

Linear Regression for Head Pose Analysis

Hui Yu, Honghai Liu, *Senior Member, IEEE*

School of Creative Technologies
University of Portsmouth
Portsmouth, PO1 2DJ, UK
{hui.yu; honghai.liu}@port.ac.uk

Abstract—Extensive research has been conducted to estimate and analyze head poses for various applications. Most existing methods tend to detect facial features and locate landmarks on a face for pose estimation. However, the sensitivity to occlusion of some face parts with key features and uncontrolled illumination of face images make the facial feature detection vulnerable. In this paper, we propose a framework for pose estimation without the need of face features or landmarks detection. Specifically, we formulate the pose estimation as a linear regression applied to the pose space. This method is based on the assumption that pose space cannot be linearly approximated in the pose subspace. The experimental results strongly support this assumption. In cases where the database does not obtain various poses in the intraclass, we propose to generate those poses through a 3D reconstruction and projection method. The experiment conducted on the CMU MultiPIE and IMM Face database has shown the effectiveness of the proposed method.

Keywords—*linear regression; pose analysis; human face; 3D face*

I. INTRODUCTION

Face pose estimation has wide applications in video surveillance, human-computer interaction, computer vision and mobile policing etc. Extensive research has been done to estimate head poses through detecting facial features and locating landmarks on faces [9, 10]. However, due to uncertainty of the quality of face images under uncontrolled environments, it is still challenging to detect facial features and landmarks. This ambiguity in facial feature detection can lead to poor performance in pose estimation.

Various pose estimation methods have proposed in recent years including geometric approaches [12, 13], regression-based method [15, 16] and template-based methods [14] etc. Those methods have achieved good results in many cases. However, most existing methods suffers from the strong requirement for detecting accurate facial features, especially when faces are away from frontal view with very limited information for detecting reliable facial features. Unreliable facial features can lead to mismatching and pronounced estimation errors.

In this paper, we propose to analyze face pose without the need for facial feature detection. The proposed method is based

on the assumption that pose space cannot be linearly approximated in the pose subspace.

II. POSE RECONSTRUCTION AND ESTIMATION

Naseem et al. [19] proposed to use Linear Regression Classification (LRC) for face recognition. It assumes that face samples in the same identity class fall in the linear subspace. During the linear regress, the dictionary data formed by training samples is highly related with the reconstruction result. Thus, the dictionary selection has become critical for the final classification performance. The assumption of linear subspace is that the test face and training faces of a subject lie on the same linear subspace [2, 8]. However, when the intraclass contains some face images with severe pose variations, the performance of the method declines.

Chai et al. [4] proposed the Locally Linear Regression (LLR) method to tackle the pose problem with the assumption of that there is an approximate linear mapping between the frontal face and its nonfrontal pose. It deal with the case of severe pose variations through sampling the nonfrontal face image to obtain many overlapped local patches. Linear regression is then applied on those individual local segments to represent the corresponding frontal counterpart. However it may not fully reconstruct the frontal face from occluded or nonfrontal face with missing information.

The LRC method is based on the concept that patterns from the same class lie on a linear subspace [3]. LRC works well when there are not severe pose variations in each class samples. In many cases, there are only a few training samples for an individual class, which is insufficient for a robust classification performance using the LRC method.

Deng et al. developed Extended SRC through intraclass variant dictionary [17]. The assumption of the Extended SRC method is that the intraclass variations of one subject can be approximated by a sparse linear combination of those of other subjects. The intraclass variation in that method refers to the difference in illumination and expression or occlusion of the same subject.

When various but not evenly sampled face poses appear in the same class, since the pose space cannot be approximated

This work was supported by EU seventh framework programme under grant agreement No. 611391, Development of Robot-Enhanced Therapy for Children with Autism Spectrum Disorders (DREAM), and the projects under grant No. DMETKF2013001 by State Key Lab of Digital Manufacturing Equipment & Technology, China.

by a linear subspace. If there are enough poses in the same class, the linear reconstruction can still have a good approximation. This suggests some relationship between the test image and the pose spaces. To better understand those relationships we apply the linear regression method to the pose space. We want to test whether there is a linear relationship among the poses and how strong the relationship is if there is. Those relationships can be used for pose estimation and analysis. In this paper, we apply linear regression across subject classes to filter the pose of the given face image.

Given N classes of faces with c_i poses in the i^{th} class, $i=1,2,3,\dots,N$. Each training sample with size $m \times n$ is represented as $I_j \in \mathfrak{R}^{m \times n}$, $j=1,2,3,\dots,N$. Each face image is transformed into a vector through column concatenation represented by $x_j \in \mathfrak{R}^{q \times 1}$ with $q = m \times n$.

The pose specific data matrix X_p is generated by stacking all those sample image vector with the same pose p across N classes,

$$X_p = [x_1^p, x_2^p, \dots, x_N^p] \in \mathfrak{R}^{q \times c_i}, p=1,2,\dots,c_i \quad (1)$$

The test face image can be transformed into the same vector representation $y \in \mathfrak{R}^{q \times 1}$. Thus, the relationship of data matrix X_p and y can be expressed as the following:

$$y = X_p \beta_p + z_p, \quad p=1,2,\dots,c_i \quad (2)$$

where $\beta_p \in \mathfrak{R}^{N \times 1}$ is the vector of regression parameters and $z_p \in \mathfrak{R}^q$ is a residual. If y is in favor of the p^{th} pose class, the residual z_p becomes a noise term with a bounded energy $\|z_p\|_2 \leq \mathcal{E}$.

The parameter vector can be solved using least-square estimation if $q > c_i$,

$$\hat{\beta}_p = (X_p^T X_p)^{-1} X_p^T y \quad (3)$$

This estimated parameter vector $\hat{\beta}_p$ can then be used to reconstruct the predicted pose of the test image,

$$\begin{aligned} \hat{y}_p &= X_p \hat{\beta}_p \\ &= X_p \hat{\beta}_p = X_p (X_p^T X_p)^{-1} X_p^T y \end{aligned} \quad (4)$$

The reconstructed \hat{y}_p is the projection of y onto the pose space spanned by the p^{th} poses of N classes. Simply, the

distance between y and the reconstructed \hat{y}_p can be calculated as

$$d_p = \|y - \hat{y}_p\|_2, \quad p=1,2,\dots,c \quad (5)$$

As face images have a poor linearity in pose subspace, the close the pose of the test face and the p^{th} poses the smaller the distance d_p is. Applying the above linear regression to each pose vector of the N given classes, we can obtain a set of distance values. The pose with minimum distance value is supposed to be the closest pose with the test image.

III. POSE SUBSPACE GENERATION

The pose space can be thought as a collection of evenly sampled local linear subspace. In some cases, we might not have all those poses available for each subject in the database. Even sometimes only one face pose is obtained. There is a practical problem with how to generate different poses from the given pose. In this paper, we assume a frontal face is obtained and we propose to create the local linear subspace through 3D reconstruction and pose projection on image space.

Given a frontal face image, we first obtain 69 facial features represented by s^{2d} using Active Appearance Model (AAM) [18]. Since the AAM method is applied to the frontal face image, the accuracy of facial landmark detection is better than cases with nonfrontal poses.

We then fit a 3D face model to these features through 2D projection and optimization. The fitting is achieved through minimizing the error between the obtained features on the given image and the projection of 3D landmarks of a set of training 3D face models.

$$E = \frac{1}{2} \|M \times s^{3d} - s^{2d}\|^2 \quad (6)$$

where M is a projection matrix, which project the 3D face shape s^{3d} formed by landmarks onto 2D space. And s^{3d} is constructed 3D landmark shape through linear regression based on the Morphable Model [1] concept.

$$s^{3d} = \bar{s} + \sum_{i=1}^k \alpha_i v_i \quad (7)$$

Where \bar{s} is the mean landmarks shape of a set of 3D training face models. And α_i is the weighting parameter for each training face model represented by its eigenvector v_i constructed by Principal Component Analysis (PCA). Assume the training 3D data subjects to a Gaussian distribution. The probability for coefficient α_i can be calculated as

$$p(\alpha_i) \sim \exp\left[-\frac{1}{2} \sum_{i=1}^{k-1} (\alpha_i / \lambda_i)^2\right] \quad (8)$$



Fig. 1. Example of pose reconstruction from single frontal view image. The left column image is the original frontal view face; the second column shows the landmarks using AAM tracking; images from column 3 to the right are reconstructed poses representing pose 75, pose 45 and pose -45 respectively.

Where λ_i is the eigenvalue of the covariance matrix of the training shapes.

Through minimizing the above equation, the 3D geometry of the given face image can be reconstructed along with the projection Matrix M . Through applying the matrix M to the full 3D geometry, we can obtain a full 2D face shape in the image space. Then the correspondence between the image pixels and 3D geometry can be established.

To generate the pose space for this given subject, we can rotate the reconstructed 3D face model and render it from the right view point. Fig. 1 shows an example of pose reconstruction from the given frontal view image using IMM Face Database [11].

IV. EXPERIMENT



Fig. 2. Examples of 13 viewpoints plus 2 additional viewpoints simulating typical surveillance camera views.

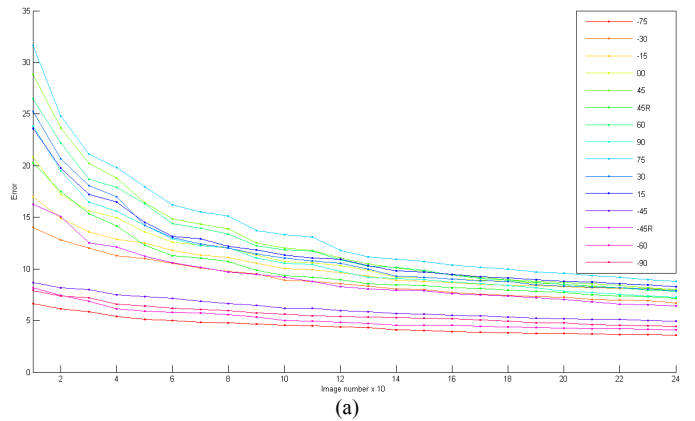
We test the proposed method on the CMU MultiPIE face database [2] and IMM Face database. CMU MultiPIE database consists of 337 subjects with each subject's face taken from 13 viewpoints plus 2 additional viewpoints simulating typical surveillance camera views. Fig. 2 shows examples of those 15 poses.

The experiment was conducted using various numbers of images for each pose. We started with 10 images and increased by 10 images in each step. Fig. 3 demonstrates reconstructed poses for pose -15 using 15 pose classes respectively. It clearly shows that the best approximated result has been achieved by using the same pose class indicated by the red frame.

Generally, the errors reduced gradually with the increase of images involved in the regression for all test poses. Fig. 4 shows two examples of using linear regression to approach 15 poses. We find that there is a strong relationship between the test pose and the pose class.



Fig. 3. Comparison of reconstructed 15 poses based on the pose space -30 for the same identity using 249 images for the regression. Three blocks (left, middle and right) demonstrate the test image and the reconstructed pose (left and right column in each block). The top row of the middle block with a red frame is the example of reconstruction using pose -30.



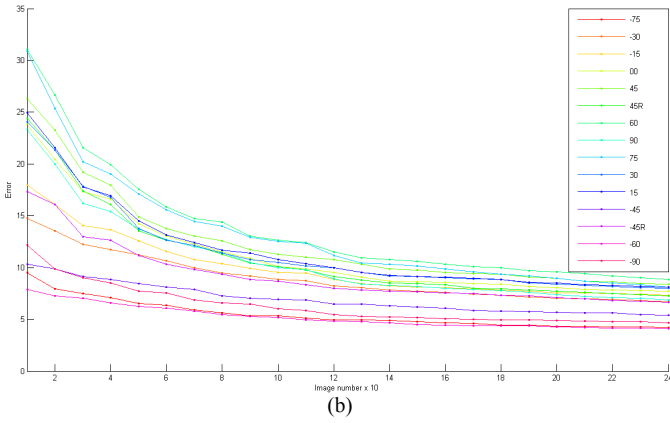


Fig.4. Error curves of two identities. Error The vertical axis is the error between the test image and the reconstructed image using 15 pose sequences in the pose space. The horizontal axis represents the number of images of the same pose category involved in the regression. (a) and (b) show error curves based on pose -75 of identity one and two respectively.

With more face images involved in the regression in each pose space, the error between the test and the reconstructed poses becomes smaller to some degree. The decrease speeds of the errors are different for different subjects and test poses. In general, the errors become more stable after 180 images in the pose class.

Fig. 5 shows a dissimilarity matrix of the pose analysis. The color-coded matrix illustrates how similar the reconstructed pose with the test image using a certain class of pose images during the regression. The red color represents a higher error value while blue color represents the smaller error. The color matrix shows that the reconstructed pose has the highest similarity with the test pose when approximated by the class of pose images with the same pose. In most occasions, the color matrix shows a basic trend that the closer of the pose between the test pose and the regression pose class, the higher similarity of the pair is. This is indicated by those square colors in the color matrix. This pattern does not show a strong linear relationship among the similarity. For example, the colors show that the similarity between pose -45R and 45R is higher than that between -45R and -45 or -60.

Fig. 6 demonstrates reconstructed errors for 15 poses using various images for regression in the pose space -30 in (a) and the neutral pose in (b). The figure shows that errors are clearly different for different poses. Even though the errors decrease with the increasing of the number of regression images, the error difference among different poses are consistent. The figure clearly shows that the in general the closer the pose space is with the test image, the smaller the error is. As can be seen from the figure, the bottom curve shows that when the test image, -30 for (a) and neutral for (b), has the same pose with the regression pose space, the error drops significantly. With more image involved in regression, the error become smaller for all poses. But the error does not change much from 200 images onwards. Fig. 7 illustrates some examples approximating to 15 poses shown in Fig. 2 with various numbers of images in each pose class involved in the regression.

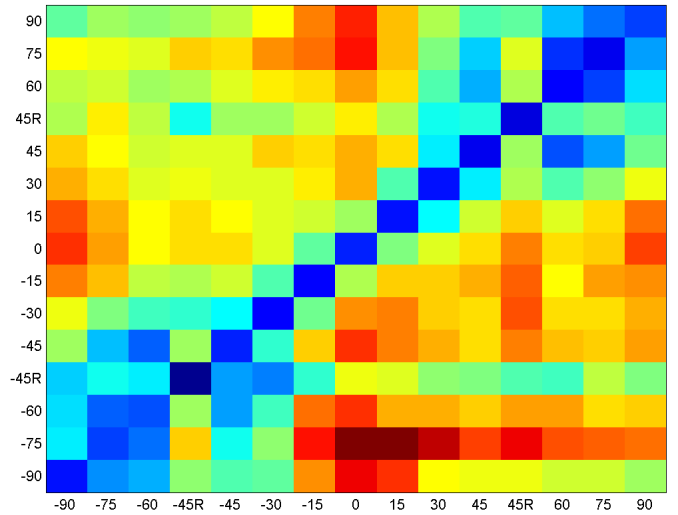


Fig. 5. Dissimilarity matrix of the pose analysis. Each color square represents the error value between the test pose and the reconstructed pose. Each horizontal color row represents a test pose. Vertical color-coded bars show errors of different test poses under the same regression pose class. Those blue squares along the anti-diagonal indicate the smallest similarity between the reconstructed poses and the probe pose.

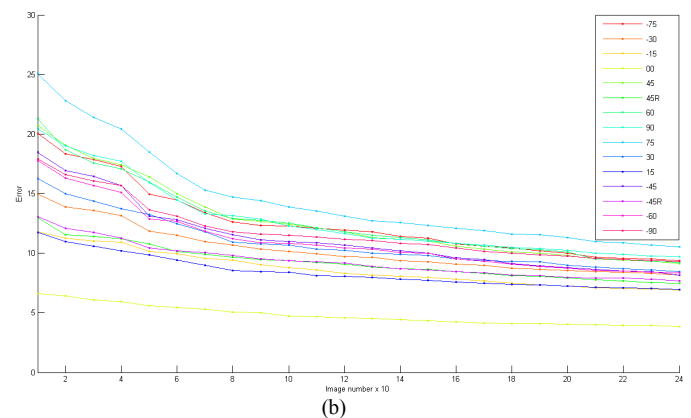
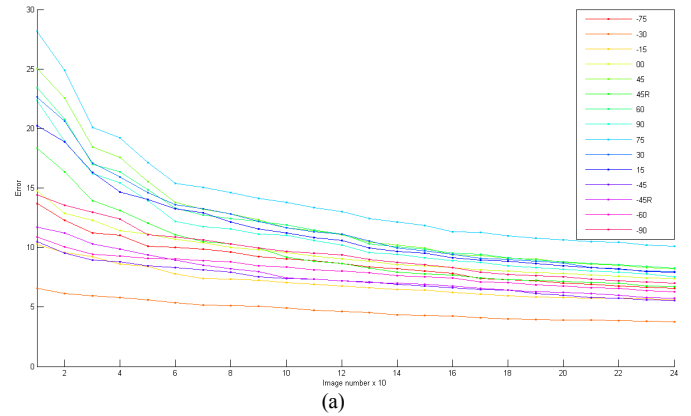


Fig. 6. Comparison of approximation errors for pose 15 and the neutral pose using 15 pose class respectively. The vertical axis is the error between the test image and the reconstructed image using 15 pose sequences in the pose space. The horizontal axis represents the number of images of the same pose category involved in the regression. (a) and (b) show error curves based on the pose -30 and the neutral pose respectively for identity one.



Fig. 7 Comparison of reconstructed 15 poses based on pose space -75 using various numbers of images in each pose class for the regression. Both blocks (left and right) show the test image (column 1) and 4 reconstructed poses (from column 2 to column 5) using the number of training images 20,100,150 and 249 respectively. The test poses for the left block are (from top to bottom) 15,30,45,60,75, 90, 45R and -45R. The test poses for the right block are (from top to bottom) -90, -75, -60,-45, -30, -15 and the neutral pose. Row 2 in the right block shows reconstructed pose -75.

V. CONCLUSION

In this paper, we have proposed to analyze head poses using a linear regression method. We have found that a face pose can be approximated by the same pose class. The closer the pose class is with the test pose, the better the approximated result is. Thus it indicates that face pose can be estimated through linear regression method. The advantage of using linear regression for pose estimation is that it avoids exactly extracting face features. This particularly can be very useful for head pose estimation under uncontrolled environments. To generate the pose space, we have developed a method for reconstructing the 3D face shape from a given frontal view face images. We have demonstrated the promising result of the method.

In future work, we would improve the regression method with less training images and experiment on the relationship among different poses. The 3D reconstruction method could also be improved to get more accurate virtual poses.

REFERENCES

- [1] V. Blanz, C. Basso, T. Poggio, and T. Vetter, "Reanimating Faces in Images and Video", Proc. Eurographics Conf., 2003.
- [2] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, & S. Baker (2008). Multi-PIE. Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition.
- [3] R. Barsi and D. Jacobs, "Lambertian Reflection and Linear Subspaces,"IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pp. 218-233, Feb. 2003.
- [4] X. Chai; S. Shan; X. Chen, W. Gao "Locally Linear Regression for Pose-Invariant Face Recognition," Image Processing, IEEE Transactions on , vol.16, no.7, pp.1716,1725, July 2007
- [5] Z. Zhou, J. HuangFu, H. Zhang, and Z. Chen, "Neural network ensemble based view invariant face recognition,"J. Comput. Study Develop., vol. 38, no. 9, pp. 1061-1065, 2001.
- [6] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 26,no. 4, pp. 449-465, Apr. 2004.
- [7] H. T. Ho; R. Chellappa, Pose-Invariant Face Recognition Using Markov Random Fields, Image Processing, IEEE Transactions on , vol.22, no.4, pp.1573,1584, April 2013

- [8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210-227, 2009.
- [9] E. Murphy-Chutorian and M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.31, no.4, pp.607,626, April 2009.
- [10] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, Volume 35 Issue 4, December 2003 Pages 399-458.
- [11] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME A Flexible Appearance Modelling Environment. *IEEE Trans. on Medical Imaging*, 22(10):1319-1331, 2003
- [12] A. Gee, R. Cipolla, Determining the gaze of faces in images, *Image and Vision Computing* 12 (10) (1994) 639-647.
- [13] T. Horprasert, Y. Yacoob, L. Davis, Computing 3D head orientation from a monocular image sequence, *IEEE International Conference on Automatic Face and Gesture Recognition*, Vermont, USA, 1996, pp. 242-247.
- [14] S. Niyogi and W. Freeman. Example-based head tracking. In *Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, pages 374-378, 1996.
- [15] Y. Ma, Y. Konishi, K. Kinoshita, S. Lao, and M. Kawade. Sparse bayesian regression for head pose estimation. In *Proc. ICPR*, pages 507-510, 2006.
- [16] Y. Li, S. Gong, J. Sherrah, and H. Liddell. Support vector machine based multi-view face detection and recognition. *Image and Vision Computing*, 22(5):413-427, 2004.
- [17] W.H. Deng, J.N. Hu and J. Guo, Extended SRC: Undersampled face recognition via intraclass variant dictionary." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864-1870, 2012
- [18] T. F. Cootes, G. J. Edwards, and C. Taylor, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 23, No. 6, June 2001
- [19] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* , vol.32, no.11, pp.2106,2112, Nov. 2010