# 3D assisted face recognition via progressive pose estimation

Wuming Zhang, Di Huang, Dimitris Samaras, Jean-Marie Morvan, Yunhong Wang, Liming Chen

# 3D ASSISTED FACE RECOGNITION VIA PROGRESSIVE POSE ESTIMATION

*Wuming Zhang[1], Di Huang[2], Dimitris Samaras[3], Jean-Marie Morvan[4], Yunhong Wang[2], Liming Chen[1]*

[1]MI Department, LIRIS Laboratory, CNRS 5205, Ecole Centrale de Lyon, 69134, Lyon, France
[2]School of Computer Science and Engineering, Beihang University, 100091, Beijing, China
[3]Image Analysis Lab, Computer Science Department, Stony Brook University, 11794, New York, USA
[4]Institut Camille Jordan, UMR 5208 du CNRS, Université Lyon 1, 69622, Lyon, France

## ABSTRACT

Most existing pose-independent Face Recognition (FR) techniques take advantage of 3D model to guarantee the naturalness while normalizing or simulating pose variations. Two nontrivial problems to be tackled are accurate measurement of pose parameters and computational efficiency. In this paper, we introduce an effective and efficient approach to estimate human head pose, which fundamentally ameliorates the performance of 3D aided FR systems. The proposed method works in a progressive way: firstly, a random forest (RF) is constructed utilizing synthesized images derived from 3D models; secondly, the classification result obtained by applying well-trained RF on a probe image is considered as the preliminary pose estimation; finally, this initial pose is transferred to shape-based 3D morphable model (3DMM) aiming at definitive pose normalization. Using such a method, similarity scores between frontal view gallery set and pose-normalized probe set can be computed to predict the identity. Experimental results achieved on the UHDB dataset outperform the ones so far reported. Additionally, it is much less time-consuming than prevailing 3DMM based approaches.

***Index Terms***— asymmetric face recognition, pose estimation, 3D morphable model, random forest

## 1. INTRODUCTION

In the light of tremendous progress that has been made in traditional FR where pose variations are strictly prohibited or controlled, it is natural to extend the research interest to unconstrained environment. Compared with conventional 2D techniques [1, 2, 3], 3D methods, which could account for pose variations, show complete dominance using comprehensive facial information [4, 5]. However, the expensive cost in acquisition, registration and calculation of 3D data makes it difficult to be widely used in FR systems.

As a trade-off between 2D and 3D techniques, 2D/3D asymmetric FR has recently become an attractive topic. The main motivation of this scenario is to integrate superiorities of both 2D and 3D based methods and avoid their drawbacks, such as unilateral enrollment of 3D data in gallery with 2D texture images in probe. In this way more robustness could be gained by applying 3D models on 2D image analysis, meanwhile acquisition of 3D data is not required in verification stage so as to largely reduce the computational cost.

To deal with such an issue, a few attempts have been made [6]. Blanz and Vetter [7] build a statistical model by a set of training data and densely fit it to a given facial image for matching, but it generally requires a long convergence process. Riccio and Dugelay [8] established a correspondence between the 3D gallery face and the 2D probe using geometric invariants on the face; Toderici et al. [9] also located some pre-defined key landmarks (eye corners and nose tip) on the facial images in different poses, and then roughly align them to a frontal 3D model for matching. Nevertheless, they both assumed that accurate localization in multi-view facial images was fulfilled, which turns out to be another tough topic. Zhang et al. proposed an asymmetric 3D-2D FR approach [10] which performs a 2D-2D matching by synthesizing 2D images from 3D models towards the same pose as probe samples, where a preprocessing pipeline for illumination normalization and pose correction as well as Oriented Gradient Maps (OGMs) based facial representation [11] are adopted. This approach was further compared and associated with work of Zhao et al. [12] as a benchmarking asymmetric 3D-2D FR system on the UHDB face database [13]. Unfortunately they both suffer from high computational cost owing to considerable complexity of pose synthesis and their pose estimation processes still lack satisfactory accuracy.

In this work, we are especially concerned with an efficient pose-independent face recognition approach. It deals with accurate pose reconstruction by introducing a progressive pose estimation processing while greatly improving computational efficiency. Experiments carried out on the UHDB dataset demonstrate our prominence compared with prevailing techniques. Fig. 1 shows an illustrative overview of the proposed method, in which we organize our system and the remainder as well by concatenating three main parts: random forest based pose estimation (Section 2), 3D morphable model based pose normalization (Section 3) and LBP-based matching (Section 4). Finally the experimental results and conclusions are presented in Section 5 and Section 6 respectively.

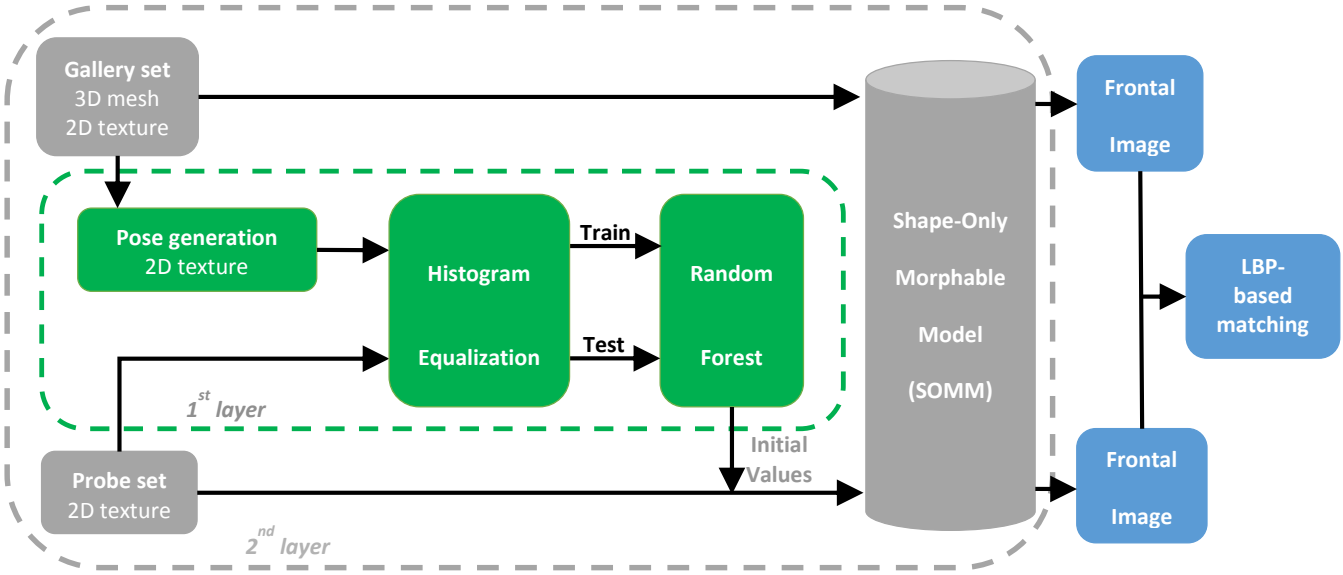


**Fig. 1.** Framework of our proposed method.

## 2. RANDOM FOREST BASED POSE ESTIMATION

As a strong classifier in machine learning, Random Forest (RF) inherits and enhances the classification capacity of a single decision tree while overcoming its inconvenience of over-fitting, therefore it is commonly used in applications towards pose estimation and feature point extraction [14, 15].

In our preprocessing stage, 3D models in the gallery set are firstly rotated to specific poses and then projected onto 2D plane to obtain a multi-pose 2D face database with accurate ground truth of pose value. As shown in Fig. 2, these images are regarded as training data of RF after a rough delighting processing. Moreover, aiming at decreasing the impact of over-fitting, only a randomly selected subset of images will be adopted for each tree, and it is thus sufficient to present the proceeding of a single tree's training.

Instead of directly utilizing the grayscale value of a randomly selected region as feature value in previous work [14], which leads to heavy computational cost and potential error, we extract LBP (Local Binary Patterns) features of each image to construct an SVM (Support Vector Machine) based classifier for each split node. The detailed algorithm is presented in Algorithm 1.

**Algorithm 1.** The outline of our RF training

1. Select a subset of images at random as training data for a new tree $T_i$, each image $I_j$ is associated with its pose vector $\theta_j = \{\theta_{yaw}, \theta_{pitch}, \theta_{roll}\}$ where the three subscripts stand for Euler angles representing conditions of 3D space rotation.
2. Judge whether current node is a leaf node or a split one. If any of these conditions is satisfied: 1) maximum depth of tree is reached, 2) remaining images are less than preselected threshold, we define this node a leaf node and jump to step 5; or we go to step 3.
3. Divide images into two classes. Histograms of LBP features for each image in current node are calculated as features to be classified by SVM. Certain binary tests in the form of (1)
$$\omega h_j - \rho < 0 \quad (1)$$
would be carried out by training certain SVMs. Note that $h_j$ is the LBP histogram of image $I_j$, ω and $\rho$ are trained weight parameter and bias offset of each SVM obtained by randomly labeling images. The final SVM coefficients are chosen by minimizing the corresponding differential entropy:
$$E = \omega_l log\big(det(\sigma_l)\big) + \omega_r log\big(det(\sigma_r)\big) \quad (2)$$
where $\omega_l$ and $\omega_r$ stand for quantity weight of the left and right image subset compared with their parent node, $\sigma_l$ and $\sigma_r$ represent the covariance matrix of the left and right subset. These SVM coefficients are stored in this node as attributes.
4. Perform steps 2, 3 iteratively until a leaf node is found.
5. Record the leaf node by storing the mean vector of all pose vectors arrived as an annotated attribute and the trace of covariance matrix.
6. Perform steps 2 to 5 iteratively until all nodes of tree $T_i$ are traversed.

After training a certain number of trees, a random forest is constructed which actually performs a projection from continuous pose variations onto discrete pose candidates in order to transform the estimation problem to a classification problem. Upon largely enhancing classification performance by gathering considerable SVM classifiers, RF is capable to tackle this problem.

When an unseen image enters, we simply release it to our well-trained RF and cluster all pose candidates obtained after discarding ones which are not informative, i.e. owning a too large trace.

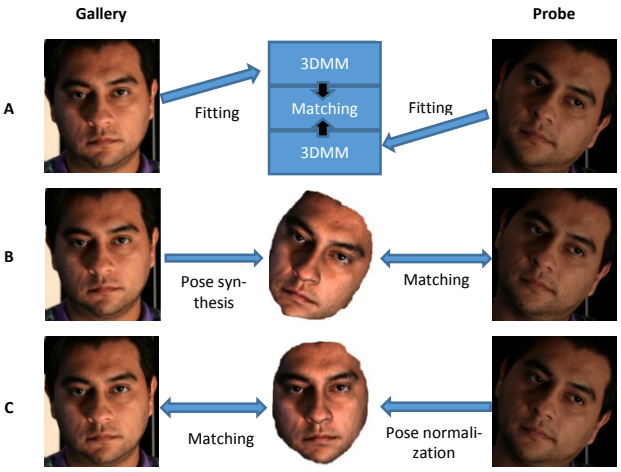


**Fig. 3.** Three categories towards model based pose-invariant face recognition: A) Fitting, B) Synthesis, C) Normalization.

## 3. 3DMM BASED POSE NORMALIZATION

Despite of RF's powerful classification capacity, it is still greatly restricted when it comes to a situation requesting high accuracy because of the intrinsic errors caused by training data's diversity and cluster effect. Another formidable tool is subsequently introduced to perfect our system.

Widely employed in 3D recovery and synthesis, 3D morphable model has always been a significant parametric modeling technique. Assuming that human face could be represented by linearly combining a train of 'prototype' face models, 3DMM offers a 3D face generation processing:

$$S_{model} = \bar{S} + \sum_{i=1}^{m-1} \alpha_i s_i, T_{model} = \bar{T} + \sum_{i=1}^{m-1} \beta_i t_i \quad (3)$$

where $S_{model}$ and $T_{model}$ are generated shape and texture of a new face; $\bar{S}$ and $\bar{T}$ are mean shape and texture of existing face models; $s_i$ and $t_i$ are eigenvectors of all models after PCA transformations; $\alpha_i$ and $\beta_i$ stand for respective weights of $i^{th}$eigenvector of shape and texture.

To date, 3DMM based FR methods could be categorized into three classes by model's usage rules:

1. Pose fitting [7, 16]: The most simple and primary method directly comparing shape and texture parameters between gallery and probe images obtained by fitting 3D models to both of them;
2. Pose synthesis [17, 18]: Estimate the pose value of probe image by the aim of generating photo-realistic images with similar pose by fitting 3D models to gallery images, and pose-invariant matching is hence achieved;
3. Pose normalization [19, 20]: Instead of rotating frontal gallery image to a specified pose, both gallery and probe images are normalized to frontal pose and then regular face recognition could be performed.

The categorization could be intuitively illustrated and compared in Fig. 2. For pose fitting methods, despite of their simplicity and intuition, the performance depends largely on their fitting accuracy. Pose synthesis methods are easier to implement and could avoid texture missing; however they require high computational cost since all gallery images are processed for each probe image. Compared with them, pose normalization tends to be more appropriate for its merits of quickness and accuracy. Nevertheless, most existing methods still adopt the same framework which aims at searching for parameters $\alpha_i$ and $\beta_i$ simultaneously or iteratively by minimizing the pixel-wise difference between the raw input image and the recovered image. Although they may finally acquire an accurate fitting result, they suffer from embarrassing computational cost and pose initialization for optimization procedure still remains a challenge.

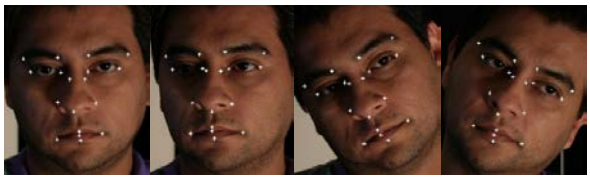

**Fig. 2.** Annotation of 18 feature points in face images under 4 different pose variations in which the first one is the frontal view image in the gallery set.

Starting from these problems, we propose a novel 3DMM based framework by introducing a reasonable pose initialization and avoiding pixel-wise comparison in order to gain a quite considerable acceleration while preserving the performance. We utilize 100 3D models in USF face dataset [21] as 'prototype' models and assume that 18 feature points are manually labeled as prior knowledge for both USF models and all images in gallery and probe. Their positions are depicted in Fig. 3. For clarity purposes, the proceeding of proposed technique is shown in Algorithm 2.

**Algorithm 2.** The outline of proposed pose normalization.

1. Calculate mean shape $\overline{s_F^{3D}}$, PCA eigenvectors $s_F^{3D} = [s_{F1}^{3D}, s_{F2}^{3D}, \dots, s_{Fn}^{3D}]$ and their corresponding eigenvalues $\sigma^{3D} = [\sigma_1^{3D}, \sigma_2^{3D}, \dots, \sigma_n^{3D}]$ of all 3D feature points labeled on the USF face models where n refers to number of eigenvectors selected in the final synthesis.
2. Minimize the cost function $F$. Instead of processing the image pixel by pixel as previous work does, we merely retain the feature points and discard texture information. The new cost function is shown below:

$$F = arg \min_{f,\varphi,\gamma,\theta,\alpha,t^{2D}} \left\| \left( fPR\left(\overline{s_F^{3D}} + \sum_{i=1}^{n} \alpha_i s_{Fi}^{3D}\right) + t^{2D}\right) - s_F^{2D} \right\|^2 + \omega \sum_{i=1}^{n} \left\| \frac{\alpha_i}{\sigma_i^{3D}} \right\|^2 \quad (4)$$

   Note that $f$ refers to scale parameter; $P$ stands for orthogonal projection; $R$ indicates 3D space rotation formed by three Euler angles $\varphi, \gamma, \theta$; and $s_F^{2D}$ are feature points of input image. Particularly, we avail ourselves of the pose values estimated by RF in Sec. 3 for initialization so that the optimization procedure develops in a relatively correct range. Furthermore, a variance-normalized constraint weighted by a pre-defined value ω is added to avoid abnormal shape parameters.
3. Synthesize normalized face shape. Since all parameters

are recovered, we can synthesize the normalized face shape of an image by simply generalizing the shape parameters to whole point clouds:

$$s_{new}^{3D} = \overline{s^{3D}} + \sum_{i=1}^{n} \alpha_i s_i^{3D} \quad (5)$$

The proposed method benefits from RF based pose estimation and much less computational cost. Notwithstanding the imperfect shape recovery accuracy shown in Fig. 4, we truly realize an effective and speedy pose normalization approach which inspires a novel pose-invariant FR system.

## 4. LBP-BASED MATCHING

Once the frontal shape model of an input face image is reconstructed, texture values could be mapped to new model by respecting the correspondence between 2D and 3D. After that, we could carry out another projection of this textured 3D model from 3D space onto 2D plane in order to gain a frontal-view 2D face image, several examples are depicted in Fig. 4. Manifold matching methods could afterwards be available for classical face recognition. In this work we calculate the chi-square distance between LBP features [22] of gallery and probe images to evaluate the recognition rate.

## 5. EXPERIMENTAL RESULTS

The proposed method aims at achieving a 2D/3D asymmetric face recognition, and the dataset applied for evaluating its performance should hence satisfy two conditions: 1) 3D models are provided in the gallery set, 2) 2D images with pose variations are included in the probe set. In respect of the fact that most existing public face datasets, such as FRGC and CMU-PIE, could hardly meet these requirements all at once, a novel dataset is adopted in our work which is offered by University of Houston, known as the UHDB face dataset.

The UHDB contains 23 3D meshes and their corresponding textures of 23 subjects in the gallery set and 1692 2D textures in the probe set. 6 illumination conditions and 12 pose conditions are covered in the probe set, considering that only pose impact is handled in our work, we merely take into account all pose conditions under neutral lighting, including 23*12 = 276 images. Several examples of one subject and their recovery results are illustrated in Fig. 4. We can infer from the figure that our method is capable to successfully recover the shape information of given image and re-map the texture onto frontal face shape to obtain a pose-normalized image. Although identity inference based on shape fitting lacks accuracy and there might occur texture missing for large pose variations (as shown in last column of Fig. 4), we could still perceive that a high-quality pose normalization is achieved and FR based on pose-normalized image is feasible.

As shown in Table. 1, the proposed method is compared with two previous work on UHDB [12]. Furthermore, another comparison experiment without RF based pose estimation is carried out as well for validating the efficiency of progressive pose estimation for pose-invariant face recognition. In the meanwhile, their computation complexities are also comparatively measured and analyzed through the average processing time for each image.

**Fig. 4.** Examples in UHDB. First row is raw image; second and third rows illustrate respectively the recovered face shape with and without estimated pose; last row shows the textured image after pose normalization. In particular, first column is gallery and others are probes.

It is clear to find out that our method outperforms the state-of-the-art 2D/3D FR approaches which report the performance on this dataset and is much more computationally efficient. In addition, it is worth mentioning that RF based pose estimation helps improve the performance of the proposed method and slightly shorten the processing time.

**Table 1.** Verification result and average processing time

| Method | Rank-1 RR | Processing Time |
|---|---|---|
| LIRIS [12] | 0.802 | 61.2s |
| UR2D [12] | 0.852 | - |
| Proposed method without RF | 0.862 | 12.5s |
| **Proposed method** | **0.888** | **10.8s** |

## 6. CONCLUSIONS

In this work, a novel 2D/3D asymmetric FR system is proposed which takes advantage of random forest based pose estimation as a preprocessing pipeline to improve the pose-invariant FR performance while limiting the use of 3D data so that a low computational cost is obtained. Compared with existing 2D/3D asymmetric methods, the proposed approach is much faster and less dependent of 3D model fitting accuracy. Furthermore, the experimental results on the UHDB prove the predominant verification power and computational efficiency of our method under unconstraint environment.

## 7. ACKNOWLEDGEMENT

This work was in part supported by the French Research Agency, Agence Nationale de Recherche (ANR), through the Biofence project, under the grant ANR-13-INSE-0004-02.

## 8. REFERENCES


[1] Z. Yi, L. Gu, and H.J. Zhang, "Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference." *Computer Vision and Pattern Recognition*. Proceedings IEEE Computer Society Conference on. Vol. 1, 2003

[2] J. Wu, and M.M. Trivedi, "A two-stage head pose estimation framework and evaluation." *Pattern Recognition,* 41.3: pp. 1138-1158, 2008

[3] X. Liu, H. Lu, and D. Zhang, "Head pose estimation based on manifold embedding and distance metric learning." *Asian Conference on Computer Vision*. Springer Berlin Heidelberg, pp. 61-70, 2010

[4] K.W. Bowyer, K. Chang and P.J. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition" *Computer Vision and Image Understanding*, vol. 101, pp. 1-15, 2006

[5] A. Scheenstra, A. Ruifrok, and R.C. Veltkamp, "A survey of 3D face recognition methods." *Audio-and Video-Based Biometric Person Authentication*. Springer Berlin Heidelberg, 2005

[6] A. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey." *Pattern Recognition Letters*, 28(14): pp. 1885-1906, 2007

[7] V. Blanz, T. Vetter, "Face recognition based on fitting a 3D morphable model." *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 25(9): pp. 1063-1074, 2003

[8] D. Riccio and J.L. Dugelay, "Geometric invariants for 2D/3D face recognition." *Pattern Recognition Letters*, 28(14): pp. 1907-1914, 2007

[9] G. Toderici, G. Passalis, S. Zafeiriou, et al. "Bidirectional relighting for 3D-aided 2D face recognition." *Computer Vision and Pattern Recognition*, IEEE Conference on, pp. 2721-2728, 2010

[10] W. Zhang, D. Huang, Y. Wang and L. Chen, "3D-Aided Face Recognition across Pose Variations." *Chinese Conference on Biometric Recognition*, 2012

[11] D. Huang, M. Ardabilian, Y. Wang and L. Chen, "Oriented Gradient Maps based Automatic Asymmetric 3D-2D Face Recognition." *International Conference on Biometrics*, 2012

[12] X. Zhao, W. Zhang, G. Evangelopoulos, et al. "Benchmarking asymmetric 3D-2D face recognition systems." *Automatic Face and Gesture Recognition (FG)*, 10th IEEE International Conference and Workshops on. IEEE, pp. 1-8, 2013

[13] UH Computational Biomedicine Lab: UHDB11 face database *http://cbl.uh.edu/URxD/datasets/*, 2009

[14] G. Fanelli, J. Gall, and G.L. Van, "Real time head pose estimation with random regression forests." *Computer Vision and Pattern Recognition*, IEEE Conference on. IEEE, pp. 617-624, 2011

[15] M. Dantone, J. Gall, G. Fanelli, et al. "Real-time facial feature detection using conditional regression forests." *Computer Vision and Pattern Recognition*, IEEE Conference on. IEEE, pp. 2578-2585, 2012

[16] O. Aldrian, and W.A.P. Smith, "Inverse rendering of faces with a 3D morphable model." *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 35(5): pp. 1080-1093, 2013

[17] K. Niinuma, H. Han, and A.K. Jain, "Automatic multi-view face recognition via 3D model based pose regularization." *Biometrics: Theory, Applications and Systems (BTAS)*, IEEE Sixth International Conference on. IEEE, pp. 1-8, 2013

[18] E. Mostafa, A. Ali, N. Alajlan, et al. "Pose invariant approach for face recognition at distance." *European Conference on Computer Vision*, Springer Berlin Heidelberg, pp. 15-28, 2012

[19] A. Asthana, T.K. Marks, M.J. Jones, et al. "Fully automatic pose-invariant face recognition via 3d pose normalization." *International Conference on Computer Vision*, IEEE International Conference on. IEEE, pp. 937-944, 2011

[20] S. Li, X. Liu, X. Chai, et al. "Morphable displacement field based image matching for face recognition across pose." *European Conference on Computer Vision,* Springer Berlin Heidelberg, pp. 102-115, 2012

[21] K. Bowyer, and S. Sarkar, USF DARPA Human-ID 3D Face Database, *http://www.csee.usf.edu/sarkar.* University of South Florida, 2001

[22] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition." *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 28(12): pp. 2037-2041, 2006