Appearance Based Generic Object Modeling and Recognition using Probabilistic Principal Component Analysis

Christopher Drexler*, Frank Mattern, and Joachim Denzler

Lehrstuhl für Mustererkennung
Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen
Phone: +49 9131 8527826
Fax: +49 9131 303811
`drexler@informatik.uni-erlangen.de`

Abstract Classifying unknown objects in familiar, general categories rather than trying to classify them into a certain known, but only similar class, or rejecting them at all is an important aspect in object recognition. Especially in tasks, where it is impossible to model all possibly appearing objects in advance, generic object modeling and recognition is crucial.

We present a novel approach to generic object modeling and classification based on probabilistic principal component analysis (PPCA). A data set can be separated into classes during an unsupervised learning step using the expectation–maximization algorithm. In contrast to principal component analysis the feature space is modeled in a *locally* linear manner. Additionally, Bayesian classification is possible thanks to the underlying probabilistic model.

The approach is applied to the COIL-20/100 databases. It shows that PPCA is well suited for appearance based generic object modeling and recognition. The automatic, unsupervised generation of categories matches in most cases the categorization done by humans. Improvements are expected if the categorization is performed in a supervised fashion.

1 Introduction

Object recognition ideally tackles the problem of identifying a certain set of objects under changing illumination, camera parameters, and viewpoints, as well as under partial occlusion. If the recognition is also unaffected by novel exemplars of a category, for which some different exemplars are already known, we call it *generic object recognition*. For example, given a generic object model (or category) of cars, the system should be able to classify never before seen cars into the category `car`. However, already known cars should get their individual class label, e.g. `BMW`. The motivation for such a coarse to fine strategy is that most tasks, for example in a service robot scenario, might be solved without exactly knowing the specific class of an object. In our work we are interested in categories that arise from appearance only and not from function.

Generic object modeling has been studied in the past (e.g. [8, 10]). Our approach differs from segmentation based approaches followed by some sort of grouping mechanism [10, 6] in that we use an appearance based approach. In contrast to image retrieval

* This work was funded by the German Science Foundation (DFG) under grant DE 735/2–1. Only the authors are responsible for the content.

techniques we do not want to classify solely into generic categories but for known objects we would like to get the specific class label, as well. This demands for a hierarchical, from coarse (category level) to fine (class level), strategy.

Our work starts with the idea of appearance based recognition using principal component analysis (PCA). We review so called *factor analysis models* [1] that are a generative way of describing the images generated from a random vector (called factors), disturbed by an arbitrary noise source — in most cases Gaussian noise. Inherently uncertainty is modeled probabilistically. In Section 2 we will summarize the theory. Starting from factor analysis models we apply the approaches from [5, 13] to the problem of generic object modeling and recognition: both introduce mixtures of factor analyzers, although in a slightly different manner. As a consequence the feature space is now approximated in a piecewise linear manner in contrast to the globally linear approximation of the PCA. In [13] restrictions on the statistical model results in the so called *probabilistic principal component analysis* (PPCA). The underlying statistical framework makes it possible to apply maximum likelihood or maximum a posteriori estimation for classification of objects. However, applying PPCA to image data for generic modeling needs some modification of the whole approach in practice (Section 3).

The benefits of the PPCA are shown in the experimental part of the paper in Section 4: using the expectation maximization algorithm (EM–Algorithm) [4] the PPCA can be estimated from a training set in an unsupervised manner, which results in categories of objects defined by similar appearance. Recognition rates for the COIL-100 database support the claim, that the approach is capable for hierarchical modeling and recognition.

2 From PCA to PPCA: Factor Analysis Models

Object models based on PCA have become a popular approach for appearance based object and face recognition in the past years [9, 2]. With respect to generic object recognition the PCA has the advantage that — as a rule of thumb — eigenvectors belonging to larger eigenvalues model the coarse appearance of an object while others are responsible for finer details (see Figure 1). Thus, generic object classes should summarize objects, whose features, when projected into the Eigenspace using eigenvectors belonging to large eigenvalues, are located close together.

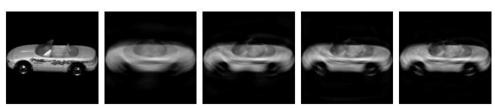

Figure 1. Backprojection of an object using different numbers of Eigenvectors. From left to right: original image, backprojection using 5, 10, 15, 20 eigenvectors corresponding to the 5, 10, 15, 20 largest eigenvalues, respectively. Details are coded in the eigenvectors belonging to smaller eigenvalues.

Disadvantages of the PCA are its global linearity assumption in the feature space and a missing underlying statistical model that would allow soft decisions about the membership of a certain object using probabilities. In the following we present a new approach for appearance based recognition using factor analysis models, that — under certain assumptions — equals PCA.

2.1 Factor Analysis

In factor analysis [1] the assumption is that an observation (for example, an image) $\boldsymbol{t}_i \in \mathrm{I\!R}^d$ is generated by a q–dimensional random vector $\boldsymbol{x}_i$ (the elements of $\boldsymbol{x}_i$ are called factors) according to the following linear mapping

$$\boldsymbol{t}_i = \boldsymbol{W}\boldsymbol{x}_i + \boldsymbol{\mu} + \boldsymbol{\epsilon} \quad . \tag{1}$$

Like in PCA the vector $\boldsymbol{t}_i$ is built up from an image by concatenating the rows or columns of the image. The vector $\boldsymbol{\mu}$ is a constant displacement vector and $\boldsymbol{\epsilon}$ is a noise vector. The assumption is that $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_q)$ as well as $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi})$ are zero mean Gaussian distributed random vectors. The matrix $\boldsymbol{I}_q$ denotes the $q \times q$–dimensional identity matrix. The covariance matrix $\boldsymbol{\Psi}$ is assumed to be diagonal. As a consequence the observation $\boldsymbol{t}_i$ is also Gaussian distributed. Given a set of n observations $\boldsymbol{t}_i$ (i.e., for example images of a certain class; we will consider the unsupervised case in Section 4) the unknown parameters of the factor model $\boldsymbol{W}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Psi}$ can be estimated using the EM algorithm. Details for the computations during the E–step and the M–step can be found in [5].

2.2 Mixture of Factor Analyzers

The described factor analysis model in (1) can be extended to a mixture model of m Gaussian distributions. The observation vectors $\boldsymbol{t}_i$ are now modeled by

$$\boldsymbol{t}_i = \sum_{k=1}^{m} \omega_k (\boldsymbol{W}_k \boldsymbol{x}_i + \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_k) \tag{2}$$

with $\boldsymbol{x}_i \sim \mathcal{N}(0, \boldsymbol{I}_q)$ and $\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \boldsymbol{\Psi}_k)$. The quantity ω_k is the weight of the kth mixture component, $\boldsymbol{\Psi}_k$ again a diagonal covariance matrix of the observation noise. Similar to the factor analysis model, the EM–algorithm can be applied to estimate the unknown parameters of the mixture model, ω_k, $\boldsymbol{W}_k$, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Psi}_k$. Due to lack of space we will not go into detail of the computations in the E–step and the M–step. The reader is referred to [5].

2.3 Probabilistic Principal Component Analysis

In [13] it has been shown that factor analysis and principal component analysis will coincide under special conditions. Coincidence means that the columns of the factor loadings matrix $\boldsymbol{W}$ contain the eigenvectors of the covariance matrix of the observations. Even if the exact correspondence is only guaranteed under specific conditions, practically there are no differences in the expression capabilities of the factor analyzer and the PCA models. However, the probabilistic approach by factor analysis has the advantage that soft decisions can be made based on probabilities. In the case of the mixture of factor analyzer, the feature space is approximated in a locally linear way, instead of a global linear one as in standard PCA.

In order to optimally approximate principal component analysis in practice without fulfilling the strict conditions, the diagonal covariance matrix $\boldsymbol{\Psi}$ is restricted to have identical diagonal elements, i.e. $\boldsymbol{\Psi} = \sigma^2 \boldsymbol{I}_d$ [13]. The log–likelihood function of the overall model

$$\mathcal{L} = \sum_{i=1}^{n} \ln p(\boldsymbol{t}_i) \quad \text{where} \quad p(\boldsymbol{t}_i) = \sum_{k=1}^{m} \omega_k p(\boldsymbol{t}_i | k) \tag{3}$$

with n observations and m Gaussian distributions (submodels) will be maximized by a few EM–Iteration steps. The overall model can be divided into categories using a ML classifier with the likelihood distribution of the observation $\boldsymbol{t}_i$ given the submodel k

$$p(\boldsymbol{t}_i|k) = (2\pi)^{-d/2}|\boldsymbol{C}_k^{-1/2}| \exp\left(-\frac{1}{2}(\boldsymbol{t}_i - \boldsymbol{\mu}_k)^T \boldsymbol{C}_k^{-1}(\boldsymbol{t}_i - \boldsymbol{\mu}_k)\right) \tag{4}$$

with $\boldsymbol{C}_k := \boldsymbol{W}_k \boldsymbol{W}_k^T + \sigma^2 \boldsymbol{I}_d$ or by a Bayes classifier with the a posteriori probability

$$p(k|\boldsymbol{t}_i) = \frac{\omega_k p(\boldsymbol{t}_i|k)}{p(\boldsymbol{t}_i)} \tag{5}$$

of the submodel k given the observation $\boldsymbol{t}_i$. These probabilities will be exploited for building a hierarchy from categories to specific classes and for performing generic object recognition (see Section 4).

3 PPCA for Generic Object Modeling and Recognition

Armed with the theory and estimation technique of the PPCA from the last section, we can now apply this framework to generic object modeling and recognition. First, some technical problems are discussed that occur when applying PPCA to large data sets of high–dimensional data. Then, we present an approach that applies first a standard dimensionality reduction by PCA, and then models the resulting features in the eigenspace by PPCA. The results of this approach with respect to generic object modeling are presented in Section 4.

Up to now, the PPCA has only been applied to observation vectors of a dimension less than 64. This prevents us from applying PPCA directly to images, since only images up to a size of 8×8 pixel could be processed.

The main reason for this restriction is that the computation of the determinant of the inverse covariance matrix depends on the dimension of the data. As a consequence, one gets numerical instabilities in the computation for large and/or small variances. For example, for an image size of 16×16 a variance larger than 16 will result in a determinant value that cannot longer be represented by a 64–bit `double` value on standard machines. Some solutions for these problems are discussed in [7].

To escape the curse of dimensionality we apply a normal PCA [9] in advance to reduce the input dimension for the PPCA algorithm to a maximum of 100, which is the maximum to be numerically manageable according to our experiments. In other words, the images are projected onto a lower dimensional space by PCA, on which a PPCA is performed in the following.

Besides the dimensionality reduction and the possibility to apply algorithms from standard eigenspace approaches one gets an additional degree of freedom in selecting features. Choosing different transformation matrixes of q successive eigenvectors, *not* starting with the first one, enables us to focus on different parts within the feature space and might be exploitable for generic object models. A systematic evaluation of such a procedure, especially on which eigenvectors to focus on, is under current investigation.

4 Experiments

We present results on experiments done with one of the standard database in the object recognition community: the COIL-20/100 databases (see [11], for the COIL-20), which contain images from 20 resp. 100 objects rotated on a turntable (72 images for each object, i.e. images taken every 5 degree). We performed the following experiments: unsupervised categorization, hierarchical model generation, and generic object recognition.

Unsupervised categorization. The objects in the COIL-20 database are modeled with mixtures of PPCA. In Figure 2 a subset of the resulting categorization is shown. The number of mixture components has been set to 20. The expectations with respect to

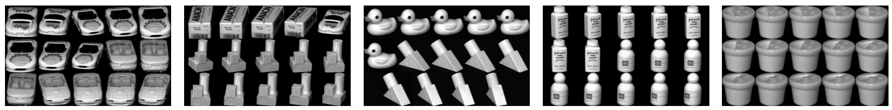

Figure 2. Subset of 5 out of 20 categories of the automatic categorization of the COIL-20 database by PPCA using 20 mixture components. For each category a representative subset of views of the contained objects is presented

unsupervised categorization holds for most object classes, for example the cars and the pots. Interesting is the subsumption of dug and the wooden part, which have similar shapes from the given viewpoint. These results show that categorization based on appearance can be done automatically with the proposed approach. Similar results have been achieved for the COIL-100 database. The reader should note, that a pure vector quantization produces similar but marginally worse results. However, we get a complete probabilistic model out of the training data, that is optimized with respect to the likelihood. This gets more important in the case of generic object recognition, as a mere vector quantization can not provide adequate information for identifying previously unseen objects. This will be shown in the experiments on generic classification later on (Figure 4 and 5).

Hierarchical model generation. We constructed up to three levels of mixtures of PPCA using all 20 objects of the COIL-20 database. For each level (0,1,2) a 5–dimensional Eigenspace is used and the 5–dimensional feature vectors of the projected training data is used as input for the PPCA algorithm. Each mixture component of the PPCA at one level is used to select those views that are modeled at the next level by an individual mixture of PPCA. With these experiments we like to show, that a coarse (category level) to fine (class level) graduation can be realized by our approach in an unsupervised manner. In Table 1 the categorization at the first and second level is shown for one of the five formed categories. The entries in the table show the number of objects filled into one category at the first level and the split into five categories at the next level. One can see, that for this example already at the second level the objects are put into distinct categories. Another example is shown in Table 2. Here, at the second level still some visually similar objects from different classes are in the same category. The cream cheese box (cat1-2) and the cars in category cat1-3 are separated. The category cat1-0 contains the cars and the medication box (anacin) that look similar at that level of representation. At the third level (not shown in the table) category cat1-0 is subdivided into another five categories where finally the anacin box is separated from the cars into separate categories. One interesting aspect can be observed during generation of the hierarchies: as soon as the objects are separated into distinct categories the PPCA starts separating the different poses of the objects. This behavior is no surprise since the manifold of the images of one object is then approximated in a piecewise linear manner by the mixture components of the PPCA.

category	objects		
	“vaseline”	“wooden part3”	“piggy bank”
cat0	14	34	48
cat0-0			26
cat0-1			9
cat0-2		34	
cat0-3	14		
cat0-4			13

Table 1. Category 0 (cat0): The entries in the table show the number of objects from the training set, classified into that category at level 0 and into the five categories at level 1.

category	objects					
	“cup”	“cream cheese”	“car1”	“anacin”	“car2”	“tylenol”
cat1	48	48	41	37	39	33
cat1-0			3	24	3	
cat1-1				5		28
cat1-2		48				
cat1-3			27		25	
cat1-4	48		11	8	11	5

Table 2. Category 1 (cat1): The entries in the table show the number of objects from the training set, classified into that category at level 0 and into the five categories at level 1.

Although we can not generalize from these results that the different objects in a training set get separated as above, it can be seen, that the categorization is able to form meaningful visual classes. This is a prerequisite for the generic recognition shown in the next paragraph.

Generic object modeling and recognition. In order to test the ability of our models to classify previously unseen objects we also used the whole COIL-20 database but we completely omitted two objects during the training stage. Both objects leave similar objects in the training set (see Figure 3). The “uncovered pot” has the “half covered pot” as a moderately similar object. The left out “car3” has two other similar cars in the training set. For these items we evaluated the ability of our models to classify seen and unseen objects.

UP P1 UC C1 C2

Figure 3. The two unknown objects “car3” (UC) and “uncovered pot” (UP) omitted from the training set together with their visually most similar objects (according to the distance within the Eigenspace) “car1” (C1), “car2” (C2) and “half covered pot” (P1) contained in the training set.

For generic object recognition the PPCA mixture models have to provide two different kinds of information: to which visual class a test image is assigned to and how well the corresponding submodel is able to model the object in terms of visual class membership. As we descend the hierarchy levels from the categories to the distinct object classes the consistency with the submodel should increase for known objects, as the models get more specialized, and it should get worse for unknown objects, as the ability of modeling the unknown appearance decreases being closer to the class level.

Figure 4 shows the log–likelihood for the hierarchy level 0 to 2 and Figure 5 does the same for the “distance from feature space” function used as quality criteria for a nearest–neighbor classification. Both diagrams show the averaged curves over all test images for the two unknown objects “uncovered pot” and “car3” as well as for the three known objects “half covered pot”, “car1” and “car2”.

In Figure 4 the two unknown objects can be easily identified with the “uncovered pot’s” log–likelihood decreasing rapidly with each level. The unknown car fits as good

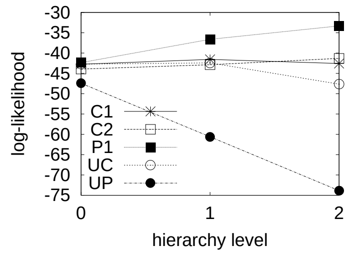

Figure 4. PPCA

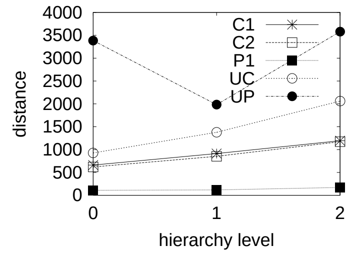

Figure 5. NN

Log–likelihood (Figure 4) and distance function plot (Figure 5) for the two cars (C1,C2) and the pot (P1) which are part of the training set as well as for the unknown car (UC) and the unknown pot (UP).

to level 1 as the known cars, but the log–likelihood also decreases at the last level. The log–likelihood for the known objects increase or remain constant through all levels, as expected.

Figure 5 demonstrates that an eigenspace approach with a nearest neighbor (NN) classification is not able to show similar behavior as the log–likelihood criteria. The distance gives no information on when to assign a given view to a category at a certain level and when to descend to the next level.

Finally experiments were done to show how well PPCA is suited for classification of known object classes. We equally divided the COIL-100 dataset into training and test. The training set has been used for unsupervised construction of the hierarchy of categories. For all images put into the same category at the lowest level we performed a standard (supervised) PCA which is used for the actual assignment to a class. With a 3–dimensional PCA in the preprocessing step (compare Section 3), the recognition rate on the test set at level 0 of the hierarchy is 80.5%, at level 1 88.2% and at level 2 91.2%. With respect to the dimension of the feature space, this appears to be a reasonable result. In the case of a 15–dimensional PCA the recognition rate at level 0 is already 98.7%, slightly decreasing to 97.0% at level 2. This unwanted phenomenon is mainly due to small numbers of examples particularly falling into one category at level 2, thus preventing a reasonable 15-dimensional PCA.

5 Conclusions

In this paper we presented a novel way of using mixtures of PPCA (MPPCA) models for hierarchical generic object recognition, where the term “generic” aims at visually similar classes and not functional ones.

During the training step the MPPCA is computed for all input images of all different classes. The single mixture components define a disjunctive partition of the training set into visual classes according the maximum likelihood of each training vector. The elements of each partition are then used as input data for creating MPPCA models at the next, more specialized, hierarchy level.

Unlike in former publications we do not subsample the training data in order to reduce the input dimension but perform a Karhunen–Loève–Transformation into a low-dimensional Eigenspace. This preserves more information on high resolution input images than using a maximum input image size of 8×8 as done up to now.

The results of our experiments show that the unsupervised partition gives reasonable classes which are appropriate for building hierarchical MPPCA models suitable for generic recognition of previously unseen objects. Classification on large data sets (COIL-100) benefits from the hierarchical approach, as a recognition rate of up to 98.7% can be achieved already with low–dimensional feature vectors.

Further, experiments on other image databases with more objects and evaluations of the presented algorithm together with a robust feature calculation for handling partially occluded objects will be performed. Also, nonlinear PCA for generic object modeling, for example Kernel PCA introduced in [12], is one topic of our further investigation.

Although beyond the scope of generic object modeling a combination of the presented approach with a contour based representation of objects, introduced in [3], seems to be very promising for object representation and segmentation in general.

6 Acknowledgments

We thank C. Schnörr for stimulating discussions and pointing us to density estimation methods in Kernel space.

References

1. D. Bartholomew. *Latent Variable Models and Factor Analysis*. Charles Griffin & Co. Ltd., London, 1987.
2. N. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
3. D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear Shape Statistics in Mumford–Shah based Segmentation. In *ECCV 2002*, to appear, 2002.
4. A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.
5. Z. Ghahramani and G. Hinton. The EM algorithm for Mixtures of Factor Analyzers. Technical Report CFG–TR–96j–1, Dept. of Computer Science,University of Toronto, February 1997.
6. Y. Keselman and S. Dickinson. Generic Model Abstraction from Examples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
7. F. Mattern. Probabilistische Hauptachsentransformation zur generischen Objekterkennung. Technical report, Diploma thesis, Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen–Nürnberg, 2001.
8. G.G. Medioni and A.R.J. Francois. 3–D Structure for Generic Object Recognition. In *Computer Vision and Image Analysis*, Volume 1, International Conference on Pattern Recognition, pages 30–37, 2000.
9. H. Murase and S. Nayar. Visual Learning and Recognition of 3–D Objects from Appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
10. R. C. Nelson and A. Selinger. Large–Scale Tests of a Keyed, Appearance–Based 3-D Object Recognition System. *Vision Research*, 38(15–16), August 1998.
11. S. Nene, S. Nayar, and H. Murase. Columbia Object Image Library (COIL–20). Technical Report CUCS–005–96, Dept. fo Computer Science, Columbia University, 1996.
12. B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10:1299–1319, 1998.
13. M. E. Tipping and C. M. Bishop. Mixtures of Probabilistic Principal Component Analysers. *Neural Computation*, 11(2):443–482, 1999.