3 WONG, W.S., NAMIKI, S., MARGALIT, M., HAUS, H.A., and IPPEN, E.P.: 'Self-switching of optical pulses in dispersion-imbalanced nonlinear loop mirrors', *Opt. Lett.*, 1997, **22**, pp. 1150–1152
4 SEONG, N.H., KIM, DUG Y., and OH, SEONG K.: 'Figure-eight laser by using a dispersion-imbalanced nonlinear optical loop mirror with lumped dispersive elements', submitted to *Opt. Lett.*,
5 GRUDININ, A.B., RICHARDSON, D.J., and PAYNE, D.N.: 'Energy quantisation in figure eight fibre laser', *Electron. Lett.*, 1992, **28**, pp. 67–68
6 DULING, I.N., III: 'Compact source of ultrashort pulses' (Cambridge University Press, 1995), pp. 11–18
7 KO, K.Y., DEMOKAN, M.S., and TAM, H.Y.: 'Transient analysis of erbium-doped fiber amplifiers', *IEEE Photonics Technol. Lett.*, 1994, **6**, pp. 1436–1438

Automatic pattern recognition by similarity representations

E. Pękalska and R.P.W. Duin

The automatic recognition of objects may benefit from using a similarity representation instead of the traditional approach based on features. It is shown that the common use of nearest neighbour classifiers for similarity representations may be improved significantly by other classification rules with respect to recognition accuracy as well as computational complexity.

Introduction: Traditionally, objects in automatic pattern recognition systems are represented by characteristic features. Classifiers are optimised on these using examples of objects for training. In some applications, however, the definition of features is not straightforward. Alternatively, similarities between new objects and prototypes may be used directly on the raw or preprocessed input measurements [1 – 4], e.g. based on template matching. For some applications like two-dimensional-shape recognition, this is much more feasible. The nearest neighbour (NN) rule is conventionally applied to classify new objects on the basis of their similarities to the set of training examples. This procedure is either computationally expensive (the similarities to all training examples have to be found) or it potentially loses accuracy when only a small number of prototypes is selected.

We propose to overcome this problem of similarity representations by replacing the traditional NN method by a more global classifier defined on the similarities to a small set of prototypes, called the representation set. The classifier will be trained by the larger training set. For its evaluation, however, just a small set of similarities to the representation set objects has to be computed. Our experiments clearly demonstrate that the trade-off between the recognition accuracy and the computational effort is significantly improved by using such classifiers (effectively a linear combination of similarities) instead of using the nearest neighbour rule.

Method: Having found a data representation of objects, a precise goal is to learn a decision rule. To construct such a rule [2], the training set T of size n (based on n objects) and the representation set R of size r will be used. R is a set of prototypes covering all categories present. We choose R to be a subset of T ($R \subseteq T$), although, in general, they might be disjunct. In this Letter we will use dissimilarities (distances) instead of similarities. In the learning process, a classifier is then built on the $n \times r$ distance matrix $D(T, R)$, relating all training objects to all prototypes. A set S of s new objects is represented by their distances to R, i.e. as an $s \times r$ matrix $D(S, R)$.

A straightforward way of dealing with dissimilarity representations leads to the NN rule [5] or more generally to the instance-based learning [6, 7]. In its simplest form, the NN rule assigns a new object to the class of its nearest neighbour in the representation set R by finding minima in the rows of $D(S, R)$. In the NN method, usually, R is equivalent to T. However, condensed variants [8] exist which reduce the size of R.

In the dissimilarity space proposed by us, each dimension corresponds with an object in the set R. The prototypes thereby constitute an r-dimensional dissimilarity space. In general, $D(x, R)$ defines a vector of r distances between the object x and the members of R. Therefore, this function can be seen as a mapping which embeds x in an r-dimensional space. The advantage of such a representation is that any traditional classifier defined for feature spaces may be used. Moreover, in contrast to the NN rule, it can be based on training sets larger than the given representation set. This does not complicate the decision rule, but increases its accuracy.

Many of the commonly used dissimilarity measures, e.g. the Euclidean distance, the mean square error and the Hamming distance, are based on sums of differences between measurements. For representation sets consisting of independent objects, the distribution of summation-based distances tends to be approximately normally distributed according to the central limit theorem. Consequently Bayesian classifiers [9] assuming normal distributions, should perform well on such dissimilarity spaces. The RLNC (regularised linear normal densities based classifier) [5], assuming equal class covariance matrices, is particularly interesting because of its simplicity.

The NN rule generalises well for large training sets. Asymptotically, its error is bounded from above by twice the Bayes error (the smallest error possible) [5]. In practice, however, the size of R is often too small to reach such an accuracy. Other classifiers, such as the RLNC mentioned above, might be more advantageous on dissimilarity representations [3, 4]. It may perform much better since they become less local in their decisions by operating on larger training sets. Note that the NN rule as we use it here, operates directly on the individual dissimilarities, while the RLNC is defined for the representation space, treating the dissimilarities as its input features.

Experiments: To compare the behaviour of the NN rule and the RLNC built on distances, a number of experiments are conducted. They are designed to observe and analyse two phenomena: a better performance of the RLNC and the positive influence of training sets larger than representation sets used for its construction. Three real world datasets are considered in our study: two versions of the subset of handwritten numerals (ten classes) of the NIST digit sets [10] (one represented by pixel-based Euclidean distances and one on the modified Hausdorff distances [11] on the contours), and a set chromosomes of 24 classes represented by differences in their DNA band profiles. Since for digits and chromosome bands no natural features arise from the application, constructing dissimilarities is an interesting possibility to tackle the automatic recognition problem.

To obtain meaningful conclusions, the experiments are performed 25 times for all datasets and the results are averaged. In a single experiment, each dataset is first randomly split into two equal-sized sets: the design set L and the test set S. L serves to obtain both the representation set R and the training set T. There exist many ways to select R. As our goal here is not to find the best R for the given problem, but to illustrate the usefulness of our approach, we will restrict ourselves to random selection. Representation sets of different sizes are studied: the initial set is chosen as a subset of L and is then gradually increased until it equals the entire set L.

The error rates for the three datasets of the two classification rules are shown in Figs. 1 – 3, relating the size n of the training set T to the size r of the representation set R. The NN results are presented (by *) on the diagonal as for this case T is identical to R. For the RLNC, the lines of constant classification error show how the same performance is reached by a larger training set represented by a smaller representation. The following observations can be made:

(i) The starting points of the RLNC curves on the diagonals have a smaller error than the equivalent NN rules. So the RLNC generalises better.

(ii) In the horizontal direction, for increasing training set sizes, classifiers have equal computational complexity for evaluation of new objects. For larger training sets the RLNC shows performances up to almost two times as good as the NN rule for the same computational complexity.

(iii) Following the curves of the RLNC, it can be observed that for larger training sets much smaller sizes of the representation sets are needed for the same performance. The RLNC sometimes demands just half the computational complexity for evaluation compared to the NN rule.

The good performance of the RLNC can be understood as follows. It is in fact just a weighted linear combination of dissimilarities between an object and the representation set. The training process emphasises prototypes which play a crucial role during discrimination, but it still allows other prototypes to influence the decision. The importance of prototypes is reflected in the weights. By combining the distance information with appropriate weights a globally sensitive classifier is built, which cannot be achieved by the NN rule.

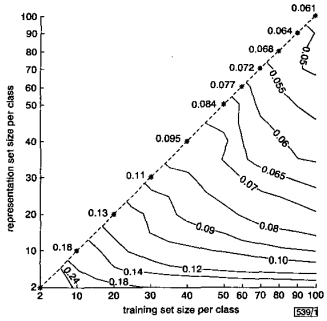

Fig. 1 *Generalisation error of RLNC and NN rule for different combinations of sizes of representation set and training set for Euclidean distances of pixel-based digits*

—— RLNC ✳ NN rule

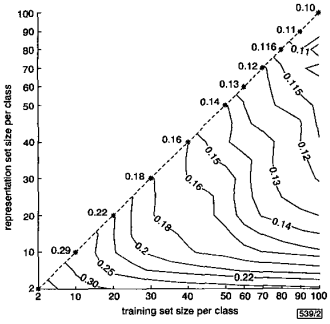

Fig. 2 *Generalisation error of RLNC and NN rule for different combinations of sizes of representation set and training set for modified Hausdorff distances of contour digits*

—— RLNC ✳ NN rule

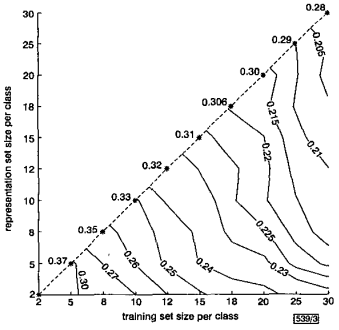

Fig. 3 *Generalisation error of RLNC and NN rule for different combinations of sizes of representation set and training set for DNA differences of chromosome bands*

—— RLNC ✳ NN rule

Conclusions: Dissimilarity representations, replacing the traditional feature-based description, open a new area for exploiting expert knowledge. The experiments manifest the strong point of our approach in contrast to the nearest neighbour method: additional objects may be used to enlarge the training set without increasing the computational complexity, allowing us to build a decision rule that generalises significantly better. In particular, we show that a linear classifier constructed on dissimilarities mostly outperforms the NN rule traditionally applied.

Our results encourage the exploration of meaningful dissimilarity information in new, advantageous ways, of which our proposal is an example. The use of other classifiers, e.g. quadratic (which by including product interactions between prototype dissimilarities may have better generalisation capabilities) and the study of representation set selection is an issue of further research.

Acknowledgments: This work was partly supported by the Dutch Organization for Scientific Research (NWO).

27 November 2000
Electronics Letters Online No: 20010121
DOI: 10.1049/el:20010121

E. Pękalska and R.P.W. Duin (*Pattern Recognition Group, Department of Applied Physics, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*)

E-mail: duin@ph.tn.tudelft.nl

References

1 EDELMAN, S.: 'Representation and recognition in vision' (MIT Press, Cambridge, 1999)
2 DUIN, R.P.W.: 'Classifiers in almost empty spaces'. Proc. 15th Int. Conf. Pattern Recognition, 2000, Vol. 2, pp. 1–7
3 DUIN, R.P.W., PĘKALSKA, E., and DE RIDDER, D.: 'Relational discriminant analysis', *Pattern Recognit. Lett.*, 1999, **20**, (11–13), pp. 1175–1181
4 PĘKALSKA, E., and DUIN, R.P.W.: 'Classifiers for dissimilarity-based pattern recognition'. Proc. 15th Int. Conf. Pattern Recognition, Barcelona, 2000, Vol. 2, pp. 12–16
5 RIPLEY, B.: 'Pattern recognition and neural networks' (Cambridge University Press, Cambridge, UK, 1996)
6 DASARATHY, B.V.: 'Nearest neighbor (NN) norms: NN pattern classification techniques' (IEEE Computer Society Press, Los Alamitos, CA, 1991)
7 AHA, D.W., KIBLER, D., and ALBERT, M.K.: 'Instance-based learning algorithms', *Mach. Learn.*, 1991, **6**, pp. 37–66
8 DEVIJVER, P.A., and KITTLER, J.: 'Pattern recognition: a statistical approach' (Prentice/Hall, London, 1982)
9 JAIN, A.K., DUIN, R.P.W., and MAO, J.: 'Statistical pattern recognition: a review', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (1), pp. 4–37
10 WILSON, C.L., and GARRIS, M.D.: 'Handprinted character database 3'. Tech. Rpt., National Institute of Standards and Technology, 1992
11 DUBUISSON, M.R., and JAIN, A.K.: 'Modified Hausdorff distance for object matching'. Proc. 12th Int. Conf. Pattern Recognition, 1994, Vol. 1, pp. 566–568

Efficient method for early detection of all-zero DCT coefficients

Shi Jun and Songyu Yu

Based on the existing method, a more efficient algorithm for detecting all-zero DCT coefficient blocks before implementation of the discrete cosine transform and quantisation in very low bit rate video coding is proposed. The problem of the existing method is shown and an improved algorithm is provided. Simulation shows that the new method improves the detecting performance by ~10–20%.

Introduction: For very low bit rate coding, especially in videophone applications, it is quite common for the coefficients of the whole block to be all-zero after motion estimation, motion compensation, implementation of the discrete cosine transform (DCT) and quantisation. A method that detected all-zero DCT coefficient blocks before DCT and quantisation would greatly improve the coding speed. An algorithm that uses the sum of absolute difference (SAD) of each motion compensation block as a criterion was proposed in [1]. We believe that although the method proposed in [1] greatly improves the coding speed, its efficiency can still be improved upon. In this Letter, we analyse the limitations of the existing method for the early detection of all-zero coefficients, and propose an improved method using SAD criteria, with the efficiency limit as a guideline. Finally, we propose an experimental rule used in H.263 [2]. Compared with the existing method, the new method improves the detecting efficiency by ~10–20%.