Benchmark between Different Feature Extraction Methods Applied to Face Recognition

Giovanni Mejía, Mauricio Martinez, Roberto A. Vázquez

Universidad La Salle, Benjamin Franklin 47, Mexico City, Mexico.
rgmb@lasallistas.org.mx, mauricio.martinez@lasallistas.org.mx, ravem@lasallistas.org.mx

Abstract. Face identification systems rely on feature extraction from still images. There exist many techniques for this purpose, each one having its own inherent limitations and impact on the system's performance in terms of recognition rate. In this paper, we describe a facial recognition system that can be used with different signal preprocessors for feature extraction, and compare different well known approaches, including the use of neural networks. We also introduce the use of two -dimensional wavelets as feature extractors. Our results show a competitive performance. We conclude that wavelet based signal preprocessors constitute an interesting alternative for facial recognition, when combined with neural network classification.

Keywords. Face recognition, wavelet transform, DFT, DCT, PCA, DWT, neural networks, image classification, computer vision.

1 Introduction

The field of image processing and computer vision is an expanding area, which presents many unsolved problems. One of the relevant problems in engineering is adapting computer vision to specific systems that imply first of all to identify a problem that can be solved with automated artificial vision. It is first necessary to comprehend the fundamental capabilities and limitations of the image processing and computer vision systems, as well as the knowledge of the fundamental image processing techniques and algorithms.

Human beings have always developed different systems to confirm the identity of people with techniques like personal identification numbers, signatures, etc. But the applications of biometric techniques such as fingerprint and retina analysis were implemented until the 1980s [1]. In this work, we use the preprocessing of images applied to face recognition; nonetheless, the techniques exposed can be applied to the solution of similar problems like object recognition, crop recognition and other areas related to digital images.

A general statement to describe the problem we are studying could be formulated in an axiomatic manner: "Given a set of images, identify or verify a person on the

Advances in Computer Science and Engineering
Research in Computing Science 63, 2013 pp. 3–12
Paper Recived 16-08-2013 and Acepted 21-10-2013

scene using a given database of faces". We pretend to find the performance of each wavelet family. To achieve this, we have proposed the following particular objectives.

1. Evaluate the performance of the wavelet transform as a feature extraction technique with different classification systems of face images.
2. Explore the representation space of face images by the application of different wavelet families.
3. Develop distance classifiers (Euclidian, Mahalanobis and Manhattan) and neural networks to perform face recognition.
4. Use the previously proposed classification techniques with preprocessed images by wavelet families, a standard database will be used (ORL database) and their performance will be evaluated.
5. Different strategies of feature extraction will be compared like DFT (Discrete Fourier Transform), PCA (Principal Component Analysis), DCT (Discrete Cosine Transform), with the selected wavelet families.

In table 1 we present some investigations with relevant results on the face recognition area.

Table 1. Comparison of remarkable results on the state of the art.

Method	*Recognition Percentage*	*Characteristic*	*Ref*
PCA	85%	High Computational Cost	[2]
Kernel PCA	84%	High Computational Cost	[2]
Kernel PCA Gauss	85%	High Computational Cost	[2]
DCT Geometric Regions	97%	Hybrid Process	[3]
Wavelet + PCA	92%	Hybrid Process	[4]
Wavelet Sym + PCA	90%	Hybrid Process	[5]
Wavelet + DCT + PCA	95%	Hybrid Process	[6]

2 Feature extraction techniques

In this section, we describe the techniques used to develop this work. If the input data to an algorithm cannot be directly processed, we have to transform it to a reduced set of characteristics. This process is known as feature extraction. If the characteristics are chosen on a careful way, we can expect those to have relevant information from the input data, which allows to perform the desired actions using this reduced representation. The specific techniques used for this investigation are as follow.

Because we will be using static images we require a version of the Fourier transform that can handle this, that is why we used the two-dimensional ***Discrete Fourier Transform***, this is possible because the pixels have indexed bidimensional coordinates. Fourier Features are used as a representation instead of using the complete image. This is very useful to process a smaller quantity of characteristics [7].

The ***Principal Component Analysis*** is a feature extraction technique used to perform pattern recognition in a given data set, and to express that data set in a way that the similarities and differences are highlighted. It is a very useful tool to transform data and make it easier to be separated on different classes, which is equivalent to transform its representation space. This technique also allows a high data compression rate [8].

The ***Discrete Cosine Transform*** is a linear transformation that can express a finite sequence of speech data as a sum of cosine functions spanning different frequencies. After the original signal has been transformed, its DCT coefficients reflect the importance of the frequencies present on it, the first coefficient refers to the lower frequency on the signal and it usually contains most of the relevant information of the signal, and the last coefficient refers to the higher frequencies of the signal. For this work we used the two-dimensional DCT, which is possible because every pixel has bidimensional indexed coordinates [7].

As with the DFT, we have the ***Discrete Wavelet Transform***, which can be calculated at some point on the time – scale space, there is no need to know the values of the function for the axis at all times, since all we need is the function values on the point in which the value is different from zero. Consequently the evaluation of the wavelet transform can be done almost in real time. The wavelet families used on this research are Daubechies, Symlets, Coiflets, Biorthogonoal and rBiorthogonal.

The discrete values of the scalar parameter a and traslation parameter b are taken in a different manner. a will be taken as 2^{-s} and b as $k2^{-s}$, where $k, s \epsilon \mathbb{Z}$ with the values of a and b, the integral of the continuous wavelet transform is transformed:

$$W_{\psi}f(k2^{-s}, 2^{-s}) = 2^{s/2} \int_{-\infty}^{\infty} f(t)(2^s t - k)dt \tag{1}$$

Now the function $f(t)$ must be discrete. By simplicity we assume the sampling rate is 1. In that case the expression can be written on this way:

$$W_{\psi}f(k2^{-s}, 2^{-s}) \approx 2^{s/2} \sum_n f(n)\psi(2^s n - k) \tag{2}$$

To calculate the wavelet Transform of a function at some point on the time – scale space, it is not necessary to know the values of the function for the complete time axis. All we need to know is the function on that values of time in which the wavelet is different from zero. Consequently the evaluation of the wavelet Transform can be done almost in real time. [5] Each wavelet Transform depends of its wavelet or mother function represented by “ψ" [9].

3 Classifiers

As it has been mentioned, pattern recognition is a discipline with the objective of classifying objects in a finite number of classes or groups. Such objects can be audio files, pictures or any type of data that can or needs to be classified. The purpose of the classifiers in this work is to separate on different classes (also known as clusters) the groups of images of the desired database. We will use three distance classifiers which are Euclidian, Mahalanobis and Manhattan as well as a Backpropagation Neural Network.

Euclidian Distance refers to the ordinary distance between two points generated when they are joint by a straight line, it is based on the Pythagoras' theorem, and when it is used as a distance it becomes a metric [10].

Mahalanobis Distance measures the distance of two groups of objects based on its means. Groups must have the same number of characteristic features but no necessarily the same quantity of elements [11].

Manhattan Distance is a metric in which the distance between two points equals the sum of the differences of its coordinates. It is also known as distance 1, norm 1 or Manhattan distance. The distance between two vectors in a vector space is the sum of the distances of the projections of line segments between its coordinate axes. [11]

The ***Artificial Neural Networks*** are an information processing paradigm inspired on the human brain. The key element on this is its structure, the Artificial Neural Networks are composed of a number of processing elements or neurons which work together to resolve a specific problem. The ARN used on this work are based on the McCulloch & Pitts model. Inside of a neural network there exist a great number of connections between the neurons, because each neuron has synaptic weighs which simulate the neural interconnections of the brain [12].

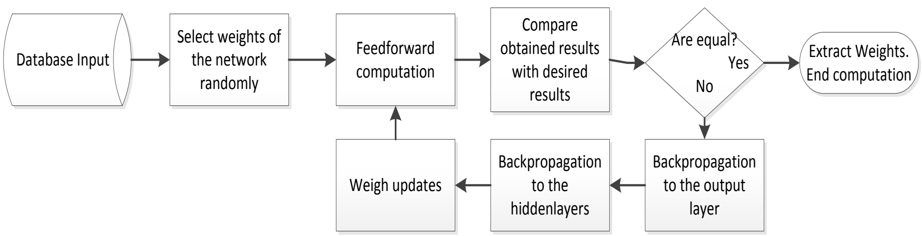

Fig. 1. Artificial Neural Network Algorithm

4 Proposed methodology

On this section we explain which tools were used to the development of this work, a general methodology is presented for wavelet feature extraction process and neural network classifier having created a benchmark based on classical feature extraction techniques (PCA, DCT and DWT). The methodology used in our research is summarized as follows:

1. Select database.
2. Apply different wavelet transforms to the database.
3. Develop and train recognition systems using a portion of the preprocessed database.

4. Test the systems with the remaining portion of the database.
5. Compare the results with the classical transforms (PCA, DCT and DWT).
6. Demonstrate based on results which wavelet transform presents the best performance.

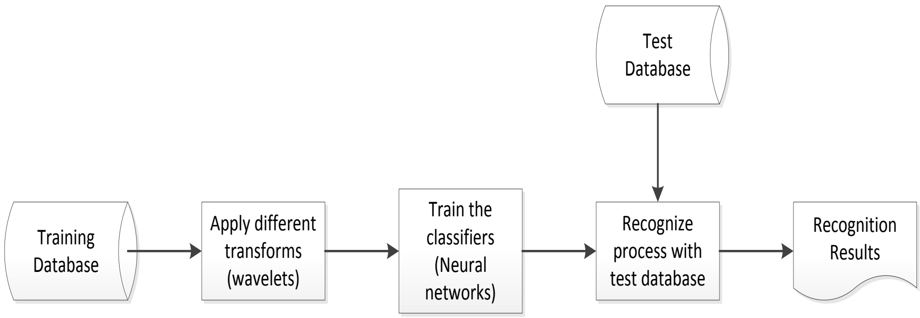

Fig. 2. General process for recognition systems.

This project is limited to implement offline face recognition on static images, with a standard database. All other implementations as a real system will be considered as future work. It should be noted that the intention of this research is to suggest a benchmark for the usage of different feature extraction techniques on images, and to test their performance with different classification systems. The main techniques used to establish the results of this work are described in the next two algorithms.

Wavelet algorithm

```
1. Obtain database specifications.
2. Obtain the list of images to transform.
3. Create the loading architecture of the images. In this
case we will use an array of three dimensions (image,
row, column).
4. Select wavelet and iteration depth.
5. Obtain the discrete wavelet transform of each image
and place it on the same architecture of three dimen-
sions.
6. If iterating, do it with the approximation coeffi-
cients.
7. Define which images will be used for training and
which will be used for testing.
8. Transform the images to row vectors.
9. Create two matrixes, the first one will contain the
vectors of the training images and the second will con-
tain the vectors of the test images.
```

Neural Network

```
1. Transform the input data on a matrix where rows are
the images and the columns are the features of the imag-
es.
2. Normalize the features between zero and one
3. Select training set.
4. Initialize weights and biases.
5. Execute in parallel the training procedure with a min-
imum of four processes with the same data but different
learning rate with the training set.
6. Iterate until the output is equal to the desired out-
put, while iterating, transfer the weights with best re-
sults to the other neural networks.
7. Insert test group to the Neural Network and classify.
Compare obtained results with desired.
```

5 Experimentation and results

This section has the objective to perform a series of experiments with the purpose of generating a benchmark between the classical feature extraction techniques (PCA, DCT, DFT) and wavelet transform families, all this is evaluated with distance classifiers and neural networks. First of all, all feature techniques were carried on to determine the quantity of features required for each technique according to a representation percentage or an iteration level. Then, using the necessary quantity of features, all distance classifiers were tested and we obtained a matrix of performance for each feature extraction process and classifier tested, based on this, the wavelets with the best results were picked and processed with a neural network classifier, with the obtained results, a specific wavelet is purposed to solve the face recognition problem with a specific database.

Our experiments have been carried out, using the ORL database. It has 400 images divided on 40 subjects with ten images each. Some images have been shot with different kinds of illumination, facial expressions, and face details like with or without sunglasses. All the images were shot with a black homogeneous background with the subjects on frontal position with soft lateral variations.

The experiments were carried using five images to train the classifiers and the five remaining to test the system. For each series, the training images were chosen randomly, and the remaining five will integrate the test group, the results reported are an average of 30 repetitions for each experiment.

First of all, the feature extraction processes DCT, DFT and PCA were used. The results showed that the quantity of characteristics to represent an image can vary from the 0.25% of the original to 14.91% depending on the technique and the percentage of representation needed.

Table 2. Quantity of features required to represent images according to the representation percentage of the image

	Quantity of features required to represent images					
	Quantity of features			*Percentage of features*		
%	DCT	DFT	PCA	DCT	DFT	PCA
75	55	55	26	0.53%	0.53%	0.25%
80	64	64	36	0.62%	0.62%	0.35%
85	74	73	52	0.72%	0.71%	0.50%
91	104	112	82	1.01%	1.09%	0.80%
92	122	141	89	1.18%	1.37%	0.86%
93	142	169	97	1.38%	1.64%	0.94%
94	168	214	106	1.63%	2.08%	1.03%
95	202	265	116	1.96%	2.57%	1.13%
96	259	365	127	2.51%	3.54%	1.23%
97	353	533	139	3.43%	5.17%	1.35%
98	535	819	155	5.19%	7.95%	1.50%
99	1035	1536	173	10.04%	14.91%	1.68%

Feature extraction with the different wavelet families can be done with 0.04% of features from the original image to 34.70%. The results with the test sets of images show that PCA method offers best percentage of face recognition (Table 3), the PCA method requires less quantity of features to train each classifier; DFT and DCT methods have high error percentages. To illustrate this, we can use the Manhattan classifier with data preprocessed by the PCA technique. With this, we were able to recognize 89% of the subjects from the database, from a representation level of 80% of the image. The disadvantage of this method is that it presents a computational complexity $O(n^2)$. For this reason, it is not functional for huge databases.

We carried out the wavelet experiments by using the DWT [9] for different wavelet families. From these experiments, we see that the methodology achieves great data compression and have less error percentage on the training step. We have established as benchmark PCA technique which has given a 6.5% of error in classification using only 173 features per image. The results according this benchmark go between a 56.5% of recognition error and 10% depending on the selected classifier and wavelet family.

We have selected the best results for each wavelet family based in the previous tests and using the classifier system ANN. In Table 4 we show the number of required features needed to represent each image and to achieve the smaller recognition errors presented in Table 5.

Table 3. Smaller error percentage obtained for each feature extraction technique using DWT with different wavelet families.

	DCT	*DFT*	*PCA*	*db 2,4*	*sym 2,4*	*coif 1,4*	*bior 2.2,4*	*Rbio 3.1,4*
Euclidian	47%	37.5%	9%	10%	10.5%	11%	10.5%	10%
Mahalanobis	78%	68%	12.5%	15%	24%	25%	15.5%	15%
Manhattan	47%	37%	6.5%	9.5%	12%	12%	10.5%	10%

Table 4. Number of features required by wavelet family to successfully represent an image.

Iter	*db2,4*	*sym2,4*	*coif1,4*	*bior2.2,4*	*rbio3.1,4*
4	0.699%	0.699%	1.068%	1.068%	0.699%

Table 5. Smaller recognition errors per wavelet family using DWT

Iter	*db2,4*	*sym2,4*	*coif1,4*	*bior2.2,4*	*rbio3.1,4*
4	15.500%	15.500%	7.000%	12.000%	13.000%

Based on these results, we can affirm that the wavelet who has the best performance to carry out face recognition is the Coiflets1on iteration level 4, this configuration only reported 7% of classification errors on the test set.

6 Conclusion

It can be concluded that based on the recognition results that images processed with PCA offer better results. Still, this procedure has a high computational cost because it requires calculating the covariance matrix of the features each time a subject is added to the database, which possesses a quadratic complexity. This cannot be applied to real applications because the quantity of features shall significantly increase, and the processing time of that matrix would be too high for real time applications with current technology.

The performance of wavelets as feature extraction technique was completely satisfactory, because superior percentages to 90% in recognition on the test sets were reached. This was achieved using a reduced quantity of features, in relation with the original image; only 1.1% of features were used to obtain excellent results on face recognition. The classifiers used in this work have been the fundamental tool to establish the comparatives between the different feature extraction methods. Its development and implementation have impacted directly on the results because on the development of this work, we have proposed a training method for neural networks which maximizes the usage of the processor, and considerably reduces the training time.

We can conclude that the neural networks method combined with wavelet transform are applicable to the implementation on real applications because, if we increment the database, we only have to train the neural network again. Nonetheless, to increment the number of subjects on PCA we have to recalculate the covariance matrix which might become incalculable. The purposed methodology shows better performance than the one presented on the state of the art, because it does not combine feature extraction techniques like wavelets + PCA, DCT + PCA to mention some. It only uses one technique of feature extraction at the time and with that the percentages of error obtained are considerably low. Based on the results, it has been established that the wavelet with best face recognition results using traditional classification systems and with neural networks is the Coiflets 1, followed by the Biorthogonal 2.2, in third place we can find rBiorthogonal 3.1 and in fourth place are Daubechies 2 and Symlets 2. For all this wavelet Families the iteration level used was 4 which can achieve a reduction of features to process inferior to 2%.

Acknowledgements

Giovanni Mejía wants to thank to CONACYT for providing the financial resources as a scholarship to pursue my Master of Science degree. Also I would like to thank to my advisor and co-advisor for their commitment in the development of this research.

The authors would like to thank to Universidad La Salle for providing the support under the projects I-064/12 and I-061/12.

References

1. Wayman J. A., Maltoni D., Maio D.: Biometric Systems. Springer, UK (2005).
2. Hala M. E.: Feature Extraction using PCA and Kernel -PCA for Face Recognition, The 8th International Conference on INFOrmatics and Systems (2012).
3. Omaia D., Poel J., Batista L.: 2D DCT Distance Based Face Recognition Using a Reduced Number of Coefficients, XXII Brazilian Symposium on Computer Graphics and Image Processing. (2009).
4. Xian L., Sheng Y., Qi W., Ming L.: Face Recognition Based on Wavelet Transform and PCA, Pacific-Asia Conference on Knowledge Engineering and Software Engineering. (2009).
5. Puyati W., Walairacht S. Walairacht A.: PCA in Wavelet Domain for Face Recognition, ICACT (2006).
6. Abouzar P., Yousefi S., Setarehdan S.: Hybrid WT Based DCT Based Face Recognition, IEEE International Conference on Signal Processing and Communications (2007).
7. Theodoridis S., Koutroumbas K.: Pattern Recognition. Elsevier, USA (2003).
8. Abdi H., Williams L. J.: Principal Component Analysis, USA (2010).
9. Mallat S.: A wavelet tour of signal processing: The sparse way, Academic Press, USA (2008).
10. Saad Y.: Iterative methods for sparse linear systems, Society for Industrial and Applied Mathematics; USA (2003).
11. Webb A. R.: Statistical Pattern Recognition, QinetiQ Ltd, UK (2002).
12. Rojas R.: Neural Networks, Springer, Berlin (1996).