

# COMPARISON OF APPEARANCE BASED METHODS FOR GENERIC OBJECT RECOGNITION

Frank Mattern\* and Joachim Denzler

Chair for Pattern Recognition, University Erlangen-Nuremberg, 91058 Erlangen, Germany

E-mail: [mattern@cs.fau.de](mailto:mattern@cs.fau.de)

**Summary:** In this paper, we present a large-scale examination of different appearance-based, segmentation-free classification methods for their usability in generic object recognition. Generic object recognition is a method to handle never before seen objects in classification by a hierarchical approach with a coarse-to-fine graduation. Unknown objects are only classified into coarse categories and rejected by classes that are too specific. Comparison of the approaches is accomplished on the basis of their recognition rate. The global generic recognition rate is computed for the best method and its robustness according to different types of noise is examined. Our experiments show that the PCA based method with Nearest Neighbor classification provides the best recognition rates, whereas models based on principal component analysis outperform the other methods in computation time and model size.

## 1. Introduction

Generic object recognition addresses the problem of classifying objects which are not known to the system. Novel exemplars are classified using hierarchically structured generic classes (or categories), where unknown objects should be assigned to a general class, and known ones should be assigned additionally to their object class. For instance, all cups, also the unknown ones, should be classified into the category “cup”, whereas a trained cup should also be recognized as that cup. This type of classification is referred to as generic object modeling and recognition and is necessary for applications where it is impossible to model all occurring objects, e.g. all objects in an office environment.

Image retrieval techniques like [1] classify objects into generic classes, too, but do not distinguish between known and unknown objects. In contrast to previous generic object modeling studies (e.g. [2, 3, 4]), we prevent image segmentation errors by using appearance based methods which operate directly on the gray values of the image. These gray values are considered as a high dimensional vector. Using the training samples, a principal component analysis (PCA) is computed to reduce the dimensions of the image vectors. by projecting them into the eigenspace.

In this paper, we compare different generic object modeling and classification techniques of the eigenspace vectors. At first, we present the supervised generic object recognition sys-

tem. In Section 3, we shortly describe the used classification techniques based on mixtures of probabilistic principal component analysis [5] (MPPCA), kernel principal component analysis (KPCA) [9] and Nearest Neighbor classification (NN) [6]. We show in Section 4 that the NN approach performs best, point out the flaws of the approaches and conclude in Section 5.

## 2. A supervised generic object recognition approach

The algorithms for automatic categorization of the training samples show that a generic classification is possible, but generally differs from human made categorizations [7]. This especially affects objects which vary more from different views than different objects vary from the same view. For example, a box and a car from the front view are more confusable than the side view and the front view of a car.

To gain meaningful generic classes, we have to use supervised approaches. The knowledge about the mapping of objects to generic classes can be used for feature selection with the Fisher transformation or for building a non-linear model for each generic class.

Figure 1 shows a sketch of the generic object recognition approach. In the training step, a PCA transformation is computed for each image set which reduces the dimensions of the images. In contrast to the unsupervised case, a model is generated for each generic class and

object class which is known to the system. Classification is done depending on the model type by maximum likelihood (ML) or NN classification. A rejection threshold is used to determine whether the objects are still represented by the generic model or not. If the object is not rejected, it will be classified in the next finer hierarchy level.

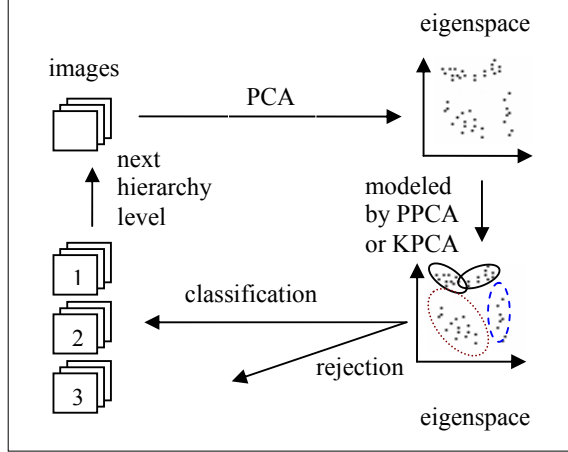


Fig. 1: Hierarchical approach for generic object recognition, dimensionality reduction with PCA, modeling with PPCA or KPCA, classification with ML or NN

### 3. Supervised generic object modeling with nonlinear methods

In general, the PCA transformed image vectors of a generic class do not form a compact cluster but a non-linear distribution. We use three methods to model these distributions: MPPCA, KPCA, and the NN approach.

The last approach just saves all trained eigenspace feature vectors of each model and calculates the nearest neighbor of the test feature vectors. The rejection threshold can be determined between zero and one using the distance measure

$$D(\mathbf{c}, \Omega_m) := \max_{\mathbf{z}_{i,m} \in \Omega_m} (\|\mathbf{c} - \mathbf{z}_{i,m}\| + 1)^{-1}, \quad (1)$$

between the test vector  $\mathbf{c}$  and the  $m$ -th model  $\Omega_m$  which contains the training vectors  $\mathbf{z}_{i,m} \forall i$ .

The MPPCA approach tries to overcome the non-linearity by combining linear PPCA models as described in [5, 7, 8]. A supervised approach is achieved using a mixture of PPCA for each generic class. On the basis of a posteriori probability the models are compared and the rejection threshold is defined.

To separate the non-linear distributed models from each other, the vectors  $\mathbf{c}$  can be projected with a non-linear function  $\varphi_m(\mathbf{c})$  into a

higher dimensional space in which the vectors can be separated linearly. The idea of the KPCA approach [9] is to avoid calculation of the projection by reducing the operations on the high-dimensional vectors to scalar products  $(\varphi_m(\mathbf{c}_x), \varphi_m(\mathbf{c}_y))$ . These can be evaluated in the original space using the so-called kernel function  $k(\mathbf{c}_x, \mathbf{c}_y) := (\varphi_m(\mathbf{c}_x), \varphi_m(\mathbf{c}_y))$ . This evaluation is only valid for kernel functions for which the mercer condition holds [9], like the polynomial kernel

$$k(\mathbf{c}_x, \mathbf{c}_y) := (\mathbf{c}_x, \mathbf{c}_y)^d, \quad d = 2, 3, \dots \quad (2)$$

and the Gaussian kernel

$$k(\mathbf{c}_x, \mathbf{c}_y) := \exp\left(-0.5\sigma^{-2}\|\mathbf{c}_x - \mathbf{c}_y\|^2\right). \quad (3)$$

For ML classification, we need the probability  $p(\mathbf{c} | \Omega_m)$  that a feature vector  $\mathbf{c}$  was generated by the non-linear model  $\Omega_m$ , which can be approximated by an energy value

$$E(\mathbf{c}, \Omega_m) \approx \ln(p(\mathbf{c} | \Omega_m)). \quad (4)$$

The energy value can be denoted by

$$E(\mathbf{c}, \Omega_m) := \tilde{\varphi}_m(\mathbf{c})^T \Sigma_{\tilde{\varphi}_m}^{-1} \tilde{\varphi}_m(\mathbf{c}) \quad (5)$$

with the covariance matrix  $\Sigma_{\tilde{\varphi}_m}^{-1}$  in the high dimensional space and the non-linear function

$$\tilde{\varphi}_m(\mathbf{c}) := \varphi_m(\mathbf{c}) - \underset{\forall i}{\text{mean}}(\varphi_m(\mathbf{z}_{i,m})), \quad (6)$$

which centers the projected vectors  $\varphi_m(\mathbf{c})$  by subtracting the mean of the projection of the training vectors  $\mathbf{z}_{i,m}$ .

The calculation of the energy value can be rewritten [10] and the scalar products of the nonlinear function can be replaced by the kernel function  $k(\mathbf{c}_x, \mathbf{c}_y)$ , which leads to the following form:

$$E(\mathbf{c} | \Omega_m) = \sum_{s=1}^r \left( \left( \sum_{i=1}^n \alpha_{i,m}^s \tilde{k}_m(\mathbf{z}_{i,m}, \mathbf{c}) \right)^2 \right. \\ \left. (\lambda_{s,m}^{-1} - \lambda_{\perp,m}^{-1}) + \lambda_{\perp,m}^{-1} \tilde{k}_m(\mathbf{c}, \mathbf{c}) \right), \quad (7)$$

where  $\lambda_{s,m}$  for  $s = 1, \dots, r$  are the  $r$  largest eigenvalues and  $\alpha_{i,m}^s$  are the components of the associated eigenvectors  $\alpha_m^s$ . The centered kernel function  $\tilde{k}_m(\cdot, \cdot)$  can be denoted by

$$\tilde{k}_m(\mathbf{c}_x, \mathbf{c}_y) := k(\mathbf{c}_x, \mathbf{c}_y) - \\ \frac{1}{n} \sum_{j=1}^n (k(\mathbf{c}_x, \mathbf{z}_{j,m}) + k(\mathbf{c}_y, \mathbf{z}_{j,m})) + \\ \frac{1}{n^2} \sum_{j,l=1}^n k(\mathbf{z}_{j,m}, \mathbf{z}_{l,m}). \quad (8)$$

The parameter  $\lambda_{\perp,m}$  is a regularization term which expresses the assumed variance in the orthogonal directions. To normalize the energy value of the Gaussian kernel, we set

$$\lim_{\|c-z_{j,m}\| \rightarrow \infty} k(c, z_{j,m}) = 0 \quad \forall z_{j,m}, \quad (9)$$

which ensures that an infinitely distant feature vector will get the same energy value. We obtain the normalization factor  $\kappa_m$  as:

$$\kappa_m = \sum_{s=1}^r \left( \left( \sum_{i=1}^n \alpha_{i,m}^s (\bar{\mu} - \mu_i) \right)^2 \right. \\ \left. (\lambda_{s,m}^{-1} - \lambda_{\perp,m}^{-1}) + \lambda_{\perp,m}^{-1} (1 + \bar{\mu}) \right) \quad (10)$$

$$\mu_i = \frac{1}{n} \sum_{j=1}^n k(z_i, z_j), \quad \bar{\mu} = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

#### 4. Experimental evaluation and results

We use the recognition rate of the first hierarchy to compare different generic object recognition approaches. Using a hierarchical approach, there are multiple definitions of a recognition rate. One is the recognition rate at a hierarchy level, which evaluates the classification into categories of that level. Another is the global generic recognition rate, where an object is classified correctly if a known object is classified into the right categories and the right object class, and an unseen object is classified into the right categories and are rejected in too specific categories where the object does not match. Thus a global generic recognition rate cannot be better than the recognition rate at the first hierarchy level.

As outlined in Fig. 1 we reduced the input dimensions using PCA and examined the recognition rates of the NN classifier, the KPCA with a Gaussian kernel as described in Sect. 3, and a mixture of 5 one-dimensional PPCA models as described in [5, 7, 8]. For testing, we divided the COIL-100 [11] database into three test sets, the training set — used for verification — a test set which contains untrained images of known objects, and a generic test set with images of untrained objects.

Comparing the best results (see Fig. 2), we determined that the NN approach is able to classify almost all objects correctly. The KPCA classifies the training set correctly and performs very well on the test set but fails on the generic test set. The PPCA model does not represent the training set completely. The rec-

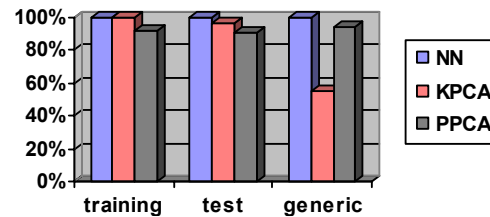


Fig. 2: Best recognition rates with different non-linear approaches using different test sample sets

ognition rate of the training and test set is about 90%. But unlike the KPCA approach the PPCA model provides good results at the generic test set.

The progression of the curves of recognition rates depending on the eigenspace dimension (Fig. 3) shows that the PPCA model breaks down at about 100 dimensions, which is due to curse of dimensionality while building the model which is described in [7]. NN and the KPCA model achieve a high recognition rate on the test set with 15 dimensions and more. On the generic test set, (Fig. 4) the NN

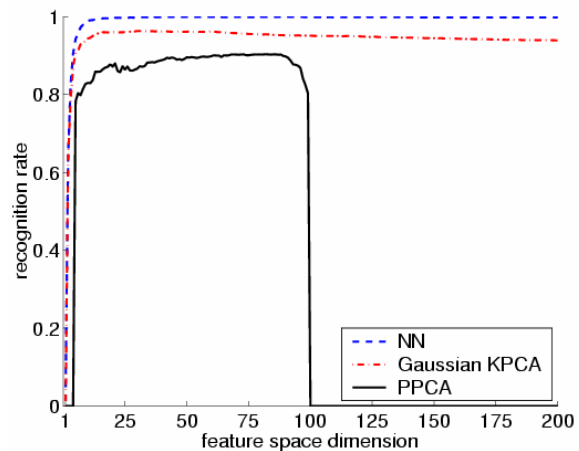


Fig. 3: Recognition rate at the first hierarchy level using the test set

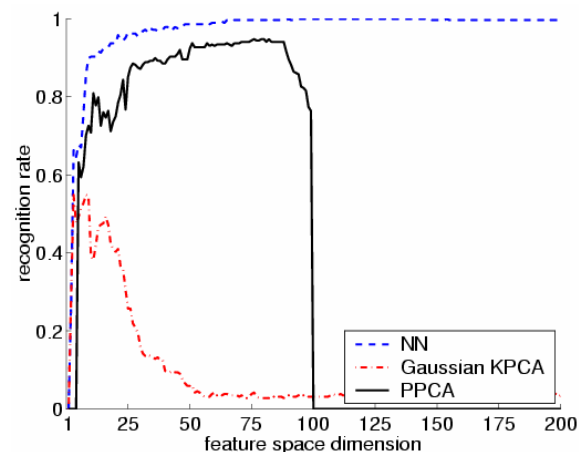


Fig. 4: Recognition rate at the first hierarchy level using the generic test set

and the PPCA model with 70 dimensions achieve good results while the KPCA model breaks down. This is because the model is specialized and does not generalize onto a generic object class.

Regarding the computation time (Table 1), we determine a linear increase for the PPCA and the NN approach and an almost constant computation time for the KPCA approach. But for 50 dimensions, the computation time of the NN approach is 26 times faster than the KPCA approach and the PPCA approach is even 400 times faster.

**Table 1: Mean computation time per test image in seconds on 50, 99 and 200 eigenspace dimensions**

dim. \ approach	50	99	200
PPCA	$3.61 \cdot 10^{-4}$	$7.36 \cdot 10^{-4}$	—
NN	$5.52 \cdot 10^{-3}$	$1.35 \cdot 10^{-2}$	$3.21 \cdot 10^{-2}$
KPCA	$1.45 \cdot 10^{-1}$	$1.51 \cdot 10^{-1}$	$1.58 \cdot 10^{-1}$

We also checked the robustness of the NN approach by disturbing the image sets with Gaussian and pixel noise and calculating the recognition rates (Table 2). We determined no significant degradation in recognition rates when using more than 10 feature dimensions.

**Table 2: Maximum degradation of recognition rate due to noise using the NN approach on eigenspace dimensions equal or greater than 10**

noise \ set	Training	test	generic test
Gaussian	0.00 %	0.26 %	0.69 %
Pixel	0.52 %	1.00 %	2.08 %

The best global generic recognition rate 97.1% is achieved by the NN approach.

## 5. Conclusion

As it was shown, the KPCA approach performs very well with known objects, but does not generalize well with unknown generic objects. Performance can be improved by varying  $\sigma$ , which is shown in the full paper. The NN approach performs very well, as the generalization is good and robust against noise. Unfortunately, the approaches cannot handle large databases, because they save all training vectors. Quantization of training vectors or leaving out unimportant vectors may solve this problem.

The classification rate of PPCA is worse than that of the other methods but can be im-

proved by using models with higher feature space dimensions or more sub-models. The approach generalizes well and is the fastest one examined. Additionally, it can handle large databases and provides a probability measure.

## References

- [1] Q. Iqbal, and J. K. Aggarwal. *Retrieval by Classification of Images Containing Large Manmade Objects Using Perceptual Grouping*, In *Pattern Recognition Journal*. Vol. 35, no. 7. pp. 1463-1479, July 2002.
- [2] G. Medioni and A. Francois. *3-D Structures for Generic Object Recognition*. In *Volume 1. Computer Vision and Image Analysis*, International Conference on Pattern Recognition, S. 30-37. 2000.
- [3] R. C. Nelson, and A. Selinger. *A Cubist approach to Object Recognition*. In *Proceedings of the 6<sup>th</sup> International Conference on Computer Vision (ICCV)*, pp. 614-621. IEEE Computer Society Press, Bombay, January 1998.
- [4] S. Dickinson, R. Bergevin, I. Biederman, J. Eklundh, R. Munck-Fairwood, A. Jain, and A. Pentland. *Panel report: The potential of geons for generic 3-d object recognition*. In *Image and Vision Computing*, 15, no. 4. pp. 277-292, April 1997.
- [5] M. E. Tipping, C. M. Bishop. *Mixtures of Probabilistic Principal Component Analysers*. In *Neural Computation*, vol. 11, no. 2: pp 443-482, 1999.
- [6] T. M. Cover, and P. E. Hart. *Nearest neighbor pattern classification*. In *IEEE Transaction on Information Theory*, vol. 13: pp. 21-27, 1967.
- [7] C. Drexler, F. Mattern, and J. Denzler. *Appearance Based Generic Object Modeling and Recognition Using Probabilistic Principal Component Analysis*. In *Proceedings of the 24th DAGM Symposium of Pattern Recognition*, pp. 100-108, Zurich, September 2002.
- [8] C. Drexler, F. Mattern, and J. Denzler. *Generic Hierarchic Object Models and Classification based on Probabilistic PCA*. In *IAPR Workshop on Machine Vision Applications*, pp. 435-438, Nara, Japan, December 2002.
- [9] B. Schölkopf, A. Smola, and K.-R. Müller. *Nonlinear component analysis as a kernel eigenvalue problem*. *Neural Computation*, vol. 10, pp. 1299-1319, 1998.
- [10] D. Cremers, T. Kohlberger, C. Schnörr. *Nonlinear Shape Statistics in Mumford-Shah Based Segmentation*. In *European Conference on Computer Vision*, pp. 93-108, Copenhagen, Denmark, May 2002.
- [11] S. A. Nene, S. K. Nayar, and H. Murase. *Columbia Object Image Library (COIL-100)*. Technical Report No. CUCS-006-96, Department of Computer Science, Columbia University, 1996.