6th German-Russian IAPR Workshop on Pattern Recognition and Image Understanding,
Novosibirski, Russion Federation, 2003, pages 38-41

IMPROVING STATISTICAL OBJECT RECOGNITION APPROACHES BY A PARAMETERIZATION OF NORMAL DISTRIBUTIONS

Ch. Gräßl[1], F. Deinzer[2], F. Mattern[3], and H. Niemann

Chair for Pattern Recognition, University of Erlangen–Nuremberg, 91058 Erlangen, Germany
{graessl, deinzer, mattern}@informatik.uni-erlangen.de

Summary. As statistical approaches play an important role in object recognition, we present a novel approach which is based on object models consisting of normal distributions for each training image. We show how to parameterize the mean vector and covariance matrix independently from the interpolation technique and formulating the classification and localization as a continuous optimization problem. This enables the computation of object poses which have been never seen during the training. For interpolation, we present four different techniques which are compared in an experiment with real images. The results show the benefit of our method both in classification rate and pose estimation accuracy.

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 the two directions of segmentation based and appearance based approaches. The segmentation based techniques have a long history in computer vision. They detect, for example, geometric features that can be used for object recognition [7] but suffer from the disadvantage that segmentation errors may occur that disturb the recognition process. Additionally they have the general problem that significant information may be lost.

Appearance based approaches [1,2,8,9] 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 [10], Gaussian mixtures for classification [1], and the eigenspace approach [8], which was extended with a statistical component in [2]. 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.

Our approach shows how one can efficiently parameterize 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. We will also detail how one can integrate different interpolation techniques (e.g. spline interpolation, radial basis functions and triangulation) into our approach. We will demonstrate how to use the presented method to extend an existing object recognition system to a continuous parameter space by means of the statistical eigenspace approach. The result will show that this extension will significantly improve the recognition and localization result.

2. Continuous Statistical Eigenspace Approach

A widely used appearance based method for feature extraction is the eigenspace approach [8] which uses a linear system to compute a feature $\boldsymbol{c} \mid \bar{\boldsymbol{u}}\boldsymbol{f}$, where the vector $\boldsymbol{f}$ comprises the pixel intensities of an image and the matrix $\bar{\boldsymbol{u}}$ contains the eigenvectors of the training images. During a training step, the features $\boldsymbol{c}_\rho^i$ and the pose parameters $\boldsymbol{\phi}_\rho^i$ of the i-th image $\boldsymbol{f}_\rho^i$ of class ρ are collected. At runtime, the feature vector of the test image is computed and all Euclidian distances to the feature vec-

[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.
[3] This work was partially funded by the German Science Foundation (DFG) under grant DE 735/2-1. Only the authors are responsible for the content.

tors of the training images are calculated. The result of classification and localization is the class and pose of the closest training image.

The usage of the Euclidian distance of eigenspace features is not well suited for a measurement of similarity of object images, because the elements of $\boldsymbol{c}$ are not as same significant. [2] shows how to improve the classical eigenspace approach by a statistical model. As the feature of an image with a small perturbation (e.g. noise or translation) is close to the feature of the original image, it is presumed that the features of perturbed images are normal distributed. Instead of features $\boldsymbol{c}_\rho^i$ a mean vector $\boldsymbol{\mu}_\rho^i$ and a covariance matrix $\boldsymbol{\Sigma}_\rho^i$ is estimated during the training for every training image. The result of classification and localization is the class and pose of the training image with the maximum likelihood of $p\left(\boldsymbol{c}\middle|\boldsymbol{\mu}_\rho^i, \boldsymbol{\Sigma}_\rho^i\right)$.

The disadvantage of the statistical eigenspace approach is that only poses can be evaluated where training images exists, which leads to a systematical pose estimation error. We have shown in [5] that the elements of $\boldsymbol{\sigma}$ and $\emptyset$ are similar for object images with similar pose and have parameterized the normal distribution dependent from class ρ an the continuous pose $\boldsymbol{\alpha}$. Classification and localization can now be formulated as a continuous optimization problem

$$\left(\bar{\rho}, \bar{\boldsymbol{\alpha}}\right) = \underset{(\rho,\boldsymbol{\alpha})}{\operatorname{argmax}}\, p\left(c \middle| \boldsymbol{\mu}(\rho,\boldsymbol{\alpha}), \boldsymbol{\Sigma}(\rho,\boldsymbol{\alpha})\right). \qquad (1)$$

Obviously the parameterization of the mean vector can be done by interpolating its components independently using $\boldsymbol{\mu}(\rho,\boldsymbol{\alpha}) = \left(\mu_1(\rho,\boldsymbol{\alpha}), \mu_2(\rho,\boldsymbol{\alpha}), \ldots\right)^T$. In contrast, the elements of the covariance matrix cannot be interpolated independently, because the positive definiteness may be lost. One possibility is to use the covariance matrix of the nearest training image. A more exact but also computational more expensive method is the usage of a *Cholesky factorization*, as every positive definite matrix can be factorized to $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$. It is possible to compute a left triangle matrix $\boldsymbol{L}$ in an offline step. The components of matrix $\boldsymbol{L}$ are interpolated independently by

$$\boldsymbol{L}(\rho,\boldsymbol{\alpha}) = \begin{pmatrix} L_{1,1}(\rho,\boldsymbol{\alpha}) & 0 & \cdots \\ L_{2,1}(\rho,\boldsymbol{\alpha}) & L_{2,2}(\rho,\boldsymbol{\alpha}) & \\ \vdots & & \ddots \end{pmatrix}$$

and the parameterized covariance matrix can be calculated using

$$\boldsymbol{\Sigma}(\rho,\boldsymbol{\alpha}) = \boldsymbol{L}(\rho,\boldsymbol{\alpha})\left(\boldsymbol{L}(\rho,\boldsymbol{\alpha})\right)^T. \qquad (2)$$

The product of a matrix and its transponent is always positive definite. Algorithms for the Cholesky factorization and the proof of the properties are presented in [6].

3. Interpolation Techniques

In principal our method of parameterization is independent from the underlying interpolation technique; nevertheless we are presenting in this chapter methods of interpolation. We have implemented a linear interpolation and a *Catmull-Rom* spline (*CRS*) interpolation, which assumes, that the pose parameters of the training images lie on a regular grid (for details consult [5]). As this property cannot always be assumed, we have investigated *trilinear* and *radial basis function* (*RBF*) interpolation for working with scattered data which are detailed in this chapter.

Interpolation of scattered data is an intensively researched field in the area of computer graphics. Most of those techniques require a two dimensional triangle net for the parameterization which means that for our purpose we are limited to a 2-D pose parameter space. It is important that the mesh generator does not create acute-angled triangles which would lead to bad interpolation results. Therefore we use a *Delaunay refinement* [11] to restructure an arbitrary triangle net and improving the suitability for interpolation. Two examples for triangulations are given in Fig.1.

One interpolation technique which uses triangle nets is the trilinear interpolation. The interpolant is defined as

$$\sigma_n(\rho,\boldsymbol{\alpha}) = u_j \sigma_{\rho,n}^j + u_k \sigma_{\rho,n}^k + u_l \sigma_{\rho,n}^l, \qquad (3)$$

where u_j, u_k, u_l are the barycentric coordinates of $\boldsymbol{\alpha}$, which is in the center of the triangle with corner $\boldsymbol{\alpha}_\rho^j, \boldsymbol{\alpha}_\rho^k$ and $\boldsymbol{\alpha}_\rho^l$. The elements of the matrix $\boldsymbol{L}$ in (2) can be treated the same way as the components of the mean vector.

Another interpolation technique for scattered data which is applicable in parameter

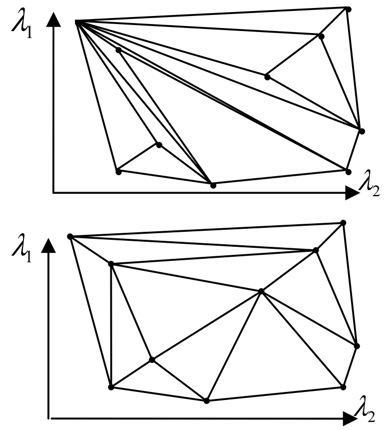

Fig.1: Top: Triangle net using a awkward triangulation algorithm. Bottom: Triangle net with delaunay refinement which is better suitable for interpolation. Dots marks pose parameters of training images.

spaces of arbitrary size is the interpolation with radial basis functions [3], which have been intensively researched in the numerical mathematic. The interpolation rule is

$$\sigma_n\left(\rho,\lambda_\rho^i\right) = \sigma_{\rho,n}^i \;\forall i,\rho \quad , \tag{4}$$

where n denotes the component of the mean vector. The interpolant is defined as

$$\sigma_n\left(\rho,\lambda\right) = \sum_i w_{\rho,n}^i h\left(d\left(\lambda,\lambda_\rho^i\right)\right) \quad , \tag{5}$$

where $h:\mathrm{IR} \to \mathrm{IR}$ is the so called radial basis function, $d(\cdot,\cdot)$ is a distance measurement (we use the Euclidian distance), and $w_{\rho,n}^i$ is a weighting coefficient. For details of the computation of $w_{\rho,n}^i$ consult [3]. There exists dozens of different radial basis functions, for our purpose we use

$$h(x) = \exp\left(-\frac{x^2}{\vartheta}\right), \tag{6}$$

because it depends on only one adjustment parameter. If the value of ϑ is large, the elements of the normal distributions which have a large distance to λ have a higher influence in comparison with a small ϑ.

5. Experiments and Results

For proving the benefit of our work, we performed experiments on the DIROKOL image database [10] which consists of 13 real objects (office and health care domain) which are presented in Fig.2. The image database consists of 1860 training images and 1860 test images per object with a resolution of 256x256. The pose parameters of the training acquisition a turntable and a robot arm were used, which allows images to be taken from a hemisphere. As three different illumination conditions have been used, we limited to one third of the image set (all with the same illumination condition). Experiments with the full set have already been presented in [5]. We used an eight dimensional eigenspace for the PCA and for the optimization problem in (1), an adaptive random search algorithm, followed by a simplex step, has been applied. The computation took place on a Linux PC with a Pentium 4 processor (2.4 GHz) and 1GB memory.

Fig.2: The DIROKOL image database [10]

We compared classification rate and pose estimation accuracy of the linear, CRS, trilinear and RBF interpolation. Thereby a Cholesky factorization has not performed in all cases; those results will be shown in the full paper. Experiments for using the discrete statistical eigenspace approach have also been done to show the improvement of the continuous model. The results, which are presented in the table, show, except for the trilinear interpolation, that the continuous approach both in classification rate and pose estimation accuracy leads to better results than the discrete approach. The accuracy is given in the so called percentile 80 values which describe the maximal localization error if the classification is correct and only the 80% best localizations are taken into account. Disadvantageous is the high computational cost of the continuous method, because a lot of mean vectors and covariance matrices have to be interpolated for optimization of (1). Further, the results show that the trilinear interpolation is not well suited for interpolation of normal distributions.

Due to lack of space, we cannot present experiments on scattered pose parameter of the

Table: Classification rate, pose estimation accuracy (percentile 80 values) in degree, and classification time using the DIROKOL image database.

Method	Classification rate	Pose estimation accuracy	Classification time
Discrete statistical	**86.1%**	**9.61°**	**< 0.1sec**
μ: linear, Σ: discrete	**90.0%**	**6.79°**	**1.6s**
μ: linear, Σ: linear	**92.6%**	**6.18°**	**3.7s**
μ: CRS, Σ: discrete	**91.1%**	**6.59°**	**3.4s**
μ: CRS, Σ: CRS	**92.5%**	**6.00°**	**6.5s**
μ: triliniar, Σ: discrete	**75.1%**	**45.28°**	**4.5s**
μ: RBF, Σ: discrete, $\vartheta=100$	**90.5%**	**6.06°**	**9.7s**

training images. Those results will be shown in the final paper.

6. Conclusion and Outlook

In this paper, we presented a novel approach for parameterization of normal distributions which is applicable for object recognition algorithms based on statistical normal distributed object models. As the parameterization of the covariance matrix is dangerous, because the positive definiteness may be lost, we proposed the usage of a Cholesky factorization. For interpolation we used Catmull-Rom splines, linear interpolation, radial basis functions and trilinear interpolation. The last two techniques are also applicable on object models where the pose parameters of the training images are scattered. Experiments which have been performed on an image database with real images show, that the continuous model is superior to the discrete model. The two best classification and pose estimation rates has been reached with the Catmull-Rom spline interpolation and linear interpolation of all elements of the normal distribution.

Further research should concentrate on other methods for parameterization of normal distributions. Also interpolation techniques based on Beziér patches like the Clough-Tocher interpolation [4] should be evaluated. The usage of other radial basis functions like the inverse multiquadratic functions may be beneficial.

References

[1] 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. Gaber (eds.)*, pages 205–212, Springer, Berlin, Germany, 1999.

[2] F. Deinzer, J. Denzler and H. Niemann. On Fusion of Multiple Views for active Object Recognition. In *23. DAGM Symposium Mustererkennung, B. Radig, F. Florczyk (eds.)*, pages 239–245, Springer, Berlin, Germany, 2001.

[3] N. Dyn and C. A. Micchelli. Interpolation by sums of radial functions, in *Numerische Mathematik,* pages 1–9, Springer, Berlin, 2001.

[4] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design*. Academic Press, Boston, 3rd edition, 1993.

[5] Ch. Gräßl, F. Deinzer, and H. Niemann. Continuous parametrization of normal distributions for improving the discrete statistcal eigenspace approach for object recognition. In V. Krasnoproschin, S. Ablameyko, and J. Soldek, editors. *PRIP 2003*, pages 73–77, Vol. 1, Minsk, Mai 2003.

[6] 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.

[7] J. Hornegger and H. Niemann. Statistical Learning, Localization, and Identification of Objects. In *Proceedings of the 5th ICCV*, pages 914–919, Boston, Massachusetts, USA, 1995. IEEE Computer Society Press.

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

[9] M. Reinhold, F. Deinzer, J. Denzler, D. Paulus, and J. Pösel. Active Appearence-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), University of Erlangen-Nuremberg, 2001.

[11] J. R. Shewchuck. Triangle: Engeneering a 2D Quality Mesh Generator and Delaunay Triangulator in *WACG 1996*, LNCS, Springer, Berlin, 1996.