FACE RECOGNITION USING WAVELET REPRESENTATIONS OBTAINED FROM DIFFERENT PRUNING STRATEGIES

Kenny Teng, James Auwaerter

Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

We present a comparison between two major pruning strategies for face recognition using wavelet packets. The first approach is top-down, beginning with the image representation of the human faces and decomposing the image selectively based on correlation energy. The second approach begins with the full wavelet decomposition, designs a correlation filter for each subspace, and prunes the tree based on classification rate. Each technique has advantages: top-down is better at verifying whether an image belongs to a particular class while the full wavelet decomposition is better at determining which class an image belongs to out of a set of different classes. We attempt to generate trees that take advantage of both strengths to see at what middle ground we maximize effectiveness at image verification and image classification, using faces as the test images. Then, we apply the tree structure to sets of human faces to determine whether the tree structure is more broadly applicable.

1. INTRODUCTION

1.1. Motivation

While it is a relatively simple matter to either identify a human face or classify between different ones, it would be very interesting to see if we could use wavelet transforms to achieve better and more robust results. One of most characteristics of wavelet transforms is their ability to represent a signal into partitions of time-frequency plane. The popular representation of wavelet transforms is a multi-resolution wavelet tree where the each sub space contains information in a time-frequency domain. Therefore, we want to design figures of merit to take advantage of those wavelet spaces to end up with different pruned trees. The main motivation is to try to design a tree that is optimal for verification and one which is good for classification of face images. Their performance at verification and classification will then be measured using the figures of merit, and the strategies will be refined based on the results. In the end, we will come up with a combined tree that is able to perform well at both verification and classification.

1.2. Previous Work and State-of-the-Art

Previous work in this area includes: Hennings *et al.* [1] who experimented with correlation filters as the pruning strategy for fingerprint classification and Phillips [2] who designed two sets of filters for his work with face verification. Each filter set was designed with a single goal in mind: verification or classification. These filters were then tested on three different types of face image sets and compared across the types.

1.3. Short Overview of the Paper

We begin by introducing the basics of biometrics, wavelet transformations, and classification methods in Section 2. Our figures of merit and proposed methodology are described in Section 3. We then describe our results in Section 4. We end this paper with a brief conclusion.

2. BACKGROUND

2.1. Biometrics

Biometrics is the use of unique physical characteristics of people to positively authenticate a user of a system or recognize a type of a pattern and then identify people based on that pattern. Examples of biometrics for authentication include iris pattern, hand vein pattern, and fingerprints, while biometrics for recognition are primarily based on facial features. By identifying which characteristics are particular to a person, that person can be recognized in the future.

Currently, biometrics is not commonly used to authenticate users. The most common use, fingerprint scanners on laptops, is still limited to after-market peripherals and a few models of IBM laptops [3]. Two of the problems that prohibit their more widespread use are the difficulty of obtaining a clean image and the greater space that it takes to store an image as opposed to an alphanumeric password. While we can assume that a user will try to give a clean image in order to maximize recognition, different lighting conditions or fingerprint pressure can greatly affect the image that is compared to the database image. Authentication is composed of two parts: recognition that an image is being scanned, and classification of that image. For the most part, we can assume that the challenge lies in the classification of the image.

2.2. Wavelet Transformations

Wavelet transformations are a method of representing signals across space and frequency. The signal is divided across several layers of division in space and frequency and then analyzed. The goal is to determine which space/frequency bands contain the most information about an image's unique features, both the parts that define an image as a particular type (fingerprint, face, etc.) and those parts which aid in classification between different images of the same type.

One type of discrete wavelet transform (DWT) is the orthogonal DWT. The orthogonal DWT projects an image onto a set of orthogonal column vectors to break the image down into coarse and fine features.

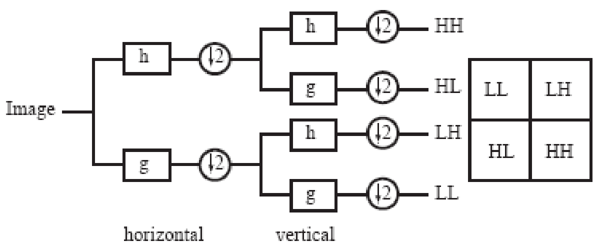

Figure 1 – A two-channel filter bank from [5].

In Figure 1, we see the order in which filters are applied to achieve a simple one-level wavelet decomposition. The filter h is a high-pass filter and the filter g is a low-pass filter.

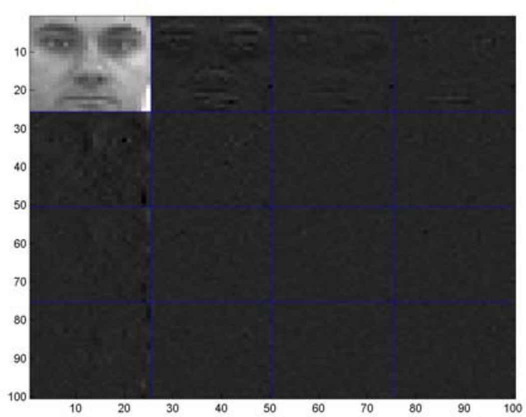

Figure 2 - An example of a full wavelet decomposition with two levels.

A typical two-level full wavelet decomposition is shown in Figure 2. Each of the subspaces is obtained by taking the original input and filtering it with a combination of high-pass and low-pass filters, designed to maximize the amount of information obtained within each subspace. This decomposition can be repeated for n-levels. The image can later be reconstructed from these subspaces. By removing subspaces that contain comparatively low amounts of information from a reconstruction, we can achieve an image that is nearly as good as the original but takes less space to store. This can be useful if we are storing a large number of similar images.

2.3. Classifier

There are many different classifiers out there that have proved to be very effective in classifying faces. We will be using advanced correlation filters, specifically the Minimum Average Correlation Energy filter (MACE). Correlation filter techniques are attractive candidates for the matching needed in face verification. According to Kumar *et al.* in [4], correlation filters can be used on any biometrics as long as they are in the form of images. Advanced correlation filters can offer a very good matching performance in the presence of variability such as facial expression and illumination changes. Furthermore, they are of less complexity and are shift invariant.

The main idea is to synthesize a filter using a set of training images that would produce correlation output that reduces the correlation values at locations other than the origin and this value at the origin is constrained to a specific peak value. When the filter is correlated with a test image that is authentic, the filter will exhibit sharp correlation peaks in the correlation plane. Otherwise the filter will output small correlation values. This correlation plane **v** can be modeled as

$$v_i = (DFT)^{-1} C_i DFTx \qquad (1)$$

where **x** is the vectorized form of the n image pixels, **DFT** is an n x n matrix containing the basis of a two-dimensional discrete Fourier transform and $\mathbf{C}_i$ is a diagonal matrix containing the correlation filter values for class i along the diagonal.

From the correlation plane, we can use a match metric from [1] by measuring the peak-to-correlation energy (PCE), which is as follows:

$$PCE(v_i) = \frac{\max(|v_i|) - mean(|v_i|)}{stdev(|v_i|)} \qquad (2)$$

The MACE filter seeks to minimize the average correlation plane energy:

$$E_{average} = \frac{1}{N}\sum_{i=1}^{N} E_i = \frac{1}{N}\sum_{i=1}^{N} h^+ D_i h = h^+ Dh\,, \quad (3)$$

where $$D = \frac{1}{N}\sum_{i=1}^{N} D_i \quad (4)$$

Minimization of $E_{average}$ is done while satisfying the linear constraints of the correlation values based on pre-specified values , that is in row vector *u*,

$$X^+ h = u \quad (5)$$

where *X* is a $d^2 \times N$ complex matrix, with the i^{th} column containing the 2-D Fourier transform of the i^{th} training image lexicographically re-ordered into a column vector. Lagrange multipliers method is used to minimize (3) while satisfying the linear constraints in (5) and this will give a closed form solution for the MACE filter as follows:

$$h = D^{-1} X (X^+ D^{-1} X)^{-1} u \quad (6)$$

3. PROPOSED WORK

3.1. Figures of Merit

We used two different figures of merit to evaluate the performance of the wavelet trees. The first one is based on the correlation energy of the filters at different wavelet sub-spaces and the other one is the false positive rate achieved by a wavelet-domain correlation filter.

3.1.1. Correlation Energy

According to Savvides in [7], if we use correlation filters as our main classifier, then a good figure of merit would be to measure the peak-to-sidelobe ratio of the correlation output to each subspace of each image. By substituting (6) into (3), we end up with:

$$\begin{aligned} \tilde{E} &= (D^{-1}X(X^+D^{-1}X)^{-1}u)^+ D(D^{-1}X(X^+D^{-1}X)^{-1}u) \\ &= u^T (X^+D^{-1}X)^{-1} u \end{aligned} \quad (7)$$

where **X** holds the spectrum of one of the training images after projecting the image onto the wavelet subspace in each column. Since better performance is obtained by lower energy, then we define our fitness metric as

$$F = 1/\tilde{E} \quad (8)$$

Therefore, the higher the fitness metric, the better the correlation filter is expected to perform. For a four-channel wavelet filter bank where a space V_0 is decomposed into four spaces, LL, LH, HL and HH, we would want to evaluate which level is performing better. In this case, we form the following inequality

$$F(V_0) > F(LL) + F(LH) + F(HL) + F(HH) \quad (9)$$

where the operator F(.) computes the fitness metric of the space. Here, LL, LH, HL and HH are the spaces of signals $x_{gg}^{(i+1)}$, $x_{gh}^{(i+1)}$, $x_{hg}^{(i+1)}$ and $x_{hh}^{(i+1)}$ respectively. If the left-hand side of the inequality is greater, we consider it as a better space to generate a correlation filter, and if the right-hand side is bigger, then four separate filters will be generated instead.

3.1.2. False Positive Rate

The second figure of merit is based on the fact that we are trying to optimize the tree for classification. Intuitively, by measuring the false positive rate that a given correlation filter yields, we can define a measure of performance. A false positive is recorded whenever, a test image is said to belong to a class other than its real class. Therefore, for a given space, V_0 and its corresponding correlation filter, the PCE value of all the classes other than the class it belongs to, are computed. Then, the performance inequality is as follows:

$$\max(PCE(V_{i\neq 0})) > \alpha E(PCE(V_0)) \quad (10)$$

where $0 < \propto < 1$ and E(.) operator is the expected value of the PCE value for space V_0. If the above inequality is satisfied, this means that the false positive rate of this correlation filter tends to be high, and therefore the discriminative power of that filter is not good enough.

3.2. Match Metric

In the evaluation stage, we need to compute a match metric when the wavelet-domain correlation filters are applied to a fully decomposed image. Therefore, after decomposing an image into a wavelet tree, the wavelet-domain correlation filters are applied to generate the correlation planes. Here, we can extend the PCE computation from (2) to calculate the sum of all the PCE values of all the wavelet sub spaces.

The formulation of this calculation is as follows. For a given image **x**, and wavelet packet correlation filters of class *i* contained in matrices $C_{i,l}$ for every subspace $_l$, the subspace correlation places are given by

$$v_{i,l} = (DFT)^{-1} C_{i,l}\ DFT\ WP_{i,l} x \quad (11)$$

and the match metric becomes

$$f(v_i) = \sum_l PCE(v_i) \qquad (12)$$

3.3. Methodology

We will be using the Cohn-Kanade AU-Coded Facial Expression database for testing our pruning strategies. For the purpose of verification and classification, Advanced Correlation filters, more specifically MACE will be used. The evaluation process will consist of two phases, namely training followed by testing phase. For the first stage, the first tree structure is going to be designed by starting in the image space and adaptively decomposing using wavelet transforms. At the same time, pruning will be done at every sub space using correlation energy based on inequality (9) as main figure of merit. The second strategy is that we start with the full decomposition and prune every single sub space using the second figure of merit defined in (10). The design of those two tree structures will be adaptive whereby we can alter the variables in our figure of merits to get those two representations as close as possible. In other words, we designed an optimal tree with the best verification rate and at the same time requiring the minimum spanning tree. The way to do that is that we build a combined tree, which has a correlation filter at a sub space on both trees. Once we have the combined tree, we can begin the testing phase. We tested it by applying those wavelet-domain correlation filters on the expression database to evaluate the effectiveness of our method. As a control test, standard correlation filters have been generated in the image-intensity domain only and their overall performances for verification and classification have been computed. We will compare our results to those values.

4. EXPERIMENTAL RESULTS

4.1. Datasets

The database we used for evaluation is the Cohn-Kanade AU-Coded Facial Expression database.

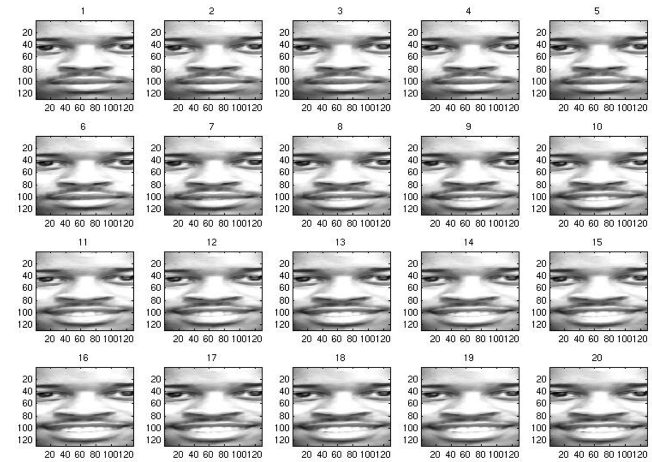

Figure 3 – An instance of a particular class from the expression database.

In the database we used, there are 20 classes (persons) with 20 different faces each, with variations facial expressions.

4.2. Evaluation

For verification evaluation of our method, a receiver operating characteristics (ROC) curve has been plotted to show the probabilities of authentic versus the probabilities of impostors. From the following graph, it is clear that our optimized tree achieved better results (blue line) than the control test (red line). The probability of having an authentic verification with zero probability of having an impostor is about 83% for wavelet-domain method and 67% for the standard filters.

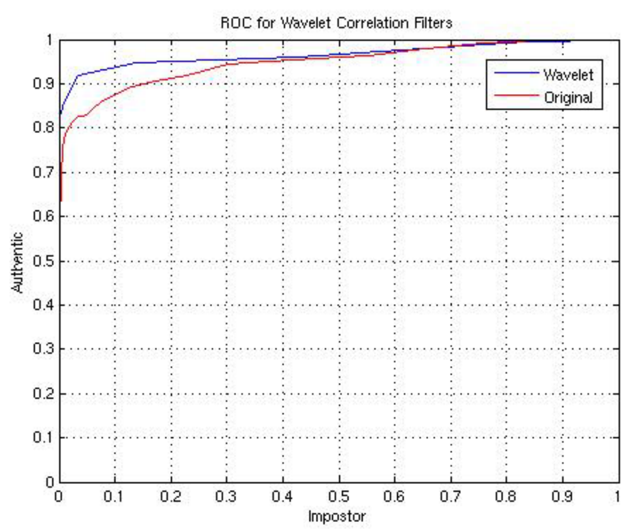

Figure 4 – ROC curve showing the results for verification for both wavelet-domain and standard correlation filters.

As for classification results, the error rates have been plotted for every single class as shown in figure 5. There is a clear reduction in classification error rate for the wavelet-domain filters (red) compared to the standard correlation filters. For class 4 where the error rate for the control test reaches 70%, it is still hard for even the wavelet-domain filters to classify efficiently, even though they yielded a better rate of 65%. However, for classes like 13 and 17 where the standard method had error in the order of 35% and 10% respectively, the wavelet-domain method achieves a zero error rate. For the overall classification rate, our wavelet-domain method yields 94.2% accuracy while the standard method achieves 90%.

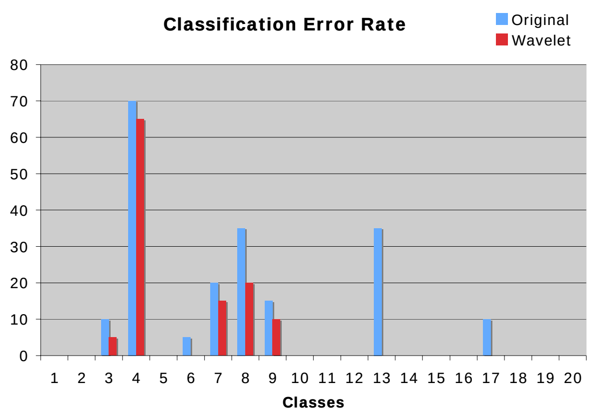

Figure 5 – Classification error rate of both wavelet-domain and standard correlation filters

CONCLUSIONS

We have discussed the basic elements of biometrics and wavelet transforms, and how correlation filters may be used to classify images within a biometric system. We explored the advantage of using wavelet packet decomposition for verification and classification and determined how to best use our figures of merit to obtain an optimal wavelet decomposition tree. The combined wavelet tree performed better than the standard correlation filters applied only in the image-intensity domain. The results show that face images have some features that remain more consistent in the wavelet sub spaces than in the spatial domain. Future work may involve designing different figure of merits that will be tailored to specific datasets.

ACKNOWLDGEMENTS

The author would like to thank Jelena Kovačević, Pablo Hennings, and Amina Chebira for their help in conducting the experiments and the class of 18-795 for their feedback during the class and presentations.

REFERENCES

[1] 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.

[2] Phillips, P.J., “Matching pursuit filters applied to face identification.” *IEEE Transactions on Image Processing*, Volume 7, Issue 8, Aug. 1998 Page(s): 1150 – 1164

[3] “IBM Introducing Fingerprint Reader into Laptop”; Germain, Jack: Written 10/04/04, Accessed 12/01/05; http://www.technewsworld.com/story/37017.html

[4] B. V. K. Vijaya Kumar, A. Mahalanobis, S.R.F. Sims, J.F. Epperson, "Unconstrained correlation filters," *Applied Optics*, vol. 33, pp. 3751-3759, 1994.

[5] A. Chebira, G. Srinivasa, J. Kovačević, “A wavelet packets approach to classification of subcellular protein images.” Unpublished work, October 2004

[6] M. Vetterli and J. Kovačevic, “Wavelets and Subband Coding,” Prentice-Hall Signal Processing Series, 1995.

[7] Marios Savvides, B.V.K. Vijaya Kumar, P.K. Khosla, "Illumination Tolerant Face Identification using Advanced Correlation Filters trained from a single face image," accepted for publication in Biometrics Consortium Conference BC2004.

[8] R. Duda, P. Hart, D. Stork, *Pattern Classification*, Wiley, New York, 2000