Technical Report

Pattern Recognition and Image Processing Group
Institute of Automation
Vienna University of Technology
Treitlstr. 3/1832
A-1040 Vienna AUSTRIA
Phone: +43 (1) 58801-8161
Fax: +43 (1) 5054668
E-mail: {ales,bis}@prip.tuwien.ac.at

PRIP-TR-47 June 30, 1997

Robust Recognition Using Eigenimages[1]

Aleš Leonardis[2]*, Horst Bischof and Roland Ebensberger*

Abstract

The basic limitations of the current appearance-based matching methods using eigenimages are non-robust estimation of coefficients and inability to cope with problems related to outliers, occlusions, and segmentation. In this paper we present a new approach which successfully solves these problems. The major novelty of our approach lies in the way how the coefficients of the eigenimages are determined. Instead of computing the coefficients by a projection of the data onto the eigenimages, we extract them by a hypothesize-and-test paradigm using subsets of image points. Competing hypotheses are then subject to a selection procedure based on the Minimum Description Length principle. The approach enables us not only to reject outliers and to deal with occlusions but also to simultaneously use multiple classes of eigenimages.

[1] This work was supported by a grant from the Austrian National Fonds zur Förderung der wissenschaftlichen Forschung (No. S7002MAT). A. L. also acknowledges the support from the Ministry of Science and Technology of Republic of Slovenia (Project J2-8829), European Union Copernicus Program (Grant 1068 RECCAD), and by U.S.–Slovene Joint Board (Project #95-158).

[2] Also at the University of Ljubljana, Faculty of Computer and Information Science, Computer Vision Laboratory, Ljubljana, Slovenia. *ales.leonardis@fri.uni-lj.si*

1 Introduction and Motivation

The appearance-based approaches to vision problems have recently received a renewed attention in the vision community due to their ability to deal with combined effects of shape, reflectance properties, pose in the scene, and the illumination conditions [1, 2]. Besides, the appearance-based representations can be acquired through an automatic learning phase which is not the case with traditional shape representations. The approach has led to a variety of successful applications, e.g., illumination planning [3], visual positioning and tracking of robot manipulators [4], visual inspection [5], "image spotting" [6], and human face recognition [7, 8].

As stressed by its proponents, the major advantage of the approach is that both learning as well as recognition are performed using just two-dimensional brightness images without any low- or mid-level processing. However, there still remain various problems to be overcome since the technique rests on direct appearance-based matching [1]. The most severe limitation of the method in its present form is that it cannot handle the problems related to occlusion and segmentation.

The approach of modular eigenspaces [9] tries to alleviate the problem of occlusion but does not solve it entirely because the same limitation holds for each of the modular eigenspaces.

Moreover, the current approaches are also not robust, where the term *robustness* refers to the fact that the results remain stable in the presence of various types of noise and can tolerate a certain portion of outliers [10, 11]. Robustness can be characterized by the concept of *breakdown point*, which is determined by the smallest portion of outliers in the data set at which the estimation procedure can produce an arbitrarily wrong estimate. For example, in current approaches even a single erroneous data point can cause an arbitrary wrong result, meaning that the breakdown point is 0%[1]. Recently some other methods which deal with the robustness issue of appearance-based matching have been proposed [12, 13].

In this paper we present a new approach which successfully solves these problems. The major novelty of our approach lies in the way how the coefficients of the eigenimages are determined. Instead of computing the coefficients by a projection of the data onto the eigenimages, we extract them by a hypothesize-and-test paradigm using *subsets* of image points. Competing hypotheses are then subject to a selection procedure based on the Minimum Description Length (MDL) principle. The approach enables us not only to reject outliers and to deal with occlusions but also to simultaneously use multiple classes of eigenimages.

The paper is organized as follows: We first review the basic concepts of the current appearance-based matching methods and point out the main limitations. In section 3 we present the basic steps of our method and outline the complete algorithm. We also suggest a modified version of the basic algorithm which is computationally less demanding and does not compromise the robustness. We first present in section 4 some results on 1-D signals where the main steps of the algorithm can easily be visualized, and in section 5 we present the results on complex image data using the standard image database from the University of Columbia [14]. We conclude with a discussion and outline the work in progress.

2 Appearance-based matching

The appearance-based methods consist of two stages. In the first stage a set of images (templates), i.e., training samples, is obtained. These images usually encompass the appearance of a single object under different orientations [5], different illumination directions [3], or multiple instances of a class of objects, e.g., faces [7]. The sets of images are normally highly correlated. Thus, they can efficiently be compressed using Principal Component Analysis (PCA) [15], resulting in a low-dimensional eigenspace.

In the second stage, given an input image, the recognition system projects parts of the input image, (i.e., subimages of the same size as training images), to the eigenspace. In the absence of specific cues, e.g., when motion can be used to pre-segment the image, the process is sequentially applied to the entire image. The recovered coefficients indicate the particular instance of an object and/or its position, illumination, etc.

We now introduce the notation. Let $\mathbf{y} = [y_1, \ldots, y_m]^T \in \mathbb{R}^m$ be an individual template, and $\mathcal{Y} = \{\mathbf{y}_1, \ldots \mathbf{y}_n\}$ be a set of templates; throughout the paper a simple vector notation is used since the

[1] This claim is of course valid only under the assumption that the outlier can attain an arbitrary value, which is in practical applications usually not the case.

extension to 2-D is straightforward. To simplify the notation we assume $\mathcal{Y}$ to be normalized, having zero mean. Let $\mathbf{Q}$ be the covariance matrix of the vectors in $\mathcal{Y}$; we denote the eigenvectors of $\mathbf{Q}$ by $\mathbf{e}_i$, and the corresponding eigenvalues by λ_i. We assume that the number of templates n is much smaller than the number of elements m in each template, thus an efficient algorithm based on SVD can be used to calculate the first n eigenvectors [1]. Since the eigenvectors form an orthogonal basis system, $<\mathbf{e}_i, \mathbf{e}_j>= 1$ when $i = j$ and 0 otherwise, where $<>$ stands for a scalar product. We assume that the eigenvectors are ordered in the descending order with respect to the corresponding eigenvalues λ_i. Then, depending on the correlation among the templates in $\mathcal{Y}$, only p, $p < n$, eigenvectors are needed to represent the $\mathbf{y}_i$ to a sufficient degree of accuracy as a linear combination of eigenvectors $\mathbf{e}_i$

$$\tilde{\mathbf{y}} = \sum_{i=1}^{p} a_i(\mathbf{y})\mathbf{e}_i \ . \tag{1}$$

We call the space spanned by the first p eigenvectors the *eigenspace*.

To recover the parameters a_i during the matching stage, a data vector $\mathbf{x}$ is projected onto the eigenspace

$$a_i(\mathbf{x}) =< \mathbf{x}, \mathbf{e}_i >= \sum_{j=1}^{m} x_j e_{i_j} \qquad 1 \leq i \leq p \ . \tag{2}$$

$\mathbf{a}(\mathbf{x}) = [a_1(\mathbf{x}), \ldots, a_p(\mathbf{x})]^T$ is the point in the eigenspace obtained by projecting $\mathbf{x}$ onto the eigenspace. Let us call the $a_i(\mathbf{x})$ coefficients of $\mathbf{x}$. The reconstructed data vector $\tilde{\mathbf{x}}$ can be written as

$$\tilde{\mathbf{x}} = \sum_{i=1}^{p} a_i(\mathbf{x})\mathbf{e}_i \ . \tag{3}$$

It is well known that PCA is among all linear transformations the one which is optimal with respect to the reconstruction error $||\mathbf{x} - \tilde{\mathbf{x}}||^2$.

2.1 Weaknesses of current appearance-based matching

In this section we analyze some of the basic limitations of the current appearance-based matching methods and illustrate them with a few examples. Namely, the way how the coefficients a_i are calculated poses a serious problem in the case of outliers and occlusions.

2.1.1 Outliers

Let us suppose that x_j in Eq. (2) is corrupted by δ. Then, $\hat{a_i} = a_i + \delta e_{i_j}$. It follows that $||\hat{\mathbf{a}} - \mathbf{a}||$ can get arbitrarily large, just by changing a single component x_j. This proves that the method is non-robust with the breakdown point 0%.

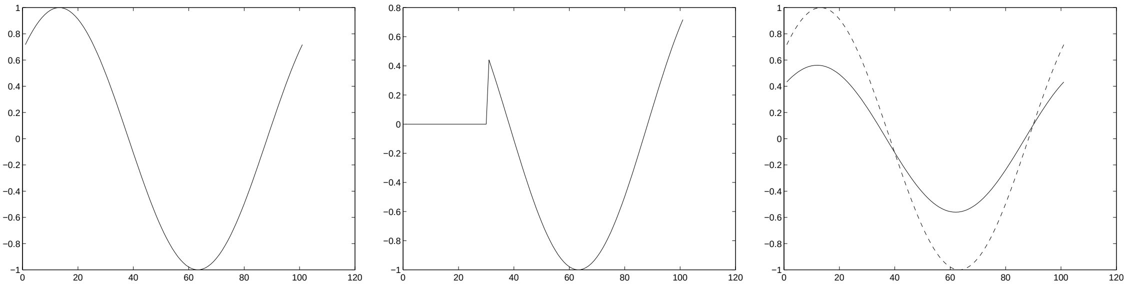
2.1.2 Occlusion

Similarly one can analyze the effect of occlusion. Suppose that $\hat{\mathbf{x}} = [x_1, \ldots, x_r, 0, \ldots 0]^T$ is obtained by setting last $m-r$ components of $\mathbf{x}$ to zero; a similar analysis holds when some of the components of $\mathbf{x}$ are set to some other values, which, for example, happens in the case of occlusion by another object. Then

$$\hat{a}_i = \hat{\mathbf{x}}^T \mathbf{e}_i = \sum_{j=1}^{r} x_j e_{i_j} \ . \tag{4}$$

The error we make in calculating a_i is

$$(a_i(\mathbf{x}) - \hat{a}_i(\hat{\mathbf{x}})) = \sum_{j=r+1}^{m} x_j e_{i_j} \ . \tag{5}$$



(a) Original signal

(b) Occluded signal

(c) Original (dashed) and reconstructed signal

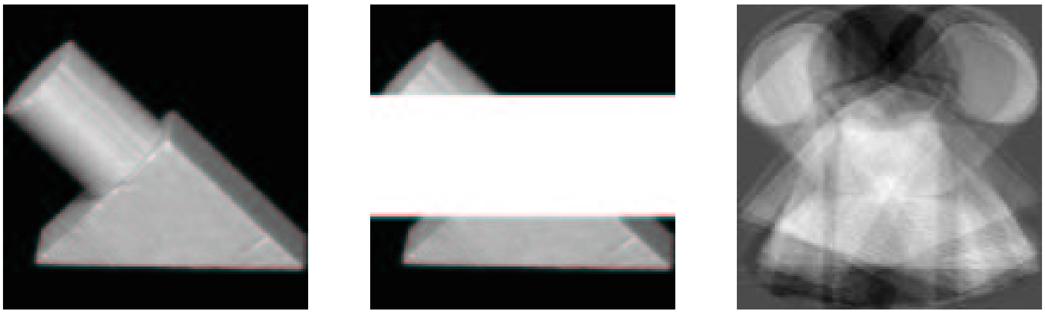
Figure 1: Demonstration of the occlusion using the standard approach for calculating the coefficients a_i.

Similarly, the additional error caused by occlusion is

$$||\sum_{i=1}^{p}(\sum_{j=r+1}^{m} x_j e_{i_j})\mathbf{e}_i||^2 \ . \tag{6}$$

This error is not localized at the occluded part but spreads over the whole vector $\tilde{\mathbf{x}}$.

Let us demonstrate the effect of occlusion on a simple 1-D example. Fig. 1a shows a test-signal taken from the set of training signals of trigonometric functions (see section 4). The signal can be exactly described by the coefficient vector $\mathbf{a} = [5.0042, 5.0470]^T$. Fig. 1b shows the occluded signal, where the first 30 elements have been set to 0. Using Eqs. (2) and (3) we get the signal shown in Fig. 1c, with the calculated coefficient vector $\mathbf{a} = [2.5658, 3.0509]^T$. This simple example nicely demonstrates the consequences of the occlusion. Of course, the same applies to the 2-D case as shown in Fig. 2.



(a) Original image

(b) Occluded image

(c) Reconstructed image

Figure 2: Demonstration of the occlusion using the standard approach for calculating the coefficients a_i.

2.1.3 Segmentation

Since in calculating the eigenimages there is no distinction made between the object and the background (which is usually assumed to be black), the effect of a varying background is, in the recognition phase, similar to occlusion. Therefore, to obtain the correct result in the case of the standard method, the object of interest should be first segmented from the background and then augmented with the original background. Our robust approach does not require this segmentation step and can thus cope with objects that appear on various backgrounds.

The problems that we have discussed arise because the complete set of data $\mathbf{x}$ is required to calculate a_i in (Eq. 2). Therefore, the method is sensitive to partial occlusions, to data containing noise and outliers, and to changing backgrounds.

In the next section we explain our new approach which has been designed to overcome precisely these type of problems.

3 Our approach

The major novelty of our approach lies in the way how the coefficients of the eigenimages are determined. Instead of computing the coefficients by a projection of the data onto the eigenimages, we extract them by a robust hypothesize-and-test paradigm using only *subsets* of image points. Competing hypotheses are then subject to a selection procedure based on the Minimum Description Length principle. More specifically, our approach, which we present in the following subsections, consists of four main steps: hypotheses generation, selection, fitting, and a final selection. We will also discuss a modification of this basic algorithm which is computationally less demanding and does not compromise the robustness.

3.1 Generating hypotheses

Let us first start with a simple observation. If we take into account all eigenvectors, i.e., $p = n$, and if there is no noise in the data x_{r_i} then in order to calculate the coefficients a_i (Eq. 2) we need only n points $\mathbf{r} = (r_1, \dots r_n)$. It is sufficient to compute the coefficients a_i by simply solving the following system of linear equations (see Fig.3):

$$x_{r_i} = \sum_{j=1}^{n} a_j(\mathbf{x}) e_{j r_i} \quad 1 \le i \le n \ . \tag{7}$$

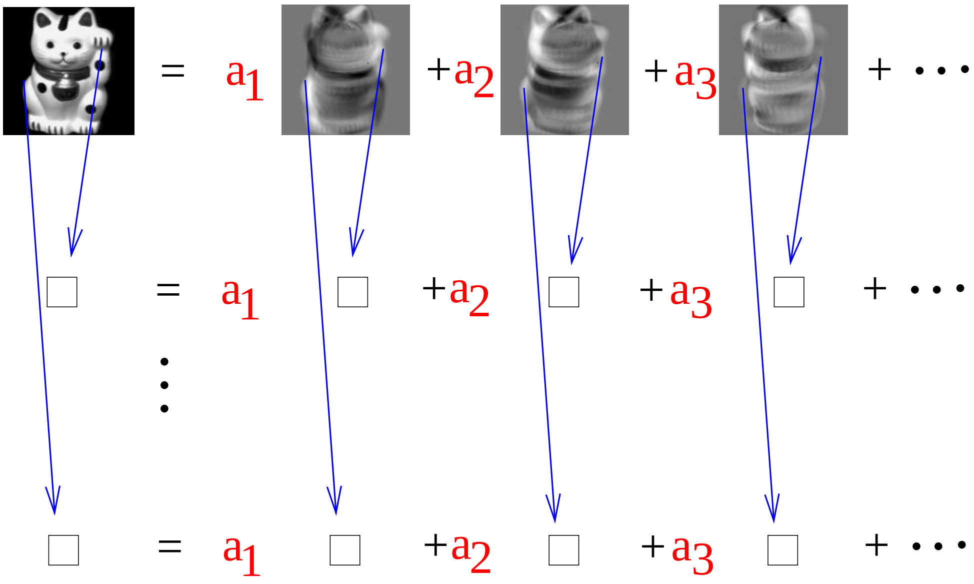


Figure 3: Illustration of using linear equations to calculate the coefficients of eigenimages.

However, if we approximate each template only by a linear combination of a subset of eigenimages, e.g., $p < n$, and there is also noise present in the data, then Eq. (7) can no longer be used, but rather we have to solve an overconstrained system of equations in the least squares sense using k data points ($p < k << m$). Thus we seek the solution vector $\mathbf{a}$ which minimizes

$$E(\mathbf{r}) = \sum_{i=1}^{k} (x_{r_i} - \sum_{j=1}^{p} a_j(\mathbf{x}) e_{j_{r_i}})^2 \ . \tag{8}$$

Of course, the minimization of Eq. (8) can only produce correct values for coefficient vector $\mathbf{a}$, if the set of points r_i does not contain outliers, i.e, not only extreme noisy points but also points belonging to different backgrounds or some other templates due to occlusion. Therefore, the solution has to be sought in a robust manner. In particular, we randomly select a set of points and then iteratively, based on the error distribution, reduce their number, which gives us the final solution of the minimum least squares problem, Eq. (8). Technical details and robustness analysis are given in Appendix A.

The coefficients a_i are then used to create a hypothesis $\tilde{\mathbf{x}}$, which is evaluated both from the point of view of the error, where we can expect for good matches an error of $\sum_{i=p+1}^{n} \lambda_i$ on the average, and from the number of compatible points. The evaluation is based on the backprojection using Eq. (3), which yields an error vector $\vec{\xi} = (\mathbf{x} - \tilde{\mathbf{x}})$. The points which are within an error margin Θ form a set of *compatible* points according to hypothesis $\tilde{\mathbf{x}}$ generated from $\mathbf{a}$. We treat a hypothesis as acceptable if it contains a minimal number of compatible points [2]. The accepted hypothesis is characterized by the coefficient vector $\mathbf{a}$, the error vector $\vec{\xi}$, and the domain of the compatible points $D = \{j | \xi_j^2 < \Theta\}$, $s = |D|$.

However, one can not expect that every initial randomly chosen set of points will produce a good hypothesis if there is one, despite the robust procedure. Thus, to further increase the robustness of the hypotheses generation step, i.e., increase the probability of detecting a correct hypothesis if there is one, we initiate a number of trials (see Appendix A). To efficiently search for subsets of data points to initialize the hypotheses one can use a data-driven masking technique [16]. This leads to a possibly redundant set of accepted hypotheses, which is then resolved by the selection procedure. Fig. 4 depicts some of the hypotheses generated. One can see, that most of the hypotheses are close to the correct solution, however not all hypotheses are valid.

3.2 Selection

The set of hypotheses which has been generated is usually highly redundant. Thus, the selection procedure has to select a subset of "good" hypotheses and reject the superfluous ones. To achieve this, we utilize the Minimum Description Length principle, which leads to the minimization of an objective function encompassing the information about the competing hypotheses [17].

The objective function has the following form:

$$F(\mathbf{h}) = \mathbf{h}^T \mathbf{C} \mathbf{h} = \mathbf{h}^T \begin{bmatrix} c_{11} & \cdots & c_{1R} \\ \vdots & & \vdots \\ c_{R1} & \cdots & c_{RR} \end{bmatrix} \mathbf{h} \ . \tag{9}$$

Vector $\mathbf{h}^T = [h_1, h_2, \ldots, h_R]$ denotes a set of hypotheses, where h_i is a *presence-variable* having the value 1 for the presence and 0 for the absence of the hypothesis i in the resulting description. The diagonal terms of the matrix $\mathbf{C}$ express the cost-benefit value for a particular hypothesis i

$$c_{ii} = \mathrm{K}_1 s_i - \mathrm{K}_2 ||\vec{\xi}_i||_{D_i} - \mathrm{K}_3 N_i \ , \tag{10}$$

where s_i is the number of compatible points, $||\vec{\xi}_i||_{D_i}$ is the error over the domain D_i, and N_i is the number of coefficients (eigenvectors). The coefficients K_1, K_2, and K_3, which can be determined automatically [17], adjust the contribution of the three terms. The off-diagonal terms handle the interaction between the overlapping hypotheses

$$c_{ij} = \frac{-\mathrm{K}_1 |D_i \cap D_j| + \mathrm{K}_2 \xi_{ij}}{2} \ , \quad \xi_{ij}^2 = \max(\sum_{D_i \cap D_j} \vec{\xi}_i, \sum_{D_i \cap D_j} \vec{\xi}_j) \ , \tag{11}$$

[2] This condition can really be kept minimal since the selection procedure will reject the false positives.

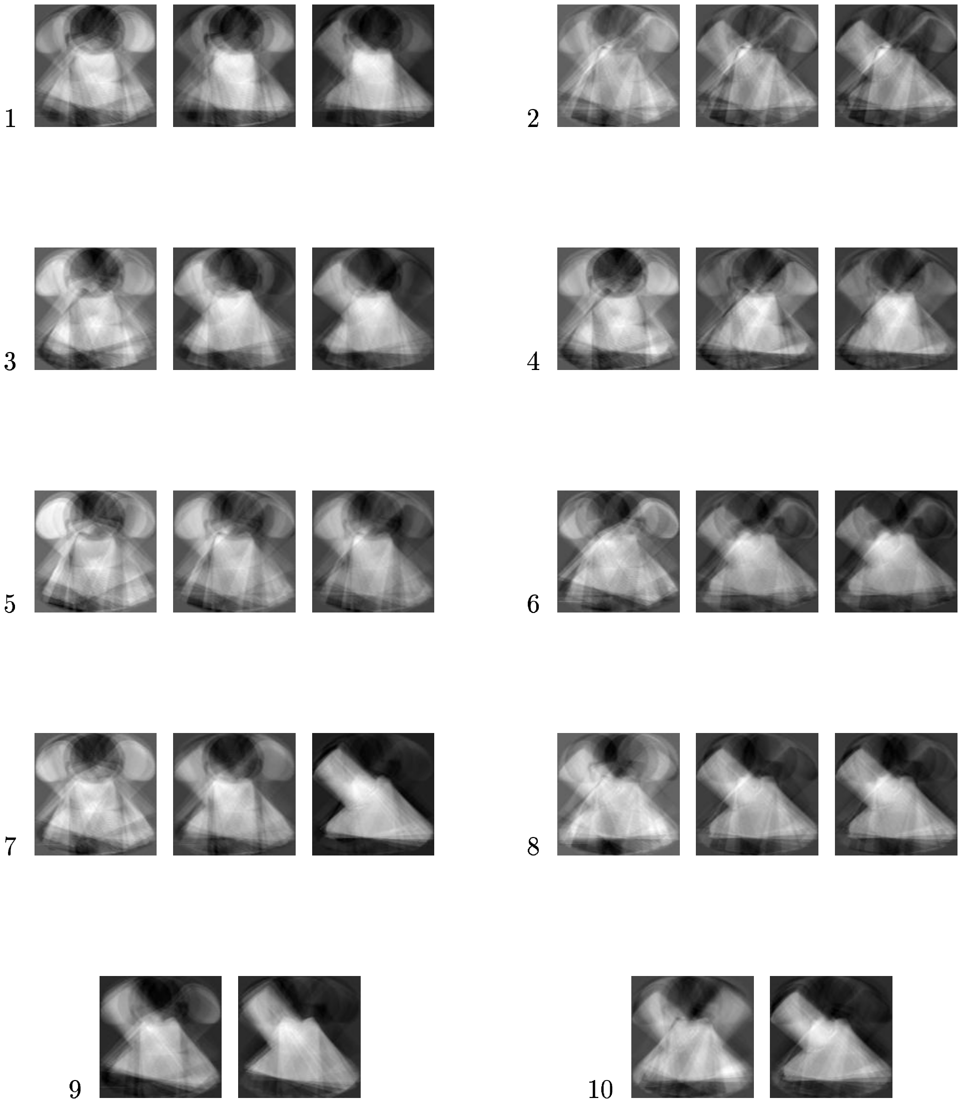


Figure 4: Some hypotheses generated by the robust method for the occluded image, Fig. 2b, with a limited number of eigenimages (for each hypothesis (1–10): reconstructed image based on initial parameters, after the first iteration, after the second iteration).

where D_i denotes the domain of the i-th hypothesis and $\sum_{D_i \cap D_j} \vec{\xi}_i$ denotes the sum of squared errors of the i-th hypothesis over the intersection of the two domains D_i, D_j.

We have formulated the problem of selection in such a way that its solution corresponds to the global extremum of the objective function. We are currently using two different methods for optimization. One is a simple greedy algorithm and the other one is Tabu search [18, 16].

One should note that the selection mechanism can be considerably simplified, when we know that there is a single object in the image or multiple non-overlapping objects. In these cases only the diagonal terms need to be considered. However, for multiple overlapping objects[3] the optimization function has to be used in its full generality (see Fig. 15).

3.3 Fitting

The selected (also the generated) hypotheses are based only on a small set of points, from which the coefficients have been computed. In order to increase the accuracy of the coefficient vector **a**, we use a (robustified) least squares approach [19], similar to the one used in the hypotheses generation step. The principle is simple: Based on the statistics of the errors of *all* compatible points, a decision is made (again related to the error margin Θ) which data points should participate in a standard least squares computation of **a**. This is repeated until the convergence is reached. Since for the selected hypotheses the initial estimates of **a** are usually quite accurate, only a few iterations are necessary for the process to converge. We could have done this already at the time of generating a hypothesis, however this would unnecessarily introduce additional computational complexity.

3.4 Final selection

Since the previous fitting step might have altered some of the hypotheses, a final selection is in some cases needed. The selection is performed in the same manner as explained in section 3.2. Note that this step is computationally not very expensive and cost only a small fraction of time compared to the first three steps, since it is applied only to a small subset of the set of all initial hypotheses.

3.5 Complete algorithm

The complete algorithm is shown in Fig. 5. One should note that we can simultaneously deal with multiple eigenspaces $\Gamma_1, \ldots, \Gamma_P$ of, for example, different objects (we call them *families*). Everything remains as described, except that at each location all of the families are used to initiate the hypotheses. The selection procedure then reasons among different hypotheses, possibly belonging to different families, and selects those that better explain the data.

Once we obtain the coefficients, to achieve the recognition we have to search for the closest point in the eigenspace. For the search we use the exhaustive search procedure implemented in SLAM [20].

The robust method performs the calculation of the coefficients as they were continuous variables. However, since our goal is object recognition (orientation estimation) we are interested only in coefficients coming from a small part of the eigenspace (i.e., discrete points, or points on a parametric manifold). We can exploit this constraint to reduce the computational complexity of the robust method as discussed in the next section.

3.6 Robust Constrained Method

In the robust constrained method we exploit the constraint that the true coefficients are either discrete points or points on a parametric manifold. Thus, instead of solving (8) iteratively to convergence, we iterate (8) only once, and then immediately search for the closest point in the eigenspace or on the parametric manifold, respectively, which gives us the coefficients of the closest training image (or an interpolation between several of these coefficients). Of course, in order to increase the robustness we have to, as in the previous case 3.5, explore multiple hypotheses, which is then followed by the selection procedure. Fig. 7 depicts some of the generated hypotheses. One can see that the majority of hypotheses are close to the correct solution, however not every hypothesis is a valid one. Since we have already a point on the parametric manifold, there is no need to perform the fitting and the final selection step of the original robust method (see Fig. 6). In section 4 we compare the speed and accuracy of the robust method versus the robust constrained method.

[3] In the above formulation we consider only the pairwise overlaps in the final solution.

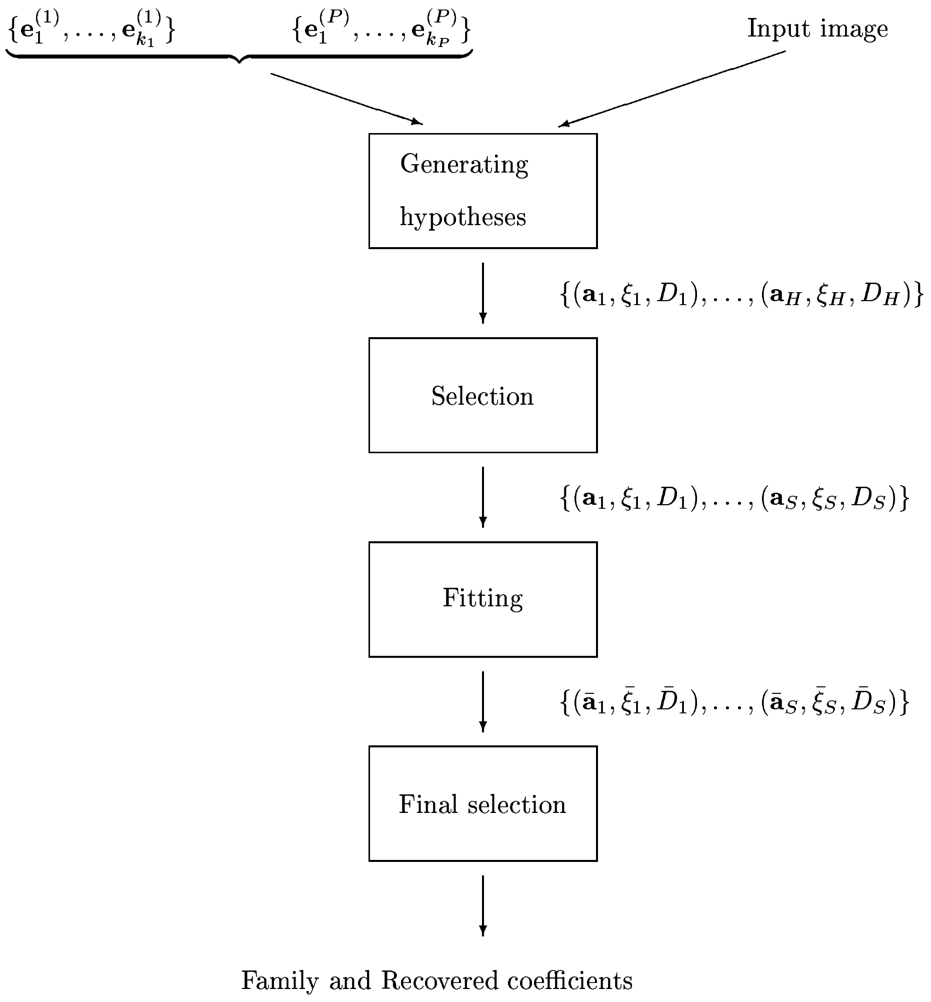

$\underbrace{\{\mathbf{e}_1^{(1)},\ldots,\mathbf{e}_{k_1}^{(1)}\} \quad \{\mathbf{e}_1^{(P)},\ldots,\mathbf{e}_{k_P}^{(P)}\}}$
Input image
Generating hypotheses
$\{(\mathbf{a}_1,\xi_1,D_1),\ldots,(\mathbf{a}_H,\xi_H,D_H)\}$
Selection
$\{(\mathbf{a}_1,\xi_1,D_1),\ldots,(\mathbf{a}_S,\xi_S,D_S)\}$
Fitting
$\{(\bar{\mathbf{a}}_1,\bar{\xi}_1,\bar{D}_1),\ldots,(\bar{\mathbf{a}}_S,\bar{\xi}_S,\bar{D}_S)\}$
Final selection
Family and Recovered coefficients

Figure 5: A schematic diagram outlining the complete algorithm.

4 Experimental results—1D Case

In order to visualize the main steps of the algorithm we first apply it to 1-D functions. We start with generating four families of functions:

- $y = a$
- $y = bx + c$
- $y = dx^2 + ex + f$
- $y = g\cos(x + h)$

For each of these families 200 training examples are generated by systematically varying their parameters $(a \ldots h)$. Fig. 8 shows some of the training samples for the trigonometric family.

With these samples the four eigenspaces are calculated (we use all eigenvectors with $\lambda_i \neq 0$). Fig. 9 shows the obtained eigenvectors.

We test our algorithm on various test-signals like the ones shown in Fig. 10. Note that these test-signals contain only portions of the original functions and are therefore not recoverable by the standard eigenspace approach (as has been demonstrated in Fig. 1).

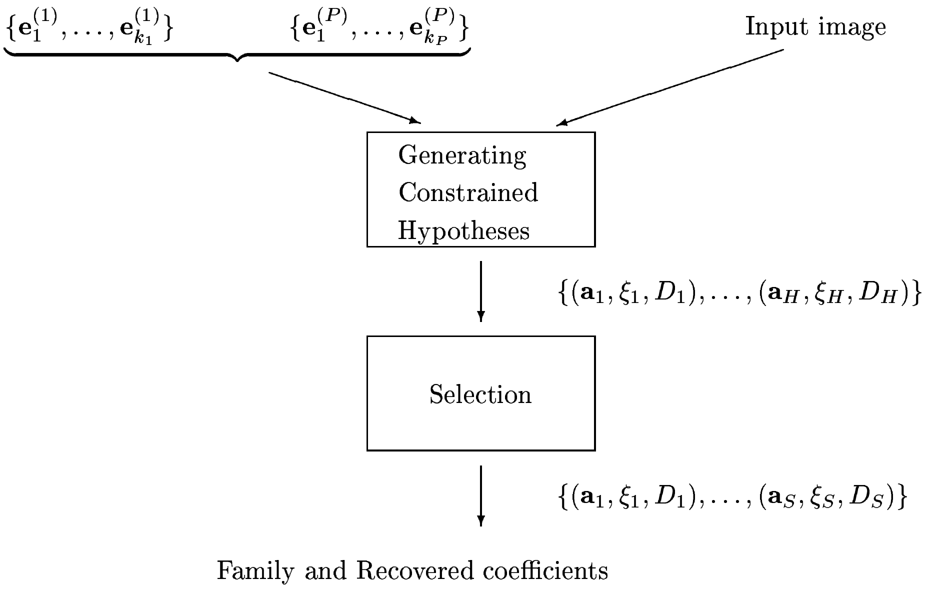

$\{\mathbf{e}_1^{(1)},\dots,\mathbf{e}_{k_1}^{(1)}\}$
$\{\mathbf{e}_1^{(P)},\dots,\mathbf{e}_{k_P}^{(P)}\}$
Input image
Generating Constrained Hypotheses
$\{(\mathbf{a}_1,\xi_1,D_1),\dots,(\mathbf{a}_H,\xi_H,D_H)\}$
Selection
$\{(\mathbf{a}_1,\xi_1,D_1),\dots,(\mathbf{a}_S,\xi_S,D_S)\}$
Family and Recovered coefficients

Figure 6: A schematic diagram outlining the complete robust constrained algorithm.

Fig. 11 shows the four steps of our algorithm applied to the test-signal depicted in Fig 10a. The members of all four families are perfectly recovered, as depicted in Fig. 11d. One can also see from this example that the fitting step did not change the hypotheses very much, therefore the final selection has not changed anything.

Fig. 12 shows the four steps of the algorithm in the case of a noisy signal. In addition we have not included the family of quadratic functions. Therefore one gets for the quadratic part an approximation consisting of two constant functions. This example demonstrates that our algorithm not only successfully deals with partial signals but also tolerates noise.

5 Experimental Results—2D Case

In this section we first present several single experiments to confirm the utility of our robust method. In the next subsection we report on extensive testing that we performed to compare the standard method with the robust and the robust constrained method. We performed all experiments on the standard set of images (Columbia Object Image Library, COIL-20) [14]. Fig. 13 shows all 20 objects in the COIL-20. Each object is represented by 72 images obtained by the rotation of the object through 360 degrees at 5 degree steps (1440 images in total).

Unless stated otherwise all experiments are performed with following parameters:

Number of eigenimages p	15
Number of hypotheses H	8
Number of initial points k	$10p = 150$
Reduction factor r	0.7
Stopping criterion s	$3p = 45$
K_1	1
K_2	0.1
K_3	50
Compatibility Threshold Θ	50

Fig. 14 demonstrates that our approach is insensitive to occlusions. One can see that both robust methods outperform the standard method considerably. The visual reconstruction from the robust con-

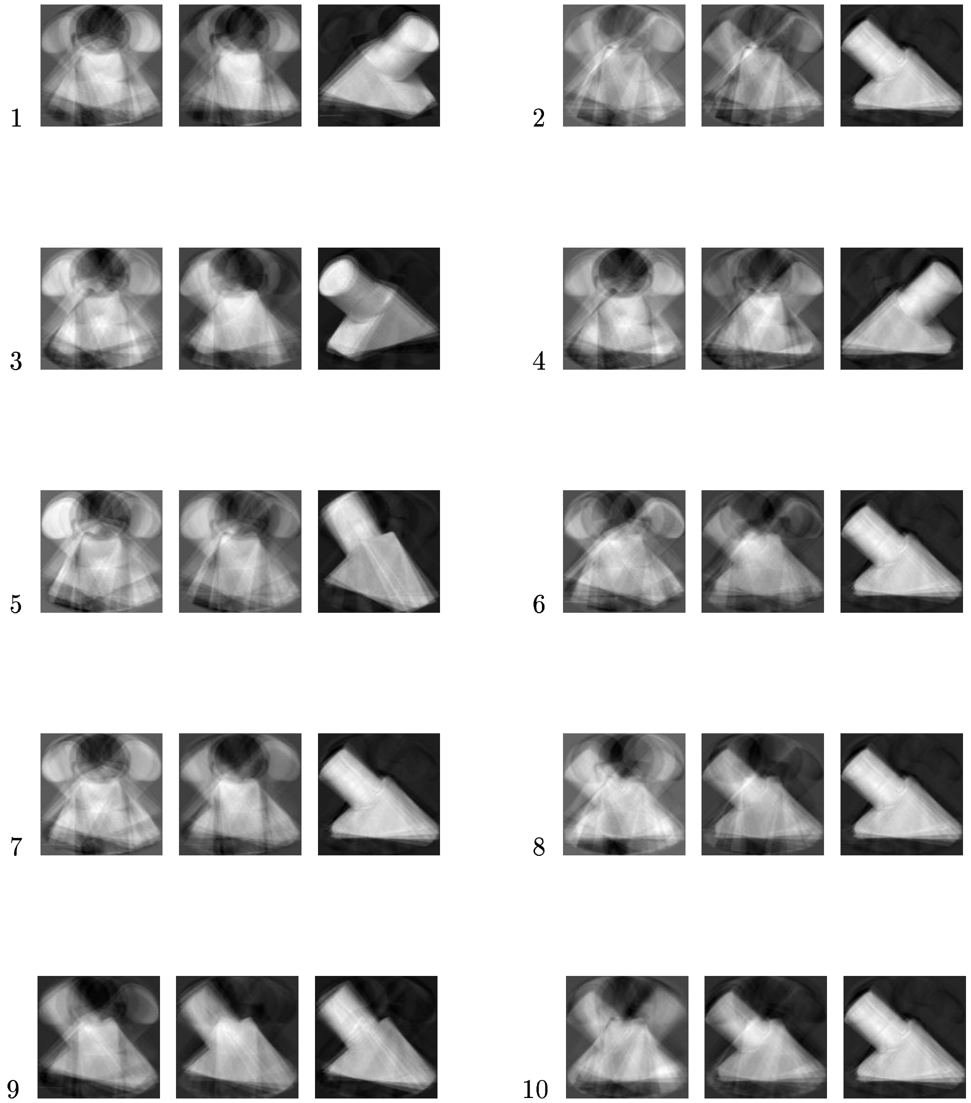


Figure 7: Some hypotheses generated by the robust constrained method for the occluded image, Fig. 2b with a limited number of eigenimages (for each hypothesis (1–10): reconstructed image based on initial parameters, after the first iteration, reconstructed image based on the parameters of the closest point on the parametric manifold).

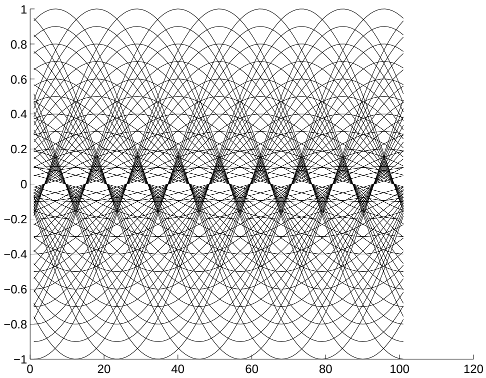
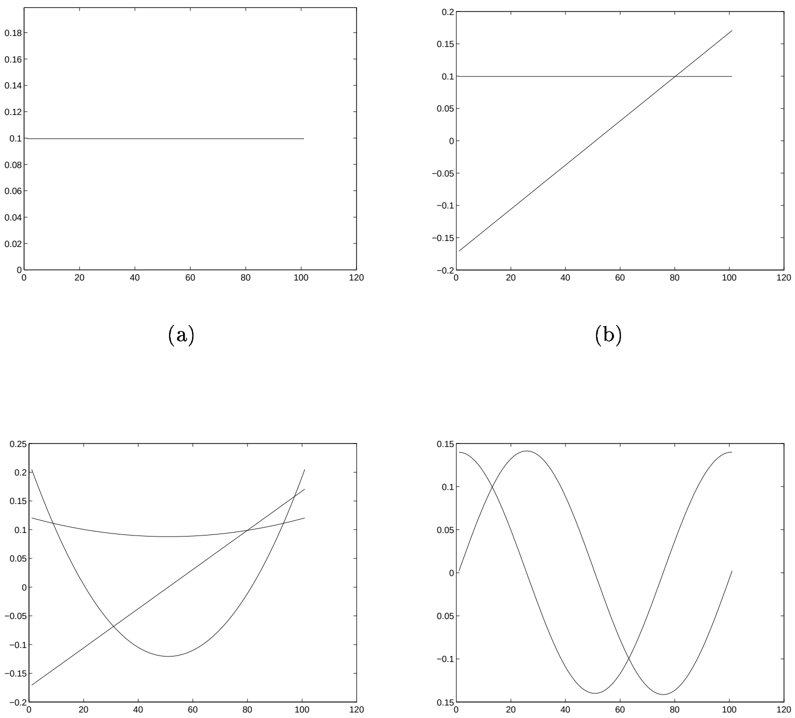


Figure 8: Training samples: Trigonometric function.



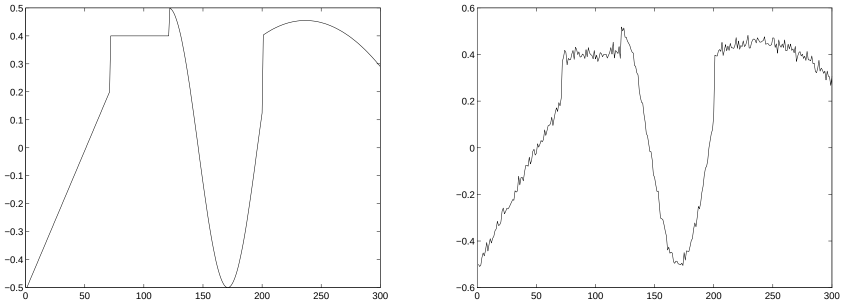
(a)

(b)

(c)

(d)

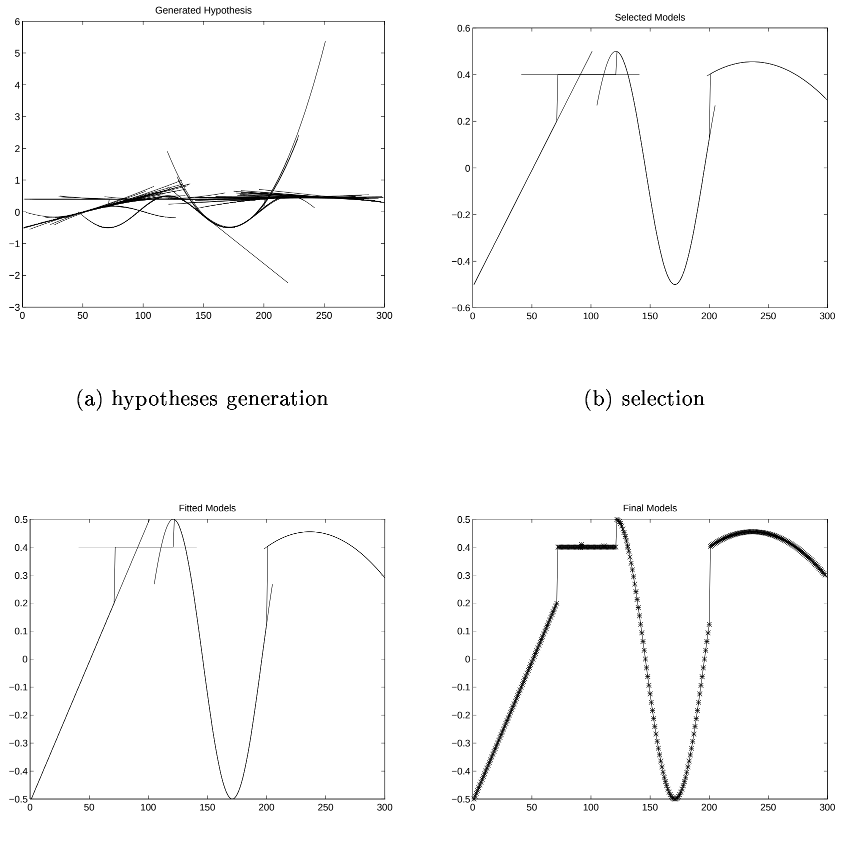
Figure 9: 1-D eigensignals: (a) constant, (b) linear, (c) quadratic, and (d) trigonometric.



(a)

(b)

Figure 10: 1-D test-signals: (a) noiseless, (b) signal with added Gaussian noise.



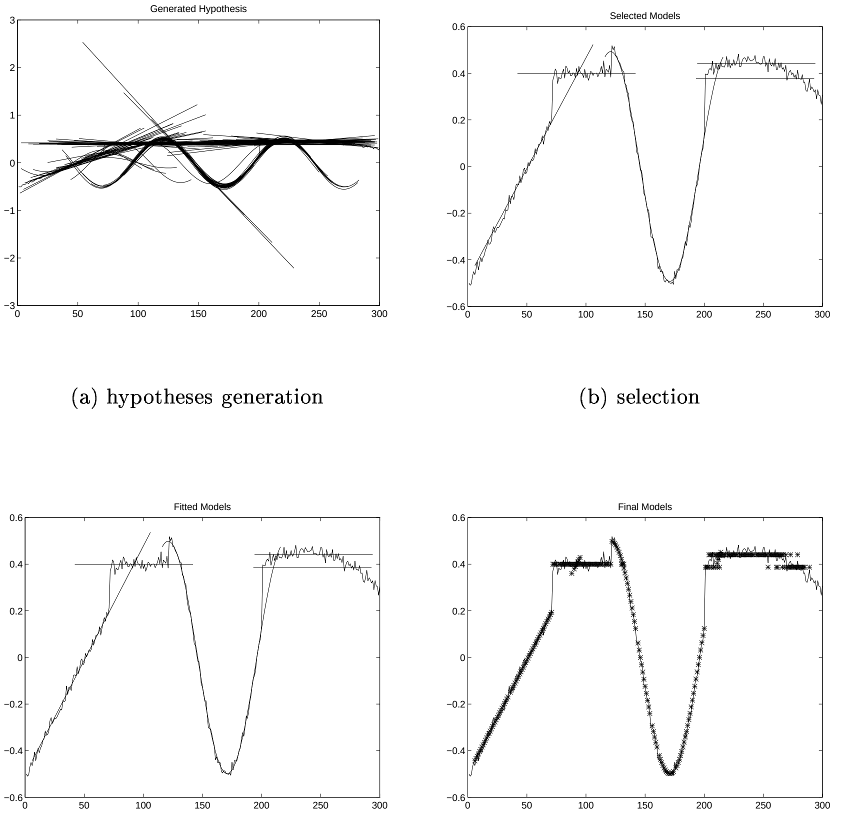
(a) hypotheses generation

(b) selection

(c) fitting

(d) final selection

Figure 11: The four steps of our algorithm shown on a 1-D signal. In Figs. (a–c) the complete extent of hypotheses is depicted while in Fig. (d) only the compatible points are plotted.



(a) hypotheses generation

(b) selection

(c) fitting

(d) final selection

Figure 12: The four steps of our algorithm shown on a noisy 1-D signal. In Figs. (a–c) the complete extent of hypotheses is depicted while in Fig. (d) only the compatible points are plotted.

strained method is better because we get the exact point on the manifold. Note that the blur visible in the reconstruction is the consequence of taking into account only a limited number of eigenimages.

Fig. 15 demonstrates that our approach can cope with situations when one object occludes another. One can see that both robust methods are able to recover both objects. One should note that in this case the selection mechanism based on the MDL principle delivers automatically that there are two objects present in the scene (i.e. we do not need to specify the number of objects in advance).

Fig. 16 shows several objects on a considerably cluttered background. All objects have been correctly recovered by both robust methods.

5.1 Comparison of the Methods

In this section we report on the extensive testings that we performed to compare the three approaches, namely the standard method, the robust method, and the robust constrained method. We compared the three methods on the level of recognition and not on the level of estimated coefficients. Therefore, the results have to be treated with caution, since sometimes the recovered coefficients are not optimal (especially with the standard method), but still the object is correctly recognized (see Fig. 17).

The tests have been performed under the following conditions (see also Fig. 18):

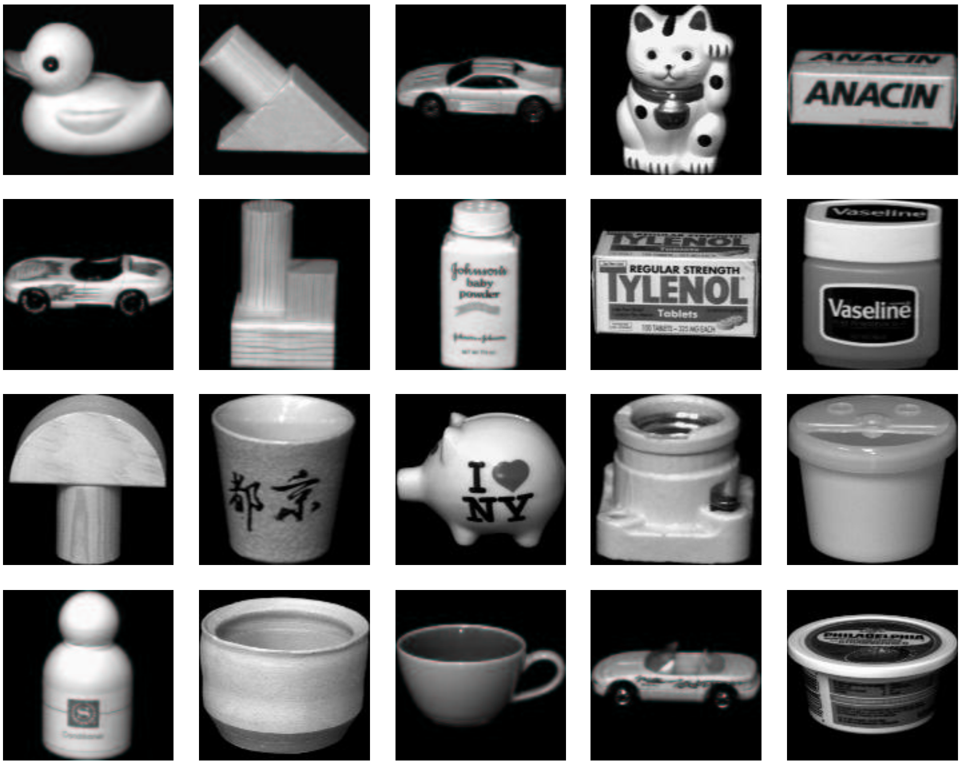
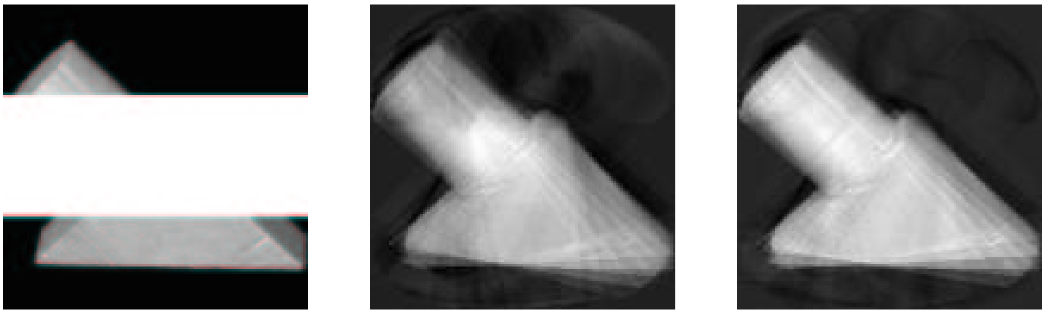


Figure 13: 20 test objects used in the experiments.

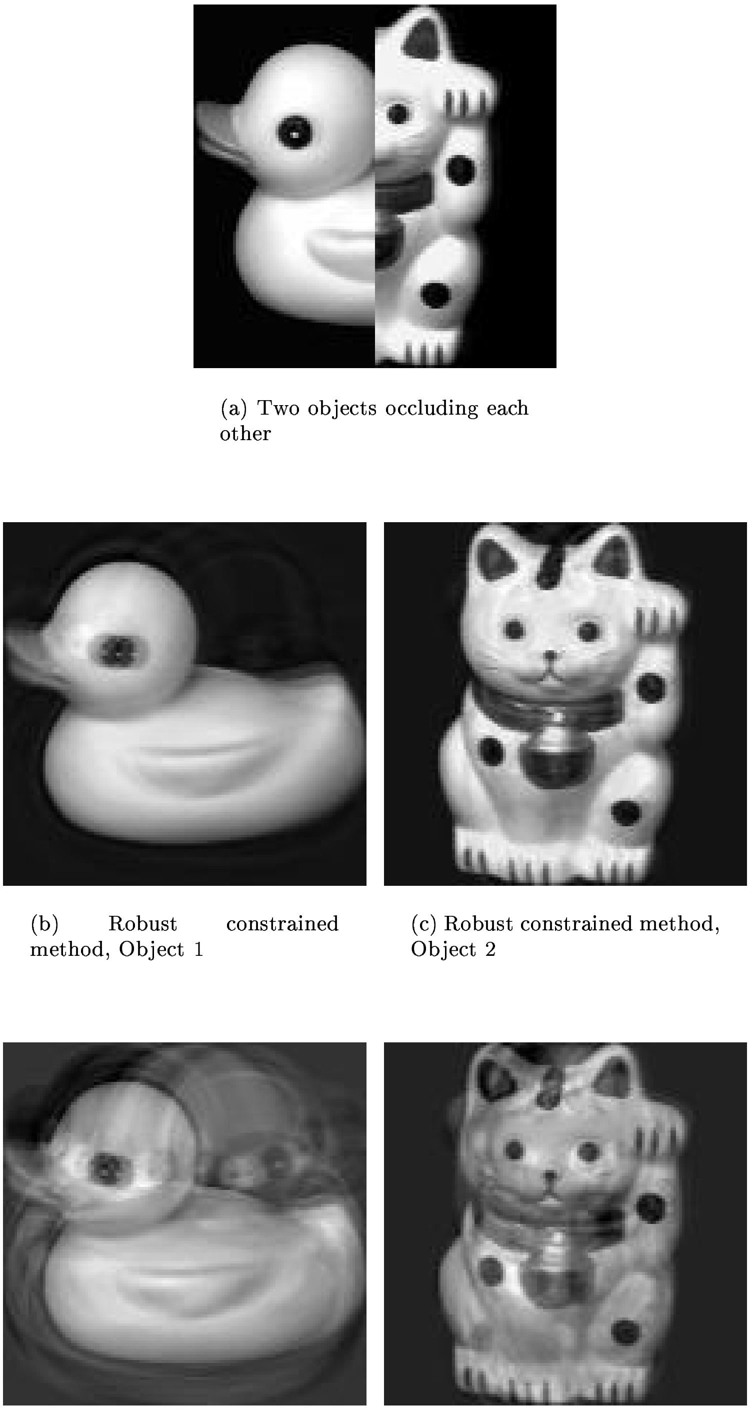


(a) Occluded image

(b) Reconstructed image, robust method

(c) Reconstructed image, robust constrained method

Figure 14: Demonstration of insensitivity to occlusions using the robust methods for calculating the coefficients a_i (compare also with Fig. 2).



(a) Two objects occluding each other

(b) Robust constrained method, Object 1

(c) Robust constrained method, Object 2

(d) Robust method, Object 1

(e) Robust method, Object 2

Figure 15: Two objects occluding each other.

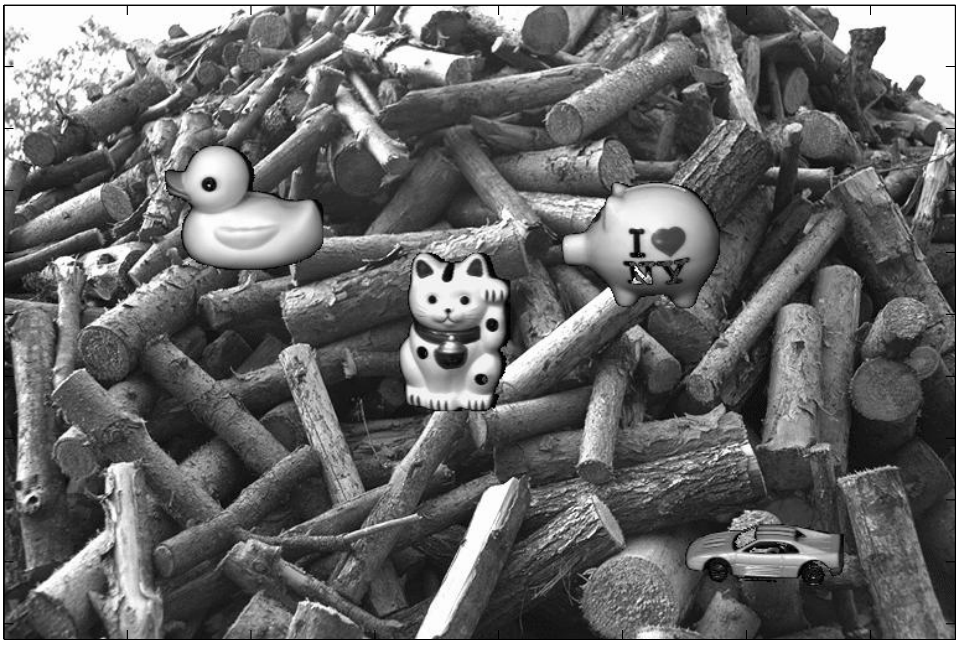


Figure 16: Testobjects on cluttered background.

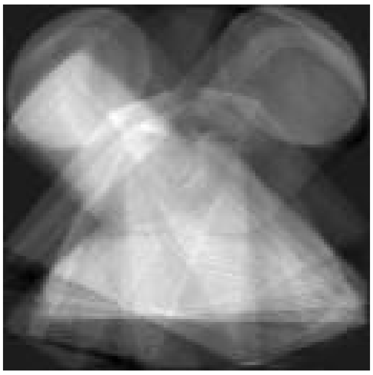
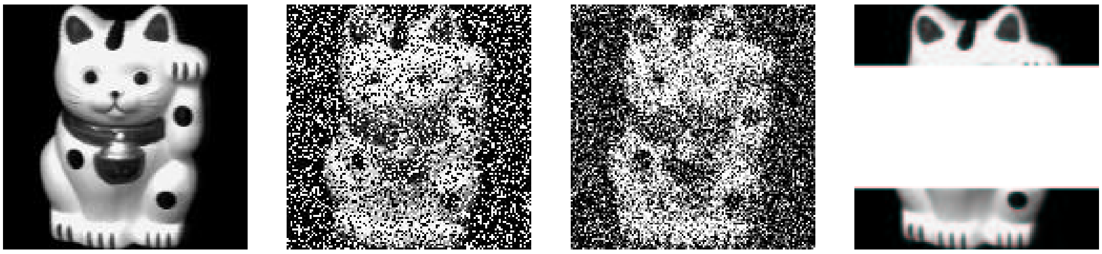


Figure 17: Reconstructed object using the coefficients determined by the standard method. Although the coefficients are far away from their true values, a correct recognition is achieved.

- Additive Salt & Pepper noise (0% – 75%),
- Additive Gaussian noise ($\sigma = 0$ – $\sigma = 300$), values less then 0 and greater than 255 are clipped to 0 and 255, respectively,
- Occlusions (0% – 60%).

We performed two types of experiments:

- Orientation estimation and
- classification.



(a) Original image

(b) 50% Salt & Pepper noise

(c) Gaussian noise $\sigma = 150$

(d) 50% Occlusion

Figure 18: Test image subjected to various types of noise and occlusion.

5.1.1 Orientation Estimation

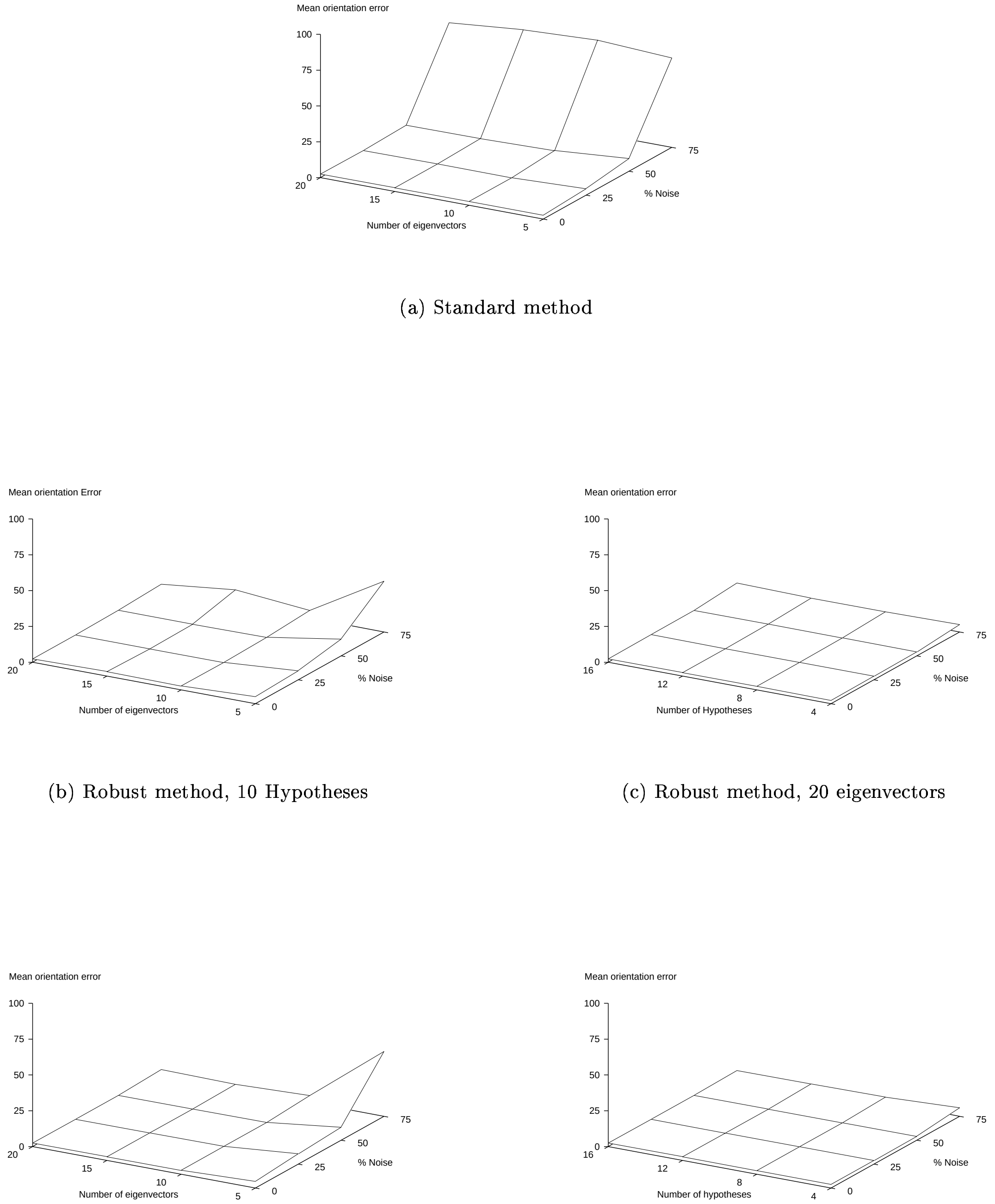
For this set of experiments we constructed the eigenspace of a single object with images from 36 orientations (each 10° apart), and used the remaining views to test the generalization ability of the methods. In addition we have varied the number of eigenvectors and the number of hypotheses. As a performance measure we used the absolute orientation error. The following plots (Figs. 19, 20, 21), show the typical results of the three methods obtained for the test set of one object under the various noise conditions (there is no significant difference between training and test images).

Fig. 19 shows that both robust methods are very robust against salt and pepper noise, only in the case of a very small number of eigenimages the results become noise sensitive. The case of Gaussian noise Fig. 20 shows that the standard method performs well (which is not surprising because it is the optimal method for Gaussian noise). For high values of σ most grey-values are already clipped, therefore we have no longer real Gaussian noise. In the case of robust constrained method we do not do any additional fitting of compatible points and thus the averaging effect is limited to a reduced set of initial data points. Thus, in the case of Gaussian noise we do not achieve good accuracy. In the case of robust method additional fitting based on all compatible points improves the accuracy. Fig. 20 clearly demonstrates this point. Fig. 21 shows the results on occluded objects, which again demonstrates the superiority of the robust methods. The higher errors can be explained by the fact, that for a high percentage of occlusion some objects already disappear in some orientations (compare also to Fig. 23).

The experiments have shown that the robust methods outperform the standard method considerably when dealing with outliers and occlusion. Table 1 summarizes the results for several objects. To generate the table, we have chosen for each method those parameter settings where the best results have been obtained averaged over all objects. The error measure is the rounded absolute orientation error given in degrees.

Table 1: Summary of orientation estimation experiments.

Method	Salt & Pepper [%]				Gaussian Noise [σ]				Occlusions [%]			
	0	25	50	75	75	150	225	300	15	30	45	60
Standard	2	3	3	48	3	3	4	24	3	25	31	45
Robust	3	3	3	5	3	4	7	15	3	3	23	36
Robust Constr.	2	3	3	4	4	5	6	10	3	3	16	29



(a) Standard method

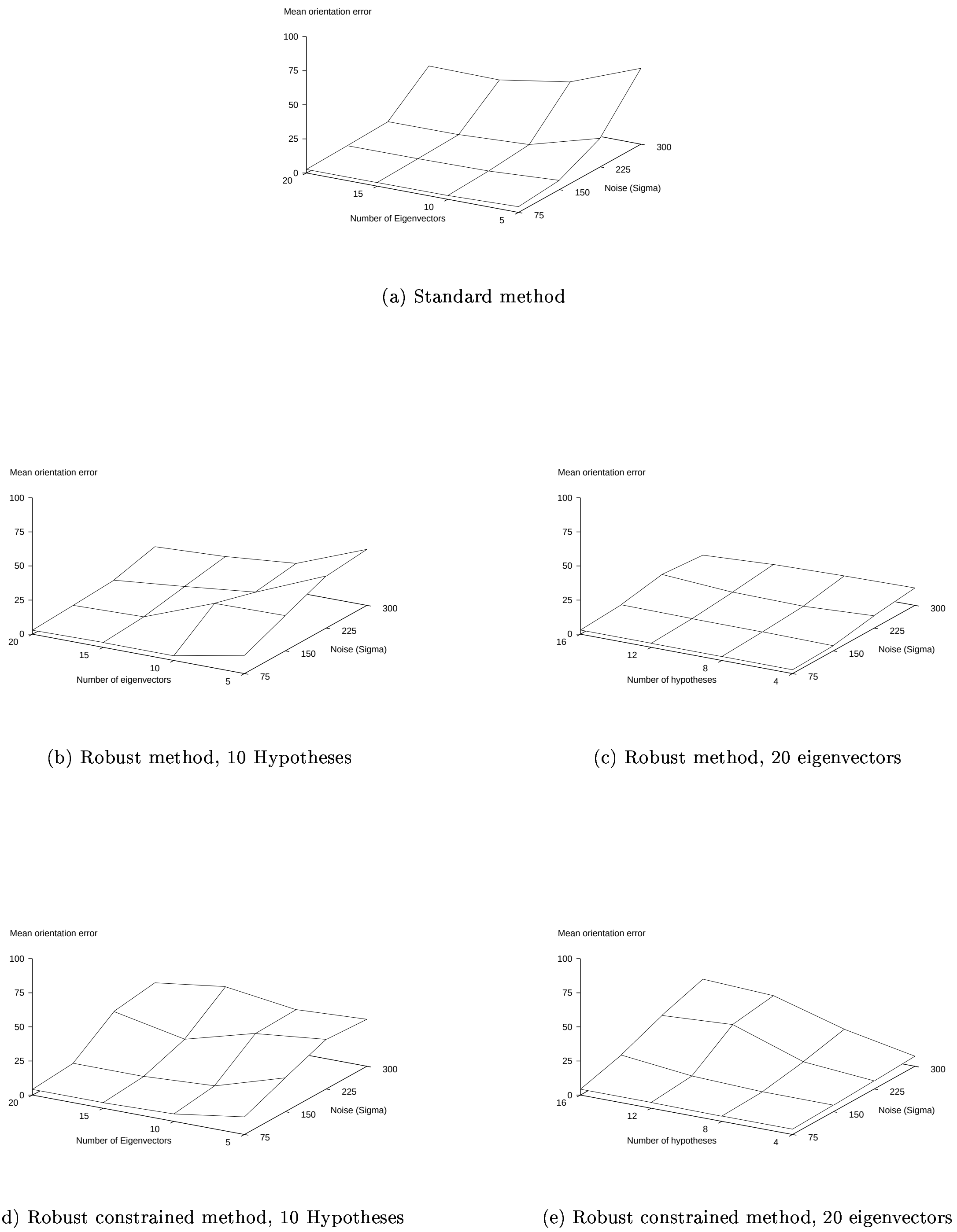
(b) Robust method, 10 Hypotheses

(c) Robust method, 20 eigenvectors

(d) Robust constrained method, 10 Hypotheses

(e) Robust constrained method, 20 eigenvectors

Figure 19: Orientation estimation results in the case of Salt & Pepper noise.



(a) Standard method

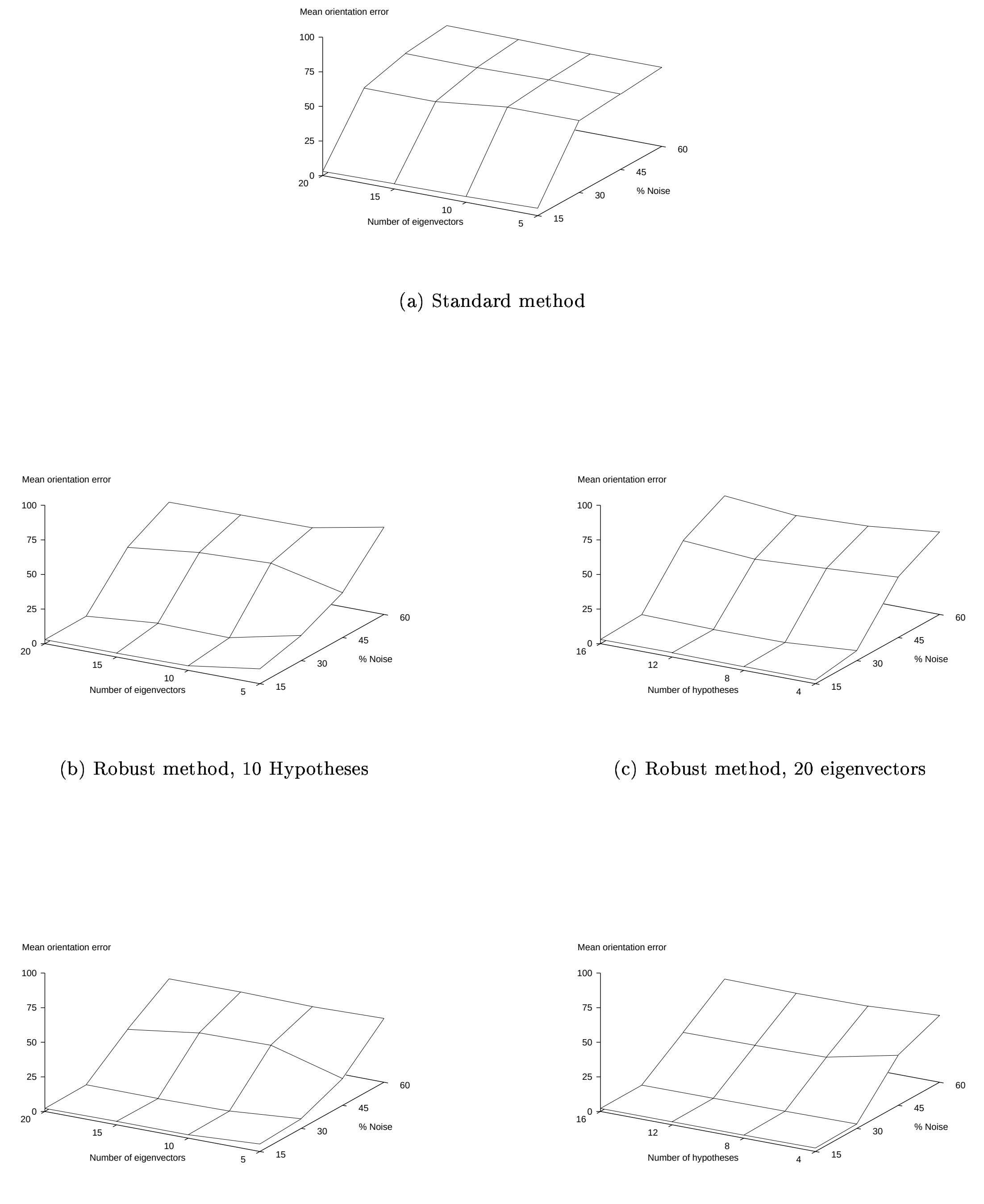
(b) Robust method, 10 Hypotheses

(c) Robust method, 20 eigenvectors

(d) Robust constrained method, 10 Hypotheses

(e) Robust constrained method, 20 eigenvectors

Figure 20: Orientation estimation results in the case of Gaussian noise.



(a) Standard method

(b) Robust method, 10 Hypotheses

(c) Robust method, 20 eigenvectors

(d) Robust constrained method, 10 Hypotheses

(e) Robust constrained method, 20 eigenvectors

Figure 21: Orientation estimation results in the case of occlusions.

5.1.2 Classification

For this set of experiments we have used all 20 objects. As a training set we used, similarly to the previous experiment, images from 36 orientations of each object. For the standard method, we calculated the universal eigenspace [1] with all training images, and when the object was correctly recognized, we used the objects' eigenspace to determine the orientation. For the robust methods, we used only the objects' eigenspaces, and in the hypotheses generation stage, we generated for each object 8 hypotheses. For all eigenspaces, 15 eigenvectors were used. As error measures we used the classification accuracy and for those objects which have been correctly recognized, we also calculated the mean absolute error in the orientation.

Table 2 shows the results for 50% Salt and Pepper noise, and Table 3 shows the results for 50% occlusions.

Table 2: Classification results on images disturbed by 50% Salt and Pepper noise.

Method	Recognition Rate	Mean absolute orientation error
Standard	46 %	22°
Robust	72 %	12°
Robust Constrained	75 %	6°

Table 3: Classification results on images 50% occluded.

Method	Recognition Rate	Mean absolute orientation error
Standard	12 %	57°
Robust	30 %	45°
Robust Constrained	66 %	29°

These results clearly indicate the superiority of the robust methods over the standard method. The higher error rates for the occlusion can be explained by the fact, that certain objects are already completely occluded in some orientations. In the Figures 22 and 23 we have depicted those objects that caused the highest error, in either orientation estimation or classification.

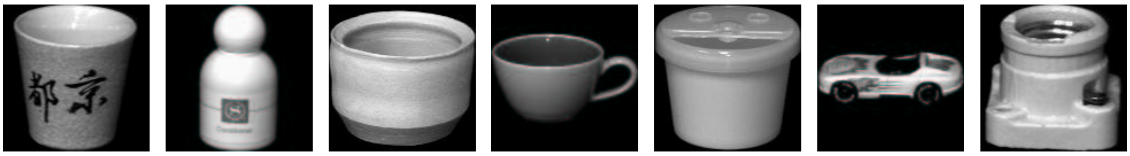


Figure 22: Gross errors in the orientation determination are mainly caused by rotation-symmetric objects.

In summary, these experiments demonstrate that our robust methods can tolerate considerable amount of noise, can cope with occluded, non-segmented multiple objects, and are therefore much wider applicable

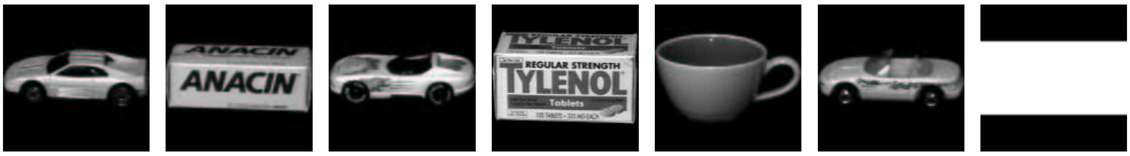


Figure 23: Gross errors in the classification are mainly caused by objects, hardly visible under occlusion.

than the standard method.

6 Discussion and Conclusions

In this paper we have presented a novel robust approach which enables the appearance-based matching techniques to successfully cope with outliers, cluttered background, and occlusions. The robust approach exploits several techniques, e.g., robust estimation and hypothesize-and-test paradigm, which combined together in a general framework achieve the goal. We have presented two robust methods. The robust constrained method is computationally less demanding and does not compromise the robustness. The robust methods have been incorporated into the SLAM package [20]. We have presented an experimental comparison of the robust methods and the standard one on a *standard database* of 1440 images. We identified the "breaking points" of different methods and demonstrated the superior performance of our robust methods. A general conclusion drawn from these experiments is as follows: The robust methods can tolerate much higher noise levels than the standard parametric eigenspace method under reasonable computational cost. In terms of speed is the standard method approximately 7 times faster than the robust constrained method, which is in turn 3 times faster than the robust method. However, in the case when we have "well-behaved" noise with a low variance in the images, then we can show, that the robust methods are faster (≈ 20 times in the robust constrained case) than the standard method. This is, because with well-behaved noise we do not need to explore many hypotheses, and we do not need to perform the selection and the back-projection of the coefficients. Also the number of initial points (see equation (8)) can be significantly reduced. What should be emphasized is that in this case the robust methods are independent of the size of the image.

It is interesting to note that the basic steps of the proposed algorithm, namely hypotheses generation, selection, fitting, and the final selection, are the same as in ExSel++ [16], which deals with robust extraction of *analytical* parametric functions from various types of data. Therefore, the method described in this paper can also be seen as an extension of ExSel++ to *learnable classes* of parametric models.

The applications of the proposed methods are numerous. Basically everything that can be performed with the classical appearance-based methods can also be achieved within the framework of our approach, only more robustly and in more complex scenes.

The proposed robust approach is a step forward, however, some problems still remain. In particular, the method is sensitive to scale, and the application of the method on a large image is computationally very demanding. In [21] we have recently demonstrated how the robust methods can be incorporated in a hierarchical framework. We have also shown how to use simple detectors to find possible candidate-points to initiate the search for objects. In addition, we have recently started to experiment with algorithms, which learn which points to select for generating hypotheses. All these methods speed up the whole process considerably without compromising the robustness of the method.

A Robust fitting

In this Appendix we first demonstrate the robustness of the way how we solve (Eq. 8), and based on that we derive how many hypotheses need to be generated in order to guarantee at least one correct estimate with a certain probability.

Starting from k points $r_1 \dots r_k$ we seek the solution vector $\mathbf{a} \in \mathrm{I\!R}^p$ which minimizes

$$E(\mathbf{r}) = \sum_{i=1}^{k}(x_{r_i} - \sum_{j=1}^{p} a_j(\mathbf{x}) e_{j_{r_i}})^2 \ . \tag{12}$$

Based on the error distribution of the set of points, we keep reducing their number by a factor of r until either all points are either within the compability threshold Θ, or the number of points is smaller than s. Therefore, the parameters are:

k	starting number of points (e.g. $k = 10p$)
r	reduction factor
s	stopping criterion (e.g. $s = 3p$).

Since it is hard to derive an analytic expression for this highly non-linear problem, we tested the robustness using a simulated Monte-Carlo approach. The procedure was as follows: We generated eigenimages from a set of test images, p eigenimages were used for projection. The true coefficients were determined by projection of the test images on the eigenspace. Then for various parameter settings and different levels of replacement noise (i.e., a point r_i is selected at random, and its value is replaced by a uniform random number between $[0 \dots 255]$) the coefficients where determined, and the distance between the recovered and the true coefficients was plotted. Fig. 24 shows a plot of the parameter setting $p = 16, k = 12p, r = 0.75, s = 3p$. One can see that more than 50% of noise can be tolerated by our method. The reason for this is that the noise amplitude deviations are limited to $[0 \dots 255]$, which is the case in images quantized to eight bits. Monte-Carlo simulations have been performed with the following ranges of parameter values: $k \in \{4p \dots 20p\}; r \in [0.4, 0.9]; s \in \{2p \dots 5p\}$. The robustness behavior of solving the equations is similar to Fig. 24 for a wide range of parameter values, i.e. $k > 7p, 0.5 < r < 1, 2p < s < 4p$. Since these parameters influence also the computational complexity of the algorithm, the parameters $k = 12p, r = 0.75, s = 3p$ are a good compromise between robustness and computational complexity.

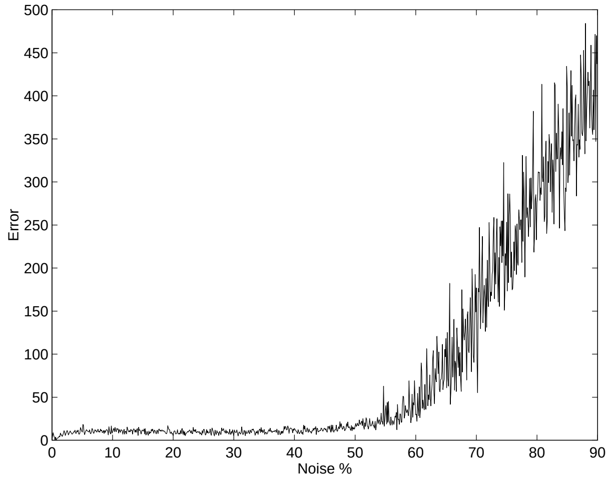


Figure 24: Monte-Carlo simulation showing the robustness of solving the equations.

A.1 How many hypotheses

Having determined the amount of noise that can be tolerated by solving the equations (denoted by τ) we can now calculate the number of hypotheses H that need to be generated for a given noise level ζ in order to find at least one good hypothesis with probability η. The derivation is straight forward:

When we generate one hypothesis the probability to find a good one is:

$$\rho(\text{good hypo}) = \rho(\text{from } k \text{ points at most } k\tau \text{ are outliers}) \ ,$$

where $\rho(a)$ denotes the probability of event a. Therefore,

$$\rho(\text{good hypo}) = \sum_{i=0}^{m=\lfloor k\tau \rfloor} \binom{k}{i} \zeta^i (1-\zeta)^{k-i} \ .$$

Now we generate H hypotheses to satisfy the following inequality $\rho(\text{at least one good hypo}) > \eta$

$$1 - (1 - \sum_{i=0}^{m=\lfloor k\tau \rfloor} \binom{k}{i} \zeta^i (1-\zeta)^{k-i})^H > \eta \ ,$$

which gives us the required number of hypotheses

$$H > \frac{\log(1-\eta)}{\log(1 - \sum_{i=0}^{m=\lfloor k\tau \rfloor} \binom{k}{i} \zeta^i (1-\zeta)^{k-i}).}$$

Figs. 25 and 26 demonstrate the behavior of H graphically, when the number of equations is $k = 150$. One can clearly see that as long as the amount of noise is within the range of the noise tolerance of solving the equations we need only a few hypotheses, however as soon as we have $\approx 5\%$ more noise, than can be tolerated by solving the equations, we would need to generate a huge number of hypotheses to guarantee to find at least one good hypothesis with probability η. Therefore, we can conclude that only a few (< 5) hypotheses need to be explored when the noise level is within the required bounds, a fact which has also been demonstrated by the experimental results.

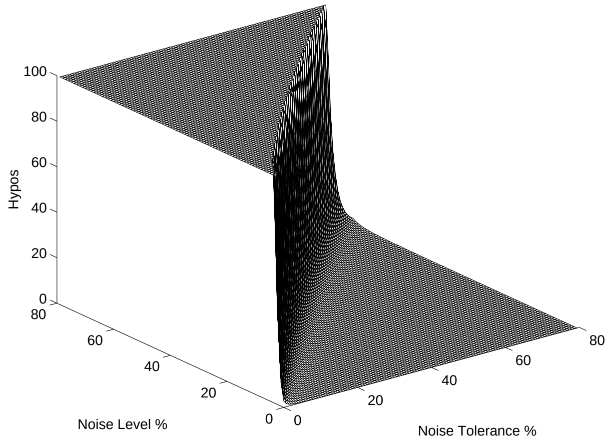


Figure 25: Plot of H versus noise and noise tolerance of solving the equations; values larger than 100 are set to 100.

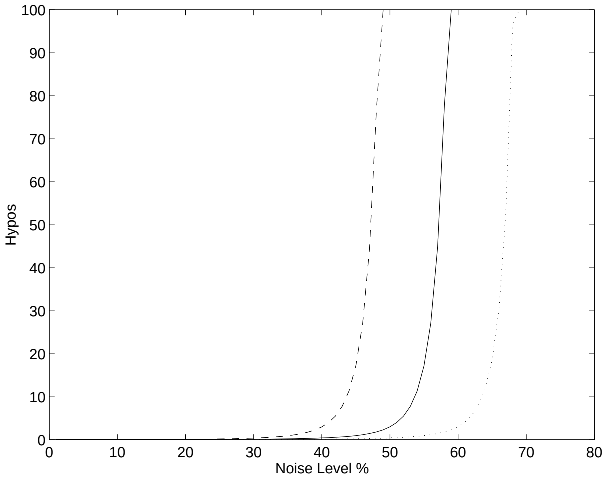


Figure 26: Plot of H versus noise the noise tolerance of solving the equations was 0.4, 0.5, 0.6; values larger than 100 are set to 100.

References

[1] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[2] B. W. Mel, "SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition," *Neural Computation*, vol. 9, no. 4, pp. 777–804, 1997.

[3] H. Murase and S. Nayar, "Illumination planning for object recognition using parametric eigenspaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1219–1227, 1994.

[4] S. K. Nayar, H. Murase, and S. A. Nene, "Learning, positioning, and tracking visual appearance," in *IEEE International Conference on Robotics and Automation*, (San Diego), May 1994.

[5] S. Yoshimura and T. Kanade, "Fast template matching based on the normalized correlation by using multiresolution eigenimages," in *Proceedings of IROS'94*, pp. 2086–2093, 1994.

[6] H. Murase and S. K. Nayar, "Image spotting of 3D objects using parametric eigenspace representation," in *The 9th Scandinavian Conference on Image Analysis* (G. Borgefors, ed.), vol. 1, (Uppsala, Sweden), pp. 323–332, June 1995.

[7] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[8] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proceedings of 5th ICCV'95*, IEEE Computer Society Press, 1995.

[9] A. Pentland, B. Moghaddam, and T. Straner, "View-based and modular eigenspaces for face recognition," Tech. Rep. 245, MIT Media Laboratory, 1994.

[10] P. J. Huber, *Robust Statistics.* New York: Wiley, 1981.

[11] P. J. Rousseuw and A. M. Leroy, *Robust Regression and Outlier Detection.* New York: Wiley, 1987.

[12] R. Rao, "Dynamic appearance-based recognition," in *CVPR'97*, IEEE Computer Society, 1997.

[13] M. Black and A. Jepson, "Eigentracking: Robust matsching and tracking of articulated objects using a view-based representation," in *ECCV96*, pp. 329–342, Springer, 1996.

[14] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Tech. Rep. CUCS-005-96, Columbia University, New York, 1996.

[15] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* New York: Wiley, 1958.

[16] M. Stricker and A. Leonardis, "ExSel++: A general framework to extract parametric models," in *6th CAIP'95* (V. Hlavac and R. Sara, eds.), no. 970 in Lecture Notes in Computer Science, (Prague, Czech Republic), pp. 90–97, Springer, September 1995.

[17] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation of range images as the search for geometric parametric models," *International Journal of Computer Vision*, vol. 14, no. 3, pp. 253–277, 1995.

[18] F. Glover and M. Laguna, "Tabu search," in *Modern heuristic techniques for combinatorial problems* (C. R. Reeves, ed.), pp. 70–150, Blackwell Scientific Publications, 1993.

[19] M. Stricker, "A new approach for robust ellipse fitting," in *Proc. of the Third International Conference on Automation, Robotics and Computer Vision*, vol. 2, (Singapore), pp. 940–945, November 1994.

[20] S. Nene, S. Nayar, and H. Murase, "SLAM: Sofware Library for Appearance Matching," Tech. Rep. CUCS-019-94, New York: Columbia University, Department of Computer Science, Sept. 1994.

[21] A. Leonardis and H. Bischof, "Computational complexity reduction in eigenspace approaches (in press)," in *Proc. CAIP 97*, Springer Verlag, 1997.