IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, pp. 265-270, VERMONT, USA, OCTOBER 1996

An Investigation into Face Pose Distributions

Shaogang Gong, Stephen McKenna and John J. Collins
Machine Vision Laboratory
Department of Computer Science
Queen Mary and Westfield College London
Mile End Road, London E1 4NS, England
sgg@dcs.qmw.ac.uk

Abstract

Visual perception of faces is invariant under many transformations, perhaps the most problematic of which is pose change (face rotating in depth). We use a variation of Gabor wavelet transform (GWT) as a representation framework for investigating face pose measurement. Dimensionality reduction using principal components analysis (PCA) enables pose changes to be visualised as manifolds in low-dimensional subspaces and provides a useful mechanism for investigating these changes. The effectiveness of measuring face pose with GWT representations was examined using PCA. We discuss our experimental results and draw a few preliminary conclusions.

1 Introduction

Techniques for computer vision-based automated face recognition can be largely divided into three categories: 3D model-based [2], 2D geometric feature-based [5, 7, 12], and 2D appearance-based matching [13, 20, 23]. We subscribe to the view that the appearance-based approach is more promising whilst neither 3D models nor 2D geometric features can be extracted and matched robustly under changing viewing conditions, in particular, face pose changes [3, 9, 26].

Face models must exhibit invariance under changes in viewing conditions if robust recognition is to be performed. Although it is possible that invariance under changes in illumination, scale, translations and small rotations in the image-plane can be achieved through a process of *normalisation* of face images, changes in face pose (rotation in depth) cannot be easily "normalised". A representation based on specific features for all face poses may be difficult to find since different image features seem to be relevant at different poses. For example, the shape of a silhouette helps distinguish poses between 3/4 view and profile view (see the last 3 frames in Figure 1) but is not of much relevance in distinguishing poses between frontal view and 3/4 view (see the first 3 frames in Figure 1). The reverse can be said about the relative position of the nose with regard to the eyes and the distance between the two eyes.

Figure 1. A face rotates in depth.

A more plausible [4] and robust [3] approach for representing face images of all poses requires the extraction of pose relevant information in a manner which is somehow holistic and independent of any judgement of specific features. However, this does not necessarily mean exhaustive representation. Appearance-based face recognition need not require every view of every person to be stored. Rather, a *canonical view* can be generalised from a range of views and the pose sphere could be represented by only a few canonical views [1, 4, 14, 21]. It is unclear though how an image representation can be chosen which would give the best measurable pose distribution of faces.

We use a Gabor wavelet transform to examine face representation. This can be regarded as part of the normalisation process and allows us to elegantly obtain invariance under scaling as well as changes in illumination conditions, skin tone and hair colour. It is also used to investigate the role of locally oriented features at a range of spatial frequencies in selecting face pose (see Figure 3). Although similar results could be obtained with Gaussian derivative filters as used by Rao and Ballard [23], the formulation of the GWT is better unified and consequently more convenient to apply.

* This research was funded by EPSRC Grant No. GR/K44657 and EC Grant No. CHRX-CT94-0636.

Principal components analysis (PCA) is widely used for reducing the dimensionality of the representation space in order to enable efficient matching [13]. However, faces represented by principal components are sensitive to illumination conditions, scale, translation or rotation in the image-plane. Whilst other studies have been concerned with these problems [6, 20], Murase and Nayar [18] have used the principal components of many views of a single object to visualise the high-dimensional manifold described by changes due to rotation in depth and illumination conditions. The object's pose could then be determined by its position on this manifold. We use PCA in a similar way to investigate the distribution of face pose in high-dimensional representation spaces. In particular, we investigate whether GWT representations are helpful for distinguishing poses.

2 GWT Face Representation

A Gabor wavelet transform (GWT) enables us to obtain image representations which are locally normalised in intensity and decomposed in spatial frequency and orientation. It thus provides a mechanism for obtaining (1) invariance under intensity transformations due to illumination, skin tone and hair colour, (2) selectivity in scale by providing a pyramid representation, and more importantly for our studies, (3) it permits investigation into the role of locally oriented features with regard to pose changes.

Figure 2. GWT kernels for 4 orientations (only real parts are shown).

We perform a GWT of an image by filtering it with a set of sinusoidally modulated Gaussian functions of different spatial frequencies and orientations, known as Gabor functions [8] (see Figure 2). We use a scheme proposed by Würtz [26] in which convolutions with Gabor kernels are performed efficiently in the Fourier domain. In this approach, a single Gabor function (the mother wavelet) is parameterised by a vector $\mathbf{k}=\binom{k_1}{k_2}$, defining variations in scale and orientation. Then a GWT in $[-\pi < \boldsymbol{\omega} = \binom{u}{v} < \pi]$ is given by [26]:

$$\mathbf{F}_{\mathbf{k}}(\boldsymbol{\omega}) = \exp\left(-\frac{\sigma^2(\boldsymbol{\omega}-\mathbf{k})^2}{2\mathbf{k}^2}\right) - \exp\left(-\frac{\sigma^2(\boldsymbol{\omega}^2+\mathbf{k}^2)}{2\mathbf{k}^2}\right)$$

The second term results in "admissibility" i.e. zero-response to spatially constant intensity. Figure 2 shows GWT kernels in the image domain at 4 orientations varying by $45°$ from $0°$ to $135°$.

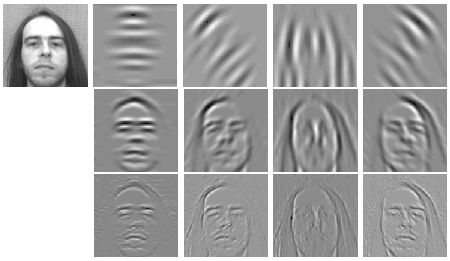

Figure 3. GWT faces are both scale and orientation sensitive. The top row shows the 4 orientational responses at a low center frequency whilst the middle and bottom rows give responses from higher frequencies.

In our studies, the GWT used was parameterised by 3 spatial frequencies and 4 orientations varying by $45°$ from $0°$ to $135°$. A GWT image representation was comprised of a set of 12 responses. At lower frequencies, images are "smoothed" to a larger extent resulting in less sensitivity to small translations in the image-plane and greater correlation between nearby images in a sequence. However, using excessively low frequencies could result in loss of relevant spatial structure (see Figure 3).

The real and imaginary parts of the kernel responses oscillate with their characteristic frequency making them highly sensitive to image-plane translations and therefore ill-suited to matching. This undesirable property can be avoided by taking the magnitude of the responses thereby removing phase information [25]. Figure 4 shows an example of the magnitude responses of the GWT. All the experiments done in this work are based on magnitude responses.

Figure 4. GWT magnitude responses of the face image shown in Figure 3.

3 Face Pose Eigenspace

Given an n-frame sequence $S = [S_0, S_1, \ldots, S_{n-1}]$ of a head rotating in depth, a Pose Eigen-Space (PES) can be calculated by applying PCA to the set of n frames. Projection of each frame onto the first few eigenvectors yields a "low-dimensional pattern vector" representation. In particular, projection onto the first three eigenvectors permits visualisation of the distribution of poses in the representation space (see Figure 5).

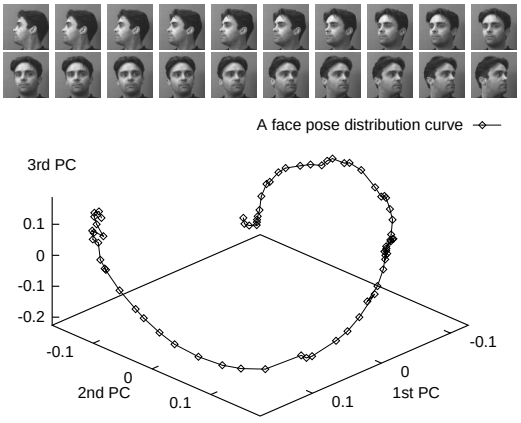

Figure 5. The PES of a face sequence of 60 frames rotating from profile-to-profile. Only 20 frames from the sequence are shown here.

The pose of a novel face image of the person can be estimated by projecting it into this PES. For example, using Euclidean distance in the PES as an approximation of Euclidean distance in the image space, the commonly used methods of minimising the sum-of-squared-difference (SSD) or maximising the correlation between images can be efficiently approximated by minimising Euclidean distance in the PES [19].

4 Face Representations for PCA

It is perhaps inappropriate to perform PCA on representations which are not invariant to changes in viewing conditions. We examine three forms of face representations for PCA. They are (1) normalised intensity faces I, (2) GWT faces $R(I)$, and (3) composite GWT faces $G(I)$ (see Figure 6).

Figure 6. Left: a normalised intensity face I. Centre: a GWT face $R(I)$. Right: a composite GWT face $G(I)$ of equal dimensionality.

An image is normalised by subtracting the mean intensity and dividing by its standard deviation. This corrected variations in overall illumination intensity, camera gain and imaging aperture[1]. A GWT face [2] $R(I)$ is obtained by superimposing the GWT responses. The result is similar to the original intensity image except that intensity distributions are *locally normalised*. A composite GWT face $G(I)$ of equal dimensionality to $R(I)$ is formed by concatenating four "oriented" 1/4 sized GWT faces, each a sub-sampled (by a factor of four) Gabor response to a different orientation (see Figure 6). Now, a principal component derived from this representation can be visualised as a composite "eigen-image" consisting of, four oriented sub-images. The magnitude of each pixel in such an eigen-image is a measure of the variability of the response of one Gabor kernel centred at the corresponding position in the original image. The magnitudes of the first eigen-image indicate *where* in the image-plane *which* orientations encode the most information about pose.

5 Experiments

5.1 Data Preparation

Two types of image sequence were captured using Datacube hardware. Firstly, several sequences of heads rotating from profile-to-profile under different lighting conditions were obtained as output from a head tracking system described elsewhere [15, 16]. These were 60 frames long and were automatically normalised with respect to translation and scale by the tracker. An example can be seen in Figure 5. Secondly, a set of labelled sequences of 12 people were obtained under controlled conditions in which subjects were asked to look at markers on the wall positioned at angles from $0°$ (frontal view) to $90°$ (right profile view) in $10°$ increments. Profile-to-profile sequences were generated by mirroring the sequences. Each labelled sequence, therefore, consisted of 19 frames of known pose. Figure 1 shows 6 frames from such a sequence. The sequences were cropped manually and illumination varied between sequences. All images were sub-sampled with spatial smoothing to 64×64 pixels.

In order to measure the effects of pose, other degrees of freedom such as image-plane translations and scale changes should be removed. An important point to note is that rotation of a head results in a horizontal translation of the face in the image-plane. This raises the problem of how to align images of different poses. Alignment of facial features results in a sequence in which the "centroid" of the head translates horizontally as the head rotates in depth. Alignment based on establishing correspondences becomes problematic due to occlusions. In the experiments described here, images

[1] This is an approximation since factors such as skin tone and hair colour also influence the first and second moments of intensity

[2] A "GWT face"-like representation could also be obtained by using symmetric filters.

are aligned approximately around the visual centroid of the head, either automatically by the tracker or manually for the labelled sequences.

5.2 Pose Manifold of Face Sequences

Initially, n-frame sequences were represented using images normalised for overall intensity. A PES was then calculated by applying PCA to the set of n frames. Three unlabelled sequences of the same person under different lighting conditions were projected onto the pose eigenspace derived from only one of these sequences. Plotted on a 3D graph in Figure 7 are the resulting 3D pattern vectors. The three curves form a fairly smooth manifold parameterised by pose and illumination. In particular, the 3rd PC seems to capture changes caused by lighting conditions. This is similar to the manifolds obtained by Murase and Nayar [18] for various non-face 3D objects under robotically manipulated pose and illumination conditions. In contrast, the face sequences used here were produced by an automatic visual tracking system with left, right and ambient lighting. As a result, the manifold shown here is less smooth, reflecting more realistic conditions.

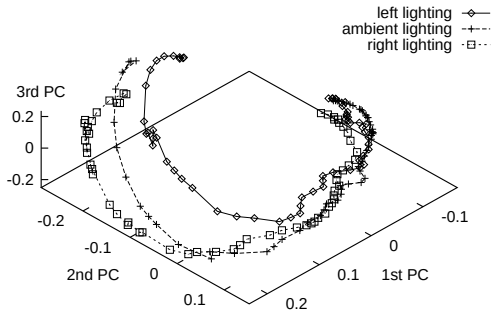

Figure 7. Manifold formed by three face sequences under different lighting conditions rotating from profile-to-profile ($-90°$ to $+90°$).

5.3 PES of Mean Intensity Faces

A straightforward way to derive a generic PES is to use a mean sequence $\bar{S} = (\bar{I}_0, \bar{I}_1, \ldots, \bar{I}_{n-1})$ formed by taking the mean of normalised intensity images at each pose angle over many different face sequences. The plot in Figure 8 shows the pose distribution of a mean sequence formed using 11 face sequences of different people. Also plotted are the projections into this mean PES of a novel face sequence and a non-face sequence of a fan rotating similarly from profile-to-profile. Now, the pose of the novel face sequence can be estimated simply by finding the nearest point along the mean curve. This is an efficient approximation to minimising SSD or maximising correlation between a novel face and a mean face of known pose. The distance of the non-face object to the faces in this PES is distinctively large for most pose angles. Furthermore, it is interesting to note that while the 1st PC separates the left and right poses, the 2nd and 3rd PCs jointly discriminate between poses from profile to frontal views reasonably well. This can also be observed from the eigen-images shown above the plot. It is worth pointing out that although we did not plot higher order PCs, it is clear that the 4th and 5th PCs capture finer changes in pose angles.

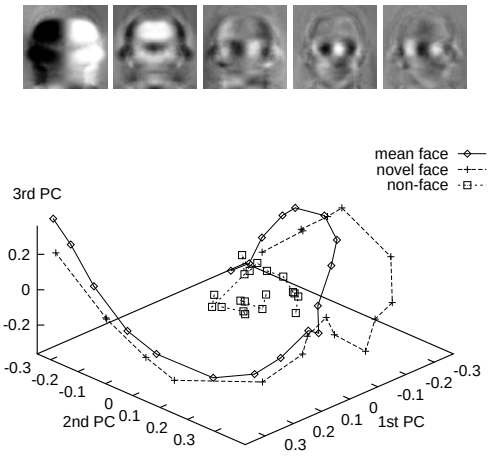

Figure 8. (1) Top row: the first 5 PC's (PCs) of the mean faces. (2) Plot: Projections onto the first 3 PC's of the mean face sequence, a novel face sequence and a non-face object (a fan) rotating from $-90°$ to $+90°$.

5.4 PES of Mean GWT Faces

We also derived a PES based on GWT face sequences. The second picture in Figure 6 shows an example of a GWT face. Similarly to the last experiment, we obtained a mean sequence $\bar{S}_r = (\bar{R}_0, \bar{R}_1, \ldots, \bar{R}_{n-1})$ by taking the mean GWT face at each pose angle over 11 sequences of different people. Figure 9 shows the pose distribution curve of this mean GWT sequence and the projections of two GWT face sequences into this PES. Compared with the PES of the mean intensity faces, the pose distributions in both 2nd and 3rd PC dimensions are more linear. This may be due to the fact that the GWT faces are less sensitive to changes in illumination and differences in local features. However, PES

of GWT faces is more sensitive to translations in the image-plane.

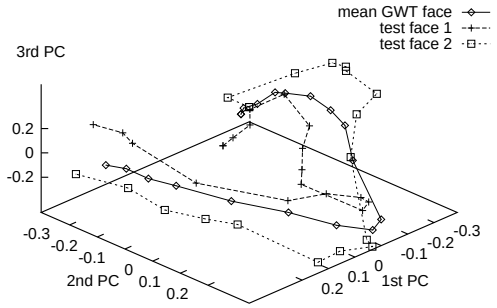

Figure 9. Pose distribution curves of (1) the mean GWT face representation of 11 face sequences (2) two test GWT face sequences. All 3 are projected into the mean GWT PES.

5.5 PES of Composite GWT Faces

We performed PCA similarly to the last two experiments with the composite GWT representation. Here, only a single spatial frequency was used to simplify the computation. Figure 10 shows the first 5 principal components of the mean composite GWT sequence. It is interesting to notice that while the sub-image of the 1st PC corresponding to horizontal orientation plays an important role in dividing the pose angles into two groups, the sub-image of the 1st PC corresponding to vertical orientation has relatively little significance. However, vertical orientation becomes a dominant factor in separating pose angles in all the other PCs. This is due to the fact that all the sequences used in our experiments are strictly based on face rotation from profile-to-profile. This suggests that when face sequences contain pose changes arising from diagonal rotations, the sub-images of PCs that correspond to $45°$ and $135°$ orientations may become more significant in separating poses. Figure 10 also shows pose distribution curves in the PES of the mean composite GWT faces. This plot reinforces our observations regarding the eigen-images. Compared to both PES of the mean intensity and GWT faces, the pose distribution curves are well linearised. As a result, the pose angles are clearly divided into two groups at the frontal view and are almost symmetrically distributed along two lines, clearly separable and easily measurable.

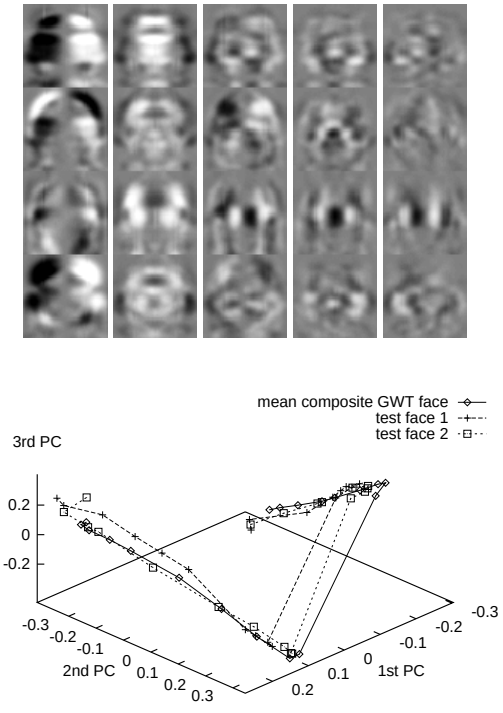

Figure 10. (1) The first 5 PC's of the mean composite sequence. The 4 sub-images correspond to Gabor responses at $0°$ (horizontal), $45°$, $90°$ (vertical) and $135°$. (2) Projections of the mean composite GWT face sequence and two test face sequences into the mean PES.

6 Conclusions

In this paper, we addressed the issue of measuring face pose. We introduced a composite face representation scheme based on a Gabor wavelet transform in order to both normalise intensity and scale and to investigate the role of locally oriented features in regularising pose distributions. We used pose eigenspaces based on principal components analysis to represent and interpret the distribution of pose changes from continuous face sequences of rotations in depth.

In particular, we have shown that pose changes of a continuous face rotation in depth form a smooth curve in pose eigenspace. Whilst the first principal component (PC) of this eigenspace divides all poses from profile-to-profile into two symmetric parts centred at the frontal view, the remaining PCs differentiate poses between profile to frontal views. The third PC also seems to capture changes in illumination.

Furthermore, it seems that the pose distribution curves of faces in the pose eigenspace are distinctively different from those of non-face objects. Although GWT representa-

tion reduces the complexity of pose distributions, it is sensitive to translational changes in the image-plane. More interestingly though, the composite GWT representation gives a highly linear pose distribution. It appears that the Gabor kernels of different orientation play some role in "regularising" pose distributions. This is computationally attractive for determining poses of novel faces. With further study, such a representation could be used to construct a simple but generic face pose eigenspace which in turn can be used to estimate poses of unknown faces. This can be done by projecting novel face images into the eigenspace and determining their positions along the pose distribution manifold by simply measuring Euclidean distance to the manifold [18]. Alternatively, the manifold could be modelled probabilistically by a set of covariance matrices at different poses before being used to measure poses based on computing Mahalonobis distance [17, 24].

As a final note, it is worth mentioning that in this paper, pose estimation has been treated essentially as a pattern recognition task. There clearly exist, however, a variety of spatial and temporal contextual cues such as body pose and continuity of pose change which could be used [1, 10, 11, 22]. This will be one of the main focuses of our future work.

References

[1] D. J. Beymer. Face recognition under varying pose. AI Memo 1461, MIT, Cambridge, Massachusetts, 1993.

[2] V. Bruce, A. Coombes, and R. Richards. Describing the shapes of faces using surface primitives. *Image and Vision Computing*, 11, 1993.

[3] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE PAMI*, 15(10), October 1993.

[4] H. Bülthoff, S. Edelman, and M. Tarr. How are three-dimensional objects represented in the brain? AI Memo 1479, MIT, Cambridge, Massachusetts, April 1994.

[5] M. Burl, T. Leung, and P. Perona. Face localization via shape statistics. In *IWAFGR*, Zurich, June 1995.

[6] N. Costen, I. Craw, and S. Akamatsu. Automatic face recognition: What representation? In *ECCV*, Cambridge, England, April 1996.

[7] I. Craw, E. Ellis, and J. Lishman. Automatic extraction of face features. *Pattern Recognition Letters*, 5, Febuary 1987.

[8] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. of Optical Society of America*, 2, 1985.

[9] S. Edelman. The features of recognition. Tech. Report CS-TR91-10, Weizmann Institutue, Israel, 1991.

[10] P. Foldiak. Learning invariance from transformation sequences. *Neural Computation*, 3, 1991.

[11] S. Gong, A. Psarrou, I. Katsoulis, and P. Palavouzis. Tracking and recognition of face sequence. In *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, Hamburg, November 1994.

[12] T. Kanade. Computer recognition of human faces. *Interdisciplinary Systems Res.*, 47, 1977.

[13] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE PAMI*, 12(1), 1990.

[14] M. Lando and S. Edelman. Generalization from a single view in face recognition. Tech. Report CS-TR95-02, Weizman Institute, Israel, 1995.

[15] S. McKenna and S. Gong. Tracking faces. Submitted to the 1996 IWAFGR, October 1996.

[16] S. McKenna, S. Gong, and H. Liddell. Real-time tracking for an integrated face recognition system. In *Second European Workshop on Parallel Modelling of Neural Operators*, Faro, Portugal, November 1995.

[17] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE ICCV*, Cambridge, Massachusetts, June 1995.

[18] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. *IJCV*, 14, 1995.

[19] S. K. Nayar, H. Murase, and S. A. Nene. Parametric appearance representations. In S. K. Nayar and T. Poggio, editors, *Early Visual Learning*, chapter 6. Oxford Univ. Press, 1996.

[20] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE CVPR*, Seattle, July 1994.

[21] T. Poggio and S. Edelman. A network that learns to recognize three-dimensioanl objects. *Nature*, 343, January 1990.

[22] A. Psarrou, S. Gong, and H. Buxton. Spatio-temporal trajectories and face signatures on partially recurrent neural networks. In *IEEE ICNN*, Perth, Australia, November 1995.

[23] R. P. N. Rao and D. H. Ballard. Natural basis functions and topographic memory for face recognition. In *IJCAI*, Montreal, 1995.

[24] K. Sung and T. Poggio. Example-based learning for view-based human face detection. Technical Report AI Memo 1512, CBCL 103, MIT, 1995.

[25] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition and gender determination. In *IWAFGR*, Zurich, 1995.

[26] R. P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*. Verlag Harri Deutsch, 1994.