependencies between single features. The mathematical framework, the training algorithms, as well as the localization and classification modules are discussed in detail. The experimental evaluation shows the usefulness of the introduced concepts on real image data.

1. INTRODUCTION

Probabilistic methods are state of the art in several areas of signal processing and analysis. Speech recognition systems, for instance, apply hidden Markov models at different levels of abstraction [4]. The application of statistical approaches for image analysis purposes shows a considerable and still increasing interest.

The central problems of statistical object recognition are: given an image; which objects appear most likely in the scene, and what are the most probable pose parameters with respect to a reference coordinate system. The motivation for a statistical setting to solve these tasks is manysided. There are both theoretical and practical reasons for probabilistic object recognition systems. Sensor data are noisy and influenced by lighting, occlusion, segmentation errors, clutter, or heterogeneous background, therefore a probabilistic framework seems natural. A statistical approach is theoretically motivated by the optimality of Bayesian classifiers with respect to misclassifications and allows the use of results and well studied methods of mathematical estimation theory.

A central problem within the statistical framework is to find a suitable probability density function which is capable to describe complex objects including all degrees of freedom. The presented approach is based on simple features computed within a segmentation step. These primitives should include the characteristics of objects which are necessary to enable the identification and localization. The consideration of relational dependencies between these features is expected to improve the discriminating power of the used primitives. Figure 1, for instance, shows a geometrical model and two sets of image features for fixed pose parameters and varied lighting conditions. A suitable binary relation between 2–D point features is given by the fact that two points might be connected with an edge or not. Due to inevitable segmentation errors this relation also demonstrates a probabilistic behavior within the observations. In the following sections it is shown how relational dependencies can be considered as aleatory variables and structural dependencies of features are estimated by 2–D observations.

The paper is divided up into eight sections. The discussion of related work (Section 2) is followed by Section 3 which treats

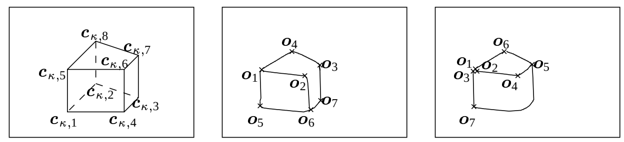

Figure 1. 3–D model (left) and 2–D segmentation results of views with varying illumination

mathematical details for statistical object modeling. The training algorithms for model densities are derived in Section 4. There, the Expectation Maximization algorithm (EM algorithm) is applied for solving the incomplete data estimation problems induced by the available 2–D training views. The usage of probabilistic object models for object localization and recognition is shown in Sections 5 and 6. In order to demonstrate the practical use of the introduced probabilistic framework, experimental results are described in Section 6. The paper concludes with a summary of the presented statistical approach and some hints for future research.

2. RELATED WORK

This work applies parameterized density functions for object modeling. Probabilistic approaches to object recognition can be distinguished with respect to the chosen statistical models and the used features. Shimshoni [5], for example, applies line features and corresponding ratios and angles for building up discrete probability measures. These are suitable for the identification and the localization of 3–D objects within gray–level images using the probabilistic peaking effect. In contrast, in [7] parametric mixture densities are used for solving 2–D object localization and classification problems. The features used in this case are restricted to oriented point features.

The methods mentioned so far are based on segmentation results and model the statistical behavior of segmented features, like straight line elements or points. The pose estimation algorithm introduced in [6], however, shows that the localization of complex objects can be done using gray–level features directly, if a 3–D model is available. This method applies the maximization of mutual information for the determination of position and orientation. Nevertheless, the subsequent theory is also based on segmentation results, since this technique was already successfully applied in [2].

3. PROBABILISTIC MODELING OF OBJECTS

In contrast to geometric models, the probabilistic description of objects has to characterize the statistical behavior of features and their relations within the image space. Since the appearance of image features might vary with the object's pose, the models should be parameterized with respect to these degrees of freedom, too. To provide a complete probabilistic setting, also the assignment of features and relational dependencies require a statistical representation as well. The usage of heuristics, for example, to solve the matching problem might work, but violates the principles of the proposed statistical framework. The probabilistic modeling of assignments, relations, and features results in model densities, which are composed density functions. The required components are introduced in the following subsections.

3.1. Mathematical Notation

The object recognition formalism has to distinguish between the D_m–dimensional model and the D_o–dimensional image space, where $D_o \le D_m$. If we deal with 3–D object recognition using 2–D sensor data, for example, we set $D_m = 3$ and $D_o = 2$. The object classes considered within the recognition experiments are represented by $\Omega_1, \Omega_2, \ldots, \Omega_K$. Since the classification of objects is based on features, we denote the n_κ model features corresponding to object class Ω_κ by $c_{\kappa,1}, c_{\kappa,2}, \ldots, c_{\kappa,n_\kappa}$. The observable features of the D_o–dimensional image space are denoted by the set $O = \{o_1, o_2, \ldots, o_m\}$. In the above example, $c_{\kappa,l}$ $(1 \le l \le n_\kappa)$ might be 3–D points and all elements of O would represent 2–D image points. The statistical parameters characterizing model features are defined by $a_{\kappa,1}, a_{\kappa,2}, \ldots, a_{\kappa,n}$. For instance, if normally distributed 3–D model features are assumed, $a_{\kappa,l}$ denotes the mean vector $\mu_{\kappa,l} \in \mathbb{R}^{D_m}$ and the covariance matrix $\Sigma_{\kappa,l} \in \mathbb{R}^{D_m \times D_m}$. The set of all parameters necessary for describing the object class Ω_κ is called B_κ. In addition to feature–specific parameters, we have to consider pose parameters, too. Primitives of the model space can be rotated, translated and projected into the image plane. This mapping is characterized by the affine transform given by $R \in \mathbb{R}^{D_o \times D_m}$ and $t \in \mathbb{R}^{D_o}$. In the following discussion, pose parameters are thus defined by R and t.

3.2. Statistical Modeling of Single Features

We start with considering the statistical modeling of single features. Assuming that the probabilistic behavior of model feature $c_{\kappa,l}$ is characterized by the parameterized density $p(c_{\kappa,l}|a_{\kappa,l})$, the integration of the affine transform, which maps model onto image features, can be done by a standard density transform.

Let us assume we have normally distributed 3–D model features and an affine transform from the model into the image space. The observable 2–D image features o_k would also be normally distributed with mean vectors $R\mu_{\kappa,l_k} + t$ and covariance matrices $R\Sigma_{\kappa,l_k}R^T$.

If the corresponding model and image features are given by the sequence of pairs $[(k, l_k)]_{1\le k\le m}$ and the features are pairwise statistically independent, the density for a set of observe features $O = \{o_1, o_2, \ldots, o_m\}$ is defined by

$$p(O|[(k,l_k)]_{1\le k\le m}, a_{\kappa,1}, \ldots, a_{\kappa,n}) = \prod_{k=1}^{m} p(o_k|a_{\kappa,l_k}, R, t) \quad (1)$$

The independency assumption, of course, is a simplification. In general, image features depend on each other. This dependency structure will be embedded in model densities using the statistical modeling of relations.

3.3. Statistical Modeling of Assignments

The corresponding image and model indices (k, l_k) are defined by the assignment function ζ_κ and are expected to be known in (1). The probabilistic description of the assignment function is based on a discrete statistical model. The discrete function

$$\zeta_\kappa : \left\{ \begin{array}{ccc} O & \to & \{1, \ldots, n_\kappa\} \\ o_k & \mapsto & l_k \end{array} \right. , \; k = 1, 2, \ldots, m \,. \quad (2)$$

induces a discrete random vector $\zeta_\kappa = (\zeta_\kappa(o_1),\ \zeta_\kappa(o_2),\ \ldots,\ \zeta_\kappa(o_m))^T \in \{1, \ldots, n_\kappa\}^m$. With each random vector, a discrete probability $p(\zeta_\kappa)$ can be associated, where the discrete probabilities sum up to one, i.e. $\sum_{\zeta_\kappa} p(\zeta_\kappa) = 1$. If, for instance, the assignments of image features are pairwise statistically independent, the factorization

$$p(\zeta_\kappa) = \prod_{k=1}^{m} p(\zeta(o_k)) \quad (3)$$

is possible. The probability of the random vector is thus given by the product of its components' probabilities.

3.4. Statistical Modeling of Relations

Relations can be defined by indicator functions. Let us assume a q–nary relation. For an observed q–tuple $(o_{k_1}, \ldots, o_{k_q})$ of observed features, we can compute the binary value

$$\chi(o_{k_1}, \ldots, o_{k_q}) = \left\{ \begin{array}{lll} 1 & , & \text{if the relation is satisfied} \\ 0 & , & \text{otherwise} \end{array} \right. . \quad (4)$$

This indicator function induces a q–dimensional binary array, which can be considered as an aleatory variable. Thus, like the assignment function, a discrete probability is associated with each array $\nu = \left(\chi(o_{k_1}, \ldots, o_{k_q})\right)_{1\le k_1,\ldots,k_q\le m}$. It is assumed that the observed relations depend on relations in the model space. Thus relations rely upon the assignment between image and model features. This statistical dependency is expressed within the likelihood $p(\nu|\zeta_\kappa)$, which is the discrete probability that the array ν is observed, if the assignment function ζ_κ is assumed.

Let us consider the neighborhood relationship, where two points are neighbors, if they are connected by a line. The binary matrix $\nu \in \mathbb{R}^{m\times m}$ for the scene shown in Figure 1 (middle) is

$$\left(\nu_{k',k''}\right)_{1\le k',k''\le 7} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} . \quad (5)$$

The segmentation results allow the observation of triples $(o_{k'}, o_{k''}, \nu_{k',k''})$, where $\nu_{k',k''} \in \{0,1\}$ is the Boolean value which indicates whether features $o_{k'}$ and $o_{k''}$ fulfill the considered relation in the image space or not. If statistically independent relations and assignments are assumed, the conditional discrete probability $p(\nu|\zeta_\kappa)$ for observing a binary matrix ν is thus given by

$$p(\nu|\zeta_\kappa) = \prod_{k',k''=1}^{m} p(\nu_{k',k''}|l_{k'}, l_{k''}) \quad , \quad (6)$$

where $l_{k'} := \zeta_\kappa(o_{k'})$ and $l_{k''} := \zeta_\kappa(o_{k''})$.

3.5. Model Densities

So far we have discussed the statistical modeling of features, assignments, and relations. Now we combine all these statistical components to describe a compound density for the characterization of object features within the image space.

The compound density for observing a $(q+1)$–tuple $(o_{k_1}, o_{k_2}, \ldots, o_{k_q}, \nu_{k_1,k_2,\ldots,k_q})$ is

$$\begin{aligned} p((o_{k_1}, o_{k_2}, \ldots, o_{k_q}, \nu_{k_1,k_2,\ldots,k_q})|\zeta_\kappa, B_\kappa, R, t) &= \\ p(\zeta_\kappa) p(\nu_{k_1,k_2,\ldots,k_q}|\zeta_\kappa) \prod_{s=1}^{q} p(o_{k_s}|a_{\kappa,s}, R, t) & \quad . \quad (7) \end{aligned}$$

The non–observable assignment function can be eliminated by marginalization, due to its statistical appearance and the presence of random vector ζ_κ. For simplicity and computational efforts, statistically independent assignments and relations are assumed.

The model density for an observed set of features O and the binary matrix ν — without knowing the assignment — results in

$$p(O,\nu|B_\kappa,R,t) \quad =$$
$$= p(\{(o_{k'},o_{k''},\nu_{k',k''})|1\le k',k''\le m\}|B_\kappa,R,t)$$
$$= \prod_{k',k''=1}^{m}\sum_{l',l''=1}^{n_\kappa} p(\zeta_\kappa(o_{k'})=l')p(\zeta_\kappa(o_{k''})=l'')p(\nu_{k',k''}|l',l'')$$
$$p(o_{k'}|a_{\kappa,l'},R,t)p(o_{k''}|a_{\kappa,l''},R,t) \qquad (8)$$

The computational complexity for evaluating the model density (8) for a given observation is obviously bounded by $\mathcal{O}(n_\kappa^2\, m^2)$.

Within this model density the parameter set B_κ includes the discrete probabilities $p(\zeta_\kappa)$, $p(\nu|\zeta_\kappa)$, and the feature–specific parameters $a_{\kappa,1}, a_{\kappa,2},\ldots, a_{\kappa,n_\kappa}$. The model generation process is now expected to determine the parameter set B_κ out of a set of observed images.

4. MODEL GENERATION FROM PROJECTED DATA

The estimation of model parameters will use the rotation and translation as well as the computed features and their relations of each 2–D training view. If N training views are available, the training data are represented by the set $\{{}^\varrho O, {}^\varrho R, {}^\varrho t|1\le\varrho\le N\}$. Since we learn the parameters for model densities of 3–D objects, the training data set is latent with respect to the range information. In addition to depth data, the assignment function ζ_κ is not part of the observation, too. The model generation process has thus to work on incomplete training data.

4.1. Incomplete Data Estimation

There exist several different methods which deal with incomplete data estimation problems. The most common used techniques apply the Expectation Maximization algorithm (EM algorithm, [1]).

Assuming that the observable random variables are denoted by X and the hidden variables by Y, we can define the parametric densities $p(X,Y|B)$, $p(Y|X,B)$, and $p(X|Y,B)$, where B represents the density parameters. The iterations of the EM algorithm expect an initialization $\widehat{B}^{(0)}$ and iteratively maximize the Kullback Leibler statistics

$$Q(\widehat{B}^{(i+1)}|\widehat{B}^{(i)}) = \int p(Y|X,\widehat{B}^{(i)})\log p(X,Y|\widehat{B}^{(i+1)})dY\,, (9)$$

which is the conditional expectation of the log–likelihood function $\log p(X,Y|\widehat{B}^{(i+1)})$. The iterations terminate, if a stationary point is reached. If multiple observations $\{{}^\varrho X|1\le\varrho\le N\}$ are available, the sum of single Kullback–Leibler statistics ${}^\varrho Q(\widehat{B}^{(i+1)}|\widehat{B}^{(i)})$ for the observation ${}^\varrho X$ has to be maximized:

$$Q(\widehat{B}^{(i+1)}|\widehat{B}^{(i)}) \quad = \quad \sum_{\varrho=1}^{N}{}^\varrho Q(\widehat{B}^{(i+1)}|\widehat{B}^{(i)}) \quad . \qquad (10)$$

The iterative optimization of the Q–function is usually done by the application of gradient methods. The convergence properties and further theoretical results on this general iterative scheme can be found in [1].

4.2. Estimation of Model Parameters

The model parameters B_κ characterizing object class Ω_κ have to be estimated using incomplete data and the EM algorithm. We assume statistically independent assignments and relations. Here, the relational dependencies are restricted to binary relations, a generalization is straight forward. The considered primitives are point features, which are assumed to be normally distributed.

4.2.1. Kullback Leibler Statistics

The computation of training formulas first requires the Kullback Leibler statistics for the model density (8) defined in Section 3. The non–observable part is restricted to the missing assignment. For one observed triple $\tau(\varrho,k',k'')=({}^\varrho o_{k'},{}^\varrho o_{k''},{}^\varrho\nu_{k',k''})$ we get the Q–function [2]

$${}^\varrho Q_{k',k''}(\widehat{B}_\kappa^{(i+1)}|\widehat{B}_\kappa^{(i)}) \quad =$$
$$= \sum_{l',l''=1}^{n_\kappa} p(l',l''|\tau(\varrho,k',k''),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t)$$
$$\log p(\tau(\varrho,k',k''),l',l''|\widehat{B}_\kappa^{(i+1)},{}^\varrho R,{}^\varrho t) \quad , \qquad (11)$$

in accordance with (9), where l' and l'' represent indices of model features. If we sum up this Q–function with respect to ϱ, k', and k'' (c.f.. (10)), we get the Kullback Leibler statistics required for the EM iterations.

4.2.2. Training of Assignments

The estimation formula for the discrete probabilities $p_{\kappa,l} := p(\zeta_\kappa(o_k)=l)$ $(1\le l\le n_\kappa)$ results from computing the partial derivatives of the Kullback Leibler statistics, where the probability constraint is enforced by Lagrange multipliers. We define $h := 2\cdot\sum_{\varrho=1}^{N}{}^\varrho m$, where ${}^\varrho m$ denotes the number of image features of the ϱ–th training view, and get the closed form reestimation formula [2]

$$\widehat{p}_{\kappa,l}^{(i+1)} \;=\; \frac{1}{h}\sum_{\varrho,k',k'',l'} p(l',l|\tau(\varrho,k',k''),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t) \quad . (12)$$

4.2.3. Training of Relations

A similar derivation gives us the training formulas for relational dependencies of single features. The term $p_\kappa(v|l',l'')$, where $v\in\{0,1\}$, and $1\le l',l''\le n_\kappa$, measures the discrete probability that two model features given by the indices l' and l'' satisfy the relation or not, which is observed within the image space. Since the assignment of image and model features is not known, the discrete probabilities can be estimated iteratively using

$$\widehat{p}_\kappa^{(i+1)}(v|l',l'') = \frac{\displaystyle\sum_{\substack{\varrho,k',k''\\ {}^\varrho\nu_{k',k''}=v}} p(l',l''|\tau(\varrho,k',k''),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t)}{\displaystyle\sum_{\varrho,v,k',k''} p(l',l''|({}^\varrho o_{k'},{}^\varrho o_{k''},v),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t)}.$$

4.2.4. Training of Mean Vectors

The parameters of the normally distributed 3–D point features, which are considered as continuous random variables, can also be computed using above Kullback Leibler statistics. The mean vectors $\mu_{\kappa,l}$ are estimated using a closed form iterative scheme, which can be derived from the Kullback Leibler statistics (see [2] for details). The result is

$$\widehat{\mu}_{\kappa,l}^{(i+1)} = \left(\sum_{\varrho,k',k'',l'} p(l',l|\tau(\varrho,k',k''),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t){}^\varrho R^T\left({}^\varrho D_{\kappa,l}\right)^{-1}{}^\varrho R\right)^{-1}$$
$$\sum_{\varrho,k',k'',l'} p(l',l|\tau(\varrho,k',k''),\widehat{B}_\kappa^{(i)},{}^\varrho R,{}^\varrho t){}^\varrho R^T\left({}^\varrho D_{\kappa,l}\right)^{-1}\left({}^\varrho o_{k'}-{}^\varrho t\right),$$

where ${}^\varrho D = R\Sigma_{\kappa,l}R^T$. If the covariance matrix is also unknown, an iterative optimization scheme is required for the maximization of the Q–function [2].

The EM algorithm is a local optimization method. Thus the initialization of the unknown parameters is crucial for its success. In our experiments, we assume uniformly distributed discrete probabilities; the mean vectors were initialized using a single 2–D reference view, and the depth components of means were set to zero.

5. STATISTICAL LOCALIZATION

The automatically generated model densities are now applied to localize and classify objects. Since the models are parameterized densities with respect to pose parameters, the localization of objects corresponds to a parameter estimation problem. Given a set of observed features O and the indicator matrix ν, the pose parameters result from the maximum likelihood estimation

$$\{\widehat{R}, \widehat{t}\} \quad = \quad \underset{R,t}{\operatorname{argmax}}\, p(O, \nu | B_\kappa, R, t) \quad . \tag{13}$$

For solving this global optimization problem, we use the adaptive random search technique based on a mixture of Gaussian and start local maximization applying the Downhill Simplex algorithm [2].

6. STATISTICAL CLASSIFICATION

The estimated pose parameters R and t allow the computation of a posteriori probabilities. For an object of class Ω_κ, the a posteriori probability for a set of observed features is

$$p(\Omega_\kappa | O, \nu) \quad = \quad \frac{p(\Omega_\kappa) p(O, \nu | B_\kappa, R, t)}{p(O, \nu)} \quad . \tag{14}$$

Thus, the classification module can apply the Bayesian decision rule:

$$\lambda \quad = \quad \underset{\kappa}{\operatorname{argmax}}\, p(\Omega_\kappa | O, \nu) \quad , \tag{15}$$

which decides for the object class with the highest a posteriori probability and minimizes the probability for misclassifications.

7. EXPERIMENTAL RESULTS

The introduced statistical framework for object modeling, localization, and classification has been tested on real image data. The tested relation between features is the neighborhood relationship. Two point features are defined to be neighbors, if they were connected by a straight line element. The images for parameter estimation were captured by a camera which is mounted on a calibrated robot's hand, i.e. for each training view the pose parameters of the object to be learned are known.

In a first series of experiments we tested the system using four 2–D objects. The model densities were trained using 300 images of each object. The test set consisted of 1000 images showing different objects with homogeneous background features. The recognition rate was 93% for both point and line features. We also considered 100 scenes which include one known object and unknown background objects (c.f. Figure 2). The rate of correct localizations for these fairly complex scenes increased from 10% to 30%.

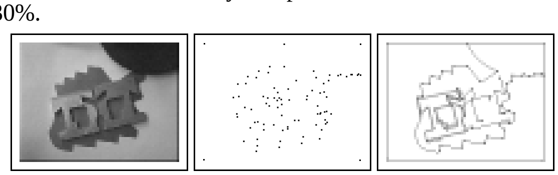

Figure 2. 2–D localization results (right) using point features (middle)

The 3–D experiments considered also four different object classes of simple, but quite similar polyhedral 3–D objects. During the training stage 400 views of each object were used. It is expected that the relations improve the localization and recognition rates. Indeed, for multiple object scenes with heterogeneous background point features are less discriminating than line features, which take relations into consideration. Figure 3 shows an example for a scene, which includes one known object. The others are considered as background features. The point features result in wrong pose parameters, whereas the correct object position could be computed using straight line features defined by binary relations of points. Tests using 100 complex scenes which include one known object with heterogeneous background features, showed an increase of correct localizations from 15% to 24%, if neighborhood relations instead of 2–D point features were used. Within the classification experiments of single objects we tested 1600 images. The recognition rate using point features was 68% and decreased to 59% if relations were taken into account. The reason for this unexpected result is the instability of line features. If a line is divided up into two small lines, the relation is no longer satisfied.

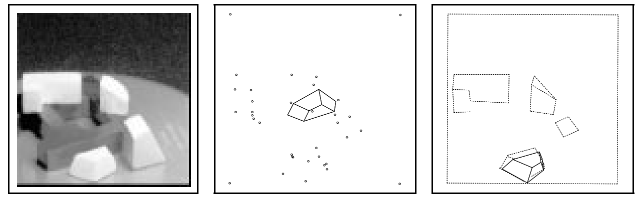

Figure 3. 3–D localization results using point (middle) and line features (right)

8. SUMMARY AND CONCLUSIONS

In contrast to classical geometric approaches for object recognition, we have shown that statistical methods are also suitable to deal with object identification and localization problems. Model generation as well as pose computations correspond to parameter estimation problems, whereas the classification is based on the Bayesian decision rule. The introduced framework allows statistical characterization of the assignment function between model and image features as well as the modeling of relations using discrete random vectors or arrays. Due to the fact that the assignment of features is neither observable during the training nor recognition stage, we have to deal with incomplete data estimation problems. Here, the EM algorithm was used to derive iterative model generation algorithms. Instead, the pose estimation was done applying adaptive random search techniques.

Future research should concentrate on the following topics: The pose computation corresponds to a parameter estimation problem, the use of multiple views will increase the size of sample data and thus the reliability of the estimated parameters. Since the quality of classifiers crucially depends upon the used features, more complicated features than point or line features should be considered within the statistical modeling.

REFERENCES

[1] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[2] J. Hornegger. *Statistische Modellierung, Klassifikation und Lokalisation von Objekten*. Shaker, Aachen, 1996.

[3] *Proceedings of the* 5^{th} *International Conference on Computer Vision (ICCV)*, Boston, June 1995. IEEE Computer Society Press.

[4] L. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[5] I. Shimshoni and J. Ponce. Probabilistic 3D object recognition. In ICCV 95 [3], pages 488–493.

[6] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. In ICCV 95 [3], pages 16–23.

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