Human Identification versus Expression Classification via Bagging on Facial Asymmetry

Yanxi Liu and Sinjini Mitra

CMU-RI-TR-03-08

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

This research is supported in part by an ONR research grant N00014-00-1-0915 (HumanID), and in part by an NSF research grant IIS-0099597.

ABSTRACT

We demonstrate a dual usage of quantified facial asymmetry for (1) human identification under expression variations and (2) expression classification across different human subjects. Our experiments show the effectiveness of using statistical bagging and feature subspace selection BEFORE applying classifiers such as Linear Discriminant Analysis. This preprocessing allows the same type but different dimensions of image features to be discriminative for two seemingly conflicting classification goals. Statistically significant improvements are found when facial asymmetry features are combined into classical classifiers.

1 Introduction

For almost all classification problems, one common concern in practice is whether there exists enough data for a classifier to draw meaningful conclusions. A large number of papers is focused on this issue (e.g. lately [8, 14]). Still, there is a lack of theoretical and practical guidance on deciding when a data set is sufficient for a specific classification problem at hand. The sufficiency of a dataset size is often discussed in the context of sample size versus the data dimension or feature dimension. The problem is that there could be some extra dimensions in the data to make the dataset unnecessarily sparse. For discrimination problems it is desirable to reduce irrelevant and redundant feature dimensions with little effects on class separation in the original data. Another concern of space reduction for many classification tasks is to land in a subspace composed of a subset of the original features for interpretation purposes. Therefore methods like principal component analysis (PCA) which minimizes reconstruction errors and combine different features are not appropriate.

Different from PCA, linear discriminant analysis (LDA) is known for its intended capability of capturing the discriminative subspace from a given labeled, multi-dimensional dataset [6]. However, LDA has its own limitations for optimal performance. LDA does not perform well on small sample sets [14] or data sets with non-Gaussian distributions. LDA is also sensitive to the initial input features, since irrelevant and redundant features will affect the orientation of the projected subspace. Sometimes feature independence assumption is used explicitly or implicitly for LDA applications [1, 14, 17]. These previous work suggests to us that a feature selection step before using classifiers like LDA should be helpful to (1) alleviate the data insufficiency problem, and (2) improve classifier performance (as a result of the curse of dimensionality).

In this work, we are interested in finding out how to effectively apply LDA when a given classification problem has small sample sizes with respect to its feature dimensions and its initial features have redundant even irrelevant dimensions. We use a combination of Bagging technique [3] with a feature subset sequential forward selection algorithm [19, 2] to achieve, simultaneously, the goals of minimizing the variance of unstable classifiers and ruling out irrelevant/redundant feature dimensions. To test out hypothesis, we apply our framework on a set of weak facial features, facial asymmetry measures in a two-fold classification problem: (1) human identification under expression variations; and (2) expression identification across human subjects.

These two classification problems are seemingly conflicting with each other in the sense that for those image features that can do well on (1), they must be insensitive to expression variations thus they would not be indicative for expressions, and therefore they are not good for classification problem (2), and vice versa. We shall show in this paper that quantified facial asymmetry can perform well on both types of classifications defined above. This is because a bagging method is used to help unstable classifiers, and our feature subspace selection approach can help to identify facial asymmetry measures in specific locations (dimensions) that provide discriminative performance respectively for either classification goals. FisherFaces [1], an LDA application on eigen values of face images, will be used as a baseline classifier for comparison.

Section 2 gives an explanation of the quantified facial asymmetry dataset used for this work. Section 3 describes our hybrid method using bagging, sequential forward feature selection and LDA for image discrimination. Sections 4 and 5 report the experimental results of the two different classification problems respectively. Section 6 discusses the findings and concludes the paper.

2 Quantification of Facial Asymmetry

Previous work has shown that facial asymmetry information is useful for human identification by human [18] as well as by machines [11, 12, 14]. While in other related fields, computer graphics for instance, it is a commonly accepted assumption that human faces are bilaterally symmetrical [15, 20]. For human identification, half faces are used [13, 7]. Some [13] reported difference in recognition rates for left and right face images, while others [7] did not.

For a given normalized face density image I (affinely deformed using three feature points: two inner eye corners and the mid point of the bottom of the nose), its vertically reflected image I', and their respective "edged" images I_e, I'_e (applying edge extraction algorithm on I, I'), using the facial asymmetry measures defined in [11, 12],

Density Difference *D-face*:

$$D(x,y) = I(x,y) - I'(x,y) \tag{1}$$

Edge Orientation Similarity *S-face*:

$$S(x,y) = \cos(\phi_{I_e(x,y),I'_e(x,y)}) \tag{2}$$

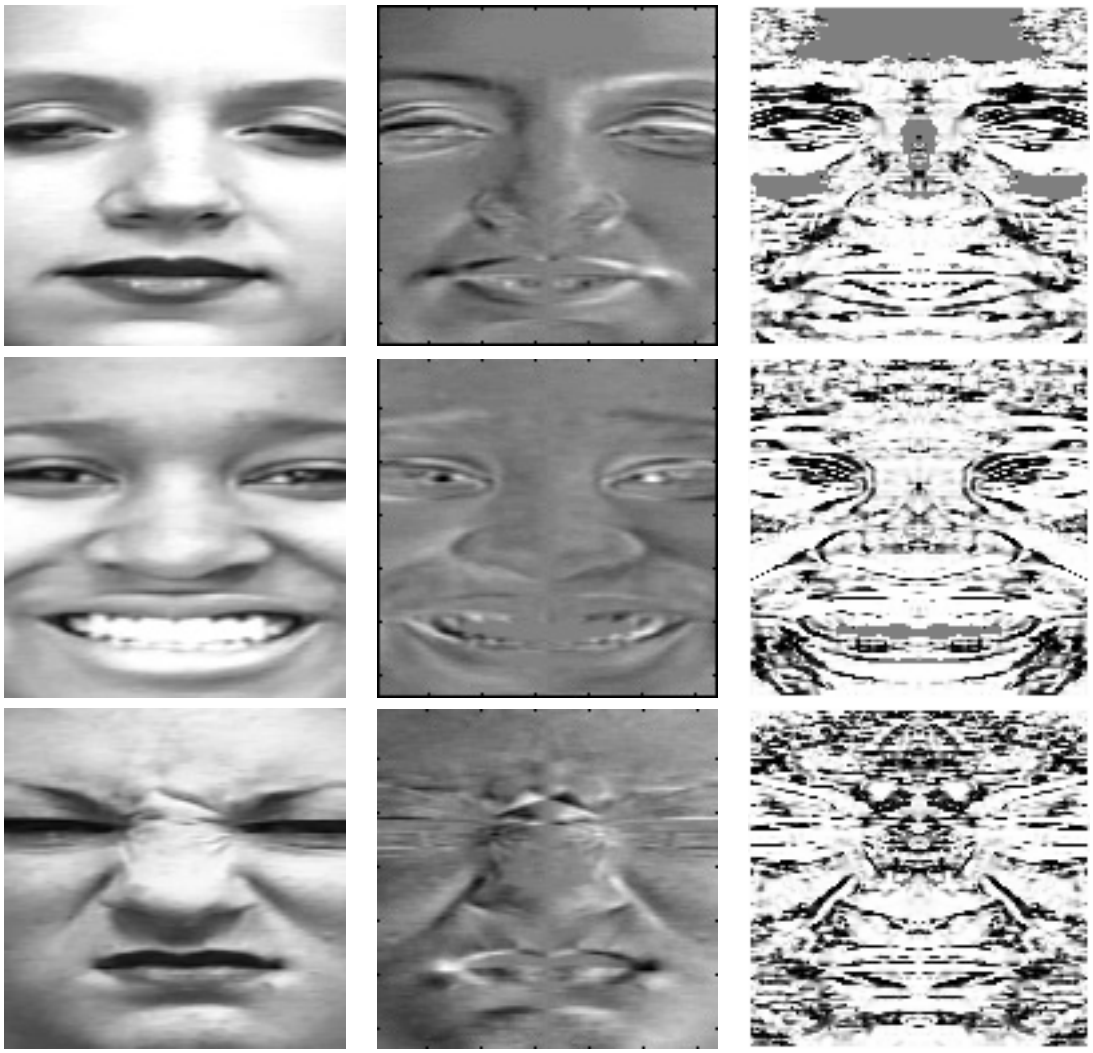
where $\phi_{I_e(x,y),I'_e(x,y)}$ is the angle between the two edge orientations of images I_e, I'_e at the same pixel point x, y. Figure 1 displays three normalized faces, and their respective *D-face* and *S-face*.

Half of the *D-* and *S-faces* contains all the information needed. PCA is applied here to reduce the feature dimensions of *D-face* and *S-face*. Table 1 contains the definition of six sets of quantified facial asymmetry measures, we call them *AsymmetryFaces*.

Table 1: Different AsymmetryFaces are defined. The Datasets used for face recognition experiments contain 55 subjects, each has three expressions, three frames (beginning, middle and ending) taken out of each expression sequence (total 495 frames).

Feature Name	**Feature Description**	**Feature Dimension**
D	top k PCs accounting for 95% variance of *D-face*	60
D_x	column-mean of *D-face*	64
D_y	row-mean of *D-face*	128
S	top k PCs accounting for 95% variance of *S-face*	100
S_x	column-mean of *S-face*	64
S_y	row-mean of *S-face*	128

For the purpose of isolating the intrinsic from extrinsic factors on facial asymmetry, the Cohn-Kanade AU-Coded Facial Expression Database [9] is used. This dataset consists of frontal view expression video sequences. The video is taken under controlled bilaterally balanced lighting. Each frame is a grey-scale image of 640x480 pixels. Each subject has 3 expression video sequences (joy, anger and disgust). 55 subjects



(A) normed Face (B) D-face (C) F-face

Figure 1: *(A) normalized face , (B)* D-face*, and (C)* F-face*.*

used in this work differ in race and gender (Figure 2). This dataset is used for five different experiments (Table 2). It turned out that the results of training on negative expressions (anger and disgust) and testing on a positive expression (joy) face (experiment #1 above) gives the worst result of all, especially when compared with training on peak expression and testing on neutral faces (and vice versa).

Table 2: Five experiment setups for Human Identification of 55 Subjects under Expression Variations

	Test Image Expressions	**Train Image Expressions**
1	**joy**	**anger and disgust**
2	anger	disgust and joy
3	disgust	joy and anger
4	neutral	peak
5	peak	neutral

In this work we use the same data set and asymmetry measures as in [12]. There are several major differences: (1) in this work we are collectively using all six facial asymmetry measures defined in Table 1 instead of a single one out of the six in previous work. (2) only three frames from each expression video sequence are selected in this work instead of all the frames used in previous work. The reason for the selective use of frames is to create a small sample size on purposely, to test the effectiveness of our classification framework. (3) the method and the nature of the classifications differ. The hybrid method we develop here is general and the goal is to achieve robust classification performance on small sampled data with redundant dimension.

3 Bagging + Feature Subset Selection for LDA

There are two important independent dimensions in a typical classification problem: the sample size n and the feature dimension d. Much attention has been paid to the relative relation between these two numbers n, d [8, 14] and various re-sampling methods [16, 3], such that even an initially ill-posed classification problem can achieve meaningful results. A less addressed issue is how to deal with redundancy in the feature dimensions at the same time as the samples are being "reproduced". In this work, we are interested in modifying both sample and feature dimensions for achieving the final goal of enhanced discriminating power of a liner discriminant analysis (LDA) classifier. We use a combination of Bagging technique [3] with a feature subset sequential forward selection algorithm [19, 2] to achieve, simultaneously, the goals of minimizing the variance of unstable classifiers and ruling out irrelevant/redundant feature dimensions.

3.1 Relevance of Bagging

Suppose we have a learning (training) data set Ł which consists of data of the form $\{(y_n, \mathbf{x}_n)|n = 1, \ldots, N\}$ where the y's are either class labels or a numerical response and N is the total number of observations in the learning set. The goal is to: for a given input $\mathbf{x}$, predict the corresponding y using a learned predictor $\varphi(\mathbf{x}, \text{Ł})$. If a sequence of learning sets $\{\text{Ł}_k\}$ can be given, each consisting of N independent observations from

Figure 2: Normalized faces from Cohn-Kanade AU-Coded Facial Expression Database [9]. Each column represents one subject (total of 7 subjects displayed) with neutral, peak joy, peak disgust and peak anger expressions in video sequences, respectively. (borrowed from [11, 12])

the same underlying distribution as Ł, one can then use these $\{Ł_k\}$'s to obtain a better predictor than the one computed from a single set of observations, $\varphi(\mathbf{x}, Ł)$. Let the predictor based on a particular $Ł_k$ be denoted by $\varphi(\mathbf{x}, Ł_k)$, the idea is to combine all the $\varphi(\mathbf{x}, Ł_k)$s. However, in practice there may be only one specific learning set (not multiple ones). In such a situation, we can generate repeated bootstrap samples $\{Ł^{(B)}\}$ from Ł (samples of size N randomly drawn from Ł *with replacement*) and form a sequence of predictors $\varphi(\mathbf{x}, Ł^{(B)})$. If y is a class label, a voting technique can be applied using the predictors $\varphi(\mathbf{x}, Ł^{(B)})$.This procedure is known as "**b**ootstrap **agg**regat**ing**" or "**Bagging**" for short [3]. Bagging averages its prediction error over a collection of bootstrap samples, thereby reducing the variance. This is one of the main reasons why bagging can help *unstable* classifiers positively. It has been empirically established that although bagging itself is not a stabilizing technique, it does improve the performance of unstable classifiers. However, the success of bagging depends on many factors, including the training sample size, the choice of the base classifier, the exact way how the training set is re-sampled, the choice of the combining rule, and the very nature of the data itself. Bagging, boosting and Random Subspace Method belong to the same type of weak classifier combination approaches [16].

"Unstable classifiers" are referring to those classifiers that are sensitive to small perturbations in the training data, such that large variance can occur in their output predictions with small changes in their input. This happens when their training sets are in *critical conditions*, for example when the number of features exceeds the number of training samples, the sample covariance matrix S (which is taken as the estimate of the actual one) becomes "ill-conditioned" (nearly singular) and that makes computation unstable. Many if not all real world computer vision and pattern recognition problems tend to have "unsatisfactory"

or critical conditioned training sets where feature dimensions are quite large compared with sample sizes. Bagging, though a popular method in statistics for regression problems [3], has been rarely used in computer vision and pattern recognition community [5].

3.2 Bagging + SFS + LDA Algorithm

Here we describe our classification procedure:
Input: Training set T of size $N \times D$, Testing set S
Output: False negative rate (FNR), false positive rate (FPR) and error variances.

1. N samples are drawn from T with replacement to form a new training set T_i';
2. compute and order each feature of T_i' in non-increasing order of its augmented variance ratio (AVR)[1] value;
3. apply sequential forward feature subset selection on the ordered features such that the remaining feature dimensions becomes $D' < D$, and the training set T_i' of size $N \times D$ is reduced to size $N \times D'$;
4. apply LDA on newly reduced training set T_i', then test on S in the projected space, record the test results;
5. repeat steps 1-4 N_i (a user defined parameter) times;
6. perform a classification using majority vote on N_i test classification results,
7. repeat steps 1-6 B times, average the final errors for all the N_is;
8. increase the value of N_i to N_{i+1} (a user defined parameter) and repeat steps 1-7 above, stop when N_i is the maximum number of bootstrapings;

In our experiments, we choose $N_i = 10, 20, 30, 40, 50, 60, 70, 100$ and $B = 20$.

4 Human Identification under Expression Variations

As shown in Table 2, we have set up five different experiments to test the discriminating power of facial asymmetry. For each experiment setup, we use FisherFaces alone, AsymmetryFaces alone and FisherFaces + AsymmetryFaces. Without bagging, we have demonstrated in our previous work that with the addition of facial asymmetry information the classification errors for experiments 1-4 can be reduced nearly to zero. However, error rate for training on Anger and Disgust expressions and testing on Joy seem to be quite persistent even when we extend the PCA components of the face eigen vectors for FisherFaces to 400 dimensions (Figure 3). Therefore, we are going to study whether bagging can help in this case.

Table 3 summarizes the classification results of human identification using the algorithm described in Section 3.2. The improvements on variances can be observed as expected. We noticed an appreciable improvement in the results. Indeed, paired-t tests [4] comparing the bagging results from the FisherFaces and that of combining FisherFaces with all the AsymmetryFaces (sampling 100 times) revealed very strong

[1] AVR is the ratio of the variance of the feature between subjects to the variance of the feature within subjects, with an added penalty for features which may have small intra-class variance but which have close inter-subject mean values.

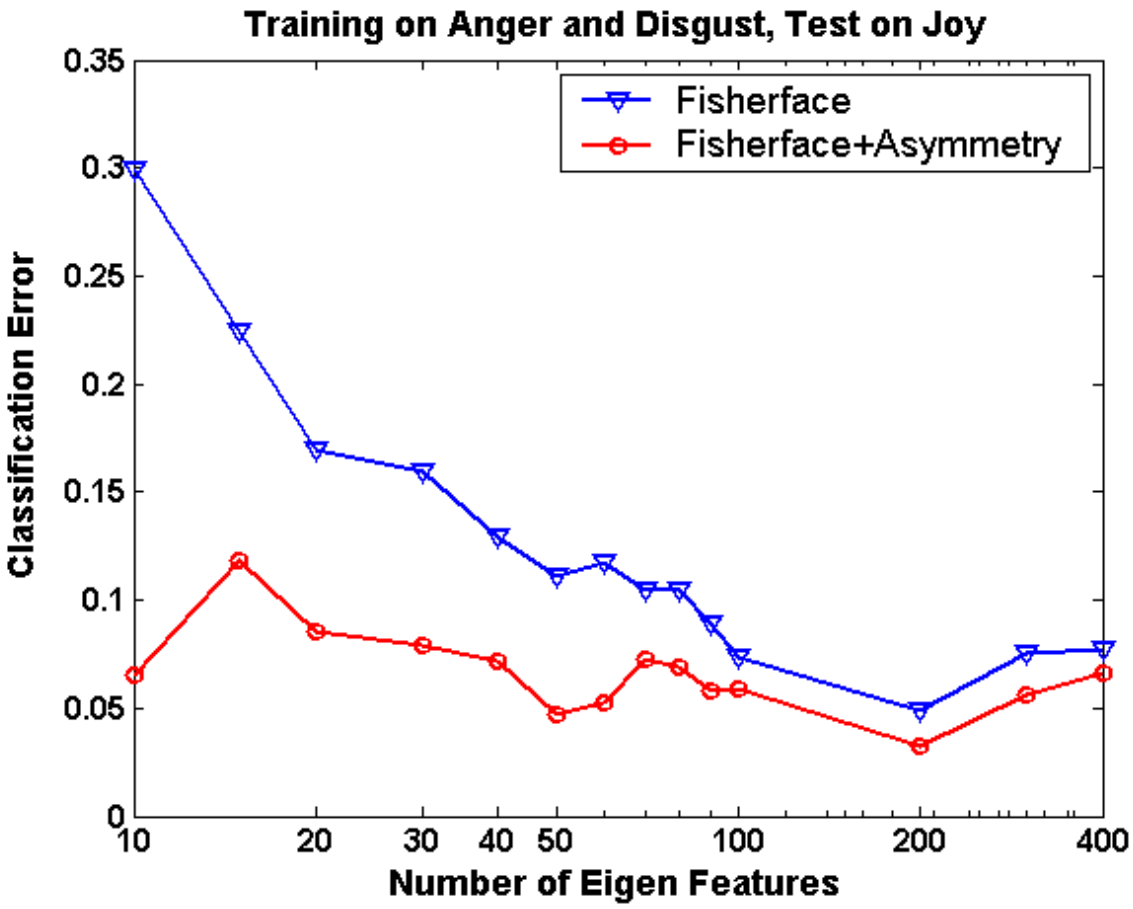
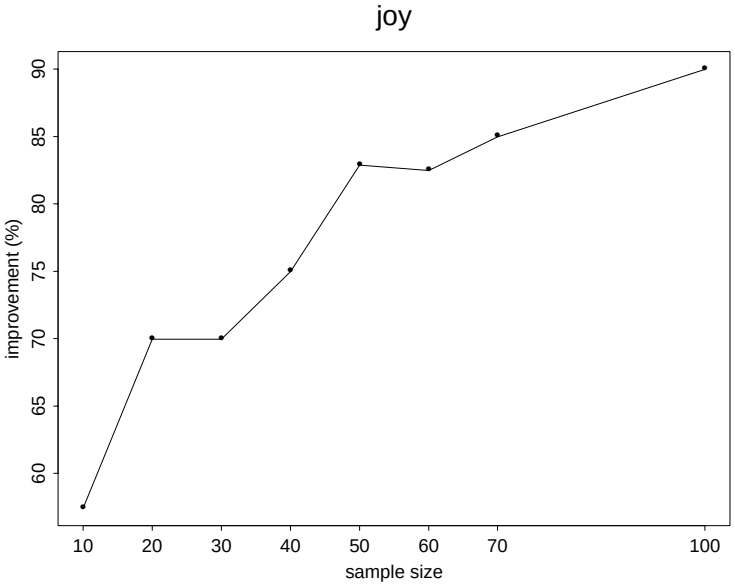


Figure 3: The combination of AsymmetryFaces with FisherFace (16 selected AsymmetryFace features plus FisherFace features) shows superior performance compared with FisherFace method alone even when extending the PCA components in FisherFace method to 400 dimensions.

Figure 4: Human Identification: The % improvement in the average FNR for testing on joy frames for 8 different times of bootstrap. These are computed as (original error-bagging error)/original error ×100 (averaged over 20 repetitions for each).



evidence of a statistically significant improvement in the error rate for testing on joy (p-value of 0). Figures 4 the increase of the improvement rates with the increasing of the number of bootstrapings.

5 Expression Identification Across Human Subjects

There are three expression classes: joy, anger and disgust (Figure 2). The training was done on all the peak frames from a randomly selected subset of 30 (out of 55) individuals, ie, 90 frames and testing was done on all the peak frames of the remaining 25 individuals, ie, on 75 frames. This random division of the total number of subjects into training and testing sets was repeated 20 times and the final results (FPRs, FNRs) were obtained by averaging over these 20 repetitions.

Table 4 shows the results of expression identification using FisherFaces alone, and FisherFaces + AsymmetryFaces. The results show that the class "joy" has both a lower FNR and a lower FPR than the other two classes which have their FPR and FNR values fairly close to one another (except the case for D_Y alone). This is something that we might expect and can be due to the fact that "anger" and "disgust" being both negative expressions can be more similar to each other than to "joy" (which is a positive expression). Also, we found that the S-face features, both the projections and the Principal Components, give better results than the corresponding D-face features, just contrary to what we found in the case of subject classifications.

Here, we also carried out combinations of feature sets, and as in the case with subject classifications, we found that the misclassification rates dropped considerably as a result of this, particularly, for the whole seven feature combinations (bottom of Table 4).

All the expression classification results using the AsymmetryFaces were significantly better than chance. Further, there is ample evidence of significant statistical difference between the results from the FisherFaces

alone and those from FisherFace+ AsymmetryFaces, a p-value of nearly 0 was obtained from the paired-t test.

6 Discussion and Conclusion

Our experimental results show that, with a novel combination of bagging method and careful feature subspace selection (note the consistency between AVR and LDA), relatively weak image features like quantified facial asymmetry features can perform well in both (1) human identification under expression variations and (2) expression identification across subjects. Figures 5, 6, 7 and 8 may provide some insight on why this is the case. These two figures show the AVR plots of facial asymmetry features under the two different classification problems. One can easily observe that the discriminative features for each type of classification is almost mutually exclusive, that is just the opposite of each other. Thus for each classification, different feature subspace is used. Therefore, there is no contradiction for the same type of features, but under different discriminative subspaces, to achieve satisfactory classification results for seemingly conflicting classification goals.

Bagging is shown to be an appropriate method for the two classification problems in this study. We have proven, using statistical analysis, that the improvements over FisherFaces using our approach is statistically significant. Bagging + Feature Selection for LDA can be very effective for small sample classification problems with redundant features. Further study is needed to explore the sufficient discriminative subspace for a given classification problem.

References

[1] P.N. Belhumeur, J.P. Jespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711–720, July 1997.

[2] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995. ISBN:0198538499.

[3] L. Breiman. Bagging predictors. *Machine Learning Journal*, 24(2):123–140, 1996.

[4] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press, Belmont, California, 1990.

[5] B. Draper and K. Baek. Baggin in computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 144–149, Santa Barbara CA., June 1998.

[6] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2001.

[7] S. Gutta, V. Philomin, and M. Trajkovic. An investigation into the use of partial-faces for face recognition. In *International Conference on Automatic Face and Gesture Recognition*, pages 33,38, Waxhington, D.C., May 2002. IEEE Computer Society.

[8] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(4):399–404, 1997.

[9] T. Kanade, J.F. Cohn, and Y.L. Tian. Comprehensive database for facial expression analysis. In *4th IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, March 1999. Publically available at http://www.ri.cmu.edu/projects/project_420.html.

[10] Y. Liu and S. Mitra. Experiments with quantified facial asymmetry for human identification. Technical Report CMU-RI-TR-02-24, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2002.

[11] Y. Liu, K.L. Schmidt, J.F. Cohn, and S. Mitra. Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Pattern Recognition Journal: special issue on Face Recognition*, to appear 2003.

[12] Y. Liu, K.L. Schmidt, J.F. Cohn, and R.L. Weaver. Facial asymmetry quantification for expression invariant human identification. In *International Conference on Automatic Face and Gesture Recognition*, May 2002.

[13] A.M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern analysis and machine intelligence*, 24(6):748–763, 2002.

[14] A.M. Martinez and A.C. Kak. PCA versus LDA. *IEEE Transactions on Pattern analysis and machine intelligence*, 23(2):228–233, 2001.

[15] S.M. Seitz and C.R. Dyer. View morphing. *SIGGRAPH*, pages 21–30, 1996.

[16] M Skurichina and R.P.W. Duin. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis and Applications*, 5(2):121–135, 2002.

[17] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831,836, August 1996.

[18] N.F. Troje and H.H. Buelthoff. How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research*, 38(1):79–89, 1998.

[19] S. Weisberg. *Applied Linear Regression*. Wiley, 1985.

[20] W.Y. Zhao and R. Chellappa. Symmetric shape-from-shading using self-ratio image. *IJCV*, 45(1):55–75, October 2001.

Table 3: This table demonstrates the test results for **Human Identification** when training on expressions Anger and Disgust and testing on Joy. Using FisherFace as a baseline classifier, the table shows the error rates of FisherFace alone, FisherFace combined with AsymmetryFaces (FF+D+S+D_Y+S_Y+D_X+S_X) but without bagging, and FisherFace combined with AsymmetryFaces with Bagging. The bagging results (using 25 PCs) in terms of the average false negative rate (FNR), false positive rate (FPR) and the corresponding standard deviation over the 55 classes under different number of bootstrap samples are computed by averaging over 20 repetitions.

Sample	Error Rates	Test on Joy
w/o AsymFaces	FNR	19.4% 15 PCs [12]
w/o AsymFaces		12.12% 25 PCs [10]
w/o baggings	FNR	9.7% 15 PCs [12]
w/o baggings	FNR	2.42% 25 PCs [10]
10	FNR	1.03%
times	(STD)	(7.10%)
	FPR	0.02%
	(STD)	(0.11%)
20	FNR	0.73%
times	(STD)	(5.39%)
	FPR	0.01%
	(STD)	(0.09%)
30	FNR	0.73%
times	(STD)	(5.39%)
	FPR	0.01%
	(STD)	(0.08%)
40	FNR	0.61%
times	(STD)	(4.49%)
	FPR	0.01%
	(STD)	(0.08%)
50	FNR	0.41%
times	(STD)	(3.07%)
	FPR	0.01%
	(STD)	(0.03%)
60	FNR	0.42%
times	(STD)	(3.15%)
	FPR	0.01%
	(STD)	(0.06%)
70	FNR	0.36%
times	(STD)	(2.70%)
	FPR	0.01%
	(STD)	(0.05%)
100	FNR	0.24%
times	(STD)	(1.80%)
	FPR	0.01%
	(STD)	(0.03%)

Table 4: These are the test results for **Expression Identification** in terms of average false negative (FNR), false positive (FPR) and stnadard deviations (STD) for combining FisherFace with AsymmetryFaces, computed over the 20 repetitions.

Dataset	Error Rates	Joy	Anger	Disgust	Average	STD
FF	FNR	6.80%	19.40%	17.60%	14.60%	9.89%
	(STD)	(5.21%)	(9.82%)	(7.72%)	(3.63%)	(4.21%)
	FPR	2.30%	8.30%	11.30%	7.30%	6.27%
	(STD)	(2.70%)	(5.16%)	(5.44%)	(1.82%)	(2.75%)
D	FNR	25.80%	42.80%	41.60%	36.73%	13.32%
		(9.67%)	(11.17%)	(10.25%)	(3.88%)	(6.85%)
	FPR	19.80%	17.20%	18.10%	18.36%	6.57%
		(7.42%)	(6.88%)	(5.05%)	(1.94%)	(3.83%)
S	FNR	15.60%	17.80%	20.00%	17.80%	7.93%
		(8.50%)	(8.15%)	(9.97%)	(3.88%)	(5.95%)
	FPR	7.10%	8.10%	11.50%	8.90%	5.07%
		(5.86%)	(3.46%)	(4.94%)	(1.94%)	(2.88%)
D_Y	FNR	44.60%	37.00%	37.20%	39.60%	10.83%
		(8.03%)	(13.29%)	(9.28%)	(2.74%)	(6.98%)
	FPR	16.30%	21.00%	23.00%	19.8%	7.41%
		(6.23%)	(7.12%)	(6.54%)	(1.37%)	(4.17%)
S_Y	FNR	14.00%	20.80%	21.20%	18.67%	7.25%
		(5.58%)	(7.06%)	(7.00%)	(3.22%)	(3.21%)
	FPR	8.30%	8.10%	11.60%	9.33%	4.14%
		(4.22%)	(3.34%)	(4.33%)	(1.66%)	(2.40%)
D_X	FNR	37.20%	57.60%	48.20%	47.67%	13.01%
		(9.37%)	(9.39%)	(10.50%)	(3.59%)	(7.44%)
	FPR	26.00%	18.10%	27.10%	23.83%	9.10%
		(7.40%)	(6.57%)	(8.30%)	(1.79%)	(3.88%)
S_X	FNR	39.20%	51.80%	41.40%	44.13%	11.62%
		(10.27%)	(10.26%)	(10.96%)	(3.74%)	(7.07%)
	FPR	24.00%	19.00%	23.20%	22.07%	9.15%
		(10.76%)	(6.94%)	(9.63%)	(1.87%)	(6.53%)
FF+S+S_Y+S_X+D+D_Y+D_X	FNR	1.80%	3.80%	5.20%	3.60%	3.82%
	(STD)	(3.55%)	(3.03%)	(5.21%)	(1.94%)	(2.55%)
	FPR	0.30%	2.60%	2.50%	1.80%	2.26%
	(STD)	(0.98%)	(2.44%)	(2.42%)	(0.97%)	(1.16%)

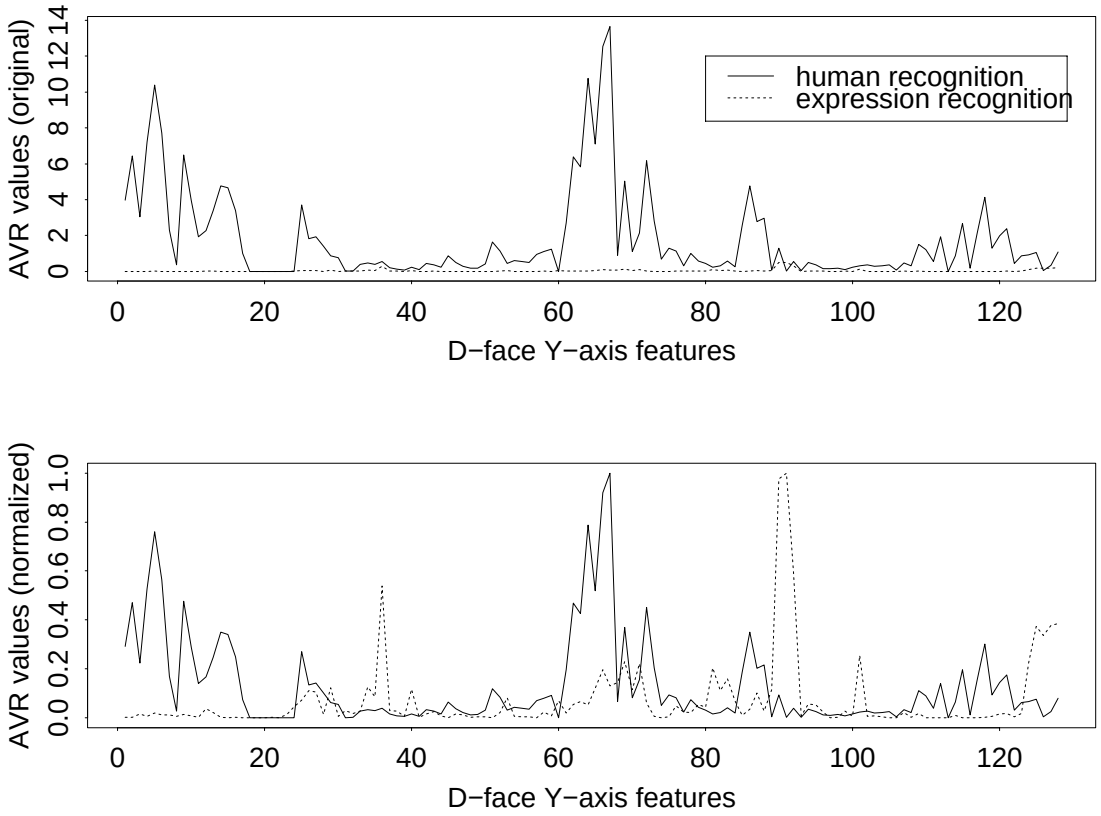


Figure 5: *Each plot shows the AVR value of each of the AsymmetryFaces D_y defined in Table 1). The higher the AVR value for a single dimension the more discriminating power that individual feature has. D_y and S_y (128 dimensions) start from forehead (left) to chin (right). D_x and S_x (64 dimensions) start from side face (left) towards the face midline (right). Solid lines indicate AVR values of respective AsymmetryFace features for human identification under expression variations. Dashed lines inidicate AVR values of respective AsymmetryFace features for Expression identification across different subjects. One can observe that the features with high weights for human identification have low weights for expression classification, and vice versa. That implies that different subset of features are selected for the two different discrimination problems.*

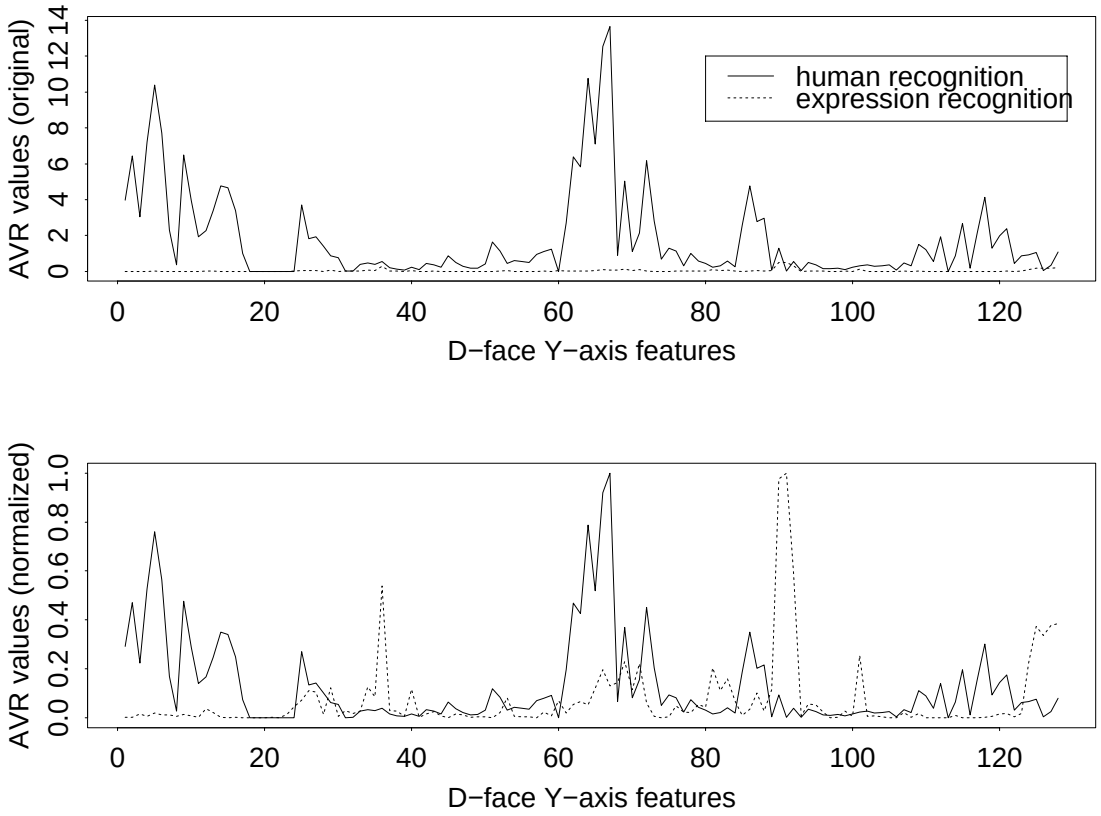


Figure 6: *Each plot shows the AVR value of each of the AsymmetryFaces D_x defined in Table 1). The higher the AVR value for a single dimension the more discriminating power that individual feature has. D_y and S_y (128 dimensions) start from forehead (left) to chin (right). D_x and S_x (64 dimensions) start from side face (left) towards the face midline (right). Solid lines indicate AVR values of respective AsymmetryFace features for human identification under expression variations. Dashed lines inidicate AVR values of respective AsymmetryFace features for Expression identification across different subjects. One can observe that the features with high weights for human identification have low weights for expression classification, and vice versa. That implies that different subset of features are selected for the two different discrimination problems.*

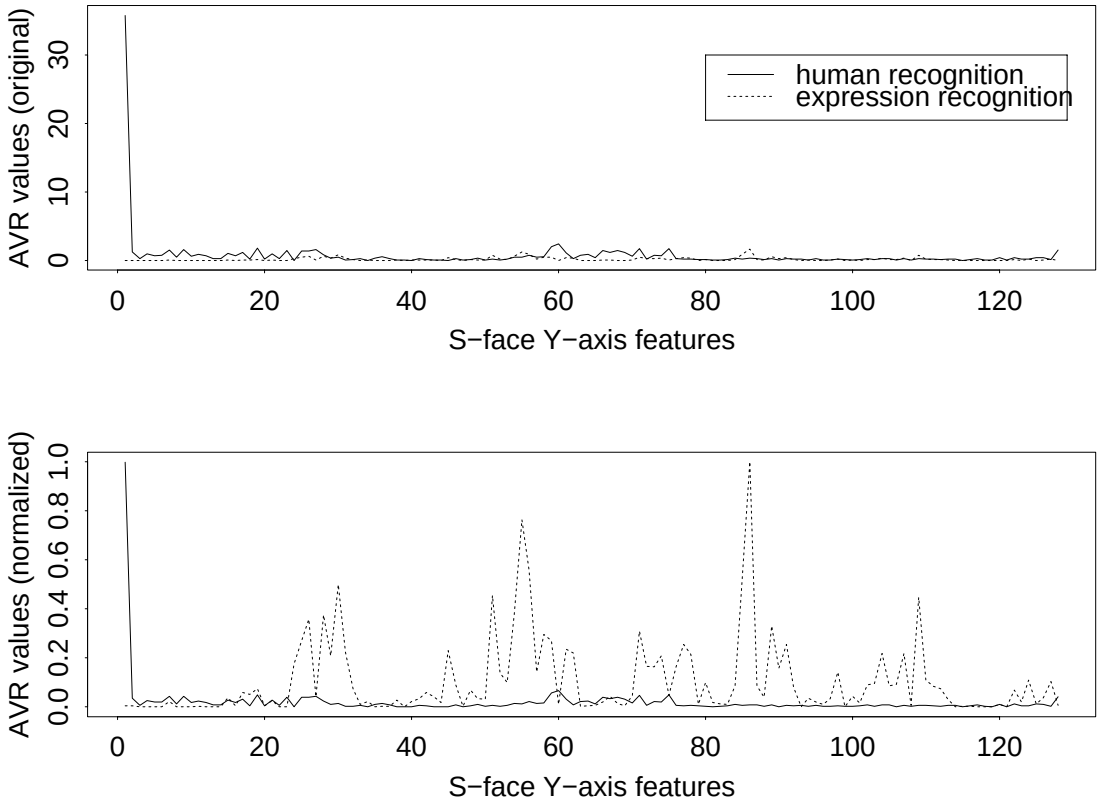


Figure 7: *Each plot shows the AVR value of each of the AsymmetryFaces S_y defined in Table 1). The higher the AVR value for a single dimension the more discriminating power that individual feature has. D_y and S_y (128 dimensions) start from forehead (left) to chin (right). D_x and S_x (64 dimensions) start from side face (left) towards the face midline (right). Solid lines indicate AVR values of respective AsymmetryFace features for human identification under expression variations. Dashed lines inidicate AVR values of respective AsymmetryFace features for Expression identification across different subjects. One can observe that the features with high weights for human identification have low weights for expression classification, and vice versa. That implies that different subset of features are selected for the two different discrimination problems.*

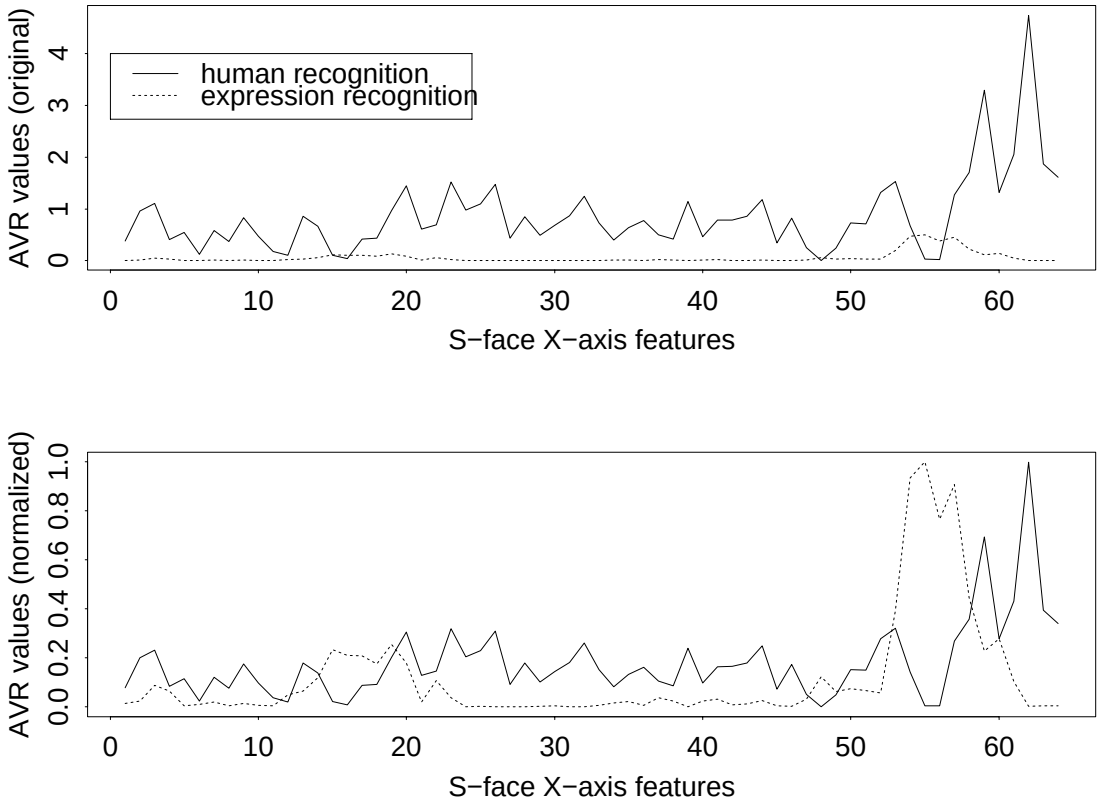


Figure 8: *Each plot shows the AVR value of each of the AsymmetryFaces S_x defined in Table 1). The higher the AVR value for a single dimension the more discriminating power that individual feature has. D_y and S_y (128 dimensions) start from forehead (left) to chin (right). D_x and S_x (64 dimensions) start from side face (left) towards the face midline (right). Solid lines indicate AVR values of respective AsymmetryFace features for human identification under expression variations. Dashed lines inidicate AVR values of respective AsymmetryFace features for Expression identification across different subjects. One can observe that the features with high weights for human identification have low weights for expression classification, and vice versa. That implies that different subset of features are selected for the two different discrimination problems.*