Expression Classification using Wavelet Packet Method on Asymmetry Faces

Kenny Teng Yanxi Liu

CMU-RI-TR-06-03

January 2006

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

© Carnegie Mellon University

Abstract

We are investigating wavelet packet method on asymmetry faces for expression classification. Two types of quantified asymmetry measures have already been defined by Liu *et al* in 2002, but they are computed only in the image intensity domain. We are proposing a novel approach to extend those asymmetry faces by using wavelet transforms to achieve better and more robust results. One of the most interesting characteristics of wavelet transforms is their ability to represent a signal into partitions of time-frequency plane. Using a random subset of 3 expressions with 55 subjects each from a normalized version of the Cohn-Kanade AU-Coded Facial Expression Database, error reduction rates of 25-95% have been achieved. Our findings show that asymmetry faces have some features that remain more consistent and discriminative in the wavelet subspaces than in the image intensity domain. We also show that certain wavelet subspaces will contribute more than others to classification accuracy.

Acknowledgements

The first author would like to thank Professor Yanxi Liu for helping him throughout the course of this project with productive discussions, valuable research materials pertaining to his research, and also help in co-authoring this paper. The first author will also like to acknowledge this amazing class of *Group Theory and its application in Robotics, Computer Vision/Graphics and Medical (18-799)* which is also taught by the second author. Through this class, the first author came to realize the importance of symmetry and its contribution to many different fields of research and hence the successful realization of this project. Finally, the first author would also like to thank the students of 18-799 for their priceless feedback about this project during class presentations.

Contents

1. Motivation

Facial attractiveness has always been closely related to the approximate bilateral symmetry of human faces. According to Thornhill *et al* [1], asymmetrical faces are considered less attractive. Psychologists and anthropologists have considered facial asymmetry as a critical factor that can be used to evaluate attractiveness and expressions [2, 3, 4], even though most of it was done qualitatively using human observers. As for human identification, half faces have been tested by Martinez [5] and different recognition rates have been reported. In literature [6], when facial asymmetry is removed from images of human faces, a small but statistically significant decrease in recognition rate is reported, therefore suggesting that facial asymmetry may be a key feature in human identification. For the first time, Liu *et al* have defined a way to quantify facial asymmetry in frontal human faces and showed that they contain discriminating information for human identification under expression variations [14]. They also showed that a synergy can be achieved by combining facial asymmetry information with classical human identification algorithms to yield significant statistical improvements [7]. Expression recognition and face recognition have a direct relationship according to [5, 13].

However, wavelet features may contain richer and more discriminative information for expression classification than spatial asymmetry features alone. The theory of wavelet analysis has been well studied in [9], where signals are represented by wavelet packet bases. This is done by dividing the time-frequency plane into regions with different resolutions. Wavelet packets have proved to be very practical in applications where time-frequency or space-frequency resolutions are needed. Daniell *et al* was one of the first to perform object recognition by applying correlation filters in the wavelet domains. Hennings *et al* have successfully used wavelet transforms to significantly improve identification and classification rates on fingerprints. They showed that wavelet subspaces contain features that are more pronounced for higher accuracy in recognizing fingerprints [10].

This is why we believe that wavelet analysis can be advantageous to classifiers applied to asymmetry faces defined in [7]. Different from previous work, we want to investigate whether asymmetry faces have consistent features that are retained in the wavelet subspaces for better classification of expressions. We will refer to those AsymmetryFaces in wavelet domain as *Wavelet AsymmetryFaces*. In particular, we use the Cohn-Kanade AU-Coded Facial Expression Database [11] as the primary test bed. In this paper we investigate (1) whether local facial asymmetry defined in [7] can be used for expression classification; (2) whether asymmetry faces in underlying wavelet subspaces contain more consistent features than the spatial domain alone; and (3) which wavelet subspace can improve overall classification rate.

The paper is organized as follows. First, we describe the data set that is going to be used for evaluation of the system (Section 2). Then, we give a review on quantification of facial asymmetry as formulated in [7] (Section 3), followed by an overview of wavelet transforms (Section 4). A review of the baseline classifier, Fisherface that will be used to discriminate between the expressions is presented in Section 5. Finally, our proposed methodology and experimental results (Section 6 and 7 respectively) are described before ending the paper with a brief discussion (Section 8) and conclusions (Section 9).

2. Data Set

One dataset is used to test the proposed method, which is a subset of the Cohn-Kanade AU-Coded Facial Expression Database [11]. The whole database consists of video sequences of subjects of different races and gender showing different facial expressions on demand. Each video sequence varies in length from 8 to 65 frames of size 640 x 480 pixels, starting from a neutral expression and ending with the requested expression. There are 55 subjects in the video data with 3 expressions: anger, disgust, and joy. For the purpose of this experiment, the same normalization process as reported in [7] is used. Each face image from the original database is subjected to an affine transformation using three points: left and right inner canthi and the philtrum. Each of those normalized faces is then cropped to 128 x 128 pixels. A subset of the normalized database is extracted to create the main test bed. Only the final frame (expression peak), with the requested expression, is selected from each of the 55 subjects. Therefore, the total sample size is 165 (3 x 55 = 165).

Figure 2. The above figure has been adopted from [13]. They are normalized faces from the Cohn-Kanade AU-Coded Facial Expression Database [11]. Each column represents a different subject displaying their peak faces. The rows show the 3 different classes: joy, disgust and anger.

3. Review of quantification of facial asymmetry

We are going to be using the same asymmetry faces defined by Liu *et al* in [7]. For clarity, we repeat the data description in this section. Let a normalized face intensity image be I, its vertically reflected image be I', and their respective "edge" images be I_e and I'_e (obtained by applying a Sobel edge extraction algorithm to I and I'). Then, two types of facial asymmetries can be defined:

Density Difference (D-face):

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

Edge orientation Similarity (S-face):

$$S(x, y) = \cos(\Phi_{I_e(x,y),I_e'(x,y)}) \quad (2)$$

where $\Phi_{I_e(x,y),I_e'(x,y)}$ is the angle between the edge orientations of two edge images I_e and I'_e at the same pixel location (x, y). Figure 3 shows some examples of the *D-face* and *S-face* of some original peak normalized images.

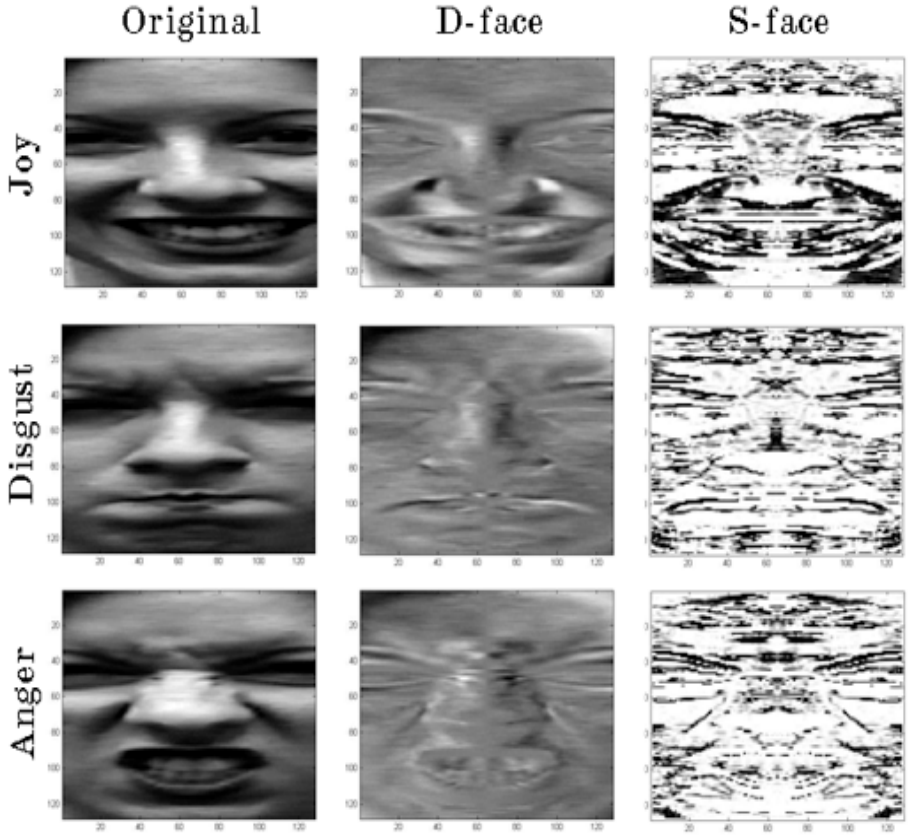


Figure 3. Some samples of original normalized faces, Difference face (D-face) and Similarity face (S-face) for the three different expressions.

D-face represents left-right relative intensity variation while *S-face* is affected by the zero-crossings of the intensity field. A high value on a *D-face* means that the face is very asymmetrical. In contrast, a symmetrical face will yield a high value of the *S-face*. From the way those facial asymmetry faces are constructed, the two halves of the *D-face* are opposite to each other while those of the *S-face* are symmetric. Therefore, only half of the D-face and S-face are retained for processing, and six projections (Table 1) can then be defined as in [7], namely **D**, D_x, D_y, **S**, S_x, S_y and they are called *AsymmetryFaces*.

Table 1: 6 projections called AsymmetryFaces are defined. D-face and S-face keep only 60 and 100 top principal components respectively to account for 95% of the variance in the data.

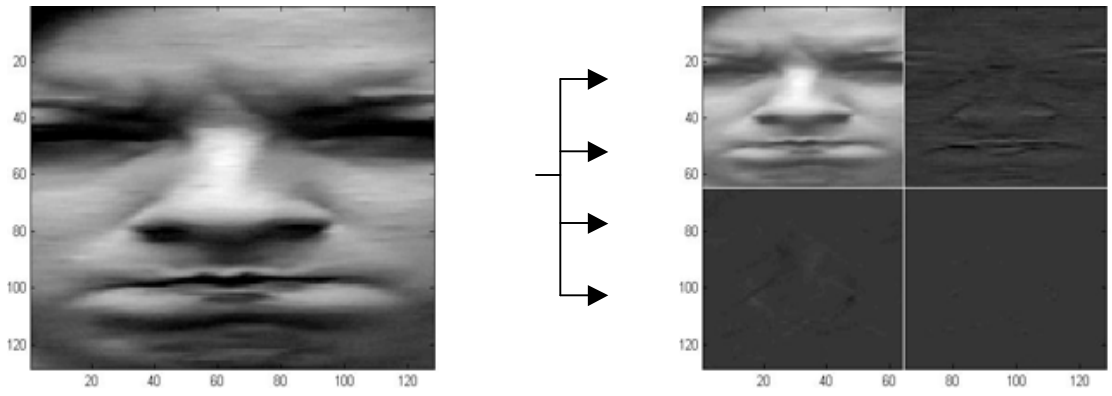
AsymmetryFaces	Description	Features
D	Top 60 Eigen vectors of *D-face*	60
D_x	Column-mean of *D-face* on X-axis	64
D_y	Row-mean of *D-face* on Y-axis	128
S	Top 100 Eigen vectors of *S-face*	100
S_x	Column-mean of *S-face* on X-axis	64
S_y	Row-mean of *S-face* on Y-axis	128

4. Wavelet Transforms

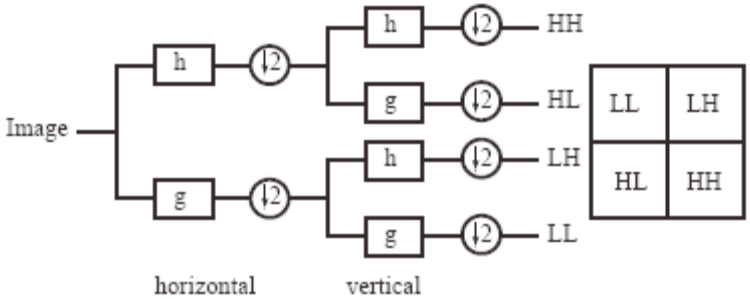
Wavelet transformations are a method of representing signals across space and frequency. For the last two decades, the theory of wavelets and multi-rate signal processing has become standard signal processing tools [9]. Filter banks generate the decomposition of signals into wavelet coefficients that provide multi-resolution into partitions of space-frequency domain. These partitions provide joint locality in both space and frequency, and therefore provide consistent features that are valuable in the analysis of biometrics.

Filter banks are the main signal processing tools that are currently used to implement wavelet transforms. As documented in [9], filter banks construct orthonormal bases for the space of finite-energy sequences. This means that we can project a signal onto spaces of low-pass and band-pass signals. The way to do such projection is by passing the signal through a low-pass filter which retrieves all the 'coarse' approximation of the original signal. At the same time the signal is passed through a complementary high-pass filter that captures the 'detailed' information from the original signal. Together, the coarse and detailed information can be used to reconstruct the original signal. The impulse responses of these filters that have been shifted by a multiple of 2 are designed to form an orthonormal basis [9]. In other words, filtered signals from either the low-pass or the high-pass are downsampled by two to avoid redundancy. This whole system can be described as an orthogonal two-channel analysis filter bank (Figure 4b). This two-channel analysis filter bank is defined as a decomposition of level 2 (with 4 subspaces). Consequently, a level 3 decomposition will have 16 subspaces at that level, and a level *n* decomposition will have 4^{n-1} subspaces.

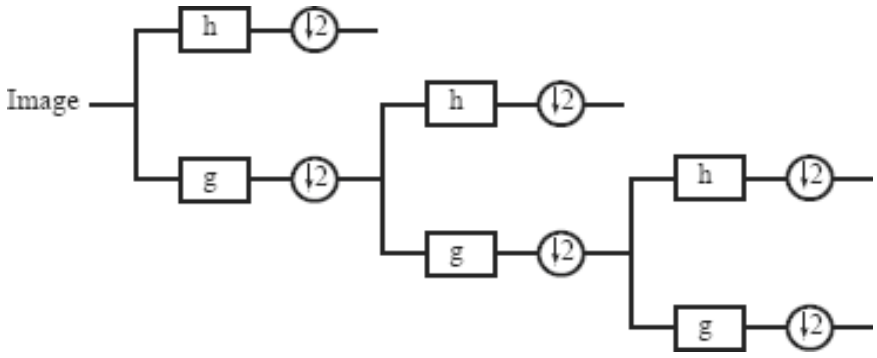
The decomposition of an image from the spatial domain into wavelet domains will form an admissible tree, which we are going to refer to as a *Wavelet tree*. (Figure 4).



(a)



(b)



(c)

Figure 4. (a) Single-level decomposition of an image into its four subspaces, LL, LH, HL and HH. (b) Quadtree representation of a single-level wavelet packet decomposition. (c) Discrete wavelet transform (DWT) which iterates only in the LL domain.

There are different flavors of wavelet transforms that can be used to represent signals. Each of them has their own interesting characteristics that can be helpful in analysis of biometric image characterization. One instance of a wavelet packet tree is called the Discrete Wavelet Transform (DWT) which iterates only in the low-pass channel (LL) shown in Figure 4c [9]. Another filter bank algorithm was introduced by Coifman *et al* which allows the high-pass channels to be decomposed as well and is called the discrete wavelet packet transform [12]. Using this generalized form of wavelet transform, it is

then possible to achieve an arbitrary time-frequency tiling. Therefore, a definition of a single-level expansion of a signal $x^{(i)}$ at level *i*, given an orthogonal low-pass filter *g* and its complementary high-pass filter *h* is as follows:

$$x^{(i)}[n] = \sum_{k \in Z} x_g^{(i+1)}[k] g[n-2k] + \sum_{k \in Z} x_h^{(i+1)}[k] h[n-2k] \quad (3)$$

where

$$x_g^{(i+1)}[n] = \sum_{k \in Z} x^{(i)}[k] g[k-2n], \quad (4)$$

$$x_h^{(i+1)}[n] = \sum_{k \in Z} x^{(i)}[k] h[k-2n] \quad (5)$$

The superscripts indicate the level of decomposition that the signals belong to and subscripts show which channel (low-pass or high-pass) is used to generate the signal. Equations (3), (4) and (5) are used to recursively wavelet packet coefficients which are stored in most cases as Quadtree representations (in 2-D signals). Therefore, for images wavelet packet decomposition will break a 2-D intensity image down into its four subspaces (LL, LH, HL and HH). In this study, we use the wavelet packet decomposition represented as a Quadtree that is for every parent subspace, four children subspaces will be generated using the two-channel orthogonal filter banks (Figure 4a). The figure below shows examples of *D-face* and *S-face* under full wavelet decomposition of level 3.

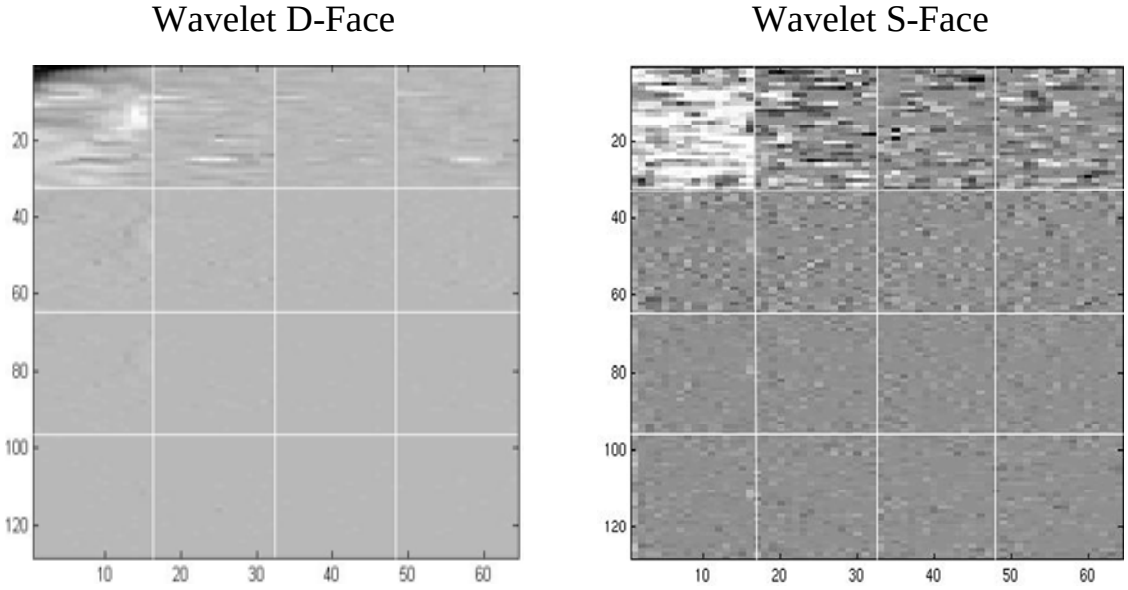


Figure 5. Examples of a wavelet packet decomposition of a *D-face* (left) and *S-face* (right) at level 2. Even though in the high frequency subspaces the remaining features are not clearly visible, there is still consistent discriminating information left in those wavelet domains.

5. Fisherface as a classifier

There are many classifiers out there that have proved to be very effective for classification of images. One of the most popular classifier used specifically in face recognition is Fisher Linear Discriminant Analysis (FLDA) or Fisherface [15]. The latter is a popular tool for multi-class pattern recognition as it takes advantage of the class scatter to make classification more reliable. The way Fisherface works is that it is a cascading of transformations of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). PCA is used for dimensionality reduction on the input data and LDA is then performed in the reduced dimensional space. LDA is the maximization of the ratio of discriminant of the projected between-class scatter matrix, S_B, to the determinant of the within-class scatter matrix, S_W.

$$S_B = \sum_{i=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T \tag{6}$$

$$S_W = \sum_{i=1}^{c} \sum_{k=1}^{N_i} (x_k^i - \mu_i)(x_k^i - \mu_i)^T \tag{7}$$

Therefore, maximize the Fisher ratio,

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{8}$$

where *w* are the projection vectors for all the different classes.

Table 2: Algorithm for performing Fisherface

1) Perform PCA for dimensionality reduction and keep N-c eigenvectors, where N is the total number of training images and c is the number of classes.
2) Project data onto these eigenvectors for dimension reductions.
3) Perform LDA in the reduced space and compute the c-1 projections where c is the number of classes.
4) Euclidean distance is then used as a nearest neighbor classifier.

6. Proposed Work

To test the effectiveness of *Wavelet AsymmetryFaces*, the normalized data set of peak faces from the Cohn-Kanade AU-Coded Facial Expression Video Sequence Database [11] is going to be used. The data set consists of 3 classes (joy, disgust and anger) with 55 subjects each. The experimental setup is as follows. The training set consists of 30 randomly selected faces from each class, therefore leaving 25 faces per class for testing

purposes. Each classification test will be computed over 20 repetitions and statistical results are going to be recorded.

6.1. Fitness metric

Feature discrimination is a very important factor to consider when trying classifying images. There are many different measures that can be used to evaluate the fitness of a specific feature dimension. To be consistent with Fisherface, variance ratios as used in [7] will be the main fitness metric. From [7], it was reported that it is possible for a feature to have small within-class scatter for a class even if the mean value of the feature is very close to the mean value of the same feature in other classes. Therefore, an augmented variance ratio (AVR) metric can be defined as follows.

$$AVR(F) = \frac{Var(S_F)}{\frac{1}{C}\sum_{i=1..C} \frac{Var_i(S_F)}{\min_{i \neq j}(|\, mean_i(S_F) - mean_j(S_F)\,|)}} \tag{9}$$

where $mean_i(F)$ is the mean of feature F's values in class *i*. AVR is the ratio of the variance of the feature between subjects to the variance of the feature within subjects, with an penalty for features that have close inter-subject mean values. Therefore, individual features with higher AVR values are more discriminating those others.

6.2. Forward feature subset selection

Feature selection algorithms can be classified in two main categories: wrapper and filter methods. Filter methods are independent of the classifier and select features based on properties that good feature sets are presumed to have such as class separability. Filter methods are computationally efficient, but may produce disappointing results. On the other hand, wrappers treat an induction algorithm as a black box that is used to evaluate each candidate feature subset. While usually giving good results in terms of the accuracy of the final classifier, wrapper approaches are computationally expensive. There are two classic greedy wrappers, namely sequential forward selection (SFS) and sequential backward elimination (SBE). Forward selection starts with an empty set of features. In the first iteration, the algorithm considers all feature subsets with only one feature. The feature subset with the highest accuracy is used as the basis for the next iteration. In each iteration, the algorithm adds to the basis each feature not previously selected and retains the feature subset that results in the highest estimated performance. The search then ends after the accuracy of the current subset cannot be improved by adding any more features. In contrast, the backward elimination works in an analogous way, starting from the full set of features and tentatively deleting each feature not deleted previously. In our investigation, we use SFS to get an optimal feature subset with reduced dimensions.

6.3. Match metric

In the evaluation stage of the investigation, we need to compute a match metric to be able to classify an expression to the appropriate class. Since Fisherface is used as the baseline classifier, an incoming test image will be first projected onto the principal components that have been defined by the training data and then LDA will be applied to get the final projected weights. Therefore, the match metric will make use of a nearest neighbor heuristic whereby the distance between test projected weights and those weights for each class will be computed.

6.4. Methodology

Since we are trying to investigate the implication of decomposing *AsymmetryFaces* into multiple subspaces to evaluate classification rate, two major steps will be conducted. The same algorithm defined in [7] will be used except that *Wavelet AsymmetryFaces* will be used. Given AsymmetryFaces *A* (which include the 6 different feature sets defined in Table 1), PCA components *P* of the normalized face images *F*, the experiments are carried out as follows.

- Step 1. AsymmetryFaces: for each of the 6 subsets of *A* and a combination of *P* along with everything of *A* (represented as **FF+AF**),
    - Compute the AVR value for each feature dimension of the training data and order the features in decreasing order of the AVR values (*P* followed by *A*).
    - Carry out sequential forward selection (SFS) on the sorted feature list and terminate with the optimal feature subset.
    - Apply Linear Discriminant Analysis (LDA) to the test data in the selected feature subspace.
    - Compute match scores by using a nearest neighbor heuristics.
    - Compute the overall False Positive Rate (FPR) and the False Negative Rate (FNR).

- Step 2. Wavelet AymmetryFaces: for each of the 6 subsets of *A* and a combination of *P* along with everything of *A*,
    - Compute the full wavelet packet transform of each training image, down to a maximum level of 3 (that is $1+4^1+4^2 = 21$ subspaces).
    - For each subspace,
        - Compute the AVR value for each feature dimension of the training data and order the features in decreasing order of the AVR values (*P* followed by *A*).
        - Carry out SFS on the sorted feature list and terminate with the optimal feature subset.
        - Apply LDA to the test data in the selected feature subspace.
        - Compute the current subspace match scores using a nearest neighbor heuristics.

        - Update the final match scores if and only if scores are improved by including current subspace.
        - Classify using overall match scores.
    - Compute the overall FPR and the FNR on test images.

Each of the 2 above steps is computed over 20 repetitions and the average FPR and FNR, and the standard deviation of FPR and FNR are computed. We are assuming that 20 are a big enough sample size to give reliable outcomes.

7. Experimental Results

In Table 3, we display all the results that have been produced following the two experimental steps (Section 6.4) on the normalized face dataset described in Section 2. Overall, the average false negative rates (FNR) for all the Wavelet AsymmetryFaces and **FF+AF** decrease compared to those without wavelet transforms. The Error Improvement Rate (EIR) defined as $(\%Error_{spatial} - \%Error_{wavelet})/\%Error_{spatial} \times 100\%$, of **S** faces (S, S_x and S_y) ranges between 6.1% (Sx for disgust) and 38.6% (S for disgust). However, the biggest change can be seen on the D faces (D, D_x and D_y) where the improvement of the average FNR ranges between 49.6% (D_x for anger) and 86.2% (D for anger).

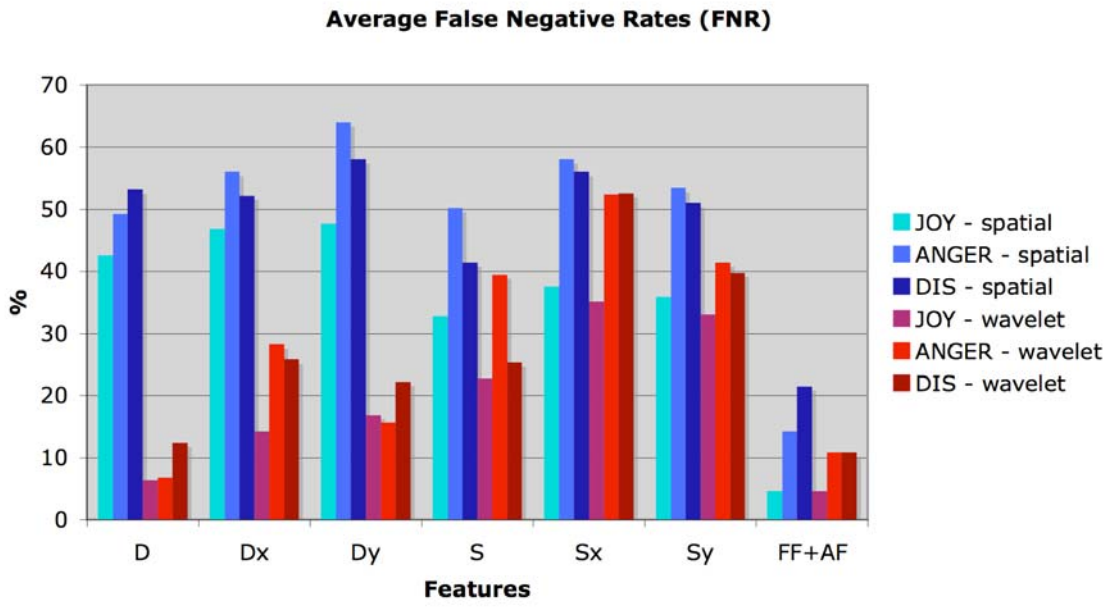
Similarly, the EIR of average false positive rates (FPR) show major improvements on the Wavelet AsymmetryFaces. Again, the D faces have the biggest overall improvement with a range between 42.9% (D_y for anger) and 93.4% (D for joy). As for S faces, the EIR is smaller between 3.3% (S_x for disgust) and 55.9% (S for joy). These results clearly show that Wavelet AsymmetryFaces have discriminative features that can help at classifying classes more accurately with reduced FNR and FRP. The most significant improvements can be seen for Joy, which reaches a maximum improvement of roughly 93%.

As for the standard deviation of both the results for FNR and FPR, there are improvements mostly for D faces, but S faces tend to show some negative results in the Wavelet AsymmetryFaces. Figures 6(b) and 7(b) show how there are both improvements and deterioration in the standard deviation in the classification of the 3 expressions. The worst case that happens is with the standard deviation of FNR of S_x face on anger. The standard deviation of that Wavelet AsymmetryFace gets worse by 66.4% as it increases from 9.6% to 15.9% in the classification of $\mathbf{S_x}$ Wavelet AsymmetryFace. Again, this shows that wavelet packets help to retrieve features mostly in the D faces and that the S faces do not contain any more features to extract even in the wavelet domains.

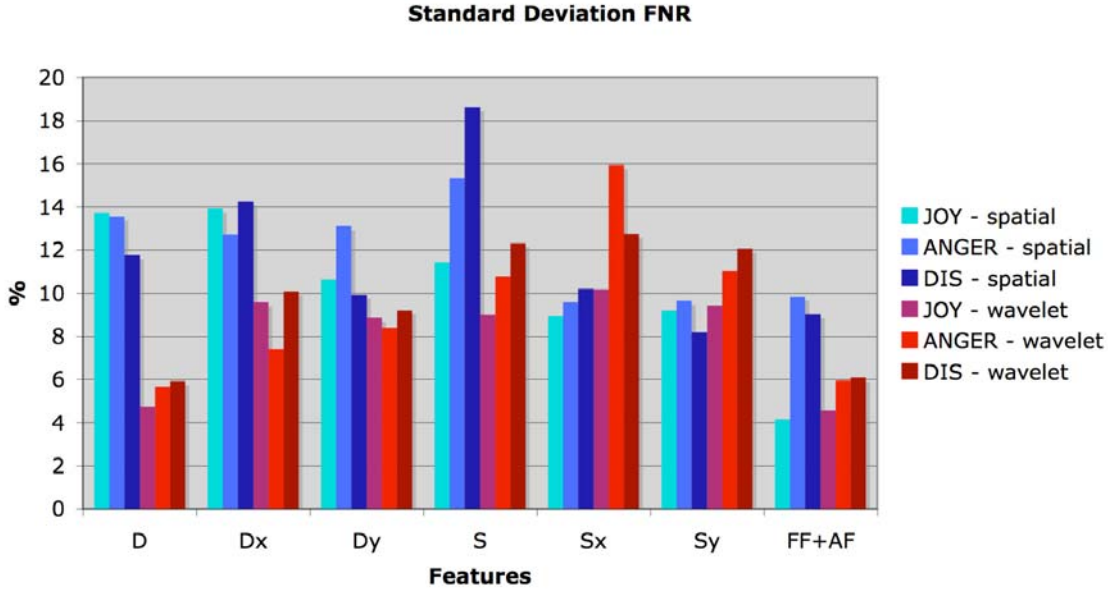
Finally, the overall classification rates of Wavelet AsymmetryFaces compared to AsymmetryFaces are 91.3% and 86.6% respectively.

Table 3: Test results for Expression Classification in terms of average false negative rate (FNR), false positive rate (FPR) and the standard deviations (STD) over 20 iterations.

Average False Negative Rates (FNR)						
	Spatial domain (step 1)			**Wavelet domain (step 2)**		
	JOY	ANGER	DIS	JOY	ANGER	DIS
D	42.6%	49.2%	53.2%	6.8%	6.8%	12.4%
D_x	46.8%	56.0%	52.2%	14.2%	28.2%	25.8%
D_y	47.7%	64.0%	58.0%	16.8%	15.6%	22.2%
S	32.8%	50.2%	41.4%	22.8%	39.4%	25.4%
S_x	37.6%	58.0%	56.0%	35.2%	52.4%	52.6%
S_y	35.8%	53.4%	51.0%	33.0%	41.4%	39.8%
FF+AF	4.6%	14.2%	21.4%	3.6%	10.8%	10.8%
Average False Positive Rate (FPR)						
	Spatial domain (step 1)			**Wavelet domain (step 2)**		
	JOY	ANGER	DIS	JOY	ANGER	DIS
D	19.7%	30.4%	22.4%	1.3%	8.5%	3.0%
D_x	25.3%	23.3%	28.9%	10.5%	13.0%	10.6%
D_y	27.3%	26.8%	30.8%	3.5%	15.3%	8.5%
S	20.4%	17.0%	24.8%	9.0%	11.7%	23.1%
S_x	25.1%	28.2%	24.5%	21.5%	22.9%	23.7%
S_y	20.5%	21.5%	28.1%	14.7%	17.7%	24.7%
FF+AF	1.0%	11.2%	7.9%	0.8%	6.6%	5.7%
Standard Deviation (FNR)						
	Spatial domain (step 1)			**Wavelet domain (step 2)**		
	JOY	ANGER	DIS	JOY	ANGER	DIS
D	13.8%	13.6%	11.8%	4.8%	5.7%	5.9%
D_x	13.9%	12.7%	14.2%	9.6%	7.4%	10.1%
D_y	10.7%	13.1%	9.9%	8.9%	8.4%	9.2%
S	11.4%	15.3%	18.6%	9.0%	10.8%	12.3%
S_x	8.9%	9.6%	10.2%	10.2%	15.9%	12.7%
S_y	9.2%	9.6%	8.2%	9.4%	11.0%	12.1%
FF+AF	4.2%	9.8%	9.0%	4.5%	6.0%	6.1%
Standard Deviation (FPR)						
	Spatial domain (step 1)			**Wavelet domain (step 2)**		
	JOY	ANGER	DIS	JOY	ANGER	DIS
D	9.2%	10.3%	8.9%	1.6%	4.4%	2.6%
D_x	10.2%	11.5%	8.1%	6.1%	5.7%	4.7%
D_y	8.9%	7.0%	7.4%	3.0%	6.2%	5.0%
S	18.1%	9.1%	11.9%	7.4%	6.0%	8.6%
S_x	6.1%	7.2%	7.8%	6.4%	5.6%	7.8%
S_y	5.3%	5.9%	7.2%	6.5%	6.2%	7.9%
FF+AF	1.2%	5.5%	4.9%	1.2%	4.0%	3.2%

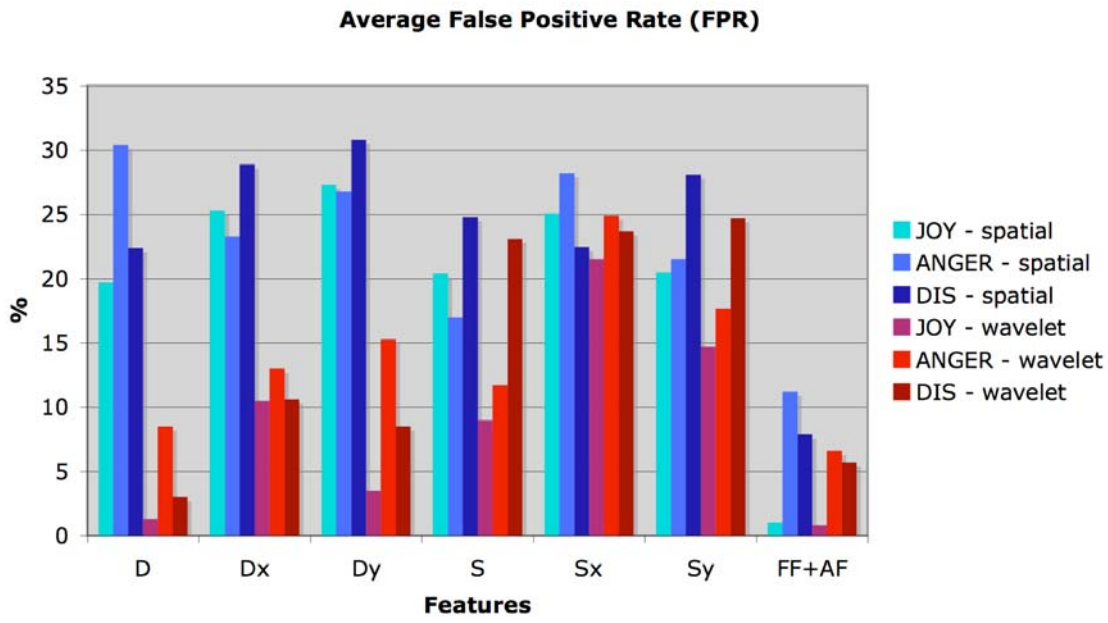


(a)

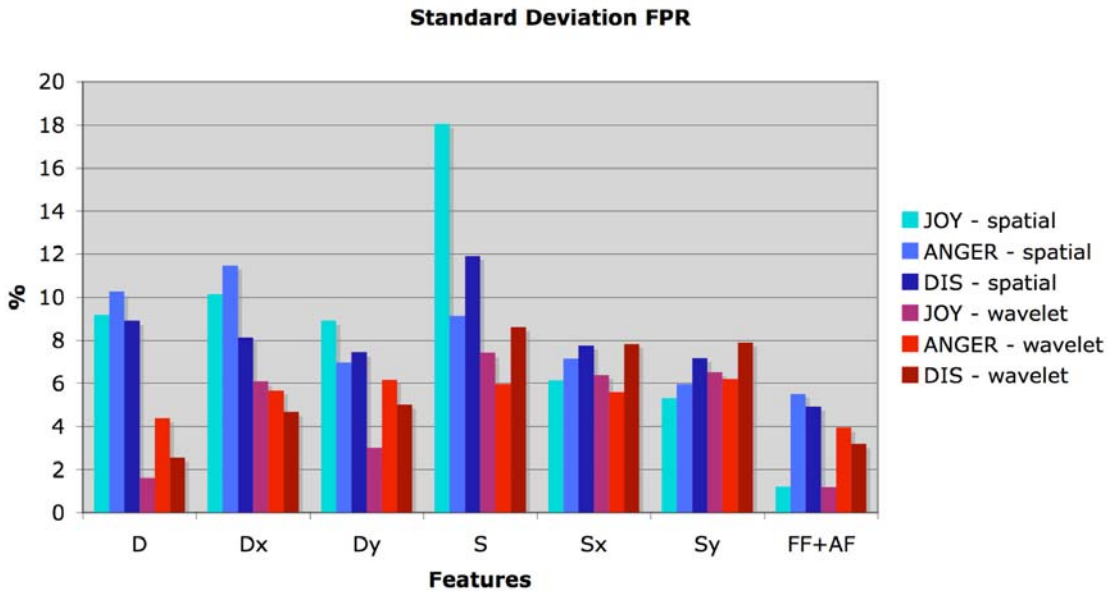


(b)

Figure 6. (a) Average False Negative Rates of each of the 6 AsymmetryFaces and FF+AF, tested over 20 iterations. Bars labeled with a 'Wave' prefix show results when wavelet transforms are used. (b) Standard deviation of FNR.



(a)



(b)

Figure 7. (a) Average False Positive Rates of each of the 6 AsymmetryFaces and FF+AF, tested over 20 iterations. Bars labeled with a 'Wave' prefix show results when wavelet transforms are used. (b) Standard deviation of FPR.

8. Discussions

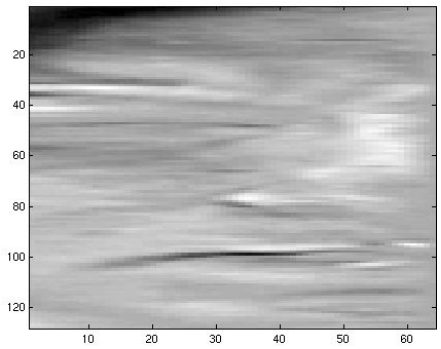
8.1. Error rate improvement in relation to the wavelet subspaces

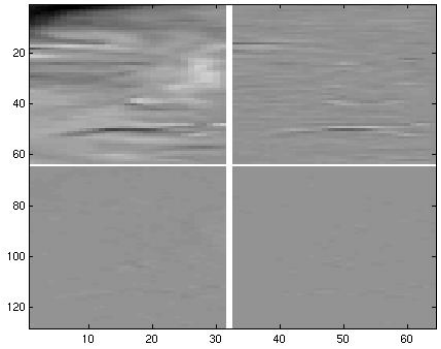
My experimental results in Table 3 show that ***D*** from the D-faces (D, Dx and Dy) of Wavelet AymmetryFaces produces the greatest improvements (86.2% on FNR and 93.4% on FPR). It is clear that the D-faces have discriminative features that are not being used in the spatial domain, but that become very prominent and useful in the wavelet domains. However, there are many different subspaces within the full wavelet tree, and some of them may be used to improve the overall classification rate while others may deteriorate the results. Then, a pertinent question can be raised: which subspace will affect the overall classification in a positive way and which one will affect it in a negative way?

We tried to answer this question by running a test on the D-faces alone. In particular, since ***D*** of the D-faces yield the best improvement, the same classification experiment defined in step 2 of Section 6.4 is run on ***D*** alone. This time, the classification rate that each subspace yield individually is recorded and the average and standard deviation over 20 repetitions are reported. Table 4 shows the results of individual subspace classification rate on ***D*** for a full wavelet tree of level 3. Several interesting points can be noted from those results:

- The subspaces that yield the smallest average classification rates are the image intensity (level 1), the LL (level 2) and the LLLL (level 3). This clearly shows that those spaces, which are the original signal, and 'coarse' versions of it, do not retain highly discriminative features.
- Significantly higher classification rates are achieved in the other channels (LH, HL and HH). Further wavelet decomposition in those channels (from level 2 to level 3) shows that those features are still retained further down the tree. Not only are they retained, but in some channels, they become even more pronounced and highly discriminative.
- If the different channels (LL, LH, HL and HH) are evaluated separately, the mean classification rate for those channels can be computed as ($\%Level2_{LL}$ + $\%Level3_{LL}$ + $\%Level3_{LH}$ + $\%Level3_{HL}$ + $\%Level3_{HH}$) / 5. Therefore, those mean rates are 62.4% for LL, 66.8% for LH, 73.5% for HL and 68.4% for HH. This shows that the higher frequency channels contain more discriminating features, with higher combined average classification rates.
- The standard deviations of the classification rates over 20 repetitions vary between 4.03% and 8.22% for most of the subspaces. Those values tell us that the discriminative features in the subspaces are consistent and that the classification results do not change by a lot over the 20 repetitions.

Table 4: Average classification rate and standard deviation (shown in parentheses) of each subspace on a full tree of level 3 for ***D* f**aces. LL means that the image was passed through the low-pass-low-pass channel and LLLL means that it was passed through LL followed by another LL.

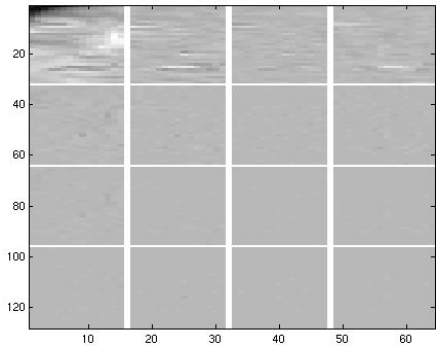
<table>
<tr><th>Wavelet spaces at each level</th><th colspan="4">Average classification rate</th></tr>
<tr><td>Level 1
  </td><td colspan="4">Image intensity Domain

52.0%
(4.03%)</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td rowspan="2">Level 2
  </td><td colspan="2">LL

48.8%
(4.33%)</td><td colspan="2">LH

68.6%
(7.72%)</td></tr>
<tr><td colspan="2">HL

77.2%
(6.64%)</td><td colspan="2">HH

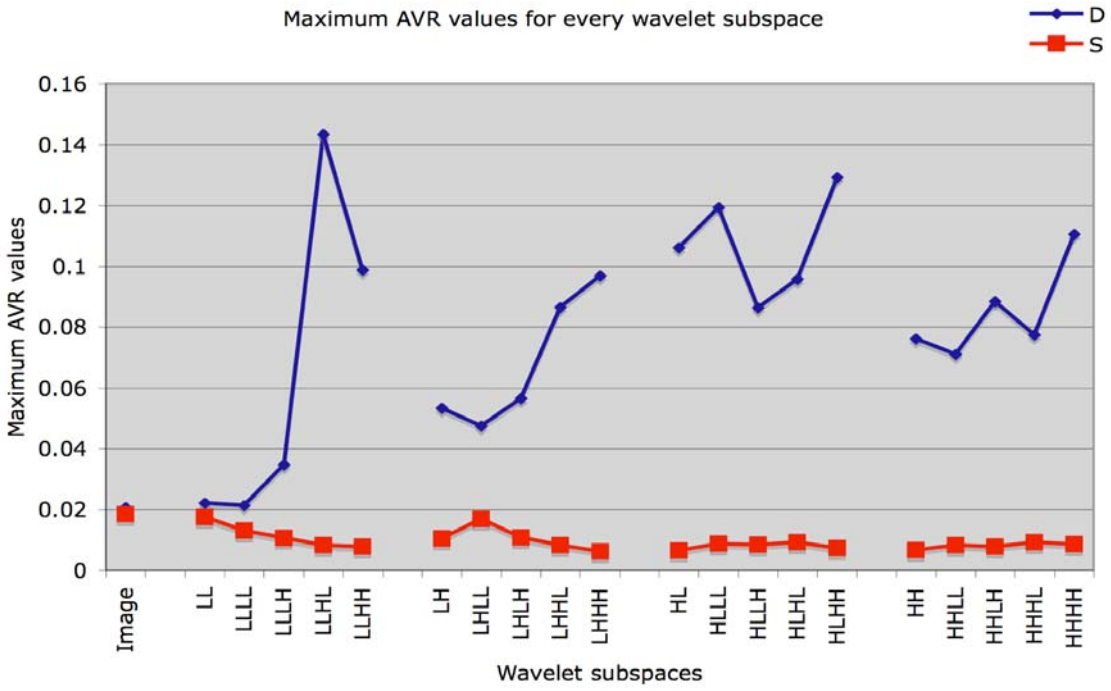
74.4%
(6.10%)</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td rowspan="4">Level 3
  </td><td>LLLL
47.8%
(4.60%)</td><td>LLLH
61.5%
(7.83%)</td><td>LHLL
60.9%
(8.08%)</td><td>LHHL
60.1%
(8.22%)</td></tr>
<tr><td>LLHL
79.1%
(4.94%)</td><td>LLHH
74.7%
(4.74%)</td><td>LHHL
74.1%
(4.88%)</td><td>LHHH
70.5%
(5.47%)</td></tr>
<tr><td>HLLL
80.8%
(6.35%)</td><td>HLLH
72.9%
(6.58%)</td><td>HHLL
72.8%
(5.61%)</td><td>HHLH
65.6%
(7.72%)</td></tr>
<tr><td>HLHL
68.1%
(4.08%)</td><td>HLHH
68.5%
(6.77%)</td><td>HHHL
62.6%
(5.09%)</td><td>HHHH
66.5%
(4.62%)</td></tr>
</table>

8.2. Why are the improvements not as significant on the S-faces as compared to the D-faces?

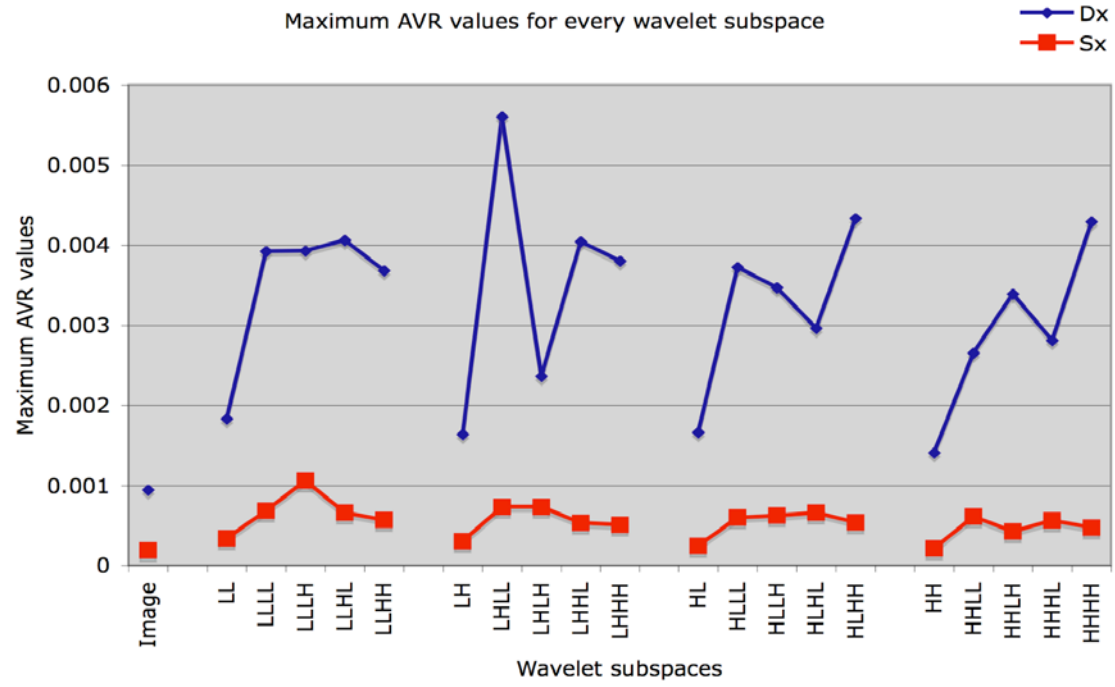
Major improvements have been noted when D Wavelet AsymmetryFaces are used in the classification compared to the **S** Wavelet AsymmetryFaces. From a high-level view of the system, one possible explanation is that the D-faces are still intensity-based images, which still contain signal information from the original intensity image. On the other hand, the way S-faces are defined in [7], they are basically a representation of the symmetry of the original intensity image. There is no signal information that is retained from the intensity images in the S-faces. Another plausible explanation for such a behavior is that S-faces are 'edge' images and therefore, the wavelet transforms which in a way retrieves the edge information (LH is vertical and HL is horizontal) from the signal is not doing anything much to the input signal. In short, applying some kind of edge filter onto an edge image is not extracting more discriminative feature from the original image.

To get a better view of how the discriminative power of the S-faces and the D-faces are behaving in the wavelet domains, the maximum of the AVR (9) values recorded at each subspace for all the Wavelet AsymmetryFaces, D, D_x, D_y, S, S_x and S_y are noted for 20 repetitions to see if the discriminative power changes by a lot down the levels of the wavelet tree. Then the average values of the maximum AVR are computed for those 20 repetitions and plotted in Figure 8 (a, b and c), which show the comparisons between D and S; D_x and S_x; and D_y and S_y respectively. From Figure 8a, D has much higher maximum AVR values compared to S, again confirming our intuitions about D having more consistent discriminative features. The maximum AVR values for D change by a lot while showing an overall increase from the image-intensity domain. The results support the explanation about why S does not produce any major improvement even in wavelet domains. As the wavelet transforms go down the levels the maximum AVR values of the S faces are almost constant (between 0.006 and 0.018) to decreasing, showing that no extra features are being extracted in the subspaces.

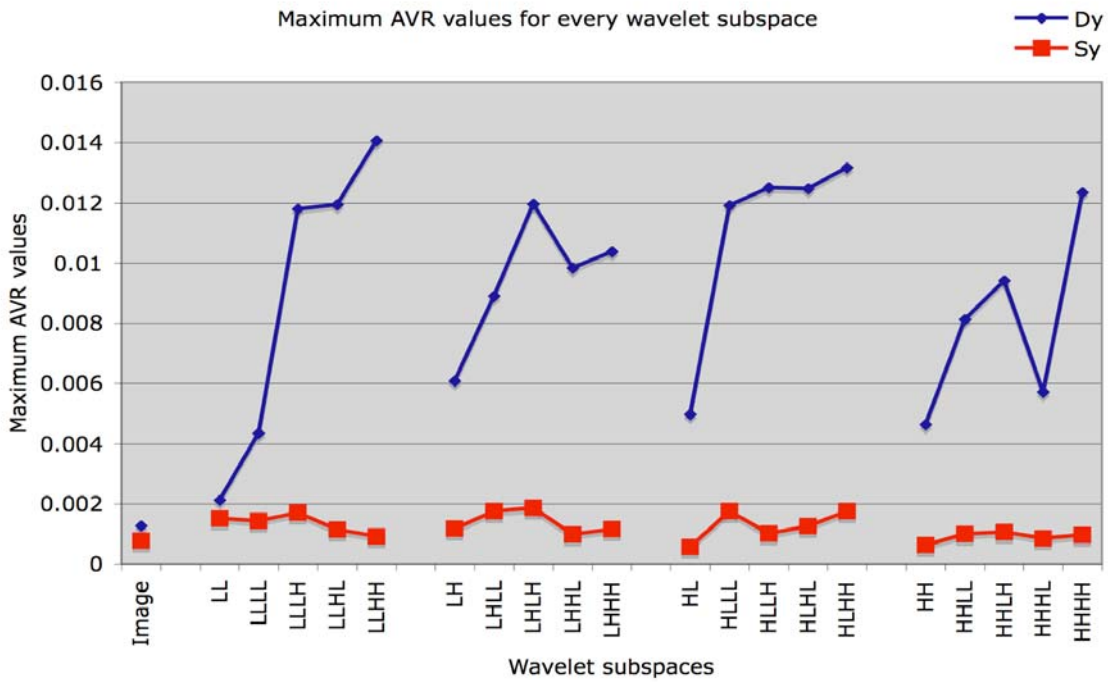
Similarly, when we compare Dx with Sx (Figure 8b) or Dy and Sy (Figure 8c), we can note similar behaviors where the maximum average AVR values for the D's increase by a lot from the spatial domain, while those for the S's show no major change. Another fact to note is that the maximum AVR values in Figure 8b and 8c are much smaller than those in Figure 8a indicating that the x and y faces of both D and S have smaller discriminative power, but enough to successfully classify expressions.


Maximum AVR values for every wavelet subspace
D
S
Maximum AVR values
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Image
LL
LLLL
LLLH
LLHL
LLHH
LH
LHLL
LHLH
LHHL
LHHH
HL
HLLL
HLLH
HLHL
HLHH
HH
HHLL
HHLH
HHHL
HHHH
Wavelet subspaces

(a)


Maximum AVR values for every wavelet subspace
Dx
Sx
Maximum AVR values
0.006
0.005
0.004
0.003
0.002
0.001
0
Image
LL
LLLL
LLLH
LLHL
LLHH
LH
LHLL
LHLH
LHHL
LHHH
HL
HLLL
HLLH
HLHL
HLHH
HH
HHLL
HHLH
HHHL
HHHH
Wavelet subspaces

(b)



(c)

Figure 8. (a) The average maximum AVR values of every subspace over 20 repetitions are reported by levels from the top level (Image intensity) to level 3 for D and S faces. (b) Comparison of AVR values between D_x and S_x faces. (c) Comparison of AVR values between D_y and S_y faces.

9. Conclusions

Previous work has already shown that AsymmetryFaces combined with standard classifier (Fisherface) have proved to be very effective at classifying expressions (joy, anger and disgust) as documented in [13]. The most important finding in this study is that Wavelet AsymmetryFaces can further improve classification of expressions. Wavelet AsymmetryFaces have discriminative features that are more prominent in certain subspaces that can enhance discrimination between the different facial expressions.

In this work, we have successfully investigated the implications of wavelet transforms on AsymmetryFaces. We have demonstrated that (1) by applying wavelet transforms on *D-faces*, a significant improvement can be achieved (Table 3); (2) certain subspaces of a wavelet tree have even more discriminative features compared to others, for instance higher frequency band (LH and HL from Table 4) of the wavelet tree; (3) the way *S-faces* are constructed, their image-intensity domain is already the optimal space with maximum discriminative features.

Wavelet transforms are definitely useful at extracting features that can be used to improve classification rates. They constitute an adaptive system that can be optimized for a specific purpose, and in this case classification of expression. Future work may involve the development of a pruning algorithm that can optimize the automatic selection of wavelet subspaces that will only contribute positively to the overall classification rate of expression.

References

[1] R. Thornhill and S. W. Gangestad. Facial attractiveness. *Trans. in Cognitive Sciences*, 3(12): 452-460, December 1999.

[2] C.K. Richardson, D. Bowers, R.M. Bauer, K.M. Heilman, C.M. Leonard, Digitizing the moving face during dynamic displays of emotion. *Neuropsychologia*, 38(7):1028-1039, 2000.

[3] R. Campbell. The lateralization of emotions: A critical review. *International Journal of Psychology*, 17:211, 219, 1982.

[4] L.G. Fakas. Facial asymmetry in healthy north American Caucasians, *Angle Orthodontist*, 51(1):70-77, 1981.

[5] 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.

[6] N.F. Troje, H.H. Buelthoff, How is bilateral symmetry of human faces used for recognition of novel views?, Vis. Res. 38 (1) (1998) 79-89.

[7] Y. Liu, K. Schmidt, J. Cohn and S. Mitra. Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding Journal*, 91(1/2):138-159, Special issue on face recognition, Martinez, Yang and Kriegman (Eds.) July/August 2003.

[8] C. Daniell, A. Mahalanobis, and R. Goodman, Object recognition in subband transform-compressed images by use of correlation filters, *Appl. Opt. 42*, 6474-6487 (2003).

[9] M. Vetterli and J. Kovačević, Wavelets and Subband Coding (Prentice-Hall, Englewood Cliffs, N.J., 1995).

[10] P. Hennings, J. Thornton, J. Kovačević, and B. Kumar, Wavelet packet correlation methods in biometrics, *Journal of Applied Optics*, special issue on Biometric Recognition Systems.

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

[12] R. Coifman, Y. Meyer, and V. Wickerhauser, Wavelet analysis and signal processing, in *Wavelets and Their Applications*, M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, eds. (Jones and Bartlett, Boston, Mass., 1992), pp. 153–178.

[13] S. Mitra and Y. Liu., Local facial asymmetry for expression classification. *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, June 2004.

[14] Y. Liu, K. Schmidt, J. Cohn and R. L. Weaver, Facial Asymmetry Quantification for expression invariant human identification. *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition (FG'02)*, May 2002.

[15] 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.