Pattern Recognition and Information Processing (PRIP'2003), Minsk, Belarus, March 2003, Volume 1, pages 73-77

CONTINUOUS PARAMETRIZATION OF NORMAL DISTRIBUTIONS FOR IMPROVING THE DISCRETE STATISTICAL EIGENSPACE APPROACH FOR OBJECT RECOGNITION

Ch. Gräßl[1], F. Deinzer[2] and H. Niemann

Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg,
Martensstraße 3, 91058 Erlangen, Germany
{graessl,deinzer}@informatik.uni-erlangen.de

Abstract. Statistical approaches play an important role in computer vision, normal distributions especially are widely used. In this paper we present a new approach for a continuous parametrization of normal distributions. Our method is based on arbitrary interpolation techniques. This approach is used to improve the discrete statistical eigenspace approach for object recognition. The continuous parametrization of normal distributions allows an estimation of object poses where no training images were available. In an experiment with real objects we will show that our continuous approach leads to better localization and classification results than the discrete approach.

1 Introduction

Within the wide area of computer vision, object recognition is still one of the main topics of current research. Approaches for object recognition can mainly be divided into two directions. Firstly segmentation based techniques which detects for example geometric features can be used for object recognition [5]. But segmentation approaches suffer from the disadvantage that segmentation errors may occur which disturb the recognition process and from the general problem that often significant information is lost.

The second direction is that of appearance based approaches [9, 2, 7, 3, 8]. They avoid these disadvantages since they directly use the image data, e.g. pixel intensities, for the recognition process. There exist some well known approaches which uses multi-resolution wavelet features [9], Gaussian mixtures for classification [2] and the eigenspace approach [7] which was extended with a statistical component in [3, 8]. One primary disadvantage of most of those approaches is, that only those poses of objects that have been seen during the training process are known. That means for the statistical eigenspace approach that only poses can be computed which have been seen during the training.

Our approach shows how one can efficiently parametrise and interpolate normal distributions in a very general way. We will show how to retain the necessary properties of normal distributions, like the positive definiteness of the covariance matrix. Finally we will use this mathematical method to extend the statistical eigenspace to allow an improved recognition and continuous localization.

2 The Discrete Statistical Eigenspace Approach

The eigenspace approach based on [7] is an appearance-based method which uses a *Karhunen-Loeve Transformation* [6] (also known as *Principal Components Analysis*) to obtain a linear system to compute a feature by $\boldsymbol{c} = \Phi(\boldsymbol{f} - \bar{\boldsymbol{f}})$. The vector $\boldsymbol{f}$ contains the intensities of the

[1]This work was partially funded by the European Commission 5th IST Programme - Project VAMPIRE

[2]This work was partially funded by the German Science Foundation (DFG) under grant SFB 603/TP B2. Only the authors are responsible for the content.

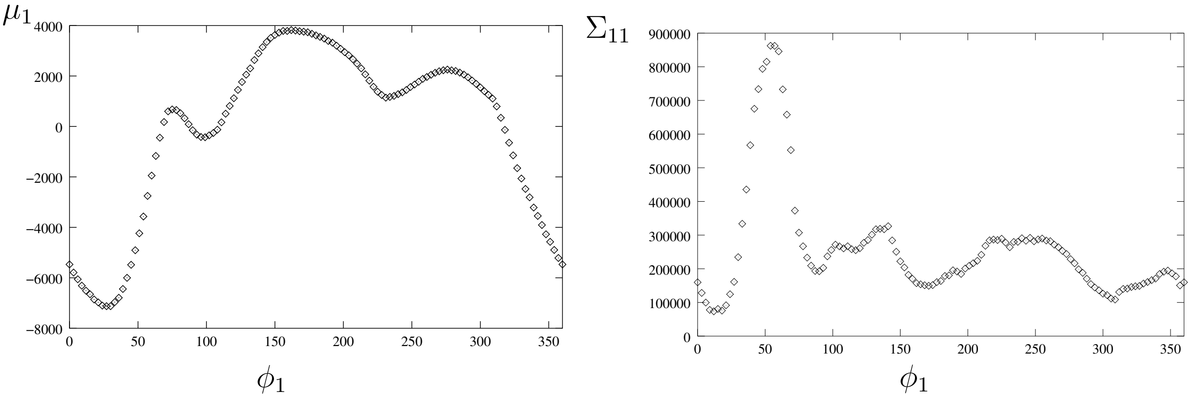

Fig. 1: The graph of an element of the mean vector (left) and of the covariance matrix (right). The parameter ϕ_1 describes the 1-D pose of an object which was placed on a turntable.

pixels of the object image. The matrix $\mathbf{\Phi}$ contains N_E eigenvectors (with the largest eigenvalues) of the covariance matrix of the training images. The average of all training images is denoted as $\bar{\boldsymbol{f}}$.

The feature $\boldsymbol{c}^i_\kappa$ and pose parameter $\boldsymbol{\phi}^i_\kappa$ of every training image $\boldsymbol{f}^i_\kappa$, where κ denotes the object's class and i is an index of the training image, are calculated during a training step. For classification and localization [7] searches for the training image whose feature has the smallest distance to the feature of the test image.

The classical eigenspace approach [7] has some disadvantages. On the one hand, the distance in the eigenspace is not well suited for our purposes, because all elements of $\boldsymbol{c}$ have the same influence on the distance, but only the first contain significant information. On the other hand, this method is not robust to noise, reflection and highlights. A more robust classification was presented in [9] where $\boldsymbol{c}^i_\kappa$ is replaced by a normal distribution

$$p(\boldsymbol{c}|\boldsymbol{B}^i_\kappa) = \mathcal{N}(\boldsymbol{c}|\boldsymbol{\mu}^i_\kappa, \mathbf{\Sigma}^i_\kappa) \tag{1}$$

where $\boldsymbol{B}^i_\kappa$ (which is estimated during a training step) consists of a mean vector $\boldsymbol{\mu}^i_\kappa$ and a covariance matrix $\mathbf{\Sigma}^i_\kappa$. Classification and localization is done by a maximum likelihood estimation.

3 Parametrisation of Normal Distributions

The major disadvantage of the discrete statistical eigenspace approach is that only poses can be estimated from where training images exists. This leads to a systematical localization error. We propose a parametrization of the parameters of the normal distribution to obtain continuous parameters for poses. Fig. 1 shows that the elements of the mean vector and covariance matrix are similar for similar object images, what is a requirement for an interpolation. So $\boldsymbol{B}$ can be parametrized by

$$\boldsymbol{B}(\kappa, \boldsymbol{\phi}) = (\boldsymbol{\mu}(\kappa, \boldsymbol{\phi})\ ,\ \mathbf{\Sigma}(\kappa, \boldsymbol{\phi}))\,. \tag{2}$$

Note that $\boldsymbol{\phi}$ is a vector with continuous values. This allows a classification and localization by solving the optimization problem

$$(\kappa^*, \boldsymbol{\phi}^*) = \underset{\kappa, \boldsymbol{\phi}}{\operatorname{argmax}}\ p\left(\boldsymbol{c}|\boldsymbol{B}(\kappa, \boldsymbol{\phi})\right)\ =\ \underset{\kappa, \boldsymbol{\phi}}{\operatorname{argmax}}\ \mathcal{N}\left(\boldsymbol{c}|\boldsymbol{\mu}(\kappa, \boldsymbol{\phi}), \mathbf{\Sigma}(\kappa, \boldsymbol{\phi})\right). \tag{3}$$

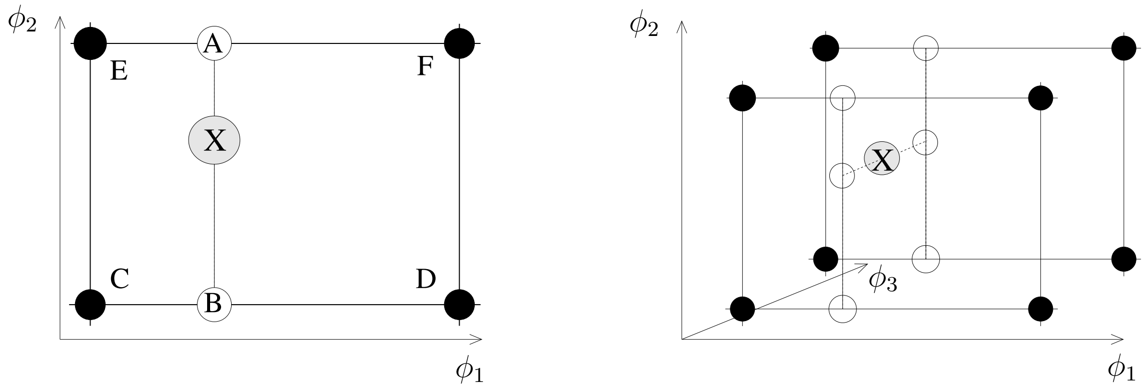

Fig. 2: Reduction of a 2-D (left) and 3-D (right) to an one-dimensional interpolation

The mean vector can be calculated by interpolating its components independently, using an interpolation technique which is presented in the next section. The components of the covariance matrix can not be interpolated independently, because the positive definiteness may be lost. So two ways of parametrizations are suggested:

- **Discrete interpolation:** Use the covariance matrix of the corresponding training image which pose $\boldsymbol{\phi}^i_\kappa$ has the lowest distance. This approach is maintainable, because the influence of the mean vector is more important.

- **Cholesky factorization:** Every positive definite matrix $\boldsymbol{\Sigma}$ can be factorized by $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$. The matrix $\boldsymbol{L}$ is a lower triangular matrix, which can be parametrized per component. The covariance matrix can be calculated by

$$\boldsymbol{\Sigma}(\kappa, \boldsymbol{\phi}) = \boldsymbol{L}(\kappa, \boldsymbol{\phi}) \left(\boldsymbol{L}(\kappa, \boldsymbol{\phi})\right)^T . \tag{4}$$

  The product of a matrix and its transponent is always positive definite. Algorithms for calculating $\boldsymbol{L}$ and the proof of the properties are presented in [4].

4 Interpolation Techniques

In the last section, a method for parametrization that is independent from the interpolation technique was presented. There exist many methods for interpolating data. We made many experiments, even with scattered pose parameters of the training images, but for lack of space, we restrict the methods to pose parameters which lies on a regular grid. We used two 1-D interpolation methods for interpolating the components of the mean vector and the left triangle matrix and extend them to a n-D interpolation: Linear interpolation and *Catmull-Rom* spline (CRS) [1] interpolation. The linear interpolation is very fast but not continuously differentiable. The CRS interpolation is also fast because of its polynomial character and continuously differentiable. A continuously differentiable interpolant is more realistic than one which is not (cf. Fig. 1) In contrast to other interpolation techniques, one does not have to solve a linear system, which makes CRS interpolation very flexible.

The 1-D interpolation can be easily enhanced to n dimensions by using a dimension-descent technique as shown in Fig. 2. The left picture shows the dimension-descent for 2-D pose parameters for the linear interpolation. The filled points are pose parameters of training images and the point *X* should be interpolated. To do this by a linear 1-D interpolation, point *A* has to be interpolated from *E* and *F*. Similar, point *B* can be interpolated from *C*

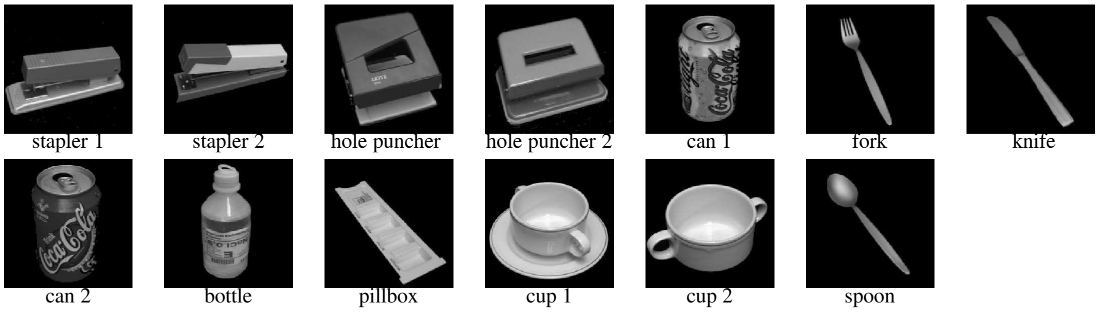

Fig. 3: The DIROKOL image database [10]

and D. Now it is possible to interpolate the searched point X by using A and B. The right picture of figure 2 shows the dimension-descent for 3-D pose parameters. Of course, the dimension-descent can be adapted to 1-D CRS.

Note that any other interpolation technique can be used for our purpose and the two examples are used to show how to use the presented method of parametrization of normal distributions. Doubtless there exist other methods of interpolation which could bring better classification rates.

5 Experiments and Results

We performed experiments on the *DIROKOL* image database [10] which consists of 13 real objects (1860 training images and 1860 test images per object) with a resolution of 256 x 256 shown in Fig. 3. For the image acquisition a turntable and a robot arm were used, which allows images to be taken from a hemisphere. Also the objects are illuminated at three different lighting conditions. For training, 1860 images per object were taken whose corresponding pose parameter which lies on a regular grid. We use a *Linux PC* with an *Athlon XP* (1.68 GHz) processor and 1GB memory for the experiments. For the optimization problem in equation (3), we applied an adaptive random search algorithm [11], followed by a simplex step.

We compared classification by using Catmull-Rom splines with linear interpolation. Additionally, we considered interpolating the covariance matrix discretely instead of using an interpolation technique. For demonstration of the improvement of the continuous methods, the discrete statistical eigenspace approach was also tested. The results are presented in table 1 and show that a continuous model leads to better results than the discrete model. The best classification rate and pose estimation can be achieved by the Catmull-Rom spline interpolation of the mean vector and the covariance matrix. Using a linear interpolation for the mean vector and a discrete interpolation for the covariance matrix shows also a very good result, but takes only $\sim$10% of the computation time. The accuracy is given with the so called *percentile* 95 values, which describe the maximal localization error if the classification is correct and only the 95% best localizations are taken into account. Note that the classification time is linearly dependent on the number of classes.

6 Conclusion and Outlook

We have presented an enhancement of the discrete statistical eigenspace approach to a continuous model. For parametrization of normal distributions, a Cholesky factorization was used to ensure that the covariance matrix keeps the positive definiteness, independently from the interpolation method. For interpolation, we used a linear interpolation and Catmull-Rom

Method	**Classification rate**	**Pose estimation accuracy**	**Classification time**
discrete statistical	97.6 %	$1.162°$	$< 0.01s$
$\boldsymbol{\mu}$: linear, $\boldsymbol{\Sigma}$: discrete	99.0%	$0.188°$	$3.38s$
$\boldsymbol{\mu}$: linear, $\boldsymbol{\Sigma}$: linear	99.3%	$0.178°$	$17.81s$
$\boldsymbol{\mu}$: CRS, $\boldsymbol{\Sigma}$: discrete	99.1%	$0.183°$	$15.73s$
$\boldsymbol{\mu}$: CRS, $\boldsymbol{\Sigma}$: CRS	99.4%	$0.149°$	$30.42s$

Table 1: Classification rate, accuracy of pose estimation (percentile 95 values) of successfully classified images and optimization time per class using the DIROKOL image database.

splines. Experiments were done on real images to prove that the continuous model is better than the discrete. It was shown that Catmull-Rom spline interpolation of all elements of the parameters of the normal distributions provides the best classification rate and pose estimation. Further work will focus on using other techniques of parametrization of normal distributions. In addition other interpolation techniques, especially for scattered pose parameters of the training images, will be tested.

References

[1] E. Catmull and R. Rom. A Class of Local Interpolating Splines. *Computer Aided Geometric Design*, pages 317–326, 1974.

[2] J. Dahmen, R. Schlüter, and H. Ney. Discriminative Training of Gaussian Mixtures for Image Object Recognition. In *21. DAGM Symposium Mustererkennung, W. Förstner, J. Buhmann, A. Faber, P. Faber (eds.)*, pages 205–212, Bonn, Germany, September 1999.

[3] F. Deinzer, J. Denzler, and H. Niemann. On Fusion of Multiple Views for Active Object Recognition. In B. Radig and Stefan Florczyk, editors, *Mustererkennung 2001, 23. DAGM-Symposium, München*, pages 239–245, Berlin, September 2001. Springer.

[4] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. The Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.

[5] J. Hornegger and H. Niemann. Statistical Learning, Localization, and Identification of Objects. In *Proceedings of the Fifth International Conference on Computer Vision (ICCV)*, pages 914–919, Boston, Massachusetts, USA, 1995. IEEE Computer Society Press.

[6] M. Loeve. *Probability Theory*. Springer, New York, 4th edition, 1978.

[7] H. Murase and S. Nayar. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[8] L. Paletta and A. Pinz. Active Object Recognition by View Integration and Reinforcement Learning. *Robotics and Autonomous Systems*, 31((1–2)):1–18, 2000.

[9] M. Reinhold, F. Deinzer, J. Denzler, D. Paulus, and J. Pösl. Active Appearance–Based Object Recognition Using Viewpoint Selection. In B. Girod, G. Greiner, H. Niemann, and H.-P. Seidel, editors, *VMV 2000*, pages 105–112. infix, Berlin, November 2000.

[10] M. Reinhold, Ch. Drexler, and H. Niemann. Image Database for 3-D Object Recognition. Technical Report LME-TR-2001-02, Informatik 5 (Lehrstuhl für Mustererkennung), Universität Erlangen-Nürnberg, Mai 2001.

[11] A. Törn and A. Žilinskas. Global Optimization. *Lecture Notes in Comuter Science, Springer-Verlag, Berlin*, 3350, 1989.