1-1-2005

Face Recognition Using Landmark-based Bidimensional Regression

Jiazheng Shi
University of Michigan, sjiazen@med.umich.edu

Ashok Samal
University of Nebraska - Lincoln, asamal1@unl.edu

David Marx
University of Nebraska - Lincoln

Follow this and additional works at: http://digitalcommons.unl.edu/cseconfwork

Part of the Computer Sciences Commons

http://digitalcommons.unl.edu/cseconfwork/30

This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in CSE Conference and Workshop Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Face Recognition Using Landmark-based Bidimensional Regression

Jiazheng Shi  Ashok Samal  David Marx

Abstract
This paper studies how biologically meaningful landmarks extracted from face images can be exploited for face recognition using the bidimensional regression. Incorporating the correlation statistics of landmarks, this paper also proposes a new approach called eigenvalue weighted bidimensional regression. Complex principal component analysis is used for computing eigenvalues and removing correlation among landmarks. We evaluate our approach using two standard face databases: the Purdue AR and the NIST FERET. Experimental results show that the bidimensional regression is an efficient method to exploit geometry information of face images.

1 Introduction
The idea of face recognition based on geometry was proposed several decades ago [1, 2, 3]. Because face recognition has been receiving more and more attention, large standard face datasets have become available, and many state-of-the-art face recognition systems typically combine geometrical features and texture information. Several decades ago [1, 2, 3]. Because face recognition has been receiving more and more attention, large standard face datasets have become available, and many state-of-the-art face recognition systems typically combine geometrical features and texture information. Because face recognition performance can not be optimal and will in general not be as good as appearance based models which exploit significantly more information encoded in a face image. However, the study of geometry-based face recognition will contribute to advances in developing hybrid methods. Also, results of the face recognition vendor test (FRVT 2002) indicate that performance of human face recognition algorithms degrades as the size of a face search space increases. Therefore, landmark-based recognition can be used as a filtering operation that removes a large fraction of faces from being considered for a more accurate and expensive matching approach.

This paper studies how biologically meaningful landmarks extracted from face images can be used for face recognition based on bidimensional regression. Bidimensional regression is the process of fitting two objects represented by a set of two-dimensional (2D) points in terms of a predefined goodness-of-fit criterion. Bidimensional regression has the potential to play an important role in object recognition, for example, determining the similarity of two geographical maps, recognizing the difference between healthy and abnormal objects in medical images. A major concern with the bidimensional regression is the measurement of similarity. A typical goodness-of-fit criterion or the similarity for 2D object superposition is the Procrustes distance. The Procrustes distance has been used widely in statistical shape analysis [4]. It is used typically to compute the mean shape of an ensemble of 2D objects. Because bidimensional regression involves two objects, each of which is described by a set of points (called configuration), the Procrustes distance can be formulated to measure their similarity by comparing the landmark distributions. The Procrustes distance, however, does not incorporate statistical correlation of landmarks. Motivated by the Mahalanobis distance, we use complex principal component analysis (CPCA) to reduce the correlation of landmarks and use the eigenvalues to normalized their distribution.

2 Bidimensional regression
Because landmarks are mapped to two-dimensional face images, it is convenient to use a complex number notation for each landmark. Let \( \{l_i \in \mathbb{C}\}_{i=1}^{N} \) be the set of \( N \) landmarks on a face image. A face \( f \) then can be modeled by a \( N \)-dimensional vector of landmarks as follows:

\[
f = [l_1, l_2, ..., l_N]^T \quad \text{with} \quad l_i = x_i + j y_i, \tag{1}\]

where \((x_i, y_i)\) are coordinates of \( l_i \) and \( j = \sqrt{-1} \).

Euclidean transform (also called linear conformal transform) is a very popular method of bidimensional regression. Let \((x_{1i}, y_{1i})\) and \((x_{2i}, y_{2i})\) be the one-to-one corresponding landmarks of face configurations \(f_1 = [l_1, l_2, ..., l_N]^T \) and \(f_2 = [l_{11}, l_{12}, ..., l_{2N}]^T \), respectively. The Euclidean transform is defined as:

\[
\begin{bmatrix}
x_{1i} \\
y_{1i}
\end{bmatrix} = \begin{bmatrix}
a \\
b
\end{bmatrix} + \beta \begin{bmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{bmatrix} \begin{bmatrix}
x_{2i} \\
y_{2i}
\end{bmatrix}, \tag{2}
\]

where \(a\) and \(b\) represent translation, \(\beta\) is a positive real number corresponding to scaling (we assume that the scaling is isotropic), \(\theta \in [0, 2\pi)\) corresponds to rotation. With
complex numbers, Equation 2 can be rewritten as:

$$l_{1i} = (a + jb) + \beta e^{i\theta}l_{2i}, \quad (3)$$

where $a + jb$ is the complex number to represent translation. The four parameter values can be determined exactly by two independent landmarks.

However, if the number of landmarks ($N$) is more than two, bidimensional regression is required for solving optimal parameter values subject to an objective fitness criterion. Mathematically, the Euclidean transform between two face configurations $f_1$ and $f_2$ is defined as

$$f_1 = (a + jb)1_N + \beta e^{i\theta}f_2 + \epsilon, \quad (4)$$

where $1_N$ is an $N$-dimensional vector of all ones and $\epsilon$ is the residual of bidimensional regression. The least-squares error is the most typical objective function. Without loss of generality, $f_1$ and $f_2$ can be centralized such that $a + ib = 0$. The least-squares function $J$ of $\beta$ and $\theta$ can then be defined as:

$$J(\beta, \theta) = \| \beta e^{i\theta} f_2 - f_1 \|_2^2, \quad (5)$$

where $\| \cdot \|_2$ is the $l_2$ norm.

For two object superposition, the goodness-of-fit of bidimensional regression of the Euclidean transform is measured by the the Procrustes distance, defined as:

$$\delta_p(f_1, f_2) = \inf_{\beta, \theta} \| J(\beta, \theta) \|_2^2 = \left\| f_1^H f_1 - \frac{f_1^H f_2 f_2^H f_1}{f_2^H f_2} \right\|_2^2, \quad (6)$$

where $\overline{f_1} = \frac{1}{N} \sum_{i=1}^{N} l_i$ and ‘$H$’ denotes the Hermitian transpose.

3 Eignevalue-weighted bidimensional regression

The normal bidimensional regression does not consider the correlation of landmarks. Motivated by the Mahalanobis distance, we propose an eigenvalue-weighted bidimensional regression method. The basic idea is to remove the correlation of landmarks and build a unit sphere by distribution for the similarity distance. The basis steps include normalizing each face configuration into a pre-shaped space, performing a complex principal component analysis, and using a refined similarity measure.

3.1 Pre-shaped space

The Procrustes distance is independent of the geometric constraints of translation, scaling, and rotation. Translation-independence can be achieved by moving the geometric center of a configuration to the origin of the coordinate system. In the matrix form,

$$f_1 = (I_N - \frac{1}{N} 1_N 1_N^H) f, \quad (7)$$

where $I_N$ is a $N \times N$ identity matrix. Scaling is the process of normalizing the centered configuration. A configuration that is free from the translation and scaling is called the pre-shape of the configuration [4]. Each face configuration is normalized into the pre-shaped space, so the following discussion is based on pre-shaped face configurations.

3.2 Complex principal component analysis

PCA has been used in various scientific fields. The essential idea behind PCA is to reduce dimensionality by removing the redundant and less significant components within a dataset governed by a large set of correlated variables [5]. PCA is applied typically to configurations that are modeled by vectors of real numbers. If a configuration (e.g., a digital image) is modeled by a matrix of real numbers, a popular method is stack either rows or columns alphabetically into a vector of real members. For the landmark-based face model $f$ with $N$ landmarks, every landmark $l_i$ is two-dimensional. In order to use PCA in real number domain, $f$ can be stacked into a $2N$ dimensional vector as

$$f = [x_1, y_1, x_2, y_2, ..., x_N, y_N]^T. \quad (8)$$

This paper, however, studies PCA in the complex number domain, in order to keep the relationship of $x$ and $y$ coordinates. Because pre-shaping removes translation of all configurations, the smallest eigenvalue is zero. Eigenvectors are composed of complex numbers and eigenvalues are nonnegative real numbers. Mathematically, complex PCA is an extension of real number PCA and have the same formulation.

3.3 Refined similarity measure

To incorporate statistical characteristics of a landmark distribution, landmarks are weighted by their eigenvalues in the pre-shaped space. Let $\lambda_1^2, \lambda_2^2, ..., \lambda_N^2$ be the $N$ eigenvalues of the covariance matrix for $\hat{l}_1, \hat{l}_2, ..., \hat{l}_N$, respectively. The eigenvalue-weighted configuration is defined as:

$$\hat{f} = \hat{l}_1 \lambda_1^{-1}, \hat{l}_2 \lambda_2^{-1}, ..., \hat{l}_N \lambda_N^{-1}^H. \quad (9)$$

The refined similarity measure for two faces $\hat{f}_1$ and $\hat{f}_2$ is defined as:

$$\delta_{e,p}(\hat{f}_1, \hat{f}_2) = \sum_{k=1}^{N} | \frac{\hat{l}_k}{\lambda_k} - \frac{\hat{l}_k}{\lambda_k} |^2. \quad (10)$$

4 Experimental results

4.1 Datasets and predefined landmarks

Two datasets are used to evaluate the performance of our approach: the Purdue AR face image database [6] and the NIST FERET database [7]. The Purdue AR database contains 70 male and 56 female young adult faces. All images were taken in two sessions with 14 days interval under strictly controlled conditions. We use the Purdue AR
dataset primarily to investigate whether landmarks collected manually are reliable or not. The FERET database is a well-known face image database used for uniform evaluation of face recognition algorithms. We use four well-defined frontal-view sets of this large database. $F_a$ (with 994 images) is designed as the gallery $G$ for training and $F_b$ (with 992 images), $D_u p$ I (with 726 images), and $D_u p$ II (with 228 images) as the three probe face sets for evaluating performance. The landmarks for the faces are collected manually. Automated extraction also could be used to locate features. Figure 1 illustrates the distributions of the 29 biologically meaningful landmarks of 994 face images in $F_a$. The 29 cross signs indicate mean landmarks of 29 clusters of landmarks. All landmark-based face configurations are normalized independently into the pre-shaped space.

4.2 Reliability of the landmarks

Before conducting any analysis, we want to evaluate the reliability of the manual feature extraction method. We use a statistical approach and pose the following null hypotheses, $H_0$: readers and repetitions have no significant impact on landmark coordinates. The null hypothesis $H_0$ is tested by using analysis of variance (ANOVA). Instead of the classic ANOVA, the mixed ANOVA procedure is applicable here in order to account for random effects.

We use the Purdue AR dataset to test the null hypothesis $H_0$. In our experiment, four human subjects were randomly invited as readers to collect the landmarks on all face images. They repeated landmark collection of the same face images after an interval of at least one week. We use the MIXED procedure of SAS, for our analysis of variance [8]. Table 1 lists the estimates of variance components and p-values of a variance component being 0. The major variances of landmark coordinates (AXIS-X and AXIS-Y) are determined by face images themselves, i.e., the ‘FACE(GENDER)’ effect. Table 1 clearly shows that the variances introduced by a reader ‘READER’ and repetitions by a reader ‘REPEAT(READER)’ are sufficiently small compared with other random effects and even the residual. Therefore, the null hypotheses $H_0$ can not be rejected and the landmark coordinates collected by readers can be used with confidence.

4.3 Performance

As the size of face datasets becomes larger, the scalability becomes an important factor. Bidimensional regression with the Procrustes distance can be used to reduce the search space by measuring how often the actual match falls in the top $K$ fraction, e.g., top 1%, top 10%, and so on. The relative rank $K$ of the face normalizes its absolute rank with the number $G$ of faces in the gallery to the range of $[\frac{k}{G}, 1]$. Figure 2(A) plots the cumulative match score $\Psi_{K}(0.01)$ for face recognition of the FERET datasets using the 29 landmarks and the Procrustes distance. The probe set $F_b$ has the highest cumulative match scores overall. As noted in [7], $D_u p$ II is the hardest probe set and has the lowest score as expected. The experimental results show that variations of facial expression have less impact on the performance than variations due to aging, if the landmark-based face configuration and the Procrustes distance are used together. When a relative rank of $K = 0.01$ (top 1%) is used, the cumulative match scores $\Psi_{K}(0.01)$ for $F_b$, $D_u p$ I, and $D_u p$ II are 0.594, 0.542, and 0.496 respectively. As the relative rank $K$ increases to 0.1 (top 10%), $\Psi_{K}(0.1)$ for $F_b$, $D_u p$ I, and $D_u p$ II increase to about 0.823, 0.784, and 0.728 respectively.

In order to study the performance of the refined similarity measure, we normalize all landmarks in $F_a$, $F_b$, $D_u p$ I, and $D_u p$ II into the pre-shaped space. The gallery $F_a$ is used as the training set for computing covariance matrix. Using CPCA, we select the 23 most significant eigenconfigurations for the refined similarity measure. Figure 2(B) plots the cumulative match score $\Psi_{K}(0.01)$ for face recognition of the FERET datasets using the refined similarity measure. All probe sets $F_b$ and $D_u p$ I have almost the same cumulative match scores. This experimental result indicates that the refined similarity measure can account for variations of facial expression and aging, which are regarded as two challenging problems by many state-of-the-art face recognition algorithms. When a relative rank of $K = 0.01$ (top 1%) is used, the cumulative match scores $\Psi_{K}(0.01)$ for $F_b$, $D_u p$ I, and $D_u p$ II are

Table 1: Estimates of variance component and p-values.

<table>
<thead>
<tr>
<th>Covariance parameter (SAS notation)</th>
<th>estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXIS-X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>READER</td>
<td>0.337</td>
<td>0.267</td>
</tr>
<tr>
<td>REPEAT(READER)</td>
<td>0.054</td>
<td>0.258</td>
</tr>
<tr>
<td>FACE(GENDER)</td>
<td>159.534</td>
<td>0.0001</td>
</tr>
<tr>
<td>LANDMARK*FACE(GENDER)</td>
<td>28.898</td>
<td>0.0001</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>15.563</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Figure 1: Landmark distributions.
0.794, 0.780, and 0.772 respectively. As the relative rank \( K \) increases to 0.1 (top 10%), \( \Psi_{\text{ep}}(0.1) \) for \( F_b, \text{Dup I}, \) and \( \text{Dup II} \) increase to about 0.928, 0.933, and 0.925 respectively. These results indicate that landmark-based face configuration combined with the refined similarity measure can reduce efficiently face search space for other face recognition algorithms.

To evaluate the sensitivity and specificity of the results, Figure 3(A) illustrates the distributions of the match and mismatch of the Procrustes distances for the \( F_b \) dataset. (The probe sets \( \text{Dup I} \) and \( \text{Dup II} \) have similar results.) We can see that the overlapping region of the match and mismatch distributions is relatively small with an area of 0.402. (Note that the area of each curve is 1.0.) Similarly, Figure 3(B) illustrates the distributions of the match and mismatch of the refined similarity measure for the \( F_b \) dataset. The overlapping region is even smaller, with an area of 0.270.

5 Conclusions

This paper studies how biologically meaningful landmarks extracted from 2D face images can be used for face recognition based on the bidimensional regression. Both the normal bidimensional regression and the eigenvalue weighted bidimensional regression are examined. Because the latter incorporates correlation statistics among landmarks, it is expected to perform better. Our investigation shows that a human face can be modeled efficiently by well-defined biologically meaningful landmarks. While we obtained the landmarks manually, we have shown using a mixed ANONA model that the landmark coordinates are minimally impacted by readers or repetitions. The Procrustes distance is an effective measure to compute the similarity for the landmark-based 2D face model and the refined similarity measure works better.

References