Published in proceedings of 3DIM'01, pages 171-178, 2001.

Robust Recognition and Pose Determination of 3-D Objects Using Range Images in Eigenspace Approach*

Danijel Skočaj and Aleš Leonardis
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, SI-1001 Ljubljana, Slovenia
{danijel.skocaj, ales.leonardis}@fri.uni-lj.si

Abstract

In this paper we propose a robust method for recognition and pose determination of 3-D objects using range images in the eigenspace approach. Instead of computing the coefficients by a projection of the data onto the eigenimages, we determine the coefficients by solving a set of linear equations in a robust manner. The method efficiently overcomes the problem of missing pixels, noise and occlusions in range images. The results show that the proposed method outperforms the standard one in recognition and pose determination.

1. Introduction

Recognition and pose determination of objects is a common problem in computer vision. A number of different approaches have been proposed (for overviews see [7, 5]).

Until recently, the objects to be recognized were most commonly modeled as 3-D object-centered representations. These representations varied in form from polygonal meshes, parametric surfaces, clouds of 3-D points, etc. Recognition and pose determination were usually performed by first extracting different features from the image of an unknown object and then matching these features to the ones associated with the models. Different features have been proposed and explored, e.g., silhouettes, edges, surface curvatures, local shapes, point features etc. [7]. This approach, which is primarily based on 3-D object-centered representations, includes two steps which prove to be difficult: first, building consistently registered 3-D object-centered model, and second, matching features of the unknown object to features in the object representation and calculating the transformation.

It turns out that these two steps, which may be unnecessary for the recognition, can be circumvented. This can be achieved by modeling objects as a set of views captured during a systematic observation of the object. The recognition can then be performed *directly* with an unknown view. This is known as view-based recognition. This approach of direct matching would be prohibitive (computationally and in terms of space) unless the views were compressed in a compact representation and the matching was performed in an efficient way. One approach is to compress the set of views in an *eigenspace* representation built from all images of objects (training images) using principal component analysis [11, 4]. Recognition of the image of an unknown object is then performed by projecting that image into the eigenspace and finding the nearest projected training image.

Campbell and Flynn [6] studied view-based approach for recognising objects from range images by testing the eigenspace approach on two databases of range images. They generated range images from full 3-D models, therefore their range images can be considered as ideal.

Due to the architecture of range image sensors, range images are usually not ideal [12, 13]. They contain *missing pixels* – pixels where depth can not be recovered. Amano et. al. [1] addressed the problem of missing pixels in the eigenspace approach. They achieved good results with their method, but the proposed solution is computationally very inefficient.

We propose a method which efficiently overcomes the problem of missing pixels. In addition, it overcomes also the problems caused by noise and occlusions. The approach has been originally designed for recognising objects from intensity images [9], but a modified approach, tailored to this specific domain, produces desired results also in the case of range images.

The paper is organised as follows: We first review the basic concepts of the view-based methods. In Section 3 we present the method for estimating eigenspace parameters of a range image containing missing pixels. In Section 4 we extend this method to handle not only missing pixels but

*This work was supported by a grant from the Ministry of Science and Technology of Republic of Slovenia (Project J2-0414).

also occlusions and outliers in range images. The experimental results are shown in Section 5. We conclude with a summary.

2. Overview of the eigenspace approach

The view-based methods consist of two stages. In the first, off-line (training) stage, a set of training images is obtained. These images have to encompass all possible appearances of the object.

Since range images usually do not contain many high frequencies, the images are highly correlated. Thus, they can efficiently be compressed using principal component analysis (PCA) [2], resulting in a low-dimensional eigenspace.

In the second, on-line (recognition) stage, an input image is projected to the eigenspace. The recognition is achieved by finding the nearest projected training image.

In a multiple-object recognition system a universal eigenspace is generated from all images of all objects. In addition, for each object an object eigenspace from all training images of that object is generated. In the recognition stage, first the universal eigenspace is used to recognize the object. Then the corresponding object eigenspace is used to determine the pose.

We now introduce the notation. We treat each range image as a 1-D vector. Let $\mathbf{y} = [y_1, ..., y_m]^T$ be a training image, and let $\mathcal{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ be a set of training images. To simplify the notation we assume $\mathcal{Y}$ to be normalised, 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 training images n is much smaller than the number of elements m in each training image; thus an efficient algorithm based on SVD can be used to calculate the first n eigenvectors [10]. Since the eigenvectors form an orthogonal basis system, $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 1$ when $i = j$ and 0 otherwise. We assume that the eigenvectors are in descending order with respect to the corresponding eigenvalues λ_i. Then, depending on the correlation among the training images 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 $\mathbf{e}_i$,

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

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 of a range image $\mathbf{x}$ is projected onto the eigenspace,

$$a_i(\mathbf{x}) = \langle \mathbf{x}, \mathbf{e}_i \rangle = \sum_{j=1}^{m} x_j e_{i,j} \ \ , \ \ 1 \le i \le p \ \ . \tag{2}$$

$a(\mathbf{x}) = [a_1(\mathbf{x}), ..., a_p(\mathbf{x})]^T$ 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

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

3. Handling missing pixels

When range images do not contain any missing data, the approach described above produces good results [6]. However, range images which are captured using a range sensor usually contain also *missing pixels*. These are pixels which belong to the object, however their depth value could not be obtained.

Missing pixels are caused by imperfections of range sensors. Namely, range sensors are usually sensitive to the surface properties of the object [12, 13]. Therefore, shadows, reflections, non-uniform albedo and texture can cause missing pixels in the range image. This is evident in Fig. 1 where an intensity image and the corresponding range image containing missing pixels are shown.

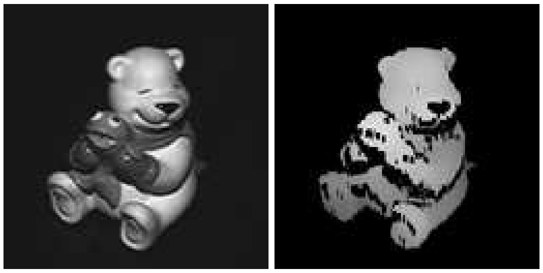


Figure 1. Intensity image of an object and range image containing missing pixels.

We make an assumption that range images in the training set do not contain missing pixels. These range images are produced in the training stage, when the computation time is not of primary concern. Therefore, these range images can be build in such a way that they do not contain missing pixels. On the other hand, during the recognition stage, the computation time can be very important. Furthermore, different types of range sensors, even low-cost ones, can be used to produce range images for the recognition phase. Thus, it is very likely that such range images contain missing pixels.

Let us analyse how the missing pixels affect the calculation of coefficients of a range image. Without loss of generality let us suppose that last $m-r$ pixels in a range image are missing pixels, thus $\hat{\mathbf{x}} = [x_1, ..., x_r, 0, ..., 0]$. Then

$$\hat{a_i} = \hat{\mathbf{x}}^T \mathbf{e}_i = \sum_{j=1}^{r} x_j e_{i,j} \quad . \tag{4}$$

The error we make in calculating a_i is

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

It follows that the reconstruction error is

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

Due to the nonrobustness of linear processing, this error affects the whole vector $\mathbf{x}$. Fig. 2 depicts the effect of missing pixels on the reconstructed image. Fig. 2a) depicts a range image without missing pixels while the Fig. 2b) depicts the same range image with some pixels turned to missing pixels. Fig. 2c) shows the reconstructed image b), which is far from the ideal image a).

If we take into account all eigenvectors, i.e., $p = n$, then computing coefficients by a projection of the data onto the eigenimages (Eq. 2) is equivalent to solving an over-constrained system of linear equations

$$x_i = \sum_{j=1}^{n} a_j(\mathbf{x}) e_{j,i} \ , \ \ i = 1 \dots m \tag{7}$$

in a least-squares sense. Thus, for calculating coefficients a_j all m pixels in a range image are used, including missing pixels, which leads us from the correct solution.

Since the size of a range image (number of pixels) is usually much bigger than the number of eigenimages, we can exclude missing pixels from the computation by living out from the set of equations Eq. 7 all equations where $x_i = 0$. Thus, if in the range image $\mathbf{x}$ k pixels are not equal to 0 ($\mathbf{r} = (r_1, ..., r_k); x_{r_i} \neq 0$) we can calculate p coefficients a_j by solving an over-constrained system of k linear equations

$$x_{r_i} = \sum_{j=1}^{p} a_j(\mathbf{x}) e_{j,r_i} \ , \ \ i = 1 \dots k \ . \tag{8}$$

Therefore, we seek the solution vector $\mathbf{a}$ which minimizes

$$E(\mathbf{r}) = \sum_{i=1}^{k} \left(x_{r_i} - \sum_{j=1}^{p} a_j(\mathbf{x}) e_{j,r_i} \right)^2 \quad . \tag{9}$$

Since depth is also not defined in the background pixels, they are excluded from the computation too. However, the background pixels define the border of the object, which is very useful for the recognition. Because of this, we perform the second step to refine the coefficients a_i. We reconstruct the image $\tilde{\mathbf{x}}$ using Eq. 3 and select all pixels from $\mathbf{x}$ where the reconstruction error $|\tilde{x}_i - x_i|$ is consistent with the distribution of the reconstruction errors of the pixels in $\mathbf{r}$. Then we add the newly selected pixels to $\mathbf{r}$ and recalculate coefficient vector $\mathbf{a}$, minimizing Eq. 9. These two steps are iterated until the convergence is reached, which happens after only a few iterations.

The coefficients obtained using the proposed algorithm are very insensitive to noise. Fig. 2d) depicts the reconstructed image from Fig. 2b) which is quite close to the ideal image. As we will show in Section 5, the recognition results are excellent even if the range image contains 80% of missing pixels.

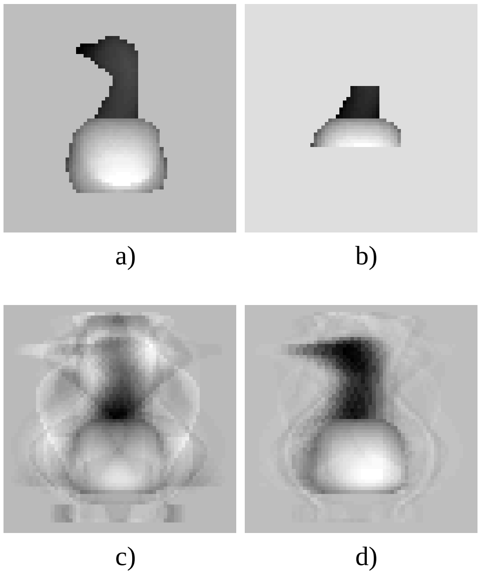


Figure 2. a) Range image without missing pixels, b) range image with missing pixels, c) reconstructed image using standard method and d) reconstructed image using robust method.

4. Handling noise and occlusions

In this section we extend the method to handle not just missing pixels but also outliers and occlusions. The main idea remains the same: not to use all pixels from a range image to obtain the coefficients, but only the “good” ones. The problem is that we do not know in advance which pixels are reliable and which are not.

Therefore, we randomly chose the set of points $\mathbf{r}$ and apply the following robust procedure to solve Eq. 9. Starting from randomly selected k points $r_1,\ldots,r_k$, we seek the solution vector $\mathbf{a}$ which minimizes Eq. 9 in a least-squares manner. Then, based on the error distribution of the set of points, we reduce their number by a factor of α (we exclude those points with the largest error) and solve Eq. 9 again with this reduced set of points. We repeat this procedure until the size of the point set $\mathbf{r}$ is reduced to a predefined number of points s.

After that, we want to include in the computation of coefficients a_i all compatible points in the same way as in the case of missing pixels. We reconstruct the image $\tilde{\mathbf{x}}$ and select all pixels from $\mathbf{x}$ where the reconstruction error $|\tilde{x}_i - x_i|$ is consistent with the distribution of the reconstruction errors of the pixels which remained in the set $\mathbf{r}$ after α-trimming. Then we recalculate the coefficient vector $\mathbf{a}$ using all compatible points by minimizing Eq. 9 in a least-squares manner. The obtained coefficients a_i are then used to create a hypothesis $\tilde{\mathbf{x}} = \sum_{i=1}^{p} a_i \mathbf{e}_i$.

Fig. 3 shows the execution steps of the proposed algorithm in the case of a good hypothesis. Figs. 3c)-h) depict the points (in grey) that were used for calculating the coefficient vector from the occluded range image shown in Fig. 3a). As one can observe, in α-trimming phase all the points in the occluded region were eliminated, and then they were not included in the set of compatible points. Therefore, coefficient vector $\mathbf{a}$ was calculated only from "good" points which results in a perfect reconstruction (Fig. 3b).

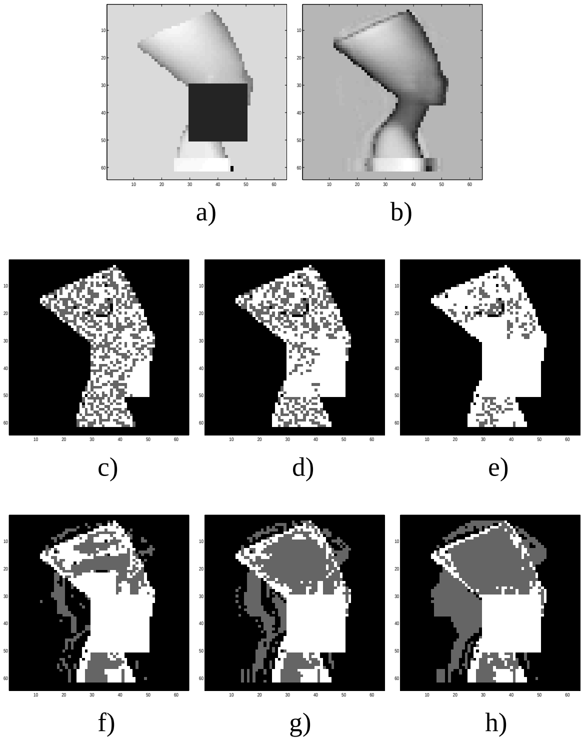
However one cannot expect that every initial randomly chosen set of points will produce a good hypothesis, despite the robust procedure. Thus, to further increase the robustness of the hypothesis generation step, i.e., increase the probability of detecting a correct hypothesis if there is one, we initiate, as in [3, 8, 9], a number of trials. Then, in the selection step, we chose the best hypothesis, i.e., the hypothesis with the smallest reconstruction error of compatible points.

5. Experimental results

In this section we present the results of experiments to demonstrate the performance of our method. We performed the experiments on two sets of images created from six 3-D models. The models (depicted in Fig. 4) were created from real objects by NRC-CNRC Institute for Information Technology[1].

The range images in the first set were generated by rotating the model around one axis in 1° steps (one degree of freedom). From each model 360 range images of the size 64×64 pixels were generated. Twelve range images generated from the first model are shown in Fig. 5.

The range images in the second set were generated by rotating a model around two axes (two degrees of freedom).


a) b)

c) d) e)

f) g) h)

Figure 3. a) Occluded image, b) reconstructed image, c)-e) three α-trimming iterations, f)-h) adding compatible points.

From each model 545 range images of the size 64×64 were created. The range images were generated from the views which were equally spaced in the object's pose space (see Fig. 6). Twelve range images generated from the first model are shown in Fig. 7.

As we mentioned above, the background pixels and the missing pixels are usually set to 0. We rather set this level

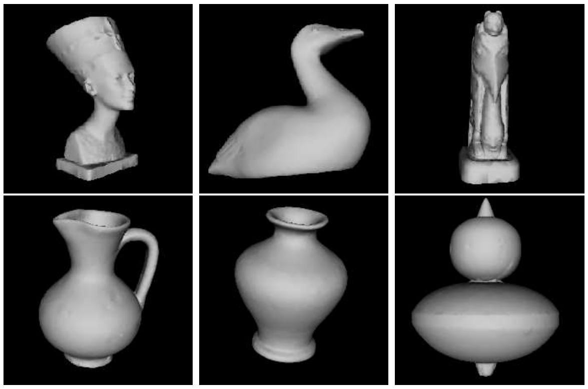


Figure 4. The models used to create range images.

[1] www.vit.iit.nrc.ca/3D/Pages_HTML/3D_Models.html

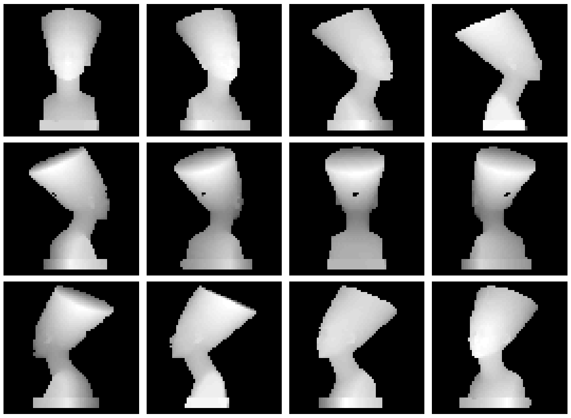


Figure 5. Twelve range images from 1 DOF example.

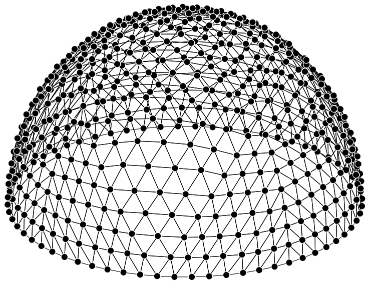


Figure 6. Viewpoints for generating range images in 2 DOF example.

to the mean value of all non-background and non-missing pixels in all images. Since the difference between missing pixels and pixels in the object is smaller, the convergence in robust minimization of Eq. 9 is much faster. Due to the same reason, the recognition results of the standard eigenspace method have also improved.

We tested the performance of our algorithms in the case of missing pixels, noise, and occlusion in range images. We simulated missing pixels by setting a number of pixel values to the level of missing pixels. These pixels were chosen in two ways: *spatially incoherent* – by randomly choosing pixels, and *spatially coherent* – by choosing the pixels concentrated in two image areas. We simulated noise in the image in the same way, except that the levels of the chosen pixels were set to random values. All four types of simulated noise are depicted in Figs. 8a)-d). Note that we can look at spatially coherent noise as an occlusion.

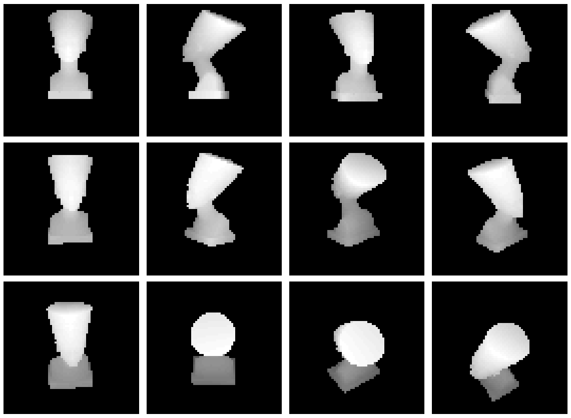


Figure 7. Twelve range images from 2 DOF example.

We first present the results of 1 DOF example. A universal eigenspace was built from 360 range images (60 images for each object). Each object eigenspace was created from 72 range images. All the images that were used were uniformly distributed in the object's pose space. In all tests we used eigenspaces of dimension 15.

First we tested the recognition of the objects. A hundred range images were randomly chosen among all generated range images of all objects and different amounts of missing pixels and noise were added (0–80%). Figs. 8e)-l) show the reconstruction of one such trial. The summary of the recognition rate is plotted in Figs. 9a)-d). It is evident that the robust method outperforms the standard one. The recognition rate of the robust method is very close to 100% even at 50% of noise.

We obtain similar results also in the case of determining the orientation of the objects. Figs. 9e)-h) plot the results of determining the orientation of the first object in Fig. 4. A hundred range images were randomly chosen among all generated range images of the object. The orientation was determined by finding the closest image in the training set (the closest point in the eigenspace). The error was defined as the difference between the orientation of the detected training image and the orientation of the training image that should be detected. For each level of missing pixels and noise the average absolute error was calculated. This error is very close to zero even at 70% of missing pixels, while the standard approach is not useful when a range image contains more than 30% of the noise.

Finally, we present the results of 2 DOF example. A universal eigenspace was built from 654 range images (109 images for each object). Each object eigenspace was created from 109 range images, which were uniformly distributed in the object's pose space. We performed only tests for spa-

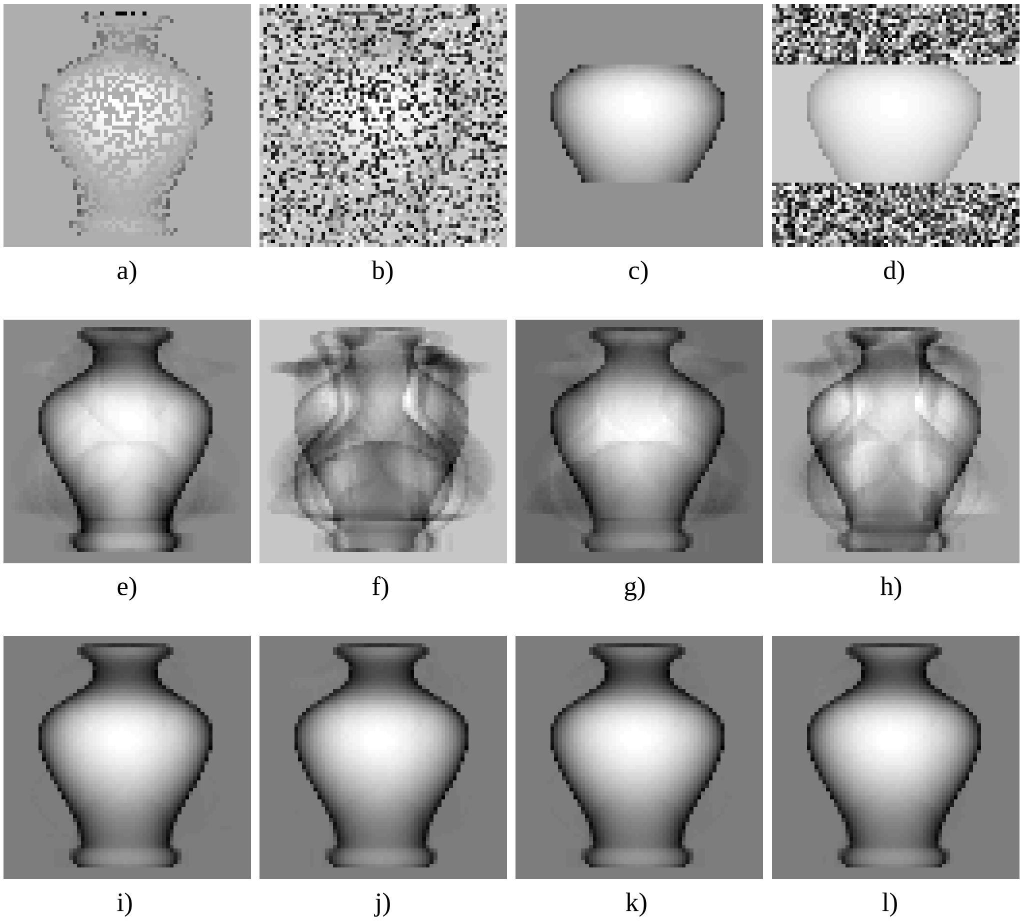


Figure 8. a) Spatially incoherent missing pixels, b) spatially incoherent noise, c) spatially coherent missing pixels, d) spatially coherent noise, e)-h) images reconstructed with the standard method, i)-l) images reconstructed with the robust method.

tially incoherent missing pixels and noise. Here, the measure for the orientation error was the angle between the detected orientation and the orientation of the object in the test image. As depicted in Fig. 10, the recognition rate and the orientation error are almost constant for all amounts of missing pixels and noise from 0–70%.

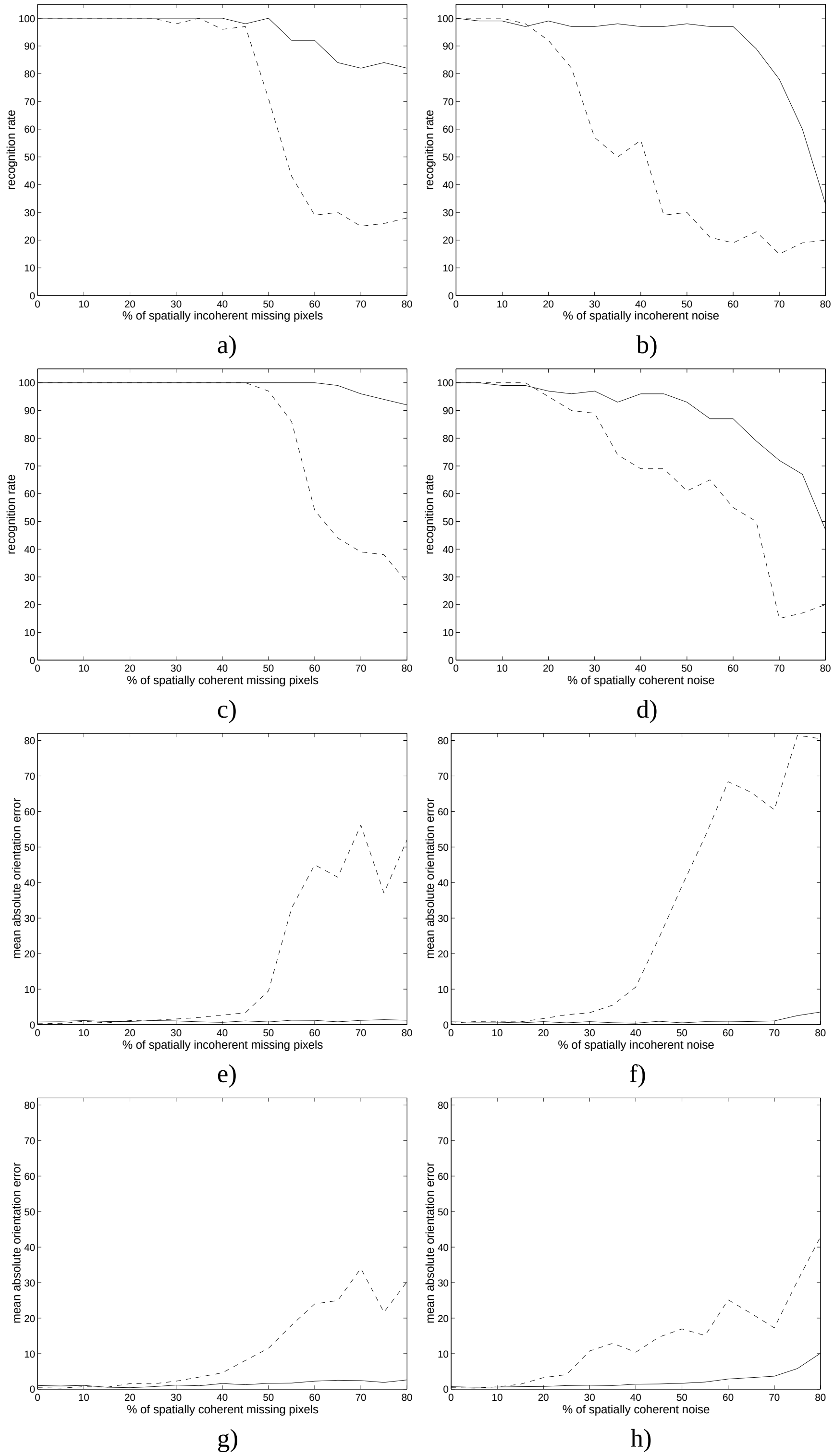
6. Conclusion

A view-based approach using range images is one of the approaches to the recognition and pose determination of objects. Due to the architecture of range image sensors, range images usually contain also missing pixels and noise.

In this paper we proposed a robust method which efficiently overcomes the problem of missing pixels. Furthermore, this method overcomes also the problems caused by the noise and occlusions. Instead of computing the coefficients by a projection of the data onto the eigenimages, we determine the coefficients by solving a set of linear equations in a robust manner.

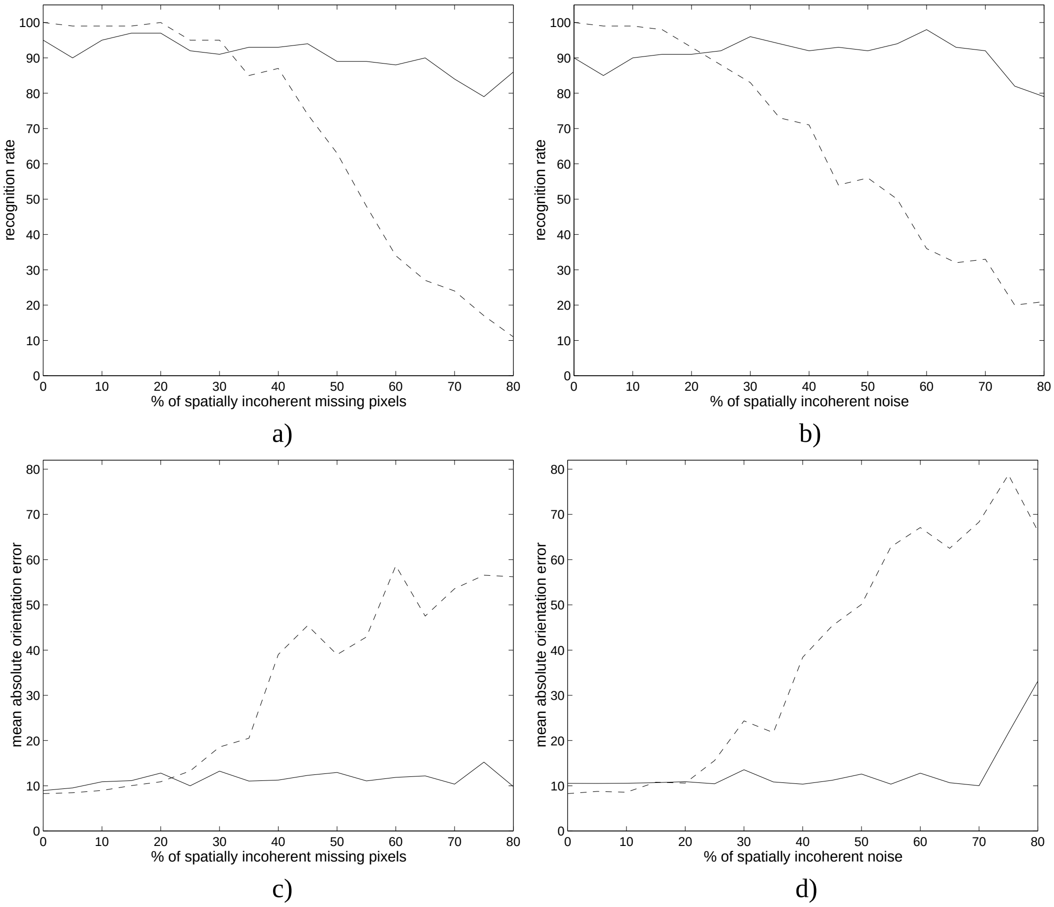
In this paper we do not address the problems, which are pertinent for the eigenspace approach (translation, scaling, parameter space explosion, etc.). However, our method for estimating the coefficients is a substitute for the standard one. Therefore it can be utilised in all algorithms that use the standard eigenspace approach providing higher robustness to missing pixels and noise.

The results show that the proposed method outperforms the standard one in recognition and pose determination. Using the proposed robust method we can reliably recognize an object and its orientation also when a range image contains very high percentage of missing pixels and noise or occlusions.



a) b) c) d) e) f) g) h)

Figure 9. 1 DOF example: a)-d) recognition rate, e)-h) mean absolute orientation error depending on the percentage of missing pixels or noise. a),e) Spatially incoherent missing pixels, b),f) spatially incoherent noise, c),g) spatially coherent missing pixels, d),h) spatially coherent noise. Dashed line for standard method, solid line for robust method.



a) b) c) d)

Figure 10. 2 DOF example: Recognition rate depending on the percentage of spatially incoherent a) missing pixels and b) noise. Mean absolute orientation error depending on the percentage of spatially incoherent c) missing pixels and d) noise. Dashed line for standard method, solid line for robust method.

References

[1] T. Amano, S. Hiura, A. Yamaguti, and S. Inokuchi. Eigen space approach for a pose detection with range images. *Proceedings of ICPR'96*, pages 622–626, 1996.

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

[3] A. Bab-Hadiashar and D. Suter. Optic flow calculation using robust statistics. *Proceedings of CVPR'97*, pages 988–993, 1997.

[4] D. H. Ballard. *Neural Computation*. MIT Press, 1997.

[5] P. J. Besl and R. C. Jain. Three-dimensional object recognition. *Computing Surveys*, 17(1):75–145, 1985.

[6] R. J. Campbell and P. J. Flynn. Eigenshapes for 3D object recognition in range data. *Proceedings of CVPR'99*, II:505–510, 1999.

[7] R. J. Campbell and P. J. Flynn. A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, in press, 2001.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commu. ACM*, 24(6):381–395, 1981.

[9] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78:99–118, 2000.

[10] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *Int. J. Comput. Vision*, 14:5–24, 1995.

[11] S. K. Nayar, H. Murase, and S. A. Nene. Parametric appearance representation. *Early Visual Learning*, pages 131–160, 1996.

[12] D. Skočaj and A. Leonardis. Acquiring range images of objects with non-uniform albedo using high-dynamic scale radiance maps. *Proceedings of ICPR'00*, pages 778–781, September 2000.

[13] M. Trobina. *Error Model of a Coded-Light Range Sensor, technical report*. Communication Technology Laboratory, ETH Zentrum, Zürich, 1995.