Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 854–858
Pattern Recognition Letters
www.elsevier.com/locate/patrec

An efficient face verification method in a transformed domain

Marcos Faundez-Zanuy [b,*], Josep Roure [b], Virginia Espinosa-Duró [b], Juan Antonio Ortega [a]

[a] *Departamento de Ingeniería Electrónica, UPC Terrassa, Spain*
[b] *Escuela Universitaria Politécnica de Mataró, Adscrita a la UPC, Avda. Puig i Cadafalch 101-111, 08303 Mataró (Barcelona), Spain*

Received 14 June 2005; received in revised form 16 June 2006
Available online 1 February 2007

Communicated P.J. Flynn

Abstract

In this paper we propose a low-complexity face verification system based on the Walsh–Hadamard transform. This system can be easily implemented on a fixed point processor and offers a good compromise between computational burden and verification rates. We have evaluated that with 36 integer coefficients per face we achieve better Detection Cost Function (6.05%) than the classical eigenfaces approach (minimum value 6.99% with 126 coefficients), with a smaller number of coefficients.

Keywords: Face verification; Transform methods; Biometrics

1. Introduction

In the last decade significant advances have been achieved on biometrics, especially on face recognition (Jain et al., 1999). This has been possible due to the increase of computational power of the state-of-the-art computers. However, there are several application scenarios where a low-complexity algorithm, which can be implemented on a low-cost processor is desirable. Some examples of this situation are mobile telephone, PDA or standalone control access systems. Probably in these situations the processor will be a fixed point one, and the number of operations per second smaller than the state-of-the-art processors used to develop the best algorithms available nowadays.

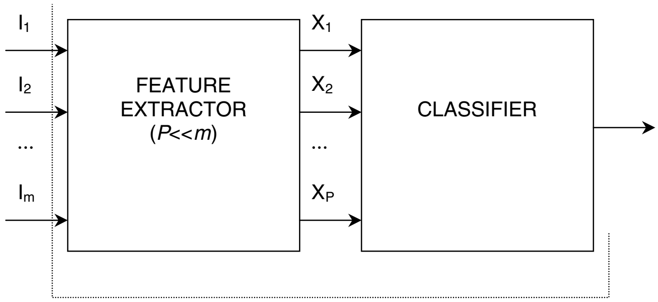
2. Face recognition

Usually, a pattern recognition system consists of two main blocks: feature extraction and classifier. Fig. 1 summarizes this scheme. On the other hand, there are two main approaches for face recognition:

(a) Statistical approaches consider the image as a high-dimension vector, where each pixel is mapped to a component of a vector. Due to the high-dimensionality of vectors some vector-dimension reduction algorithm must be used. Typically the Karhunen–Loeve transform (KLT) is applied with a simplified algorithm known as eigenfaces (Turk and Pentland, 1991). However, eigenfaces algorithm is suboptimal approximation to KL transform. Nowadays, with the improvements on computational speed and memory capacities, it is possible to compute the KLT directly, but computational burden and memory requirements are still important. In order to alleviate this problem we have resized the original images from 112×92 to 56×46 for the KLT results. KLT approach is similar to eigenface assuming zero-mean (using a correlation matrix instead of a covariance matrix to do eigen analysis and find eigenvector basis representation).
(b) Geometry-feature-based methods try to identify the position and relationship between face parts, such as eyes, nose, mouth, etc., and the extracted parameters are measures of textures, shapes, sizes, etc. of these regions.

* Corresponding author. Tel.: +34 93 757 44 04; fax: +34 93 757 05 24.
E-mail address: faundez@eupmt.es (M. Faundez-Zanuy).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2006.12.005



In this paper, we mainly focus on the study of the feature extraction for the face recognition using statistical approaches.

3. Walsh–Hadamard transform

The Walsh–Hadamard transform basis functions can be expressed in terms of Hadamard matrices. A Hadamard matrix H_n is a $N \times N$ matrix of ± 1 values, where $N = 2^n$.

In contrast to error-control coding applications, in signal processing it is better to write the basis functions as rows of the matrix with increasing number of zero crossings.

The ordered Hadamard matrix can be obtained with the following equations (Gonzalez and Woods, 1993):

$$H(x,u) = \frac{1}{N}(-1)^{\sum_{i=0}^{n-1} b_i(x)p_i(u)}$$

where $b_k(x)$ is the kth bit in the binary representation of x.

$$\begin{aligned}
p_0(u) &= b_{n-1}(u)\\
p_1(u) &= b_{n-1}(u) + b_{n-2}(u)\\
p_2(u) &= b_{n-2}(u) + b_{n-3}(u)\\
&\vdots\\
p_{n-1}(u) &= b_1(u) + b_0(u)
\end{aligned}$$

where the sums are performed in modulo-2 arithmetic.

For example, for $n = 3$ the ordered Hadamard matrix is

$$H_3 = \frac{1}{\sqrt{8}}\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1\\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1\\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1\\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1\\
1 & -1 & -1 & 1 & -1 & 1 & 1 & -1\\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1\\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1
\end{bmatrix}$$

The two-dimensional Hadamard transform pair for an image U of $2^n \times 2^n$ pixels is obtained by the equation $T = H_n U H_n$. We have zero padded the 112×96 images to 128×128. Thus, in our experiments $n = 7$.

Gibson et al. (1998) summarizes two measures that indicate the performance of transforms in terms of energy packing efficiency and decorrelation efficiency. It can be observed that the performance of the Walsh–Hadamard transform (WHT) is just a little bit worse than discrete cosine transform (DCT) and Karhunen–Loeve transform (KLT).

The WHT is a fast transform that does not require any multiplication in the transform calculations because it only contains ± 1 values. This is very suitable for fixed point processors because no decimals are produced using additions and subtractions. Table 1 compares the computational burden of KLT, DCT and WHT (Jain, 1989). It is interesting to observe that when dealing with DCT and WHT, basis functions are known in advance (they are not data dependent). In addition, it is important to emphasize that referent to performance gain, the transform choice is important if block size is small (Jain, 1989), say $N < 65$. This is not our case, because we consider each image as a block of about 10,000 components.

Table 2 provides execution times using a Pentium 4 processor at 3 GHz.

We can define a zonal mask as the array $m(f_1, f_2) = \begin{cases} 1, & f_1, f_2 \in I_t \\ 0, & \text{otherwise} \end{cases}$, and multiply the transformed image by the zonal mask, which takes the unity value in the zone to be retained and zero on the zone to be discarded. In image coding it is usual to define the zonal mask taking into account the transformed coefficients with largest variances. Then the zonal mask is applied to the transformed image (or blocks of the image) and only the nonzero elements are encoded. In our case we will not take into account the variances of the transformed coefficients and we will just define the zonal mask in the following easy ways:

(a) Rectangular mask: it will be a square containing $N' \times N'$ pixels.
(b) Sectorial mask: it will be a sector of 90° of a r radius circle.

Fig. 2 shows one example of each situation.

Table 1
Computational burden of KLT, DCT and WHT for images of size $N \times N$

Transform	Basis function computation	Image transformation
KLT	$O(N^3)$ (to solve 2 $N \times N$ matrix eigenvalue problems)	$2N^3$ multiplications
DCT	0	$N^2 \log_2(N)$ multiplications
WHT	0	$N^2 \log_2(N)$ additions or subtractions

Table 2
Execution time for KLT, DCT and WHT

Transform	Basis function computation	Image transformation (s)
KLT	347.78 s	0.23
DCT	0	0.0031
WHT	0	0.0003

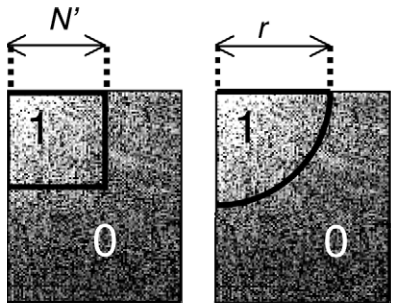


Fig. 2. Example of rectangular and sectorial masks.

This definition lets to easily obtain the coefficients. The dimension of the resulting vector is $N' \times N'$ for the rectangular mask and, for the sectorial mask, the number of pixels meeting the following condition:

$$\vee f_1, f_2, \text{ if } \sqrt{(f_1 - c_1)^2 + (f_2 - c_2)^2} < \text{radius}$$
$$\text{then } m(f_1, f_2) = 1; \text{ else } m(f_1, f_2) = 0$$

where the coordinates of the center are the frequency origin ($c_1 = c_2 = 0$).

In our experiments we have obtained similar results with both kinds of masks. Thus, we have chosen the rectangular mask because it is easier to implement.

It is interesting to observe that in image coding applications the image is split into blocks of smaller size, and the selected transformed coefficients of each block are encoded and used for the reconstruction of the decoded image. In face recognition all the operations are performed over the whole image (it is not split into blocks) and all the computations are done in the transformed domain. Thus, it is not necessary to perform any inverse transform. On the other hand in image coding the goal is to reduce the amount of bits without appreciably sacrificing the quality of the reconstructed image, and in image recognition the number of bits is not so important. The goal is to reduce the dimensionality of the vectors in order to simplify the complexity of the classifier and to improve recognition accuracy.

4. Results

This section evaluates the results achieved using the WHT and compares them with the classical KLT, eigenface, and DCT methods.

4.1. Database

The database used is the ORL (Olivetti Research Laboratory) faces database (Samaria and Harter, 1994). This database contains a set of face images taken between April 1992 and April 1994 at ORL. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department.

There are ten different images of each of the 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

Our proposed algorithm is mainly though for low-cost fixed-point processors, implemented in systems with reduced amount of storage capacity. Thus, the application would be restricted to a reasonable reduced amount of users. For this reason, we have selected the well-known ORL database. In fact, we have already executed this algorithm on an ARM processor at 400 MHz. However, it must be studied whether this database can produce statistically significant results or not. In (Guyon et al., 1998) the minimum size of the test data set, N, that guarantees statistical significance in a pattern recognition task is derived. The goal in the abovementioned work is to estimate N so that it is guaranteed, with a risk α of being wrong, that the error rate P does not exceed that estimated from the test set, $\widehat{P}$, by an amount larger than $\varepsilon(N, \alpha)$, that is,

$$\Pr\{P > \widehat{P} + \varepsilon(N, \alpha)\} < \alpha$$

Letting $\varepsilon(N, \alpha) = \beta P$ and supposing recognition errors as Bernoulli trials (i.i.d. errors), we can derive the following relation after some approximations:

$$N \approx \frac{-\ln \alpha}{\beta^2 P}$$

For typical values of α and β($\alpha = 0.05$ and $\beta = 0.2$), the following simplified criterion is obtained:

$$N \approx \frac{100}{P}$$

If the samples in the test data set are not independent (due to correlation factors that may include variations in recording conditions, in the type of sensors, etc.), then N must be further increased. The reader is referred to Guyon et al. (1998) for a detailed analysis of this case, where some guidelines for computing the correlation factors are also given.

In our experiments, we are making for each user, all other users' samples as impostor test samples, so we finally have, that $N = 40 \times 5$ (client) $+ 40 \times 39 \times 5$ (impostors) $=$ 8000. So, with 95% confidence, our experiments guarantee statistical significance in experiments with an empirical error rate, $\widehat{P}$, down to 1.25%, which is certainly suitable for our experiments.

4.2. Conditions of the experiments

Our results have been obtained with the ORL database in the following situation: 40 people, faces 1–5 for training, and faces 6–10 for testing.

We obtain one model from each training image. During testing each input image is compared to all the models within the database ($40 \times 5 = 200$ in our case) and the model close to the input image (using for instance the Mean Square Error criterion) indicates the recognized person.

4.3. Reduction of dimensionality using DCT, WHT and eigenfaces

The first experiment consisted of the evaluation of the identification rates as function of the vector dimension. Thus, 200 tests (40 people × 5 test images per person) are performed for each vector dimension (92 different vector dimensions for DCT and WHT, and 200 for the eigenfaces and KLT method) and the corresponding identification rates are obtained. The possible vector lengths for the rectangular mask are $(N')^2 = 1, 4, 9, 16$, etc. For the eigenfaces and KLT methods we have just retained a number of eigenfaces ranging from 1 to 200. Fig. 3 compares the achieved results. We have used the minimum value of the detection cost function (DCF) for comparison purposes. This parameter is defined as Martin et al. (1997):

$$\mathrm{DCF} = C_{\mathrm{miss}} \times P_{\mathrm{miss}} \times P_{\mathrm{true}} + C_{\mathrm{fa}} \times P_{\mathrm{fa}} \times P_{\mathrm{false}}$$

where C_{miss} is the cost of a miss (rejection), C_{fa} is the cost of a false alarm (acceptance), P_{true} is the a priori probability of the target, and $P_{\mathrm{false}} = 1 - P_{\mathrm{true}}$. We have used $C_{\mathrm{miss}} = C_{\mathrm{fa}} = 1$.

Taking into account that we are looking for a low-complexity face recognition system, the classifier consists of a nearest neighbor classifier using the mean square error (MSE) or the mean absolute difference (MAD) defined as

$$\mathrm{MSE}(\vec{x}, \vec{y}) = \sum_{i=1}^{(N')^2} (x_i - y_i)^2$$

$$\mathrm{MAD}(\vec{x}, \vec{y}) = \sum_{i=1}^{(N')^2} |x_i - y_i|$$

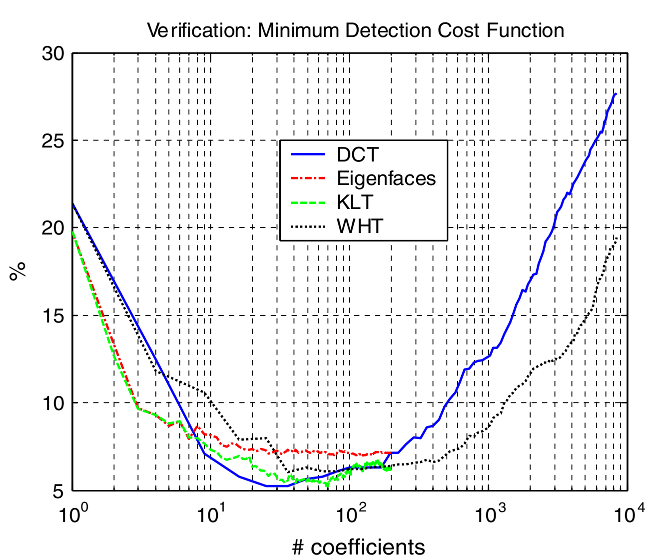


Fig. 3. Minimum DCF vs number of coefficients for several dimensionality reduction methods.

Table 3
Minimum detection cost function (DCF) for several transforms

Transform	# Coefficients	Min (DCF)
Eigenfaces	126	6.99%
KLT	68	5.24%
DCT	25	5.23%
WHT	36	6.05%

In our simulations better results are obtained using the MAD criterion. Thus, we have chosen the MAD criterion in our simulations.

Table 3 shows the optimal number of coefficients for each transform, and the associated DCF value.

These results can be improved (Faundez-Zanuy, 2004) offering several trials to each user. This is a data fusion scheme (Faundez-Zanuy, 2005) where a user is accepted if there is, at least, one face successfully verified in several snapshots. In our simulations, using this procedure and five photos per test, we have reduced the DCF more than two times.

It is important to observe that our new proposal, in addition to efficiency, presents another advantage over the classical eigenfaces method: the transformation is not data dependent, so it is not necessary to find any projection vector set. This same kind of solution is also preferred in image coding algorithms (JPEG, MPEG, etc.) that use DCT instead of KLT, because it is a fast transform that requires real operations and it is a near optimal substitute for the KL transform of highly correlated images, and has excellent energy compactation for images.

On the other hand, the KLT or its practical implementation (eigenfaces) implies that perhaps the set of projection vectors is too fitted to the training images that have been used to extract them, and can present poor generalization capabilities when representing the test images not used during training. For this reason, we can check in Table 3 that DCT and WHT require a smaller number of coefficients than KLT.

Table 4 summarizes the results using the FERET database, which consist of $N = 992$ (client) + 993 × 992

Table 4
Recognition rates using FERET database

Input signal	Transform	Identif. (%)	Min (DCF) (%)
R	DCT	73.08	5.60
	WHT	66.53	6.21
G	DCT	68.55	5.99
	WHT	63.51	6.63
B	DCT	65.63	6.09
	WHT	61.69	6.68
Y	DCT	69.46	5.76
	WHT	63.51	6.64
Score fusion: R + G + B	DCT	71.57	5.31
	WHT	68.55	6.05
Score fusion: 0.3R + 0.59G + 0.11B	DCT	70.97	5.45
	WHT	67.94	6.18

(impostors) = 986048. So, with 95% confidence, our experiments guarantee statistical significance in experiments with an empirical error rate $\widehat{P}$, down to 0.01%. We can observe the same conclusions as using the ORL database.

5. Conclusions

We have proposed a new approach to face recognition based on the Walsh–Hadamard transform, which can be easily implemented on a fixed point processor (Faundez-Zanuy et al., 2005). The experimental results reveal that it is competitive with the state-of-the-art statistical approaches to face recognition. Taking advantage of the minor differences of using different transforms (see Jain, 1989, p. 517), emphasis is focused on this items:

(a) We check that WHT performs reasonably good using small and large size databases (ORL, FERET).
(b) We check the differences on execution time, which justify the utility of WHT jointly with the possibility to implement it on a fixed point processor.

Acknowledgements

This work has been supported by FEDER and MEC, TIC-2003-08382-C05-02, TEC-2006-13141-C03-02.

References

Faundez-Zanuy, M., 2004. On the vulnerability of biometric security systems. IEEE Aerospace Electron. Systems Mag. 19 (6), 3–8.

Faundez-Zanuy, M., 2005. Data fusion in biometrics. IEEE Aerospace Electron. Systems Mag. 20 (1), 34–38.

Faundez-Zanuy, M., Espinosa, V., Ortega, J.A., 2005. A low-cost webcam and personal computer opens doors. IEEE Aerospace Electron. Systems Mag. 20 (11), 23–26.

Gibson, J.D. et al., 1998. Digital Compression for Multimedia, Principles and Standards. Morgan Kaufmann.

Gonzalez, R.C., Woods, R.E., 1993. Digital Image Processing. Addison Wesley.

Guyon, I., Makhoul, J., Schwartz, R., Vapnik, V., 1998. What size test set gives good error rate estimates?. IEEE Trans. Pattern Anal. Machine Intell. 20 (1) 52–64.

Jain, A.K., 1989. Fundamentals of Digital Image Processing. Prentice Hall.

Jain, A.K., Bolle, R., Pankanti, S., 1999. Biometrics. Personal Identification in Networked Society. Kluwer Academic Publishers.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., Przybocki, M., 1997. The DET curve in assessment of detection performance. In: European Speech Processing Conf. Eurospeech, vol. 4, pp. 1895–1898.

Samaria, F., Harter, A., 1994. Parameterization of a stochastic model for human face identification. 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, December.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cognitive Neurosci. 3 (1), 71–86, Massachusetts Institute of Technology.