Face Verification via ECOC

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]@eim.surrey.ac.uk

Abstract

We develop a novel approach to face verification based on the Error Correcting Output Coding (ECOC) classifier design concept. In the training phase the client set is repeatedly divided into two ECOC specified sub-sets (superclasses) to train a set of binary classifiers. The output of the classifiers defines the ECOC feature space, in which it is easier to separate transformed patterns representing clients and impostors. The proposed method exhibits superior verification performance on the well known XM2VTS data set as compared with previously reported results.

1 Introduction

In many security applications involving person authentication, the level of security can be considerably enhanced by means biometrics. Existing commercial systems are exploiting a myriad of biometric modalities including voice characteristics, iris scan and finger print. However, as a source of biometric information, the human face plays a particularly important role as facial images (photographs) not only can easily be acquired but also they convey discriminatory features which are routinely used for recognition by humans without the need for specialist training. This opens the possibility for a close human - machine interaction and cooperation.

Although a considerable progress has been made over recent years, face verification is still a challenging task. For this reason one of the recent paradigms has been to use multiple modalities to achieve robustness and improved performance. Typically, one would combine voice and face data [2] to achieve better verification rates (lower false rejection and false acceptance rates). However, the merits of the combination of other modalities including face profile, lip dynamics and 3D face information to name but a few have also been investigated. Although the multimodal approach has been demonstrated to achieve significant improvements, there is still the need to improve the performance of the constituent biometric subsystems to drive the error rates even lower. Some advances recently reported in this context include [9].

As an alternative solution to gain performance improvements, attempts have been made to combine the outputs of several decision making systems. This approach draws on the results in multiple classifier fusion [10]. By combining several opinions one can reduce the error variance of the outputs of the individual experts and achieve better error rates. 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 %. However,

such ad hoc designs of multiple expert systems may not necessarily produce the best solutions.

In this paper we propose a novel method for designing multiple expert face verification systems. It is based on the error correcting output codes (ECOC) approach developed for channel coding. The basic idea is to allocate additional bits over and above the bits required to code the source message in order to provide error correcting capability. In the context of pattern classification the idea implies that each class is represented by a more complex code than the conventional code $Z_{ij} = 0, \ \forall i \neq j$ and $Z_{ij} = 1, \ i = j$. The implementation of such error resilient code requires more than the usual number of classifiers.

The main difficulty in applying the ECOC classification method to the problem of face verification is that verification is a two class problem and ECOC is suited exclusively to multiclass problems. We overcome this difficulty by proposing a two stage solution to the verification problem. In the first stage we view the verification task as a recognition problem and develop an ECOC design to generate class specific discriminants. In fact we need only the discriminant for the class of the claimed identity. In the second stage we test the hypothesis that the generated discriminant is consistent with the distributions of responses for the particular client.

The proposed scheme leads to an effective design which exhibits the attractive properties of ECOC classifiers but at the same time it is applicable to the two class personal identity verification problem. The design approach has been tested on the XM2VTS face database using the Lausanne protocol. The false rejection and false acceptance rates achieved are superior to the best reported results on this database to date [14].

The paper is organised as follows. In Section 2 we describe how face images are represented. In Section 3 we outline the Error Correcting Output Code method and adapt it to the verification problem. The results of the proposed method obtained on the XM2VTS face database are reported in Section 4 which is followed by conclusions in Section 5.

2 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 [18, 15], 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 $x_i, i = 1, \ldots, M$, $x_i \in R^D$, each belonging to one of **c** classes $\{C_1, C_2, \ldots, C_c\}$, we compute the between-class scatter matrix, S_B,

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

and within-class scatter matrix, S_W

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{2}$$

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]:

$$S_B W - S_W W \Lambda = 0 \tag{3}$$

where Λ is a diagonal matrix whose elements are the eigenvalues of matrix $S_W^{-1} S_B$. The column vectors $\vec{w}_i$ ($i = 1, \ldots, c - 1$) 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 (3). 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:

$$W_{opt} = W_{lda} * W_{pca} \tag{4}$$

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

$$W_{lda} = arg \max_W \frac{|W^T W_{pca}^T S_W W_{pca} W|}{|W^T W_{pca}^T S_B W_{pca} W|} \tag{5}$$

3 ECOC Verification Method

Error-Correcting Output Coding (ECOC) is an information theoretic concept which suggests that there may be advantages in employing ECOC codes to represent different signals which should be distinguished from each other after being corrupted while passing through a transmission channel. Dieterich and Bakiri [4] suggest that classification can be modelled as a transmission channel consisting of "input features", "training samples", and "learning paradigm". Classes are represented by *code words* with large Hamming distance between any pair. ECOC is believed to improve performance both by decomposing the multi-class problem as well as by correcting errors in the decision-making stage [5]. The binary values in the code word matrix are determined by the code generation procedure; it is possible to choose values that provide a meaningful decomposition [19], but

usually there is no meaning attached [5, 6, 21, 11]. There are a few methods to find a set of code words with a guaranteed minimum distance between any pair, the most popular being the BCH codes [5, 17], which we use in our experiments.

To understand the ECOC algorithm, consider a $k \times b$ code word matrix Z (k is the number of classes) in which the k rows represent code words (labels), one for each class. In the training phase, for each column, the patterns are re-labelled according to the binary values ("1s" and "0s"), thereby defining two *super classes*. A binary classifier is trained b times, once for each column. Each pattern can now be transformed into ECOC feature space by the b classifiers, giving a vector

$$\underline{y} = [y_1, y_2, ... y_b]^T \tag{6}$$

in which y_j is the real-valued output of jth classifier. In the test phase, the distance between output vector and label for each class is determined by

$$L_i = \sum_{j=1}^{b} |Z_{ij} - y_j| \tag{7}$$

and a pattern is assigned to the class corresponding to the code word having minimum distance to $\underline{y}$.

In the case of verification, the task is somewhat different. 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. Incidently, 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. We propose to represent each client i by a set Y_i of N ECOC classifier output vectors, i.e.

$$Y_i = \{\underline{y}_i^l | l = 1, 2, ..., N\} \tag{8}$$

where N is the number of $i - th$ client patterns available for training. 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 robustnes to outliers as in (7), i.e.

$$d_i(\underline{y}) = \frac{1}{N} \sum_{l=1}^{N} \sum_{j=1}^{b} |y_j^l - y_j| \tag{9}$$

where y_j is the jth binary classifier output for the test pattern, and y_j^l is the jth classifier output for the lth 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, i.e.

$$d_i(\underline{y})\{ \begin{matrix} \leq t & accept\ claim \\ > t & reject\ claim \end{matrix} \tag{10}$$

596

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

$$k_i(\underline{y}) = \sum_\alpha w_\alpha \exp\{-\frac{d_\alpha(\underline{y})}{\sigma^2}\} \tag{11}$$

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:

$$k_i(\underline{y})\{ \begin{array}{ll} \geq 0.5 & accept\ claim \\ < 0.5 & reject\ claim \end{array} \tag{12}$$

4 Experiments on XM2VTS Data Base

The aim of the experiments reported in this section is to evaluate the proposed approach to personal identity verification and to compare the results with other verification methods. We use the XM2VTS face database for this purpose 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 [16]. [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 [14]. 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

---

[1] *http://www.ee.surrey.ac.uk/Research/VSSP/xm2fdb.html*

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 2, so that each pattern is represented by a vector with 199 elements. There are 200 clients, so from the identification viewpoint we are facing a 200 class problem. We use a BCH equi-distant codes codebook containing 200 entries (compound labels) 511 bit long. The codes were generated by means of the *over produce and select* strategy. The Hamming distance between any pair of labels is 256 bits. The choice of code and the advantages of equi-distant codes are discussed in [21]).

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. For the identification task, we used an MLP with three hidden nodes.

As explained in Section 3, the outputs of the MLPs define an ECOC feature vector, and from equation (9), $d_i(\underline{y})$ 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.

4.2.1 Distance based combination

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 using the evaluation set data 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 2 we report error rates when seeds from both the evaluation and training sets are used to set the thresholds. Even though generalisation

---

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

598

search step	FR(Ev)	FA (Ev)	FR(Ts)	FA(Ts)
.25	0	0	13.2500	0.1078
.2	0	0	10.5000	0.1422
.1	0	0	6.5000	0.2772
.05	0	0	5.2500	0.4130
.01	0	0	4.7500	0.6540
.005	0	0	4.7500	0.7111
.001	0	0	4.5000	0.7391

Table 1: Result of verification when the clients in the evaluation set are used as seeds

has improved, it is not clear from the evaluation set performance how to select the best step size. One possibility is to combine the results from all step sizes, and the final row of table 2 shows the result of majority vote combination.

search step	FR(Ev)	FA(Ev)	FR(Ts)	FA(Ts)
.2	0	0.065	6.75	.1676
.1	0	0	4.50	.2174
.05	0	0	3.25	.3668
.01	0	0	1.25	.6495
.005	0	0	1.25	.7038
.001	0	0	1.25	.7482
combining	0	0	1.25	.6603

Table 2: Result of verification when the clients in the evaluation and training sets are used as seeds

To demonstrate the effectiveness of ECOC we report in Table 3 the result of applying the exhaustive search method directly to the original 199 dimensional feature vectors. Comparing Tables 2 and 3, 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.

search step	FR(Ev)	FA (Ev)	FR(Ts)	FA(Ts)
.25	1.67	0.89	16.75	1.105
.2	0.83	1.07	15.25	1.144
.01	0.167	0.33	8.0	1.180
.005	0.167	0.31	8.0	1.239
.001	0.167	0.2925	8.0	1.310

Table 3: Result of verification in the fisher face features space

4.2.2 Kernel combination

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 5.1. 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, along with the results of combining by majority vote are shown in Table 4. In comparison with Table 2, 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	0.7500	0.8833
50	0	0	0.5000	0.8786
100	0	0	0.7500	1.2455
combining	0	0	0.7500	0.8596

Table 4: Result of verification using the kernel score with different numbers of centres for the impostors

4.3 Comparison with other methods

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 5.

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

The results show a number of interesting features. First of all, by comparing the Euclidean metric performance with the proposed distance $d_i(\underline{y})$ in Table 3 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. Most importantly, all the ECOC based results are decisively superior to the decision making in the original Fisher face space. Finally, the combination of ECOC multiple classifier outputs by means of the relative similarity score in (12) appears to yield slightly better results than using the distance based score $d_i(\underline{y})$. The implication of this finding and of the work reported elsewhere is that the choice of decision (score) function plays an extremely important role in the design of verification systems and should receive more attention in the future.

5 Conclusion

We described a novel approach to face identification and verification based on the Error Correcting Output Coding (ECOC) classifier design concept. In the proposed method, during the training phase the client set is repeatedly divided into two ECOC specified sub-sets (super-classes) to train a set of binary classifiers. The output of the classifiers defines the ECOC feature space, in which it is easier to separate transformed patterns representing clients and impostors. For the matching score computation we investigated two alternatives. Either we use the first order Minkowski distance in the ECOC feature space or a kernel based similarity measure. We argued that it was more effective to use the actual rather than the target responses of the training samples to represent each client and devised a novel threshold selection method to separate authentic probe images from impostors. The proposed method was shown to exhibit superior verification performance on the well known XM2VTS data set as compared with previously reported results.

Acknowledgements

References

[1] P N Belhumeur, J P Hespanha, and D J Kriegman. 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, J Luettin, K Jonsson, J Matas, and J Kittler. Audio-visual person verification. In *Computer Vision and Pattern Recognition*, pages 580–585, Los Alamitos, California, June 1999. IEEE Computer Society.

[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. pages 572–577. Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), 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. pages 203–206, Barcelona,Spain, September 2000. International Conference on Pattern Recognition(ICPR2000).

[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. Banff,Canada, 1997. Int. Conf. on Artificial Intelligence and soft computing. 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 Luettin 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. Kotroupolous, 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. Luettin, 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 Applications of Computer Vision*, Sarasota,Florida, 1994. http://mambo.ucsc.edu/psl/olivetti.html.

[19] T.J. Sejnowski and C.R. Rosenberg. Parallel networks that learn to pronounce english text. *Complex systems*, 1(1):145–168, 1987.

[20] L. Sirovich and M. Kirby. Low-dimensitional procedure for the characterization of human face. *Journal .Opt.Soc. Am. A*, 3(4):519–524, 1987.

[21] T. Windeatt and R. Ghaderi. Binary codes for multi-class decision combining. volume 4051, pages 23–34, Florida,USA, April 2000. 14th Annual International Conference of Society of Photo-Optical Instrumentation Engineers (SPIE).