IMPROVED HUMAN FACE IDENTIFICATION USING FREQUENCY DOMAIN REPRESENTATION OF FACIAL ASYMMETRY

Sinjini Mitra

Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

Marios Savvides

Electrical and Computer Engineering Department
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

This paper explores the role of facial asymmetry in identification tasks using a frequency domain representation. Satisfactory results are obtained for two different tasks, namely, human identification under extreme expression variations and expression classification, using a PCA-type classifier which establishes the robustness of these measures to intra-personal distortions. We next demonstrate that it is possible to even improve upon these results by simple means. In particular, we use two methods, namely, feature set combination and statistical resampling methods like bagging, which attains perfect classification results (0% error rate) in some cases. Both these methods require very few additional resources in terms of computing power, hence they are useful for practical applications as well.

1. INTRODUCTION

A commonly accepted notion in computer vision is that human faces are bilaterally symmetric ([1]) and [2] reported no differences whatsoever in recognition rates while using only the right and left halves of the face. However, a well-known fact is that manifesting expressions cause a considerable amount of facial asymmetry, they being more intense on the left side of the face ([3]). Indeed [4] found differences in recognition rates for the two halves of the face under a given facial expression.

Human faces have two kinds of asymmetry − intrinsic and extrinsic. The former is caused by growth, injury and age-related changes, while the latter is affected by external factors such as viewing orientation and lighting direction. Of the two, intrinsic asymmetry is more interesting since it is directly related to the individual face structure while extrinsic asymmetry can be controlled and removed to an extent by normalization. Psychologists have observed that the more asymmetric a face, the less attractive it is ([5]). Furthermore, the less attractive a face is, the more recognizable it is ([6]). The role of asymmetry in automatic identification was first studied by Liu ([7]), who investigated the role of spatial intensity-based asymmetry features in human identification tasks. This was followed by more in-depth studies ([8], [9]) which further investigated the role of asymmetry measures for human as well as expression classifications. The seminal work on using frequency domain representation of facial asymmetry in automatic identification was [10] which serves as a baseline for this work.

The paper is organized as follows. Section 2 describes the dataset used. Section 3 introduces the frequency domain asymmetry representation along with identification results. Section 4 presents two simple techniques for improving classification performance of these features along with the results, and finally a discussion appears in Section 5.

2. DATA

The dataset used is a part of the “Cohn-Kanade Facial Expression Database” ([11]), consisting of images of 55 individuals expressing three different emotions - joy, anger and disgust. The data thus consist of video clips of people showing an emotion, each clip being broken down into several frames. The raw images are normalized using an affine transformation, the details being included in [8]. Some normalized images from our database are shown in Figure 1.

Fig. 1. Sample images from our database.

We use a total of 495 frames − 3 frames from each emotion for each subject ($55 \times 3 \times 3$), same as that used in [10] to facilitate comparison. These frames are chosen from the most neutral (the beginning frame), the most peak (the final frame) and a middle frame in the entire sequence.

3. THE FREQUENCY DOMAIN

Many engineering applications in involve the frequency-domain representation of signals. The frequency spectrum consists of two components, the *magnitude* and *phase*. In 2D images particularly, the phase component captures more of the image intelligibility than magnitude and hence is very significant for performing image reconstruction ([12]). [13] showed that correlation filters built in the frequency domain can be used for efficient face-based recognition. Recently, the significance of phase has also been used in identification problems. [14] proposed correlation filters based only on the phase component of an image, which performed as well as the original filters. All these indicate the benefits of considering classification features in the frequency domain for potentially improved results.

Symmetry properties of the Fourier transform are often very useful ([15]). Any sequence $x(n)$ can be expressed as a sum of a *symmetric* part $x_e(n)$ and an *asymmetric* part $x_o(n)$. Specifically,

$$x(n) \;=\; x_e(n) + x_o(n),$$

where $x_e(n) = \frac{1}{2}(x(n) + x(-n))$ and $x_o(n) = \frac{1}{2}(x(n) - x(-n))$. When a Fourier transform is performed on a real sequence $x(n)$, the even part ($x_e(n)$) transforms to the real part of the Fourier transform and the odd part ($x_o(n)$) transforms to its imaginary part (Fourier transform of any sequence is generally complex-valued). The Fourier transform of a real and even sequence is thus real; that of a real and odd sequence is purely imaginary. This implies that the imaginary component of the Fourier transform can therefore be considered as a measure of facial asymmetry in the frequency domain and the real component a measure of facial symmetry. However, these relations hold for one-dimensional sequences alone, and hence we define asymmetry features based on the Fourier transforms of row slices of the images. This is intuitively clear because we are interested in facial symmetry only in the horizontal direction (left-right) which is preserved by 1D Fourier transforms. 2D transforms create symmetry in a diagonal direction with respect to the origin that destroys actual symmetry information, and hence does not offer a valid representation.

3.1. The Asymmetry Biometric

Following the notion presented above, we define two asymmetry biometrics for the images in our database as follows:

- **I-face**: frequency-wise imaginary components of Fourier transforms of each row slice.
- **R-face**: frequency-wise real parts of the Fourier transforms of the 1D row slices of the edged images I_e.

Both these feature sets are of the same dimension as the original images (128×128 for our database). A higher value of I-face signifies greater asymmetry between the two sides of a face whereas a higher R-face value signifies greater symmetry between the edges on the two sides of the face ("edged" images computed using a standard edge-detection algorithm). However, one half of both I-faces and R-faces contain all the relevant information owing to symmetry properties (I-face has the same magnitude but opposite sign across the face midline while R-face has both the same magnitude and sign on the two sides).

3.2. Identification Results

Of the different classification methods tried (including LDA, Fisher faces, SVM), the best results are obtained with the individual PCA (IPCA) approach ([16]). The IPCA method is different from the global PCA approach ([17]) where a subspace W is computed from all the images regardless of identity. In individual PCA, on the other hand, subspaces W_p are computed for each person p and each test image is projected onto each individual subspace using $y_p \;=\; W_p^T(x - m_p)$. The image is then reconstructed as $x_p \;=\; W_p y_p + m_p$ and the reconstruction error is computed as: $||e_p||^2 \;=\; ||x - x_p||^2$. The final classification chooses the subspace with the smallest $||e_p||^2$.

For human identification, the training is done on the neutral frames of the 3 emotions of joy, anger and disgust from all the 55 individuals in the dataset and testing on the peak frame of all the 3 emotions from all the people. For expression classification, we train on the peak expression frames for a randomly selected subset of 30 individuals and test on those from the remaining 25 individuals. This is repeated 20 times and the final error rates are obtained by averaging over those from these 20 repetitions. The results appearing in Table 1 are fairly good although there seems to be plenty of room for improving upon their performance, an issue we explore in the next section. The I-faces perform better than the R-faces for both identification problems.

Features	Human Identification	Expression Classification
I-face	3.64%	26.93% (4.18%)
R-face	10.30%	27.07% (6.36%)

Table 1. Misclassification rates using asymmetry measures.

4. IMPROVING CLASSIFICATION PERFORMANCE

Given the sensitive nature of applications of face recognition technology today, say in homeland security, it is desirable to have as accurate algorithms as possible. Hence in this section, we focus our effort to attain as near perfect performance (100% accuracy) as possible with the help of two different but simple techniques: (i) combination of feature sets, and (ii) use of statistical resampling methods.

4.1. Feature Set Combination

Here, we concatenate the I-faces and the R-faces to yield a two-dimensional feature vector per frequency and perform both human identification and expression classification in the same way as before. The idea was to investigate if the two sets of features could complement each other and result in improved performance. The results in Table 2 clearly show that improvement occurs in their performances for both classification tasks. For human identification, the R-face results improve significantly whereas the amount of improvement in the I-face results is not that significant which is expected given the already good results from using them alone. As to expression classification, both the feature sets show significant improvements as a result of the combination. The relative improvement figures shown in Table 3 are calculated as Imp. $= \frac{\text{Individual error rate - Combined error rate}}{\text{Individual error rate}} \times 100$. We consider this instead of absolute improvement figures because this gives a more realistic idea of the actual gain, since it is much harder to improve upon an error rate of 1% than one of 15%. This fact is reflected in the relative rates in a more precise manner.

Classification	Error rates
Human	2.78%
Expression	18.90%

Table 2. Error rates for feature set combination I-face+R-face.

Classification	Imp. for I-face	Imp. for R-face
Human	23.63%	73.01%
Expression	29.82%	30.18%

Table 3. The improvements rates for feature set combinations over the individual feature sets.

4.2. Bagging

Statistical resampling methods are well-known as effective means of improving performance of several classifiers in a fairly simple way. Two different resampling methods are used - Bagging and Random Subspace Method (RSM), and we apply these to the individual feature sets as well as to their combinations.

Bagging was introduced as a method for increasing the accuracy of *unstable* predictors, that is, if results from the underlying predictor are significantly affected by small perturbations of the training set ([18]). On the other hand, it is less effective if the underlying predictor is sufficiently stable, and can even do worse in such a scenario. According to [19], linear classifiers built on large training sets are stable. Hence, when the training sets are large, bagging will not improve results for them. Bagging is useless for very small training samples as well, since small training sets often represent the actual distribution poorly and the resultant classifiers are likely to be equally poor. However, when the training sample size is "critical" (the number of training samples is comparable to the number of features), linear classifiers can be quite unstable. So bagging linear classifiers such as PCA might be beneficial for high-dimensional data in general, which prompts us to apply this to our problem at hand.

The methodology of bagging consists of generating independent replications with replacement from the given training set and developing a classifier based on each of the bootstrap samples by treating them as separate training sets. The final results are obtained by aggregating the results from the replication by a simple majority voting rule. The number of replications is application-dependent and needs to be chosen suitably by the user.

We apply the procedure of bagging only to the human identification problem for the time being, but we wish to extend this to expression classification in a similar way. We use different number of replications for all the feature sets using IPCA as the base classifier, and the entire resampling procedure for each replication size is repeated 20 times (to remove selection bias) so that the final bagging errors are obtained by averaging over these 20 iterations. Figure shows how the bagging errors for human identification based on I-faces, of which the optimal number of replication is chosen to be 150 (that produce the lowest error rate). Note that this is considerably higher than the convention of 50 replication used in most statistical applications, the reason being that our feature sets are much higher dimensional than most standard statistical problems. The figures for the other feature sets are not included for space constraints. Table 4 shows the final bagging errors along with the relative improvements computed in the same way as for the feature set combinations for all the feature for human identification. 100% classification accuracy is obtained with the I-faces and the combination of I-faces and R-faces, and significant improvement (close to 90%) was also acheived for the R-faces.

	I-faces	R-faces	I-faces+R-faces
Errors	0%	1.25%	0%
Improv.	100%	87.86%	100%

Table 4. Final bagging errors and the relative improvements for the three feature sets for human identification.

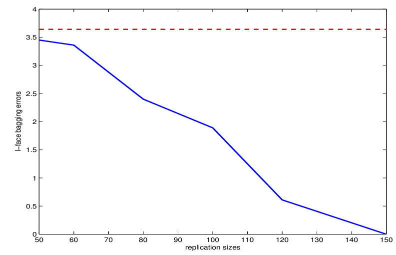

Fig. 2. Bagging errors for different replication sizes for human identification using the I-faces. The red dashed line shows the pre-bagging error rate of 3.64%.

5. DISCUSSION

We have thus shown in this paper that facial asymmetry measures in the frequency domain offer a promising potential as an useful biometric in practice, especially, in the presence of expression variations in face images. The initial results were further improved by the use of two simple methods of feature set combination and the statistical method of bagging. The latter was especially effective and produced 100% classification accuracy. This is a tremendous achievement and thus this method should be very useful in practical applications that seek accurate algorithms, such as in biometric identification including surveillance in airports. Also the implementation of both these techniques are very straightforward and requires very few additional resources and hence these are attractive from the viewpoint of practical viewpoint. Finally, these tools also helped demonstrate the true potential of these frequency-based asymmetry features in identification tasks.

The next direction of research will consist of applying bagging to expression classification, exploring other resampling methods like Random Subspace method (RSM; [20]) and boosting. We also wish to investigate the robustness of these frequency domain asymmetry features to other types of distortions like illumination, and extension to a larger database.

6. REFERENCES

[1] Seitz, S.M. and Dyer, C.R., "View morphing," *SIGGRAPH*, pp. 21–30, 1996.

[2] Gutta, S., Philomin, V., and Trajkovic, M., "An investigation into the use of partial-faces for face recognition," in *International Conference on Automatic Face and Gesture Recognition*, Washington D.C., 2002, pp. 33–38.

[3] Borod, J.D., Koff, E., Yecker, S., Santschi, C., and Schmidt, J.M., "Facial asymmetry during emotional expression: gender, valence and measurement technique," *Psychophysiology*, vol. 36, no. 11, pp. 1209–1215, 1998.

[4] Martinez, A.M., "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *PAMI*, vol. 24, no. 6, pp. 748–763, 2002.

[5] Thornhill, R. and Gangstad, S. W., "Facial attractiveness," *Transactions in Cognitive Sciences*, vol. 3, no. 12, pp. 452–460, 1999.

[6] Troje, N. F. and Buelthoff, H. H., "How is bilateral symmetry of human faces used for recognition of novel views?," *Vision Research*, vol. 38, no. 1, pp. 79–89, 1998.

[7] Liu, Y., Schmidt, K., Cohn, J., and Weaver, R.L., "Human facial asymmetry for expression-invariant facial identification," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

[8] Liu, Y., Schmidt, K., Cohn, J., and Mitra, S., "Facial asymmetry quantification for expression-invariant human identification," *CVIU*, vol. 91, no. 1/2, pp. 138–159, 2003.

[9] Mitra, S. and Liu, Y., "Local facial asymmetry for expression classification," in *Proceedings of CVPR*, 2004.

[10] Mitra, S. and Savvides, M., "Analyzing asymmetry biometric in the frequency domain for face recognitio," in *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP'05)*, 2005, pp. 953–956.

[11] Kanade, T., Cohn, J.F., and Tian, Y.L., "Comprehensive database for facial expression analysis," in *4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 46–53.

[12] Hayes, M.H., "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *ASSP*, vol. 30, no. 2, pp. 140–154, 1982.

[13] Savvides, M., Vijaya Kumar, B.V.K., and Khosla, P., "Face verification using correlation filters," in *3rd IEEE Automatic Identification Advanced Technologies*, Tarrytown, NY, 2002, pp. 56–61.

[14] Savvides, M. and Kumar, B.V.K., "Eigenphases vs.eigenfaces," in *IEEE International Conference in Pattern Recognition*, 2004.

[15] Oppenheim, A.V. and Schafer, R.W., *Discrete-time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[16] Turk, M. and Pentland, A., "Eigenfaces for recognition," *Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–96, 1991.

[17] Liu, X., Chen, T., and Vijaya Kumar, B.V.K., "Face authentication for multiple subjects using eigenflow," *Pattern Recognition, Special issue on Biometrics*, vol. 36, pp. 313–328, 2003.

[18] Breiman, L., "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[19] Skurichina, M. and Duin, R.P.W., "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Analysis and Applications*, vol. 5, pp. 121–135, 2002.

[20] Ho, T. K., "The random subspace method for constructing decision trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.