

Face Verification Using Error Correcting Output Codes

J Kittler, R Ghaderi, T Windeatt and J Matas
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, Surrey GU2 7XH, UK
[J.Kittler, T.Windeatt, J.Matas]@eim.surrey.ac.uk

Abstract

The Error Correcting Output Coding (ECOC) approach to classifier design decomposes a multi-class problem into a set of complementary two-class problems. We show how to apply the ECOC concept to automatic face verification, which is inherently a two-class problem. The output of the binary classifiers defines the ECOC feature space, in which it is easier to separate transformed patterns representing clients and impostors. We propose two different combining strategies as the matching score for face verification. The first uses the first order Minkowski metric, and requires a threshold to be set. The second is a kernel-based method and has no parameters to set. The proposed method exhibits better performance on the well known XM2VTS data set compared with previous reported results.

1 Introduction

Automation of a system that performs personal identity verification may use a variety of biometric modalities including facial features, voice characteristics, iris scan, fingerprints. One approach to improving accuracy and robustness of such systems is by combining different modalities, for example voice and face data as in [2]. Although this type of multimodal approach can achieve significant performance improvement, it is still desirable to concentrate on biometric subsystem performance in order to further reduce error rates. Facial images are a popular source of biometric information since they are relatively easy to acquire, and provide discriminatory features used by humans for recognition. However automated face verification systems often have poor levels of performance and improving them is known to be a difficult task. Some advances recently reported in this context include those described in [9].

A different approach to increasing accuracy is provided by the method of combining multiple experts within a single modality. For example, in [8] it was shown that by combining the scores of several diverse face verification systems the error rate of the best expert could be reduced by more than 42 percent. This approach draws on the results

in multiple classifier fusion [10]. Informally, the idea is that for some complex problems it may be better to combine relatively simple multiple experts with somewhat differing opinions rather than designing a single complex expert. If experts are not too well correlated and a suitable combining rule can be found, it has been shown that simpler and more accurate systems may result. Several different techniques exist to ensure diversity among experts, the Error Correcting Output Codes (ECOC) method being just one of them.

In this paper we report on the novel use of ECOC for designing multiple experts for face verification. Use of ECOC for decomposing a multi-class problem into a set of complementary two class problems is a well established method in many applications [4, 5, 6, 7, 11, 21, 23, 22, 24]. Such a decomposition means that attention can be focused on developing an effective technique for the two-class classifier, without having to consider explicitly the design and automation of the multi-class case. It is also hoped that the parameters of a simple expert run many times may be easier to determine than a complex expert run once. When first suggested ECOC was based on the idea of using error-correcting codes as class labels, so that individual classification errors propagated from a set of binary classifiers could potentially be corrected [4]. For a two-class problem, classification errors can be one of two types, either predicted class C_1 for target class C_2 or predicted class C_2 for target class C_1 .

At first sight, it may seem that ECOC, which is aimed at multi-class problems, is not suited for face verification which has a single class of clients and a single class of impostors. However in this paper we show how to divide the verification task into two stages, the first being a multi-class recognition problem.

The paper is organised as follows. The original ECOC classification method is explained in Section 2, along with a discussion of how clients can be represented in ECOC feature space and two methods of checking identity. In section 3 we describe the representation of patterns for face verification. The results over XM2VTS face data base come in section 5, followed by conclusions.

2 ECOC method and verification

The original motivation for encoding multiple classifiers using an error-correcting code is based on the idea of modelling the prediction task as a communication problem, in which class information is transmitted over a channel [5]. Errors introduced into the process arise from various aspects of the learning algorithm, including features selected and finite training sample. From error-correcting theory, we know that a matrix designed to have d bits error-correcting capability implies that there is a minimum Hamming Distance $2d + 1$ between any pair of code words. Assuming each bit is transmitted independently, it is then possible to correct a received pattern having fewer than d bits in error, by assigning the pattern to the code word closest in Hamming distance. The ability to detect and possibly correct errors is dependent on the assumption that each error is independently produced. While in practice some errors will be correlated, the experimental evidence reported in [5] is that application of ECOC principles does lead to reduced error.

In the ECOC method, a $k \times b$ binary code word matrix Z has one row (code word) for each of k classes, with each column defining one of b sub-problems that use a different labelling. Specifically, for the j th sub-problem, a training pattern with target class C_i ($i = 1 \dots k$) is re-labelled as class C_1 if $Z_{ij} = x$ and as class C_2 if $Z_{ij} = \bar{x}$ (where x is a binary variable, typically zero or one). One way of looking at the re-labelling is to consider the k classes as being arranged into two super-classes.

A summary of the original ECOC Classification algorithm is as follows:

Summary of Training: for $j = 1 : b$

- re-label training patterns into two classes (super-classes) according to binary element corresponding to each class for column j
- train a binary classifier using the re-labelled training set

Summary of Testing:

- apply pattern to the b trained classifiers forming vector

$$\bar{y} = [y_1, y_2, \dots, y_b]^T \quad (1)$$

in which y_j is the real-valued output of j th base classifier

- compute distance between output vector and code word for each class

$$L_i = \sum_{j=1}^b |Z_{i,j} - y_j| \quad (2)$$

- assign pattern to class corresponding to closest code word $ArgMin(L_i)$

The main constraint in designing Z is the distance between rows. Indeed, if Z is an equidistant code, the combining strategy is identical to the Bayesian decision rule [23]. Reference [18] explains the complexities involved in designing matrices with well separated rows. The ECOC method has an additional constraint which requires that distance between columns be high, in order to ensure diversity among experts. In our experiments we use BCH coding method with *allzero* code word removed, and we obtain equidistant rows by the *over produce and select* strategy. Further details, and a comparison of BCH with random codes can be found in [21].

From another perspective, we can view equation 1 as providing posterior probability of super-class membership. The decomposition defined by Z can then be interpreted as a transformation between spaces representing probabilities of individual class and super-class membership [24]. In fact, a solution to recovering individual class probabilities is based on method of least squares, providing $Z Z^T$ is non-singular. Reference [7] can be consulted for a discussion of the importance of the super-class probabilities as source of effectiveness of ECOC. The classifier outputs represent the estimates of super-class probabilities and these are the estimates that we use to represent clients and impostors for face verification.

We propose to represent each client i by a set Y_i of N ECOC classifier output vectors, i.e.

$$Y_i = \{y_i^l | l = 1, 2, \dots, N\} \quad (3)$$

where N is the number of i -th client patterns available for training.

Two methods of combining are proposed.

2.1 Distance based combination

We wish to ascertain whether the classifier outputs are jointly consistent with the claimed identity. This could be accomplished by setting a threshold on the distance of the outputs from the client code. However, the compound code represents an idealised target, rather than the real distribution of these outputs. Thus measuring the distance from the client code could be misleading, especially in spaces of high dimensionality. One alternative would be to adopt the *centroid* of the joint classifier outputs to characterise each client and to measure the consistency of a new client claim from this representation. Incidentally, the use of centroid in the context of ECOC classifiers is also advocated in [7]. However, as we have only a very small number of training samples, the estimated centroid would be very unreliable.

In order to test the hypothesis that the client claim is authentic we adopt as a test statistic the average distance between vector \underline{y} and the elements of set Y_i . The distance is measured using first order Minkowski metric rather than Euclidean metric in order to achieve better robustness to outliers as in (2), i.e.

$$d_i(\underline{y}) = \frac{1}{N} \sum_{l=1}^N \sum_{j=1}^b |y_j^l - y_j| \quad (4)$$

where y_j is the j th binary classifier output for the test pattern, and y_j^l is the j th classifier output for the l th member of class i . The distance is checked against a decision threshold, t . If the distance is below the threshold, client's claim is accepted, otherwise it is rejected.

2.2 Kernel combination

Note that instead of measuring the distance between points, we could measure a between point similarity which can be expressed by a kernel function that assumes a maximum when the distance is zero and monotonically decreases as the distance increases. The design of the decision function cannot involve any training as the number of points available is extremely small. We simply use exponential kernels with fixed width σ . The centres do not need to be explicitly determined because we use $d_i(\underline{y})$ in the exponent of the kernel to measure similarity of \underline{y} to class i . We allocate one kernel per client and a number of kernels for each imposter. We measure the relative similarities of a test vector to the claimed identity and to the impostors as

$$k_i(\underline{y}) = \sum_{\alpha} w_{\alpha} \exp\left\{-\frac{d_{\alpha}(\underline{y})}{\sigma^2}\right\} \quad (5)$$

where index α runs over all imposter kernel placements and client i , and the weights w_{α} are estimated. The client claim test is carried out as follows:

$$k_i(\underline{y}) \begin{cases} \geq 0.5 & \text{accept claim} \\ < 0.5 & \text{reject claim} \end{cases} \quad (6)$$

3 Face Image Representation

Normalisation or standardisation is an important stage in face recognition or verification. Face images differ in both shape and intensity, so *shape alignment* (geometric normalisation) and *intensity correction* (photometric normalisation) can improve performance of the designed system. Our approach to geometric normalisation has been based on eye position. Four parameters are computed from the eye coordinates (rotation, scaling and translation in horizontal and vertical directions) to crop the face part from the original image and scale it to any desired resolution. Here we use

“manually localised” eye coordinates to eliminate the dependency of the experiments on processes which may lack robustness. In this way, we can focus our investigation on how the performance is affected by the methodology of verification and in particular by the ECOC technique. For photometric normalisation we have used histogram equalisation as it has exhibited better performance in comparison with other existing methods[12].

Although it is possible to use gray levels directly, as demonstrated in earlier experiments [19, 16], normally features are first extracted. There are many techniques in the pattern recognition literature for extracting and selecting effective features that provide maximal class separation in the feature space [3]. One popular approach is *Linear Discriminant Analysis (LDA)* which is used in our experiments. We briefly review the theory of LDA, and how it is applied to face recognition or verification. Further details may be found in [3, 1].

Given a set of vectors $x_i, i = 1, \dots, M, x_i \in R^D$, each belonging to one of c classes $\{C_1, C_2, \dots, C_c\}$, we compute the between-class scatter matrix, S_B ,

$$S_B = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^T \quad (7)$$

and within-class scatter matrix, S_W

$$S_W = \sum_{i=1}^c \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (8)$$

where μ is the grand mean and μ_i is the mean of class C_i .

The objective of LDA is to find the transformation matrix, W_{opt} , that maximises the ratio of determinants $\frac{|W^T S_B W|}{|W^T S_W W|}$. W_{opt} is known to be the solution of the following eigenvalue problem [3]:

$$S_B W - S_W W \Lambda = 0 \quad (9)$$

where Λ is a diagonal matrix whose elements are the eigenvalues of matrix $S_W^{-1} S_B$. The column vectors \vec{w}_i ($i = 1, \dots, c - 1$) of matrix W are referred to as *fisherfaces* in [1].

In high dimensional problems (e.g. in the case where x_i are images and D is $\approx 10^5$) S_W is almost always singular, since the number of training samples M is much smaller than D . Therefore, an initial dimensionality reduction must be applied before solving the eigenvalue problem in (9). Commonly, dimensionality reduction is achieved by Principal Component Analysis [20][1]; the first $(M - c)$ eigenprojections are used to represent vectors x_i . The dimensionality reduction also allows S_W and S_B to be efficiently calculated. The optimal linear feature extractor W_{opt} is then defined as:

$$W_{opt} = W_{lda} * W_{pca} \quad (10)$$

where W_{pca} is the PCA projection matrix and W_{lda} is the optimal projection obtained by maximising

$$W_{lda} = \arg \max_W \frac{|W^T W_{pca}^T S_W W_{pca} W|}{|W^T W_{pca}^T S_B W_{pca} W|} \quad (11)$$

4 Experiments on XM2VTS Data Base

We use the XM2VTS face database as it is known to be challenging and several results of experiments, carried out according to an internationally agreed protocol using other verification methods, are readily available in the literature.

4.1 Database and experimental protocol

The extended M2VTS (XM2VTS) database contains 295 subjects. The subjects were recorded in four separate sessions uniformly distributed over a period of 5 months, and within each session a number of shots were taken including both frontal-view and rotation sequences. In the frontal-view sequences the subjects read a specific text (providing synchronised image and speech data), and in the rotation sequences the head was moved vertically and horizontally (providing information useful for 3D surface modelling of the head). Further details of this database can be found in [17].¹

The experimental protocol (known as Lausanne evaluation protocol) provides a framework within which the performance of vision-based (and speech-based) person authentication systems running on the extended M2VTS database can be measured. The protocol assigns 200 clients and 95 impostors. Two shots of each session for each subject's frontal or near frontal images are selected to compose two configurations. We used the first configuration which is more difficult as the reported results show [15]. In this configuration, for each client there are 3 training, 3 evaluation and 2 test images. The impostor set is partitioned into 25 evaluation and 70 test impostors. Within the protocol, the verification performance is measured using the false acceptance and the false rejection rates. The operating point where these two error rates equal each other is typically referred to as the equal error rate point. Details of the this protocol can be found in [13].²

4.2 System description

All images are projected to a lower dimensional feature space as described in Section 3, so that each pattern is represented by a vector with 199 elements. There are 200 clients,

¹<http://www.ee.surrey.ac.uk/Research/VSSP/xm2fdb.html>

²http://www.idiap.ch/~m2vts/Experiments/xm2vtsdb_protocol_october.ps

so from the identification viewpoint we are facing a 200 class problem. We use equi-distant codes generated by the BCH method, containing 200 entries (compound labels) and 511 bit long. The codes were generated as explained in section 2 The Hamming distance between any pair of labels is 256 bits.

For the verification task, the level-zero classifier is a Multi-Layer Perceptron (MLP) with one hidden layer containing 199 input nodes, 35 hidden nodes and two output nodes. The Back-propagation algorithm with fixed learning rate, momentum and number of epochs is used for training. The dual output is mapped to a value between "0" and "1" to give an estimation of probability of super-class membership.

The outputs of the MLPs define an ECOC feature vector, and from equation (4), $d_i(y)$ for the claimed identity i is calculated by averaging over respective class images. Both distance and similarity based rules for combining the outputs of the ECOC multiple classifiers have been investigated. Of the two decision functions, the distance based rule is the only one that depends on a parameter, the decision threshold, that has to be selected.

4.3 Two Combining Methods

Normally one would use the evaluation set data to compute the Receiver Operating Characteristics (ROC) curve which plots the relationship of false rejection rate and false acceptance rate as a function of threshold. A suitable threshold is then selected to achieve the required behaviour. For instance, one can specify the threshold that delivers equal false rejection and false acceptance rates. The threshold can be selected for each client separately, or globally by averaging the errors over all the clients.

One of the difficulties encountered with our ECOC based approach was that because the level-zero classifier was "too powerful", the FR and FA errors on the evaluation set were zero for a large range of thresholds. In such circumstances the ROC curve is not very useful in threshold setting. This problem was circumvented by the following procedure. Starting from $t = 0$ we successively increased the threshold in fixed steps to find the point where the total error (the sum of FR and FA errors) is minimum. If the total error was zero for several such increments the selected threshold would correspond to the point just before the total error would start rising.

The results obtained with the above threshold selection procedure are given in Table 1 as a function of step size. As different step sizes terminate the threshold selection procedure at different destinations from the impostors in the evaluation set the test set performance varies. In Table 1 we see that when the evaluation set by itself is used to set thresholds, error rates are higher than when using the com-

step size	FR(e)	FA (e)	FR(e+tr)	FA(e+tr)
.25	13.250	0.108	8.75	.384
.2	10.500	0.142	6.75	.168
.1	6.500	0.277	4.5	.217
.05	5.250	0.413	3.25	.367
.01	4.750	0.654	1.25	.649
.005	4.750	0.711	1.25	.704
.001	4.500	0.739	1.25	.748

Table 1: False Rejection and Acceptance rates on test sets when clients of (i) evaluation set (e) and (ii) evaluation and training set (e+tr) are used to set thresholds

bined training/evaluations sets. Even though generalisation has improved, there is still zero error rate performance on the evaluation set so that it cannot be used to select the step size. This may suggest that we could incorporate the evaluation set into ECOC training, but we did not try it.

To demonstrate the effectiveness of ECOC we report in Table 2 the result of applying the exhaustive search method directly to the original 199 dimensional feature vectors. Comparing Tables 1 and 2, the benefits of mapping the input data onto the ECOC output vectors are clearly visible. Note also that in this case the evaluation set error rates are non zero, i.e. the populations of clients and impostors are overlapping. In this particular case the ROC curve could have been computed but we did not pursue this particular scheme as it was clearly inferior to the ECOC based approach.

step size	FR(Ev)	FA (Ev)	FR(Ts)	FA(Ts)
.25	1.67	.89	16.75	1.105
.2	0.83	1.07	15.25	1.44
.1	.5	.35	12.75	.735
.01	.167	.33	8	1.18
.005	.167	.31	8	1.239
.001	.167	.292	8	1.311

Table 2: False Rejection and Acceptance rates on evaluation (Ev) and test (Ts) sets when features are applied directly

Although the kernel combination method requires no thresholds, there are design parameters that can be varied to control the behaviour of the method. In particular, we can choose different ways to represent impostors. Each of the 25 evaluation impostors has 4 sets of 2 images as explained in Section 4.1. Therefore, as an alternative to 25 centres averaged over 4 sets we can choose 50 centres averaged over 2 sets or 100 centres averaged over 1 set. The error rates for 25, 50, 100 impostor centres, are shown in Table 3. In comparison with Table 1, there is a different trade-off between false acceptance and false rejection rates.

impostor centres	FR(Ev)	FA(Ev)	FR(Ts)	FA(Ts)
25	0	0	0.750	0.883
50	0	0	0.500	0.879
100	0	0	0.750	1.245

Table 3: False Rejection and Acceptance rates(on evaluation and test sets) using kernel combination with various number of impostor centres

For comparison we are including the results obtained using three other methods on the same data set and with the same protocol. The methods use the same representation of image data in terms of 199 fisher face coefficients. They employ three different scores for decision making in this feature space. In particular, we use the Euclidean metric, s_E , Normalised correlation, s_N , and Gradient metric, s_O , as detailed in [9]. The results are summarised in Table 4.

Score	Evaluation set			Test set		
	FR	FA	TE	FR	FA	TE
s_E	7.83	7.83	15.66	5.50	7.35	12.85
s_N	2.50	2.50	5.00	2.25	2.56	4.81
s_O	1.74	1.74	3.48	1.75	1.70	3.45

Table 4: Performance of the three baseline matching scores on manually registered images

By comparing Table 4 and Table 1 it would appear that the more robust metric used in $d_i(\underline{y})$ combined with the multiple representation of clients may be more effective than the Euclidean distance based score. From Table 1, Table 2 and Table 3 we see that all the ECOC based results are decisively superior to the decision making in the original Fisher face space. Also, from Table 1 and Table 3 the combination of ECOC multiple classifier outputs by means of the relative similarity score in (6) appears to yield slightly better results than using the distance based score $d_i(\underline{y})$. We conclude that if we can transform patterns representing clients and impostors into a feature space where they are more easily separated, then verification performance may improve. The first stage of the ECOC classification algorithm provides such a transformation, and both the combining methods that we tried in the experiments performed well.

5 Conclusion

We have described a face verification system based on the concept of Error Correcting Output Coding (ECOC), which decomposes a multi-class problem into a set of complementary two-class problems that can be solved with binary

classifiers. The output of the classifiers defines the ECOC feature space, in which we show that it is easier to separate transformed patterns representing clients and impostors. Face verification is inherently a two-class problem, and we proposed a solution that first generates a discriminant from the multi-class recognition problem defined by the ECOC matrix. For verification, the generated discriminant is checked for consistency with the distribution of responses for the particular client. Verification performance using two different combining methods on the XM2VTS face database was shown to be better than previous reported results ([15]).

Acknowledgments

The support received from OmniPerception Ltd, EPSRC Grant GR/M61320 and EU Framework V Project Banca is gratefully acknowledged.

References

- [1] P. et al Belhumeur. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *Proc. of ECCV'96*, pages 45–58, Cambridge, United Kingdom, 1996.
- [2] S Ben-Yacoub and S Luetttin, J Jonsson, K Matas, and J Kittler. Audio-visual person verification. In *Computer Vision and Pattern Recognition*, pages 580–585, Los Alamitos, California, Jun 1999.
- [3] P.A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.
- [4] T.G Dietterich and G. Bakiri. Error-correcting output codes: A general method for improving multiclass inductive learning programs. In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, pages 572–577. AAAI Press, 1991.
- [5] T.G. Dietterich and G Bakiri. Solving multi-class learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.
- [6] R. Ghaderi and T. Windeatt. Circular ecoc, a theoretical and experimental analysis. In *International Conference of Pattern Recognition(ICPR2000)*, pages 203–206, Barcelona,Spain, September 2000.
- [7] G. James. *Majority Vote Classifiers: Theory and Applications*. PhD thesis, Dept. of Statistics, Univ. of Stanford, May 1998. <http://www-stat.stanford.edu/gareth/>.
- [8] J Kittler, Y P Li, and J Matas. Face verification using client specific fisher faces. In J T Kent and R G Aykroyd, editors, *The Statistics of Directions, Shapes and Images*, pages 63–66, 2000.
- [9] J Kittler, Y P Li, and J Matas. On matching scores for lda-based face verification. In M Mirmehdi and B Thomas, editors, *British Machine Vision Conference*, pages 42–51, 2000.
- [10] J Kittler and F Roli. *Multiple Classifier Systems*. Springer-Verlag, Berlin, 2000.
- [11] E.B. Kong and T.G. Diettrich. Probability estimation via error-correcting output coding. In *Int. Conf. of Artificial Intelligence and soft computing*, Banff,Canada, 1997. <http://www.cs.orst.edu/tgd/cv/pubs.html>.
- [12] Y.P. Li. *Linear Discriminant Analysis and its application to face Identification*. PhD thesis, School of Electronic Engineering, Information technology and Mathematics, University of Surrey, Guildford, Surrey, U.K. GU2 7X, September 2000.
- [13] J Luetttin and G. Maitre. *Evaluation Protocol For The Extended M2VTS Database (XM2VTS)*. Dalle Molle Institute for Perceptual Artificial Intelligence, P.O. Box 592 Martigny, Valais, Switzerland, July 1998. IDIAP-Com 98-05.
- [14] J. Matas, M. Hamouz, M. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, w. Gerstner, S. Ben-Yacoub, Y. Abduljaoued, and Y. Majoraz. Comparison of face verification results on the xm2vts database. In A. A Sanfeliu, J.J Villanueva, M. Vanrell, R. Alqueraz, J. Crowley, and Y. Shirai, editors, *Proceedings of the 15th ICPR*, volume 4, pages 858–863, Los Alamitos, USA, September 2000. IEEE Computer Soc Press.
- [15] J Matas, K Jonsson, and J Kittler. Fast face localisation and verification. *IVC*, 17(8):578–581, June 1999.
- [16] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. of AVBPA'99*, pages 72–77, 1999.
- [17] W.W. Peterson and JR. Weldon. *Error-Correcting Codes*. MIT press, Cambridge,MA, 1972.
- [18] F.S. Samaria and A.C. Harter. Parameterisation of a stochastic model for human face identification. In *Proceeding of the 2nd IEEE Workshop on application of computer vision*, Sarasota,Florida, 1994. <http://mambo.ucsc.edu/psl/olivetti.html>.
- [19] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization on human face. *Journal .Opt.Soc. Am. A*, 3(4):519–524, 1987.
- [20] T. Windeatt and R. Ghaderi. Binary codes for multi-class decision combining. In *14th Annual International Conference of Society of Photo-Optical Instrumentation Engineers (SPIE)*, volume 4051, pages 23–34, Florida,USA, April 2000.
- [21] T. Windeatt and R. Ghaderi. Binary labelling and decision level fusion. *Information Fusion*, 2(2):103–112, 2001.
- [22] Ghaderi R. Windeatt T. Multi-class learning and error-correcting code sensitivity. *Electronics Letters*, 36(19):1630–1632, Sep 2000.
- [23] Ghaderi R Windeatt T. Least squares and estimation measures via error correcting output code. In *2nd Int. Workshop Multiple Classifier Systems,Lecture notes in computer science, Springer-V erlag*, pages 148–157, Jul 2001.