

# SCALE-INVARIANT IMAGE RECOGNITION BASED ON HIGHER ORDER AUTOCORRELATION FEATURES<sup>‡</sup>

M. KREUTZ, B. VÖLPEL and H. JANSSEN

Institut für Neuroinformatik, Lehrstuhl für Theoretische Biologie  
Ruhr-Universität Bochum, 44780 Bochum, Germany  
e-mail: kreutz@neuroinformatik.ruhr-uni-bochum.de

**Abstract**—We propose a framework and a complete implementation of a translation and scale invariant image recognition system for natural indoor scenes. The system employs higher order autocorrelation features of scale space data which permit linear classification. An optimal linear classification method is presented, which is able to cope with a large number of classes represented by many, as well as very few samples. In the course of the analysis of our system, we examine which numerical methods for feature transformation and classification show sufficient stability to fulfill these demands. The implementation has been extensively tested. We present the results of our own application and several classification benchmarks.

Image recognition      Face recognition      Scale invariancy      Scale space      Higher order  
autocorrelation      Optimal linear classification

## 1. INTRODUCTION

The task of visual recognition which was defined by Marr<sup>(1)</sup> with the question: "What objects are where in the environment?" is still difficult to fulfill.

The difficulties in answering that question are due to the changes the image of an object undergoes as a result of variations in viewing angle, viewing distance, illumination, etc. and the noise introduced by the imaging process. Thus, robust object recognition requires a high degree of invariancy with regard to these variations. There are plenty of approaches to cope with these difficulties. However, according to Ullman<sup>(2)</sup> a broad characterization of the different approaches can be done on account of the concept the invariancy capabilities are achieved:

First, there are the algorithms which rely on *invariant properties*<sup>(3, 4)</sup>.

The second class of algorithms aims at *object decomposition*, where an object is defined by its constituent parts.

Third there is the *alignment approach* which compensates for the mentioned variations by a matching transformation of either the image or the data base of models stored in memory.

The last class has proven to be very powerful if some a priori knowledge about the objects to be classified is available (faces, cars in traffic scenes, etc.) and the number of models in the data base is relatively small<sup>(5, 6, 7)</sup>. On account of the linear dependency of the number of transformations to be performed on the size of the data base and the limited set of allowable matching transformations these algorithms become unsuitable if either a priori knowledge is lacking or a huge number of different objects has

to be stored.

When designing a visual recognition system one should well define the capabilities it aims at. Technically spoken the parameter space representing the transformations the image of an object can undergo should be restricted.

The specific task we had in mind is a module for our autonomous robot vehicle MARVIN, which is equipped with an active stereo camera system<sup>(8)</sup>. It is designed as a preprocessing module for scene recognition and visual orientation. However, the module can easily be used for a variety of applications. Our demands are as follows:

- translation invariancy along the support of the image plane,
- scale invariancy within a range of factor 4,
- robustness against small rotations and slight changes of lighting conditions,
- ability to store a large number of different classes (order of 100) and to work with a low number of samples per class ( $\geq 2$ ).
- no a priori knowledge about the objects.

Thus our algorithm is a representative of the *invariant properties* algorithms.

Otsu and Kurita<sup>(9)</sup> proposed and extended<sup>(3)</sup> a recognition system which already meets some of our requirements and serves as the basis for our work. We present a framework and a complete implementation of an optimal linear classification scheme. This framework employs higher order autocorrelation functions of scale space data of natural images.

Due to the use of higher order autocorrelation functions the shift invariancy is guaranteed. The unsolved problems we focus on are:

- examine and extend scale invariancy,

<sup>‡</sup>Supported by the German Federal Department of Research and Technology (BMFT), NAMOS project and the KOGNET fellowship program of the *Deutsche Forschungsgesellschaft (DFG)*

- determine an appropriate preprocessing stage in addition with a reasonable feature vector extraction,
- develop an optimal as well as numerically stable linear classifier, which is not trivial due to the size of the feature space,
- proof the usefulness of the system by extensive experiments and benchmark testing.

The remainder of this paper is organized as follows. The next section presents the preprocessing and feature extraction steps of our approach. Section 3 discusses theoretical as well as numerical problems of optimal linear classification. The current implementation, our own results and several classification benchmarks are discussed in section 4. Section 5 summarizes and concludes.

## 2. FROM IMAGES TO FEATURES

### 2.1. SCALE INVARIANCY BY MEANS OF SCALE SPACE DATA

Scale space representations are widespread tools in the field of computer vision<sup>(10, 11)</sup>. They allow to speed up algorithms<sup>(12)</sup> and provide a high degree of flexibility with regard to changing algorithmic requirements. For our system we considered gaussian, laplacian and gabor pyramids. Note, that in addition to representing different frequency bands on separate scales as laplacian pyramids do, gabor pyramids split up the image along the orientation dimension too<sup>(13)</sup>.

Fig. 1 illustrates the idea of scale invariant object recognition by means of a bandpass octave scale space (laplace, gabor). Three different object distances  $d = \{1m, 2m, 4m\}$  are drawn at a time. The object consists of a frontoparallel plate, monofrequently textured with frequency  $k_0$  corresponding to the mean frequency of the coarsest scale. The resulting images for each object distance and the corresponding scale space representation are shown schematically. As the object distance is increased by a factor of 2 the area of the object on the camera image shrinks by a factor of 4. In addition the frequency increases by a factor of 2. This is,  $d \propto k_i$  with  $k_i$  denoting the image frequency. Hence, the representation of the object in the scale space is of same size just shifted to a higher frequency scale. A following recognition stage (see section 2.2), which treats object representations regardless of their position in the scale space will therefore be translation and scale invariant.

Obviously the preceding assumptions are valid only in the borders of appropriate scale sampling, i.e. the Nyquist theorem has to be obeyed. If this holds, even object representations distributed over two neighbouring scales, resulting from in between distances in the example of Fig. 1, could be used for interpolation. If the Nyquist theorem is violated aliasing would occur.

To achieve appropriate sampling lowpass filtering of the image spectrum (which is the *signal* in scale direction) has to be applied before sampling. Lowpass filtering and sampling of the image spectrum is merged in a single processing step by filtering the image with a set of bandpass filters with different mean frequencies. Laplacian pyramid filter characteristics are a close approximation to these requirements.

However, interpolation is very expensive. Therefore, we solely use  $\frac{1}{2}$ -octave scale space representations with 6 scales to ensure scale invariancy.

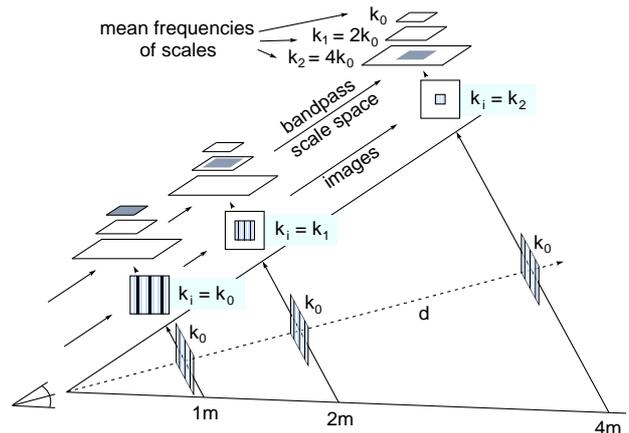


Figure 1: Shift of object representation along scale dimension under changing object distance. For further explanations see text.

### 2.2. FEATURES: AUTOCORRELATION FUNCTIONS

Based on Otsu *et al.*<sup>(3, 9)</sup> we use higher order autocorrelation functions as primitive features. Higher order autocorrelation functions are defined by:

$$R^n(\mathbf{a}_1, \dots, \mathbf{a}_n) = \int_P f(\mathbf{x}) f(\mathbf{x} + \mathbf{a}_1) \dots f(\mathbf{x} + \mathbf{a}_n) d\mathbf{x} \quad (1)$$

where  $n$  denotes the order of the autocorrelation function,  $\mathbf{x}$  is the image coordinate vector, and the  $\mathbf{a}_i$  are the displacement vectors, respectively.  $f(\mathbf{x})$  stands for the image intensity function on the retinal plane  $P$ . Autocorrelation functions are shift invariant; for a detailed analysis of the properties of higher order autocorrelation functions see McLaughlin and Raviv<sup>(14)</sup>. Our extension with respect to the work of Otsu and Kurita is that we accumulate feature vectors of different scales to a single feature vector of same dimension, instead of concatenating the feature vectors of all scales. Furthermore, we consequently treated the idea of a three dimensional object representation space by not only correlating along the two spatial dimensions, but also along the scale dimension. However, we tested our system with the other possible setups as well. Note, that in both cases the feature vectors are independent of the scale the single features stem from. Hence, we achieve our feature vector to be scale invariant as claimed in section 2.1.

However, there are some problems we have to cope with:

1. Otsu and Kurita emphasized the need to reduce the number of evaluated displacements  $\mathbf{a}_i$  for reasons of computation time. Using second order autocorrelation in 2D and exploiting the symmetry of the autocorrelation function one gets 29 features for a  $3 \times 3$  displacement region. A  $5 \times 5$  displacement region results in 205 features and a three-dimensional displacements in a  $3 \times 3 \times 3$  displacement region leads to 264 features.

Each feature  $R_j^2$  is characterized by a vector consisting of 3 2D or 3D displacement vectors

$$R_j^2 = R^2(a_{1j}, a_{2j}, a_{3j}) \quad (2)$$

respectively.  $R_j^2$  is independent of the sequence of the  $\mathbf{a}_i$ .

The increase of features forbids the evaluation of larger displacements as well as autocorrelations of order  $n > 2$ . Note, that this is not only because of the increased computation time, but also on account of the numerical problems which will be discussed in section 3.

2. Fig. 2 depicts the problems arising from 3D-displacement vectors. Correlation along different scales leads to drastic boundary and shear effects. Furthermore, if information present at one scale becomes distributed on two scales due to a distance shift of the object, the feature vector response evoked by that information will be shifted to other features.

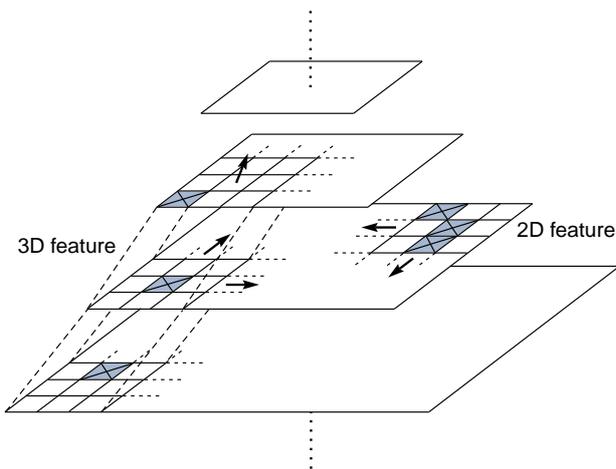


Figure 2: Scale sampling with exemplary second order autocorrelation features. A 2D feature corresponding to  $R(a_1, a_2, a_3) = R((0, 0)^T, (-1, 1)^T, (0, -1)^T)$  and a 3D feature corresponding to  $R(a_1, a_2, a_3) = R((0, 0, 0)^T, (0, 1, -1)^T, (-1, -1, 1)^T)$  are drawn. Displacement vectors are marked as dark squares. The features have to be shifted inside the scale space representation (black arrows).

### 3. CLASSIFICATION

The task of classification is to assign each image represented by its feature vector to one of a finite number of classes. The design of the classifier strongly depends on the used features. There is a trade-off between the complexity of feature extraction and classification<sup>(15)</sup>.

The Bayes classifier is known to be optimal in terms of minimizing the risk of misclassification<sup>(15)</sup>. However, it requires the full knowledge of all underlying probability distributions and leads in general to a non-linear model.

It is reasonable to examine the applicability of a linear classification scheme first. In the case of a linear Bayes classifier the feature vectors as proposed in section 2.2 have to be tested for the necessary and sufficient conditions. In particular multinormal distribution of classes with equal covariance matrices is a sufficient condition. Since a complete statistical analysis makes little sense due to the limited sample set we solely perform a test for normal distribution. The results are presented in section 4.1.

In order to obtain comparative results another linear classifier based on linear regression analysis is employed. By this approach the mean squared error is minimized.

When applying these methods directly to the feature vectors severe numerical problems arise. This is due to the small number of samples per class compared to the high dimension of the feature space. In section 3.2 we give a complete design of a linear classifier which is not only optimal in theory but also robust in practice even for extreme cases.

#### 3.1. OPTIMAL LINEAR CLASSIFICATION

Given  $M$  classes with a priori probabilities  $P_i$  and let  $N_i$  be the number of samples of class  $i$  we can define the following positive semi-definite scatter matrices

$$\mathbf{S}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} (\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})^T \quad (3)$$

$$\mathbf{W} = \sum_{i=1}^M P_i \mathbf{S}_i \quad (4)$$

$$\mathbf{B} = \sum_{i=1}^M P_i (\bar{\mathbf{x}}^{(i)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(i)} - \bar{\mathbf{x}})^T \quad (5)$$

$$\mathbf{T} = \mathbf{W} + \mathbf{B} \quad (6)$$

where  $\mathbf{x}^{(i)}$  denotes an  $n$ -dimensional feature vector belonging to class  $i$  with mean  $\bar{\mathbf{x}}^{(i)}$  and total mean  $\bar{\mathbf{x}}$ , respectively. In our case individual scatter matrices  $\mathbf{S}_i$  of class  $i$  can not be estimated separately due to the sparse sample data. In the remainder the individual scatter matrices are approximated by the mixture scatter matrix  $\mathbf{W}$  and  $P_i$  is estimated by  $1/N_i$ .  $\mathbf{B}$  and  $\mathbf{T}$  denote the between-class scatter matrix and the total scatter matrix, respectively.

The decision rule of a linear Bayes classifier in terms of the shortest Mahalanobis distance results in:

$$\begin{aligned} &\text{Assign } \mathbf{x} \text{ to class } i, \text{ if} \\ &(\mathbf{x} - \bar{\mathbf{x}}^{(i)})\mathbf{W}^{-1}(\mathbf{x} - \bar{\mathbf{x}}^{(i)})^\top = \quad (7) \\ &\min_{j=1,\dots,M} (\mathbf{x} - \bar{\mathbf{x}}^{(j)})\mathbf{W}^{-1}(\mathbf{x} - \bar{\mathbf{x}}^{(j)})^\top. \end{aligned}$$

If  $\mathbf{W}$  is transformed according to

$$\Phi^\top \mathbf{W} \Phi = \mathbf{I} \quad (8)$$

the Mahalanobis distance becomes equal to the Euclidian distance which greatly simplifies the evaluation of the decision rule. The result of this transformation is shown in Fig. 3. The derivation of the transformation matrix  $\Phi$  is postponed to section 3.2.

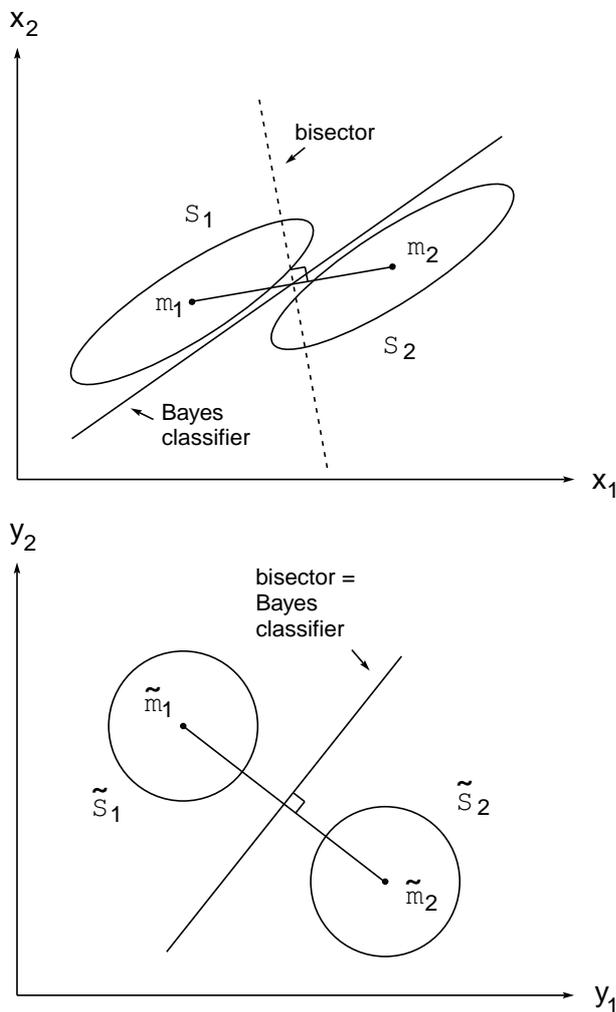


Figure 3: Effect of transformation  $\Phi$  (Eq. 8) on the feature space. Due to this transformation the Mahalanobis distance becomes equal to the Euclidian distance.

Since we assume a normal distribution, a subspace containing the  $M$  class centroids preserves the same class separability as the original feature space<sup>(16)</sup>. Thus, the dimension of this discriminant subspace is limited to  $M - 1$ . The Linear Discriminant Analysis (LDA)<sup>(17, 18, 19, 15)</sup> performs an axis rotation to

maximize the discriminatory power of each variable along these axes. The corresponding rotation matrix  $\Phi$  is obtained by solving the generalized eigenequation

$$\mathbf{B}\Phi = \mathbf{W}\Phi\Lambda, \quad (9)$$

where  $\Lambda$  is the diagonal matrix of the eigenvalues and the rows of  $\Phi$  contain the corresponding eigenvectors (for details see Peters and Wilkinson<sup>(20, 21)</sup>).

If the transformation defined by  $\Phi$  is applied to  $\mathbf{W}$  and  $\mathbf{B}$  Eq. 9 simplifies to a standard eigenequation

$$\Phi^\top \mathbf{B} \Phi \Upsilon = \Upsilon \Lambda. \quad (10)$$

Note, that Eq. 9 and Eq. 10 yield the same solution for  $\Lambda$  since eigenvalues are invariant under any one-to-one transformation.

If we represent the  $M$  classes by an arbitrary set of  $M$  orthogonal vectors  $\mathbf{y}_i$ , Linear Regression Analysis (LRA)<sup>(22, 18)</sup> determines a linear mapping  $\hat{\mathbf{y}} = \mathbf{A}^\top \mathbf{x} + \mathbf{b}$  so that the mean squared error between  $\hat{\mathbf{y}}_i$  and  $\mathbf{y}_i$  is minimized. The closed form solution is given by

$$\begin{aligned} \mathbf{A} &= \mathbf{T}^{-1} \mathbf{S}_{xy} \\ \mathbf{b} &= \bar{\mathbf{y}} - \mathbf{A}^\top \bar{\mathbf{x}}, \end{aligned} \quad (11)$$

where  $\mathbf{S}_{xy}$  denotes the cross-scatter matrix of  $\mathbf{x}_i$  and  $\mathbf{y}_i$ . If  $\mathbf{y}_i$  is the canonical unit vector  $\mathbf{e}_i$  the decision rule results in:

$$\begin{aligned} &\text{Assign } \mathbf{x} \text{ to class } i, \text{ if} \\ &\|\mathbf{A}^\top \mathbf{x} + \mathbf{b} - \mathbf{y}_i\| = \min_{j=1,\dots,M} \|\mathbf{A}^\top \mathbf{x} + \mathbf{b} - \mathbf{y}_j\|. \end{aligned} \quad (12)$$

### 3.2. TREATMENT OF NUMERICAL PROBLEMS

Both methods described above require an inversion of a scatter matrix ( $\mathbf{W}$  or  $\mathbf{T}$ ) which must not be singular. This holds if the number of linearly independent samples  $N$  is at least equal to or greater than the dimension  $n$  of the feature vectors. Practical applications require  $N \gg n$ . In the case of high dimensional feature vectors it is sometimes difficult or impossible to obtain a sufficient number of samples. If this cannot be achieved,  $\mathbf{W}$  or  $\mathbf{T}$  possibly become singular or ill-conditioned. A lot of proposed methods do not treat that problem<sup>(23, 24, 25, 26)</sup>. Hence, they fail when applied to our database introduced in section 4.1. Because a numerical stable method is essential for our classification task we study the problem of  $\mathbf{W}$  being singular or ill-conditioned in greater detail. The results derived for  $\mathbf{W}$  hold for  $\mathbf{T}$ , as well.

The same problem occurs in the context of orthogonal linear discriminant analysis, which was treated by others:

- a) Hong and Yang<sup>(27)</sup> circumvented the problem by adding a small singular value perturbation to  $\mathbf{W}$  resulting in  $\tilde{\mathbf{W}}$  such that  $\tilde{\mathbf{W}}$  becomes non-singular. The shortcomings with the perturbation method are: First, the perturbation of  $\mathbf{W}$



discriminant space which is two-dimensional in this case. Fig. 4 depicts the distribution of the projected feature vectors.

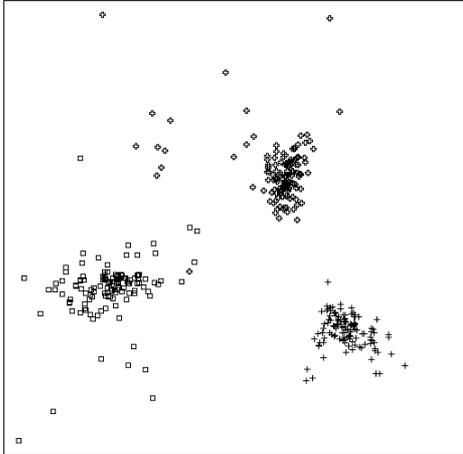


Figure 4: Projection of all feature vectors of three classes (out of 25 natural objects, see section 4.1) to the discriminant space. The features were extracted over 3 gaussian scales using a  $5 \times 5$  displacement for the autocorrelation function. The classifier was trained with half of the samples.

#### 4.1.2. RECOGNITION RATES

Tab. 1 shows the recognition rates for different features. The computation times for each processing step are shown in Tab. 2.

computation times	
<i>preprocessing</i>	
pyramid with 3 scales, 1 octave	< 0.1 sec
pyramid with 6 scales, $\frac{1}{2}$ octave	0.4 sec
2nd order acf $3 \times 3$ , 29 feat.	0.9 sec
2nd order acf $5 \times 5$ , 205 feat.	8.3 sec
2nd order acf $3 \times 3 \times 3$ , 264 feat.	9.1 sec
1st order acf 221/254 feat.	0.3 sec
<i>training</i>	
3200 samples (25 classes, 87 feat.)	4.7 sec
3200 samples (25 classes, 205 feat.)	41.7 sec
3200 samples (25 classes, 264 feat.)	83.2 sec
<i>classification</i>	
3200 samples (25 classes, 87 feat.)	1.2 sec
3200 samples (25 classes, 205 feat.)	3.6 sec
3200 samples (25 classes, 264 feat.)	4.1 sec

Table 2: Computation times for each processing step ( $256 \times 256$  images) on a SUN sparcl0-512 workstation.

We also tested Laplacian pyramids and Gabor filter pyramids. The results are not shown because the recognition rates were poor. Since Laplacian pyramids and Gabor filter pyramids split up the image in different frequency bands and different orientations,

respectively, the response of a single filter is very sparse. Autocorrelations on local regions cannot extract sufficient information here, so they are not the appropriate features for sparse data.

The recognition rates for 3D-feature vectors are significantly lower than the rates for 2D-feature vectors while using nearly the same number of features. The reason is the low sampling rate in scale dimension leading to severe discretization errors.

It is conspicuous that the concatenated feature vectors as used by Otsu *et al.* lead to poorer recognition rates. They are not independent of the position of the object representation in the scale space so they yield no scale invariancy.

The proposed feature extraction allows the use of a simple linear classifier due to the linear separability in feature space. Thus, the high recognition rates are mainly achieved by an appropriate image representation. In addition to the employed linear classifiers we tested some other simple classification schemes like nearest neighbor rule and minimal Euclidian distance to the class mean. The results were poor ( $< 10\%$ ) due to high linear correlations between the components of the feature vectors which are not handled by these methods.

#### 4.2. 86 HUMAN FACES

This database, taken from Lades *et al.*<sup>(35)</sup>, contains 86 human faces with 6 images each (standard portrait,  $15^\circ$  and  $30^\circ$  rotation, 75% and 50% size, grimace).

Compared to the dimensionality of the feature vector, the number of samples is extremely small.

Although this database was designed to test systems which require very few samples per class our system remains stable on this data. Tab. 3 shows the recognition rates.

pyramid	features	% rate
gauss, 3 sc, 1 oct	2nd ord. acf $3 \times 3^*$	89.7
gauss, 6 sc, $\frac{1}{2}$ oct	2nd ord. acf $5 \times 5$	92.1
gauss, 6 sc, $\frac{1}{2}$ oct	2nd ord. acf $3 \times 3 \times 3$	72.7

Table 3: Recognition rates for faces obtained by the Bayes classifier. The overall recognition rate is estimated by the *leaving-one-out* test. Comparability with the recognition rates in Lades *et al.*<sup>(35)</sup> is limited since this system stores only one sample per class and is adjusted to reject all positive false recognitions. The rates in the reference are between 79% and 88%. \*As used by Otsu *et al.*<sup>(3)</sup>.

#### 4.3. BENCHMARKS

Finally, we apply our linear classifier to some public benchmarks. All benchmarks except the last two are available via FTP on server ics.uci.edu.

- Breast Cancer Database, University of Wisconsin Hospitals<sup>(36, 37)</sup>,

pyramid	autocorr.	# feat.	% rate a)		% rate b)	
			Bayes	LRA	Bayes	LRA
1 scale	2nd order acf 3×3	29	52.9	67.6	72.9	70.7
3 gauss scales, 1 octave	2nd order acf 3×3	29	50.0	65.9	71.4	72.7
3 gauss scales, 1 octave	2nd order acf 3×3*	87	61.1	75.0	69.3	71.8
3 gauss scales, 1 octave	2nd order acf 5×5	205	89.9	95.9	87.0	87.2
6 gauss scales, 1/2 octave	2nd order acf 3×3	29	46.9	64.1	70.5	71.8
6 gauss scales, 1/2 octave	2nd order acf 3×3**	174	56.4	70.1	66.6	67.4
6 gauss scales, 1/2 octave	2nd order acf 5×5	205	89.8	97.3	86.2	86.6
6 gauss scales, 1/2 octave	2nd order acf 3×3×3	264	54.8	65.8	67.2	66.5

Table 1: Recognition rates for different features. The test sets contained a) the two longest object distances, b) the two longest and two shortest object distances. \*As used by Otsu *et al.*<sup>(3)</sup>. \*\*Concatenated feature vectors similar to Otsu *et al.*

- b) Ionosphere database, Johns Hopkins University<sup>(38)</sup>,
- c) Pima Indians Diabetes Database<sup>(39)</sup>,
- d) NASA Satellite Images,
- e) Image Segmentation Data, University of Massachusetts,
- f) Vehicle Silhouettes<sup>(40)</sup>,
- g) Human Faces<sup>(29)</sup>,
- h) Our own database.

In Tab. 4 the results of the different classifiers are compared. Among the other linear classifiers listed in the table our classifier gives the highest recognition rates. One reason is the high numerical stability of our method which yields accurate results especially in extreme cases. Another reason is that some proposed methods are not appropriate for classification. The method of orthogonal discriminant vectors proposed by Sammon<sup>(41)</sup> and extended by Foley and Sammon<sup>(24)</sup> is inadequate with regard to classification<sup>(42)</sup> as used by Cheng *et al.*<sup>(28, 29)</sup> and Okada *et al.*<sup>(25)</sup>. Additional orthogonal vectors do not convey any new information and the information is spread over all orthogonal axes. Moreover, this method is very time consuming since it requires an eigenvalue decomposition for each axis.

used data set			% recognition rate		
set	# class	# total/train	we	<sup>(29)</sup>	ref.
a)	2	369/184	95.1	94.6	93.5
b)	2	351/200	90.7	90.0	90.7
c)	2	768/576	79.7	77.1	76.0
d)	7	6435/4435	84.0	77.1	–
e)	7	2310/1155	90.3	49.6	–
f)	4	846/423	78.0	71.1	–
g)	3	24/6	100.0	94.4	–
h)	25	3200/1600	99.6	51.1	–

Table 4: Recognition rates of different classifiers. The reference rates are obtained by a) 2 pairs of parallel hyperplanes<sup>(36)</sup>, b) linear perceptron (92.0% with a nonlinear perceptron)<sup>(38)</sup>, c) ADAP<sup>(39)</sup>.

## 5. CONCLUSION

We presented an image recognition system, based on higher order autocorrelation features, which is translation- as well as scale-invariant over a wide range. The proposed feature extraction allows the use of a linear classifier due to the linear separability in the feature space. The classification method used is not only optimal in terms of linear classification, but also turned out to be numerically stable under extreme conditions. We extensively tested our classifier as well as our complete application. With a database created with regard to the demands the system aims at we demonstrated its excellent recognition invariance. This also holds for tests with a human faces database. To demonstrate its generality we applied our classification method to many public benchmarks. The superiority of our method compared to other linear methods is revealed.

## REFERENCES

1. D. Marr, *Vision*. W. H. Freeman and Company, New York (1982).
2. S. Ullman, Aligning pictorial descriptions: An approach to object recognition, *Cognition* **32**, 193–254 (1989).
3. T. Kurita, N. Otsu, and T. Sato, A face recognition method using higher order local autocorrelation and multivariate analysis, *11th IAPR International Conference on Pattern Recognition*, Vol. 2, pp. 213–216, The Hague (1992).
4. M. Seibert and A. M. Waxman, Adaptive 3-D object recognition from multiple views, *IEEE Trans. Pattern Anal. Machine Intell.* **14**, 107–124 (1992).
5. M. Schwarzinger, D. Noll, and W. von Seelen, Object recognition with deformable models using constrained elastic nets, *Informatik aktuell, Mustererkennung 1992*, pp. 96–104 (1992).

6. S. Ullman and R. Basri, Recognition by linear combinations of models, *IEEE Trans. Pattern Anal. Machine Intell.* **13**, 992–1005 (1991).
7. R. P. Würtz *et al.*, Recognition of human faces by a neuronal graph matching process, H. Schuster, editor, *Applications of Neural Networks*, pp. 181–200 (1992).
8. S. Bohrer *et al.*. An active system for task-specific information processing. H. G. Schuster, editor, *Applications of Neural Networks*. VCH-Verlagsgesellschaft, (1992).
9. N. Otsu and T. Kurita, A new scheme for practical flexible and intelligent vision systems, *Proc. IAPR Workshop on Computer Vision*, pp. 431–435, Tokyo (1988).
10. T. Lindeberg, Scale-space for discrete signals, *IEEE Trans. Pattern Anal. Machine Intell.* **12**, 234–254 (1990).
11. E. Rosenfeld, *Multiresolution Image Processing and Analysis*. Springer, Berlin (1984).
12. D. Terzopoulos, The computation of visible-surface representations, *IEEE Trans. Pattern Anal. Machine Intell.* **10**, 417–438 (1988).
13. T. D. Sanger, Stereo disparity computation using gabor filters, *Biol. Cybern.* **59**, 405–418 (1988).
14. J. A. McLaughlin and J. Raviv, Nth-order autocorrelations in pattern recognition, *Information and Control* **12**, 121–142 (1968).
15. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. Wiley, New York (1973).
16. K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd Printing, New York (1974).
17. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Academic Press, 7th Printing (1989).
18. M. M. Tatsuoka, *Multivariate Analysis: Techniques for educational and psychological research*. Wiley, New York (1971).
19. P. A. Lachenbruch, *Discriminant Analysis*. Hafner Press, New York (1975).
20. G. Peters and J. H. Wilkinson,  $Ax = \lambda Bx$  and the generalized eigenproblem, *SIAM J. Numer. Anal.* **7**, 479–492 (1970).
21. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford (1988).
22. M. Kendall, *Multivariate Analysis*. Charles Griffin & Company Ltd, London (1980).
23. K. Fukunaga and W. L. G. Koontz, A criterion and an algorithm for grouping data, *IEEE Trans. Comput.* **C-19**, 917–923 (1970).
24. D. H. Foley and J. W. Sammon Jr., An optimal set of discriminant vectors, *IEEE Trans. Comput.* **C-24**, 281–289 (1975).
25. T. Okada and S. Tomita, An optimal orthonormal system for discriminant analysis, *Pattern Recog.* **18**, 139–144 (1985).
26. Y. Hamamoto *et al.*, A note on the orthonormal discriminant vector method for feature extraction, *Pattern Recog.* **24**, 681–684 (1991).
27. Z.-Q. Hong and J.-Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, *Pattern Recog.* **24**, 317–324 (1991).
28. Y.-Q. Cheng, Y.-M. Zhuang, and J.-Y. Yang, Optimal Fisher discriminant analysis using the rank decomposition, *Pattern Recog.* **25**, 101–111 (1992).
29. K. Liu, Y.-Q. Cheng, and J.-Y. Yang, A generalized optimal set of discriminant vectors, *Pattern Recog.* **25**, 731–739 (1992).
30. Q. Tian, Comparison of statistical pattern-recognition algorithms for hybrid processing, II: eigenvector-based algorithms, *J. Opt. Soc. Am. A* **5**, 1670–1672 (1988).
31. Y. T. Chien, Selection and ordering of feature observations in a pattern recognition system, *Information and Control* **12**, 395–414 (1968).
32. Y. T. Chien and K. S. Fu, On the generalized Karhunen-Loève expansion, *IEEE Trans. Inform. Theory* **IT-13**, 518–520 (1967).
33. I. N. Bronstein and K. A. Semendjajew, *Taschenbuch der Mathematik*. Verlag Harri Deutsch (1989).
34. M. H. DeGroot, *Probability and Statistics*. Addison-Wesley, Reading, Massachusetts (1975).
35. M. Lades *et al.*, Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* **C-42**, 300–311 (1993).
36. W. H. Wolberg and O. L. Mangasarian, Multi-surface method of pattern separation for medical diagnosis applied to breast cytology, *Proceedings of the National Academy of Sciences, U.S.A.*, Vol. 87, pp. 9193–9196 (1990).
37. O. L. Mangasarian and W. H. Wolberg, Cancer diagnosis via linear programming, *SIAM News* **23**, 1–18 (1990).

38. V. G. Sigillito *et al.*, Classification of radar returns from the ionosphere using neural networks, *Johns Hopkins APL Technical Digest* **10**, 262–266 (1989).
39. J. W. Smith *et al.*, Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, *Proc. Symposium on Computer Applications and Medical Care*, pp. 261–265 (1988).
40. J. P. Siebert. Vehicle recognition using rule based methods. Technical report, Turing Institute Research Memorandum TIRM-87-018, (1987).
41. J. W. Sammon Jr., An optimal discrimination plane, *IEEE Trans. Comput.* **C-19**, 826–829 (1970).
42. J. Kittler, On the discriminant vector method of feature selection, *IEEE Trans. Comput.* **C-26**, 604–606 (1977).

**About the Author**—Martin Kreutz was born in Hannover, Germany, on 28 January 1968. He received the Diploma in Computer Science from the University Dortmund, Germany, in 1994. He is currently a Research Assistant and pursuing the Doctoral degree at the Institut für Neuroinformatik, Ruhr-Universität Bochum. His current research interests are in the areas of computer vision, image recognition, and evolutionary algorithms.

**About the Author**—Bernd Völpel was born in Worms, Germany, on 5 July 1965. He received the Diploma in Electrical Engineering from the RWTH Aachen, Germany, in 1991. He is currently a Research Assistant and pursuing the Doctoral degree at the Institut für Neuroinformatik, Ruhr-Universität Bochum. His current research interests are in the areas of 3D-depth reconstruction and object recognition.

**About the Author**—Herbert Janßen was born in Duisburg, Germany, on 19 April 1961. He received the Diploma in Physics from the Westfälische-Wilhelms-Universität Münster, Germany, in 1989. In 1992 he was a Visiting Researcher in the Department of Mathematical Engineering, University of Tokyo. Since 1989 he is a Research Assistant at the Institut für Neuroinformatik, Ruhr-Universität Bochum and is currently pursuing the Doctoral degree. His current research interests are in the areas of active vision, saccades, and behavioral control.