# EQUIVALENT TLU- AND ΣΠ-NETWORKS FOR INVARIANT PATTERN RECOGNITION

Helmut Glünder
Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität, Gœthestraße 31, D-8000 München 2, FRG

## Abstract

Two universal types of networks for the invariant recognition of pictorial patterns are compared with respect to function, structure and costs. The main stage of both networks serves for the extraction of features that are invariant under certain types of *unrestricted* geometric transformations, e.g. rigid translations. Both approaches are conceptualized for *unequivocal* class definitions and thus for the feasibility of *perfect* pattern reconstructions. Although the networks are structurally different, they are to a high degree functionally equivalent. The costs, i.e., the number of weights per class that must be adjusted in order to obtain ideal and invariant classification, turn out to be almost the same for both approaches as well as for the reference network (list classifier). In practice, however, the ΣΠ-network is superior to the TLU-network; it is more robust and even single invariant features are unequivocally defined. The investigations reported here do not concern any aspects of learning.

## Introduction

Most network approaches to geometrically invariant recognition of pictorial patterns are based on invariant classification, i.e., explicit representations of invariant features do not occur. The most powerful approach of this kind that can be adapted to meet arbitrary demands for invariance and pattern reconstruction, is the so-called list classifier. This straightforward method uses many holistic template-matching (sub)classifiers for each invariant class in parallel – one for every transformed version of the class prototype. A class-specific decision $z_{List,j}$ is obtained by the summation of the corresponding subclassifier outputs. Each of the holistic subclassifiers is a threshold logic unit (TLU) that computes the inner product of the normalized signal vector $\mathbf{x}$ and the normalized weighting vector $\mathbf{w}_{jv}$ and finally thresholds the resulting cross correlation coefficient $y_{jv}$ according to

$$u_{jv} = \begin{cases} 1 & \text{if } y_{jv} = 1 \\ 0 & \text{else} \end{cases} \qquad \text{with} \quad y_{jv} = (\mathbf{x}^T/|\mathbf{x}|)(\mathbf{w}_{jv}/|\mathbf{w}_{jv}|)\,. \tag{1}$$

Obviously, invariant list classification allows for unequivocal pattern characterization and thus reconstruction which, of course, does not include the description parameters under which the classification is invariant. The advantages of this approach are opposed by the total lack of generalization which can only be attained at unreasonably high costs.

In this paper, however, two network approaches to geometrically invariant pattern recognition are presented and compared which essentially deal with the generation of representations of invariant features that may subsequently be linearly classified. This decision making second stage consists of TLUs – one for every class j –, whose thresholds can be replaced by a single maximum detector if generalization is demanded. Unlike most of the existing investigations that deal with 'learning' and recognition rates, the main goal of this contribution is to determine the necessary and sufficient conditions (structure and costs) under which the following demands for a classifier can be met:

1. invariance under certain types of *unrestricted* geometric transformations
2. *unequivocal* class definitions and thus *perfect* reconstructability also of nonbinary patterns
3. no *a priori* knowledge about the kind and number of the pattern classes

Owing to these demands, more economic but less universal standard approaches, such as linear reduction of signal dimensions, as well as subclass techniques, must be discarded.

The investigations reported here are exemplified for unrestricted 2D pattern translations on a toroidal image array of n pixels. Consequently, the maximum number of translation invariant classes that exist for signals of m levels, is $k_{max} \gtrsim m^n/n$. – Comparable considerations hold for unrestricted rotations and scale changes which can be imagined as converted into 2D translations by log-polar mapping (Brousil and Smith, 1967; Casasent and Psaltis, 1976).

(1990) In: Kohonen T. (ed.) Proc. INNC-90, Paris. Kluwer, Dordrecht.

## Direct Invariant Classification

A list classifier that is invariant under unrestricted translations of signals on a toroidal image array of n pixels typically requires n TLUs per class – one for every shift position. Because the TLUs comprise holistic templates, i.e., nD vectors $\mathbf{w}_{ji}$, the costs of ideal invariant class definitions, expressed by the number of weighting coefficients per class, are $N \approx n^2$ (see Fig.1a, roman inscriptions).

Direct invariant classification by a linear classifier, i.e., by cross correlations and subsequent maximum detection, is generally impossible. Thus, nonlinearities, at least of the polynomial degree p=2, must be applied to the signal components before linear classification becomes possible. Owing to the involved thresholds $\theta$, list classification relies on nonlinearities of degree $p \rightarrow \infty$. Hence p=2 is a necessary and $p \rightarrow \infty$ a sufficient condition for ideal and invariant class definitions in such networks.

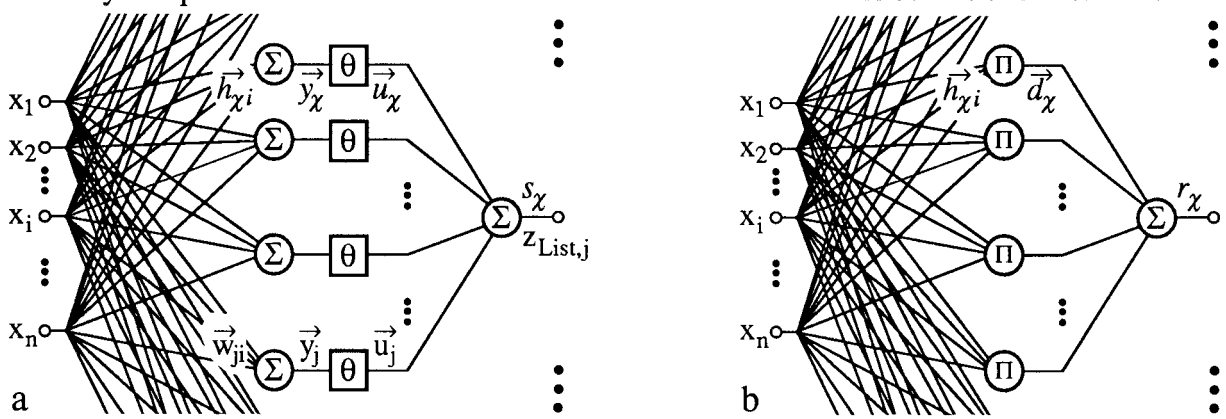


**Figure 1a, b.** Network structures for list classification (roman inscriptions) and for polynomial pattern decomposition (PPD) (italic inscriptions) (**a**), as well as for multilinear pattern decomposition (MPD) (**b**)

## Polynomial Pattern Decomposition

Instead by holistic comparisons, one can try to describe patterns in terms of their composition from pattern elements. For this purpose the weighting vectors consist of only q<<n nonzero components and shall be called mask vectors ${}^q\mathbf{h}_\chi$. As with list classification, each type of mask (denoted by $\chi$, with $1 \leq \chi \leq \kappa$) must be applied at every position if shift invariance is demanded, and the desired invariant features must be determined from the outputs of each of the thereby defined $\kappa$ sets that consist of typically n TLUs. Commonly the number of suprathreshold signals, i.e., their sum $s_\chi$, is used for this purpose which leads to a network structur that is known from the list classifier (see Fig.1a, italic inscriptions).

Unfortunately separate normalization of signal and mask vectors according to Eq.(1) cannot be applied for the unequivocal detection of pattern elements. Because normalization is only feasible at the enormous costs of $\kappa n$ individual operations which are difficult to realize within network structures, normalization is sacrificed. Consequently, the threshold of the TLUs must be replaced by a nonlinear and strictly monotonic transfer characteristic of non-negative derivative which leads to so-called generalized TLUs. If the nonlinearity is expressed by a polynomial of degree p, then every invariant feature is a sum of typically n polynomials that are computed from a linear discriminant function $y_{\chi i}$ (cf. Eq. (1)). Therefore, this kind of extraction of invariant features shall be called a polynomial approach to pattern decomposition (PPD).

$$ {}_p s_\chi = \sum_{i=1}^{n} {}_p u_{\chi i} = \sum_{i=1}^{n} \sum_{\rho=0}^{p} \alpha_\rho \cdot (y_{\chi i})^\rho \qquad (2) $$

In conjunction with a final linear classifier stage the PPD network represents a *perceptron* network with one hidden layer (Lippman, 1987; Rosenblatt, 1962).

Obviously, non-binary masks do not make much sense without normalization. Therefore, and in or-

der to prevent the combinatorial explosion of mask types, all further investigations are confined to 0/1-valued masks – which turns out to be an appropriate strategy. Irrespective of this restriction, the fundamental question, how many masks of what type are necessary and sufficient for perfect translation invariant pattern descriptions, is not yet answered. Empirical investigations reveal that they are not achieved if masks with q<3 are applied, or if the nonlinearity is of polynomial degree p<3. Therefore, a necessary condition and, with the fact that a subsequent template-matching classifier accesses κD features, the costs for translation invariant classification with ideal class definitions are

$$(p = 3) \wedge (q = 3) \Rightarrow (N = {}^3\kappa \approx n^2/6) \qquad \text{for} \quad n >> 3 \ . \tag{3}$$

Although more specific than the statements that were given at the end of the previous section, it does not yet clarify whether p=3 and q=3 are sufficient for unequivocal pattern reconstructions.

## Multilinear Pattern Decomposition

From the previous conditions it must be concluded that only a single term of every polynomial ${}^3{}_3u_{\chi i}$ (cf. Eq.(2)) is actually responsible for perfect pattern descriptions, namely the trilinear term ${}^3d_{\chi i}$ that comprises those q=3 signal values which are selected by mask ${}^3\mathbf{h}_{\chi i}$. Hence, the hereby defined invariant features ${}^3r_\chi$ are sums of n well-chosen trilinear terms, i.e., special trilinear forms. The corresponding q-linear form is defined by

$${}^qr_\chi = \sum_{i=1}^{n} {}^qd_{\chi i} = \sum_{i=1}^{n} \prod_{\sigma=1}^{q} h_{\chi i\sigma} \cdot x_{\chi i\sigma} \ , \tag{4}$$

with the selected signal values $x_{\chi i\sigma}$ that may be weighted by the corresponding mask components $h_{\chi i\sigma}$. For 0/1-masks the features are autocorrelation coefficients of order q of a signal **x**. Obviously, pattern elements can be defined through the multiplication schemes that are given by 0/1-masks, although the element detection relies on mechanisms that fundamentally differ from mask matching. Owing to the multilinear forms, this kind of extraction of invariant features shall be called a multilinear approach to pattern decomposition (MPD). Figure 1b shows a network for the computation of a single feature. The generalized TLUs of Fig.1a are replaced by so-called product units (Durbin and Rumelhart, 1989; Giles and Maxwell, 1987; Glünder, 1986). According to Eq.(4), this network belongs to the category of the ΣΠ-networks (Rumelhart *et al.*, 1986).

Merely the minimal order q that guarantees perfect invariant pattern descriptions remains to be determined. Fortunately it has been proven by several authors – firstly, by McLaughlin and Raviv (1968), for binary pattern signals by Minsky and Papert (1969), and for complex-valued signals by Lohmann and Wirnitzer (1984) –, that any pattern of finite extent is perfectly described by its complete triple-autocorrelation function, except for its translatory position (translation invariance). Actually, all possible ${}^3\kappa$ and ${}^2\kappa$ trilinear and bilinear autocorrelation coefficients, in conjunction with a subsequent template-matching stage, permit ideal translation invariant classifications. Therefore, the necessary and sufficient condition and the costs for ideal MPD classification are

$$(q = 2) \wedge (q = 3) \Rightarrow (N = {}^2\kappa + {}^3\kappa \approx n^2/6) \qquad \text{for} \quad n >> 3 \ . \tag{5}$$

This result expresses the primary role polynomial order q plays in ideal invariant classification and it demonstrates that the polynomial degree p is relevant only insofar, as multilinear terms are to be generated through polynomials, e.g. by generalized TLUs. Features that are invariant under other geometric transformations can be evaluated via generalized autocorrelation (Glünder, 1987).

## Conclusions

Somewhat surprisingly, the costs for list, PPD, and MPD classification turn out to be almost the same, namely $N \approx n^2$ weighting coefficients per class that must be adjusted or 'learned' in order to obtain unrestricted translation invariant and unequivocal class definitions. While list classifiers suffer from the mentioned shortcoming, the PPD and MPD approaches appear functionally equivalent,

at least theoretically. In practice, however, PPD networks must cope with a high signal dynamic that is caused by the large number of polynomial terms that constitute each feature. On the other hand, the processing must be highly accurate in order to represent the crucial trilinear terms which otherwise are lost in the polynomials. In this regard, MPD networks are more economic and robust.

All three approaches permit a proportional decrease in the number of templates, masks, or multiplication schemes if the variance is restricted. This property is a consequence of the fact that invariance is exclusively due to the averaging of nonlinearly weighted correlation coefficients, or multilinear terms which are extracted at different positions, according to the extent of the geometric variance, as it was already pointed out by Pitts and McCulloch (1947).

Because every single PPD or MPD feature is invariant, a reduction of the number of features does not affect the invariance but the precision of the pattern descriptions. In this sense, good pattern descriptions even require certain types of variance, namely all those geometric transformations that relate the pixels of a pattern. That is why 2D translations, as well as expansions in conjunction with rotations, are well suited for ideal pattern characterizations. – For example, a reduction of the number of mask types by a factor n, to all $^2\kappa \approx n/2$ possible features, implies that patterns with identical second order autocorrelation functions become indistinguishable.

Unfortunately, the decrease in descriptive power that is caused by fewer features happens more rapidly for PPD than for the MPD features, at least if non-negative signals are considered. This is due to the principal fact that PPD features cannot unambiguously signal the presence or absence of form elements in patterns which in turn is a consequence of the normalization problem. Owing to the coincidence-detection character of multiplication, MPD features are ideally suited for such decisions, especially if they are computed from sparsely coded signal representations, i.e., those that contain only a few nonzero values. The MPD approach even permits unequivocal reconstructions of binary pattern versions from binarized multilinear terms. Consequently, the replacement of the multiplication nodes by real coincidence detectors drastically improves the noise immunity of the feature extraction process. PPD systems lack this valuable property because pattern information is essentially coded in the amplitudes of polynomials.

## References


Brousil, J.K. and Smith, D.R. (1967) A threshold logic network for shape invariance. *IEEE Trans. Electronic Computers*, **EC-16**, 818-828

Casasent, D. and Psaltis, D. (1976) Position, rotation and scale invariant optical correlation. *Applied Optics*, **15**, 1795-1799

Durbin, R. and Rumelhart, D.E. (1989) Product units: a computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, **1**, 133-142

Giles, C.L. and Maxwell, T. (1987) Learning, invariances, and generalization in high-order neural networks. *Applied Optics*, **26**, 4972-4978

Glünder, H. (1986) Neural computation of inner geometric pattern relations. *Biol. Cybern.*, **55**, 239-251

Glünder, H. (1987) Invariant description of pictorial patterns via generalized autocorrelation functions. In Meyer-Ebrecht, D. (ed.), *ASST '87*, Springer Verlag, Berlin, 84-87

Lippmann, R.P. (1987) An introduction to computing with neural nets. *IEEE ASSP Magazine*, **4**, 4-22

Lohmann, A.W. and Wirnitzer, B. (1984) Triple correlations. *Proc. IEEE*, **72**, 889-901

McLaughlin, J.A. and Raviv, J. (1968) Nth-order autocorrelations in pattern recognition. *Information and Control*, **12**, 121-142

Pitts, W. and McCulloch W.S. (1947) How do we know universals. The perception of auditory and visual forms. *Bull. Math. Biophys.*, **9**, 127-147

Rosenblatt, F. (1962) *Principles of Neurodynamics – Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, Washington/DC

Rumelhart, D.E.; Hinton, G.E. and McClelland, J.L. (1986) A general framework for parallel distributed processing. In Rumelhart, D.E. and McClelland, J.L. (eds.), *Parallel Distributed Processing 1*, The MIT Press, Cambridge/MA, 45-76

# INNC 90 PARIS

## Volume 2

# INTERNATIONAL NEURAL NETWORK CONFERENCE

JULY 9-13, 1990
PALAIS DES CONGRES - PARIS - FRANCE

**UNDER THE PATRONAGE OF THE COMMISSION OF THE EUROPEAN COMMUNITIES**

**THE INTERNATIONAL NEURAL NETWORK SOCIETY (INNS), THE IEEE NEURAL NETWORK COUNCIL COOPERATING SOCIETIES**

DG13 Telecommunications Information Industries and Innovation

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

KLUWER ACADEMIC PUBLISHERS
DORDRECHT / BOSTON / LONDON