PERGAMON

Pattern Recognition 35 (2002) 1225–1235

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Localization and classification based on projections

Joachim Hornegger[a,*,1], Volkmar Welker[b,2], Heinrich Niemann[c,3]

[a] *Siemens Medical Solutions, AXE 1, Siemensstraße 1, D-91301 Forchheim, Germany*
[b] *Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, D-35032 Marburg, Germany*
[c] *Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, Martensstr. 3, D-91058 Erlangen, Germany*

Received 13 October 2000; accepted 4 June 2001

Abstract

Due to the loss of range information, projections as input data for a 3-D object recognition algorithm are expected to increase the computational complexity. In this work, however, we demonstrate that this deficiency carries potential for complexity reduction of major vision problems. We show that projections provide a reduction of feature dimensions, and lead to structures exhibiting simple combinatorial properties. The theoretical framework is embedded in a probabilistic setting which deals with uncertainties and variations of observed features. In statistics marginal densities and the assumption of independency prove to be the key tools when one encounters projections. The examples discussed in this paper include feature matching, pose estimation as well as classification of 3-D objects. The final experimental evaluation demonstrates the practical importance of the marginalization concept and independency assumptions.

Keywords: Statistical object recognition; Pose estimation; Matching; Marginal densities

1. Introduction

It is well known that the recognition of 3-D objects based on gray-level images as input data is computationally hard. In practice, several obstructions avert an exaxt solution of the problem by simple and efficient algorithms, e.g. sensor data are noisy projections of real world, objects may be rotated and translated, self occlusion occurs. Thus feasible algorithms have to be based on simplifying and idealizing assumptions. For example, feature selection algorithms are applied and mutual statistical independency of features is assumed in order to decrease the dimension of the parameter space [1], segmentation downsizes the input data [2]. On the other hand the representation of data in higher dimensions, e.g. representation of points and projections maps in homogeneous coordinates, can also lower the computational complexity [3, Chapters 2 and 10]. However, for automatic model generation the absence of range data obstructs the discriminating power of features. As a consequence, most computer vision systems that work with gray-level images concentrate on the computationally expensive 3-D reconstruction. These facts show that, while both projections and reconstruction are applied for pattern recognition purposes, usually projections are seen as a major source for time consuming computations.

Nevertheless, this paper demonstrates that for object recognition and pose estimation purposes the use of

* Corresponding author. Tel.: +49-9191-189572; fax: +49-9191-189523.
E-mail addresses: joachim@hornegger.de (J. Hornegger), welker@mathematik.uni-marburg.de (V. Welker), niemann@informatik.uni-erlangen.de (H. Niemann).

[1] Supported by Deutsche Forschungsgemeinschaft (DFG) through a "Forschungsstipendium" (grant Ho 1791/2-1) while visiting Stanford University.
[2] Supported by Deutsche Forschungsgemeinschaft (DFG), "Heisenberg Stipendium".
[3] Supported by Deutsche Forschungsgemeinschaft (DFG), SFB 603, TP B2.

0031-3203/02/$22.00
PII: S0031-3203(01)00122-4

projections can be advantageous and can lead to more efficient algorithms. The key ideas are:

- the projection of 2-D image features on coordinate axes reduces the dimension of the search space for pose parameters [4], and
- 2-D projections have to match properly on the 1-D intersection of the projection planes [5].

Geometry tells us that there are pairs of different 3-D objects that cannot be distinguished by their 2-D projections. Therefore, 1-D projections are even more likely to cause the occurrence of ambiguous features that make classification impossible. Obviously, there exists a trade off between efficiency and the discriminating power of features. This important issue will be addressed in the experimental part in detail. We also argue that the use of real image data makes the application of statistical methods indispensable. In this setting, objects are not given by their exact shape in space but rather modeled by probability density functions which give a probability measure on the space of object positions.

In the theoretical part we introduce a method which combines tools from discrete and computational geometry with statistics. The projection of features is formalized in a probabilistic framework using marginal densities. We also show that for *generic* 3-D objects we can efficiently decide from 1-D projection data, whether two 2-D projections belong to the same set of 3-D points.

The paper is divided into six parts: The introduction is followed by a summary of basic facts about projections and marginals, including a discussion of advantages and disadvantages of using marginal densities. Section 3 briefly introduces the basic ideas of statistical object modeling, and the fundamental problems related to pose estimation and object recognition. The description of methods for localization and identification based on projections takes up most of Section 4. The paper closes with a discussion of experimental data, conclusions derived from the results and algorithms, and motivation and directions for further research.

2. Marginal densities

Seen as objects from geometry, projections map a vector lying in some high-dimensional space to its *shadow* in a low-dimensional subspace. In a probabilistic setting vectors are considered as random vectors. Their statistical behavior is characterized by a density function, and thus projections induce density transforms. In general, for bijective mappings of random variables the transform is easily computed [6, pp. 128–132]. Since a projection is bijective if and only if it is the identity, we have to resort to marginalization which allows the reduction of dimensions. The concept of marginalization is defined via integration. In all applications of this approach the dimension n of random vectors will be 2 or 3, and $(X_1,\dots,X_n)^{\mathrm{T}}$ will describe a point in $\mathbb{R}^n$. In terms of generality we state the method for n. Assume we have a parametric density function $p(X_1,\dots,X_n;B)$ for the random vector $X=(X_1,\dots,X_n)^{\mathrm{T}}$, where B denotes the set of parameters. Let us consider the projection

$$\pi_j:\begin{cases}\mathbb{R}^n & \to \mathbb{R}^{n-1}\\ (X_1,\dots,X_n) & \mapsto (X_1,\dots,X_{j-1},X_{j+1},\dots,X_n)\end{cases}. \quad (1)$$

The density of the projected random vector results from marginalization. An arbitrary random variable X_j can be eliminated by considering the integral over X_j, i.e.,

$$p(X_1,\dots,X_{j-1},X_{j+1},\dots,X_n;B) = \int p(X_1,\dots,X_n;B)\,\mathrm{d}X_j. \quad (2)$$

Example. Let the random vector $X=(X_1,\dots,X_n)^{\mathrm{T}}$ be normally distributed. Thus the parameter set B includes the mean vector $\boldsymbol{\mu}\in\mathbb{R}^n$ and the $(n\times n)$-covariance matrix $\boldsymbol{\Sigma}$. A projection can be defined, for instance, by an affine transform. The matrix $R\in\mathbb{R}^{m\times n}$ $(m\leqslant n)$ and the vector $t\in\mathbb{R}^m$ map the original n-dimensional random vector to the m-dimensional vector $Y=(Y_1,Y_2,\dots,Y_m)^{\mathrm{T}}$, where $Y=RX+t$. This vector is again normally distributed with mean vector $R\boldsymbol{\mu}+t\in\mathbb{R}^m$ and covariance matrix $R\boldsymbol{\Sigma}R^{\mathrm{T}}\in\mathbb{R}^{m\times m}$ [7, pp. 27,28].

Potential applications of marginal densities are, for instance:

- If the observable features are incomplete, the statistical behavior of the available features is characterized by the marginal density where the missing components are eliminated by integration.
- Marginalization reduces the number of parameters of the density function and thus leads to parameter estimation problems in lower dimensional spaces and thus of lower complexity.

An important drawback of marginals is related to the fact that projections reduce the discriminating power of the features. Another disadvantage is due to the fact that it is a major problem to find probability density functions which describe the statistics of features in image space properly. In the next section we discuss this issue and propose a statistical modeling scheme for objects and their appearance in the image plane.

3. Statistical object and scene modeling

The following discussion is restricted to point features which are computed automatically using a standard

segmentation algorithm [8]. A gray-level image f is transformed into a set of 2-D points $O=\{o_1,o_2,\ldots,o_m\}$ where $o_k\in\mathbb{R}^2$. These points are the input data of subsequent recognition and pose estimation algorithms. The appearance and the position of point features in the image plane depend on illumination conditions, sensor noise, the selected viewing direction, and the chosen point detector. For that reason, 2-D point features show a probabilistic behavior, and are therefore considered as 2-D random vectors. The application of the Kolmogorov–Smirnov test proves that point features are approximately normally distributed [9], and we associate a Gaussian probability density function with each 2-D point feature. Mean vectors and covariance matrices, however, are not sufficient to characterize the set of observable point features. The appearance of points varies with the objects' pose and depend on self-occlusion. Additional parameters are required which incorporate pose parameters of the object with respect to a reference world coordinate system. Another parameter set has to describe the probability of point appearance.

We observe 2-D point features in the image plane that are projections of 3-D points. Here rotation, translation, and projection are defined by an affine transform given by a rotation matrix $R\in\mathbb{R}^{2\times3}$ and a translation vector $t=(t_1,t_2,t_3)^{\mathrm{T}}\in\mathbb{R}^2$. The 3-D rotation has three degrees of freedom: the rotation angles φ_x, φ_y and φ_z around the x, y and z-axis of the world coordinate system. If we postulate that the 3-D point features are normally distributed, the 2-D projections are also Gaussian (mean vector and covariance matrix are given by the formulas from Section 2). If the pose parameters—implicitly given by R and t—and the corresponding normal distribution of an observed point feature are known, a density value can be computed. Let us assume that $C=\{c_1,c_2,\ldots,c_n\}$ denotes the set of 3-D point features $c_l\in\mathbb{R}^3$ corresponding to an object. We assume that these features are statistically characterized by $p(c;a_l)$, where a_l is the parameter associated with the lth model feature. Thus the rotated, translated, and projected point $o\in\mathbb{R}^2$ has the augmented probability density function $p(o;a_l,R,t)$.

It is the basic hypothesis of our statistical modeling scheme that 3-D point features are mutually statistically independent and normally distributed. Note that we are considering rigid objects and therefore the mean value of the position of each point is determined by its relative position to the other points. Nevertheless, this fact does not obstruct our independency assumption. Our model resembles the situation in solid state physics, where the positions of the atoms are assumed to obey independent probability distributions whose mean values form a rigid lattice. Instead of a rigid lattice, we use the original 3-D structure of the considered object. If the corresponding 3-D model and 2-D image features are given by the sequence of pairs $[(k,l_k)]_{1\leqslant k\leqslant m}$ where l_k denotes the index of the corresponding 3-D point c_{l_k} and the features are mutually independent, the density for a set of observed 2-D points $O=\{o_1,o_2,\ldots,o_m\}$ is defined by

$$p(O\,|\,[(k,l_k)]_{1\leqslant k\leqslant m};a_1,\ldots,a_n,R,t)=\prod_{k=1}^{m}p(o_k;a_{l_k},R,t). \tag{3}$$

The corresponding image and model indices (k,l_k) are formally given by the assignment function ζ which assigns features o_k to the index of the corresponding 3-D feature c_{l_k}. The description of the assignment function in terms of probabilities is based on a discrete statistical modeling scheme introduced in Refs. [9,4]. The discrete mapping

$$\zeta:\begin{cases}O\ \to\ \{1,\ldots,n\}\\ o_k\ \mapsto\ l_k,\ k=1,2,\ldots,m.\end{cases} \tag{4}$$

induces a discrete random vector $\boldsymbol{\zeta}=(\zeta(o_1),\zeta(o_2),\ldots,\zeta(o_m))^{\mathrm{T}}\in\{1,\ldots,n\}^m$. With each random vector, a probability $p(\boldsymbol{\zeta})$ can be associated, where the discrete probabilities sum up to one, i.e., $\sum_{\zeta}p(\boldsymbol{\zeta})=1$. This probability is a statistical measure for the appearance of a special matching between image and model features. Of course, feature assignment is not an independent process. In 3-D object recognition there are, for instance, features which occlude each other. For that reason, we know if one feature is visible, the other cannot be part of the observation. Another statistical property of features is represented in their probabilistic modeling: due to segmentation errors the probability of feature appearance can vary. The detection of some features can be more robust, and therefore the probability of their appearance and assignment is higher compared to other features.

So far we have discussed the statistical modeling of features and assignments. Now we combine the introduced statistical components to describe a compound density for the characterization of object features within the image space. The non-observable assignment function ζ can be eliminated by projection, i.e. marginalization due to its statistical appearance. The probability density function of image features with latent assignments is therefore

$$p(O;B,R,t)=\sum_{\zeta}p(O,\boldsymbol{\zeta};B,R,t)=\sum_{\zeta}p(\boldsymbol{\zeta})\prod_{k=1}^{m}p(o_k;a_{\zeta(o_k)},R,t). \tag{5}$$

Obviously the evaluation complexity of this sum is in $\mathcal{O}(m\,n^m)$ and thus the usage of this density is computationally prohibitive.

The statistical modeling of assignment functions shows several degrees of freedom. For simplicity and computational efforts, statistically independent assignments are assumed. For instance, if the assignments of

image features are mutually independent, the factorization

$$p(\zeta) = \prod_{k=1}^{m} p(\zeta(o_k)) \tag{6}$$

is possible. Using this fact, the probability of the random vector is given by the product of its component probabilities. The model density for an observed set of features O simplifies to [9]

$$p(O;B,R,t) = \sum_{\zeta} p(O,\zeta;B,R,t)$$

$$= \prod_{k=1}^{m} \sum_{l=1}^{n} p(\zeta(o_k) = l)\, p(o_k; a_l, R, t). \tag{7}$$

For this model density the computational complexity for the evaluation of the model density (7) of a given observation is obviously bounded by $\mathcal{O}(n\,m)$. It allows the computation of density values for a given observation and a set of pose parameters in polynomial time.

In the following the model densities are assumed to be given, i.e., the parameter set B is known (see Ref. [4] for training algorithms). We restrict the algorithmic part of this paper to classification and pose estimation tasks. Pose estimation using above density functions corresponds to a parameter estimation problem which induces the global optimization task:

$$\{\hat{R}, \hat{t}\} = \operatorname*{argmax}_{R,t} p(O;B,R,t). \tag{8}$$

The classification applies the Bayesian decision rule, which decides for that class with highest a posteriori probability. This rule is known to minimize the rate of misclassification based on a 0/1 cost function [7, pp. 57, 58].

4. Recognition via projection

The algorithms for classification and pose estimation of 3-D objects are dominated by two different search problems [10]: the *continuous search* in the pose space and the *discrete search* considering various matchings of image and model features. In this section we also consider these strategies and describe two approaches to 3-D object recognition via 1-D projections. Instead of using 2-D data to reconstruct 3-D information, we apply further projections into 1-D space. In contrast to the elimination of assignments by marginalization (see Eq. (7)), the following ideas are easily motivated by geometric arguments.

4.1. Projection on coordinate axes

The L-shaped object shown in Fig. 1 is represented by 2-D point features. Rotation and translation of the original 3-D object induce a transform in the 2-D projection, and have to be known for classification purposes provided only point features are used. If 3-D object points are mapped into the image plane by orthographic projection, the pose is defined by the five parameters [3, p. 51] φ_x, φ_y, φ_z, t_1, and t_2 introduced in Section 3. Note that the orthographic projection is invariant with respect to translations $t = (0, 0, t_3)$ perpendicular to the image plane. Further variables can be eliminated by additional projections: considering Fig. 1 it is obvious that the 1-D projection of 2-D image points on the x-axis is invariant with respect to translation t_2 along the y-axis and rotation φ_x around the x-axis. Analogous properties are valid for the 1-D projection on the y-axis. These 1-D point features do not change with translation along the x-axis and rotation around the y-axis.

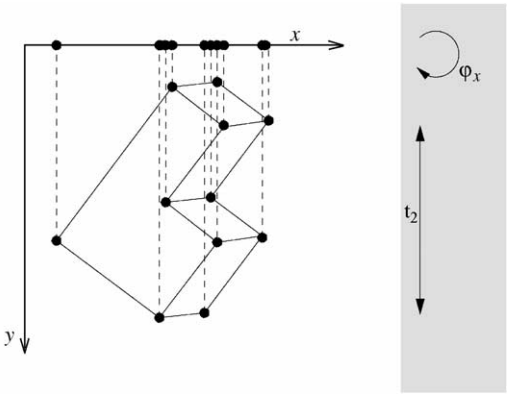

Fig. 1. Projections on the x-axis.

Now we are in position to apply projection techniques to efficient pose estimation. The density functions of 1-D point features are computed by marginalization. We integrate out the x- or y-coordinates of the 2-D point features. The resulting model densities show three degrees of freedom for pose parameters. Obviously, the projections reduce the search space from five to three dimensions, but they also decrease the discriminating power of features. There exists an infinite number of 2-D point configurations which share the same 1-D projection. For that reason, it is not sufficient to solve the associated optimization problem in the projection only.

Usually, a global optimization problem is divided up into two steps: first, a set of distinguished points is selected in the search space. Then in a second step, local optimization is performed in each point. For the selection of the distinguished points several procedures are feasible. For instance, points on regular grids or randomly chosen points are used. Within the proposed method, we take advantage of projections. We use local maxima of lower dimensional search spaces to guide the global optimization in higher dimensions. We suggest the following four-stage search algorithm [4] for solving the pose

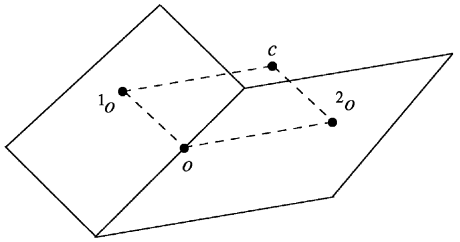

Fig. 2. Projection on the intersection of image planes.

estimation problem. The search is based on a parametric probability density function $p(O;B,R,t)$ for the 2-D point features in the image plane and its marginals over x and y coordinates:

(i) We compute the locations $(\varphi_y,\varphi_z,t_2)$ of the maxima in the projection of the probability density to the x-axis.
(ii) For the parameters φ_z that occur in the list of maxima in step (i), we compute the maxima $(\varphi_x,\varphi_z,t_1)$ in the projection of the probability density to the y-axis, i.e. we solve a 2-D optimization problem.
(iii) We perform local optimization in the points $(\varphi_x,\varphi_y,\varphi_z,t_1,t_2)$ whose coordinates $(\varphi_x,\varphi_z,t_1)$ and $(\varphi_y,\varphi_z,t_2)$ are in the solution sets of step (i) and step (ii), respectively.
(iv) We select the highest maximum of step (iii) as an approximation of the global maximum.

This procedure allows the estimation of pose parameters without knowing the matching between model and image features. The global optimization in the 5-D pose space is guided by the local maximization in 1-D projections on the coordinate axes.

4.2. Projection on the intersection of image planes

The basic theoretical problem that pops up in the second application of 1-D projections is the following: Decide whether two given sets of n distinct points in the real plane are images of the same set of points in $\mathbb{R}^3$ under a different orthogonal projection without explicit 3-D knowledge. The results presented in Ref. [5] concerning perspective projection will not be used in this paper.

In Fig. 2 the basic idea of the following theoretical discussion is illustrated. A 3-D model point $c\in\mathbb{R}^3$ is projected into two different image planes. The resulting projections are ${}^1o,{}^2o\in\mathbb{R}^2$. The orthographic projections of those 2-D image points on the intersection line of both image planes coincide in $o\in\mathbb{R}$. We conclude: if different sets of 2-D points are orthogonal projections of the same 3-D points, the corresponding 1-D projections on intersection lines are identical. This result is introduced in Ref. [5]. Their work is motivated by a different albeit not unrelated situation: In order to simplify the presentation and provide a clear understanding of the basic geometric ideas, we first confine ourselves to the basic setting as defined in Ref. [5]. Thus, at the beginning we assume that the point features show no instabilities and are exact. Later we incorporate statistical considerations which allow to deal with uncertainties. Thereby, we answer a question raised in Ref. [5] for an analogous algorithm in the setting of configurations of points whose position is characterized by a probability density function.

4.2.1. Geometric setup

For a subspace $V\subset\mathbb{R}^3$ we denote by π_V the orthographic projection onto V. In Ref. [5] the authors start with the following intuitive geometric observation we have already illustrated above:

Lemma 4.1. *Let V and W be two non-parallel planes in $\mathbb{R}^3$ and $\ell=V\cap W$. Assume $P\subseteq V$ and $Q\subseteq W$ are point-sets of cardinality n. There is a point-set $R\subseteq\mathbb{R}^3$ of cardinality n such that $\pi_V(R)=P$ and $\pi_W(R)=Q$ if and only if $\pi_\ell(P)=\pi_\ell(Q)$ counting multiplicities.*

For our purposes an important conclusion of the preceding lemma is the following: If P is a set of n points in the plane, then in order to decide, whether some set of m points Q is likely to be the image of the *same* set of points in $\mathbb{R}^3$, it suffices to check the existence of lines ℓ' and ℓ'' in the plane such that the intersection $\pi_{\ell'}(P)\cap\pi_{\ell''}(Q)$ of the orthographic projections is *large*. Here, we implicitly identify ℓ' and ℓ'' by a suitable orientation preserving isometry of the plane mapping ℓ' to ℓ''. Moreover, if the projections are generic, i.e., no two points of either set are mapped to the same point on the corresponding line, we also match the points corresponding to the same preimage for no extra cost. Of course, knowing ℓ' and ℓ'' still leaves freedom for the position of the planes V and W, and hence for the point-set S in $\mathbb{R}^3$ that projects onto P and Q. Indeed, for fixing the point-set S, up to affine transformations, in a generic situation, we need three projections.

The key fact employed in Ref. [5] in order to reduce the complexity of the algorithm is that the projections of a set of n points in the plane onto a line can advantageously be partitioned into $\mathcal{O}(n^2)$ equivalence classes. This observation and the definition of a suitable data structure are borrowed from Computational Geometry [11, pp. 29–32].

For a set P of n points in the plane one defines the *circular sequence* circ (P) corresponding to P as the set of permutations of the n points that occur as a projection of P onto a line. More precisely, label the points in P in an arbitrary but fixed way by numbers $\{1,\ldots,n\}$. Then fix a line ℓ_0 through the origin. Let ℓ_ρ be the image of ℓ_0 under clockwise rotation by the angle $\rho\in[0,\pi]$ around the origin. The set of angles $[0,\pi]$ is subdivided

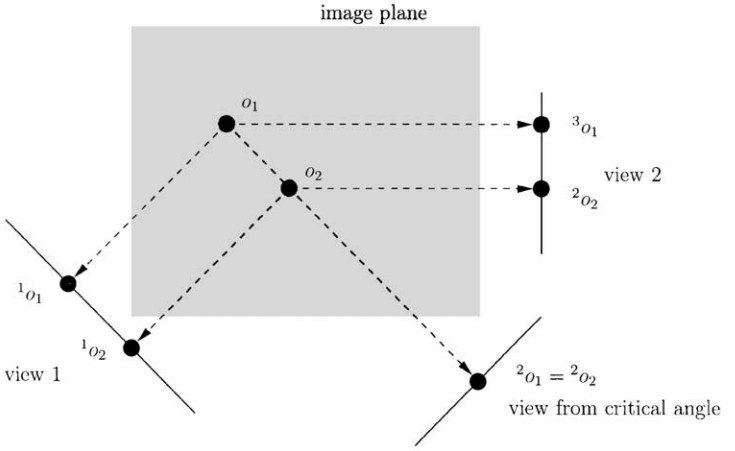

Fig. 3. Different 1-D projections and critical angle.

into open intervals and critical angles according to the order in which the points of P project on the line ℓ_ρ. Each *critical angle* ρ corresponds to the situation when at least two points of P map to the same point on the line ℓ_ρ corresponding to the angle ρ. Thus, when the lines ℓ_ρ pass through a critical angle at least two points swap their position. Fig. 3 illustrates this situation.

Obviously, there are at most $\binom{n}{2}$ different critical points and hence at most $\binom{n}{2}+1$ open intervals. The critical points are determined by the $\leqslant \binom{n}{2}$ lines through pairs of points in P. Thus circ (P) can be stored in an $n \times \binom{n}{2}$ array. Note, that for computational purposes only swaps have to be stored. Hence, circ (P) can be computed in time $\mathcal{O}(n^2 \log n)$—sort the angles determined by the $\leqslant \binom{n}{2}$ lines through pairs of points in P.

Before, we proceed to the part of the algorithm responsible for checking existence and finding lines satisfying the criterion of Lemma 4.1, we make certain additional genericity assumptions on possible point configurations. A point-set P in $\mathbb{R}^2$ is called *generic* if no two subsets of four points can be transformed into each other by an affine transformation and no three points lie on a line. Note that the set of non-generic configurations is a set of measure 0 in the set of all point configurations of n distinct points.

The next lemma gives a criterion for the uniqueness of the lines ℓ' and ℓ'' [5].

Lemma 4.2. *Let P and Q be two generic configurations of four points in $\mathbb{R}^2$. Then there is at most one pair of lines ℓ' and ℓ'' for which there is an orientation preserving isometry of $\mathbb{R}^2$ identifying ℓ and ℓ' such that $\pi_{\ell'}(P) = \pi_{\ell''}(Q)$.*

In order to check existence and find the lines ℓ' and ℓ'' we have to solve a system of four *homogeneous linear equations* in four variables. Since we are dealing with orthographic projections we may assume that the lines ℓ', ℓ'' are passing through the origin. Let $P = \{p^1, \ldots, p^4\}$ and $Q = \{q^1, \ldots, q^4\}$. Here the indexing is chosen according to occurrence of the points in the circular sequences circ(P) and circ(Q). Thus, we have to solve the system

$$v' p^i = v'' q^i, \quad i = 1, \ldots, 4 \tag{9}$$

for non-zero solutions v' and v'' of equal length, which then span the lines ℓ' and ℓ''. Note that any pair v' and v'' of non-zero solutions of the system solves the recognition problem for scaled orthographic projection with scaling factor given by the ratio of its lengths (see discussion in Section 6).

By our genericity assumptions the coefficient matrix of the system has either rank three, i.e., there is a solution and the solution—the lines determined by v' and v''—is unique, or rank four, i.e., there is no solution.

Even though we assume for this paragraph that point positions are given exactly, we here describe an algorithm for solving the system in case of numerical instability of the input data. In this situation the coefficient matrix will *usually* have rank four. Note, that the set of matrices of rank $\leqslant 3$ is a set of measure 0 in the set of all (4×4)-matrices. Thus we need a method for deciding, whether our data *probably* come from a situation where a non-trivial solution exists. Here we propose the following approach: Let A be a (4×4)-matrix and $x \in \mathbb{R}^4$. We want to find non-trivial solutions of $Ax = 0$. First, we calculate the singular value decomposition $A = U \operatorname{diag}(w) V^{\mathrm{T}}$ of A, where U and V are orthogonal matrices and $\operatorname{diag}(w)$ is a diagonal matrix with diagonal entries from the vector $w \in \mathbb{R}^4$. The components of w are called the singular values of A. Now A is a rank four matrix if and only if w has no non-zero entries. Let w' be the vector

obtained from w by replacing the singular value with least absolute value by 0, and define $A' = U\,\mathrm{diag}(w')V^{\mathrm{T}}$. It is fairly easy to see that if A is of rank four then A' is the rank three matrix that minimizes the distance to A in the spectral norm. Recall, for a matrix B the spectral norm is the largest eigenvalue of the square root of BB^{T}. Thus if A is a (4×4)-matrix of rank four then its deviation from being a rank three matrix can be measured by its smallest singular value. If the smallest singular value of A falls below a certain threshold—depending on the singular values of A—then we replace A by A'. Assume we have set the fourth entry of w in w' to zero. Let e_4 be the fourth unit vector, then we solve the system by the formula

$$x = Ve_4. \tag{10}$$

Otherwise we say that there exists no non-trivial solution.

Finally, we reduce the number of pairs of sets of cardinality four that have to be considered. A counting argument [5] shows that there must be four columns in $\mathrm{circ}(P)$ such that the total number of quadruples of points in these columns is $\mathcal{O}(n)$. Four columns satisfying this property can be found in time $\mathcal{O}(n^2)$—find the four columns with the least number of points.

Now we are in position to formulate the algorithm proposed in Ref. [5] for orthographic projections. For two generic sets P and Q of points in $\mathbb{R}^2$; $|P| = n$, $|Q| = m$, we proceed as follows:

(i) Compute circular sequences $\mathrm{circ}(P)$ and $\mathrm{circ}(Q)$; this is in $\mathcal{O}(m^2 \log m + n^2 \log n)$
(ii) Find four columns in $\mathrm{circ}(P)$, (resp., in $\mathrm{circ}(Q)$) such that the total number of quadruples of points in these columns is $\mathcal{O}(n)$ (resp., $\mathcal{O}(m)$); the search is in $\mathcal{O}(m^2 + n^2)$.
(ii) For each of the $\mathcal{O}(nm)$ pairs of quadruples compute the lines ℓ' and ℓ''—if they exist—such that the projection of the corresponding quadruples on the lines coincide, which is bounded by $\mathcal{O}(mn)$.
(iii) Among the pairs of lines constructed in Step (iii), select the pair (ℓ', ℓ'') such that $\pi_{\ell'}(P) \cap \pi_{\ell''}(Q)$ is maximal; this step is obviously bounded by $\mathcal{O}(\max(n, m))$.

Note, that in our situation we can assume that $\mathrm{circ}(Q)$ and the four columns in $\mathrm{circ}(Q)$ containing $\mathcal{O}(m)$ quadruples have been pre-computed. The complexity of the algorithm is dominated by Steps (iii) and (iv). In the worst case Step (iv) has to be performed for all instances of Step (iii). Thus the algorithm runs in time $\mathcal{O}(mn \max(m, n))$.

Inaccuracy of real world data obstructs a straightforward implementation of the introduced algorithm. But as mentioned before the statistical behavior of segmented point features can be modeled by Gaussian distributions [9]. It remains to incorporate these statistical methods in the algorithm.

4.2.2. Statistical setup

It is the basic hypothesis of our statistical modeling scheme that 3-D point features are mutually independent and normally distributed (c.f. Section 3). We observe 2-D projections of 3-D points. Since the projection is orthographic, rotation, translation and projection are characterized by an affine mapping, the resulting 2-D features are also normally distributed. The above introduced geometric algorithm requires at least two 2-D projections for classification. It is important to note that the 3-D structure of objects is not necessarily required for the recognition of the 3-D object. Generally, we assume that any two projections are statistically independent. The reasoning from Ref. [7, pp. 27,28] also applies to the projection of the 2-D points onto the lines determined by Eq. (10). If line ℓ is defined by its normalized spanning vector $\boldsymbol{v} = (v_1, v_2)^{\mathrm{T}}$ where $||\boldsymbol{v}|| = 1$, then the 2-D point o' resulting from the orthographic projection of o to the line ℓ is given by

$$o' = \begin{pmatrix} o_1' \\ o_2' \end{pmatrix} = \begin{pmatrix} 1 - v_1^2 & -v_1 v_2 \\ -v_1 v_2 & 1 - v_2^2 \end{pmatrix} \begin{pmatrix} o_1 \\ o_2 \end{pmatrix} = Mo. \tag{11}$$

Also the 1-D coordinate of the projection is normally distributed, since the 1-D coordinate $\pi_\ell(\boldsymbol{o})$ is defined by the affine mapping:

$$\pi_\ell(\boldsymbol{o}) = -\frac{\boldsymbol{v}}{||\boldsymbol{v}||^2}\,\boldsymbol{o} = -\,\boldsymbol{v} \in \mathbb{R}. \tag{12}$$

The point positions on the lines are normally distributed 1-D random variables. The mean value is $-v\,M\boldsymbol{\mu} \in \mathbb{R}$ and the covariance $\boldsymbol{v}M\boldsymbol{\Sigma}(\boldsymbol{v}M)^{\mathrm{T}} \in \mathbb{R}$.

The lines determined by Eq. (10) are computed using the detected 2-D features. Clearly, the statistical behavior of the 2-D point features influence the computation of the lines.

We have to neglect these instabilities, since we do not know of any result that relates the distribution of the entries of the matrix A to the behavior of the matrix V from Eq. (10). For recent results that show how difficult it is to obtain such estimates we refer the reader to Ref. [12]. Moreover, we think that this omission is justified by the fact that our method for solving the system with instable parameters is based on a method that optimizes the accuracy of the solution (see Section 4.2.1).

The above transform shows how to compute densities for 1-D projections on lines. Now we are in position to formulate the final algorithm for classification.

Assume we are given a database of objects which are described by their density functions in a 2-D projection—these density functions will be Gaussian distributions described by their mean vector and covariance matrix. These densities can result from 3-D models as well as from simple 2-D models as for instance used in appearance based approaches to vision [13]. Within a

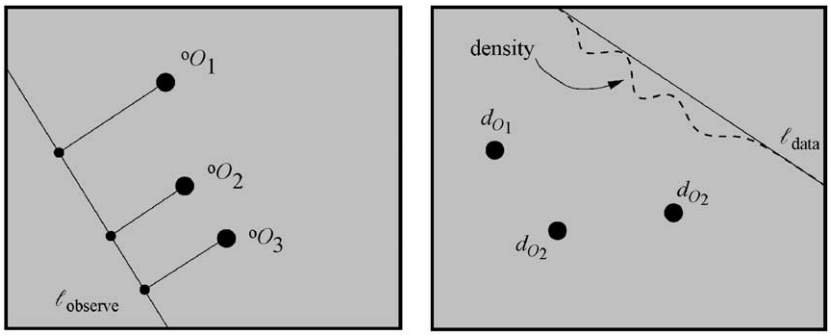

Fig. 4. 1-D projection of observed point features and a database density function.

pre-processing step, gray-level images are transformed into segmentation results which provide geometric features including the 2-D point features attached to the object. In the next step we give a probabilistic estimate for an object from the database which is most likely to be the origin of the observation. To a given object in the database we associate the point feature that is given by the mean vector of the density function. We run the algorithm described in Section 4.2 to check the existence and compute lines ℓ_{data} in the plane of the database object and ℓ_{observe} in the observation plane for which the respective projections of a pair of quadruples of point features match. Then we calculate the marginal density of the database object on the line ℓ_{data}. We project the point features of the observed object onto ℓ_{observe} (see Fig. 4).We identify the line ℓ_{data} with the line ℓ_{observe} by identifying the centers of gravity of the quadruples that where used to compute the lines (orientation described by the order of points). Now we evaluate the marginal density function on ℓ_{data} at the point features on ℓ_{observe}. We maximize an objective function calculated from density values over the objects in the database. The object where the maximum is obtained will then provide the estimate. In the experimental evaluation in Section 5 we have used as the objective function the weighted matching between observed points and points of the database object multiplied with the ratio $\leqslant 1$ of observed and model points. The weights are given by the density values of the observed points.

5. Experimental evaluation

According to the theoretical description, the experimental evaluation is divided up into two parts: The results concerning pose estimation using 1-D projections on coordinate axes (see algorithm in Section 4.1), and the evaluation of the recognition algorithm based on 1-D point features in the intersection of two projection planes (see algorithm in Section 4.2). Two important things which have to be taken into consideration are the impact of ambiguities and the relation between achieved speed-up and reduction of discriminating power.

5.1. Pose estimation experiments

In this section we describe pose estimation experiments using an implementation of the algorithm proposed in Section 4.1. Let us consider a probability density function associated with an object characterized by ten 3-D points. Each point is assumed to be normally distributed. Obviously, the global optimization of this function is a non-trivial problem. Even by visual inspection of this multi-model density there is no maximum with a "large" area of attraction. This is even worse using 1-D projection. Nevertheless, it is advantageous to start the global optimization based on 1-D features due to the reduced dimension of the search space. But not only the dimension is decreased. Different projections lead to optimization problems in different sub-spaces. Unfortunately, these sub-spaces depend on each other, and therefore a final verification stage using the original 2-D projections is required. An appreciated side-effect, however, is due to the fact that the time required for density evaluations is reduced by 15% using 1-D instead of 2-D features. For that reason the search related to 1-D features proceeds in a lower dimensional parameter space and with more efficient function evaluation. Since there exist no theoretical mathematical tool to compare the estimates based on 2-D with estimates using 1-D projection we run an experimental comparison. We use 400 2-D views and measure the rate of correct estimates. The optimization using 1-D projections is done by an adaptive random search technique combined with local optimization using the downhill simplex algorithm [14, Chapter 10.4]. The local optimization in 5-D search space is also done by the simplex method. Experiments show that 87% of all global optimizations succeed in finding the global maximum. The average number of function evaluations using 1-D features is 5400 and in case of 2-D features about 260 evaluations are sufficient. Compared to a pure 5-D search the

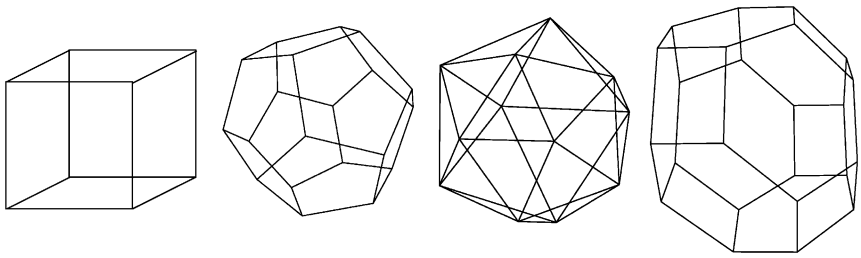

Fig. 5. Polytopes: cube, dodecahedron, icosahedron, and 3-D permutahedron.

Table 1
Models, generated quadruples, and average runtime

Object #	# points	# quadruples	Runtime (s)
1	8	28	47
2	24	277	48
3	12	67	108
4	20	191	372

Table 2
Recognition rates (in %) for varying variances

Object #	$\sigma^2 = 0$	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$	$\sigma^2 = 1.0$	$\sigma^2 = 1.5$
1	100	100	100	99	95
2	100	100	100	100	100
3	100	100	100	100	96
4	100	98	90	93	88

three-stage optimization method is twice as fast. In the presence of additional point features, which do not belong to the object, the global maximization based on 1-D projections is even four times as fast.

5.2. Recognition experiments

Before we analyze the algorithm for classification from Section 4.2 on real data, we run experiments based on synthetic data; here we use some standard polytopes. [4] The objects are shown in Fig. 5, and the average width and height are 30 pixels. We use 100 random views of each object, where the point features are normally distributed. We run the classification procedure based on the 1-D projections introduced in Section 4.2.1. Since we have prior 3-D models and the presented method matches two 2-D views we also generate the second view randomly. The interesting question to compute the most discriminating model view for a given observation was not part of this investigation. The number of model points, the cardinality of generated quadruples, and the runtime on an SGI O_2, R10000 are shown in Table 1. The recognition rates for varying variances of the 1-D point features are summarized in Table 2.

Sample views of real objects used for classification experiments are shown in Fig. 6. The distance to the camera is kept constant such that no scaling appears and perspective distortion is minimal. For corner detection we use a standard algorithm which is based on the maximization curvatures of detected lines. The segmentation results show that half of the observed point features belong to artifacts and show no correspondence to a real 3-D point of the object. The achieved recognition rate using nine views of each object is more than 86%. [5] The number of points detected in our pictures by the segmentation process varies from 5 to 60. Correspondingly, the runtime of the algorithm is in the range from 80 s to 4 h.

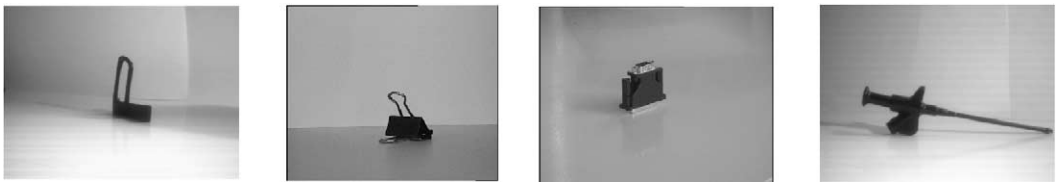

Fig. 6. Real objects.

[4] Coordinate calculations and visualization were performed using the software package POLYMAKE [15].

[5] The data files used for experiments can be found in http://www.mathematik.uni-marburg.de/ ~welker/hwn.

Due to the fact that the segmentation results are not improved by manual manipulation and that there is no need to compute 3-D data, the recognition rate is remarkably high on real data.

The obtained experimental results prove that the usage of 1-D point features is useful for both recognition and pose estimation.

6. Summary and conclusions

The main result of the paper is that 1-D projections can simplify the combinatorial structure of the pose estimation and classification problem. Even though the presented algorithms still carry the drawback that projections reduce the discriminating power of the features, the simplification of the combinatorial structure leads to a remarkable speed up of existing algorithms for object recognition.

Our current approach is restricted to point features. This restriction requires the robust detection of the point features. Thus future research on classification algorithms relying on 1-D projections will concentrate on features avoiding this prerequisite. A potential approach could be based on the projection of gray-level images on the 1-D intersection of planes by integrating over gray levels. In a second thread it is desirable to design hybrid mechanisms that combine the advantages of existing algorithms using 2-D and 1-D projections, automatically selecting optimal strategies. In a separate thread it seems worthwhile to investigate the potential of our approach from Section 4.2.1 for scaled orthographic projections. This approach appears to be promising since it is known that the error made by replacing a perspective projection by a scaled orthographic is negligible under certain circumstances [16].

References

[1] A.K. Jain, D. Zongker, Feature selection: evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 153–158.

[2] H. Niemann, Pattern Analysis and Understanding, Springer Series in Information Sciences, Vol. 4, Springer, Heidelberg, 1990.

[3] O. Faugeras, Three-Dimensional Computer Vision—A Geometric Viewpoint, MIT Press, Cambridge, MA, 1993.

[4] J. Hornegger, H. Niemann, Probabilistic modeling and recognition of 3-D objects, Int. J. Comput. Vision 39 (3) (2000) 229–251.

[5] D. Huttenlocher, J.M. Kleinberg, Comparing point sets under projection, Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, Arlington, Virginia, January, SIAM, Philadelphia, PA, 1994, pp. 1–7.

[6] P. Bremaud, An Introduction to Probabilistic Modeling, Undergraduate Texts in Mathematics, Springer, Heidelberg, 1988.

[7] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, Boston, 1990.

[8] R. Beß, D. Paulus, M. Harbeck, Segmentation of lines and arcs and its application to depth recovery, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 4, IEEE Computer Society Press, Munich, April 1997, pp. 3161–3165.

[9] W.M. Wells III, Statistical approaches to feature-based object recognition, Int. J. Comput. Vision 21 (2) (1997) 63–98.

[10] A.K. Jain, P.J. Flynn (Eds.), Three-Dimensional Object Recognition Systems, Elsevier, Amsterdam, 1993.

[11] H. Edelsbrunner, Algorithms in Combinatorial Geometry, Springer, Heidelberg, 1987.

[12] A. Edelman, The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law, J. Multivariate Anal. 60 (2) (1997) 205–232.

[13] J. Mundy, A. Liu, N. Pillow, A. Zisserman, S. Abdallah, S. Utcke, S. Nayar, C. Rothwell, An experimental comparison of appearance and geometric model based recognition, in: J. Ponce, A. Zisserman, M. Hebert (Eds.), Object Representation in Computer Vision, Lecture Notes in Computer Science, Vol. 1144, Springer, Heidelberg, 1996, pp. 257–272.

[14] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C—The Art of Scientific Computing, Cambridge University Press, New York, 1990.

[15] E. Gawrilow, M. Joswig, POLYMAKE, Technical Report http://www.math.tu-berlin.de/diskregeom/polymake, Fachbereich Mathematik, Technische Universität, Berlin, 1997.

[16] G. Xu, Z. Zhang, Epipolar Geometry in Stereo, Motion and Object Recognition—A Unified Approach, Computational Imaging and Vision, Vol. 6, Kluwer Academic Press, Dordrecht, 1996.

About the Author—JOACHIM HORNEGGER graduated 1992 (diploma) and received his Ph.D. degree in Computer Science 1996 at the Universität Erlangen-Nürnberg, Germany, for his work on statistical object recognition. Joachim was a research and teaching associate at the Universität Erlangen-Nürnberg, a visiting scientist at the Technion, Israel, and at the Massachusetts Institute of Technology (MIT), USA. During 1997/98 he was a visiting scholar at Stanford University, USA. His major research interests are 3-D computer vision, 3-D object recognition and statistical methods applied to image analysis problems. Joachim has taught computer vision, medical image processing, and pattern recognition at the Universität Erlangen-Nürnberg, Germany, at the Catholic University of Eichstätt, Germany, at the University of Seville, Spain, and at Stanford University, USA. He is the author and coauthor of more than 50 scientific publications including a monography on applied pattern recognition. Joachim was awarded the runner up DAGM Prize (1996), the DAGM Prize (1997), and the "Innovationspreis für Medizintechnik" (2000) by the German Ministry of Science and Education. Currently Joachim is with Siemens Medical Solutions working on 3-D reconstruction. He is also an appointed lecturer at the Universität Mannheim, Germany.

About the Author—VOLKMAR WELKER received his diploma and Ph.D. in Mathematics in 1988 and 1990 at Universität Erlangen-Nürnberg, Germany, and his Habilitation in Mathematics from Universität Essen, Germany, in 1996. He was "Wissenschaftlicher Angestellter" at the universities in Erlangen, Germany (1989–1992, computer science), and Essen, Germany (1993–1994, 1996–1998, mathematics). During the academic year 1991/92 he spent 5 months at Mittag-Leffler Institute in Stockholm, Sweden, on a scholarship of the Royal Swedish Academy of Sciences. From 1992 till 1993 he was visiting scholar at the department of applied mathematics of Massachusetts Institute of Technology (MIT), USA, funded by Deutsche Forschungsgemeinschaft (DFG). From 1994 till 1996 he held a "Habilitations-Stipendium" of DFG working in Essen, Stockholm and at MSRI, Berkeley, and from 1998 till 1999 a Heisenberg scholarship of DFG working at Technische Universität Berlin, Germany. Since 1999 he is professor for discrete mathematics at Philipps-Universität Marburg. In 1995 he was awarded the Heinz-Maier-Leibnitz price by the German Ministry of Science and Education and DFG. He has been an invited plenary speaker on conferences in Germany, USA, Poland and Japan, and has organized several conferences in Germany and USA. His research areas include algebraic, enumerative and geometric combinatorics with a view toward applications in physics, chemistry and computer science.

About the Author—HEINRICH NIEMANN obtained the degree of Dipl.-Ing. in Electrical Engineering and Dr.-Ing. at Technical University Hannover in 1966 and 1969, respectively. During 1966/67 he was a graduate student at the University of Illinois, Urbana. From 1967 to 1972 he was with Fraunhofer Institut für Informationsverarbeitung in Technik und Biologie, Karlsruhe, working in the field of pattern recognition and biological cybernetics. During 1973–1975 he was teaching at Fachhochschule Giessen in the department of Electrical Engineering. Since 1975 he has been Professor of Computer Science at the University of Erlangen-Nürnberg, where he was dean of the engineering faculty of the university from 1979–1981. Since 1988 he is also head of the research group 'Knowledge Processing' at the Bavarian Research Institute for Knowledge Based Systems (FORWISS), and since 1998 he is speaker of a 'special research area' (SFB) entitled 'Model-Based Analysis and Visualization of Complex Scenes and Sensor Data' funded by the German Research Foundation (DFG). His fields of research are speech and image understanding and the application of artificial intelligence techniques in these fields. He is on the editorial board of Signal Processing, Pattern Recognition Letters, Pattern Recognition and Image Analysis, Journal of Computing and Information Technology, and Computers and Electrical Engineering.