The following manuscript was published in

Extraction and estimation of facial expression by means of wavelet decomposition and active shape template

Yung-Sheng Chen* and Hang-Ping Kuo
Department of Electrical Engineering, Yuan Ze University

ABSTRACT
This paper presents an automatic approach for extracting a face and estimating the facial expression from a gray image. To improve the execution time and the tolerance degree, the technique of wavelet decomposition is used in our approach for feature extraction. The main steps of the proposed approach are as follows. First, each input image is preprocessed with normalization, which includes translation, rotation, scaling, and light source adjustment. Secondly, the wavelet decomposition transform is applied to making image into different levels, and the information extracted from these levels is used to estimate the face region with a template-matching method. Finally, the facial expression is estimated from the found face region by using the technique of active shape template. Experimental results have confirmed the feasibility of the proposed approach.

Keywords: Face detection, Wavelet decomposition, Active shape template.

1. INTRODUCTION
Face recognition has revived from 1993, and prevails since 1997. We can find that pattern recognition journals and conference papers often issue the related topics. It represents that face recognition is an important application in pattern recognition field. The goal of this paper is to present an automatic approach for extracting a face and estimating the facial expression from a gray image. We follow the method of Cootes et al.’s work [1-3] to locate face feature position under different variance. That the spline actively searches feature until energy is the smallest, where the energy concept is based on Kass et al.’s work [4]. Next, we involve wavelet decomposition [5] into Cootes’s method in order to improve the searching speed and accuracy. At the wavelet lowest level, we can extract face-like objects speedy from the image, and the other levels are used to estimate the face scale, rotation, and each feature position. After locating each feature position, we can find the similar expression coefficients from each person in database.

In our approach, at training stage and before applying image for processing, each image is normalized by the following steps: translation, rotation, scaling, and light source adjustment. Since these preprocessed steps are known in an image processing handbook, they will not be presented in this paper. The other parts of this paper are organized as follows. Section 2 reviews simply the adopted fast wavelet transform theorem. Section 3 presents the strategy of our database training. Section 4 presents how to apply principal components analysis to our shape templates. Section 5 details our approach that uses active shape template to estimate the facial expression. Section 6 shows the experimental results. Section 7 finally gives the conclusion of our work.

2. WAVELET DECOMPOSITION
Because wavelet samples the signal by power of two, input signal length must be \( 2^n \). Thus the image space should be sampled to \( 2^n \times 2^n \) for a 2D image. In our system, the input image size is 512×512 pixels, then \( L_1 = L_2 = 9 \) and we can decompose the image into 10 levels. Our goal is to make the image decompose frequency into different levels under translation condition by wavelet transform. Different wavelet depends on different basis function. We choose coefficients 2 wavelet basis functions in order that the sampling interval consists of DC value which we calculate only one time at this interval. These theorems can be found in reference [5]. We can apply one-dimensional discrete wavelet transform twice (transform first by rows, and then transform by columns) to obtain the two-dimensional wavelet space, the lowest wavelet coefficient places at the top left square. Consider the \( p \)th level frequency component, if we keep the

* eyschen@ee.yzu.edu.tw; phone 886 3 4638800 ext 409; fax 886 3 4639355; http://www-zion.eed.yzu.edu.tw/; Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 320, Taiwan.
area information of 
\( \left( \{2^{p-1} \leq x < 2^p \} \cap \{2^{p-1} \leq y < 2^p \} \right) \cup \left( \{0 \leq x < 2^p \} \cap \{2^{p-1} \leq y < 2^p \} \right) \cup \left( \{2^{p-1} \leq x < 2^p \} \cap \{0 \leq y < 2^p \} \right) \) and the remainder are set to zero, then we can use previous setting wavelet plane pass through inverse wavelet transform. Where we define \( w^{(p)}(x, y) \) to be the value at \((x, y)\) on the reconstructed plane \( RW^{(p)} \) with level \( p \). If we keep the area information of \( \left( \{2^0 \leq x < 2^p \} \cap \{2^0 \leq y < 2^p \} \right) \) and the remainder are set to zero, then we can use previous setting wavelet plane pass through inverse wavelet transform. Where we define \( c^{(p)}(x, y) \) to be the value at \((x, y)\) on the reconstructed plane \( RC^{(p)} \) with level \( p \). Note here that the defined \( w^{(p)}(x, y) \) and \( c^{(p)}(x, y) \) has the following relationship.

\[
c^{(p)}(x, y) = w^{(p)}(x, y) + w^{(p-1)}(x, y) + \cdots + w^{(0)}(x, y)
\]  

(1)

3. DATABASE TRAINING

A model built from face features has the strongest edges. In our face model, we take off two pupils, two ears, and face upper outline, as shown in Fig. 1(a). Since there are no tools could make this model, we have to program an algorithm to build these models. Our training method is to make each image to be very close to the mean shape with upon 3D pose and different expression variance in the database. Thus we need to adjust scale, rotation, \( x \)-translation and \( y \)-translation for each shape. This algorithm mainly includes two stages, that is, training stage and extracting information stage.

3.1 Training stage

A shape data can be expressed by 
\[
X_i = (x_{i\alpha}, y_{i\alpha}, x_{i\beta}, y_{i\beta} ..., x_{in}, y_{in})' = \begin{bmatrix} x_{i\alpha} \\ y_{i\alpha} \end{bmatrix}, \quad k = 0, 1, ..., n - 1 \]. Where \( x_{in} \) means the \( i \)th shape with a landmark point \( n \) at \( x \) coordinate, and \( y_{in} \) means the \( i \)th shape with a landmark point \( n \) at \( y \) coordinate, respectively. Before training, we choose first shape as the mean shape (the first shape must be front-viewed and nonexpression), and define nose tip to be the shape center denoted by \((x_m, y_m)\). Whenever we adjust the shape by translation, scaling, and rotation operations, the nose tip plays an origin-point of axes. We adjust rotation of each shape first that makes angle \( \theta = 0 \) (\( \theta \) is an included angle between the left eye’s right-corner to the right eye’s left-corner line and horizontal line).

In the first training iteration, we use the first shape as the mean shape and compare it with all shapes from the database. Once all shapes have adjusted completely, we can calculate a new mean shape and compare it with all shapes again. If the subtraction between the previous mean shape from the last mean shape is smaller than 1, then we stop the training stage and go to next stage.

Fig. 1. (a) Our face model. (b) The effect of four 3D-pose and expression parameters.
3.2 Extracting information stage
First, we normalize each gray level image from its trained shape parameters ($x$-translation, $y$-translation, scale, and rotation). Next, we apply wavelet transform to decompose these normalized images. After that, the profile value on the object boundary can be extracted from the wavelet different decomposition plane and denoted by $w_{ijkw}$. The normalizing form can be written as $w_{ijkw} = w_{ijkw}/\sqrt{\sum_{k=1}^{n_p} w_{ijkw}^2}$. Where $w$ is the wavelet coefficient at the profile, $i \in [1,n]$ is the number of landmark points, $j \in [1,N]$ is the number of images, and $k \in [1,n_p]$ with $n_p$ is the profile length. For the covariance matrix, we follow [4] and use a simpler method to set the nondiagonal elements to zeros. The covariance matrix may be expressed as

$$ W = \begin{bmatrix} (w_{ik})^2 & 0 & \ldots & 0 \\ 0 & (w_{jk})^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & (w_{nk})^2 \end{bmatrix}, \text{ where } (w_{ik})^2 = \frac{1}{N} \sum_{i=1}^{N} (w_{ik} - \bar{w})^2. \quad (2) $$

Where $w_{ik}$ wavelet coefficient represents the frequency information, and its vector form can be written as $w^{(p)}$ with length $n$ (profile length) at $p$ level. For extracting the reconstructed wavelet information, we use the previous method extracting frequency information of $w$ replace by $c$. Thus, wavelet reconstruct information $c^{(p)}$ can also be extracted, and its covariance matrix is $C^{(p)}$.

4. PRINCIPAL COMPONENT ANALYSIS
The principal components give a new set of linearly combined measurements. Thus, we can use it to analyze our database and interpret the main variance. Thirty images (3 persons and 10 images/person) are used for the principal component analysis in our work. After training stage, the normalized shapes and their mean shapes are used to find covariance matrix. Let $X_{xcl}$ denote a cluster with $n$-dimension. There are $N$ samples in this cluster, and its mean can be written as

$$ \bar{X}_{2xcl} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad (3) $$
and its $n \times n$ covariance matrix $S$ is expressed as

$$ S_{2x2n} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^T. \quad (4) $$
Next, we choose the first 6 eigenvalues and rearrange them from the largest to the smallest. Thus the reconstructed form of $X$ can be expressed as

$$ X_{2xcl} = \bar{X}_{2xcl} + P_{2x6d}b_{6d}. \quad (5) $$
The eigenvector $P_{2x6d}$ denotes the landmark point’s variance, and parameter $b_{6d}$ denotes the weight. We can use equation (5) adjusting weight $b_{6d}$ to create a new shape, and give suitable limits.

$$ -3 \sqrt{\lambda_k} \leq b_k \leq 3 \sqrt{\lambda_k}, \quad k = 1,2,\ldots,6. \quad (6) $$
Where $k = 1$ indexes the largest variance of our training set, and $k = 6$ indexes the smallest one. Figure 1(b) illustrates some variances ($k = 1, 2, 3, 4$) of our database.

5. ACTIVE TEMPLATE SEARCH ON MULTIRESOLUTION WAVELET DECOMPOSE LEVEL
Although the wavelet decompose the $512 \times 512$ image into 10 levels, in our approach, only 5 levels (starting from level 5, and ending at level 9) are used to search object boundary. At the lowest level, the face position should be first located. At the higher level, we deform the shape and make it fit the object boundary. When the shape is deformed to the object boundary, we must adjust the translation, scale, rotation, and $b$ parameters. The translation operation is to control shape being moved to the selected position. Hence it includes $x$-direction and $y$-direction parameters.
5.1 Searching on level 5

First, we apply mean shape as the template that moves and compares it with wavelet coefficients on level 5 \( (p = 5) \). We define a template first as shown in Fig. 2(a). This template expresses common property, when a face decomposes on lower level. This template shows that two eyes, mouth, and chin belong to dark region; whereas two cheeks, nose, and forehead belong to bright region. Let \( D_i \) denote the dark region of average gray scale, and \( B_j \) denote the bright region of average gray scale. We can define respectively the contrast energy of \( E_1 \) and \( E_2 \) as

\[
E_1 = \sum_{i} D_i \quad \text{and} \quad E_2 = \sum_{j} B_j \exp(D_j - B_j).
\]

The profile match is measured by Mahalanobis distance, which is expressed respectively as

\[
E_3 = D_{\text{Mahalanobis}} = \sum_{r=1}^{N} (w_i^{(p)} - \bar{w}_i^{(p)})^T W^{-1} (w_i^{(p)} - \bar{w}_i^{(p)}), \quad \text{and} \quad E_4 = D_{\text{Mahalanobis}} = \sum_{r=1}^{N} (c_i^{(p)} - \bar{c}_i^{(p)})^T C^{-1} (c_i^{(p)} - \bar{c}_i^{(p)})
\]

Thus, the total energy is

\[
E = 10(E_1 + 0.1E_2) + 0.01(E_3 + E_4)
\]

This template is used to examine the whole image. The location of face should possess the minimum energy \( E \). Where the variances of \( W^{-1} \) and \( C^{-1} \) like to a weight of profile. If the variances are larger then we will obtain more variance of this profile. While the shape is moving to the selected position, then we need to update the shape parameters. Since the wavelet is sampled by power of 2, we need to multiply 2 of translation parameters for the next level. Thus the parameters translated from level 5 to level 6 must be updated as

\[
\begin{bmatrix}
    dx_{new}^{(5)} \\
    dy_{new}^{(5)}
\end{bmatrix}
= 
\begin{bmatrix}
    2dx_{old}^{(5)} \\
    2dy_{old}^{(5)}
\end{bmatrix}
\]

5.2 Searching on level 6

Now, we not only adjust translation but also estimate scale of the shape, and estimate the translation by smaller region. We reuse Mahalanobis distance defined in equation (8) to test the selected region. The energy is given by:

\[
E = E_1 + E_4
\]

Next, parameters on level 6 could be updated as (12) to the next level.

\[
\begin{bmatrix}
    dx_{new}^{(6)} \\
    dy_{new}^{(6)}
\end{bmatrix}
= 
\begin{bmatrix}
    2dx_{old}^{(6)} + 2dx^{(5)} + 2(x_0^{(6)} - s^{(6)}x_0^{(6)}) \\
    2dy_{old}^{(6)} + 2dy^{(5)} + 2(y_0^{(6)} - s^{(6)}y_0^{(6)})
\end{bmatrix}, \quad \text{and} \quad s_{new} = s_{old}^{(6)}
\]

Where \( (x_0^{(6)}, y_0^{(6)}) \) is the nose tip coordinate at the level 6.

5.3 Searching on level 7

On this level, the previous measure method involving the measure of rotation estimation is used. Then the parameters could be updated as (13) to the next level.
\[
\begin{bmatrix}
dx^{(7)}_{\text{new}} \\
dy^{(7)}_{\text{new}}
\end{bmatrix}
= \begin{bmatrix}
2dx^{(7)}_{\text{old}} + 2dx^{(6)} + 2(x^{(7)} - s^{(7)}x^{(7)}) \\
2dy^{(7)}_{\text{old}} + 2dy^{(6)} + 2(y^{(7)} - s^{(7)}y^{(7)})
\end{bmatrix}, \quad s^{(7)}_{\text{new}} = s^{(6)}_{\text{old}}, \quad \text{and } \theta^{(7)}_{\text{new}} = \theta^{(7)}_{\text{old}}
\] (13)

5.4 Searching on level 8

On this level, we will deform the shape from equation (5) to find an initial estimation of 3D-pose and expression. Besides, we have to use some different measures because high frequency components may appear at higher level. When we accept a shape deformation, the energy will be the smallest. The energy function is

\[
E_{\text{image}} = d_{\text{line}}E_{\text{line}} + d_{\text{similar}}E_{\text{similar}} + d'_{\text{contract}}E_{\text{contract}}
\]

\[
= -d_{\text{line}}\left(\frac{c_{\text{line}}^{(p)} - c_{\text{line}}^{(p')}}{c_{\text{line}}^{(max)} - c_{\text{line}}^{(min)}}\right) + d_{\text{similar}}E_{\text{similar}} + d'_{\text{contract}}\left(\sum_{k=0}^{n_{\text{sim}}^2} c_{k}^{(p)} - \sum_{k=0}^{n_{\text{sim}}^2} c_{k}^{(p')}\right)
\]

\[
d_{\text{contract}} = \text{sign}(f)d'_{\text{contract}}
\] (15)

Where \(d\) is the weight, \(n_{\text{p}}\) is the profile length, and \(n\) is the number the landmark point. Line energy term is used to control curve fit to the darkest line edge, and its range is from 0 to \(c_{\text{line}}^{(p)}\). We use \(3 \times 3\) local area to find local maximum and local minimum, which are denoted by \(c_{\text{line}}^{(p)}\) and \(c_{\text{line}}^{(p')}\), respectively. Besides, \(c_{\text{line}}^{(p)}\) is the current wavelet reconstructed value of \(3 \times 3\) center. Equation (15) possesses direction property depending on different curve as indexed in Fig. 2(b), where the corresponding sign value is listed in Fig. 2(c).

Then the parameters could be updated as equation (16) to the next level.

\[
\begin{bmatrix}
dx^{(8)}_{\text{new}} \\
dy^{(8)}_{\text{new}}
\end{bmatrix}
= \begin{bmatrix}
2dx^{(8)}_{\text{old}} + 2dx^{(7)} + 2(x^{(8)} - s^{(7)}x^{(7)}) \\
2dy^{(8)}_{\text{old}} + 2dy^{(7)} + 2(y^{(8)} - s^{(7)}y^{(7)})
\end{bmatrix}, \quad s^{(8)}_{\text{new}} = s^{(7)}_{\text{old}}, \quad \theta^{(8)}_{\text{new}} = \theta^{(7)}_{\text{old}}\text{, and } b_{\text{b1}(\text{new})} = b_{\text{b1}(\text{old})}
\] (16)

Where

\[
b_{\text{b1}} = \mathbf{P}_{\text{b1}}^{(8)}(\mathbf{X} - \overline{\mathbf{X}})\]

(17)

can be derived from equation (5).

5.5 Searching on level 9

First, we search a small range of each parameter. Next, we adjust single landmark point to the strongest object boundary with a small range. The single landmark point moving direction is perpendicular to the curve. We move each landmark point to the selected position that the energy function (14) is minimal. Next, we estimate the shape \(b\) parameters by equation (17) in order to avoid the shape deformed to the ill shape. Thus we use equation (18) as the limit, which can be obtained from equation (6).

\[
D_{m}^2 = \sum_{k=1}^{e} \frac{b_{k}^2}{\lambda_{k}} \leq D_{\text{max}}^2
\]

(18)

Where \(D_{\text{max}} = 3\), and \(e\) is the number of eigenvalues. If \(D_{m} < D_{\text{max}}\), then this shape is acceptable. Otherwise \(b\) parameters are adjusted by

\[
b_{\text{new}} = \left(\frac{D_{\text{max}}}{D_{m}}\right) b_{\text{old}}
\]

(19)

6. EXPERIMENTAL RESULTS

In our experiments, we use size of \(512 \times 512\) with gray level image as our input image, and use the tool of Watcom C to implement our algorithms. We use face images of Lanitis’s [1] database, which can be downloaded from the World Wide Web site of PEIPA (http://peipa.essex.ac.uk/). The database includes 300 training images, but only 3 persons per 10 images having different expressions, orientations, and races are used for training in our work. These training patterns include the conditions as: two males and one female, image background is white, face expression variable, head 3D pose variable, light source is fixed, head size variable, and no shelter. Where the test images include conditions are: PEIPA and our database are of 600 and 100 images respectively, sex mixed, image background variable, face expression
variable, head 3D pose variable, light source variable, head size variable, shelter variable, and each image contains one head. The presented experiments are of two folds: locating face region from image by shape template matching and positioning the face features by deforming active contour. Some results are shown in Fig. 3.

The proposed algorithms are implemented on a Pentium II 350MHZ Machine. It takes 2 seconds to obtain the face region, and takes 2~3 seconds to obtain face features. Thus the total time to locate the face features is about 4~5 seconds. Our shape template deformed to a new shape depends on the database limit. Hence if there is a strange object boundary, then our template will attempt to deform itself, which near this object boundary as close as possible, but its shape style will be limited by the database. Moreover, if the object has some shelters or the boundary not clear, then spline will be difficult for deforming to the object boundary.

7. CONCLUSION

Face feature extraction is an important but not easy topic since it has many variances, such as lighting effect, 3D-pose, expression and shelters. In this paper, we select a method of active template to locate face features. This template consists of different splines, and we use landmark points and database limits to control these splines deformed to object boundary. However, if we apply this template to searching object boundary on a whole image will spend much time. Hence, we apply wavelet deposition to control the template search from coarse to fine, which can improve the template searching efficiency and accuracy. As a conclusion, a feasible approach of locating face region and face feature from a gray image has been proposed in this paper. The improvements of correlation algorithm, the body recognition, and the pose analysis, will be our near feature works.

ACKNOWLEDGMENTS

This work was supported partially by National Science Council of Republic of China under Grants NSC89-2213-E-155-032.

REFERENCES


Fig. 3. Experimental results.