A NEURAL NETWORK FOR CALCULATING ADAPTIVE SHIFT AND ROTATION INVARIANT IMAGE FEATURES

Sabine Kröner

Technische Informatik I

Technische Universität Hamburg-Harburg

21071 Hamburg, Germany

Tel/Fax: +49 [40] 7718 2539 / 7718 2911

e-mail: kroener@tu-harburg.d400.de

ABSTRACT

Shift and rotation invariant pattern recognition is usually performed by first extracting invariant features from the images and second classifying them. This poses the problem of not only finding suitable features but also a suitable classifier.

Here a structured invariant neural network architecture (SINN) is presented that performs adaptive invariant feature extraction and classification simultaneously. The network is sparsely connected and uses shared weight vectors. As a result features especially well suited for a given application are calculated with a computational complexity of $\mathcal{O}(N)$ for $N = 2^n$ input elements. Experiments show the recognition ability of the invariant neural network on synthetic and real data.

1 Introduction

Most image processing systems for shift and rotation invariant pattern recognition achieve their invariance property by first calculating invariant features and second classifying them with either standard classifiers or neural networks. For the feature extraction usually general methods like moment invariants or geometric invariants are used.

However, the computation of the invariants with these methods turns out to be very costly since the calculated features are not adapted to the patterns in the present application. Often the separation quality of each feature for the underlying pattern set is not known in advance; this may lead to the calculation of numerous unsuitable features.

The problem can be overcome by calculating adaptive features which are especially appropriate for a given application. Here a neural network architecture is presented that performs this task. From the patterns adaptive shift and rotation invariant features are extracted and classified simultaneously. The invariance property is built into the network architecture by use of shared weight vectors and a sparse connection structure.

The paper is organized as follows: First the architecture of the structured invariant neural network (SINN)

for the calculation of shift and rotation invariant image features is presented. Then the realization of the node functions is explained together with a learning algorithm for the weights. Experiments show the recognition ability of the invariant neural network compared to a standard feature extraction method using geometric invariants on synthetic and real data. Finally the results are summarized in a conclusion.

2 The Network Architecture

The architecture of the shift and rotation invariant neural network is a feedforward architecture resembling a binary tree. Its size is determined by the number of input elements: it has $\mathrm{ld}\,N$ layers for an input vector of $N = 2^n$ elements. The realization for shift invariance only corresponds to one branch in the connection structure of the signal flow graph of the shift invariant transforms of the class $\mathcal{C}\,T$ [1]. Fig. 1 shows the one-dimensional shift invariant network architecture. The
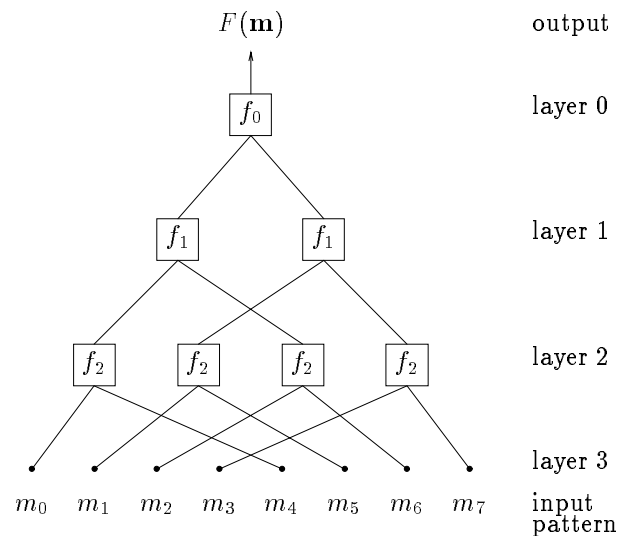

Figure 1: Architecture of the structured shift invariant neural network for 8 input elements

sparsely connected network architecture with only indegree two of all nodes is necessary for the invariance
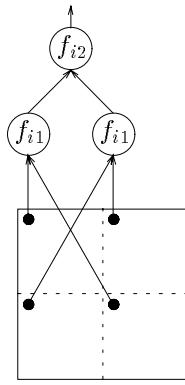
Figure 2: Connection of four input elements on layer i in a structured shift and 90 degree rotation invariant network

property of the network output. Another prerequisite for the invariance is the symmetry of the node functions f_i. Moreover all nodes of the same layer are coupled. This means that the nodes of one layer share all their weights and also their transfer functions. Therefore the action of layer i in this architecture can completely be described by the node function f_i. Due to this definition of the architecture the network is called structured invariant neural network (SINN).

The network architecture for one-dimensional patterns can be extended to an architecture for the recognition of two-dimensional patterns. With the connection structure shown in Fig. 2 additional invariance with respect to rotations of multiples of 90 degrees is integrated into the same architecture without any extra costs. For invariance with respect to shifts and general rotations the network has to be implemented in several different angular orientations so that the networks cover regularly the full circle of 360 degrees. Caused by the good generalization ability of the network it is sufficient for most applications to implement the two-dimensional shift- and 90-degree rotation invariant network in the four angular orientations of 0, 22.5, 45, and 67.5 degrees. After the input has been processed in parallel by the four networks the results are combined to the final output again using the connection structure of Fig. 2.

The invariance property of the architecture is achieved stepwise. It can be shown that it increases layer by layer up to final shift and rotation invariance of the image [3]. The proof is based on results from the theory of groups.

3   The Node Functions

The main requirement for the node functions in the SINN is symmetry with respect to the two inputs. It can be solved by use of twin nodes performing commutative calculations with respect to the weights and the input arguments. A possible realization of the node function

f_i in the neural nodes of layer i is shown in Fig. 3 with the equation

$$f_i = u_i \left(w_{i3} \left(t_i(w_{i1}m_1 + w_{i2}m_2 + w_{i0}) \right) + w_{i3} \left(t_i(w_{i2}m_1 + w_{i1}m_2 + w_{i0}) \right) \right)$$

where w_{ij} are the weights on the connectional links, w_{i0} is the threshold, and t_i and u_i are hard limiter or sigmoid transfer functions.
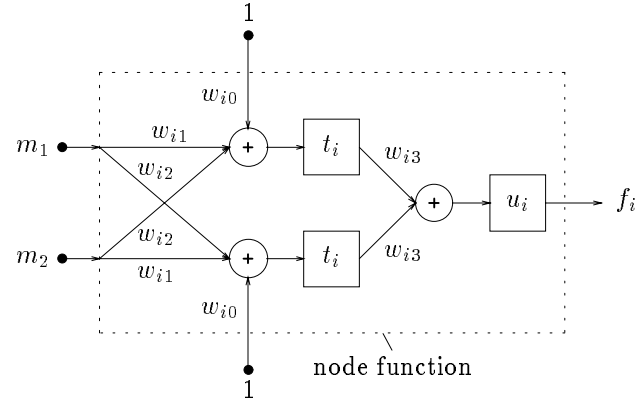

Figure 3: Node function f_i

When using hard limiter transfer functions only two pattern classes can be distinguished by the network output. So for separating k classes at least k networks are needed, one for the recognition of each class. The result is a binary output vector with a one in that entry that corresponds to the recognized pattern and zero in all others.

4   The Learning Algorithm

Assume two pattern classes A and B. The weights have to be chosen so that the percentage of misclassified patterns of both classes A and B is minimized in layer 0. So the error function E can be formulated as

$$E = \min_{\forall \text{weights}} \left(\frac{\# \text{ incorrect patt. class A}}{\# \text{ patterns class A}} + \frac{\# \text{ incorrect patt. class B}}{\# \text{ patterns class B}} \right) . \quad (1)$$

For the training of the weights a global learning algorithm like a backpropagation algorithm modified for the training of structured architectures [4] can be used if differentiable transfer functions are given.

Another possibility are local optimization techniques. Such methods are applicable since the strong weight coupling extremely reduces the dimension of the input space. In fact the dimension of the input space is determined by the indegree of the nodes in each layer. So instead of representing a N-dimensional pattern as a point in a N-dimensional input space it is equivalent with the architecture of a SINN to represent it as $N/2$ points in

a 2D-input space. With the symmetric node functions all cyclic permutations of the input elements can be represented in parallel in the same input space. Different patterns are marked as different sets of points in this space. The weights associated with the node function of a layer then determine the form of the decision line between the points of two pattern classes. This means that only the few parameters of one node function have to be adjusted on every layer.

The network training consists in successively determining the weights so that as many points of patterns of class A as possible are separated from those of class B by the decision line of the current layer. Binary transfer functions limit the number of weight combinations of the w_i, which result in different outputs, to six on all layers i, $0 \leq i \leq n - 2$. It can be shown that the weight combinations giving the minimum of the local error functions also lead to a minimum of the global error function on layer 0. On the first layer where the input elements are of the interval $[0, 1]$ the evaluation of more than six weight combinations is suitable. Here also quadratic decision lines can be applied.

Once the weights are trained the computational complexity to recognize unknown patterns is of order $\mathcal{O}(N)$.

5   Experiments

First the robustness of the adaptive features calculated with the neural network with respect to statistical pattern distortions is investigated on synthetic grey scale images. Four grey value images of size 64×64 pixels are used (Fig. 4), one with constant grey value, the others showing rectangles of different size. They are all equal with respect to the grey value sum. A total of 280 noisy images are generated by adding gaussian noise with mean 0 and a variance leading to a fixed signal-to-noise-ratio. The recognition is performed with four shift and 90°-rotation invariant SINNs with $n = 12 = \mathrm{ld}\, 64^2$ layers. The weights are determined by training the original patterns only.

The results are compared with the standard feature extraction method of invariant integration [2] followed by classification. Here twelve different shift and rotation invariant grey scale features based on monomials of up to order three are used. They are evaluated with the Bayes classifier under the assumption that the features are classwise normally distributed, the weighted nearest neighbour classifier (WNC) (i.e. the euclidean distance weighted with the variance of the features), and the nearest neighbour classifier (NC) (i.e. the simple euclidean distance), respectively. The training set for the three classifiers consisted of 20 images per class (including noisy images with different SNR). It has to be emphasized that no preprocessing (e.g. smoothing, segmentation) is performed on the images neither for the SINN nor for the method of invariant integration.

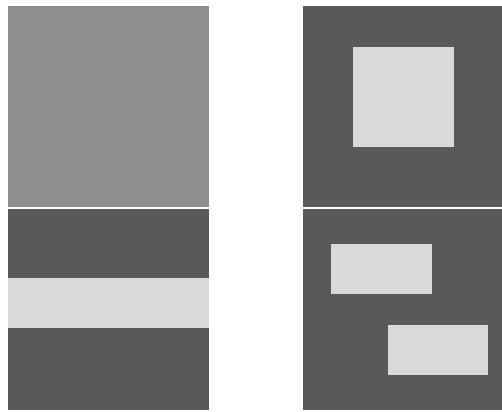The upper part of Table 1 shows the recognition rates

Figure 4: Synthetic images Q1, Q2, Q3, and Q4 of size 64^2 with equal sum of grey values.

for the different signal-to-noise-ratios (SNR) achieved by using the SINN whereas the lower part shows the recognition rates for the three different classifiers based on the invariant grey scale features. Table 1 shows that due to the adaptivity of the SINNs the classification of the four patterns from Fig. 4 is very robust with respect to statistical pattern distortions. Using the grey scale features together with the nearest neighbour (NC) or weighted nearest classifier (WNC) does not allow a reliable separation of the classes. However, the Bayes classifier together with the grey scale features also permits the robust separation of the classes.

In the second experiment the invariance property of the SINN is shown with respect to arbitrary shifts and rotations. Moreover the separation ability is demonstrated in comparison to the standard feature extraction method. For this experiment subimages of size 256×256 from scanned grey scale images belonging to 4 different classes are used. Each class consists of 10 arbitrarily shifted and rotated versions of the same object. Fig. 5 shows one example image of each class.

The background is modified with gaussian noise so that the mean value of the images is approximately the same (otherwise the separation problem could be solved just by comparing the mean grey values). Again only one image of each class is used for training the four different structured shift and rotation invariant neural networks. In the first layer a quadratic decision line is applied and in all other layers a linear one. Internally a representation with four different rotation angles is used as described in section 2. However these representations do not coincide with all the occuring rotation angles and shifted positions it can be seen from the recognition rates of the patterns of the test set in Table 2 that the network has a good generalization ability with respect to general rotations and shifts.

These results are compared with the same standard feature extraction method of invariant integration as in

SNR		16	18	20	22	24	27	30
SINN	Q1	100	100	100	100	100	100	100
	Q2	100	100	100	100	100	100	100
	Q3	100	100	100	100	100	100	100
	Q4	100	100	100	100	100	100	100
NC	Q1	100	75	75	75	100	100	100
	Q2	0	0	0	25	100	100	100
	Q3	0	0	0	0	0	0	0
	Q4	100	100	100	60	40	0	0
WNC	Q1	100	100	100	100	100	100	100
	Q2	0	0	0	25	100	100	100
	Q3	0	40	0	20	0	0	0
	Q4	100	100	100	60	0	0	0
Bayes	Q1	100	100	100	100	100	100	100
	Q2	100	100	100	100	100	100	100
	Q3	100	100	100	100	100	100	100
	Q4	100	100	100	100	100	100	100

Table 1: Recognition rates in percent for the patterns shown in Fig. 4 for a structured shift and 90°-rotation invariant neural network and for invariant grey scale features and three different classifiers.

	apple	pear	mushr.	tomato
SINN	100	97.2	100	91.6
NC on 5 features	85.7	28.5	57.1	57.1
NC on 12 features	100	100	100	100

Table 2: Recognition rates in percent for the patterns shown in Figure 5 for a structured shift and rotation invariant neural network and for 5 and 12 invariant grey scale features.

If a SINN is used instead good separation results are achieved with much lower costs. Moreover there is no need to find a suitable classifier.

the first experiment. Two sets of features consisting of 5 features based on monomials of up to order two and of 12 features based on monomials of up to order three are used. Since the number of images is too small to train a Bayes classifier recognition rates are only presented for the nearest neighbour (NN) classifier, s. Table 2. The training set consisted of three images per class.

The recognition rates show that a complete separation of the images can only be achieved by use of monomials of order three with the method of invariant integration. The computational costs for the calculation of these features is rather high especially because the selection of features needed for separation is not known in advance.
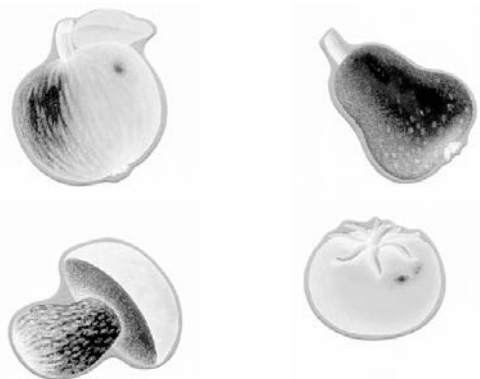
6 Conclusion

A neural network has been presented for the calculation and simultaneous classification of cyclic shift and rotation invariant image features. The invariance property is achieved by a sparsely connected structured invariant neural network architecture with layerwise coupled node functions. During the learning process the whole network is adapted to the characteristics of the pattern set in a given application. This reduces the total number of invariant features necessary for the separation of the patterns and leads to better robustness with respect to global disturbances. Moreover by the simultaneous adaptive feature extraction and classification the usual problem of suitably choosing the feature extraction method and the design of the classifier to obtain good recognition results is avoided. Present investigations focus on the development of enhanced node functions, and the implementation of higher order characteristics into the network architecture.

Figure 5: Scanned grey value images of apple, pear, mushroom, and tomato of size 256^2.

References

[1] H. Burkhardt, *Transformationen zur lageinvarianten Merkmalgewinnung*, VDI-Fortschrittbericht, Reihe 10, Nr. 7, Oct. 1979

[2] H. Schulz-Mirbach, *Invariant features for gray scale images*, 17. DAGM Symposium Mustererkennung 1995, Sagerer, G; Posch, S.; Kummert, F. (Ed.), S. 1-14, Springer Verlag, Bielefeld, Sept. 1995

[3] S. Kröner, H. Schulz-Mirbach, *Adaptive features for position invariant pattern recognition*, 17. DAGM Symposium Mustererkennung 1995, Sagerer, G.; Posch, S.; Kummert, F. (Ed.), S. 23-35, Springer Verlag, Bielefeld, Sept. 1995

[4] D.E. Rumelhart et al., *Learning internal representations by error propagation*, in Parallel Distributed Processing, vol. 1, ch. 8, Rumelhart, D.E.; McClelland, J.L. (Eds.), Cambridge, MA, M.I.T. Press, 1986