

Wavelets and Face Recognition

Dao-Qing Dai and Hong Yan

*Sun Yat-Sen (Zhongshan) University and City University of Hong Kong
China*

1. Introduction

Face recognition has recently received significant attention (Zhao et al. 2003 and Jain et al. 2004). It plays an important role in many application areas, such as human-machine interaction, authentication and surveillance. However, the wide-range variations of human face, due to pose, illumination, and expression, result in a highly complex distribution and deteriorate the recognition performance. In addition, the problem of machine recognition of human faces continues to attract researchers from disciplines such as image processing, pattern recognition, neural networks, computer vision, computer graphics, and psychology. A general statement of the problem of machine recognition of faces can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces.

In identification problems, the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.

The solution to the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face regions, recognition or verification. Robust and reliable face representation is crucial for the effective performance of face recognition system and still a challenging problem.

Feature extraction is realized through some linear or nonlinear transform of the data with subsequent feature selection for reducing the dimensionality of facial image so that the extracted feature is as representative as possible.

Wavelets have been successfully used in image processing. Its ability to capture localized time-frequency information of image motivates its use for feature extraction. The decomposition of the data into different frequency ranges allows us to isolate the frequency components introduced by intrinsic deformations due to expression or extrinsic factors (like illumination) into certain subbands. Wavelet-based methods prune away these variable subbands, and focus on the subbands that contain the most relevant information to better represent the data.

In this paper we give an overview of wavelet, multiresolution representation and wavelet packet for their use in face recognition technology.

2. Introduction to wavelets

Wavelets are functions that satisfy certain mathematical requirements and are used in presenting data or other functions, similar to sines and cosines in the Fourier transform. However, it represents data at different scales or resolutions, which distinguishes it from the Fourier transform.

2.1 Continuous wavelet transform

Wavelets are formed by dilations and translations of a single function $\psi(t)$ called mother wavelet so that the dilated and translated family

$$\left\{ \psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \right\}_{(a,b) \in \mathbb{R} \setminus \{0\} \times \mathbb{R}}$$

is a basis of $L^2(\mathbb{R})$. The normalization ensures that $\|\psi_{a,b}(t)\|$ is independent of the scale parameter a and the position parameter b . The function ψ is assumed to satisfy some admissibility condition, for example,

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (1)$$

where $\Psi(\omega)$ is the Fourier transform of ψ . The admissibility condition (1) implies

$$\Psi(0) = \int \psi(t) dt = 0 \quad (2)$$

The property (2) motivates the name wavelet. The “diminutive” appellation comes from the fact that ψ can be well localized with arbitrary fine by appropriate scaling. For any $f(t) \in L^2(\mathbb{R})$, the continuous wavelet transformation (CWT) is defined as

$$CWTf(a,b) = \langle f, \psi_{a,b}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{a,b}(t)} dt$$

However, in signal processing, we often use discrete wavelet transform (DWT) to represent a signal $f(t)$ with translated version of a lowpass scaling function $\phi(t)$ and the dilated and translated versions of mother wavelet $\psi(t)$ (Daubechies, 1992).

$$f = \sum_k c_{j_0,k} \phi_{j_0,k} + \sum_{j^3} \sum_{j_0}^{\infty} \sum_k d_{j,k} \psi_{j,k}, \quad c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle, \quad d_{j,k} = \langle f, \psi_{j,k} \rangle,$$

where the functions $\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k)$ and $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$, $j, k \in \mathbb{Z}$, form an orthonormal basis of $L^2(\mathbb{R})$.

The partial sum of wavelet $\sum_{k=-\infty}^{+\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k}$ can be interpreted as the approximation of f at the resolution 2^j . The approximation of signals at various resolutions with orthogonal projections can be computed by multiresolution which is characterized by a particular discrete filter that governs the loss of information across resolutions. These discrete filters provide a simple procedure for decomposing and synthesizing wavelet coefficients at different resolutions (Mallat, 1999).

$$c_{j,k} = \sum_{\ell} \overline{h_{\ell-2k}} c_{j+1,\ell}, \quad d_{j,k} = \sum_{\ell} \overline{g_{\ell-2k}} c_{j+1,\ell}$$

$$c_{j+1,k} = \sum_{\ell} h_{k-2\ell} c_{j,\ell} + g_{k-2\ell} d_{j,\ell}$$

where $\{h_k\}, \{g_k\}$ are discrete filter sequences, they satisfy respectively

$$\phi(t) = \sum_k h_k \phi(2t - k), \quad \psi(t) = \sum_k g_k \phi(2t - k), \quad g_k = (-1)^k \overline{h_{1-k}}$$

The two-channel filter bank method parallelly filters a signal by the lowpass filters h and highpass filter g followed by subsampling. The filter h removes the high frequencies and retains the low frequency components, the filter g removes the low frequencies and produces high frequency components. Together, they decompose the signal into different frequency subbands, and downsampling is used to keep half of the output components of each filter. For the wavelet transform, only the lowpass filtered subband is further decomposed.

2.2 Two-dimensional wavelet transform

The two-dimensional wavelet can also be constructed from the tensor product of one-dimensional ϕ and ψ by setting:

$$\phi(x, y) = \phi(x)\phi(y), \quad \psi^H(x, y) = \phi(x)\psi(y),$$

$$\psi^V(x, y) = \psi(x)\phi(y), \quad \psi^D(x, y) = \psi(x)\psi(y)$$

where $\psi^H(x, y), \psi^V(x, y), \psi^D(x, y)$ are wavelet functions. Their dilated and translated family $\{\psi_{j,k_1,k_2}^\lambda(x, y) : j, k_1, k_2 \in \mathbb{Z}, \lambda = H, V, D\}$ and $\{\phi_{j,k_1,k_2}(x, y) : j, k_1, k_2 \in \mathbb{Z}\}$ forms an orthonormal basis of $L^2(\mathbb{R}^2)$. For every $f \in L^2(\mathbb{R}^2)$, it can be represented as

$$f = \sum_{k \in \mathbb{Z}^2} c_{j_0,k} \phi_{j_0,k} + \sum_{j^1, j_0, k \in \mathbb{Z}^2, \lambda = H, V, D} d_{j,k}^\lambda \psi_{j,k}^\lambda$$

$$c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle, \quad d_{j,k}^\lambda = \langle f, \psi_{j,k}^\lambda \rangle$$

Similar to one-dimensional wavelet transform of signal, in image processing, the approximation of images at various resolutions with orthogonal projections can also be computed by multiresolution which characterized by the two-channel filter bank that governs the loss of information across resolutions. The one-dimensional wavelet decomposition is first applied along the rows of the images, then their results are further decomposed along the columns. This results in four decomposed subimages L_1, H_1, V_1, D_1 . These subimages represent different frequency localizations of the original image which refer to Low-Low, Low-High, High-Low and High-High respectively. Their frequency components comprise the original frequency components but now in distinct ranges. In each iterative step, only the subimage L_1 is further decomposed. Figure 1 (Top) shows a two-dimensional example of facial image for wavelet decomposition with depth 2.

The wavelet transform can be interpreted as a multiscale differentiator or edge detector that represents the singularity of an image at multiple scales and three different orientations — horizontal, vertical, and diagonal (Choi & Baraniuk, 2003). Each image singularity is represented by a cascade of large wavelet coefficients across scale (Mallat, 1999). If the singularity is within the support of a wavelet basis function, then the corresponding wavelet

coefficient is large. Contrarily, the smooth image region is represented by a cascade of small wavelet coefficients across scale. Some researchers have studied several features of wavelet transform of natural images (Mallat, 1999) (Vetterli & Kovachov, 1995) (Choi & Baraniuk, 2003):

- Multiresolution: Wavelet transform analyzes the image at different scales or resolutions.
- Locality: Wavelet transform decomposes the image into subbands that are localized in both space and frequency domains.
- Sparsity: A wavelet coefficient is large only if the singularities are present in the support of a wavelet basis function. The magnitudes of coefficients tend to decay exponentially across scale. Most energy of images concentrate on these large coefficients.
- Decorrelation: Wavelet coefficients of images tend to be approximately decorrelated because of the orthonormal property of wavelet basis functions.

These properties make the wavelet domain of natural image more propitious to feature extraction for face recognition, compared with the direct spatial-domain.

2.3 Wavelet-packet

There are complex natural images with various types of spatial-frequency structures, which motivates the adaptive bases that are adaptable to the variations of spatial-frequency. Coifman and Meyer (Coifman & Meyer 1990) introduced an orthonormal multiresolution analysis which leads to a multitude of orthonormal wavelet-like bases known as wavelet packets. They are linear combinations of wavelet functions and represent a powerful generalization of standard orthonormal wavelet bases. Wavelet bases are one particular version of bases that represent piecewise smooth images effectively. Other bases are constructed to approximate various-type images of different spatial-frequency structures (Mallat, 1999).

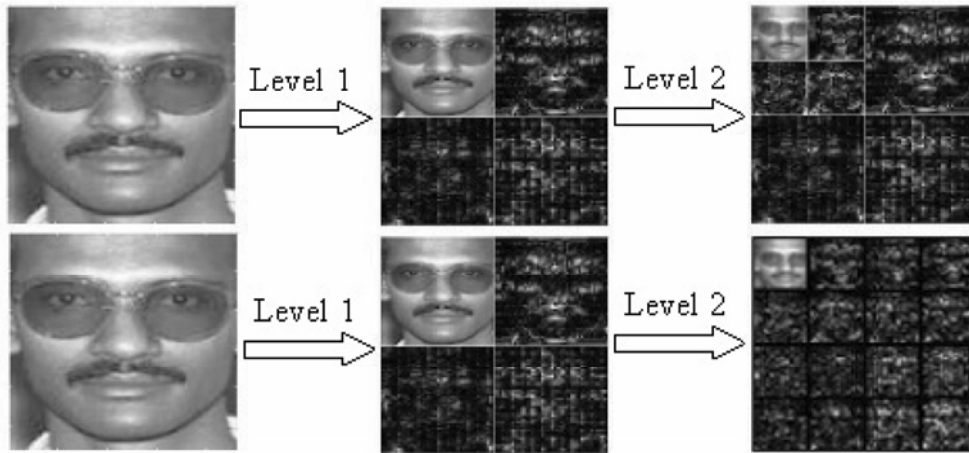


Figure 1. (Top) Two-dimensional wavelet decomposition of facial image with depth 2.
(Bottom) Two-dimensional wavelet packet decomposition of facial image with depth 2

As a generalization of the wavelet transform, the wavelet packet coefficients also can be computed with two-channel filter bank algorithm. The two-channel filter bank is iterated over both the lowpass and highpass branch in wavelet packet decomposition. Not only L_1 is further decomposed as in wavelet decomposition, but also H_1 , V_1 , D_1 are further decomposed. This provides a quad-tree structure corresponding to a library of wavelet packet basis and images are decomposed into both spatial and frequency subbands, as shown in Fig 1.

3. Preprocessing: Denoising

Denoising is an important step in the analysis of images (Donoho & Johnstone 1998, Starck et al. 2002). In signal denoising, a compromise has to be made between noise reduction and preserving significant signal details. Denoising with the wavelet transform has been proved to be effective, especially the nonlinear threshold-based denoising schemes. Wavelet Transform implements both low-pass and high-pass filters to the signal. The low-frequency parts reflect the signal information, and the high-frequency parts reflect the noise and the signal details. Thresholding to the decomposed high-frequency coefficients on each level can effectively denoise the signal.

Generally, denoising with wavelet consists of three steps:

- Wavelet Decomposition. Transform the noisy data into wavelet domain.
- Wavelet Thresholding. Apply soft or hard thresholding to the high-frequency coefficients, thereby suppress those coefficients smaller than certain amplitude.
- Reconstruction. Transform back into the original domain.

In the whole process, a suitable wavelet, an optimal decomposition level for the hierarchy and one appropriate thresholding function should be considered (Mallat 1999). But the choice of threshold is the most critical.

3.1 Wavelet Thresholding

Assuming the real signal $f[n]$ of size N is contaminated by the addition of a noise. This noise is modeled as the realization of a random process $W[n]$. The observed signal is

$$X[n] = f[n] + W[n], \quad n=0, \dots, N-1$$

The signal f is estimated by transforming the noisy data X with a *decision operator* Q . The resulting estimator is

$$\tilde{F} = QX$$

The goal is to minimize the error of the estimation, which is measured by a *loss function*. The square Euclidean norm is a familiar loss function. The risk of the estimator $\tilde{F}$ of f is the average loss:

$$r(Q, f) = E\{\|f - QX\|^2\}.$$

The noisy data

$$X = f + W \tag{3}$$

is decomposed in a wavelet basis $B = \{b_m\}_{0 \leq m < N}$. The inner product of (3) with b_m gives

$$X_B[m] = f_B[m] + W_B[m]$$

where $X_B[m] = \langle X, b_m \rangle$, $f_B[m] = \langle f, b_m \rangle$, $W_B[m] = \langle W, b_m \rangle$.

A diagonal estimator of f from (3) can be written

$$\tilde{F} = QX = \sum_{m=0}^{N-1} \rho_m(X_B[m])b_m,$$

where ρ_m are thresholding functions.

A wavelet thresholding is equivalent to estimating the signal by averaging it with a kernel that is locally adapted to the signal regularity. A filter bank of conjugate mirror filters decomposes a discrete signal in a discrete orthogonal wavelet basis. The discrete wavelets $\psi_{j,k}[n] = \psi(t)[n - N2^j k]$ are translated modulo modifications near the boundaries. The support of the signal is normalized to $[0, 1]$ and has N samples spaced by N^{-1} . The scale parameter 2^j thus varies from $2^L = N^{-1}$ up to $2^J < 1$:

$$B = [\{\psi_{j,k}[n]\}_{L < j \leq J, 0 \leq k < 2^{-j}}, \{\phi_{l,k}\}_{0 \leq k < 2^{-l}}].$$

A thresholding estimator in this wavelet basis can be written

$$\tilde{F} = \sum_{j=L+1}^J \sum_{k=0}^{2^{-j}} \rho_T(\langle X, \psi_{j,k} \rangle) \psi_{j,k} + \sum_{k=0}^{2^{-l}} \langle X, \phi_{l,k} \rangle \phi_{l,k},$$

where ρ_T is a hard thresholding or a soft thresholding.

A hard thresholding estimator is implemented with

$$\rho_T(x) = \begin{cases} x & \text{if } |x| > T \\ 0 & \text{if } |x| \leq T \end{cases}$$

A soft thresholding estimator is implemented with

$$\rho_T(x) = \begin{cases} x - T & \text{if } x \geq T \\ x + T & \text{if } x \leq -T \\ 0 & \text{if } |x| \leq T \end{cases}$$

The threshold T is generally chosen so that there is a high probability that it is just above the maximum level of the noise. When W_B is a vector of independent Gaussian random variables of variance σ^2 , the maximum amplitude of the noise has a very high probability of being just below $T = \sigma\sqrt{2\ln N}$. So we often choose the threshold $T = \sigma\sqrt{2\ln N}$. In this case, the soft thresholding guarantees with a high probability that $|\langle \tilde{F}, \psi_{j,k} \rangle| = |\rho_T(\langle X, \psi_{j,k} \rangle)| \leq |\langle f, \psi_{j,k} \rangle|$. The estimator $\tilde{F}$ is at least as regular as f because its wavelet coefficients have a smaller amplitude. This is not true for the hard thresholding estimator, which leaves unchanged the coefficients above T , and which can therefore be larger than those of f because of the additive noise component.

Face images with noise can be estimated by thresholding their wavelet coefficients. The image $f[n_1, n_2]$ contaminated by a white noise is decomposed in a separable two-dimensional wavelet basis. Figure 2 (a) is the original image, Figure 2 (b) is the noise image. Figure 2 (c, d) are obtained with a hard thresholding and a soft thresholding in a Symmlet 4 wavelet basis.

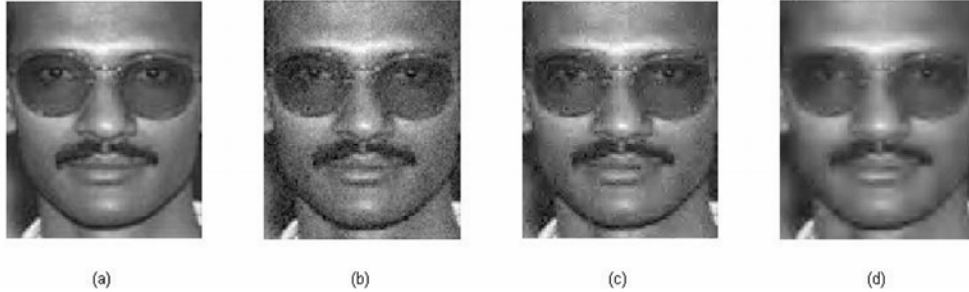


Figure 2. (a) Original image, (b) Noisy image (SNR = 19.95), (c) Estimation with a hard thresholding in a separable wavelet basis (Symmlet 4), (SNR = 22.03), (d) Soft thresholding (SNR = 19.96)

The threshold $T = \sigma\sqrt{2\ln N}$ is not optimal, especially, when the noise W is not white, the variance of the noise depends on each vector b_m of the basis. Thresholding estimators can be adapted.

3.2 Multiscale SURE Thresholds

Piecewise regular signals have a proportion of large coefficients $|\langle f, \psi_{j,k} \rangle|$ that increases when the scale 2^j increases. Indeed, a singularity creates the same number of large coefficients at each scale, whereas the total number of wavelet coefficients increase when the scale decreases. To use this prior information, one can adapt the threshold choice to the scale 2^j . At large scale 2^j , the threshold T_j should be smaller in order to avoid setting to zero too many large amplitude signal coefficients, which would increase the risk.

3.3 Translation Invariance

Thresholding noisy wavelet coefficients create small ripples near discontinuities. Indeed, setting to zero a coefficient $\langle f, \psi_{j,k} \rangle$ subtracts $\langle f, \psi_{j,k} \rangle \psi_{j,k}$ from f , which introduces oscillations whenever $\langle f, \psi_{j,k} \rangle$ is non-negligible. These oscillations are attenuated by a translation invariant estimation, consequently, can significantly improve the SNR. Thresholding wavelet coefficients of translated signals and translating back the reconstructed signals yields shifted oscillations created by shifted wavelets that are set to zero. The averaging partially cancels these oscillations, reducing their amplitude. Design of a translation invariant pattern recognition based on wavelets is still demanded.

4. Wavelets for feature extraction

Feature extraction in the sense of some linear or nonlinear transform of the data with subsequent feature selection is commonly used for reducing the dimensionality of facial image so that the extracted feature is as representative as possible. The images may be represented by their original spatial representation or by frequency domain coefficients. Features that are not obviously present in one domain may become obvious in the other domain. Unfortunately, Heisenberg uncertainty theorem implies that the information can not be compact in both spatial and frequency domain simultaneously. So, neither approach is ideally suited for all kinds of feature distribution. It motivates the use of the wavelet

transform which represents both the spatial and frequency domain simultaneously. Moreover, multiresolution analysis makes it more appropriate to represent and extract features across different scales.

The wavelet transform or the wavelet packet transform have been used for feature extraction in face recognition. These are used in three ways:

- Direct use of wavelet coefficients.
- From combination of wavelet coefficients.
- Searching the best feature in the wavelet packet library.

4.1 Direct use of wavelet coefficients

The simplest application of the wavelet transform for face recognition uses directly wavelet coefficients as features. The wavelet transform can locally detect the multiscale edges of facial images, the lineament edge information exists in the lowest spatial-frequency subband, while finer edge information presents in the higher spatial-frequency subband.

The waveletface (Chien & Wu, 2002) is a wavelet based approach. It uses the wavelet transform to decompose the image data into four subimages via the low-pass and high-pass filters with respect to the column vectors and the row vectors of array pixels. Then the low spatial-frequency subimage is selected for further decomposition. The three-level lowest spatial-frequency subimage with a matrix of $(n_{row}/8) \times (n_{col}/8)$ is extracted as the feature vector, referred to as waveletface, where $n_{row} \times n_{col}$ is the resolution of facial image. Generally, low frequency components represent the basic figure of an image, which is less sensitive to image variations. These components form the most informative subimage gearing with the highest discriminating power. The waveletface can be expressed by a form of linear transformation: $y = W^T_{wavelet} x$, where $W^T_{wavelet} x$ is composed of impulse responses of the low pass filter h . Different from some statistics based methods, such as eigenface and fisherface, see (Zhao et al 2003), the waveletface can be independently extracted without the effect of new enrolled users. Waveletface is an efficient method because no extra computation is needed.

4.2 From combinations of wavelet coefficients

The direct use of wavelet coefficients may not extract the most discriminative features for two reasons:

- There is much redundant or irrelevant information contained in wavelet coefficients.
- Can not recover new meaning underlying features which has more discriminative power.

In order to overcome the deficiency of direct use of wavelet coefficients, it is possible to construct features from the combinations of wavelet coefficients to produce a low-dimensional manifold with minimum loss of information so that the relationships and structure in the data can be identified. These can be done in two ways:

- Use the statistical quantum of wavelet coefficients in each spatial-frequency subband as discriminative features.
- Employ traditional transforms (e.g., PCA, LDA, ICA, AM, Neural Networks) to enhance and extract discriminative features in one or several special spatial-frequency subbands.

4.2.1 Use the statistical measures as discriminative features

The statistical measures, e. g., mean, variance, are usually helpful to represent features or characteristics of data, it is simple and requires less computation load.

Garcia et al. (Garcia et al., 1998) present a wavelet-based framework for face recognition. Each face is described by a subset of subband filtered images containing wavelet coefficients after two-level wavelet packet transform. These coefficients characterize the face texture and a set of simple statistical measures are used to reduce dimensionality and characterize textural information, which forms compact and meaningful feature vectors $\nu = \bigcup_{i=0}^{16} \{\mu_i, \sigma_i^2\}$. After the extraction of all the vectors of the training set, only the components with a mean value above a predefined threshold are considered for feature vector formation. It is supposed that each component pair is independent from the other component pairs of the feature vector. Then, the Bhattacharyya distance between two feature vectors ν_k and ν_l is computed on a component-pair basis

$$D(\nu_k, \nu_l) = \sum_{i=0}^n D_i(\nu_k, \nu_l), \quad D(\nu_k, \nu_l) = \frac{1}{4} \frac{(\mu_{ik} - \mu_{il})^2}{(\sigma_{ik}^2 + \sigma_{il}^2)} + \frac{1}{2} \ln \left[\frac{(\sigma_{ik}^2 + \sigma_{il}^2)}{2\sqrt{\sigma_{ik}^2 \sigma_{il}^2}} \right]$$

in order to classify the face feature vectors into person classes.

In fact, other statistical measures, e. g., other kinds of moments can be used in the above wavelet-based framework for face recognition. Moreover, the discrete density function of whole wavelet coefficients in each subband can be evaluated. The similarity measure of density function can be computed by some relative entropy, such as Kullback-Leibler divergence or J-divergence.

4.2.2 Employ traditional transform in special subbands

Generally, the wavelet coefficients are deficient to be good discriminative features, a further discriminant analysis is adopted to recover new meaningful underlying features which has more discriminative power. The traditional transforms (e.g., PCA, LDA, ICA, AM, Neural Networks) are very popular for their simplicity and practicality. They can be performed on one or several special spatial-frequency subbands which may be chosen by certain criterion. We (Feng et al. 2000) proposed a wavelet subband approach in using PCA for human face recognition. Three-level wavelet transform is adopted to decompose an image into different subbands with different frequency components. A midrange frequency subband is selected for PCA representation. The experiments show that it has low computation and higher accuracy, comparing with using original PCA directly in spatial domain.

In (Dai & Yuen, 2006) we used a wavelet enhanced regularized discriminant analysis to solve the small sample size problem and applied it to human face recognition. We analyzed the role of the wavelet transform, low-pass filtering will reduce the dimension of input data but meanwhile increases the magnitude of the within-class covariance matrix so that the variation information plays too strong a role and the performance of the system will become poorer. It also overcomes the difficulty in solving a singular eigenvalue problem in traditional LDA. Moreover, a wavelet enhanced regularization LDA system for human face recognition is proposed to adequately utilize the information in the null space of withinclass scatter matrix (Dai & Yuen, 2003).

Ekenel et al. (Ekenel & Sankur, 2005) introduced a ternary-architecture multiresolution face recognition system. They used the 2D discrete wavelet transform to extract multiple

subband face images. These subband images contain coarse approximations of the face as well as horizontal, vertical and diagonal details of faces at various scales. Subsequently, The PCA or ICA features are extracted from these subbands. They exploit these multiple channels by fusing their information for improved recognition. Their experiments show that it has good performance, especially against illumination perturbations.

In (Zhang et al., 2004), they proposed a modular face recognition scheme by combining the techniques of wavelet subband representations and kernel associative memories. By the wavelet transform, face images are decomposed and the computational complexity is substantially reduced by choosing a lower spatial-frequency subband image. Then an kernel associative memory (KAM) model are built up for each subject, with the corresponding prototypical images without any counter examples involved. Multiclass face recognition is thus obtained by simply holding these associative memories. When a probe face is presented, the KAM model gives the likelihood that the probe is from the corresponding class by calculating the reconstruction errors or matching scores.

Illumination compensation is always a problem important but difficult to solve in face recognition. The wavelet transform decomposes the data into different frequency ranges which allows us to isolate the frequency components introduced by illumination effects into certain subspaces. We can use the subspaces that do not contain these illumination-based frequency components to better represent our data, so as to eliminate the influence of the illumination changes, before a face image is recognized. In (Zhang et al., 2005), a face compensation approach based on wavelet and neural network is proposed. A rough linear illumination compensation was first performed for the given face image, which can only compensate the lower frequency features and the effect is limited. The higher frequency features are not be compensated. But it can reduce the learning pressure of the neural network, accelerate the convergent rate and improve the learning accuracy as well as the extensibility of the network. The method can compensates the different scale features of the face image by using the multi-resolution characteristic of the wavelet and the self-adaptation learning and good spread ability of BP neural network. Their experiments show that it can solve the problem of illumination compensation in the face recognition process.

4.3 Search local discriminant basis/coordinates in wavelet packet library

As a generalization of the wavelet transform, the wavelet packet not only offers us an attractive tool for reducing the dimensionality by feature extraction, but also allows us to create localized subbands of the data in both space and frequency domains. A wavelet packet dictionary provides an over-complete set of spatial-frequency localized basis functions onto which the facial images can be projected in a series of subbands. The main design problem for a wavelet packet feature extractor is to choose which subset of basis functions from the dictionary should be used. Most of the wavelet packet dictionary methods that have been proposed in the literature are based on algorithms which were originally designed for signal compression such as the best basis algorithm (Coifman & Wickerhauser, 1992), or the matching pursuit algorithm (Mallat & Zhang, 1993).

Saito and Coifman introduced the local discriminant basis (LDB) algorithm based on a best basis paradigm, searching for the most discriminant subbands (basis) that illuminates the dissimilarities among classes from the wavelet packet dictionary (Coifman & Satio, 1994) (Satio & Coifman, 1995). It first decomposes the facial images in the wavelet packet dictionary, then facial image energies at all coordinates in each subband are accumulated for

each class separately to form a spatial-frequency energy distribution per class on the subband. Then the difference among these energy distributions of each subband is measured by a certain “distance” function (e.g., Kullback-Leibler divergence), a complete local discriminant basis (LDB) is selected by the difference-measure function using the best basis algorithm (Coifman & Wicherhauser, 1992), which can represent the distinguishing facial features among different classes. After the basis is selected, the loadings of their coordinates are fed into a traditional classifier such as linear discriminant analysis (LDA) or classification tree (CT). Finally, the corresponding coefficients of probes are computed and fed to the classifier to predict their classes.

Unfortunately, the energies may not be so indicative for discrimination sometimes, because not all coordinates in the LDB are powerful to distinguish different subjects. Many less discriminant coordinates may add up to a large discriminability for the LDB. An example of artificial problem was used to validate that it may be fail to select the right basis function as a discriminator (Saito & Coifman, 2002). So Saito and Coifman suggested a modified version of the LDB (MLDB) algorithm which uses the empirical probability distributions instead of the space-scale energy as their selection strategy to eliminate some less discriminant coordinates in each subband locally (Saito & Coifman, 2002). It estimates the probability density of each class in each coordinate in all subbands. Then the discriminative power of each subband is represented by the first N_0 most discriminant coordinates in terms of the “distance” among the corresponding densities (e.g., Kullback-Leibler divergence among the densities). This information is then used for selecting a basis for classification as in original LDB algorithm. Although the MLDB algorithm may overcome some shortage of LDB, the selection of coordinates is only limited to each subband so that the coordinates in different subbands are still incomparable. Therefore, the MLDB algorithm gives an alternative to the original LDB.

This LDB concept has become increasingly popular and has been applied to a variety of classification problems. Based on LDB idea, Kouzani et al. proposed a human face representation and recognition system based on the wavelet packet method and the best basis selection algorithm (Kouzani et al. 1997). An optimal transform basis, called the face basis, is identified for a database of the known face images. Then it is used to compress all known faces within the database in a single pass. For face recognition, the probe face image is transformed, and the compressed face is then compared against the database. The best filter and best wavelet packet decomposition level are also discussed there.

Since features with good discriminant property may locate in different subbands, it is important to find them among all subbands instead of certain specific subbands. We proposed a novel local discriminant coordinates (LDC) method based on wavelet packet for face recognition to compensate for illumination, pose and expression variations (Liu et al. 2007). The method searches for the most discriminant coordinates from the wavelet packet dictionary, instead of the most discriminant basis as in the LDB algorithm. The LDC idea makes use of the scattered characteristic of best discriminant features. In the LDC method, the feature selection procedure is independent of subbands, and only depends on the discriminability of all coordinates, because any two coordinates in the wavelet packet dictionary are comparable for their discriminability which is computed by a *maximum a posterior* logistic model based on a dilation invariant entropy. LDC based feature extraction not only selects low frequency components, but also middle frequency components whose judicious combination with low frequency components can improve the performance of face recognition greatly.

4.4 Robust issue

It is known that a good feature extractor of face recognition system is claimed to select as more as possible the best discriminant features which are not sensitive to arbitrary environmental variations. Nastar et al. (Nastar & Ayach, 1996) investigated the relationship between variations in facial appearance and their deformation spectrum. They found that facial expressions and small occlusions affect the intensity manifold locally. Under frequency-based representation, only high-frequency spectrum is affected. Moreover, changes in pose or scale of a face and most illumination variations affect the intensity manifold globally, in which only their low-frequency spectrum is affected. Only a change in face will affect all frequency components. So there are no special subbands whose all coordinates are not sensitive to these variations.

In each subband, there may be only segmental coordinates have enough discriminant power to distinguish different person, the remainder may be sensitive to environmental changes, So some methods based on the whole subband may also use these sensitive features which maybe affect their performance for face recognition.

Moreover, there may be no special subbands containing all the best discriminant features, because the features not sensitive to environmental variations are always distributed in different coordinates of different subbands locally. So methods based on the segmental subbands may lose some good discriminant features.

Furthermore, in different subbands, the amount and distribution of best discriminant coordinates are always different. Many less discriminant coordinates in one subband may add up to a larger discriminability than another subband whose discriminability is added up with few best discriminant coordinates and residual small discriminant coordinates. So the few best discriminant coordinates may be discarded by some methods which search for the best discriminate subbands, but usually only the few best discriminant coordinates are needed.

So the best discriminant information selection should be independent of their seated subbands, and only depends on their discriminability for face recognition. In addition, there may be some redundancy or collinearity in features which will affect the performance for face recognition. However, another limitation of using wavelet for face recognition is that the wavelet transform has no property of translation invariance. Mallat (Mallat, 1996) discussed that the wavelet representation not only contains spatial and frequency information, but also phase information. When the phase information varies with small translations, it will cause difficulties with matched filtering applications. For achieving translation invariance, it should contain some redundant information in the representing features.

The wide-range variations of human face, due to pose, illumination, and expression, require the wavelet transform to extract features that are translation invariant and to a certain extent scale invariant. This constitutes a trade-off between the amount of possible invariance and the sparseness of the wavelet representation. So a robust wavelet feature extractor should select a best discriminant features group with appropriate redundancy or co-linearity. However, searching such a wavelet feature extractor is a difficult task and needs further research.

5. Conclusion

Wavelets have been successfully used in image processing. Their ability to capture localized spatial-frequency information of image motivates their use for feature extraction. We give an overview of using wavelets in the face recognition technology. Due to limit of space the use of Gabor wavelets is not covered in this survey. Interested readers are referred to section 8.3 for references.

6. Acknowledgements

This project is supported in part by NSF of China (60175031, 10231040, 60575004), the Ministry of Education of China (NCET-04-0791), NSF of Guangdong (05101817) and the Hong Kong Research Grant Council(project CityU 122506).

7. References

- Chien, J. T. & Wu, C. C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 12, (Dec. 2002) pp. 1644–1649.
- Choi, H. & Baraniuk, R. G. (2003). Wavelet Statistical Models and Besov Spaces, In: *Nonlinear Estimation and Classification*, Denison, D. D.; Hansen, M. H., et al. (Eds.), pp. 9-29, Springer-Verlag, NewYork.
- Coifman, R. R. & Meyer, Y. (1990). Orthonormal wavelet packet bases, Preprint.
- Coifman, R. R. & Wickerhauser, M. V. (1992). Entropy-based algorithm for best basis selection, *IEEE Trans. Infor. Theory*, Vol. 38, No. 2, (March, 1992) pp. 713–718.
- Dai, D. Q. & Yuen, P. C. (2003). Wavelet-based 2-parameter regularized discriminant analysis for face recognition, In: AVBPA 2003, LNCS 2688, J. Kittler and M.S. Nixon (Eds.), pp. 137–144, Springer-Verlag, Berlin Heidelberg.
- Dai, D. Q. & Yuen, P. C. (2006). Wavelet based discriminant analysis for face recognition, *Applied Mathematics and Computation*, 175(April 2006), 307-318
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, New York, 1992.
- Donoho D. L. & Johnstone I. M. (1998). Minimax estimation via wavelet shrinkage, *Annals of Statistics*, 26 (3) (JUN 1998) pp. 879-921.
- Ekenel, H. K. & Sanker, B. (2005). Multiresolution face recognition, *Image and Vision Computing*, Vol. 23, (May 2005) pp. 469–477.
- Feng, G. C., Yuen, P. C. & Dai, D. Q. (2000). Human face recognition using PCA on wavelet subband, *Journal of Electronic Imaging*, Vol. 9, No. 2, (April 2000) pp. 226–233.
- Garcia, C., Zikos, G. & Tziritas, G. (1998). A wavelet-based framework for face recognition, Proc of the Workshop on Advances in Facial Image Analysis and Recognition Technology, *5th European Conference on Computer Vision (ECCV'98)* , pp. 84-92, Freiburg Allemagne.
- Jain A. K., Ross R. & Prabhakar S. (2004) , An introduction to biometric recognition, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, no. 1 (Jan. 2004), pp. 4-20.
- Kouzani, A. Z., He, F. & Sammut, K. (1997). Wavelet packet face representation and recognition, *IEEE Int Conf. Systems, Man, and Cybernetics*, Vol. 2, (Oct. 1997) pp. 1614–1619.

- Liu C. C., Dai D. Q. & Yan H. (2007). Local discriminant wavelet packet coordinates for face recognition, *Journal of Machine Learning Research*, Vol. 8 (May 2007) 1165-1195.
- Mallat, S. & Zhang, Z. (1993). Matching pursuit in a time-frequency dictionary, *IEEE Transactions on Signal Processing*, Vol. 41, pp. 3397-3415.
- Mallat, S. (1996). Wavelets for a vision, *Proc. IEEE*, Vol. 84, No. 4, pp. 604-614.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing*, Academic Press (2nd Edition), ISBN : 0-12-466606-X, San Diego.
- Nastar, C. & Ayach, N. (1996). Frequency-based nonrigid motion analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, (Nov. 1996) pp. 1067-1079.
- Saito, N. & Coifman, R. R. (1994). Local discriminant bases, *Proc. SPIE 2303*, pp 2-14.
- Saito, N. & Coifman, R. R. (1995). Local discriminant bases and their applications, *J. Math. Imaging Vision*, Vol. 5, No. 4, pp. 337-358.
- Saito, N., Coifman, R. R., Geshwind, F. B. & Warner, F. (2002). Discriminant feature extraction using empirical probability density estimation and a local basis library, *Pattern Recognition*, Vol. 35, (Dec. 2002) pp. 2841-2852.
- Starck J. L., Candes E. J., Donoho D. L. (2002). The curvelet transform for image denoising, *IEEE Transactions on Image Processing* 11 (6)(JUN 2002) pp. 670-684
- Vetterli, M. & Kovalevich, J. (1995). *Wavelets and Subband coding*, Prentice Hall, Englewood Cliffs, NJ.
- Zhang, B. L., Zhang, H. H. & Ge, S. S. (2004). Face recognition by applying wavelet subband representation and kernel associative memory, *IEEE Trans. Neural Networks*, Vol. 15, No. 1, (Jan 2004) pp. 166-177.
- Zhang, Z. B., Ma, S. L. & Wu, D. Y. (2005). The application of neural network and wavelet in human face illumination compensation, *Proc. Advances in Neural Networks*, pp. 828-835.
- Zhao W., Chellappa R. Phillips P. J. & Rosenfeld A. (2003). Face recognition: A literature survey, *ACM Comput. Surv.* Vol. 35, no. 4, (Dec. 2003) pp. 399-459.

8. Further readings

8.1 Face detection

- Huang, L. L. & Shimizu, A. (2006). A multi-expert approach for robust face detection, *Pattern Recognition*, Vol. 39, No. 9, (SEP 2006) pp. 1695-1703.
- Liu, C. J. (2003). A Bayesian discriminating features method for face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 6, (JUN 2003) pp. 725-740.
- Le, D. D. & Satoh, S. (2006). A multi-stage approach to fast face detection, *IEICE Transactions on Information and Systems*, E89D No. 7, (JUL 2006) pp. 2275-2285.
- Shih, P. C. & Liu, C. J. (2006). Face detection using discriminating feature analysis and Support Vector Machine, *Pattern Recognition*, Vol. 39, No. 2, (FEB 2006) pp. 260-276.
- Wang, J. W. & Chen, W. Y. (2006). Eye detection based on head contour geometry and wavelet subband projection, *Optical Engineering*, Vol. 45, No. 5, (MAY 2006).

8.2 Face recognition

- Bebis, G.; Gyaourova, A.; Singh, S. & Pavlidis, I. (2006). Face recognition by fusing thermal infrared and visible imagery, *Image and Vision Computing*, Vol. 24, No. 7, (JUL 2006) pp. 727-742.
- Kwak, K. C. & Pedrycz, W. (2004). Face recognition using fuzzy integral and wavelet decomposition method, *IEEE Transactions On Systems Man And Cybernetics Part B Cybernetics*, Vol. 34, No. 4, (AUG 2004) pp. 1666-1675.
- Kruger, V. & Sommer, G. (2002). Wavelet networks for face processing, *Journal of the Optical Society of America A-Optics Image Science and Vision*, Vol. 19, No. 6, (JUN 2002) pp. 1112-1119.
- Li, B. & Liu, Y. H. (2002). When eigenfaces are combined with wavelets, *Knowledge-Based Systems*, Vol. 15, No. 5-6, JUL 2002 pp. 343-347.
- Ngo, D. C. L.; Teoh, A. B. J. & Goh, A. (2006). Biometric hash: High-confidence face recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, No. 6 (JUN 2006) pp. 771-775.
- Pal, H. S.; Ganotra, D. & Neifeld, M. A. (2005). Face recognition by using feature-specific imaging, *Applied Optics*, Vol. 44, No. 18, (JUN 2005) pp. 3784-3794.
- Perlibakas, V. (2004). Face recognition using principal component analysis and wavelet packet decomposition, *Informatica*, Vol. 15, No. 2, pp. 243-250.
- Rajwade, A. & Levine, M. D. (2006). Facial pose from 3D data, *Image and Vision Computing*, Vol. 24, No. 8, (AUG 2006) pp. 849-856.
- Shih, F. Y. & Cheng, S. X. (2005). Improved feature reduction in input and feature spaces, *Pattern Recognition*, Vol. 38, No. 5, (MAY 2005) pp. 651-659. Wavelets and face recognition 17
- Tay, D. B. H. (2002). Parametric Bernstein polynomial for least squares design for all other of 3-D wavelet filter banks, *IEEE Transactions on Circuits and Systems I-Fundamental Theory and Applications*, Vol. 49, No. 6, (JUN 2002) pp. 887-891.
- Wijaya, S. L.; Savvides, M. & Kumar, B. V. K. V. (2005). Illumination-tolerant face verification of low-bit-rate JPEG2000 wavelet images with advanced correlation filters for handheld devices, *Applied Optics*, Vol. 44, No. 5, (FEB 2005) pp. 655-665.

8.3 Using Gabor wavelets

- Arca, S.; Campadelli, P. & Lanzarotti, R. (2006). A face recognition system based on automatically determined facial fiducial points, *Pattern Recognition*, Vol. 39, No. 3, (MAR 2006) pp. 432-443.
- Gokberk, B.; Irfanoglu, M. O.; Akarun, L. & Alpaydin, E. (2005). Selection of location, frequency, and orientation parameters of 2D Gabor wavelets for face recognition, *Advanced Studies in Biometrics*, Vol. 3161, pp. 138-146.
- Jeon, I. J.; Nam, M. Y. & Rhee, P. K. (2005). Adaptive gabor wavelet for efficient object recognition, *Knowledge-Based Intelligent Information and Engineering Systems, Part 2*, Vol. 3682, pp. 308-318.
- Kamarainen, J. K.; Kyrki, V. & Kalviainen, H. (2006). Invariance properties of Gabor filterbased features - Overview and applications, *IEEE Transactions on Image Processing*, Vol. 15, No. 5, (MAY 2006) pp. 1088-1099.

- Kim, D. S.; Jeon, I.; Lee, S. Y.; Rhee, P. K. & Chung, D. J. (2006). Embedded face recognition based on fast genetic algorithm for intelligent digital photography, *IEEE Transactions on Consumer Electronics*, Vol. 52, No. 3, (AUG 2006) pp. 726-734.
- Liu, C. J. & Wechsler, H. (2002). Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition, *IEEE Transactions on Image Processing*, Vol. 11, No. 4, (APR 2002) pp. 467-476.
- Liu, C. J. & Wechsler, H. (2003). Independent component analysis of Gabor feature's for face recognition, *IEEE Transactions on Neural Networks*, Vol. 14, No. 4, (JUL 2003) pp. 919-928.
- Liu, C. J. (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, (MAY 2004) pp. 572-581.
- Shen, L. L. & Bai, L. (2006). A review on Gabor wavelets for face recognition, *Pattern Analysis and Applications*, Vol. 9(2-3)(Oct. 2006), pp. 273-292.
- Shin, H. C.; Choi, H. C. & Kim, S. D. (2006). Directionally classified eigenblocks for localized feature analysis in face recognition, *Optical Engineering*, Vol. 45, No. 7, (JUL 2006).
- Singh, R.; Vatsa, M. & Noore, A. (2005). Textural feature based face recognition for single training images, *Electronics Letters*, Vol. 41, No. 11, (MAY 2005) pp. 640-641.
- Yan, S. C.; He, X. F.; Hu, Y. X.; Zhang, H. J.; Li, M. J. & Cheng, Q. S. (2004). Bayesian shape localization for face recognition using global and local textures, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1, (JAN 2004) pp. 102-113.
- Yu, J. G. & Bhanu, B. (2006). Evolutionary feature synthesis for facial expression recognition, *Pattern Recognition Letters*, Vol. 27, No. 11, (AUG 2006) pp. 1289-1298.
- Yu, W. W.; Teng, X. L. & Liu, C. Q. (2006). Face recognition fusing global and local features, *Journal of Electronic Imaging*, Vol. 15, No. 1, (JAN-MAR 2006).
- Zhang, H. H.; Zhang, B. L.; Huang, W. M. & Tian, Q. (2005). Gabor wavelet associative memory for face recognition, *IEEE Transactions on Neural Networks*, Vol. 16, No. 1, (JAN 2005) pp. 275-278.
- Zheng, W. M.; Zhou, X. Y.; Zou, C. R. & Zhao, L. (2006). Facial expression recognition using kernel canonical correlation analysis (KCCA), *IEEE Transactions on Neural Networks*, Vol. 17, No. 1, (JAN 2006) pp. 233-238.