SYMMETRY-AIDED FRONTAL VIEW SYNTHESIS FOR POSE-ROBUST FACE RECOGNITION

Daniel González-Jiménez and José Luis Alba-Castro

Departamento de Teoría de la Señal y Comunicaciones
Universidad de Vigo, Spain
{danisub,jalba}@gts.tsc.uvigo.es

ABSTRACT

This paper tackles the problem of pose variations in a 2D face recognition scenario. Using a training set of sparse face meshes, we built a Point Distribution Model and identified the parameters which are responsible for controlling the apparent changes in shape due to turning and nodding the head, namely the pose parameters. Given a test image and its associated mesh, the pose parameters are set to typical values of frontal faces, thus obtaining a virtual frontal mesh. Taking advantage of *facial symmetry*, we overcome problems due to self-occlusion and virtual frontal faces are synthesized via Thin Plate Splines-based texture mapping. These corrected images are then fed into a recognition system that makes use of Gabor filtering for feature extraction. The CMU PIE database is used to assess the performance of the proposed method in a closed-set identification scenario where large pose variations are present, achieving state-of-the-art results.

Index Terms— Face Recognition, Image generation, Spline functions

1. INTRODUCTION

This paper addresses one of the major issues within the general face recognition problem: dealing with *pose changes*. 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. Up to now, the most successful algorithms are those which make use of prior knowledge of the class of faces. In [2], Beymer and Poggio extend the earlier attempt presented in [1] (whose main drawback was that images from different view-points were needed for every client). From a single image of a subject and making use of face class information, virtual views facing different poses are synthesized and used in a view-based recognizer. In [6], Blanz and Vetter propose a 3D Morphable Model, where each face can be represented as a linear combination of 3D face examplars. Given an input image, the 3D Morphable Model is fitted, recovering shape and texture parameters following an analysis-by-synthesis scheme. Several approaches make use of the 3D Morphable Model to perform recognition. The main drawback of these methods is the high computational complexity needed to recover image parameters. In [7], Romdhani and Vetter report high recognition rates on the CMU PIE database [15], by means of the 3D Morphable Model and a fitting algorithm that uses linear relations to update the shape and texture parameters, which are then employed for recognition purposes. Blanz et al. also use the 3D Morphable Model in [8] to synthesize frontal faces from non frontal views, which are 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 in each of the regions in which the face is divided, and in [5] where a 3D model is used. In [9], Samaras and Zhang combine the strengths of Morphable models to capture the variability of 3D face shape and a spherical harmonic representation for the illumination. In [10], Gross et al. propose to estimate the eigen light-fields of the subject's head, using them for recognition across pose and illumination changes with tests on the CMU PIE database.

Using a dataset containing sparse face meshes (62 points per image), we built a Point Distribution Model and from the main modes of variation, the parameters responsible for controlling the apparent changes in shape due to turning and nodding the head (so-called pose parameters) were identified, similar to the research by Lanitis et al. [11], where the pose of the face was estimated using those parameters. We propose a method in which the pose parameters are set to typical values of frontal faces, so that a virtual frontal mesh is obtained. Afterwards, taking advantage of facial symmetry, we overcome problems due to self-occlusion and synthesize a virtual frontal face by sampling texture from the original image onto the new mesh, using Thin Plate Splines-based warping. There exist similarities between our method and the works of Blanz et al. [8] and Xiujuan Chai et al. [4], [5] as all of them try to generate frontal images. However, and among other differences, facial symmetry will be taken into account, leading to important improvements in system performance. Once the virtual face is obtained, Gabor responses are extracted and used for comparison. The paper is organized as follows. Next section briefly reviews the concept of Point Distribution Models. In section 3, we introduce the method to cope with pose variations, describing the synthesis of virtual faces and the use of facial symmetry. Section 4 explains feature extraction on corrected images through Gabor filtering. Section 5 shows identification experiments on the CMU PIE database. Finally, conclusions and future research lines are drawn in section 6.

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 coordinates 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 coordinates of the j-th landmark in the i-th training image. Principal Components Analysis is performed to find the most important modes of shape variation. As a consequence, any training shape $\mathbf{X}_i$ can be approximately reconstructed:

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

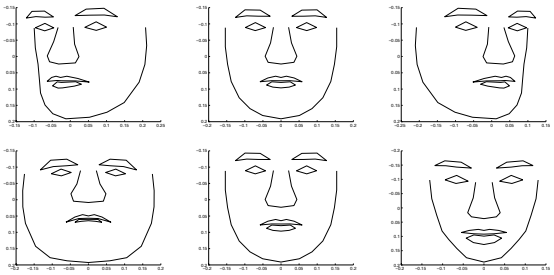

Fig. 1. Effect of varying the pose parameters of the PDM. First row: The parameter responsible for horizontal rotations in depth is varied. Second row: Elevation parameter is modified. The middle column shows the average shape of the training set.

where $\bar{\mathbf{X}}$ stands for the mean shape, $\mathbf{P}$ 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 $\mathbf{b}$ ($\mathbf{b}_k, k = 1, 2, \ldots, t$) weighs the k-th mode of variation. Examining the shapes generated by varying $\mathbf{b}_k$ within suitable limits, we find those parameters responsible for controlling the apparent changes in shape due to turning and nodding the head, as indicated in figure 1. Let $\mathbf{b}^{pose}$ be the set of parameters which accounts for pose variation. Also, since the columns of $\mathbf{P}$ are orthogonal, we have that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, and thus $\mathbf{b} = \mathbf{P}^T\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)$, i.e. given any shape, it is possible to obtain its vector of parameters $\mathbf{b}$ and, in particular, we are able to find its pose (i.e. $\mathbf{b}^{pose}$). We built a 62-point PDM using a set of manually annotated landmarks[1].

When a new image containing a face is presented to the system, the vector of shape parameters that fits the data, $\mathbf{b}$, should be computed automatically. There exist several techniques like [12] to deal with this problem. In this work we have used manual annotations instead, which allows us to test the classification performance alone, without the effect of landmark detection errors.

3. POSE CORRECTION

Given a test image I_{test} with unknown identity and a training image I_{train} of a given client, the system must output a measure of similarity (or dissimilarity) between them. Straightforward texture comparison between I_{test} and I_{train} may not produce desirable results as differences in pose could be quite important. So, in order to deal with these differences, we apply an algorithm that makes use of the pose parameters.

3.1. Normalize to Frontal Pose and Warp (NFPW)

The method we propose aims to synthesize frontal faces from non frontal views. Once the meshes have been fitted to I_{train} and I_{test}, their respective vectors of shape parameters, $\mathbf{b}_{train}$ and $\mathbf{b}_{test}$ are computed, and only the subset of parameters that account for pose variations are fixed to typical values of frontal faces (as the average shape corresponds to a frontal face, we decided to fix pose parameters to zero, i.e. $\mathbf{b}^{pose}_{train} = \mathbf{b}^{pose}_{test} = \vec{0}$), obtaining the modified vectors of shape parameters $\hat{\mathbf{b}}_{train}$ and $\hat{\mathbf{b}}_{test}$. New mesh coordinates are computed using equation (1), and virtual images, $\hat{I}_{train}$ and $\hat{I}_{test}$, must then be synthesized by warping the original faces onto the new shapes.

3.2. Thin Plate Splines texture mapping

For the synthesis of virtual faces, we used a method developed by Bookstein [14], based on thin plate splines. Provided the set of correspondences between the original mesh X and the corrected one $\hat{X}$, the original face I is allowed to be deformed so that the original landmarks are moved to fit the new shape, as it can be seen in figure 2. Thin Plate Splines are a class of non-rigid spline mapping functions $f(x, y)$ with several desirable properties for our application. They are globally smooth, easily computable, separable into affine and non-affine components and contain the least possible non-affine warping component to achieve the mapping. By using two separate thin plate spline functions f_x and f_y which model the displacement of the landmarks in the x and y direction we arrive at a vector-valued function $\mathbf{F} = (f_x, f_y)$ which maps each point of the source image onto a new point in the target image: $(x, y) \rightarrow (f_x(x, y), f_y(x, y))$. This spline defines a global warping of space, and is therefore used to warp the entire source image onto the target mesh.

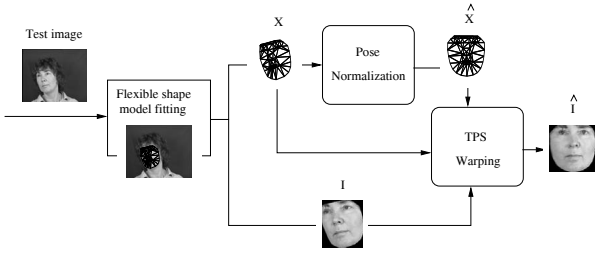

Fig. 2. Block diagram for pose normalization using *NFPW*. TPS stands for Thin Plate Splines.

3.3. Taking advantage of facial symmetry

As explained before, the synthesis of a virtual image is accomplished by sampling texture from the original one. The problem arises when, due to self-occlusion, some face regions become not visible, i.e. texture is not available, and hence the corresponding regions in the pose normalized image do not represent subject's appearance correctly. In order to overcome this drawback, we take advantage of the vertical symmetry of the face. For a horizontal rotation in depth of the head and once the mesh has been fitted, the parameter controlling the azimuth angle indicates whether the face is showing mostly its right or its left side. Whenever a frontal face is synthesized from a non-frontal view, we warp the original image and its mirror version onto the pose-corrected frontal mesh and then blend the two virtual images, using simple masks that weigh the two sides of the face appropriately (according to the current rotation -left or right- of the head), as it can be seen in figure 3.

4. FEATURE EXTRACTION

The recognition engine is based on Gabor filtering. Gabor filters are biologically motivated convolution kernels in the shape of plane waves restricted by a Gaussian envelope. Our system uses a set of 40 Gabor filters with the same configuration as in [3]. The region surrounding a pixel in the image is encoded by the convolution of the image patch with these filters, and the set of responses

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

is called a jet, $\mathcal{J}$. So, a jet is a vector with 40 coefficients, and it provides information about an specific region of the image. At each of the nodes of the pose-normalized mesh, a Gabor jet is extracted and stored for comparison. Given two images to be compared, say I_1 and I_2 with node coordinates $\mathcal{P} = \{\vec{p_1}, \vec{p_2}, \ldots, \vec{p_N}\}$ and $\mathcal{Q} = \{\vec{q_1}, \vec{q_2}, \ldots, \vec{q_N}\}$, their respective sets of jets are computed: $\{\mathcal{J}_{\vec{p_i}}\}_{i=1,\ldots N}$ and $\{\mathcal{J}_{\vec{q_i}}\}_{i=1,\ldots N}$. Finally, the score between the two images is given by:

$$\mathcal{S} = f_N \{< \mathcal{J}_{\vec{p_i}}, \mathcal{J}_{\vec{q_i}} >\}_{i=1,\ldots,N} \quad (2)$$

where $< \mathcal{J}_{\vec{p_i}}, \mathcal{J}_{\vec{q_i}} >$ represents the normalized dot product between correspondent jets, but taking into account that only the moduli of jet coefficients are used. In equation (2), f_N stands for a generic combination rule of the N dot products (for this work, we chose $f_N \equiv$ mean).

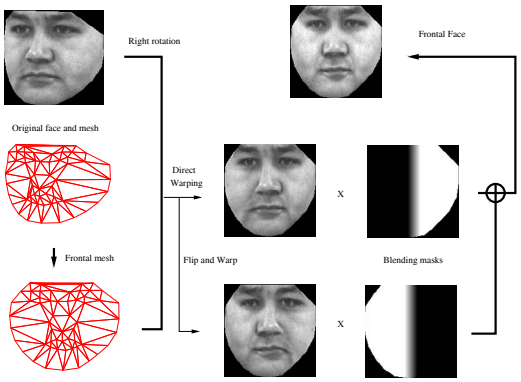

Fig. 3. Taking advantage of facial symmetry for image synthesis

5. FACE IDENTIFICATION ON THE CMU PIE DATABASE

In [16], the *NFPW* method (without symmetry) was tested on the XM2VTS database [13], achieving good authentication results. However, the XM2VTS is mainly a frontal face database and hence, it is not suitable to assess the performance of the method under large pose variations. Moreover, we want to test whether there exist improvements when facial symmetry is taken into account, and to compare our technique with other pose-robust face recognition methods. For these reasons, a more appropriate database was used: the CMU PIE (Pose, Illumination and Expresion) database[15].

Fig. 4. Images taken from all cameras of the CMU PIE database for subject 04006.

5.1. Database and experimental setup

The CMU PIE database consists of face images from 68 subjects recorded under different combinations of poses and illuminations. Figure 4 shows the images taken for subject 04006 from all cameras under neutral illumination. In this paper, we use a subset of the database, namely the images taken from cameras 11, 29, 27, 05 and 37 with neutral illumination. All of them (a total of 68×5 images) were manually annotated with 62 points per image. We distinguish between gallery (training) and probe (testing) images. The gallery contains images of known individuals, which are used to build templates, and the probe set contains images of subjects with unknown identity, that must be compared against the gallery. A *closed universe model* is used to assess system performance, meaning that every subject in the probe set is also present in the gallery. We did not restrict ourselves to work with frontal faces as gallery. Instead, the performance of the system was computed for all possible $(gallery, probe)$ combinations.

5.2. Results

Table 1 shows the baseline results when no pose correction is applied. The average recognition rate is 68.38%. When the *NFPW* method is used, the correct identification rate increases to 78.46% (table 2). However, results are poor for completely different viewpoints. As it can be seen from table 3, performance is clearly improved if facial symmetry is taken into account, leading to an average recognition rate of 87.50%.

Table 1. Identification rates (%) on the CMU PIE database: No pose correction

Probe Pose / Gallery Pose	c11	c29	c27	c05	c37
c11	–	94.12	63.24	48.53	25.00
c29	97.06	–	92.65	66.18	39.71
c27	79.41	91.18	–	92.65	51.47
c05	67.65	80.88	98.53	–	88.23
c37	23.53	38.24	51.47	77.94	–

Table 2. Identification rates (%) on the CMU PIE database: *NFPW* without facial symmetry

Probe Pose / Gallery Pose	c11	c29	c27	c05	c37
c11	–	97.06	77.94	55.88	19.12
c29	98.53	–	98.53	73.53	44.12
c27	89.71	98.53	–	100	85.29
c05	67.65	85.29	100	–	97.06
c37	36.76	54.41	91.18	98.53	–

Table 3. Identification rates (%) on the CMU PIE database: *NFPW* plus facial symmetry

Probe Pose / Gallery Pose	c11	c29	c27	c05	c37
c11	–	97.06	88.23	80.88	66.18
c29	98.53	–	95.59	85.29	67.65
c27	94.12	98.53	–	100	89.71
c05	80.88	80.88	98.53	–	100
c37	70.59	64.71	94.12	98.53	–

Table 4 presents the recognition rates achieved with the use of the 3D morphable model [6] and the LiST fitting algorithm [7]. The

average recognition rate is 88.45%. As we can see, *NFPW* plus symmetry achieves comparable performance over the set of considered poses.In [4] and [5], only frontal images were used as gallery. The recognition rates for these two methods are shown in the first two rows of table 5, with averages of 85.5% and 94.87% respectively. For the same gallery, *NFPW* plus symmetry obtains 95.59% of correct recognition rate. The remaining rows from this table present the results achieved with two different versions of the eigen light-field (ELF) approach [10]: the 3-point ELF (3 points - eyes and mouth - are used to warp the face image) and the Complex ELF (where a set of manually annotated points is used for the normalization). Due to the use of manual landmarks, the last one is specially suitable for comparison with our method. We can see that *NFPW* plus symmetry outperforms the Complex ELF in the range of considered poses[2] (88.79% compared to 82.5% correct recognition rate).

Table 4. Identification rates (%) on the CMU PIE database: 3D Morphable Model with LiST fitting algorithm [7]

Probe Pose Gallery Pose	c11	c29	c27	c05	c37
c11	–	94	94	74	65
c29	96	–	96	78	68
c27	93	97	–	99	94
c05	88	90	99	–	93
c37	82	82	93	94	–

Table 5. Identification rates (%) on the CMU PIE database: Other results

Method	Probe Pose Gallery Pose	c11	c29	c27	c05	c37
[4]	c27	76.5	95.6	–	91.2	77.9
[5]	c27	95	97	–	98	89
ELF 3-point	c27	76	85	–	89	75
ELF 3-point	c37	73	66	80	80	–
ELF Complex	c27	76	90	–	94	90
ELF Complex	c37	70	69	83	88	–

6. CONCLUSIONS

Based on a subset of the modes of a Point Distribution Model, namely the pose parameters, we have proposed a method for pose correction that makes use of the facial symmetry of the face. The identification experiments on the CMU PIE database show that the proposed method achieves comparable results to the 3D morphable model and outperforms other approaches in the set of considered poses. Moreover, it is confirmed that taking advantage of facial symmetry does clearly improve system performance. Currently, the major drawback of our method is that it relies on manually annotated landmarks. Hence, the next step will be to test the pose correction stage with automatic fitting. Moreover, we must extend the algorithm in order to cope with near-profile views (where almost half of the landmarks are not visible) by defining correspondences between landmarks (and views) from different angles. Another possible improvement could be to learn a view-based weight function, so that depending on the current pose of the face, some regions get more importance than others in the computation of the final similarity score.

[2] At the moment of writing this paper, only numerical results with poses c27 and c37 as gallery could be obtained from the authors of [10].

7. REFERENCES

[1] Beymer, D.J., "Face Recognition under varying pose," in *Proceedings CVPR 1994*, 756–761.

[2] Beymer, D.J. and Poggio, T., "Face Recognition from One Example View," in *Proceedings ICCV 1995*, 500–507.

[3] Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C., "Face recognition by Elastic Bunch Graph Matching," in *IEEE Trans. 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*, Southampton, UK, 2006, 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 *Proceeding 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] Romdhani, S., Blanz, V., and Vetter, T., "Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions," in *Proceedings ECCV 2002*, Copenhagen, Denmark, 2002, 3–19.

[8] Blanz, V., Grother, P., Phillips, P.J., and Vetter, T., "Face Recognition Based on Frontal Views Generated from Non-Frontal Images," in *Proceedings CVPR 2005*, 454–461.

[9] Zhang, L., Samaras, D., "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," in *IEEE Trans. on PAMI,* Vol. 28, No. 3, March 2006, pp. 351 - 363.

[10] Gross, R., Matthews, I. and Baker, S., "Appearance-Based Face Recognition and Light-Fields," in *IEEE Trans. on PAMI*, Vol. 26, No. 4, April 2004, pp. 449–465.

[11] Lanitis, A., Taylor, C.J. and Cootes, T.F.,"Automatic Interpretation and Coding of Face Images Using Flexible Models,"in *IEEE Trans. of PAMI*, Vol. 19, No. 7, pp. 743–756, 1997.

[12] Cootes, T., Edwards, G., and Taylor, C., "Active Appearance Models," in *IEEE Trans. on PAMI*, Vol. 23, No. 6, 2001, pp. 681–685.

[13] Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. XM2VTSDB: "The extended M2VTS database," in *Proceedings AVBPA 1999*, 72–77.

[14] Bookstein, Fred L.: "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," in *IEEE Trans. on PAMI*, Vol. 11, No. 6, April 1989, pp. 567–585.

[15] Sim, T., Baker, S., and Bsat, M.,"The CMU Pose, Illumination, and Expression Database," in *IEEE Trans. on PAMI*, Vol. 25, No. 12, December, 2003, pp. 1615 – 1618.

[16] González-Jiménez, D. and Alba-Castro, J.L., "Pose Correction and Subject-Specific Features for Face Authentication", in *Proceedings ICPR 2006*, Vol. 4.