Point Distribution Models for Pose Robust Face Recognition: Advantages of Correcting Pose Parameters Over Warping Faces to Mean Shape

Daniel González-Jiménez and José Luis Alba-Castro[⋆]

Departamento de Teoría de la Señal y Comunicaciones
Universidad de Vigo (Spain)

Abstract. In the context of pose robust face recognition, some approaches in the literature aim to correct the original faces by synthesizing virtual images facing a standard pose (e.g. a frontal view), which are then fed into the recognition system. One way to do this is by warping the incoming face onto the average frontal shape of a training dataset, bearing in mind that discriminative information for classification may have been thrown away during the warping process, specially if the incoming face shape differs enough from the average shape. Recently, it has been proposed a method for generating synthetic frontal images by modification of a subset of parameters from a Point Distribution Model (the so-called pose parameters), and texture mapping. We demonstrate that if only pose parameters are modified, client specific information remains in the warped image and discrimination between subjects is more reliable. Statistical analysis of the verification experiments conducted on the XM2VTS database confirm the benefits of modifying only the pose parameters over warping onto a mean shape.

1 Introduction

It is well known that the performance of face recognition systems drops drastically when pose differences are present within the input images, and it has become a major goal to design algorithms that are able to cope with this kind of variations. Some of these approaches aim to synthesize faces across pose in order to cope with viewpoint differences. One of the earliest attempts was done by Beymer and Poggio [1]: from a single image of a subject and making use of face class information, virtual views facing different poses were synthesized. For the generation of the virtual views, two different techniques were used: linear classes and parallel deformation. Vetter and Poggio also took advantage of the concept of linear classes to synthesize face images from a single example in [8]. Blanz et al. employed the 3D Morphable Model [6] to synthesize frontal faces from non frontal views in [7], which were then fed into the recognition system. In this same direction, other researchers have tried to generate frontal faces from non frontal views, like the works proposed by Xiujuan Chai et al. in [4], via linear regression

[⋆] This work was supported with funds provided partially by the Spanish ministry of education under project TEC2005-07212/TCM and the European sixth framework programme under the Network of Excellence BIOSECURE (IST-2002-507604)

González-Jiménez D. and Luis Alba-Castro J. (2007).
Point Distribution Models for Pose Robust Face Recognition: Advantages of Correcting Pose Parameters Over Warping Faces to Mean Shape.
In *Proceedings of the 7th International Workshop on Pattern Recognition in Information Systems*, pages 138-147
DOI: 10.5220/0002428001380147

in each of the regions in which the face is divided, and in [5] where a 3D model is used. More recently, González-Jiménez and Alba-Castro [9] have proposed an approach for generating frontal faces via modification of a subset of parameters from a Point Distribution Model (so-called pose parameters) and texture mapping. Another possibility is to warp the face image onto a frontal standard shape (e.g. the average frontal shape of a training dataset) prior to recognition. This solution is adopted by methods that use holistic features for face representation (e.g. Eigenfaces [2]), and that need all images to be embedded into a constant reference frame. For example, Lanitis et al. [10] deformed each face image to the mean shape using 14 landmarks, extracted shape and appearance parameters and classified using the Mahalanobis distance.

The difference between the synthetic images obtained using the methods described in [9] and [10] relies on the generation of the synthetic frontal shapes onto which the original faces must be warped. It seems rather safe to think that warping faces onto a mean shape may provoke discriminative information reduction, specially if the incoming face shape differs enough from the average shape (see Figure 1). On the other hand, [9] did not analyze whether the modification of the pose parameters had any influence on non-rigid factors (such as expression and identity), which could provoke non desirable effects in the synthesized faces. The goal of this paper is two-fold:

1. Show that, if the training set is chosen appropriately, pose parameters do not contain important non-rigid (expression/identity) information, and
2. Propose an empirical comparison of the synthetic images obtained with the methods described in [9] and [10] respectively. Obviously, we need a non-subjective way to compare the two approaches and, to this aim, we conducted face verification experiments on the XM2VTS database [11] using Gabor filtering for feature extraction.

Fig. 1. Images from subject 013 of the XM2VTS. Left: Original image. Right: Image warped onto the average shape. Observe that subject-specific information has been reduced (specially in the lips region).

The paper is organized as follows. Next section briefly reviews Point Distribution Models, and introduces the concepts of Pose Eigenvectors and Pose Parameters. The two techniques used to generate frontal face images are presented in Section 3. Section 4 shows the results of the verification experiments conducted on the XM2VTS database [11]. Finally, conclusions are drawn in Section 5.

2 A Point Distribution Model For Faces

A point distribution model (PDM) of a face is generated from a set of training examples. For each training image I_i, N landmarks are located and their normalized coordinates (by removing translation, rotation and scale) are stored, forming a vector $\mathbf{X}_i = (x_{1i}, x_{2i}, \ldots, x_{Ni}, y_{1i}, y_{2i}, \ldots, y_{Ni})$. The pair (x_{ji}, y_{ji}) represents the normalized coordinates of the j-th landmark in the i-th training image. Principal Components Analysis (PCA) is performed to find the most important modes of shape variation. As a consequence, any training shape $\mathbf{X}_i$ can be approximately reconstructed:

$$\mathbf{X}_i = \bar{\mathbf{X}} + \mathbf{Pb}, \tag{1}$$

where $\bar{\mathbf{X}}$ stands for the mean shape, $\mathbf{P} = [\phi_1|\phi_2|\ldots|\phi_t]$ is a matrix whose columns are unit eigenvectors of the first t modes of variation found in the training set, and $\mathbf{b}$ is the vector of parameters that define the actual shape of $\mathbf{X}_i$. So, the k-th component from $\mathbf{b}$ ($b_k, k = 1, 2, \ldots, t$) weighs the k-th eigenvector ϕ_k. Also, since the columns of $\mathbf{P}$ are orthogonal, we have that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, and thus:

$$\mathbf{b} = \mathbf{P}^T \left(\mathbf{X}_i - \bar{\mathbf{X}}\right), \tag{2}$$

i.e. given any shape, it is possible to obtain its vector of parameters $\mathbf{b}$. We built a 62-point PDM using manually annotated landmarks (some of them were provided by the FGnet project[1], while others were manually annotated by ourselves). Figure 2 shows the position of the landmarks on an image from the XM2VTS database [11].

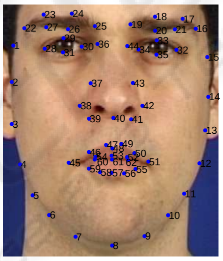

Fig. 2. Position of the 62 landmarks used in this paper on an image from the XM2VTS database.

2.1 Pose Eigenvectors and Pose Parameters

Among the obtained modes of shape variation, the authors of [9] identified the eigenvectors that were responsible for controlling the apparent changes in shape due to rigid facial motion. However, it was not analyzed whether the modification of these pose parameters had any influence on non-rigid factors (such as expression and identity), which

[1] Available at `http://www-prima.inrialpes.fr/FGnet/data/07-XM2VTS/xm2vts-markup.html`

could provoke non desirable effects on the synthesized faces. In this section, we show that if the training data are appropriately chosen, the identified pose parameters do only account for pose variations.

Clearly, the eigenvectors (and their relative position) obtained after PCA strongly depend on the training data and hence, the choice of the examples used to build the PDM is critical. In fact, if all training meshes were strictly frontal, there would not appear any eigenvector explaining rotations in depth. However, if we are sure that pose changes are present in the training set, the eigenvectors explaining those variations will appear among the first ones, due to the fact that the energy associated to rigid facial motion should be higher than that of most expression/identity changes (once again, depending on the specific dataset used to train the PDM). With our settings, it turned out that ϕ_1 controlled up-down rotations (see Figure 3) while ϕ_2 (Figure 4) was the responsible for left-right rotations. Hence, given a mesh X with a vector of shape parameters $\mathbf{b} = [b_1, b_2, \ldots, b_t]^T$, we can change the values of b_1 and b_2 (i.e. the pose parameters), and use equation (1) to generate a synthetic mesh X_2 facing a different pose.

A major problem, inherent to the underlying PCA analysis, relies on the fact that a given pose-eigenvector may not only contain rigid facial motion (pose) but also non-rigid (expression/identity) information, mostly depending on the training data used to build the PDM. The reconstructed shapes in Figure 4, show that expression changes are not noticeable when sweeping b_2. Regarding ϕ_1, it has been shown [15] that there exists a dependence between the vertical variation in viewpoint (nodding) and the perception of facial expression, as long as faces that are tilted forwards (leftmost shape in Figure 3) are judged as happier, while faces tilted backwards (rightmost shape in Figure 3) are judged as sadder. Apart from this subjective perception, we provide visual evidence in the next section suggesting that the influence of non-rigid factors within ϕ_1 and ϕ_2 is small.

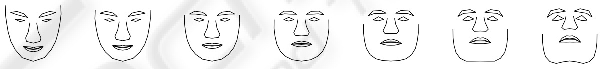

Fig. 3. Effect of changing the value b_1 on the reconstructed shapes. ϕ_1 controls the up-down rotation of the face.

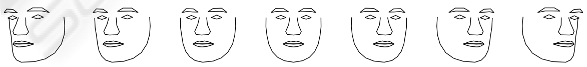

Fig. 4. Effect of changing the value b_2 on the reconstructed shapes. ϕ_2 controls the left-right rotation of the face.

2.2 Experiment on a Video-sequence: Decoupling of Pose and Expression

In order to demonstrate that the presence of non-rigid factors within the identified pose-eigenvectors is minimal, we used a manually annotated video-sequence of a man during conversation[2] (hence, rich in expression changes). For each frame f in the video, the vector of shape parameters

$$\mathbf{b}\left(f\right) = \left[b_1\left(f\right), b_2\left(f\right), \ldots, b_t\left(f\right)\right]^T$$

of the corresponding mesh $X(f)$ was calculated and splitted into the rigid (pose) part

$$\mathbf{b}_{pose}\left(f\right) = \left[b_1\left(f\right), b_2\left(f\right), 0, \ldots, 0\right]^T$$

and the non-rigid (expression) part

$$\mathbf{b}_{exp}\left(f\right) = \left[0, 0, b_3\left(f\right), \ldots, b_t\left(f\right)\right]^T$$

Finally, we calculated the reconstructed meshes $X_{pose}\left(f\right)$ and $X_{exp}\left(f\right)$ using equation (1) with $\mathbf{b}_{pose}\left(f\right)$ and $\mathbf{b}_{exp}\left(f\right)$ respectively. Ideally, $X_{pose}\left(f\right)$ should only contain rigid mesh information, while $X_{exp}\left(f\right)$ should reflect changes in expression and contain identity information. As shown in Figure 5, it is clear that although there exists some coupling (specially in the seventh row with small eyebrow bending in $X_{pose}(f)$), $X_{exp}(f)$ is responsible for expression changes and identity information (face shape is clearly encoded in $X_{exp}(f)$) while $X_{pose}(f)$ does mainly contain rigid motion information. For instance, the original shapes from the first and second rows share approximately the same pose, but differ substantially in their expression. Accordingly, the X_{pose}'s are approximately the same while the X_{exp}'s are clearly different.

3 Synthesizing Frontal Faces

Given two faces I_A and I_B to be compared, the system must output a measure of similarity (or dissimilarity) between them. Straightforward texture comparison between I_A and I_B may not produce desirable results as differences in pose could be quite important. In this section, we describe the two approaches proposed in [10] and [9] for frontal face synthesis. These two methods share one common feature: given one face image I, the coordinates of its respective fitted mesh, X, and a new set of coordinates, X_2, a synthetic face image must be generated by warping the original face onto the new shape. For this purpose, we used a method developed by Bookstein [13], based on thin plate splines. Provided the set of correspondences between X and X_2, the original face I is allowed to be deformed so that the original landmarks are moved to fit the new shape.

[2] `http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html`

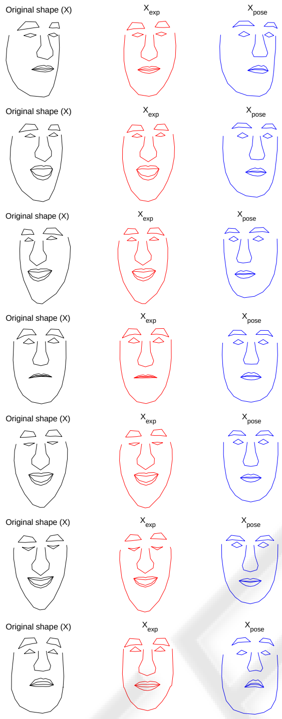

Fig. 5. Experiment on the video sequence. Each row shows, for a given frame f, the original shape $X(f)$ and the reconstructed shapes ($X_{exp}(f)$ and $X_{pose}(f)$) using $b_{exp}(f)$ and $b_{pose}(f)$ respectively. Clearly, $X_{exp}(f)$ controls expression and identity while $X_{pose}(f)$ is mostly responsible for rigid changes.

3.1 Warping to Mean Shape (WMS) [10]

Once the meshes have been fitted to I_A and I_B, both faces are warped onto the average shape of the training set, $\bar{\mathbf{X}}$, which corresponds to setting all shape parameters to 0, i.e. $\mathbf{b}_A = \mathbf{b}_B = \mathbf{0}$. Thus, the images are deformed so that a set of landmarks are moved to coincide with the correspondent set of landmarks on the average shape, obtaining $\bar{I}_A$ and $\bar{I}_B$. The number of landmarks used as "anchor" points is another variable to be fixed. For the experiments, we used two different sets:

- The whole set of 62 points.
- The set of 14 landmarks used in [10].

As the number of "anchor" points grows, the synthesized image is more likely to present artifacts because more points are forced to be moved to landmarks of a mean shape (which may differ significantly from the subject's shape). On the other hand, with few "anchor" points, little pose correction can be made.

3.2 Normalizing to Frontal Pose and Warping (NFPW) [9]

Once demonstrated that the pose parameters do only account for pose variations (Sections 2.1 and 2.2), we suggest that normalizing only these parameters should produce better results than warping images onto a mean shape, as long as we are not modifying identity information. In Figure 6, we can see a block diagram of this method. Given $\mathbf{b}_A$ and $\mathbf{b}_B$, only the pose parameters are fixed to the typical values of frontal faces from the training set (as the average shape corresponds to a frontal face, we fixed pose parameters to zero, i.e. $\mathbf{b}_A^{pose} = \mathbf{b}_B^{pose} = \mathbf{0}$). New coordinates are computed using Equation 1, and virtual images, $\hat{I}_A$ and $\hat{I}_B$, are synthesized.

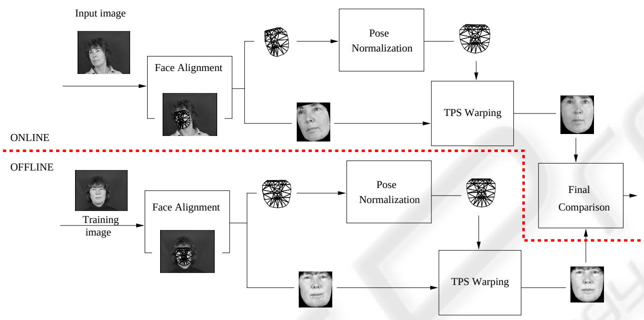

Fig. 6. Block diagram for pose correction using *NFPW*. After face alignment, the obtained meshes are corrected to frontal pose (*Pose Normalization* block), and virtual faces are obtained through Thin Plate Splines (TPS) warping. Finally, both synthesized images are compared. It is important to note that the processing of the training image could (and should) be done offline, thus saving time during recognition.

4 Face Authentication on the XM2VTS Database

Using the XM2VTS database [11], authentication experiments were performed on configuration I of the Lausanne protocol [12] in order to confirm the advantages of modifying only pose parameters over warping onto a mean shape.

4.1 Feature Extraction

Holistic approaches such as eigenfaces [2] need all images to be embedded into a constant reference frame (an average shape for instance), in order to represent these images as vectors of ordered pixels. This constraint is violated by the faces obtained through $NFPW$, leading us to the use of local features: Gabor jets as defined in [3] are extracted at each of the pose corrected mesh coordinates and stored for further comparison.

4.2 Database and Experimental Setup

The XM2VTS database contains face images recorded on 295 subjects (200 clients, 25 evaluation impostors, and 70 test impostors) during four sessions taken at one month intervals. The database was divided into three sets: a training set, an evaluation set, and a test set. The training set was used to build client models, while the evaluation set was used to estimate thresholds. Finally, the test set was employed to assess system performance.

We compared the performance of the different methods presented in section 3: **a)** *WMS*_14: Warping images onto a mean shape using the same set of 14 “anchor” points employed in [10], **b)** *WMS*_62: Warping images onto a mean shape using the full set of 62 “anchor” points, and **c)** *NFPW*: Normalizing only the subset of pose parameters to generate a frontal mesh. Table 1 shows the False Acceptance Rate (FAR), False Rejection Rate (FRR) and Total Error Rate (TER=FAR+FRR) over the test set for the above mentioned methods. Moreover, the last row from this table presents the baseline results when no pose correction is applied (Baseline).

Table 1. False Acceptance Rate (FAR), False Rejection Rate (FRR) and Total Error Rate (TER) over the test set for different methods.

METHOD	FAR(%)	FRR(%)	TER(%)
*WMS*_14	2.31	5.00	7.31
*WMS*_62	2.64	4.50	7.14
NFPW	2.17	2.75	4.92
Baseline	2.93	4.25	7.18

Table 2. Confidence interval around $\Delta_{HTER} = HTER_A - HTER_B$ for $Z_{\alpha/2} = 1.645$.

METHOD	*WMS*_62	*NFPW*	Baseline
*WMS*_14	$[-1.15\%, 1.32\%]$	$[0.07\%, 2.32\%]$	$[-1.16\%, 1.29\%]$
*WMS*_62		$[0.02\%, 2.20\%]$	$[-1.21\%, 1.17\%]$
NFPW			$[-2.20\%, -0.06\%]$

4.3 Statistical Analysis of the Results

In [14], the authors adapt statistical tests to compute confidence intervals around Half Total Error Rates ($HTER = TER/2$) measures, and to assess whether there exist statistically significant differences between two approaches or not. Given methods A and B with respective performances $HTER_A$ and $HTER_B$, we compute a confidence interval (CI) around $\Delta_{HTER} = HTER_A - HTER_B$. Clearly, if the range of obtained values is symmetric around 0, we can not say the two methods are different. The confidence interval is given by $\Delta_{HTER} \pm \sigma \cdot Z_{\alpha/2}$, where

$$\sigma = \sqrt{\frac{FAR_A(1-FAR_A)+FAR_B(1-FAR_B)}{4 \cdot NI} + \frac{FRR_A(1-FRR_A)+FRR_B(1-FRR_B)}{4 \cdot NC}} \tag{3}$$

and

$$Z_{\alpha/2} = \begin{cases} 1.645 \text{ for a } 90\% \text{ CI} \\ 1.960 \text{ for a } 95\% \text{ CI} \\ 2.576 \text{ for a } 99\% \text{ CI} \end{cases} \tag{4}$$

In equation (3), NC stands for the number of client accesses, while NI stands for the number of imposter trials. For each comparison between the methods presented in table 1, we calculated confidence intervals which are shown in table 2. From both tables we can conclude:

1. Although pose variation is not a major characteristic of the XM2VTS database, it is clear that the use of *NFPW* significantly improved system performance compared to the baseline method.
2. Warping images to a mean shape suffers from the greatest degradation in performance. It is clear that synthesizing face images with *WMS* does seriously distort the "identity" of the warped image, as long as the performances of the baseline algorithm and the two *WMS*'s methods are very similar (robustness to pose provokes subject-specific information supression, leading to no improvement at all). Furthermore, we assess that significant differences are present when comparing *WMS* with *NFPW*, as the confidence intervals do not include 0 in their range of values.
3. There are no statistically significant differences between *WMS_14* and *WMS_62*, as the confidence interval is symmetric around 0.

It was previously stated that *NFPW* was not suitable for holistic feature extraction, but this is not the case of *WMS*. In order to assess the performance of a (baseline) holistic method on WMS faces, we applied eigenfaces [2] and obtained a TER of 16.27%, which is significantly worse than that of the local feature extraction on *WMS* faces, with a confidence interval around Δ_{HTER} of $[2.95\%, 6.01\%]$.

5 Conclusions

We have demonstrated that, if the training set is appropriately chosen, the pose eigenvectors as introduced in [9] are mostly responsible for rigid mesh changes, and do not contain important non-rigid (expression/identity) information that could severely distort the synthetic meshes. Moreover, we have confirmed, with experimental results on the XM2VTS database, the advantages of modifying these pose parameters over warping faces onto a mean shape.

References

1. Beymer, D.J. and Poggio, T., "Face Recognition from One Example View," in *Proceedings ICCV 1995*, 500–507.
2. Turk, M. and Pentland, A., "Eigenfaces for recognition," in *Journal of Cognitive Neuroscience* Vol. 3, 1991, pp. 7286.
3. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C., "Face recognition by Elastic Bunch Graph Matching," in *IEEE Transactions on PAMI*, Vol. 19, No.7, 1997 pp. 775–779.

4. Xiujuan Chai, Shiguang Shan, Xilin Chen, Wen Gao, “Local Linear Regression (LLR) for Pose Invariant Face Recognition,” in *Proceedings of 7th International Conference on AFGR’06*, 631–636.
5. Xiujuan Chai, Laiyun Qing, Shiguang Shan, Xilin Chen, Wen Gao, “Pose Invariant Face Recognition under Arbitrary Illumination based on 3D Face Reconstruction,” in *AVBPA 2005*, NY, USA, 2005, 956–965.
6. Blanz, V. and Vetter, T., “A Morphable model for the synthesis of 3D faces,” in *Proceedings SIGGRAPH 1999*, 187-194.
7. Blanz, V., Grother, P., Phillips, P.J., and Vetter, T., “Face Recognition Based on Frontal Views Generated from Non-Frontal Images,” in *Proceedings IEEE Conference on CVPR 2005*, 454–461.
8. Vetter T., and Poggio T., “Linear Object Classes and Image Synthesis from a Single Example Image,” IEEE PAMI vol. 19, pp. 733–742, 1997
9. González-Jiménez, D. and Alba-Castro, J.L., “Pose Correction and Subject-Specific Features for Face Authentication”, in *ICPR 2006*, Vol. 4.
10. Lanitis, A., Taylor, C.J. and Cootes, T.F.,“Automatic Interpretation and Coding of Face Images Using Flexible Models,”in *IEEE PAMI*, Vol. 19, No. 7, pp. 743–756, 1997.
11. Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. XM2VTSDB: “The extended M2VTS database,” in *Proceedings AVBPA 1999*, 72–77.
12. Luttin, J. and Maître, G. “Evaluation protocol for the extended M2VTS database (XM2VTSDB),” *Technical report RR-21*, IDIAP, 1998.
13. Bookstein, Fred L.: “Principal Warps: Thin-Plate Splines and the Decomposition of Deformations,” in *IEEE PAMI*, Vol. 11, No. 6, April 1989, pp. 567–585.
14. Bengio, S., Mariethoz, J. “A statistical significance test for person authentication”, in *Proceedings Odyssey 2004*, 237–244.
15. Lyons, M.J., Campbell, R., Plante, A., Coleman, M., Kamachi, M., and Akamatsu, S., “The Noh Mask Effect: Vertical Viewpoint Dependence of Facial Expression Perception,” in *Proceedings of the Royal Society of London B 267: 2239-2245 (2000)*.