Pose Estimation Based on Gaussian Error Models

Xiujuan Chai[1], Shiguang Shan[2], Laiyun Qing[2], and Wen Gao[1,2]

[1] School of Computer Science and Technology, Harbin Institute of Technology, 150001 Harbin, China
[2] ICT-ISVISION Joint R&D Lab for Face Recognition, ICT, CAS, 100080 Beijing, China
{xjchai, sgshan, lyqing, wgao}@jdl.ac.cn

Abstract. In this paper, a new method is presented to estimate the 3D pose of facial image based on statistical Gaussian error models. The basic idea is that the pose angle can be computed by the orthogonal projection computation if the specific 3D shape vector of the given person is known. In our algorithm, Gaussian probability density function is used to model the distributions of the 3D shape vector as well as the errors between the orthogonal projection computation and the weak perspective projection. By using the prior knowledge of the errors distribution, the most likely 3D shape vector can be referred by the labeled 2D landmarks in the given facial image according to the maximum posterior probability theory. Refining the error term, thus the pose parameters can be estimated by the transformed orthogonal projection formula. Experimental results on real images are presented to give the objective evaluation.

1 Introduction

Human head pose estimation is the key step towards the multi-view face recognition[1] and other multimedia applications, such as the passive navigation, industry inspection and human-computer interface and so on [2]. With these applications more and more techniques are investigated to realize the robust pose estimation.

Existing pose estimation algorithms can be classified into two main categories, one is the model-based algorithm, and the other is the appearance-based method. Model-based methods first assume a 3D face model to depict face. Then erect the relation of the features between 2D and 3D, finally the conventional pose estimation techniques are used to recover the pose information. Appearance-based algorithms suppose that there is one and only correlation between the 3D pose and the characteristics of 2D facial image. So the aim is to find this mapping relation from lots of training images with the known 3D poses. Here, the characteristics of the facial image conclude not only the intensities, color but also the intensity gradient and all kinds of image transformations etc.

Many appearance-based approaches have been reported on pose estimation. Hogg proposed a method to construct the mapping relation between 2D facial image and the 3D face pose by using artificial Neural Networks [3]. Later, Darrell performed face

D. Zhang and A.K. Jain (Eds.): ICB 2006, LNCS 3832, pp. 136 – 143, 2005.
© Springer-Verlag Berlin Heidelberg 2005

detection and pose estimation by eigen-space method [4]. A separate eigen-space was erected for every pose of each training face. Given an image, projecting it to each eigen-space, the face and its pose were determined by the eigen-space which has the minimum error term. The similar idea was also appeared in paper [5]. Exclusive correlation between the 3D pose and its projection to the eigen-space is the potential assumption of this kind of eigen-space methods. A skin color model based pose estimation algorithm was proposed in [6], where the head was modeled by the combination of the skin/hair regions. In summary, the appearance based methods usually need lots of facial images under many poses for different persons to do training. They are simple in computing, however, not very accurate since many of them require interpolation.

Many model-based approaches have also been reported in the literature. Most of them model a face with some feature, for example the cylinder, the eclipse, or some key feature points. Then the 2D features are matched to the corresponding 3D feature to get the face pose. Nikolaidis determined the face pose by the equilateral triangle composed by the eyes and mouth [7]. Similarly, Gee used a facial model based on the ratios of four world lengths to depict the head [8,9]. Under the assumption of the weak perspective projection, the ratio of the 2D/3D lengths and the plane skew-symmetry are investigated to compute the normal and estimate the head pose finally. Except these methods, more complicated models were also proposed to tackle the pose estimation problem. Lee used a general 3D face model to synthesis eight different poses facial images [10]. The correlation between the input image and the modeled images were calculated to give the pose estimation results. More complicated, Ji and Hu assumed that the shape of a 3D face could be approximated by an ellipse and the aspect ratio of 3D face ellipse was given in advance [11]. So the ratio of the detected major axis and minor axis was used to calculate the face pose. To sum up, these model-based methods are more reliable and robust if the features can be detected accurately.

Our pose estimation method is also a model-based algorithm. In this paper, face is modeled by five landmarks. Using the MAP theory, the specific 3D shape vector corresponding to the given face is inferred and then used to get the accurate 3D pose.

The remaining parts of the paper are organized as follows: In Section 2, a simple pose estimation idea based on orthogonal projection is introduced. Then against the two problem existed in the above method, we propose a novel pose estimation algorithm based on the Gaussian error models in Section 3. Some pose estimation results of our algorithm are presented in Section 4, followed by short conclusion and discussion in the last section.

2 Pose Estimation Based on Simple Orthogonal Projection

We know that the head can be approximated as a 3D rigid body within the 3D coordinate system, hence the pose variation also satisfied the regular pattern of rigid motion. The face images under different poses can be regarded as the different projections in 2D image plan for different rotations around the head center.

In this paper, the pose variation is denoted by the tilt-yaw-pitch rotation matrix. The definition of the rotation angles is illustrated as Fig.1.

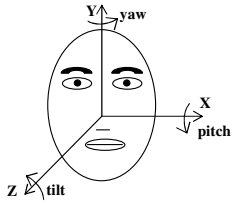

Fig. 1. The definition of the three rotation angles

Thus the rotation matrix $\mathbf{R}$ can be represented by:

$$\mathbf{R} = \mathbf{R}_Z(\gamma)\mathbf{R}_Y(\beta)\mathbf{R}_X(\alpha) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}. \tag{1}$$

In our method, five landmarks are used to model the head, which are the left and right iris centers, nose tip, left and right mouth corners respectively. The five points of 2D facial image can be written as a 2×5 matrix $\mathbf{S}_f$, where $\mathbf{S}_f = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{pmatrix}$. In a similar way the corresponding 3D points can be reshaped into a 3×5 matrix $\mathbf{S}$. Basing on the orthogonal projection theory, the following equation holds:

$$\mathbf{S}_f = c\mathbf{PRS} + \mathbf{T}, \tag{2}$$

where c is the scale factor, $\mathbf{T}$ is the 2D translation vector on x and y orientation. And $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ is a transformation matrix to throw away the z information. We can obtain the pose parameters from equation (2) if the 3D head model $\mathbf{S}$ is known. Because the $\mathbf{S}$ is unknown for a specific given face, the average 3D face model can be used to substitute the specific $\mathbf{S}$ to get the approximate pose angles.

3 Gaussian Error Models Based Pose Estimation Algorithm

The above method will get good solutions for the faces whose 3D structures are similar with the average 3D face model. While, it will lead to large errors for those faces having remarkable different 3D structures compared with the general face. We think that there are two major factors introducing the deviations:

- ✧ The 3D shape vector $\mathbf{S}$ is different from each other. The using of the average shape $\mathbf{S}$ inevitably imports the deviation more or less.
- ✧ The facial images we estimate are almost generated by weak perspective projection. The orthogonal projection computation with the feature landmarks of real facial image will also generates indeterminate deviations.

Considering these factors, we modify the equation (2) as:

$$\mathbf{S}_f = \mathbf{PRS} + \mathbf{e}. \tag{3}$$

In this equation, the 2D shape vector $\mathbf{S}_f$ and the 3D shape vector $\mathbf{S}$ is aligned to the same standard position and scale to statistic the error distribution. Error term $\mathbf{e}$ is a 2×5 matrix. The distribution of the error terms can be modeled by a Gaussian probability density function.

Our pose estimation based on Gaussian error models algorithm consists of 2 steps: statistical error models computation and the pose estimation of facial image. In the following paragraphs, the two steps will be described in turn.

3.1 Learning the Gaussian Error Models

Our training set includes 100 laser-scanned 3D faces selected from the USF Human ID 3-D database [12]. The 3D shape vectors can be denoted as $\{\mathbf{S}_1 \quad \mathbf{S}_2 \quad \cdots \quad \mathbf{S}_n\}$, where $n=100$. $\mathbf{S}_i$ is a 3×5 matrix. The mean vector and the covariance matrix of these vectors can be computed by : $\boldsymbol{\mu}_{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_i$ and $C_{\mathbf{s}} = \frac{1}{n}\sum(\mathbf{S}_i - \boldsymbol{\mu}_{\mathbf{s}})(\mathbf{S}_i - \boldsymbol{\mu}_{\mathbf{s}})^T$.

To simplify the statistical procedure, the error term $\mathbf{e}_{\mathbf{R}}^n$ under each sampling pose for a face is computed by the imaging formula directly. Computing the orthogonal projection and the weak perspective projection for the five points respectively, we get the two vectors: orthogonal projection vector $\mathbf{V}_{orth}$ and perspective projection vector $\mathbf{V}_{per}$. In order to normalize these two vectors, we align them in scale and make them have the same barycenter. Then we have error term $\mathbf{e}_{\mathbf{R}}^n$ by $\mathbf{e}_{\mathbf{R}}^n = \mathbf{V}_{per}^n - \mathbf{V}_{orth}^n$, where n is the index of different training shape. Under each sampling pose, we statistic the error mean vector $\boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}}$ and the covariance matrix $\mathbf{C}_{\mathbf{e}}^{\mathbf{R}}$. Having these statistical Gaussian error models, the concrete pose estimation algorithm is described later.

3.2 Pose Estimation Based on the Gaussian Error Models

When given a facial image, first, let average 3D shape $\overline{\mathbf{S}}$ be $\mathbf{S}$ and the error term $\mathbf{e}$ be zero. So the approximate pose $\mathbf{R}_0$ can be computed by equation (3): $\mathbf{S}_f = \mathbf{PRS} + \mathbf{e}$. Set the $\mathbf{R} = \mathbf{R}_0$, the specific 3D shape of the given face is computed by the maximum posterior probability and the error term $\mathbf{e}$ can be calculated subsequently. In the first place, the mean vector $\boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}}$ and the covariance matrix $\mathbf{C}_{\mathbf{e}}^{\mathbf{R}}$ of the error are refined by the simple neighborhood weighted strategy.

After refining the mean and covariance of error, we can recover the specific 3D shape $\mathbf{S}$ for the given face. As we all know that $\mathbf{S}_{MAP} = \arg\max_S(P(\mathbf{S}\,|\,\mathbf{S}_f))$. It is difficult to compute the $\arg\max_S(P(\mathbf{S})\,|\,\mathbf{S}_f)$ directly, so we use the Bayes' rule $P(\mathbf{S}\,|\,\mathbf{S}_f)P(\mathbf{S}_f) = P(\mathbf{S}_f\,|\,\mathbf{S})P(\mathbf{S})$ to simplify $\mathbf{S}_{MAP}$. As the $\mathbf{S}_f$ is definite, the $P(\mathbf{S}_f)$ is a constant, thus we have:

$$\mathbf{S}_{MAP} = \arg\max P(\mathbf{S}_f\,|\,\mathbf{S})P(\mathbf{S}) \tag{4}$$

where $P(\mathbf{S})$ is the Gaussian probability density function we have learned in advance. From equation (3), if the $\mathbf{S}$ is fixed, and then $P(\mathbf{S}_f)$ is also a Gaussian probability density function with mean $(\mathbf{PRS}+\boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}})$ and covariance matrix $\mathbf{C}_{\mathbf{e}}^{\mathbf{R}}$. So, we have:

$$\mathbf{S}_{MAP} = \arg\max_{\mathbf{S}} \left(Gauss\left(\mathbf{PRS}+\boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}}, \mathbf{C}_{\mathbf{e}}^{\mathbf{R}}\right) \times Gauss\left(\boldsymbol{\mu}_{\mathbf{S}}, \mathbf{C}_{\mathbf{S}}\right) \right) \tag{5}$$

Using log probability for the right segment of equation (5), then set the first derivative with respect S to 0 to get the maximum probability, we get:

$$-(\mathbf{PR})^T \cdot (\mathbf{C}_{\mathbf{e}}^{\mathbf{R}})^{-1} \cdot \left(\mathbf{S}_f - \mathbf{PRS} - \boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}}\right) + \left(\mathbf{C}_{\mathbf{S}}\right)^{-1}\left(\mathbf{S}-\boldsymbol{\mu}_{\mathbf{S}}\right) = 0 \tag{6}$$

Rearranging the equation (6), we can obtain the following linear equation: $\mathbf{A} * \mathbf{S} = \mathbf{T}$, where, $\mathbf{A} = (\mathbf{PR})^T \left(\mathbf{C}_{\mathbf{e}}^{\mathbf{R}}\right)^{-1} (\mathbf{PR})$, $\mathbf{T} = (\mathbf{PR})^T \left(\mathbf{C}_{\mathbf{e}}^{\mathbf{R}}\right)^{-1} \left(\mathbf{S}_f - \boldsymbol{\mu}_{\mathbf{e}}^{\mathbf{R}}\right) + \left(\mathbf{C}_{\mathbf{S}}\right)^{-1} \boldsymbol{\mu}_{\mathbf{S}}$.

Thus, the specific 3D shape vector $\mathbf{S}$ for the given face is recovered. And finally, the accurate pose angle can be calculated according to the equation (3).

4 Experiments and Results

Pose estimation is an opened problem so far. It is difficult to estimate the accurate angles for only given one facial image. Through many experiments, we think that the orthogonal projection computation (OPC) is a reasonable solution to this problem. So in our experiments, we compare our results with those of the orthogonal projection computation using average 3D shape vector.

4.1 Experiments with Single Image

First, we carry our experiment on some images in FERET database [13] and the results examples are given by Fig.2. To present the visualized evaluation, a 3D face

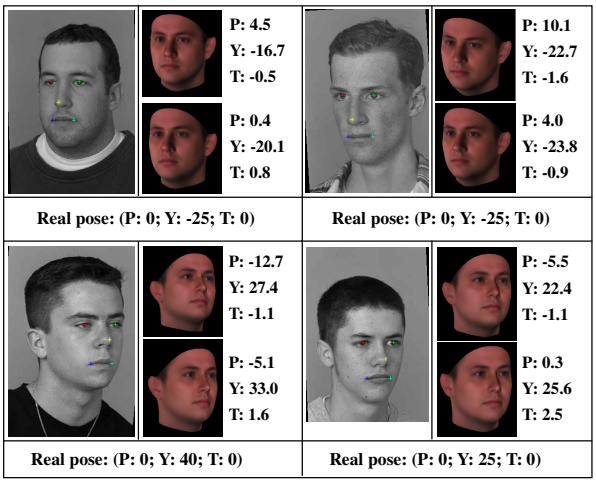

Fig. 2. The pose estimation results for the images in FERET database

model is rendered according to the pose estimated by our Gaussian error models (GEMs) algorithm and the orthogonal projection computation (OPC) algorithm respectively. The estimated pose angles are listed right to the rendered faces. For each test image, the upper rendered pose face is for the OPC results and the lower one is for the result of our algorithm. The real image poses are also given below the input images to be the references. From these results, we can see that our Gaussian error models based pose estimation improves the two major problems in orthogonal projection computation and achieves good performance.

4.2 Experiments with Image Series

We also take our experiment on an image series that captured the variation of the face turning from left to right. The image series is recorded by a real-time image captured system frame by frame. At the same time, the real pose angles can be provided by the special sensor equipment.The example images of pose variations are shown in Fig.3.

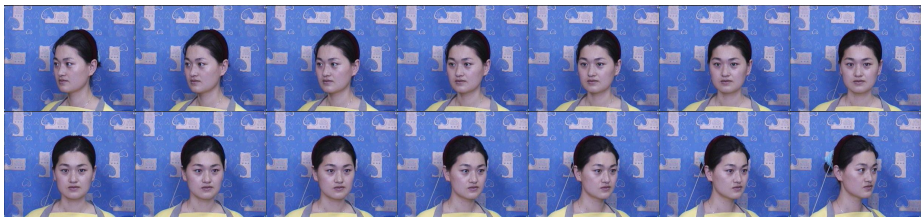

Fig. 3. The examples of the pose image series

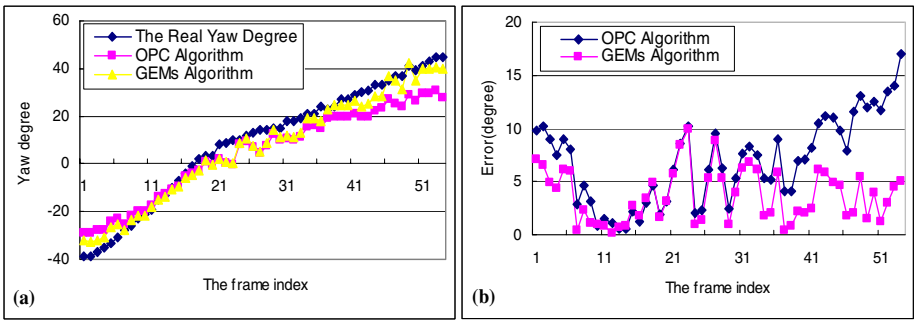

Fig. 4. (a) is the pose estimation results and (b) is the estimation deviation

In our test series, there are 54 images. The pose changes from left 39 degree to right 45 degree. The pitch is maintained nearly horizontal so only the yaw angle is statistic here. In this experiment, we also compare the results between the orthogonal projection computation (OPC) and the Gaussian error models algorithm (GEMs). The pose estimation results are given by Fig.4 (a) and the deviations to real yaw angles are presented by Fig.4 (b). The quantitative deviations of this image series for OPC

algorithm and GEMs algorithm are 6.9 degree and 3.6 degree respectively. From these experimental results, we can see that the estimation pose angles by our Gaussian error models method are close to the real degrees and the deviations are small enough for many related applications.

5 Conclusion

In this paper a novel Gaussian error models based algorithm is proposed to perform pose estimation. Five key points are used to model the face. Assuming the 2D landmarks of given facial image have been located, orthogonal projection computation can be used to compute a coarse pose by using a general average 3D model. For considering the difference of specific face and the error term between the orthogonal projection and weak perspective projection, we use Gaussian probability density function to model the distribution of the two variables respectively. Based on the prior knowledge, the specific 3D shape vector corresponding to the given face can be inferred by MAP theory. Finally, the more accurate pose angles can be calculated easily using the transformation of the orthogonal projection formula. The experimental results show that our pose estimation algorithm is robust and reliable for estimating the pose of real facial images.

We should note that the locations of five landmarks in 2D images are necessary for pose estimation, hence the many efforts in the future, for example, the more accurate feature alignment, will make our algorithm more practicable in daily applications.

References

1. S.Y.Lee, Y.K. Ham, R.H.Park, Recognition of Hman Front Faces using Knowledge-based Feature Extraction and Neuro-Fuzzy Algorithm. Pattern Recognition 29(11), (1996) 1863-1876.
2. Shinn-Ying Ho, H.L.Huang, “An Analytic Solution for the Pose Determination of Human Faces from a Monocular Image”, Pattern Recognition Letters, 19, (1998) 1045-1054.
3. T. Hogg, D. Rees, H. Talhami. Three-dimensional Pose from Two-dimensional images: a Novel Approach using Synergetic Networks. IEEE International Conference on Neural Networks. 2(11), (1995) 1140-1144.
4. T. Darrell, B. Moghaddam, A. P. Pentland. Active Face Tracking and Pose Estimation in an Interactive Room. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (1996) 67-72.
5. H. Murase, S. Nayar. Visual Learning and Recognition of 3-d Objects from Appearance. International Journal of Computer Vision, 14, (1995) 5-24
6. Q. Chen, H.Wu, T. Shioyama, T. Shimada. A Robust Algorithm for 3D Head Pose Estimation. IEEE International Conference on Multimedia Computing and Systems. (1999) 697-702.
7. A. Nikolaidis, I. Pitas. Facial Feature Extraction and Determination of Pose. Pattern Recognition, 33, (2000) 1783-1791.
8. A. Gee, R. Cipolla, “Determining the Gaze of Faces in Images”, Image and Vision Computing 12, (1994) 639-647.

9. A. Gee, R. Cipolla. Fast Visual Tracking by Temporal Consensus. Image and Vision Computing. 14, (1996) 105-114.
10. C. W. Lee, A. Tsukamato. A Visual Interaction System using Real-time Face Tracking. The 28th Asilomar Conference on Signals, Systems and Computers. (1994) 1282-1286.
11. 11.Q. Ji, R. Hu. 3D Face Pose Estimation and Tracking from a Monocular Camera. Image and Vision Computing. (2002) 1-13.
12. 12.V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces", In Proceedings, SIGGRAPH'99, (1999) 187-194.
13. 13.P. Phillipse, H. Moon, S. Rizvi and P. Rauss, "The FERET Evaluation for Face-Recognition Algorithms", IEEE Trans. on PAMI, 22, (2000) 1090-1103.