Gabor Feature Constrained Statistical Model for Efficient Landmark Localization and Face Recognition

Sanqiang Zhao [a,*], Yongsheng Gao [a], and Baochang Zhang [a,b]

[a] *School of Engineering, Griffith University, Nathan Campus, Brisbane, QLD 4111, Australia*
[b] *School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China*

Abstract: *Feature extraction and classification using Gabor wavelets have proven to be successful in computer vision and pattern recognition. Gabor feature based Elastic Bunch Graph Matching (EBGM), which demonstrated excellent performance in the FERET evaluation test, has been considered as one of the best algorithms for face recognition due to its robustness against expression, illumination and pose variations. However, EBGM involves considerable computational complexity in its rigid and deformable matching process, preventing its use in many real-time applications. This paper presents a new Constrained Profile Model (CPM), in cooperation with Flexible Shape Model (FSM) to form an efficient localization framework. Through Gabor feature constrained local alignment, the proposed method not only avoids local minima in landmark localization, but also circumvents the exhaustive global optimization. Experiments on CAS-PEAL and FERET databases demonstrated the effectiveness and efficiency of the proposed method.*

Keywords: *Face recognition, Elastic Bunch Graph Matching, Flexible Shape Model, Constrained Profile Model, Gabor wavelet.*

1. Introduction

Over the past few decades, the issue of automatic face recognition, where computers are still inferior to humans, has attracted great attention in the research community (Zhao et al.,

---

[*] Corresponding author. Tel.: +61 7 3735 3652; fax: +61 7 3735 5198.
 E-mail address: s.zhao@griffith.edu.au (S. Zhao).

2003). This is in large due to its extensive applications in various areas such as law enforcement, identity authentication, video surveillance and human-computer interfaces.

Existing technologies for face recognition can be roughly classified into holistic approaches and feature-based approaches. Using information derived from the whole face image, holistic approaches, such as Eigenface (Turk and Pentland, 1991) and Fisherface (Belhumeur et al., 1997), are conceptually simple and easy to implement, but their performance is affected by facial expression, pose and illumination changes in practice. On the other hand, feature-based approaches, such as Elastic Bunch Graph Matching (EBGM) (Lades et al., 1993; Wiskott et al., 1997), Line Edge Map (LEM) (Gao and Leung, 2002) and Directional Corner Point (DCP) (Gao and Qi, 2005), extract local information from salient facial features to distinguish faces. Represented by a set of low dimensional local feature vectors, these methods have the advantage of robustness to environmental variations.

The EBGM algorithm showed good identification accuracy in the FERET evaluation (Phillips et al., 2000), demonstrating the successful application of Gabor wavelets for face representation (Shen and Bai, 2006). In EBGM, face geometry is represented by an object-adapted (elastic) topology graph, where each node refers to a predefined fiducial point (or landmark). Local features are modeled by a set of Gabor coefficients (known as a *jet*) and labeled on each node. The comparison of different individuals is conducted through a graph matching strategy. As matching with each model graph is inefficient for large galleries, Wiskott et al. (1997) developed a stack-like structure, called Face Bunch Graph (FBG), to avoid such a process. A *bunch* is a set of jets taken from the same node over a small number of model graphs. To build the representing graph for a test face image, a model graph is first placed at an initial location and then is iteratively matched using Gabor jets to optimize its similarity with FBG, instead of with all the model graphs in the gallery. This matching process consists of two consecutive stages: 1) *rigid matching*: a model graph is scaled and shifted inside the test image while keeping its grid

rigid, which aims to account for global transformations; and 2) *deformable matching*: local distortions due to rotations in depth or expression variations are encoded by deforming individual nodes and evaluated by a graph similarity function within a topological constraint. To ensure accurate landmark localization, the time-consuming Gabor convolution is conducted on all the nodes and iteratively repeated in both stages. Although attractive identification accuracy was reported, the expensive computational cost prevents the use of EBGM in many applications.
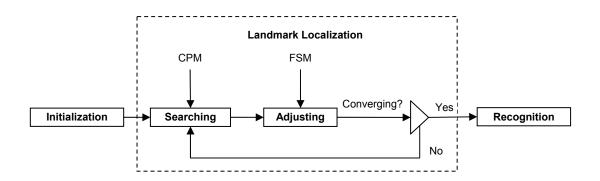
Many efforts have been made to improve the accuracy and robustness of EBGM. Würtz (1997) presented a system in which several practical procedures were integrated to increase the robustness of elastic graph matching against translations, deformations and changes in background. Tefas et al. (2001) employed support vector machines to derive optimal coefficients that weigh the local similarity values at the nodes of an elastic graph according to their discriminatory power. Zafeiriou et al. (2007) applied discriminant analysis techniques at multiple phases of elastic graph matching for face verification. Shin et al. (2007) extended elastic graph matching to a generalized EGM (G-EGM) to obtain enhanced performance on globally misaligned faces. Although attractive identification accuracies were reported, the expensive computational cost makes EBGM not suitable to many real-time applications. However, the efficiency issue of EBGM received less attention in the research community. Considering that most computationally expensive work of EBGM comes from Gabor convolution, Jiao et al. (2003) utilized E-M algorithm to model the Gabor feature distribution for landmark localization. The method achieved a performance better than Active Shape Model (ASM) (Cootes et al., 1995), but inferior to EBGM. Choi et al. (2008) presented a simplified version of Gabor wavelets (SGWs) and an efficient (but non-EBGM) algorithm for face recognition. Since the SGW was generated through quantizing its corresponding Gabor wavelet, there was a trade-off between computation time and approximation accuracy.

In this paper, we present a new Constrained Profile Model (CPM) to reduce the

computation time of face landmark localization and recognition. The landmarks are separated into several groups according to the distribution of salient facial features. Two *control points* are selected from each group to constrain other *non-control points* in the same group, preventing them from sticking into local minima during optimization. In CPM, each control point is modeled by a Gabor bunch, while all the non-control points are modeled by Normalized Derivative Profile (NDP) (Cootes et al., 1994). Employing both CPM and Flexible Shape Model (FSM) (Lanitis et al., 1997), we form an efficient localization framework (see Fig. 1), in which landmarks can be localized iteratively via two major steps, *searching* and *adjusting*. Based on a rough face detection result, the mean shape model of FSM is placed in a test image as the initial state. In the first searching step, a new suggested movement for each landmark is computed. Gabor jets are utilized to find the movements of control points, while the movements of non-control points are calculated by an efficient searching method based on NDP along the normal of the shape boundary (Cootes et al., 1994). Because NDP is a unidirectional texture model and has less accurate localization characteristics than Gabor jets, an extra procedure, called *Gabor feature constrained local alignment*, is conducted to refine the positions of non-control points in each group. In the second adjusting step, the pose parameters (i.e., scaling, rotation and translation) and shape parameters of FSM are adjusted in order to move each landmark as close as possible to the new suggested position. With the new parameters this process is repeated until the shape difference from two consecutive cycles of iteration is less than a predefined threshold. Different from EBGM, Gabor convolution in the proposed method applies only to a few control points, not to non-control points. Therefore the computational cost is significantly reduced.

Finally, face recognition is conducted using Gabor jets of control points as these jets have been calculated previously in the localization process. Gabor features from other landmarks, such as non-control points and interpolated locations, may be introduced into recognition to yield a higher recognition rate with extra computation. However, our experiments revealed that without

further using more landmarks, the proposed method is capable of obtaining comparable recognition accuracy to the standard EBGM algorithm with considerably less computation time.

Fig. 1. Framework of the proposed method.

The rest of this paper is organized as follows. Section 2 presents a detailed description of the process of landmark localization covering Flexible Shape Model, Constrained Profile Model, CPM searching and FSM adjusting. Section 3 describes the face representation and recognition. Section 4 reports our comparative experimental results on landmark localization and face recognition. The last section concludes the paper.

2. Landmark localization

The proposed framework incorporates both Flexible Shape Model (FSM) and Constrained Profile Model (CPM) for landmark localization. In this section, we first give a brief introduction of FSM, and then present the algorithm of building CPM. The two major steps in landmark localization, CPM searching and FSM adjusting, are described at last.

2.1. Flexible Shape Model

Flexible Shape Model (FSM) (Lanitis et al., 1997) is generated by a statistical analysis of the positions of the feature points from the training set to represent face shape variation due to differences between individuals as well as changes from environmental conditions. It ensures that

the instances of the shape model can only deform in ways found in a training set.

Mathematically, given a set of training images, the labeled N landmarks in each face image can be represented as a vector

$$\mathbf{x} = \left[x_1, y_1, x_2, y_2, \cdots, x_N, y_N \right]^T . \tag{1}$$

These shape vectors are first aligned using an iterative modified Procrustes algorithm (Gower, 1975). Principal Component Analysis (PCA) (Jolliffe, 1986) is then applied to get a set of orthogonal bases $\mathbf{P}$ to capture the statistics of the aligned landmarks. Each aligned shape can be approximated as

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} , \tag{2}$$

where $\mathbf{b}$ is the vector of eigenvalues, representing the shape parameters of each object. By modifying the values in $\mathbf{b}$, new instances of the shape model can be manipulated. When the elements of $\mathbf{b}$ are kept within certain limits (typically three times of the standard deviation), the new shape instance remains similar to those of the training set (Cootes et al., 1995). The shape parameters $\mathbf{b}$ can be retrieved from a given aligned shape $\mathbf{x}$ via

$$\mathbf{b} \approx \mathbf{P}^T \left(\mathbf{x} - \bar{\mathbf{x}} \right) . \tag{3}$$

2.2. Constrained Profile Model

As we employ an iterative process to obtain the optimized landmark locations, a suggested movement for each landmark has to be calculated based on the current shape estimate in each cycle of iteration. To find a suggested movement for each landmark, a proper local appearance modeling is required. Cootes et al. (1994) proposed a Normalized Derivative Profile (NDP) to model local gray-level information around each landmark in the training set of images, and suggested movements could be made along the normal of the shape boundary. NDP modeling is considerably fast. However, subsequent research (Wang et al., 2002) demonstrated that when the object class is inconsistent in shape and intensity appearance, this modeling method does not

perform well due to its ambiguous local-texture model and unidirectional searching strategy. Based on the observation that Gabor features are robust against illumination change, distortion and scaling (Kamarainen et al., 2006), Wiskott et al. (1997) proposed a stack-like Face Bunch Graph (FBG) to perform the coarse-to-fine Elastic Bunch Graph Matching (EBGM) and obtained noticeable localization accuracy. Each node of FBG is attached with a Gabor bunch to cover the local appearance variations. The representing graph is iteratively extracted from a test face image until a maximum similarity with FBG is reached. Although EBGM exhibits robustness against expression, illumination and pose variations, it involves considerable computational cost, as the optimized image graph is obtained through iteratively computing Gabor jets of all the nodes and distorting FBG in various sizes and positions over the entire image area.

2.2.1. Face landmark grouping

The proposed CPM is also a local appearance model obtained from the neighborhood of each landmark. It is represented by Gabor bunches on the control points and statistical descriptions of NDP on the non-control points. Specifically, the landmarks are classified into several groups based on the distribution of salient features in a face area (Zhao et al., 2004). Two control points are selected from each group for Gabor feature extraction. Because Gabor features are robust against local distortions, control points are able to constrain other non-control points in the same group, preventing them from sticking into local minima during optimization. Fig. 2 illustrates examples of face landmark grouping on two faces. Each face has seven groups of landmarks corresponding to two eyebrow regions, two eye regions, one nose region, one mouth region and one contour region. Landmarks in different groups are marked with different colors, while control points in all groups are identified by red color. Note that only one control point in the nose region is selected (Fig. 2) as the mid point of the two inner eye corners (the 45th and 49th control points in Fig. 6(a)) is calculated and used as the other control point.
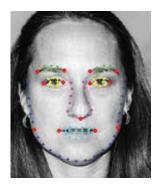
Fig. 2. Face landmark grouping.

2.2.2. Gabor feature extraction

Gabor wavelets are biologically motivated convolution kernels in the shape of plane waves restricted by a Gaussian envelope function. The 2D Gabor kernel functions used for feature extraction are

$$\psi_j(\mathbf{z}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 z^2}{2\sigma^2}\right)\left[\exp\left(i\mathbf{k}_j\mathbf{z}\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right], \tag{4}$$

where

$$\mathbf{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos\varphi_\mu \\ k_v \sin\varphi_\mu \end{pmatrix}, \quad k_v = 2^{-\frac{v+2}{2}}\pi, \quad \varphi_\mu = \mu\frac{\pi}{8}. \tag{5}$$

The index $j = \mu + 8\nu$ covers a discrete set of five different frequencies $\nu = 0,\cdots,4$ and eight orientations $\mu = 0,\cdots,7$. The width of the Gaussian is controlled by the parameter $\sigma = 2\pi$. A *jet J* is defined as the set $\{J_j\}$ of 40 convolution coefficients for kernels of different frequencies and orientations obtained at one image pixel $\mathbf{z} = (x, y)$ in an image $I(\mathbf{z})$:

$$J_j(\mathbf{z}) = \int I(\mathbf{z}')\psi_j(\mathbf{z} - \mathbf{z}')d^2\mathbf{z}'. \tag{6}$$

The set of jets referring to one feature point over a small number of model faces is called a *bunch*. Fig. 3 illustrates the visualized magnitudes and phases of the 40 Gabor kernel functions.
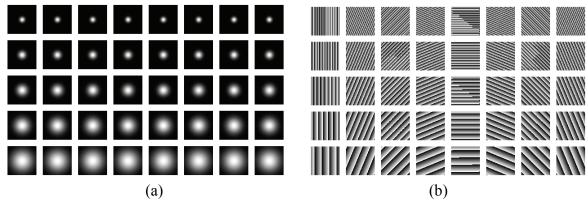
Fig. 3. Visualization of 40 Gabor kernels. (a) Gabor Magnitude. (b) Gabor Phase.

2.2.3.Building CPM

Different from EBGM where all the landmarks are modeled by Gabor features, the proposed CPM extracts Gabor features from control points only. Because the number of the control points is very small (13 in our experiments), the overall computational effort is significantly reduced. For the non-control points, the efficient NDP modeling is utilized. Eventually, each control point is modeled by a Gabor bunch to represent a range of possible variations in the local appearance of facial features; and each non-control point is modeled by a mean and a covariance matrix of NDP over all the training images. Algorithm 1 demonstrates the detailed process of building CPM. Note that when extracting a Gabor bunch from each control point, we only store the Gabor jet that is less similar to the current bunch, i.e., its magnitude similarity with all the existing jets in the current bunch is less than a predefined similarity threshold ($S_{mag}(J^{ik}, G_k) = \max_{b}\{S_{mag}(J^{jk}, J^{bk})\} < T_J$). This ensures that the extracted bunch includes jets from various conditions and hence covers different local distortions. The magnitude similarity function S_{mag} for measuring two jets will be introduced in the next subsection.

9

Algorithm 1. Building Constrained Profile Model

Notations:

I_i — the ith image, $i = 1, \cdots, M$

$\mathbf{z}_{ik}$ — the kth landmark in the ith image, $k = 1, \cdots, N$

J^{ik} — jet extracted from the kth landmark in the ith image

G_k — Gabor bunch of the kth landmark composed of a set of jets $\{J^{bk}\}$, $b = 1, \cdots, B$

T_J — magnitude similarity threshold of two jets

$\mathbf{p}_{ik}$ — NDP of the kth landmark in the ith image

Input: M training images and each labeled with N predefined landmarks.

Output: CPM (Gabor bunches from control points; means and covariance matrices of NDP from non-control points).

Procedures:

$b = 0$

for $k = 1 : N_{control\ point}$

    $G_k = \varnothing$

end for

for $i = 1 : M$

    for $k = 1 : N$

        if $\mathbf{z}_{ik}$ is control point

            if $b < B$

$$J_j(\mathbf{z}_{ik}) = \int I(\mathbf{z}')\psi_j(\mathbf{z}_{ik} - \mathbf{z}')d^2\mathbf{z}'; \quad J^{ik} = \{J_j(\mathbf{z}_{ik})\}, \quad j = 0, \cdots, 39$$

                if $b == 0$ $\|$ $\max_b\{S_{mag}(J^{jk}, J^{bk})\} < T_J$

                    $G_k += \{J^{ik}\}; \quad b++$

                end if

            end if

        else

            Calculate $\mathbf{p}_{ik}$

        end if

    end for

end for

for $k = 1 : N$

    if $\mathbf{z}_{ik}$ is non-control point

        $\bar{\mathbf{p}}_k = \frac{1}{M}\sum_{i=1}^{M}\mathbf{p}_{ik}; \quad \mathbf{C}_k = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{p}_{ik} - \bar{\mathbf{p}})(\mathbf{p}_{ik} - \bar{\mathbf{p}})^T$

    end if

end for

2.3. CPM searching

After having obtained CPM and FSM from the training set of images, landmark localization is implemented in an iterative two-step framework: CPM searching and FSM adjusting. The efficiency of CPM searching is realized through modeling Gabor bunches from a small number of control points. The accuracy of CPM searching is enhanced through conducting an extra procedure called Gabor feature constrained local alignment, which utilizes Gabor feature's accurate localization characteristics to refine the positions of non-control points. By increasing the number of control points, the localization accuracy can be improved with a cost of more time. To suit different circumstances in practice, one can adjust the number of control points to balance the computational efficiency and localization accuracy.

2.3.1.Individual landmark searching

Given initial positions of face landmarks, two different algorithms are employed to compute the suggested movements. The movement of non-control point is calculated by an efficient searching method based on NDP along the normal of the shape boundary; while the movement of control point is calculated by a Gabor jet localization algorithm. The jet localization algorithm performs displacement estimation based on jet phase information. For a test face image, a *test* jet is first extracted at the position of control point, and then compared one by one with all the *model* jets contained in the control point's Gabor bunch from CPM. The most similar model jet is selected using jet magnitude similarity for measurement:

$$S_{mag}\left(J,J'\right) = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}}, \tag{7}$$

where two jets J and J' are represented in the polar form $J_j = a_j \exp(i\phi_j)$. The magnitude similarity function is used here because jet magnitudes $a_j(\mathbf{z})$ vary slowly with spatial position,

11

which provides robustness to small position variation.

With the selected model jet and the test jet, the Gabor jet localization algorithm is conducted to find the suggested movement of the current control point. Since jet phases $\phi_j(\mathbf{z})$ vary quickly with location and rotate with a rate set by the spatial frequency $\mathbf{k}_j$ of the kernels, they provide a way of accurate displacement estimation (Wiskott et al., 1997). The jet displacement estimation is based on a displacement-compensated phase similarity function

$$S_{pha}(J,J') = \frac{\sum_j a_j a_j' \cos\left(\phi_j - \phi_j' - \mathbf{d}(J,J') \cdot \mathbf{k}_j\right)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}, \tag{8}$$

where $\mathbf{d}(J,J') = (d_x, d_y)^T$ is the relative displacement between two locations that jets J and J' refer to. The term $\mathbf{d}(J,J') \cdot \mathbf{k}_j$ is used to compensate for the phase shifts. The cosine terms $\cos(\phi_j - \phi_j' - \mathbf{d}(J,J') \cdot \mathbf{k}_j)$ in Equation (8) can be approximated by a two-term Taylor expansion:

$$S_{pha}(J,J') \approx \frac{\sum_j a_j a_j' \left(1 - 0.5(\phi_j - \phi_j' - \mathbf{d}(J,J') \cdot \mathbf{k}_j)^2\right)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}. \tag{9}$$

By maximizing Equation (9) via setting $\partial S_{pha}/\partial d_x = \partial S_{pha}/\partial d_y = 0$, we obtain the following equation for $\mathbf{d}(J,J')$:

$$\mathbf{d}(J,J') = \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \frac{1}{\Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx}} \begin{pmatrix} \Gamma_{yy} & -\Gamma_{yx} \\ -\Gamma_{xy} & \Gamma_{xx} \end{pmatrix} \begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix}, \quad \text{if } \Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx} \neq 0 \tag{10}$$

where $\Phi_x = \sum_j a_j a_j' k_{jx}(\phi_j - \phi_j')$, $\Phi_y = \sum_j a_j a_j' k_{jy}(\phi_j - \phi_j')$, $\Gamma_{xx} = \sum_j a_j a_j' k_{jx} k_{jx}$, $\Gamma_{xy} = \sum_j a_j a_j' k_{jx} k_{jy}$, $\Gamma_{yx} = \sum_j a_j a_j' k_{jy} k_{jx}$ and $\Gamma_{yy} = \sum_j a_j a_j' k_{jy} k_{jy}$. This yields a straightforward calculation for estimating the suggested movements of all the control points using the selected model jet and the test jet.

2.3.2. Gabor feature constrained local alignment

When the new suggested positions of two control points in each group are identified, an affine transformation is conducted to constrain non-control points in the same group. This process is considered as a local alignment between two sets of positions (one set before and one set after the local alignment), and is defined as a scaling (s), a rotation (θ) and a translation ($\mathbf{t_x} = (t_x, t_y)$) which minimize the sum of squared distances between pairs of positions. Such a coordinate transformation $A(s, \theta, \mathbf{t_x})[\ ]$ can be represented as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A(s, \theta, \mathbf{t_x})[\mathbf{x}] = \begin{pmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = \begin{pmatrix} a & -b & t_x \\ b & a & t_y \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{11}$$

Since we have obtained two pairs of coordinates from control points (before and after the jet displacement estimation respectively), the parameters of the affine transformation in Equation (11) can be calculated through

$$\begin{pmatrix} x_1 & -y_1 & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ x_2 & -y_2 & 1 & 0 \\ y_2 & y_2 & 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \end{pmatrix}. \tag{12}$$

Using the affine parameters from control points in each group, non-control points in the same group are locally aligned through Equation (11). Fig. 4 illustrates the effectiveness of Gabor feature constrained local alignment on landmarks in different regions.
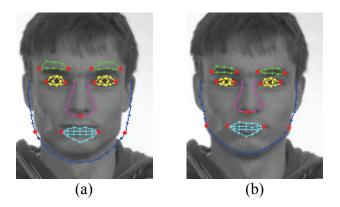

(a)   (b)

Fig. 4. Gabor feature constrained local alignment. (a) Before alignment. (b) After alignment.

2.4. FSM adjusting

When a test face image is presented to the proposed algorithm with a rough face detection result, the mean shape model $\overline{\mathbf{x}}$ of FSM is placed in the image as the initial state. Starting from the mean shape model $(\mathbf{x}^i = \overline{\mathbf{x}})$, all the landmarks are first moved to a better location $(\mathbf{x}^i + d\mathbf{x}^i)$ individually from their current locations $(\mathbf{x}^i)$ through CPM searching. Using a least-squared approximation method (Cootes et al., 1995), these shape deformations are then transformed into adjustments to the pose parameters $(s^i(1+ds^i),\ \theta^i + d\theta^i,\ \mathbf{t}_{\mathbf{x}}^i + d\mathbf{t}_{\mathbf{x}}^i)$ and the shape parameters $(\mathbf{b}^i + d\mathbf{b}^i)$ of FSM such that the new estimated shape $\hat{\mathbf{x}}^i = A(s^i(1+ds^i),\theta^i + d\theta^i, \mathbf{t}_{\mathbf{x}}^i + d\mathbf{t}_{\mathbf{x}}^i)[\overline{\mathbf{x}} + \mathbf{P}(\mathbf{b}^i + d\mathbf{b}^i)]$ moves the landmarks as close as possible to the previous local deformations, i.e. $\hat{\mathbf{x}}^i \approx \mathbf{x}^i + d\mathbf{x}^i$, while still satisfying the shape constraints of FSM $(-3\sigma_{\mathbf{b}_k} \leq \mathbf{b}_k^i + d\mathbf{b}_k^i \leq 3\sigma_{\mathbf{b}_k})$. With the estimated shape $\hat{\mathbf{x}}^i$, the above process is repeated $(\mathbf{x}^{i+1} = \hat{\mathbf{x}}^i)$ until the shape difference from two consecutive cycles of iteration is less than a threshold $(\| \hat{\mathbf{x}}^i - \hat{\mathbf{x}}^{i-1} \| < T_{\mathbf{x}})$. This dynamic shape-fitting process is summarized in Algorithm 2.

Algorithm 2. Landmark localization

Notations:

$\overline{\mathbf{x}}$ — mean face shape vector in FSM

$\mathbf{P}$ — principal components matrix in FSM

$\mathbf{x}^i$ — initial face shape for the ith iteration

$d\mathbf{x}^i$ — suggested movement for $\mathbf{x}^i$

$\mathbf{b}^i$ — initial weighting vector for $\mathbf{x}^i$

$d\mathbf{b}^i$ — additional weighting change for $\mathbf{b}^i$

$\sigma_{\mathbf{b}_k}$ — standard deviation of $\mathbf{b}_k$ in FSM, $k = 1, \cdots, N_{principal\ component}$

$\hat{\mathbf{x}}^i$ — estimated face shape from the ith iteration

$T_{\mathbf{x}}$ — shape difference threshold for two adjacent iterations

Input: One test face image with rough detection result.

Output: N landmark locations of the test image.

Procedures:

$\mathbf{b}^0 = \mathbf{0}$

$\hat{\mathbf{x}}^0 = \overline{\mathbf{x}}$

$i = 0$

do

$\qquad i{+}{+}$

$\qquad \mathbf{x}^i = \hat{\mathbf{x}}^{i-1}$

$\qquad$ Individual landmark searching from $\mathbf{x}^i$

$\qquad\qquad$ Search each control point using Gabor jet localization algorithm

$\qquad\qquad$ Search each non-control point using NDP method

$\qquad$ Gabor feature constrained local alignment for each group and get $d\mathbf{x}^i$

$\qquad$ FSM adjusting such that $A(s^i(1+ds^i), \theta^i + d\theta^i, \mathbf{t}_{\mathbf{x}}^i + d\mathbf{t}_{\mathbf{x}}^i)[\overline{\mathbf{x}} + \mathbf{P}(\mathbf{b}^i + d\mathbf{b}^i)] \approx \mathbf{x}^i + d\mathbf{x}^i$

$\qquad s^i * = 1 + ds^i$; $\theta^i {+}{=} d\theta^i$; $\mathbf{t}_{\mathbf{x}}^i {+}{=} d\mathbf{t}_{\mathbf{x}}^i$

$\qquad$ **for** $k = 1 : N_{principal\ component}$

$\qquad\qquad$ **if** $\mathbf{b}_k^i + d\mathbf{b}_k^i > 3\sigma_{\mathbf{b}_k}$

$\qquad\qquad\qquad \mathbf{b}_k^i = 3\sigma_{\mathbf{b}_k}$

$\qquad\qquad$ **else if** $\mathbf{b}_k^i + d\mathbf{b}_k^i < -3\sigma_{\mathbf{b}_k}$

$\qquad\qquad\qquad \mathbf{b}_k^i = -3\sigma_{\mathbf{b}_k}$

$\qquad\qquad$ **else**

$\qquad\qquad\qquad \mathbf{b}_k^i {+}{=} d\mathbf{b}_k^i$

$\qquad\qquad$ **end if**

$\qquad$ **end for**

$\qquad \hat{\mathbf{x}}^i = A(s^i, \theta^i, \mathbf{t}_{\mathbf{x}}^i)[\overline{\mathbf{x}} + \mathbf{P}\mathbf{b}^i]$

while $\| \hat{\mathbf{x}}^i - \hat{\mathbf{x}}^{i-1} \| \geq T_{\mathbf{x}}$

return $\hat{\mathbf{x}}^i$
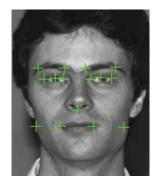
3. Face representation and recognition

For the simplicity of labeling images in the training process, most landmark are placed on the borders of facial features, such as eye corners, mouth corners and face contour. These features have clear and exact meaning and refer to the same positions across different faces. Besides, with homogeneous background the face contour is easier to be localized. The selection of control points is based on two basic principles. First, one control point should stay apart from the other in each group to avoid Gabor feature overlapping; and second, the neighborhood of two control points should contain rich feature information for better representation. In Davies et al. (2002), an automatic method for building statistical shape models was proposed, which can be used for automatic landmark selection.

In this study, a face image is represented by Gabor features of control points. This is mainly due to the fact that 1) most control points are located in the interior of a face where the most stable and informative features reside; 2) Gabor features of control points have been computed previously in the landmark localization process; and 3) selecting a small number of control points for representation reduces storage space and increases executive speed of the algorithm. Exceptionally, the centers of two eyes are also added into the representation landmark set, considering their importance for face recognition. For a test face image I^T with N' Gabor jets J_k^T, $k = 1, \cdots, N'$, a classification function between the test image I^T and a model image I^M is defined as the average over the magnitude similarity measures between pairs of corresponding jets:

$$S\left(I^T, I^M\right) = \frac{1}{N'} \sum_{k=1}^{N'} S_{mag}\left(J_k^T, J_k^M\right). \tag{13}$$

Face recognition is performed through calculating the classification function between the test face and each model face in the database. The model with maximum average similarity is considered as the correct match. Fig. 5 illustrates all the landmarks for face representation and
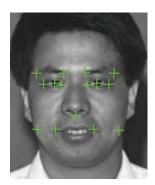
recognition in our experiments.

Fig. 5. Landmarks used for face representation and recognition.

Introducing Gabor features from more landmarks, such as some non-control points or interpolated locations, into recognition may yield a better performance with extra computational cost. However, our experiments in Section 4 demonstrate that, without further employing more landmarks, the proposed algorithm is able to obtain equivalent recognition accuracy to the standard EBGM method with significantly less computation time.
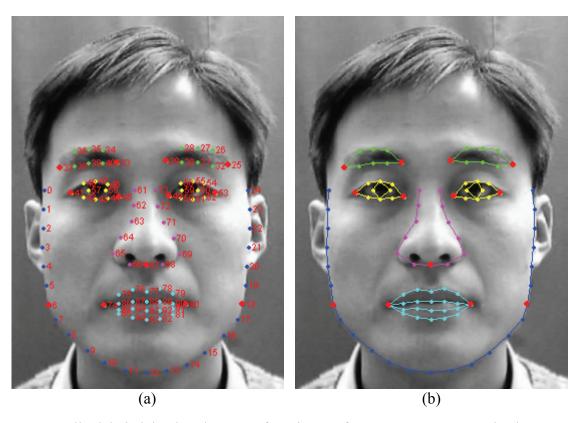
4. Experimental results

In order to validate the proposed algorithm, we separated our experiments into two parts: localization and recognition. Landmark localization is a critical step in a feature-based face recognition approach and is also important to achieve successful results in other applications such as expression analysis and face animation. Two publicly available face databases, CAS-PEAL database (Gao, et al., 2008) and FERET database (Phillips et al., 2000), were used to evaluate localization and recognition performance respectively.

4.1. Landmark localization performance

CAS-PEAL-R1 face database (Gao, et al., 2008) contains 9,060 frontal images of 1,040 subjects, with different expressions, accessories, illuminations, etc. The first 500 frontal faces under neutral expression and controlled lighting condition were used in this experiment to evaluate the landmark localization performance. Each image was manually labeled with 96

landmark points (see Fig. 6). Among these images, the first 300 images were used for training FSM and CPM, and the rest 200 images were used for localization testing. The manually labeled landmarks served as ground truth to assess the accuracy of the automatically located positions. As shown in Fig. 6, only 13 control points were selected for Gabor feature extraction. For each control point, the Gabor bunch size was 70. The length of NDP was 10 pixels, which is slightly longer than that (8 pixels) in Cootes et al. (1994) for a wider-range image search. The number of principal components in FSM was 45, which represents 95% of the variation in the training set, the same percentage as used in Lanitis et al. (1997). The jet magnitude similarity threshold for selecting models jets in each Gabor bunch was 0.95. The average shape difference threshold for two adjacent iterations was 1 pixel.

(a)   (b)

Fig. 6. Manually labeled landmarks on a face image from CAS-PEAL-R1 database. Control points are marked with red color. (a) 96 indexed landmarks. (b) A face shape defined by connecting landmarks.

Similar to Jiao et al. (2003), Lanitis et al. (1997) and Wang et al. (2002), the mean of point-to-point Euclidean distances between automatically located positions and the ground truth

landmarks was calculated to evaluate the accuracy of a localization algorithm on each test image. This mean distance provides a quantitative description of the localization error of the algorithm. The average localization error on all the test images is represented as

$$E = \frac{1}{P}\sum_{i=1}^{P}\left(\frac{1}{N}\sum_{j=1}^{N}\sqrt{(x_{ij}-x'_{ij})^2+(y_{ij}-y'_{ij})^2}\right),\qquad(14)$$

where P is the number of the test images, N is the number of the landmarks in each image, (x_{ij},y_{ij}) and (x'_{ij},y'_{ij}) are the jth ground truth landmark and the jth automatically located position in the ith test image, respectively. Based on Equation (14), the average localization error of the proposed CPM algorithm was compared with that of ASM method. A factor

$$m = \frac{E_{ASM}-E_{CPM}}{E_{ASM}}\times100\%\qquad(15)$$

was calculated to measure the improvement percentage of the proposed algorithm over ASM. When m is positive, the proposed algorithm outperforms ASM.

To evaluate the system localization performance under good and rough initialization conditions, we conducted the comparative experiments twice, with initialization using manually labeled two eye positions and Viola-Jones face detector (Viola and Jones, 2001) respectively. We used the Viola-Jones face detector in OpenCV library (OpenCV, 2008). When a face was detected, a bounding box containing the approximate location and size of the face was used for the initialization of our system. Table 1 shows the comparative localization performances on 200 test images. When manually labeled eye locations are used for initialization, the average localization error of ASM is 3.72 pixels, while CPM localization error is only 1.95 pixels. The 47.6% improvement of the proposed method indicates that CPM modeling and searching have the capability of obtaining much more accurate landmark localization than ASM. When rough automatic face detection results are used for initialization, the localization performance of ASM becomes noticeably worse with an average error of 11.63 pixels. Nevertheless, the performance

improvement of CPM over ASM increases from 47.6% to 54.5%, indicating the proposed method

has better landmark localization ability than ASM under rough initialization condition.

Table 1. Landmark localization performance

Method	ASM	Proposed CPM	Improvement
Average localization error using manually labeled two eyes (pixels)	3.72	1.95	47.6%
Average localization error using Viola-Jones face detector (pixels)	11.63	5.29	54.5%

4.2. Face recognition performance

The experiment of face recognition was performed on the FERET face database (Phillips et

al., 2000), which consists of a total of 14,051 gray-level face images representing 1,199

individuals. The images contain variations in lighting, facial expressions, pose angle, etc. Only

frontal faces were considered in this experiment. These frontal images are divided into five sets,

with details listed in Table 2. The FERET database provides locations of two eyes of all the

images, which were used for the algorithm initialization. The CPM and FSM obtained from

CAS-PEAL-R1 database were employed in this experiment. A training process conducted on

another database can provide objective evaluation of the algorithm performance on the FERET

database. Face recognition was performed using Gabor features of 15 landmarks (13 control
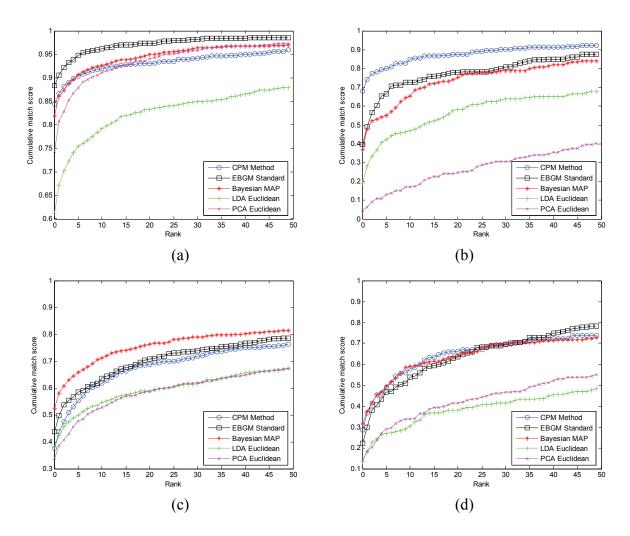
points plus 2 eye centers).

Table 2. FERET datasets

Datasets	Image Number	Description
FA	1,196	Regular facial expression, used as a gallery set.
FB	1,195	Alternative facial expression to FA.
FC	194	Different lighting conditions.
Dup I	722	Aging of subjects, i.e. images taken later in time.
Dup II	234	Difficult subset of Dup I. Taken at least one year later.

To conduct an objective comparison, the performances of the proposed algorithm were compared with the published results of the Colorado State University (CSU) Face Identification Evaluation System (Beveridge et al., 2005; CSU Face Identification Evaluation System, 2007), which is obtained on the FERET database as well. The CSU results provide identification accuracies of four face recognition algorithms, i.e., EBGM, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Bayesian intrapersonal/extrapersonal image difference classifier.

The cumulative match scores of these methods are plotted in Fig. 7, and the top-one recognition rates are listed in Table 3. Compared with standard EBGM, the proposed method yields slightly inferior top-one recognition accuracy on FERET FB and Dup I probe sets, each dropped by 3% and 5% respectively. This is not unexpected because our method makes use of Gabor features from only 15 landmarks, while the CSU implementation involves a total of 80 landmarks for identification (Beveridge et al., 2005; CSU Face Identification Evaluation System, 2007). On the other hand, our results on FC and Dup II sets are noticeably higher than EBGM, increased by 29% and 7% respectively. This is an interesting result especially in the sense of a small number of landmarks used for face recognition. This may be because when the test images are of significant detrimental quality (FC and Dup II are taken from different illumination and

after a considerable time respectively), Gabor feature based recognition performed on more landmarks could be worse than that from fewer reliable landmarks. The average recognition results revealed that the proposed method not only is capable of obtaining comparable recognition accuracy to the standard EBGM algorithm, but also outperforms the other three algorithms contained in CSU system, i.e. PCA, LDA and Bayesian algorithms. Through training of CPM and FSM on the FERET gallery set FA, the performance of the proposed method could be further improved.

Fig. 7. The cumulative match scores of the proposed CPM method and CSU implementations of four algorithms performed on the (a) FB, (b) FC, (c) Dup I and (d) Dup II probe sets.

Table 3. The top-one face recognition rates

Method	FB	FC	Dup I	Dup II	Average
CPM Method	0.84	0.68	0.38	0.29	0.55
EBGM Standard	0.87	0.39	0.43	0.22	0.48
Bayesian MAP	0.82	0.37	0.52	0.32	0.51
LDA Euclidean	0.61	0.19	0.38	0.14	0.33
PCA Euclidean	0.74	0.05	0.34	0.14	0.32

4.3. Computational efficiency

Table 4 lists the average computation time consumed in one cycle of iteration in our experiments. These results were obtained on a Windows-platform PC with 2.8GHz CPU and 1GB RAM. Operating on 96 landmarks, the NDP searching method spends less than 1ms to finish a cycle, while the EBGM method needs more than 1 second. With the computation time of 131ms, the proposed CPM method dramatically reduces the time of EBGM by 90.9%. This result indicated that with equivalent recognition performance, the proposed method can be around 10 times faster than EBGM. The implementation of our method converged within less than 6 cycles. The total computation time for one match is less than 1 second in average.

Table 4. Average computation time in one cycle of iteration

Method	EBGM Standard	CPM Method	Time Saved
Average Time (ms)	1440	131	90.9%

5. Conclusions

Gabor wavelets have demonstrated outstanding localizability, orientation selectivity and spatial frequency characteristics in image processing and pattern recognition. Gabor feature based

EBGM algorithm is one of the most successful approaches for face recognition due to its robustness against local distortions caused by variance of expression, illumination and pose. However, its considerable complication and computation prevents it from extensive applications in practice. We have presented an efficient Gabor feature constrained algorithm which can be used for locating facial features as well as identifying the individual in an image. Experimental results for both landmark localization and face recognition are promising. The recognition performance of the proposed method is comparable to the standard EBGM algorithm, and its computation time per cycle of iteration is 90.9% less than EBGM. We expect this method could be extended to non-rigid object recognition and applied to other HCI-related applications in the future.

Acknowledgements

References

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 711-720.

Beveridge, J.R., Bolme, D., Draper, B.A., Teixeira, M., 2005. The CSU Face Identification Evaluation System: its purpose, features, and structure. Machine Vision and Applications, 16(2): 128-138.

CSU (Colorado State University) Face Identification Evaluation System, 2007. http://www.cs.colostate.edu/evalfacerec/index.html.

Choi, W.P., Tse, S.H., Wong, K.W., Lam, K.M., 2008. Simplified Gabor wavelets for human face recognition. Pattern Recognition, 41(3): 1186-1199.

Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J., 1994. Use of active shape models for locating structures in medical images. Image and Vision Computing, 12(6): 355-365.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—Their training and application. Computer Vision and Image Understanding, 61(1): 38-59.

Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J., 2002. A minimum description length approach to statistical shape modeling. IEEE Transactions on Medical Imaging, 21(5): 525-537.

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D., 2008. The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 38(1): 149-161.

Gao, Y., Leung, M.K.H., 2002. Face recognition using line edge map. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(6): 764-779.

Gao, Y., Qi, Y., 2005. Robust visual similarity retrieval in single model face databases. Pattern Recognition, 38(7): 1009-1020.

Gower, J.C., 1975. Generalized procrustes analysis. Psychometrika, 40(1): 33-51.

Jiao, F., Li, S., Shum, H.-Y., Schuurmans, D., 2003. Face alignment using statistical models and wavelet features. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 321-327.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer.

Kamarainen, J.K., Kyrki, V., Kälviäinen, H., 2006. Invariance properties of Gabor filter-based features—Overview and applications. IEEE Transactions on Image Processing, 15(5): 1088-1099.

Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R.P., Konen, W., 1993. Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42(3): 300-311.

Lanitis, A., Taylor, C.J., Cootes, T.F., 1997. Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 743-756.

OpenCV (Open Computer Vision Library), 2008. http://sourceforge.net/projects/opencvlibrary/.

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., 2000. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(10): 1090-1104.

Shen, L., Bai, L., 2006. A review on Gabor wavelets for face recognition. Pattern Analysis and Applications, 9(2-3): 273-292.

Shin, H., Kim, S.D., Choi, H.C., 2007. Generalized elastic graph matching for face recognition. Pattern Recognition Letters, 28(9): 1077-1082.

Tefas, A., Kotropoulos, C., Pitas, I., 2001. Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(7): 735-746.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1): 71-86.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 511-518.

Wang, W., Shan, S., Gao, W., Cao, B., Yin, B., 2002. An improved active shape model for face alignment, In: Proceedings of the IEEE International Conference on Multimodal Interfaces, pp. 523-528.

Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 775-779.

Würtz, R.P., 1997. Object recognition robust under translations, deformations, and changes in background. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 769-775.

Zafeiriou, S., Tefas, A., Pitas, I., 2007. The discriminant elastic graph matching algorithm applied to frontal face verification. Pattern Recognition, 40(10): 2798-2810.

Zhao, S., Gao, W., Shan, S., Yin, B., 2004. Enhance the Alignment Accuracy of Active Shape Models Using Elastic Graph Matching, In: Proceedings of the International Conference on Biometric Authentication, pp. 52-58.

Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: a literature survey. ACM Computing Surveys, 35(4): 399-459.