# Active Object Recognition in Parametric Eigenspace *

Hermann Borotschnig, Lucas Paletta, Manfred Prantl and Axel Pinz
Institute for Computer Graphics and Vision,
Technical University Graz,
Münzgrabenstr. 11, A-8010 Graz, Austria,
[last_name]@icg.tu-graz.ac.at

**Abstract**

We present an efficient method within an active vision framework for recognizing objects which are ambiguous from certain viewpoints. The system is allowed to reposition the camera to capture additional views and, therefore, to resolve the classification result obtained from a single view. The approach uses an appearance based object representation, namely the parametric eigenspace, and augments it by probability distributions. This captures possible variations in the input images due to errors in the pre-processing chain or the imaging system. Furthermore, the use of probability distributions gives us a gauge to view planning. View planning is shown to be of great use in reducing the number of images to be captured when compared to a random strategy.

## 1 Introduction

Most computer vision systems found in the literature perform object recognition on the basis of the information gathered from a single image. Typically, a set of features is extracted and matched against object models stored in a database and much research in computer vision has gone in the direction of finding features that are capable of discriminating objects [1]. However, this approach faces problems once the features available from a single view are simply not sufficient to determine the identity of the observed object. Such a case happens, for example, if there are objects in the database which look very similar from certain views (*ambiguous objects*); a difficulty that is compounded when we have large object databases.

A solution to this problem is to utilize the information contained in multiple sensor observations. *Active recognition* [2, 12] provides the framework for collecting evidence until we obtain a sufficient level of confidence in one object hypothesis. There have been several approaches to explore the merits of this framework, for example, to the tasks of remote-sensing [10] and object recognition [3, 9]. From the viewpoint of active recognition the work most closely related to ours is probably that by Hutchinson and Kak [4] and by Callari and Ferrie [2]. Hutchinson and Kak describe an active object recognition

* This work was supported by the Austrian 'Fonds zur Förderung der wissenschaftlichen Forschung' under grant S7003, and the Austrian Ministry of Science (BMWV Gz. 601.574/2-IV/B/9/96).

BMVC 1998 doi:10.5244/C.12.63

system based on Dempster-Shafer belief accumulation. The action which minimizes a newly defined measure of ambiguity is performed next. They use various actively controlled sensors, most prominently a range-finder and a CCD-camera. The experiments are performed in a blocks-world environment. Callari and Ferrie base their active object recognition system on model-based shape, pose and position reconstructions from range data. They estimate Bayesian probabilities with neural nets and choose those steps that minimize the expected ambiguity measured by Shannon entropy. Previous work has also been reported in planning sensing strategies. Murase and Nayar [6] have presented an approach for illumination planning in object recognition by searching regions in eigenspace where object-manifolds are best separated. A conceptually similar strategy will be followed below. However, Murase and Nayar have limited their approach to an off-line planning phase using no active steps.

Active recognition accumulates evidence collected from a multitude of sensor observations and thus, in a sense, moves the burden of object recognition slightly away from the process used to recognize a single view to the process of integrating the classification results of multiple views. Nevertheless, the system still has to provide tentative object hypotheses for a single view (see [11] for a recent review on research in object recognition).

The system we present in this paper uses a modified version of Murase's and Nayar's [7] appearance based object recognition system to provide object classifications for a single view and augments it by active recognition components. Murase's and Nayar's method was chosen because it does not only result in object classification but also gives accurate pose estimation (a prerequisite for active object recognition).

## 2 Object recognition in parametric eigenspace

Appearance based approaches to object recognition, and especially the eigenspace method, have experienced a renewed interest in the computer vision community [13, 8, 7] due to their ability to handle combined effects of shape, pose, reflection properties and illumination. Furthermore, appearance based object representations can be obtained through an automatic learning procedure and do not require the explicit specification of object models. As the eigenspace object recognition method proposed by Murase and Nayar forms the basis for our recognition system, we shall give a brief description of their approach in the remainder of this section (more detail can be found in [7]).

The eigenspace approach requires an off-line learning phase during which images of all objects from many different views are used to construct the eigenspace. In subsequent recognition runs the test images are projected into the learned eigenspace and the closest model point is determined. In a preprocessing step it is ensured that all images of all objects are of the same size and that they are normalized with regard to overall brightness changes due to variations in the ambient illumination or aperture setting of the imaging system. These normalized images can then be written as a vector $\mathbf{x}$ by reading pixel brightness values in a raster scan manner, i.e., $\mathbf{x} = (x_1, \ldots, x_N)^T$ with $N$ being the number of pixels in an image. $\mathbf{X} := (\mathbf{x}_{1,1}, \mathbf{x}_{2,1}, \ldots, \mathbf{x}_{M,P})$ denotes a set of images with $M$ being the number of models (objects) and $P$ being the number of images used to sample the variations in pose or other parameters for each model[1]. Next, we define the $N \times N$ *covariance matrix* $\mathbf{Q} := \mathbf{X}\mathbf{X}^T$ and determine the eigenvectors $\mathbf{e}_i$ and the

[1] In order to simplify notation we assume $\mathbf{X}$ having zero mean.

corresponding eigenvalues $\lambda_i$. We assume that the number of images, $MP$, is much smaller than the number of pixels in an image, $N$, and thus efficient methods to calculate the first $MP$ eigenvectors can be used (see [7] for a discussion of various numerical techniques). Since $\mathbf{Q}$ is real and symmetric, we may assume that $< \mathbf{e}_i, \mathbf{e}_j >= \delta_{ij}$. We sort the eigenvectors in descending order of eigenvalues. The first $k$ eigenvectors are then used to represent the image set $\mathbf{X}$ to a sufficient[2] degree of accuracy: $\mathbf{x}_{i,j} \approx \sum_{s=1}^{k} g_s \mathbf{e}_s$, with $g_s =< \mathbf{e}_s, \mathbf{x}_{i,j} >$. We call the vector $\mathbf{g}_{i,j} := (g_1, \ldots, g_k)$ the projection of $\mathbf{x}_{i,j}$ into the eigenspace. Under small variations of the parameters $j$ for a fixed object $i$ the image $\mathbf{x}_{i,j}$ will usually not be altered drastically. Thus for each object $i$ the projections of consecutive images $\mathbf{x}_{i,j}$ are located on piece-wise smooth manifolds in eigenspace parameterized by $j$.

In order to recover the eigenspace coordinates $\mathbf{g_y}$ of an image vector $\mathbf{y}$ during the recognition stage, $\mathbf{y}$ is projected into the eigenspace, i.e., $\mathbf{g_y} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k)^T \mathbf{y}$. The object $m$ that has minimum distance $d_m$ between its manifold $\mathbf{g}_{i,j}$ and $\mathbf{g_y}$ is assumed to be the object in question: $d_m = \min_i \min_j \|\mathbf{g_y} - \mathbf{g}_{i,j}\|$. This gives us both: an object hypothesis and a pose estimation. In order to improve the pose estimation Nayar and Murase have suggested to use also individual object eigenspaces that are built by taking only images from one specific object for all values of the parameters $j$. Once the object hypothesis has been obtained using the universal eigenspace the image $\mathbf{y}$ is also projected into the eigenspace of object $m$ and a better estimate of the parameter $j$ is obtained.

# 3 Probability distributions in eigenspace

Before going on to discuss active fusion in the context of eigenspace object recognition we extend Nayar and Murase's concept of manifolds by introducing probability densities in eigenspace (Moghaddam and Pentland [5] also used probability densities in eigenspace for the task of face detection and recognition). Let us denote by $p(\mathbf{g}|o_i, \varphi_j)$ the likelihood of ending up at point $\mathbf{g}$ in the eigenspace of all objects projecting an image of object $o_i$ with pose parameters $\varphi_j$. [3] The likelihood is estimated from a set of sample images with fixed $o_i, \varphi_j$. The samples capture the inaccuracies in the parameters $\varphi$ such as location and orientation of the objects, fluctuations in imaging conditions such as moderate light variations, pan,tilt and zoom errors of the camera and segmentation errors. With the rule of conditional probabilities we obtain

$$P(o_i, \varphi_j|\mathbf{g}) = \frac{p(\mathbf{g}|o_i, \varphi_j) P(\varphi_j|o_i) P(o_i)}{p(\mathbf{g})} \tag{1}$$

In our experiments $P(\varphi_j|o_i)$ and $P(o_i)$ will be uniformly distributed. Given the vector $\mathbf{g}$ in eigenspace the conditional probability for seeing object $o_i$ is

$$P(o_i|\mathbf{g}) = \sum_j P(o_i, \varphi_j|\mathbf{g}). \tag{2}$$

Nayar and Murase's approach consists in finding an approximate solution for $o_m = arg\max_i P(o_i|\mathbf{g})$ by searching for the minimum distance to the next manifold. We can

[2] Sufficient in the sense of *sufficient for disambiguating various objects*. Quantitatively we demand $\sum_{i=1}^{k} \lambda_i / Trace(\mathbf{Q}) > threshold$.

[3] We use capital $P$ to indicate probabilities and lower case $p$ to indicate probability densities.

restate this approach in the above framework and thereby make explicit the underlying assumptions. We obtain Nayar and Murase's algorithm if we

1. estimate $P(o_i, \varphi_j|\mathbf{g}) = f(\|\mathbf{g}_i(\varphi_j) - \mathbf{g}\|)$ with $f(x) > f(y) \Leftrightarrow x < y$. Thus they assume that the mean of the distribution lies at the one captured or interpolated position $\mathbf{g}_i(\varphi_j)$. The distributions are radially symmetric and share the same variance for all objects $o_i$ and all poses $\varphi_j$. With this estimation the search for minimum distance can be restated as a search for maximum posterior probability:

$$arg \max_{i,j} P(o_i, \varphi_j|\mathbf{g}) = arg \min_{i,j} \|\mathbf{g}_i(\varphi_j) - \mathbf{g}\| .$$

2. In the calculation of the object hypothesis the sum in equation (2) is approximated by its largest term:

$$P(o_i|\mathbf{g}) \approx \max_j P(o_i, \varphi_j|\mathbf{g}) \Rightarrow arg \max_i P(o_i|\mathbf{g}) = arg \min_i \min_j \|\mathbf{g}_i(\varphi_j) - \mathbf{g}\| .$$

The first approximation may be error-prone as the variance and shape of the probability distributions in eigenspace may differ from point to point. In our experiments we have found it necessary to model $p(\mathbf{g}|o_i, \varphi_j)$ with a multivariate normal distribution. The second approximation may lead to mistakes in case only a few points of the closest manifold lie near to $\mathbf{g}$ while a lot of points of the second-closest manifold are located not much further away.

# 4 Active object recognition

Active steps in object recognition will lead to striking improvements if the object database contains objects that share similar views. The key process to disambiguate such objects is a planned movement of the camera to a new viewpoint from which the objects appear distinct. We will tackle this problem now within the framework of eigenspace based object recognition. In order to emphasize the major ideas only one degree of freedom (rotation around z-axis) is assumed in the following. Note, however, that the following discussion on active object recognition is quite general and can be extended to any number of degrees of freedom. Furthermore, it is not limited to the eigenspace recognition approach but its overall steps can also be adapted to other appearance based object recognition techniques.

## 4.1 View classification and pose estimation

During active recognition step number $n$ a camera movement is performed to a new viewing position at which an image $I_n$ is captured. The viewing position $\psi_n$ is known to the system through $\psi_n = \Delta\psi_1 + .. + \Delta\psi_n$ where $\Delta\psi_k$ indicates the movement performed at step $k$. Processing of the image $I_n$ consists of figure-ground segmentation, normalization (in scale and brightness) and projection into the eigenspace, thereby obtaining the vector $\mathbf{g}_n = \mathbf{g}_n(I_n)$.

Given input image $I_n$ we seek from the object recognition system on the one hand a classification result for the object hypotheses $P(o_i|I_n)$ while on the other hand a possibly separate pose estimator should deliver $P(\varphi_j^n|o_i, I_n)$. The superscript $\varphi_j^n$ is used

to indicate that the estimation has been obtained after processing image $I_n$ at hardware position $\psi_n$. In eigenspace recognition we obtain through eq. (1) the quantity $P(o_i, \varphi_j^n | I_n) := P(o_i, \varphi_j^n | \mathbf{g}_n)$ from the probability distributions in the eigenspace of all objects. From that quantity we can calculate $P(o_i|I_n) := P(o_i|\mathbf{g}_n)$ as indicated by eq. (2). For the experiments described below the pose estimation obtained by using the eigenspace of all objects has turned out to be sufficiently accurate. Therefore we use $P(\varphi_j^n | o_i, I_n) = P(o_i, \varphi_j^n | I_n)/P(o_i|I_n)$.

In order to ensure consistency when fusing pose estimations obtained at different hardware-positions each pose estimation has to be transformed to a fixed set of coordinates of the hardware setup. We use the quantity $P(\varphi_j | o_i, I_n, \psi_n)$ to denote the probability of having the object pose $\varphi_j$ at the fixed hardware position $\psi = 0$ after processing of $I_n$. Since the real image $I_n$ has been captured at position $\psi_n$ this probability is related to $P(\varphi_j^n | o_i, I_n)$ through $P(o_i, \varphi_j | I_n, \psi_n) := P(o_i, \varphi_j^n + \psi_n | I_n)$. It is $P(o_i, \varphi_j | I_n, \psi_n)$ that will be used for fusion. For ease of notation we will omit the dependence on $\psi_n$ in the following and distinguish the two probabilities $P(o_i, \varphi_j | I_n)$ and $P(o_i, \varphi_j^n | I_n)$ solely through the superscript on $\varphi_j^n$.

## 4.2 Information integration

The currently obtained probabilities $P(o_i|I_n)$ and $P(\varphi_j|o_i, I_n)$ for object hypothesis $o_i$ and pose hypothesis $\varphi_j$ are used to update the overall probabilities $P(o_i|I_1, .., I_n)$ and $P(\varphi_j|o_i, I_1, .., I_n)$. For the purpose of updating, the outcome of individual observations is assumed to be conditionally independent given $o_i$ and we obtain (assuming equal priors)

$$P(o_i|I_1, .., I_n) \propto P(o_i|I_1, .., I_{n-1}) P(o_i|I_n) \quad (3)$$

$$P(\varphi_j|o_i, I_1, .., I_n) \propto P(\varphi_j|o_i, I_1, .., I_{n-1}) P(\varphi_j|o_i, I_n) \quad (4)$$

$$P(o_i, \varphi_j|I_1, .., I_n) = P(\varphi_j|o_i, I_1, .., I_n) P(o_i|I_1, .., I_n) \quad (5)$$

## 4.3 View planning

View planning consists in attributing a score $s_n(\Delta\psi)$ to each possible movement $\Delta\psi$ of the camera. The movement obtaining the highest score will be selected next:

$$\Delta\psi_{n+1} := arg \max_{\Delta\psi} s_n(\Delta\psi) \quad (6)$$

The score measures the utility of action $\Delta\psi$, taking into account the expected reduction of entropy for the object hypotheses. We denote entropy by

$$H(o_i|\mathbf{g}_1, .., \mathbf{g}_n) := -\sum_{o_i} P(o_i|\mathbf{g}_1, .., \mathbf{g}_n) \log P(o_i|\mathbf{g}_1, .., \mathbf{g}_n) \quad (7)$$

where $P(o_i|\mathbf{g}_1, .., \mathbf{g}_n) = P(o_i|I_1, .., I_n)$ as each of the real images $I_k$ translates deterministically into an eigenspace vector $\mathbf{g}_k = \mathbf{g}_k(I_k)$. Other factors may be taken into account such as the cost of performing an action. For the purpose of demonstrating the principles of active fusion in object recognition, we have restricted attention to the average entropy reduction using

$$s_n(\Delta\psi) := \sum_{o_i, \varphi_j} P(o_i, \varphi_j|I_1, .., I_n) \, \Delta H(\Delta\psi, o_i, \varphi_j, I_1, .., I_n). \quad (8)$$

The term $\Delta H$ measures the entropy loss to be expected, if $o_i, \varphi_j$ were the correct object and pose hypotheses and step $\Delta\psi$ is performed. In the calculation of the score $s_n(\Delta\psi)$ this entropy loss is weighted by the probability $P(o_i, \varphi_j | I_1, .., I_n)$ for $o_i, \varphi_j$ being the correct hypotheses. The expected entropy loss is again an average quantity given by

$$\Delta H(\Delta\psi, o_i, \varphi_j, I_1, .., I_n) := \\ H(o_i|\mathbf{g}_1, .., \mathbf{g}_n) - \int_\Omega p(\mathbf{g}|o_i, \varphi_j + \psi_n + \Delta\psi) H(o_i|\mathbf{g}_1, .., \mathbf{g}_n, \mathbf{g}) d\mathbf{g} \qquad (9)$$

The integration runs over the whole eigenspace $\Omega$. Note that $H(o_i|\mathbf{g}_1, .., \mathbf{g}_n, \mathbf{g})$ on the right hand side of eq. (9) implies a complete tentative fusion step performed with the hypothetically obtained eigenvector $\mathbf{g}$ at position $o_i, \varphi_j + \psi_n + \Delta\psi$. Usually evaluation of the averages in eq. (8) and (9) will be time consuming. Restricting attention only to the most probable initial positions $o_i, \varphi_j$ in eq. (8) and efficient sampling of the average in eq. (9) are reasonable simplifications. Once a set of samples $\mathbf{g}_s$ has been fixed the likelihoods on the right hand side of eq. (9) and the probabilities $P(o_i, \varphi_j|\mathbf{g}_s)$ needed to calculate $H(o_i|\mathbf{g}_1, .., \mathbf{g}_s)$ can be computed off-line, thereby increasing the speed of planning considerably. Alternatively one may implicitly learn a useful strategy from experience [9] but this requires an extensive training phase beforehand.

In order to avoid capturing views from similar viewpoints over and over again the score as calculated with eq. (8) is multiplied by a mask that is zero at the already visited locations and rises to 1 as the distance from these locations increases.
The process terminates if entropy $H(o_i|\mathbf{g}_1, .., \mathbf{g}_n)$ is lower than a pre-specified value or no more reasonable actions can be found (maximum score too low).

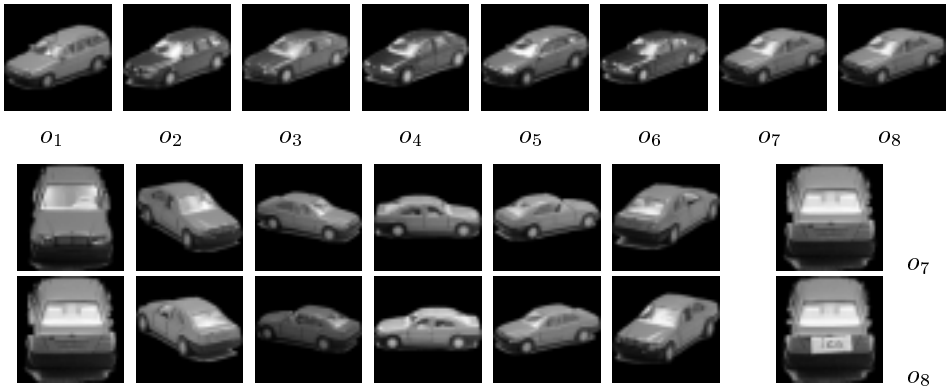



Figure 1: Each of the objects (*top row*) is modeled by a set of 2-D views (*below*, for object $o_1$). A marker is attached at the rear side of object $o_8$ to discriminate it from object $o_7$ (*bottom right*).

# 5 Experiments

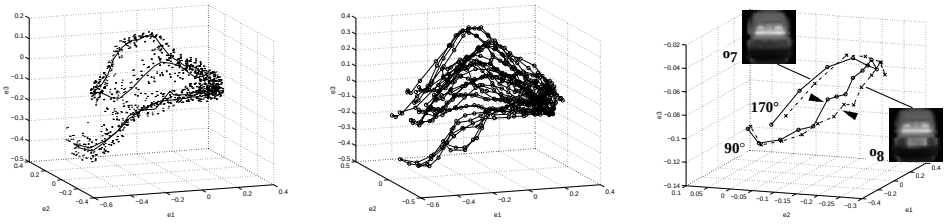
The proposed recognition system was tested with 8 objects (Figure 1) of similar appearance concerning shape, reflectance and color. For comparison reasons, two objects $o_7$

and $o_8$ are identical but discriminated by a white marker which is attached at the rear side of object $o_8$. The items were rotated on a computer-controlled turn-table by $5^\circ$ intervals, with constant illumination and fixed distance to the camera. The object region is automatically segmented from the background using a combined brightness and gradient threshold operator, while pixels classified as background are set to zero gray level. Images are then rescaled to $100 \times 100$ pixels and projected to an eigenspace of dimension 3 (see section 6 for comments on the unusually low dimensionality). For each view, additional samples are collected, emulating possible segmentation errors. The object region in the normalized image is shifted into a randomly selected direction by $3\%$ of the image dimension, as proposed in [6].

The sample distribution in eigenspace, drawn for a single object, is depicted in Figure 2a. The significant overlap between manifolds of all objects, computed by interpolation between the means of pose distributions (Figure 2b), visualizes the overall ambiguity in the representation.

For a probabilistic interpretation of the data, the likelihood of a sample $\mathbf{g}$, $p(\mathbf{g}|o_i, \varphi_j)$, given specific object $o_i$ and pose $\varphi_j$, is modeled by a multivariate Gaussian density $N(\mu_{i,j}, \Sigma_{i,j})$, where mean $\mu_{i,j}$ and covariance $\Sigma_{i,j}$ are estimated from the data that have been corrupted by segmentation errors. From this estimate both object (eq. (2), (3)) and pose (eq. (2), (4), (5)) hypotheses are derived, assuming uniform probability of the priors.

For object $o_7$ the ambiguity in the classification of object membership is demonstrated

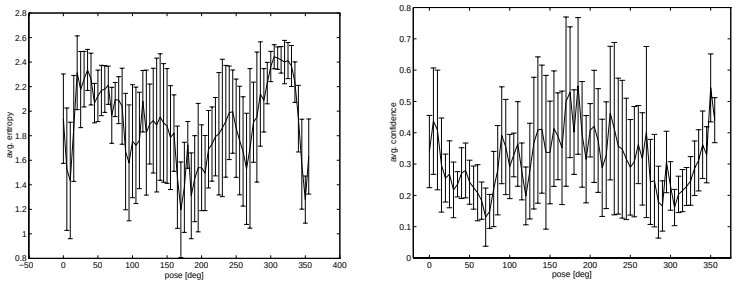



(a) Sample distribution (b) Manifolds (c) Marker distance

Figure 2: Eigenspace representation (*a*) of object 1 with the interpolated manifold (*line*) and samples (*dots*, emulating segmentation errors); (*b*) manifolds of all 8 objects; (*c*) distance between the manifolds of two similar objects introduced by a discriminative marker feature.

in the distribution of entropies over different poses in Figure 3a. The minimum at $\approx 170^\circ$ indicates the most discriminative view which contributes to the distinction from object $o_8$ due to the attached marker. The local minimum at $\approx 0^\circ$ specifies a pose where the similar objects $o_7$ and $o_8$ are favored whereas the others are sorted out. (Figure 3b).

Table 1 depicts the probabilities for the object hypotheses in a selected run that finished after three steps obtaining an entropy of $0.17$ (threshold $0.2$) and the correct object and pose estimations. Object $o_7$ had been placed on the turn-table with pose $0^\circ$. Note that the run demonstrates a hard test for the proposed method. The initial conditions have been chosen such that the first image when projected into the $3(!)-$dimensional eigenspace does not favor the correct hypotheses. Consequently, object recognition relying on a single image would favor erroneously object $o_3$ at pose $0$ (pose estimations are




(a) Entropy (b) Confidence

Figure 3: (*a*) The average entropy of the estimated posteriors $p(o_i|\varphi_j, \mathbf{g})$ of the samples $\mathbf{g}$ at each pose $\varphi_j$ of object $o_7$ (error bars denote standard deviations). (*b*) The corresponding means of the confidences in $o_7$ indicate the evidence provided by the views.

not depicted in table 1). Only additional images can clarify the situation.

A set of test runs has been performed. The average number of steps required to find the correct answer has been $2.6$. This number has to be compared to $12.8$, the average number of steps required to find the correct answer when using a random strategy. The high number of steps in the random strategy is due to the highly overlapping probability distributions in three dimensions. Random placement often ends in regions where no object hypothesis is clearly favored and thus "confuses" the system.

# 6 Conclusions

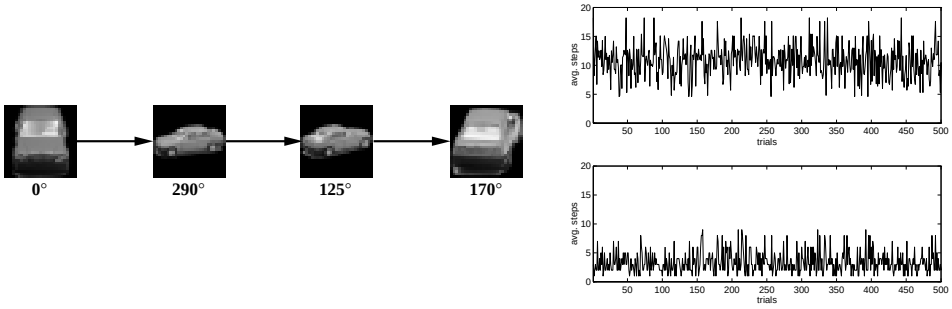
We have presented an active object recognition system for single object scenes. Depending on the uncertainty in the current object classification the recognition task acquires new sensor measurements in a planned manner until the confidence in a certain hypothesis obtains a pre-defined level or another termination criterion is reached. The well known object recognition approach using eigenspace representations was augmented by probability distributions in order to capture possible variations in the input images due to errors in the pre-processing chain. Furthermore, probabilistic object classifications (instead of hard decisions) can be used as a gauge to perform view planning. View planning is based on the expected reduction in Shannon entropy over object hypotheses given a new viewpoint.

The experimental results lead to the following conclusions:

1. The dimension of the eigenspace can be lowered considerably, if active recognition is guiding the object classification phase. This fact opens the way to the use of very large object databases containing e.g. $500$ objects at $72$ (i.e., $5^\circ$ rotation) views where eigenspace construction is feasible with todays computing resources only if its dimension stays low enough (around $200 = 0.5\%$ of $36,000$ images). The traditional eigenspace methods are likely to face problems if the eigenspace dimensionality is too low relative to the number of objects represented (due to the overlapping manifolds).

2. Even objects sharing most of their views can be disambiguated by an active movement that places the camera such that the differences between the objects become apparent. That is, active recognition tries to identify those regions in eigenspace where the manifold representations are well separated.

3. The planning phase is necessary and beneficial as random placement of the camera leads to distinctively worse results.




(a) Sample sequence (b) Performance

Figure 4: (*a*) Sample pose sequence actuated by the planning system (see Table 1). A comparison of averaged trial lengths (*b*) between a random (*top*) and the presented look-ahead policy (*below*) illustrates the improved performance.

| | $\psi_0 = 0°$ | | $\psi_1 = 290°$ | | $\psi_2 = 125°$ | | $\psi_3 = 170°$ | |
|---|---|---|---|---|---|---|---|---|
| $o_i$ | $P(o_i\|\mathbf{g}_0)$ | $P_f$ | $P(o_i\|\mathbf{g}_1)$ | $P_f$ | $P(o_i\|\mathbf{g}_2)$ | $P_f$ | $P(o_i\|\mathbf{g}_3)$ | $P_f$ |
| 1 | 0.001 | 0.001 | 0.000 | 0.000 | 0.139 | 0.000 | 0.000 | 0.000 |
| 2 | 0.026 | 0.026 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.314 | 0.314 | 0.097 | 0.203 | 0.055 | 0.074 | 0.091 | 0.013 |
| 4 | 0.027 | 0.027 | 0.096 | 0.017 | 0.097 | 0.011 | 0.002 | 0.000 |
| 5 | 0.000 | 0.000 | 0.098 | 0.000 | 0.335 | 0.000 | 0.032 | 0.000 |
| 6 | 0.307 | 0.307 | 0.015 | 0.031 | 0.009 | 0.001 | 0.224 | 0.000 |
| 7 | 0.171 | 0.171 | 0.354 | 0.403 | 0.224 | 0.597 | 0.822 | 0.967 |
| 8 | 0.153 | 0.153 | 0.338 | 0.344 | 0.139 | 0.315 | 0.032 | 0.019 |

Table 1: Probabilities for object hypotheses in one run (see Figure 4). $P_f$ are the fused probabilities $P(o_i|\mathbf{g}_1, .., \mathbf{g}_n)$.

# References

[1] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 2(94):115–147, 1987.

[2] F. G. Callari and F. P. Ferrie. Autonomous Recognition: Driven by Ambiguity. In *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pages 701–707, 1996.

[3] K.D. Gremban and K. Ikeuchi. Planning Multiple Observations for Object Recognition. *International Journal of Computer Vision*, 12(2/3):137–172, 1994.

[4] S.A. Hutchinson and A.C. Kak. Multisensor Strategies Using Dempster-Shafer Belief Accumulation. In M.A. Abidi and R.C. Gonzalez, editors, *Data Fusion in Robotics and Machine Intelligence*, chapter 4, pages 165–209. Academic Press, 1992.

[5] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, July 1997.

[6] H. Murase and Shree K. Nayar. Illumination Planning for Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12):1219–1227, 1994.

[7] Hiroshi Murase and Shree K. Nayar. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision*, 14(1):5–24, January 1995.

[8] S. Nayar, H. Murase, and S. Nene. General Learning Algorithm for Robot Vision. In SPIE, editor, *Neural and Stochastic Methods in Image and Signal Processing*, volume 2304, July 1994.

[9] L. Paletta, M. Prantl, and A.Pinz. Reinforcement Learning for Autonomous Three-Dimensional Object Recognition. In *Proc. 6th Symposium on Intelligent Robotic Systems*. Edinburgh, UK,, 1998. in print.

[10] A. Pinz, M. Prantl, H. Ganster, and Hermann Kopp-Borotschnig. Active Fusion - A New Method Applied to Remote Sensing Image Interpretation. *Pattern Recognition Letters*, 17(13):1349–1359, 1996. Special issue on 'Soft Computing in Remote Sensing Data Analysis'.

[11] Arthur R. Pope. Model-Based Object Recognition: A Survey of Recent Research. Technical Report 94-04, The University of British Columbia - Department of Computer Science, January 1994.

[12] M. Prantl, H. Borotschnig, H. Ganster, D. Sinclair, and A. Pinz. Object Recognition by Active Fusion. In *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling*, volume 2904, pages 320–330. SPIE, 1996.

[13] M. Turk and A. Pentland. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, 3(1), 1991.