3D FACE POSE DISCRIMINATION USING WAVELETS

Mukesh.C.Motwani and Qiang Ji †

Dept. of Computer Science
University of Nevada
† Department of Electrical, Computer, and System Engineering
Rensselaer Polytechnic Institute

ABSTRACT

This paper describes a robust method for discriminating 3D face pose (face orientation) from a video sequence featuring views of a human head under variable lighting and facial expression conditions. Wavelet Transform is used to decompose the image into multiresolution face images containing both spatial and spatial-frequency information. Principal Component analysis (PCA) is used to project a low-resolution sub-band face pose onto a pose eigen-space where the first 3 eigen coefficients are found most sensitive to pose and follow a trajectory as the pose changes. Any unknown pose of an query image can then be estimated by finding the Euclidean distance of the first 3 eigen coefficients of the query image from the estimated trajectory. Wavelet Transform reduces the computational load on the PCA and makes the algorithm robust against illumination changes and facial expression. An accuracy of 84% was obtained for test images unseen during training under different environment conditions, facial expressions, and even different human subjects.

1. INTRODUCTION

The problem of human body parts recognition and detecting their pose in the 3D space has been around for quite a while. Determining human head pose is just one of many aspects of the mentioned problem. Estimating head orientation is central in vision-based animation, gaze estimation and is a component of inferring the intentions of agents from their actions.A real-time pose estimator can be used to drive graphical models for applications such as virtual teleconferencing. It can also be used to index a more detailed view specific representation for identity recognition and expression analysis. In addition, pose prediction is useful for overcoming display lags in real-time interactive and visual communication applications. A human head rotating in depth (out of the image plane) induces nonlinear transformations in the projected image of the face. Facial features become occluded and the outline of the face alters its shape causing interference with the background. Pose estimation is therefore a difficult task. We propose a robust algorithm which is computationally cheap to make real-time pose estimation possible without using specialized hardware. The novelty of the method is to use wavelet transform to make the intensity image a parameter of pose only and make it insensitive to other parameters such as facial expression,illumination which also change the intensity image besides pose. A major problem in automatic real-time pose tracking is that one does not know where in the current image the face is located. Our method relies on an existence of a preprocessor routine that segments out the face for us.

2. RELATED WORK

Automatic face pose estimation by computer can be divided into two approaches, namely, feature based and appearance based. In feature based approach, estimation based on the relationship between human facial features [1] [2] relies highly on the accuracy of the facial feature detection schemes. Azarbayejani et al.[3] use feature point tracking projected on an ellipsoidal model to track the head position. Gong et el [4] also characterized facial feature points using Gabor Wavelets and used special hardware to make the algorithm work in real time. Some model based methods [5][6] [7] estimate the 3D pose from 2D pose by detecting features and matching model features with them.

Appearance based approach attempts to capture and define the face as a whole. The face is treated as a two-dimensional pattern of intensity variation. Rae et al.[8] describe a neural network based system to estimate pan and tilt of a person. SVM's have also been used for pose discrimination [9]. Niyogi and Freeman [10] match image templates of entire heads at different viewing angles. 3D model based head pose estimation based on color [11] have also been used. In [12], Pose estimation is done using Volumetric Frequency Repre-

0-7803-6725-1/01/$10.00 ©2001 IEEE

sentation (VFR) model constructed from range data of a person and using gray level images to index into the model. Volker et al.[13] used Gabor Wavelet Networks combining the advantages of RBF networks with the advantages of Gabor Wavelets to improve the performance. J. Sherrah and et al [14] used orientation-selective Gabor filters to enhance differences in pose. However, we use wavelets for pre-processing of the images to obtain invariance under different facial expressions, lighting, distance from camera changes. We try to solve the problem of building a robust real-time system without using specialized hardware.

Under the face-based approach Principal Component Analysis (PCA) represents a face as a linear combination of weighted eigenvectors, known as eigenfaces. Faces rotating across views form continuous manifolds in a Pose EigenSpace (PES) [15]. Since the eigenspace is optimal for computing correlation between images, we can project the current image to eigenspace to obtain a 3 dimensional point and look for the closest point on the appearance manifold. The matching problem then is to find the minimum distance between the eigenspace point and the discrete manifold points and the closest manifold point determines the pose. Darrell et al. [16] describe a view-based PCA approach to face pose estimation. Under this approach a separate set of eigenfaces is computed for each possible pose. The pose is identified by computing the eigenspace projection of the input image onto each eigenspace and selecting the one with the lowest residual error. The resulting ensemble is a highly complex and non-separable manifold. The appearance of the face is the combined effect of its shape(shape changes because of expressions,small occlusion such as glasses), sensor parameters, pose in the scene and the illumination conditions. We use "parametric eigenspace" [17] representation where the face appearance is parameterized by the variables, namely face pose, facial expressions and illumination. By using wavelets appearance manifolds are reduced to curves parameterized by just face pose.

3. PROPOSED METHOD

The system consists mainly of two stages,namely, training and estimation stage.

Training Stage:

3.1. Acquisition of training data

In the training stage, a MPEG video clip of a person in various poses is decompressed into frames.The training data typically consists of 50 to 100 images. If

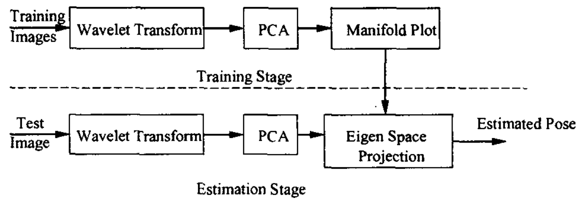

Fig. 1. Block Diagram

smaller number of learning poses are used, discrimination tends to be unreliable when the test images correspond to poses that lie in between the learning poses. If the pose increments used in the learning stage are small, we obtain a larger number of learning samples and hence a larger number of points on the parametric manifold. If the person slows down the motion of the head in a particular pose more closely spaced points are obtained in that region on the manifold. Since our database involves face sequences,there are no standard face sequences yet available for comparison.

3.2. Discrete Wavelet Transform

Discrete Wavelet Transform of depth 3 is performed on all the images. Low frequency LL subband of resolution 30x44 is selected for further processing. The 3-level Wavelet Transform using CDF (2,2) bi-orthogonal wavelet is implemented using Lifting Scheme [18]. Gabor Wavelets seem to be the most probable candidate for feature extraction. But they suffer from certain limitations viz., they cannot be implemented using Lifting Scheme and secondly the Gabor Wavelets form a non-orthogonal set thus making the computation of wavelet coefficients difficult and expensive. Special hardware is required to make the algorithm work in real time. Thus choosing a wavelet for face pose estimation depends on a lot of trial and error. Experiments were done with Haar and Daub4 wavelets and no improvement in performance was observed.Haar wavelet wasnt used because it is not a smooth wavelet. Images with aspect ratio other than 1:1 are considered carefully taking into account the boundary conditions.

In choosing the WT subband, we have the following criteria:

1. Reduce the computational complexity:

The computational complexity of PCAbased method is in the cubic order of image resolution or number of training images, depending on which value is smaller. To minimize the computational complexity, we prefer to work with low resolution subbands which have sufficient information for pose estimation. In turn, working on a lower resolution image will reduce the computational complexity dramatically.We compute the energy in the subbands to measure the information content in the subbands. Previously low-resolution information has been used for face pose estimation [10],recognition [19] and gender classification [20].

2. Insensitive to minor changes in facial appearance, lighting and occlusion: When there is a change in human face, all frequency components will be affected. Whereas changes in illumination affect the intensity manifold globally, in which only the low frequency spectrum is affected.

3.3. Principal Component Analysis

In the second stage of training, PCA [21] is applied on the low-frequency LL subband of resolution 30 x 44, instead of the original image resolution of 240x352. The output of this step will be a set of eigenvectors and eigenvalues. These eigenvectors constitute the dimensions of the eigenspace.

3.4. Manifold Plot

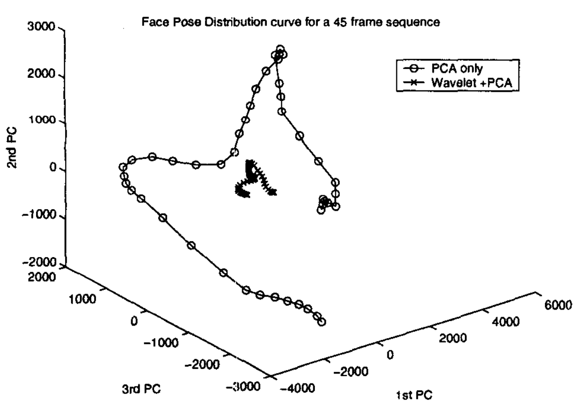

Fig. 2. A 3D plot of manifold formed by using wavelet+PCA and PCA only

By arranging the eigenvalues in a descending order, we select 3 eigenvectors with the largest eigenvalues. The first 3 eigenvectors encode information about the pose.The rest of the eigen vectors encode information about facial expression and other finer details. Thus all images in the training set are represented by a linear combination of 3 representational bases by projecting then into the Pose Eigen Space. All training images are projected on the eigenspace to obtain a set of points in a 3D eigenspace. These points lie on a manifold that is parameterized by pose.

3.5. Estimation Stage

In the estimation stage, to estimate the pose of query image we find the wavelet transform of the image and project the LL subband of resolution 30x44 onto Pose Eigen Space using 3 eigenfaces. We use the Euclidean distance to find the shortest distance between the first 3 eigen coefficients of the query image from the manifold points in the 3D space. We can divide the face poses of training images into 5 classes viz. left,mid-left,center,mid-right,right and identify the class of the query image as the one, which has the closest match to the image in a given class.During the training stage, the images are manually chosen and assigend to a class. When a test image is discriminated to the closest image in the database, a simple mapping is done to assign the class to the test image.

4. PERFORMANCE

A number of experiments were done to test the robustness of the algorithm and increase the estimation accuracy. In the figures shown below we discriminate face pose with different facial expressions and illuminations.

The algorithm performed fairly well with an accuracy of 84% as compared to 71% using PCA only for images under different environment conditions and not included while training. An accuracy of 100% was observed for poses taken from the same video sequence and not included in the training dataset.We also tested the system on people who were not in the training set. The identity of the closest match head constantly changed, but the pose of the best match generally matched well within the pose of the input image. When the distance between the camera changes the algorithm faired poorly if either LH,HL or HH subband is used. LL subband gave the optimum performance. The results are not perfect but they are reasonable. The proposed system requires 17.26 seconds while the traditional PCA method requires 337.6 seconds for training of 50 images. This shows that the proposed system is more computational efficient than that of traditional PCA method. All experiments are performed on a Ultra 360 MHz Sun Workstation.

5. CONCLUSION

This paper proposed a wavelet subband approach in using PCA for face pose estimation. Wavelet transform is adopted to decompose an image into different subbands

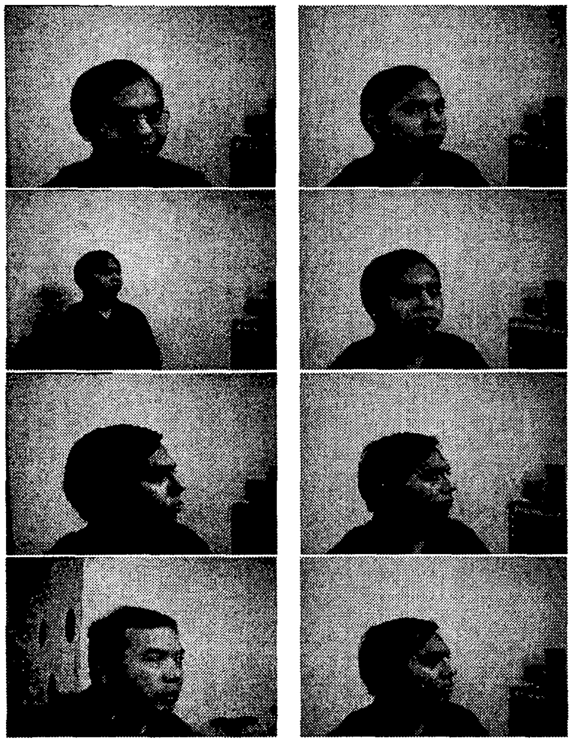

Fig. 3. The images on the left are the test images. The closest match is shown on the right. Query images with varying facial expression, distance from camera, illumination and persons with different identity were used to test the robustness

with different frequency components. A low frequency subband is selected for PCA representation. Results show that the proposed method gives better accuracy and class separability than applying PCA on the whole original image that contains all frequency components. Nevertheless, a combination of frequency bands, can even give better performance. Further studies in face detection simultaneously with pose discrimination will be our future direction.

6. REFERENCES

[1] A.H. Gee and R.Cippola, "Determning the gaze of faces in images," *Image and Vision Computing*, vol. 12, no. 10, pp. 639–647, 1994.

[2] R. Brunelli, "Estimation of pose and illuminant direction for face processing," *MIT Artificial Intelligence Laboratory*, vol. Technical Report 1499, 1994.

[3] A. Azarbayejani T. Starner B. Horowitz and A. Pentland, "Visually controlled graphics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 602–605, 1993.

[4] S. McKenna and S. Gong, "Real time face pose estimation," *International Journal on Real Time Imaging, Special Issue on Real-time Visual Monitoring and Inspection*, vol. 4, pp. 333–347, 1998.

[5] Francoise Preteux Malciu Marius and Vasile Buzuloiu, "3d global head pose estimation : A robust approach," *International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, 1999.

[6] Thanarat Horprasert Yaser, "Computing 3-d head orientation from a monocular image sequence," *Proc. of Int. Conf. on Automatic Face and Gesture Recognition*, pp. 242–247, 1996.

[7] Shinichi Matsumori Kazuyuki Hattori and Yukio Sato, "Estimating pose of human face based on symmetry plane using range and intensity images," *14th International Conference on Pattern Recognition*, vol. 2, 1998.

[8] R. Rae and H. Ritter, "Recognition of human head orientation based on artificial neural networks," *IEEE Transactions on Neural Networks*, vol. 9, no. 2, pp. 257–265, 1998.

[9] J. Huang, X. Shao, and H. Wechsler, "Face pose discrimination using support vector machines," *Proc. of 14th Int'l Conf. on Pattern Recognition (ICPR'98)*, pp. 154–156, 1998.

[10] S.Niyogi and W.T. Freeman, "Example-based head tracking," *Proc.2nd Int. Conf. on automatic Face and Gesture Recognition*, 1996.

[11] P.A. Beardsley, "Pose estimation of the human head by modelling with an ellipsoid," *IEEE Conf. on Automatic Face and Gesture Recognition*, 1998.

[12] J. Ben-Arie and D. Nandy, "A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 449–457, 1998.

[13] V. Kruger and G. Sommer, "Gabor wavelet networks for object representation," *Proc. of the Int. Dagstuhl 2000 Workshop, 2000*, 2000.

[14] S. Gong J. Sherrah and E-J. Ong., "Understanding pose discrimination in similarity space," *Proc. British Machine Vision Conference, Nottingham.*, 1999.

[15] S. Gong, S. McKenna, and J. Collins, "An investigation into face pose distributions," *Proc. of the 2nd Int. Worksh. on Aut. Face- and Gesture-Recog*, pp. 265–270, 1996.

[16] T. Darrell, B. Moghaddam, and A. Pentland, "Active face tracking and pose estimation in an interactive room," *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 1996.

[17] S. Nayar, H. Murase, and S. Nene, "Parametric appearance representation," *In Early Visual Learning. Oxford University Press*, 1996.

[18] Wim Sweldens and Peter Schrder, "Building your own wavelets at home," *Wavelets in Computer Graphics, ACM SIGGRAPH Course Notes*, 1996.

[19] Y. Dai and Y. Nakano, "Recognition of facial images with low resolution using a hopfield memory model," *Pattern Recognition*, vol. 31, pp. 159–167, 1998.

[20] S. Tamura, H. Kawai, and H. Mitsumoto, "Male/female identification from 8x6 very low resolution face images by neural network.," *Pattern Recognition*, vol. 29, no. 2, pp. 331–335, 1996.

[21] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.