Understanding Pose Discrimination in Similarity Space

Jamie Sherrah, Shaogang Gong, Eng-Jon Ong
Department of Computer Science
Queen Mary and Westfield College
London E1 4NS, UK
[jamie|sgg|ongej]@dcs.qmw.ac.uk

Abstract

Identity-independent estimation of head pose from prototype images is a perplexing task, requiring pose-invariant face detection. The problem is exacerbated by changes in illumination, identity and facial position. Facial images must be transformed in such a way as to emphasise differences in pose, while suppressing differences in identity. We investigate appropriate transformations for use with a similarity-to-prototypes philosophy. The results show that orientation-selective Gabor filters enhance differences in pose, and that different filter orientations are optimal at different poses. In contrast, PCA was found to provide an identity-invariant representation in which similarities can be calculated more robustly. We also investigate the angular resolution at which pose changes can be resolved using our methods. An angular resolution of $10°$ was found to be sufficiently discriminable at some poses but not at others, while $20°$ is quite acceptable at most poses.

1 Introduction

Head pose, closely related to gaze, is an important visual cue for interpretation of human behaviour and intentions. Estimation of head pose from video sequences is a highly complex task, since it implicitly requires face detection at arbitrary pose.

Our approach to identity-independent pose estimation over a wide range of angles is based on similarities to prototypes [4, 6]. The approach uses second-order similarity to obtain robust similarity measures from sparse data [1]. Our current system estimates head yaw (azimuth) ϕ on the range $[0°, 180°]$ and tilt (elevation) θ on the range $[60°, 120°]$. The face prototype database consists of facial images from several different people at different poses in $10°$ increments. Given a novel face image and hypothesised head pose, the distance of each prototype to the novel face is calculated. The novel face is then represented in similarity space as a vector of dis-similarities: $\mathbf{s}_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,N}]^{\mathrm{T}}$, where N is the number of human subjects in the database, and $d_{i,j}$ is the distance from the face i to subject j at the same hypothesised pose. The principle is that distances in similarity space are smoother than in the original high-dimensional image space.

BMVC 1999 doi:10.5244/C.13.52

The method primarily relies on a general assumption that **different people at the same pose look more similar than the same person at different poses**. In other words, pose is a stronger indicator of image-space similarity than identity. This assumption is here referred to as the *pose similarity assumption.* As one can imagine, this assumption is valid only for significant changes in pose. Nevertheless, even for significant pose differences the assumption may be invalid because intensity images are sensitive to variations in illumination and mis-alignment. To validate the assumption, the facial images must be transformed to compensate for these variations, and to emphasise differences in pose over differences in identity.

The work presented here investigates the following two issues. First, for a given pose, what transformation of the images is optimal to exaggerate differences in pose and suppress differences in identity? Second, what is the minimum angular separation that can be resolved using similarity-based methods?

The most obvious transformation for images is to apply an image filter. The optimal filtering of prototype images is expected to be different at each pose angle, because different features are important at different poses [3]. The most natural filtering of images for this task is to use orientation-selective features. Gabor filters are particularly appropriate because they incorporate smoothing which reduces sensitivity to spatial mis-alignments. Recent studies on Gabor filters have shown that these filters are approximately the basis functions for natural images [7].

While filtering may enhance pose-specific features, it is expected to provide only small invariance to identity. Intuitively, a representation of the images is required that encodes only very coarse-scale intensity variations with pose. It has been shown in [3] that principal component analysis (PCA) can be used to discard identity information while maintaining pose information. PCA has the extra advantage that similarity measures in a low-dimensional space are more robust and easier to compute than in a high-dimensional space.

In this work, we define a criterion to quantify the goodness of a given transformation method for pose prototypes. The criterion is then used in a series of experiments. In the first experiment, Gabor filters are examined as a method for enhancing pose differences at each pose angle. In the second experiment, PCA is used to represent prototypes and its identity-invariant properties are examined. In the third experiment, the criterion is used to determine the angular resolution at which neighbouring poses can be resolved.

2 The Pose Similarity Ratio

When matching images from various poses to a group of prototype images, it is desirable to calculate similarity in a space that is invariant to identity and sensitive to differences in pose. To select a good transformation, a criterion is required to allow us to compare image representations. The criterion should be based on the pose similarity assumption, that differences in pose are more significant than differences in identity. Our criterion is defined as the following ratio:

$$r\left(\phi, \theta, f(\cdot)\right) = \frac{\bar{d}\left(\phi, \theta, f(\cdot)\right)}{\bar{d}\left(\phi \pm \delta\phi, \theta \pm \delta\theta, f(\cdot)\right)} \tag{1}$$

This ratio shall be referred to as the *pose similarity ratio* where:

- $f(\cdot)$ is a transformation function that maps the images to some other representation either with the same dimensionality, *e.g.:* an image filter, or with lower dimensionality, *e.g.:* linear projection.
- $\bar{d}(\phi, \theta, f(\cdot))$ is the average distance between f-transformed prototypes of varying identity at a given pose:

$$\bar{d}(\phi, \theta, f(\cdot)) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d\left(f(\mathbf{x}_{\phi,\theta}^{i}), f(\mathbf{x}_{\phi,\theta}^{j})\right)}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 1} \tag{2}$$

where $\mathbf{x}_{\phi,\theta}^{i}$ is the prototype image of subject i at pose angles (ϕ, θ), and $d(\mathbf{x}_1, \mathbf{x}_2)$ is the distance between two points in high-dimensional space.

- $\bar{d}(\phi \pm \delta\phi, \theta \pm \delta\theta, f(\cdot))$ is the average distance between f-transformed prototypes at the given pose and prototypes of varying identity and pose over the given range of neighbouring poses:

$$\bar{d}(\phi \pm \delta\phi, \theta \pm \delta\theta, f(\cdot)) = \tag{3}$$

$$\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{y=\phi-\delta\phi}^{y=\phi+\delta\phi} \sum_{t=\theta-\delta\theta}^{t=\theta+\delta\theta} d\left(f(\mathbf{x}_{y,t}^{i}), f(\mathbf{x}_{y,t}^{j})\right) .\delta(y-\phi, t-\theta)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{y=\phi-\delta\phi}^{y=\phi+\delta\phi} \sum_{t=\theta-\delta\theta}^{t=\theta+\delta\theta} \delta(y-\phi, t-\theta)}$$

where $\delta\phi$, $\delta\theta$ are the sizes of the yaw and tilt neighbourhoods, and $\delta(y - \phi, t - \theta)$ is a delta function to discount the distance of a prototype to itself:

$$\delta(a, b) = \begin{cases} 0 & \text{if } a = 0 \text{ and } b = 0; \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

The ratio can be interpreted as follows: when the ratio is small, faces at the given pose are more similar to each other than to faces at neighbouring poses, and the pose similarity assumption is valid. For large ratio values, faces at neighbouring poses are more similar than at the same pose, and the assumption is invalid. At a given pose, the ratio can be minimised with respect to $f(\cdot)$.

We now describe three experiments using the ratio criterion. All results are based on a database of 30×30 images collected from $N = 8$ subjects at poses over the pose sphere of range $\phi \in [0°, 10°, \ldots, 180°]$ and $\theta \in [60°, 70°, \ldots, 120°]$. In all experiments, the distance function used was Euclidean distance. Images were always post-normalised by subtracting the mean intensity from each pixel and dividing by the intensity standard deviation.

3 Filtering for Pose Discrimination

Consider what sort of image filters would be appropriate for discriminating different poses. It is expected that different image features are important at different poses, and that those features will be oriented differently. For example, the mouth may be important at frontal poses and the nose at profile poses. Therefore filters that highlight oriented features are appropriate. In this section, we investigate whether Gabor filters are useful for discriminating pose.

Gabor filters are oriented sinusoidal filters modulated by a Gaussian envelope. Examples of Gabor filters are shown in Figure 1 for angles 0, 30, 60, 90, 120 and 150 degrees. The real and imaginary parts are shown on the left and right respectively. These filters have a natural application for pose estimation because pose estimation involves variations in orientation [3].

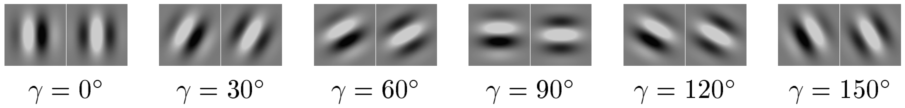
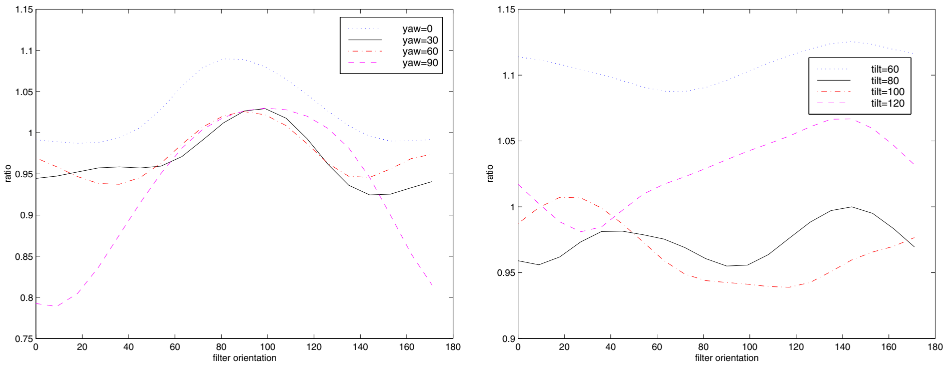


Figure 1: Gabor filters at different orientations γ. The real part is on the left, imaginary on the right.

To see whether Gabor filters are useful for discriminating pose, let us evaluate the pose similarity ratio of Equation(1) at a fixed pose but with varying Gabor filter orientation. The filter orientation γ is varied from 0° to 180° in 9° increments. The tilt angle is fixed at 90° (frontal view) and is not varied in the calculation of the ratio, *i.e.:* $\delta\theta = 0$. The yaw neighbourhood $\delta\phi$ is set to 30°and the size of the filters is 13×13. The result is a series of ratio values versus filter orientation. The process has been repeated at different fixed poses with yaw varying over the range $[0°, 90°]$, and tilt fixed at 90°. The results are shown in Figure 2(a). Clearly the ratios vary smoothly with filter orientation, and there are well-defined minima in the curves. The implication is that Gabor filters reveal oriented features in the facial images that are specifically appropriate for discrimination at a given pose.



(a) Varying yaw, with tilt fixed at 90°. The neighbourhood is based on yaw only.

(b) Varying tilt, with yaw fixed at 90°. The neighbourhood is based on tilt only.

Figure 2: Pose similarity ratios for varying head pose and filter orientation.

In Figure 2(b), the correlations between filter orientations and pose variations in tilt are presented. Yaw is fixed at 90°, and the pose neighbourhood is $\delta\phi = 0°$, $\delta\theta = 10°$. Tilt is varied over 60° to 120°. Again it is observed that the ratios vary smoothly with filter orientation, and that the curves contain well-defined minima. We can conclude that features at a specific orientation are important for discriminating poses. This raises the question: does the best filter orientation vary with pose, and if so, how does the orientation vary across the pose sphere?

Let us now proceed to examine the best single orientation-selective Gabor filter for each pose by minimising the pose similarity ratio at each pose. To determine the best filter size, the average minimum ratio for a range of filter sizes is shown in Table 3. It can be noted that the ratio decreases monotonically with the filter size. Taking 13×13 as the filter size, and using a neighbourhood of $\delta\phi = 20^\circ$, $\delta\theta = 10^\circ$, the optimal filter orientations and corresponding ratios are shown in Figure 3.

Filter Size (in pixels)	9	11	13	15
Average Best Ratio	0.974617	0.964791	0.958090	0.953220

Table 1: Average minimum pose similarity ratios for filters of different sizes.

Examining Figure 3(a), it is clear that different orientations are optimal for different poses. Considering that the pose database itself contains spatial and pose mis-alignments, the filter orientations vary gradually with pose angle. There is also a fair degree of symmetry in the orientations about central yaw, $\phi = 90^\circ$. In Figure 3(b), the minimum ratios are represented as intensities, with darker colours denoting lower (better) ratios. The pose angles containing an "×" have a ratio greater than 1. The results show that Gabor filters are able to discriminate faces from neighbouring poses except at some poses on the fringe of the pose sphere.

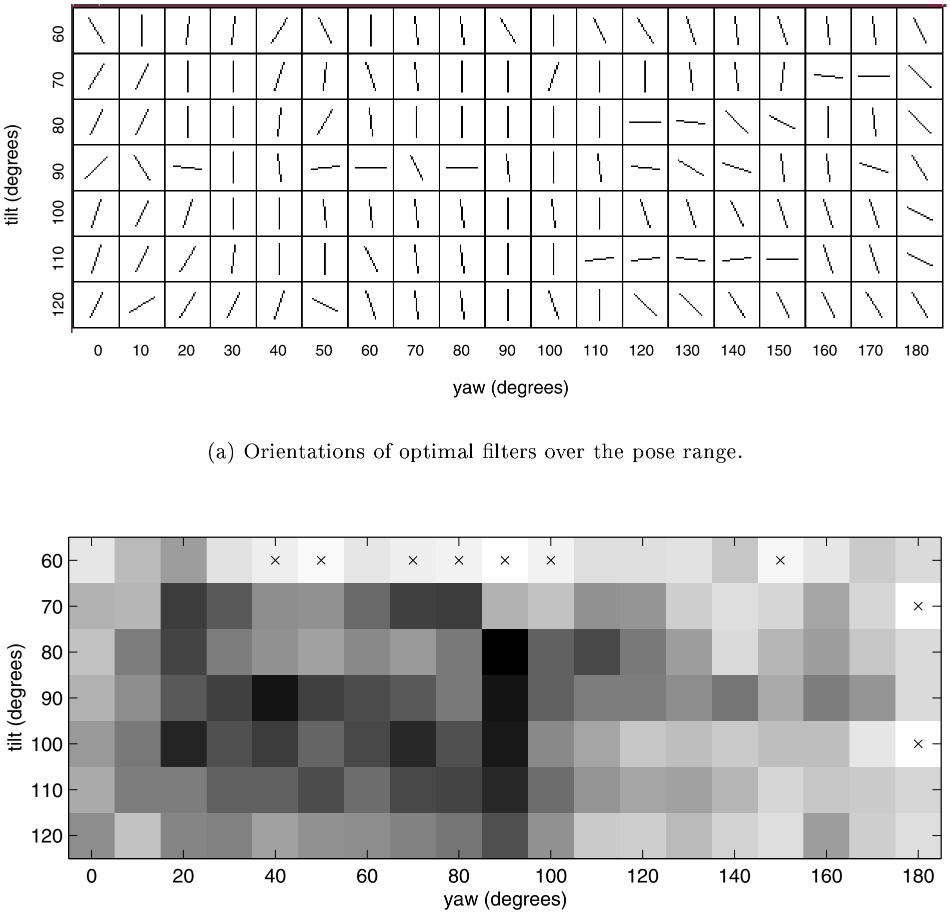
There are a few other points of interest from Figure 3(b). The lowest ratios are at frontal yaw, reinforcing the intuition that pose discrimination is easier at frontal views. The ratios when the subject is looking upwards are generally worse than when looking downwards. This could either indicate that the database acquisition system is less accurate at low tilts, or it could be a natural phenomenon. The asymmetry in ratios about central yaw is due to mis-alignments and varying illumination conditions in the database.

To summarise, orientation-specific features are found in facial images at different poses, and Gabor filters can be used to find these features. We now proceed to look at transformation for identity invariance.

4 Identity Invariance through PCA

We have seen how orientation-selective filters can emphasise differences in pose, but can we also suppress identity? To obtain some invariance to identity, we investigate the use of principal component analysis (PCA) on the pose data. A previous investigation into pose distributions in PCA space found that continuous changes in yaw result in smooth manifolds in eigen-space with identity collapsed [3]. Here we extend the study by calculating the pose similarity ratio based on similarities calculated in the PCA space.

To examine pose manifolds in PCA space, two PCA bases are calculated: one from images with tilt fixed at 90° and yaw varying from 0° to 90°, and the other with yaw fixed at 90° and tilt varying from 60° to 120°. The range of poses for the investigation is restricted so that the PCA bases are based on a manageable

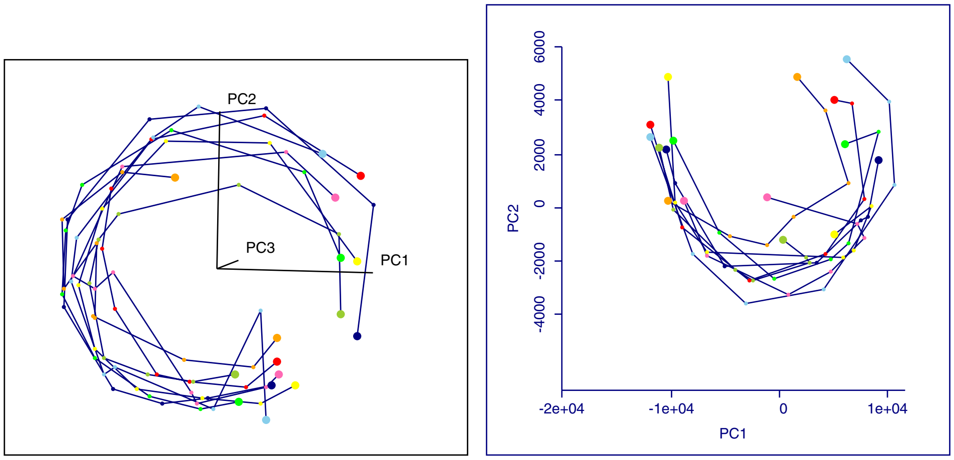


(a) Orientations of optimal filters over the pose range.

(b) Corresponding minimum pose similarity ratios with whiter cells corresponding to higher ratios. Ratios greater than one are denoted by "×".

Figure 3: Results for best filters of size 13 × 13.

range of intensity variations. In each case, prototypes from all 8 subjects are used to construct a PCA basis, and all images are blurred and normalised before use. Figure 4 shows the prototypes of varying pose projected onto the first major principal components. Prototypes belonging to the same person are joined by a line in order of pose. In Figure 4(a) for varying yaw, the curves form a horseshoe shape, but the identities are clustered fairly tightly. The first two principal components account for 54% of the variance in the data. In Figure 4(b) as tilt is varied, the same manifold shape is observed, and the first two components account for 55% of the variance. The first two principal components largely describe changes in pose, while the remaining components primarily encode changes in identity and facial expression. Therefore, projection onto the first two principal components provides a representation that is invariant to identity but sensitive to pose.



(a) Prototypes at $\theta = 90°$ and $\phi = [0, 10, \ldots, 90]$ projected onto first 3 principal components.

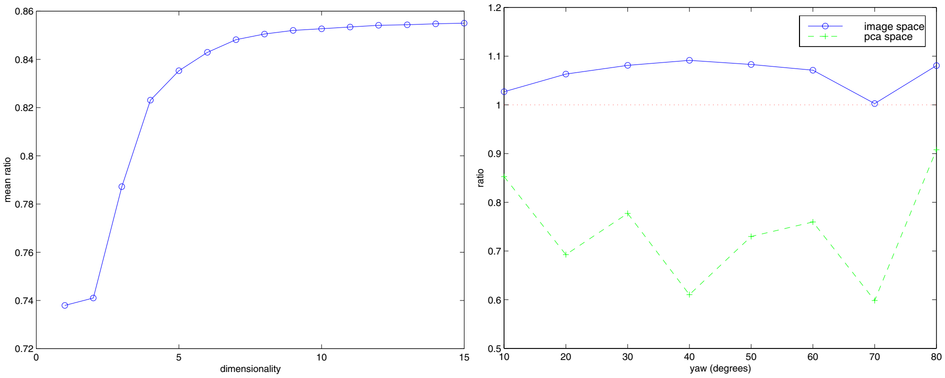
(b) Prototypes at $\phi = 90°$ and $\theta = [60, 70, \ldots, 120]$ projected onto first 2 principal components.

Figure 4: Pose manifolds in PCA space.

The PCA bases look appealing, but do they maintain sufficient discernibility between poses? To investigate, we calculate the pose similarity ratio *with distances calculated in PCA space.* The ratio is calculated at a range of poses covered by the PCA bases using a neighbourhood only in the axis of pose variation. For varying yaw, the neighbourhood is $\delta\phi = 10°$, $\delta\theta = 0°$, and for varying tilt the neighbourhood is $\delta\phi = 0°$, $\delta\theta = 10°$. The average best ratio is plotted versus the number of principal components used, where the average is over varying yaw in Figure 5(a), and over varying tilt in Figure 6(a). Comparing with the mean ratios for Gabor filters in image space shown in Table 3, the PCA-based ratios are much lower. Therefore PCA not only maintains good pose discrimination, it does so much more effectively than in the image space.

Using only the first two principal components to calculate similarities, the ratios are plotted for varying yaw and tilt in Figures 5(b) and 6(b). On the same axes, the ratios are plotted for similarities measured in image space (no PCA, but blurred and normalised). Comparing the ratios with and without PCA, it is clear that PCA is a much more appropriate representation for the similarity calculations. Relating these results back to the Gabor filters, the non-PCA ratios plotted here are consistently higher than those obtained using Gabor filters, emphasising the need for Gabor filtering to exaggerate pose differences in image space.

In summary, PCA is an appropriate representation for pose similarity prototypes because it suppresses identity variations while maintaining sensitivity to pose. We have also seen that pose similarity ratios both in PCA space and after Gabor filtering in image space are better than those based on the original images. The fact that lower ratios are obtained with PCA than when using the Gabor fil-



(a) Lowest ratio averaged over yaw versus number of PCA coefficients.

(b) Ratios versus yaw angle for similarities calculated in image space and in PCA space using the first 2 coefficients.

Figure 5: Comparison of pose similarity ratios calculated in image space and PCA space, for varying yaw angles.

ters does not necessarily mean the orientation-selective filters are no longer needed. Such a comparison is unfair because distance calculations are generally less robust in the high-dimensional image space due to the curse of dimensionality. Gabor filters are also expected to improve the smoothness of the PCA representation by reducing the first two components' sensitivity to illumination changes.

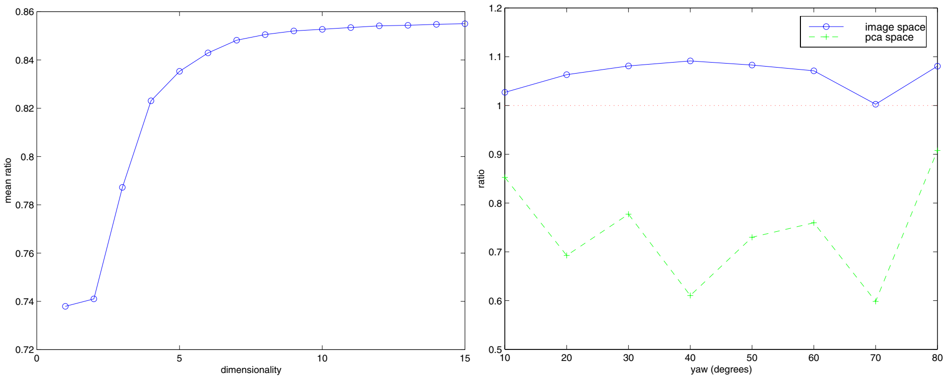
5 Valid Angular Resolution

Logically there is a limit to the angular resolution with which poses can be discriminated using similarity-based methods. For example, at differences of 1° yaw, two 30×30 facial images would look so similar as to be indistinguishable. So the question arises: what is the minimum angular resolution at which pose differences can be discerned in the presence of varying identity and illumination? To find out, we modify the denominator of the pose similarity ratio. Equation(3) becomes:

$$\bar{d}\left(\phi \pm \delta\phi, \theta \pm \delta\theta, f(\cdot)\right) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{y=\phi\pm\delta\phi} \sum_{t=\theta\pm\delta\theta} d\left(f(\mathbf{x}_{y,t}^{i}), f(\mathbf{x}_{y,t}^{j})\right)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{y=\phi\pm\delta\phi} \sum_{t=\theta\pm\delta\theta} 1} \qquad (5)$$

Here the ratio only involves neighbouring poses **at** $\phi \pm \delta\phi$, rather than all poses in the range of $\phi - \delta\phi, \ldots, \phi + \delta\phi$, and similarly for θ. This is akin to sampling the database at a lower angular resolution. Now we can plot the modified ratio versus angular resolution to find the minimum acceptable resolution.

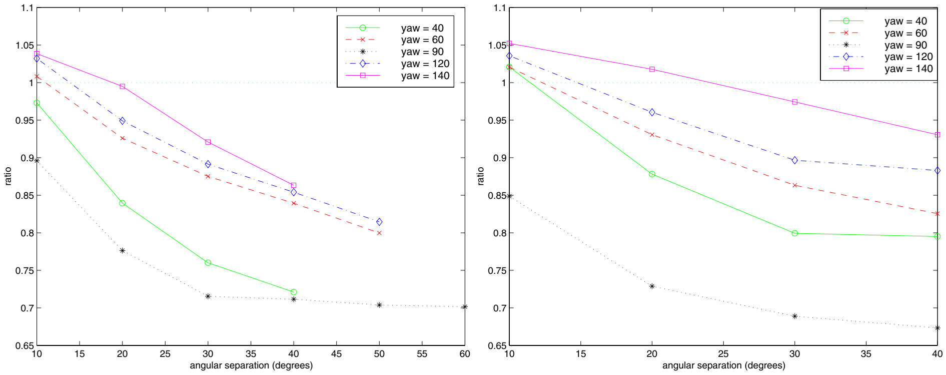
At a range of yaws and two different tilts, the pose similarity ratio is calculated for the optimal filter (see Section 3) at varying yaw resolution $\delta\phi \in [10°, 60°]$ but with no tilt neighbourhood, $\delta\theta = 0$. The results are shown in Figure 7, with the



(a) Lowest ratio averaged over tilt versus number of PCA coefficients.

(b) Ratios versus tilt angle for similarities calculated in image space and in PCA space using the first 2 coefficients.

Figure 6: Comparison of pose similarity ratios calculated in image space and PCA space, for varying tilt angles.



(a) $\theta = 90°$

(b) $\theta = 110°$

Figure 7: Ratio versus different angular resolution across yaws at different tilts.

$r = 1$ threshold marked as a dotted line. As expected, the ratios monotonically decrease with angular separation, because it is easier to discriminate larger changes in pose. For each tilt angle, 10° angular separation is not sufficient for some yaw angles because the ratio exceeds 1. At 20°, however, the angular separation is generally sufficient. The fact that the ratio is less than 1 at some yaws but greater than 1 at others implies that different angular resolution may be required at different poses. This requirement may arise because the problem is harder at these poses, or because the noise in the acquisition system is higher at these poses.

6 Conclusion

We have presented an analysis of face similarity distributions under varying head pose for different types of image transformation, with the aim of understanding pose in similarity space. Orientation-selective Gabor filters were found to detect features for pose discrimination. Dimensionality reduction through PCA was found to provide invariance to identity while accurately describing pose changes. PCA also has the advantages of being understandable through visualisation, and more computationally efficient, since similarities are calculated in the low dimensional space. The lowest angular separation at which pose differences can be feasibly detected was also investigated. A greatest lower bound of approximately 20° was determined, and the actual minimum resolution may be 10° or lower at some poses.

Overall, this work has shown that pose differences can be enhanced and identity similarities suppressed within a similarity-space framework using inexpensive algorithms. Such findings should facilitate the development of real-time pose estimation systems. Some remaining issues are:(i) The optimal filter orientation at each pose is not necessarily unique. Indeed, Gabor filters may not be the optimal filters for pose estimation. A more general approach could be taken by adapting the filter orientation locally within the facial images [2]. (ii) The benefits of pose-selective filters and PCA need to be combined. The main difficulty lies in creating PCA bases from images that have been filtered differently. (iii) PCA may not be the best linear projection for removal of identity information. For instance, linear discriminant analysis could be used to find the projection that maximises discrimination between faces at different poses. Such an approach has previously been adopted to achieve invariance to illumination conditions and facial expression [5].

References

[1] S. Edelman. *Representation and Recognition in Vision.* MIT Press, June 1999.

[2] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE PAMI*, 13(9):891–906, October 1991.

[3] S. Gong, S. McKenna, and J. Collins. An investigation into face pose distributions. In *IEEE Int. Conf. on Face & Gesture Recognition*, pages 265–270, Vermont, USA, October 1996.

[4] S. Gong, E. Ong, and S. McKenna. Learning to associate faces across views in vector space of similarities to prototypes. In *British Machine Vision Conference*, volume 1, pages 54–64, Southampton, UK, September 1998.

[5] Baback Moghaddam and Alex Pentland. Probabilistic visual learning for object representation. *IEEE PAMI*, 19(7):696–710, July 1997.

[6] E. Ong, S. McKenna, and S. Gong. Tracking head pose for inferring intention. In *European Workshop on Perception of Human Action*, Freiburg, June 1998.

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