Generalization in Holistic versus Analytic Processing of Faces

M. Bicego*
DEIR - University of Sassari,
via Torre Tonda, 34
07100, Sassari (Italy)

A.A. Salah
Centrum voor Wiskunde en Informatica,
PNA4 - Signals and Images,
94079, Amsterdam (The Netherlands)

E. Grosso
DEIR - University of Sassari,
via Torre Tonda, 34
07100, Sassari (Italy)

M. Tistarelli
DAP - University of Sassari,
piazza Duomo, 6
07041, Alghero (Italy)

L. Akarun
Perceptual Intelligence Lab.
Boğaziçi University,
34342 Bebek, İstanbul (Turkey)

Abstract

The distinction between holistic and analytical (or feature-based) approaches to face recognition is widely held to be an important dimension of face recognition research. Holistic techniques analyze the whole face in order to recognize a subject, whereas analytical methodologies are devoted to the processing of different local parts of the face. This paper proposes a principled experimental comparison between these two approaches. Local and global face processing architectures that have access to similar feature representations and classifiers are implemented and tested under the same training and testing conditions. The analysis is performed with a recognition scenario on the difficult BANCA dataset, containing images acquired in degraded and adverse conditions. Different classifiers of increasing complexity are used in each scenario, and different classifier fusion methods are used for combining the local classifers. Our results show that holistic approaches perform accurately only with complex classifiers, whereas feature-based approaches work better with simple classifiers. We were able to show a clear boosting effect by fusing a large number of simple classifiers.

1. Introduction

Automatic analysis of face images is an active research area, for both scientific and industrial reasons. A plethora of techniques have been proposed in the past years to tackle the difficult problems of face recognition (for a review see [24]). These techniques can be grouped according to different criteria, each taxonomy serving to highlight a different aspect of the problem.

Perhaps the most fundamental distinction is made between holistic and feature-based approaches. In the former approach, also called *global, configural, relational,* and *monolithic* faces are perceived as units, and the extracted features pertain to the entire face. The most important algorithm proposed for holistic face recognition is the eigenface methodology [20]. The second class of approaches are called feature-based, but also *analytic, local, piecemeal, part-based, componential* and *fine grained*. In these approaches, fragments of the face images are analysed, and integrated. Typically, holistic approaches contain suitable transformations to subspaces of managable dimensionality, which is important when dealing with large databases. Feature-based methods are less sensitive to variations in illumination and viewpoint, but a part of the face may not contain sufficient information for identification purposes. These also require precise and reliable feature localization. For a review of the most important methods in both classes, see [24].

From a human perceptual point of view, both types of processing are relevant, but the issue of their relative contributions to human face recognition is an open problem. In an early study demonstrating the so-called *inversion effect*, Yin showed that inverting faces impairs their recognition far more than it would impair the recognition of any other object [23]. Since the individual features of the face are not affected too much by inversion, it is argued that inversion disrupts the configural processing, which is of special importance in human face recognition. There are other findings that support this claim; features of the face are much easier recognized in the presence of other facial features (*face superiority effect*) [19], and inversion of isolated features affects their recognition rates relatively less [14]. These studies suggest that human face recognition is a holistic process [19].

However, other experimental settings demonstrate that feature-based processing co-exists with holistic processing. In [22], an attempt has been made to single out the contributions from the holistic and feature-based components of human facial image processing. The configural processing breaks down when the facial parts are scrambled, and the feature based system breaks down when faces are blurred. Ullman *et al.* note that small fragments of images corresponding to simple features processed in the early to intermediate stages of the visual system can be used for object detection in general, and face detection in particular [21].

Comparing feature-based and holistic approaches in the context of machine face recognition is therefore very interesting. Although it may be possible to group existing results (especially from recent face recognition competitions such as FRGC, FRVT, BANCA) according to their processing type and say something about the relative success of these paradigms, the issue deserves a principled inspection. The goal of this paper is to provide such a comparison, by implementing local and global processing architectures that have access to similar feature representations and classifiers, under the same training and testing conditions.

Local and global features were previously compared for face recognition in [9], where the main classifier was a Support Vector Machine. However, the choice of the classifier has a bearing on the issue. In the present paper, we test several classifiers of different complexity with local and global features. The holistic approach attempts to analyse the face as a whole, while the analytic approach fuses results obtained from different local experts, each performing independent classification by looking at a particular region of the face.

The analysis is performed on the difficult BANCA dataset [2], which contains images from 52 subjects acquired in 12 sessions. The sessions are grouped into controlled, degraded and adverse conditions. The graded and adverse conditions provide an excellent way to determine the generalization performance of a particular system. Our results show that the choice of the classifier is relevant to the choice between analytic and holistic processing. Local approaches are better with simple classifiers, which are less prone to overlearning. We were able to obtain a clear boosting effect by employing a large number of simple classifiers. Global approaches perform accurately only with complex classifiers that can match the increased complexity of the features. It is difficult to balance data complexity and model complexity, consequently the analytic approaches seem to be more promising.

2. Methodologies

In this section we present two architectures for comparing holistic and analytic (feature-based) processing of faces. The feature-based architecture is described in Section 2.1, and the holistic architecture is described in Section 2.2, respectively. The different classifiers we use in each method are summarized in Section 2.3.

2.1. Feature-based architecture

The idea under the proposed feature-based architecture is to create a set of local experts, each one devoted to a particular zone of the face. In the training phase, the face is sampled in order to obtain a set of patches. Each local expert is then trained with the patches that belong to its zone, with the aim of discriminating between subjects based on that zone. In the testing phase, the unknown face is sampled in a similar manner, and each patch is fed to the nearest local expert. The number of patches determines the number of local classification results for a particular face, which are subsequently fused to obtain the final decision.

The first step of this method is the extraction of a set of patches. There are many possibilities: picking them up at random positions (like in patch-based image classification [8, 5, 4] or image characterization [12]), or extracting them at relevant points, such as edges [3] or ridge and valleys [1]; the technique adopted here follows [15, 16], in which the face image is regularly sampled, extracting a set of partially overlapped sub-windows positioned at the nodes of a regular grid. The size of the patch is related to the size of the image.

The second step is to assign each patch to a particular local expert. The number of experts determines a trade-off between complexity and robustness, but also changes the representation for each expert. With a small number of experts, the input to the classifier comes from a broader area, necessarily containing more variation. Based on preliminary studies, we have selected a grid of 16 points as the hypothetical locations of local experts, and used the Euclidean distance to the experts in deciding which expert is responsible from the classification of each patch. We are aware that there are many other alternatives to define the local expert positions. For example we can link them to precise facial landmark positions, or we can extract salient points on the training set and then perform a clustering. The advantage of a regular grid is that we have uniform coverage, and similar training conditions for all local experts.

Feature extraction is a far more important issue in this architecture, as there are many possible candidate features tested in the literature. We have examined simple gray level features, Gabor wavelet coefficients, DCT coefficients, and their feature-level combinations. A concatenation of gray level features and the most important DCT coefficients is selected as the more robust scheme [16].

In the training phase, one classifier is trained for each local expert, using all the training patches belonging to that expert. We summarize the classifiers we have used in the next section. The classifier is trained in a discriminative manner, solving the full n-class problem. Thus, each classifier attempts to decide on the identity of the subject from a single patch, producing a set of posterior probabilities. During testing, the maximum a posteriori (MAP) criterion is used to select the class of the test image. Since the posteriors are at our disposal, fusion is possible at the score level or at the decision level:

- Score level: The fused classifier combines the confidence values (also called matching scores). We have applied the SUM, MAX and PROD rules [13].

- Decision level: In this case the fusion result is obtained by combining the decisions of each classifier. We have applied simple voting and Borda count methods [10]. In the ordinary Borda count, the decision of each local expert is used to rank the classes, and each class receives points from 1 to N_C (number of classes) from each local expert, depending on their rank. The highest rank class receives N_C points, the next class receives $N_C - 1$ points, up to the last class, receiving 1 points. The class with the highest final score is retained. In a modified version of the Borda count, the scores are proportional to $1/r$, where r is the rank.

  With these methods, poor classes that do not have competitive scores still influence the ranking. In this paper we introduce a more intuitive scheme, where these classes are trimmed from the scoring. In what we call the *trimmed Borda count*, classes that have not received a rank-1 result with any of the local classifiers are trimmed. We have shown that this procedure leads to better results with sufficient number of classifiers. The exact number of classifiers for which this method produces better results than the ordinary Borda count depends on the application, but our results show that the trimmed version is better for a larger range of models. For smaller number of experts, this method can also be implemented by trimming the worst classes after ordinary Borda count, and performing a second scoring.

2.2. Holistic architecture

With this architecture the whole face is used for the analysis. In particular the same type of features used for the local analysis are extracted, namely DCT and graylevel values. In order to alleviate the curse of dimensionality, a principal component analysis (PCA) is applied to the gray level features, and the projected features are concatenated with the most important DCT coefficients as in the analytical architecture. When we consider the overlapping input to the local classifiers, the total amount of input features fed to the classifier architecture is smaller in the holistic architecture. However, the amount of features per classifier is larger. The input dimensionality works two ways, on the one hand, it allows more complex inferences that can draw on more features, with the promise of better classification. On the other hand, the increased model complexity necessitates a proportionally larger training set, making learning and generalization more difficult.

2.3. Classifiers

With these two architectures, the problem is cast in a classification context, where any classifier could be used. Since we are interested in isolating the effects of using feature-based and holistic approaches, we will use the same classifiers in both architectures. We have selected several classifiers of different complexity [11, 6, 7]:

- Logistic Linear Classifier (LOGLC): This classifier computes a linear discriminant between classes of the dataset by maximizing the likelihood criterion using the logistic (sigmoid) function.

- Linear Bayes Normal Classifier (LBNC): In this method a linear classifier is implemented, where the distribution of the data items falling under a class is assumed to be isonormal Gaussian distributed. A joint covariance matrix is computed by taking a weighted (by a priori probabilities) average of the class covariance matrices. The a priori probabilities are approximated with the item frequencies of each class in the training dataset.

- Perceptron (PERC): In this case the classes are discriminated with a hyperplane, which is iteratively updated as a function of the distances of the misclassified patterns from it. A crucial parameter is the learning rate, indicating how much impact the estimation error should have on the update of parameters.

- Multi Layer Perceptron (MLP): We have used a feed forward neural network, with a single hidden layer. With an MLP, we can fit non-linear decision boundaries to the dataset. In this context, MLP is a complex classifier, as the face recognition problem is high-dimensional.

- Fisher Linear Classifier (FISHER): This classifier finds the linear discriminant function between the classes by minimizing the errors in the least square sense. Since we have a multi-class problem, the Fisher discriminants are found for each class, separating the samples of the class from all the other samples.

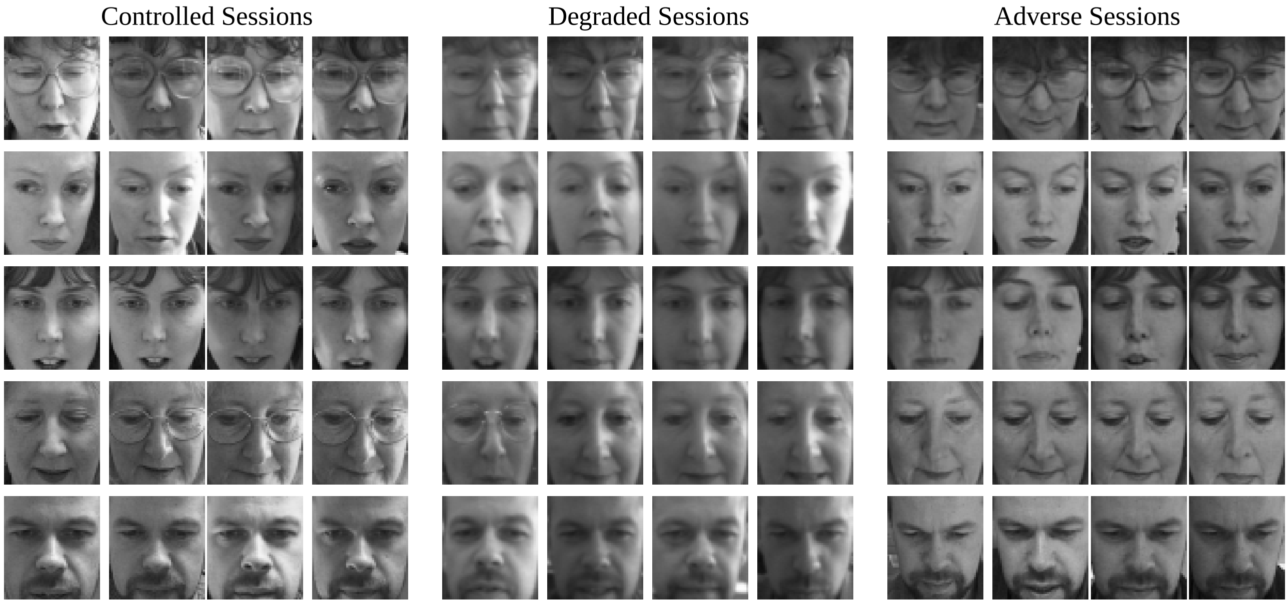

Figure 1. Examples from the BANCA Database: for each subject, one example for each session is shown. The sessions are divided in three classes: controlled sessions, degraded sessions and adverse sessions.

3. Experimental Evaluation

3.1. The database and the experimental protocol

The experimental evaluation was performed using the cropped faces from the English version of the BANCA dataset [2]. This database contains 52 subjects (26 female and 26 male). For each subject, 12 different sessions recorded under different conditions are available (four controlled, four degraded, and four adverse). Some examples are presented in Fig. 1, where one image from each of the 12 sessions of five different subjects is displayed. It is apparent from these samples that especially for the adverse conditions, the recognition is quite challenging.

The BANCA protocol, described in [2], defines different configurations for testing, all nevertheless thought for the authentication scenario, where the test image comes with a claim, and is tested against the gallery image(s) of a single class. Here we are interested in the recognition scenario, which is a more difficult problem, in that one needs to compare the test image against all the gallery images. Consequently, we define a different protocol: The training is performed using images from the first three sessions, whereas the test images come from the remaining sessions. This enables us to test the behaviour of the algorithms under controlled, degraded and adverse conditions with the largest possible training set size.

3.2. Results

We have tested each architecture with a range of parameter settings. Due to space constraints, some of our preliminary results are omitted here. The final presented results have the following parameters: In the feature-based approach, patches were of dimension 11×11, extracted with a 50 per cent horizontal and vertical overlap. We have used 10 coefficients for the DCT analysis. Regarding the holistic approaches, the PCA space was calculated using the training images and the BANCA World model, retaining the 99 per cent of the variance. Also in this case 10 coefficients were retained from the DCT transformation. Regarding the classifiers, the learning parameters (if applicable, e.g. the number of hidden units, learning rate, and convergence criteria in the neural network models) were optimized on a subset of the training set. The test set is only used after these parameters are determined once and for all. One third of the training dataset was used as a validation set to control the complexity of the trained model during the actual training.

The results derived from the application of the proposed approaches are shown in Table 1. In particular, results are divided by sessions, grouped after the test condition as “controlled,” “degraded,” and “adverse”. For the local schemes, only the best fusion result is displayed.

	Controlled		Degraded		Adverse	
Classifier	Local	Global	Local	Global	Local	Global
LOGLC	97.88%	100%	28.12%	59.81%	36.20%	60.24%
LBNC	97.12%	66.92%	31.39%	9.47%	36.20%	4.18%
PERC	99.62%	94.42%	67.50%	27.36%	53.22%	29.18%
MLP	91.92 %	95.19%	25.72%	52.64%	21.59%	44.52%
FISHER	99.62%	97.69%	62.74%	41.25%	60.43%	43.27%

Table 1. Recognition accuracies for different classifiers, with local and global architectures, for different testing conditions.

3.3. Discussion

Looking at our results, we have several observations.

- From the presented results, it is not possible to decide on a clear supremacy of either local or global architectures. The choice of classifier obviously has a bearing on this issue. For complex classifiers (such as Logarithmic Classifiers and Multi Layer Perceptrons), the global approach is better than the local approach. This means that the complex classifier is actually able to harness the holistic information, and the associations between different dimensions of the data. On the other hand, simple classifiers (such as the linear perceptron or Fisher's discriminant) are more suitable for local approaches. A simple classifier can only learn simpler, e.g. local associations, and the fusion succeeds most when the classifiers in the ensemble have different error patterns. Combination of complex local experts quickly leads to overtraining. This is consistent with the basic idea of the boosting approach [17, 18]. On the average, the local architectures show a slight supremacy over the global architectures. The LBNC classifier has a very poor generalization performance in both architectures, we are currently investigating the reasons.

- Looking at the best performers in the three class of sessions, we can notice that the feature-based approach with Fisher's linear classifier and the holistic approach with the logistic classifier are the best in the controlled and adverse sessions (with comparable accuracies). In the degraded scenario, the best technique is the local architecture with the perceptron classifier.

- Looking at the general classification rates, it is clear that the classification task is quite difficult. In order to have a baseline result, we have also computed the accuracy of the eigenface method with the same configurations. The accuracies were 95.77 per cent for the controlled, 40.57 per cent for the degraded, and 32.31 per cent for the adverse conditions, respectively.

- Regarding the local architectures, we have observed that the best results are produced by the trimmed version of the Borda count fusion scheme.

4. Conclusions

In this paper a principled experimental evaluation of holistic versus feature-based approaches for face recognition has been proposed. The aim of our evaluation was to estimate differences in the classification accuracies in a difficult datasets, constraining holistic and analytic approaches to use exactly the same classifier and the same feature representation. The results show that the performances strictly depend on the classifier complexity: holistic approaches work properly with complex methodologies that are able to properly model complicated features, whereas local methods are better with simple models, showing a clear boosting effect when fusing the results. A novel fusion scheme called the *trimmed Borda count* is proposed for score fusion, with which we have obtained the best results for the feature-based architecture.

5. Acknowledgements

This work is supported by funds from the 6^{th} Framework Programme European Network of Excellence BioSecure, by the Dutch BSIK/BRICKS project, and Boğaziçi University project BAP-03S106.

References

[1] D. Gonzalez Jimenez, and J. Alba-Castro. Shape contexts and gabor features for face description and authentication. In *Proc. IEEE Int. Conf. on Image Processing*, volume 2, pages 962–965, 2005.

[2] E. Bailly-Baillire, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Marithoz, J. Matas, K. Messer, V. Popovici, F. Pore, B. Ruiz, and J.-P. Thiran. The BANCA database and evaluation protocol. In *Proc. Int. Conf. on Audio- and Video-*

Based Biometric Person Authentication (AVBPA03), pages 625–638. Springer-Verlag, 2003.

[3] M. Bicego, G. Brelstaff, L. Brodo, E. Grosso, A. Lagorio, and M. Tistarelli. Distinctiveness of faces: a computational approach. *ACM Trans. on Applied Perception*, accepted, 2007.

[4] G. Csurka, C. Dance, C. Bray, L. Fan, and J. Willamowski. Visual categorization with bags of keypoints. In *Proc. Workshop Pattern Recognition and Machine Learning in Computer Vision*, 2004.

[5] G. Dorko and C. Schmid. Selection of scale-invariant parts for object class recognition. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 634–640, 2003.

[6] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, second edition, 2001.

[7] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. De Ridder, and D. Tax. Prtools4, a matlab toolbox for pattern recognition, 2004. Delft University of Technology.

[8] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, page 264, 2003.

[9] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*, 91:6–21, 2003.

[10] T. Ho, J. Hull, and S. Stihari. Decision combination in multiple classifier systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(1):66–75, 1994.

[11] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[12] N. Jojic, B. Frey, and A.Kannan. Epitomic analysis of appearance and shape. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 34–41, 2003.

[13] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[14] S. Rakover and B. Teucher. Facial inversion effects: Parts and whole relationship. *Perception & Psychophysics*, 59:752–761, 1997.

[15] A. Salah, E. Alpaydın, and L. Akarun. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(3):420–425, 2002.

[16] A. Salah, M. Bicego, L. Akarun, E. Grosso, and M. Tistarelli. Hidden markov model-based face recognition using selective attention. In *Proc. Int. Conf. on Human Vision and Electronic Imaging*, 2007.

[17] R. Schapire. The boosting approach to machine learning: An overview. In *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[18] R. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictors. *Machine Learning*, 37(3):297–336, 1999.

[19] J. Tanaka and M. Farah. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A:225–245, 1993.

[20] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[21] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5:682–687, 2002.

[22] C. Wallraven, A. Schwaninger, and H. Bülthoff. Learning from humans: computational modeling of face recognition. *Network: Computation in Neural Systems*, 16(4):401–418, 2005.

[23] R. Yin. Looking at upside-down faces. *Journal of Experimental Psychology*, 81:141–145, 1969.

[24] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399 – 458, 2003.