3D Aided Face Recognition across Pose Variations

Wuming Zhang[1,2], Di Huang[2], Yunhong Wang[2], and Liming Chen[1]

[1]MI Department, LIRIS, CNRS 5205, Ecole Centrale de Lyon, 69134, Lyon, France.
[2]IRIP, School of Computer Science and Engineering, Beihang Univ., 100191, Beijing, China.
wuming.zhang@hotmail.com; dhuang@buaa.edu.cn.

Abstract. Recently, 3D aided face recognition, concentrating on improving performance of 2D techniques via 3D data, has received increasing attention due to its wide application potential in real condition. In this paper, we present a novel 3D aided face recognition method that can deal with the probe images in different viewpoints. It first estimates the face pose based on the Random Regression Forest, and then rotates the 3D face models in the gallery set to that of the probe pose to generate specific gallery sample for matching, which largely reduces the influence of head pose variations. Experiments are carried out on a subset of the FRGC v1.0 database, and the achieved performance clearly highlights the effectiveness of the proposed method.

Keywords: Face recognition; pose estimation; random regression forests; LBP.

1   Introduction

Due to its scientific challenges and application potential, machine-based face recognition has always been an active topic in the field of computer vision and pattern recognition [1]. Compared with other biometrics, e.g. fingerprint and iris, recognition based on the face is more in accord with the nature of human; moreover, it can be achieved without physical contact which endues it with an extra important advantage.

In the past several decades, 2D image based face recognition has rapidly developed and a great number of milestone techniques have been proposed and studied, such as PCA [2], LDA [3], ICA [4], FDA [5], EBGM [6], LBP [7], SIFT [8], etc. However, despite the great progress made in this domain, 2D face images do not remain reliable when affected by changes of lighting, pose and expression. Recently, 3D face recognition has emerged as a major solution to deal with the unsolved issues in 2D domain, i.e. lighting and pose variations [9, 10]. Unfortunately, 3D face recognition approaches are currently limited by their high acquisition, registration and computation cost.

More recently, 3D aided face recognition has attracted increasing interests, since it is expected to limit the use of 3D data where it really helps to improve the face recognition accuracy [11], i.e. aiming to handle the problem of illumination and pose in 2D area. For example, based on a generic 3D face model, re-lighting or de-lighting techniques [12] are adopted to reduce the influences caused by lighting variations. While, in this study, we address the problem of pose changes. In order to deal with such an issue, a few attempts have been made. Blanz and Vetter [13] build a statistical model using a set of training data (also named as 3D morphable model) and densely fit it to a given facial image for matching, but it generally requires a long convergence process.

Toderici et al. [14] first locate some pre-defined key landmarks (eye corners and nose tip etc.) on face images in different poses, and then roughly align them to a frontal 3D model for the recognition step. Nevertheless, to achieve accurate localization in multi-view facial images involves in another tough topic.

In this paper, we propose a novel method for 3D aided face recognition, aiming to improve the tolerance of 2D face recognition against pose variations. It first estimates the face pose status based on Random Regression Forest, and then rotates the 3D face models in the gallery set to the one of the probe pose achieved previously to generate specific gallery samples for the matching step. In contrast to the existing methods that only process probe data for matching, the proposed approach operates on the enrolled data, and thanks to the Random Regression Forest algorithm, pose variations of probe faces are estimated and largely reduced before matching. Experiments are carried out on a subset of the FRGC v1.0 database, and performance clearly highlights the effectiveness of the proposed method.

The rest part of the paper is organized as follows: an overview of the proposed approach is presented in Section 2. Section 3 describes the process of pose estimation in detail, and Section 4 introduces LBP based face recognition. Experimental results are shown and analyzed in Section 5, and Section 6 concludes the paper.

2 Framework Overview and Data Preparation

An entire framework is shown by two flowcharts (in Fig. 1 and Fig. 3), demonstrating the training stage and the test stage respectively.
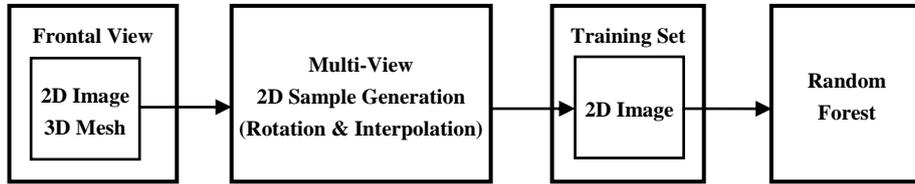

Fig. 1. Framework of training stage.

At the training stage, a collection of textured frontal 3D face models, each of which consists of one 3D mesh and its texture counterpart, is required. We select some models from the FRGC v1.0 dataset, and based on our previous work [15], the nose tip of each face can be localized automatically. For the lack of 2D facial images in arbitrary viewpoint, we have to generate samples for training as follows:

1. **Loading**. First of all, we read a face model in the FRGC v1.0 dataset, including its 3D coordinates in real space and its texture information. In addition, we record the position of its nose tip calculated previously.
2. **Rotation.** The rotation of human head runs according to three degrees-of-freedom (DoF), and it can thus be described by three angles, i.e. *yaw*, *pitch* and *roll*, namely egocentric rotation angles [16] as shown in Fig. 2. We then rotate the face model to another viewpoint according to these three angles using rotation matrix defined as:

$$R_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta_x & -sin\theta_x \\ 0 & sin\theta_x & cos\theta_x \end{bmatrix} = exp\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \theta_x \\ 0 & -\theta_x & 0 \end{bmatrix} \quad (1)$$

$$R_y(\theta_y) = \begin{bmatrix} cos\theta_y & 0 & sin\theta_y \\ 0 & 1 & 0 \\ -sin\theta_y & 0 & cos\theta_y \end{bmatrix} = exp\begin{bmatrix} 0 & 0 & -\theta_y \\ 0 & 0 & 0 \\ \theta_y & 0 & 0 \end{bmatrix} \quad (2)$$

$$R_z(\theta_z) = \begin{bmatrix} cos\theta_z & -sin\theta_z & 0 \\ sin\theta_z & cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} = exp\begin{bmatrix} 0 & \theta_z & 0 \\ -\theta_z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3)$$

where θ_x is pitch angle; θ_y is yaw angle; θ_z is roll angle. In our system, we generate diverse angles, with yaw angle from -90⁰ to +90⁰ and pitch angle from -45⁰ to +45⁰, aiming to cover all possible poses. Some examples are shown in Fig. 3.
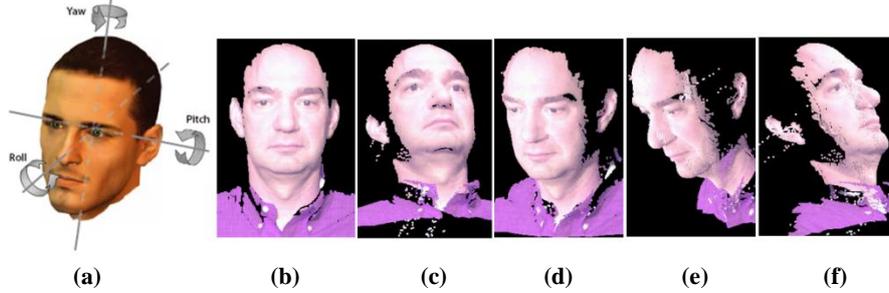

Fig. 2. (a) The three DoF of human head can be described by the egocentric rotation angles, i.e. *pitch*, *roll*, and *yaw* [16]; (b) The original frontal model whose name in FRGC is "02463d456"; (c)-(f) pairs of yaw angle and pitch angle: (-30⁰, -15⁰), (30⁰, 15⁰), (-60⁰, -30⁰), and (60⁰, 30⁰).

3. **Interpolation.** As in Fig. 2, there occurs inevitably some change concerning distribution of points on the facial area, leading to holes on the produced texture map. In order to avoid this imperfection, interpolation techniques are adopted.
4. **Cropping.** For the purpose of reducing calculation burden, we simply crop the face area according to the nose tip position.
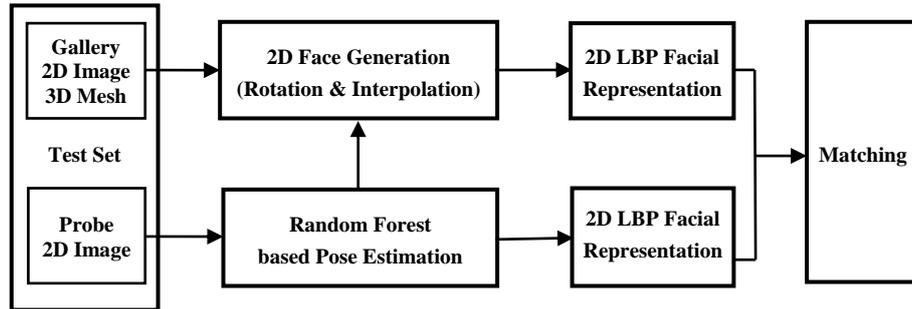

Fig. 3. Framework of test stage.

We hence obtain the 2D images with different views to make up of the training set. With the help of random regression forests, we are able to generate a set of regression trees which should be regarded as a classifier for pose estimation.

At the test stage, each probe facial image will be traversed through the random forest pre-trained to estimate the pose status. As its pose is determined, all 3D face models in the gallery set are rotated to generate a facial image under this pose. The identification operates between the original probe face image and the new produced gallery one by comparing their similarity in the LBP feature space.

3 Pose Estimation

3.1 Problem Statement

Head pose variations can incur serious change in the appearance of human faces, and thus introduce a quite difficult problem in the domain of 2D face recognition. Without a proper solution to handle pose changes, even the most sophisticated face recognition systems probably fail. In the framework of the proposed 3D aided face recognition, to a 2D probe facial image, we estimate its pose status. Compared with 3D model based methods, pose estimation based on 2D facial images is a much greater challenge since 3D data are continuously distributed and offer distinctive geometry features.

In literature, there exist several studies on 2D image based pose estimation, which can be roughly categorized into four streams. **Geometry based methods** [17] [18] use the pre-detected facial feature points, such as inner and outer corners of eyes and nose tip, to calculate head pose directly on the basis of the prior knowledge of their relative configuration. Its advantages are their simplicity and rapidness, only a few geometric cues are required and no training process is used; however, their performance highly relies on the feature point detection accuracy. **Model based methods** [19] [20] seek a non-rigid face model which conforms to the facial structure so that it can be fit by the face image, and then head pose estimation can be achieved. These models cost rather low computation and are invariant to head localization error, yet they depend largely on the facial features and they are robust neither to illumination variation nor to far-field head pose estimation. **Manifold embedding methods** [21] [22] hunt for optimal low-dimensional manifolds which can describe the intrinsic pose variations in order to embed the new images into these manifolds and thus estimate their pose. These methods successfully reduce the dimensionality and the computation cost, but it remains a problem separating pose variations from appearance variations, such as identity, scale and illumination. **Learning based methods** [23] [24] aim to map the input image to discrete or continuous head poses by using machine learning tools, for instance, SVM and Neural Networks. It has attracted increasing attention over the last few years for its powerful classification capacity and robustness to appearance variation. The disadvantage of these methods lies in the problem of overfitting, which means the performance is vulnerable to noise in the training data. A proposed remedy to this problem is the application of random regression forests [25] which avoid the influence of noise data by introducing a set of decision trees. It has proved to achieve outstanding performance in 3D head pose estimation [26] [27]. In this paper we further investigate its effectiveness on 2D pose estimation.

3.2 Random Forest

Derived from classification and regression trees, random regression forests have successfully optimized the problem of overfitting by introducing a series of trees at random while keeping the powerful capacity in handling large datasets.

In our work, all training images with different poses are divided into patches. Each patch is annotated with a vector $\theta = \{\theta_x, \theta_y, \theta_{yaw}, \theta_{pitch}\}$, here θ_x and θ_y represent an offset vector pointing from the center of the patch to nose tip, and $\theta_{yaw}, \theta_{pitch}$ record respectively the yaw and pitch angle of image which the patch belongs to.

At the beginning of the training step, a number of patches are randomly selected as input data for each tree, and at each split node there will be a set of binary tests defined with similar style:

$$|R_1|^{-1} \sum_{q \in R_1} RGB^c(q) - |R_2|^{-1} \sum_{q \in R_2} RGB^c(q) > \tau \tag{4}$$

where R_1 and R_2 are two rectangles randomly selected inside the patch, c is one out of the three channels R, G or B, $RGB^c(q)$ represents the texture value of pixel q in channel c, finally τ is a random threshold.

Secondly, it is essential to determine an appropriate test for each node which could maximize the distinctiveness of the actual node. Here, we adopt the concept of information gain defined as the difference between the differential entropy of these patches at the parent node and one of patches at the children nodes as value of distinctiveness:

$$IG = H(S) - (\omega_L H(S_L) + \omega_R H(S_R)) \tag{5}$$

where H is the abbreviation of differential entropy, S, S_L, S_R represent the set of patches at the parent node, at the left child node and at the right child node respectively, ω_L and ω_R are weight for each child node defined as the ratio between number of patches at the child node and the one at the parent node.

Finally, besides the split nodes, there exist also the nodes which store the result of training, namely leaf nodes. A node should be regarded as a leaf node if at least one of the two conditions is achieved: number of patches arriving at this node is smaller than pre-defined threshold; or the tree has attained its maximum depth. Once a leaf node is created, it will be annotated with mean and covariance of patches reaching it.

In this way, we are capable to obtain a collection of trees randomly generated, for each tree a test at every split node is recorded and so are mean and covariance at every leaf node, these values will serve to our test process.

3.3 Pose Estimation

Given an unseen 2D image of a face, patches with the same scale of those in the training set are extracted and sent to each tree well trained.

For each patch, it could generate the same number of leaves as the number of trees used, all the leaves will be gathered and the ones with a covariance larger than threshold will be firstly abandoned because they are much less informative.

The rest of leaves have the honor to be called a "vote" for our last test; they will be clustered to discard the noise and select the most centralized area on the vector plan.

Finally, we sum up all the leaves remaining active and calculate their mean vector which indicates the final estimation of test image's pose. Fig. 4 shows some results.
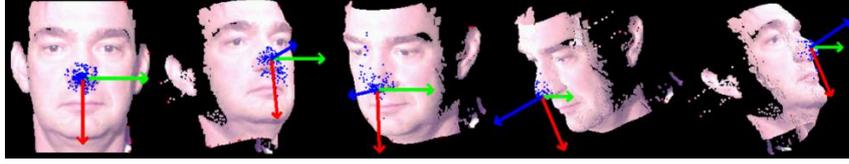
Fig. 4. Some results of pose estimation (x axis in green; y axis in red; z axis in blue).

4 LBP based Multi-view Face Recognition

As one of the most distinguished texture descriptors, the LBP operator [28] has been widely used in numerous applications. It has turned out to be a highly discriminative operator and its core advantages, i.e. its invariance to monotonic illumination changes and computational efficiency, make it reasonable for undertaking the responsibility of representing faces. This powerful operator labels each pixel of an image by thresholding its 3x3-neighbourhood with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor.

Afterwards, in order to fit in textures with different sizes, the LBP operator was extended to the neighborhoods of different sizes [29]. Using circular neighborhoods and the bilinear interpolation technique, the pixel values allow any radius and number of pixels within the neighborhood. We use the notation *(P, R)* for neighborhoods which means P sampling points on a circle of radius of R.

Furthermore, it has been shown that among 2^p possible binary patterns there exist certain patterns which contain more information than the others, and we thus come up with another extension: uniform patterns. This concept proposed by Ojala et al. represents the patterns that contain at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular.

After pose estimation, we rotate each 3D face model in the gallery set and generate its texture map I_g in the pose of the probe I_p for face matching. The LBP face image is separated firstly in m regions from each of which we extract a histogram, and we then combine them to construct final histograms encoding both the local texture and spatial information. At last, the Chi square distance is exploited to decide the similarity between the final vectors H_g and H_p of gallery and probe face.

5 Experimental Results

The dataset for technique evaluation is based on FRGC v1.0. 50 3D face models from different individuals are randomly selected for training. While another 100 out of the rest subjects, each of which possesses more than two face models, are used for testing. For the 100 subjects used in the test step, their first models make up of the gallery set and their second models are regarded as probes. For the 50 training samples as well as the 100 probes, we rotate them with an interval 15 °between -90 °and 90 °in yaw angle (totally 13 poses) and between -45 °and 45 °in pitch angle (totally 7 poses), leading to a considerable capacity of $150 \times 13 \times 7 = 14k$ 2D facial images of various poses.

In our experiments on pose estimation, there are three main parameters that influence estimation performance: i.e. number of trees, threshold of angle error and threshold of nose error. It should be noted that the pose estimation accuracy in this study indicates ratio between the number of samples that are correctly estimated and the total number with respect to a pre-defined threshold angle error or nose error as [26]. The result of pose estimation is depicted in Fig.4.

From Figure 4 (a), we can infer that following the increase of number of trees, both nose error and angle error become smaller which highlights the superiority of random regression forests compared with the standard decision tree. Considering that increase the number of trees produces much more calculation amount, we have set tree number at 10 to achieve a compromise between accuracy and computation cost.

According to the results shown in Figure 4(b) and 4(c), we can achieve remarkable estimation performance even if the thresholds of angle and nose error are limited in a very narrow range. As we increase the threshold, our performance is improved.
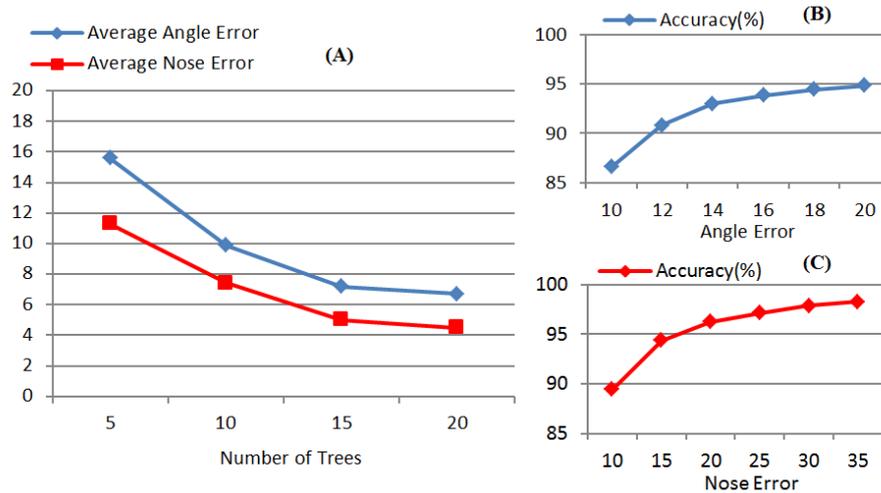

Fig. 5. (a) average error of angle and nose with respect to number of trees; (b) estimation accuracy with respect to angle error threshold; (c) estimation accuracy with respect to the nose error threshold in *mm*.

In face recognition, we rotate the 3D face models in the gallery set to the estimated pose of the probe to generate 2D face images for matching. We calculate recognition performance based on different thresholds of nose error and angle error, and the result is displayed in Table. Considering that the profiles do not contain enough information for identification, we discard the faces -90 °and 90 °in yaw angle, composing a probe set of 7,700 facial images.

From Table 1, we can see that although slightly influenced by the accuracy of head pose estimation, we are still able to achieve very high recognition rates (around 90%). It is worthy of noting that the LBP descriptor used in this experiment is a quite basic one (its radius is set at 1 and the number of neighbors is set at 8), the effectiveness of the propsoed is hence emphasized.

Table 1. Recognition rates using LBP and influence of pose estimation accuracy.

Nose Error Threshold	Angle Error Threshold	Pose Estimation Accuracy	Rank-one Recognition Rate
10	10	0.8660	0.8441
15	12	0.9082	0.8688
20	14	0.9297	0.8958
25	16	0.9383	0.8992
30	18	0.9441	0.9000
35	20	0.9483	0.9014

6   Conclusions

This paper presents a novel 3D aided face recognition method which owns the capacity to handle the 2D probes in different viewpoints. It first estimates the pose status by introducing the approach of Random Regression Forest, and then rotates the 3D face models in the gallery set to that of the probe pose to generate specific gallery samples for the final LBP based matching. The proposed method largely reduces the influence caused by pose variations. Experiments are carried out on a subset randomly extracted from the FRGC v1.0 database, and the achieved performance in both pose estimation and face recognition clearly highlights the effectiveness of the proposed method.

Acknowledgment

References

1. Zhao, W., Chellappa, R., Rosenfeld, A. and Phillips, P. J.: Face recognition: a literature survey. ACM Computing Surveys, vol. 35, pp. 399--458 (2003)
2. Turk, M. and Pentland, A.: Eigenfaces for recognition. J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71--86 (1991)
3. Yu, H. and Yang, J.: A direct LDA algorithm for high-dimensional data-with application to face recognition. PR, vol. 34, no. 10, pp. 2067--2070 (2000)
4. Comon, P.: Independent component analysis, a new concept? Signal Processing, Elsevier, vol. 36, no. 3, pp. 287--314 (1994)
5. Huang, J., Yuen, P., Chen, W. and Lai, J.: Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations. IEEE TSMC-B, vol. 37, no. 4, pp. 847--862 (2007)
6. Wiskott, L. H., Fellous, J. M., Kuiger, N. and von der Malsburg, C.: Face recognition by elastic bunch graph matching. In: IEEE TPAMI, vol. 19, pp. 775--779 (1997)

7. Ahonen, T., Hadid, A. and Pietikäinen, M.: Face recognition with local binary patterns. In: ECCV (2004)
8. Lowe, D. G.: Distinctive image features from scale-invariant keypoints, In: IJCV, vol. 60, no. 2, pp. 91–110 (2004)
9. Bowyer, K. W., Chang, K. and Flynn, P. J.: A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. In: CVIU, vol. 101, pp. 1--15 (2006)
10. Scheenstra, A., Ruifrok, A. and Veltkamp, R. C.: A survey of 3d face recognition methods, In: AVBPA (2005)
11. Huang, D., Ardabilian, M., Wang, Y. and Chen, L.: Oriented gradient maps based automatic asymmetric 3D-2D face recognition. In: ICB (2012)
12. Zhang, L, Wang, S, and Samaras, D: Face synthesis and recognition under arbitrary unknown lighting using a spherical harmonic basis morphable model. In: CVPR (2005)
13. Blanz, V. and Vetter, T.: Face recognition based on fitting a 3D morphable model. In: IEEE TPAMI, vol. 25, no. 9, pp. 1063--1074 (2003)
14. Toderici, G., Zafeiriou, S., Tzimiropoulos, G., Petrou, M., Theoharis, T. and Kakadiaris, I. A.: Bidirectional relighting for 3D-aided 2D face recognition. In: CVPR (2010)
15. Szeptycki, P., Ardabilian, M. and Chen, L.: A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking. In: BTAS (2009)
16. Murphy-Chutorian, E. and Trivedi, M: Head pose estimation in computer vision: a survey. In: IEEE TPAMI, pp. 442--449 (2008)
17. Beymer, D.: Face recognition under varying pose. In Proc. CVPR (1994)
18. Liang, X., Tong, W.: Face pose estimation using near-infrared images. In: CSNT (2012)
19. Krüger, K. N., Pötzsch, M. and Malsburg, C. v. d.: Determination of face position and pose with a learned representation based on labeled graphs. In: IVC, vol. 15, no. 8, pp. 665--673 (1997)
20. Cootes, T. F., Edwards, G. and Taylor, C. J.: Active Appearance Models. In: IEEE TPAMI (2001)
21. Wu, J and Trivedi, M.: A two-stage head pose estimation framework and evaluation. In: PR, vol. 41, no. 3, pp. 1138--1158 (2008)
22. Liu, X., Lu, H. and Zhang, D.: Head pose estimation based on manifold embedding and distance metric learning. In: ACCV (2009)
23. Murphy-Chutorian, E., Doshi, A. and Trivedi, M. M.: Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In: Proc. IEEE Int. Transp. Syst. Conf., pp. 709--714 (2007)
24. Osadchy, M., Cun, Y. L., and Miller, M. L.: Synergistic face detection and pose estimation with energy-based models. In: JMLR. Vol. 8, pp. 1197--1215 (2007)
25. Breiman, L., Friedman, J., Olshen, R. and Stone, C.: Classification and regression trees. Wadsworth and Brooks, Monterey, CA (1984)
26. Fanelli, G., Gall, J. and Van Gool, L: Real time head pose estimation with random regression forests. In: CVPR (2011)
27. Tang, Y., Sun, Z. and Tan, T.: Real-time head pose estimation using random regression forests. In: CCBR (2011)
28. Ojala, T., Pietikäinen, M. and Maenpaa, T.: Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. In: IEEE TPAMI, vol. 24, no. 7, pp. 971--987 (2002)
29. Huang, D., Ardabilian, M., Wang, Y. and Chen, L.: Asymmetric 3D/2D face recognition based on LBP facial representation and canonical correlation analysis. In: ICIP (2009)