Pose Independent Face Recognition by Localizing Local Binary Patterns via Deformation Components

Iacopo Masi, Claudio Ferrari, Alberto Del Bimbo
MICC, University of Florence
Italy
http://www.micc.unifi.it/vim

Gérard Medioni
USC IRIS Lab, University of Southern California
Los Angeles, USA
http://iris.usc.edu/USC-Computer-Vision.html

***Abstract*—In this paper we address the problem of pose independent face recognition with a gallery set containing one frontal face image per enrolled subject while the probe set is composed by just a face image undergoing pose variations. The approach uses a set of aligned 3D models to learn deformation components using a 3D Morphable Model (3DMM). This further allows fitting a 3DMM efficiently on an image using a Ridge regression solution, regularized on the face space estimated via PCA. Then the approach describes each profile face by computing Local Binary Pattern (LBP) histograms localized on each deformed vertex, projected on a rendered frontal view. In the experimental result we evaluate the proposed method on the CMU Multi-PIE to assess face recognition algorithm across pose. We show how our process leads to higher performance than regular baselines reporting high recognition rate considering a range of facial poses in the probe set, up to $\pm 45°$. Finally we remark that our approach can handle continuous pose variations and it is comparable with recent state-of-the-art approaches.**

I. Introduction

Face recognition has been considered a key problem in computer vision for decades. Even if frontal face recognition seems a issue nearly solved if addressed in constrained conditions, the general problem is still open for faces captured in the wild. A "face in the wild" typically means that the subject is captured under challenging conditions such as aging, pose, expression and illumination variations. Considering these challenges, the one that mostly affects recognition performance is pose variation. In fact it is demonstrated [1] that, when the face is in a non-frontal view, face recognition performance drops drastically because discriminative descriptors, such as Local Binary Pattern (LBP) and Gabor filters, suffer from misalignment issues. In addition to these, face recognition across pose also leads to another subtle problem which is the ambiguity of landmarks caused by the self-occlusion of the face: when the face assumes a profile pose, landmark detectors respond with the same number of landmarks with respect to the ones detected on a frontal face, but with different semantic meaning (if the right part of the jaw is occluded, the detector will return a landmark on a cheek instead of on the jaw). This problem also becomes harder if we consider that in the gallery we have just one exemplar image to describe each subject.

In this paper we address the problem of pose invariant face recognition with a gallery set containing *one* frontal face image per enrolled subject, while the probe set is composed by just a face image undergoing pose variations. This scenario, defined as is, is an ill-posed problem considering the gap between the kind of information present in the gallery and the one available in the probe. Considering these issues, the main contributions are the following:

- our approach solves the landmark ambiguity reported previously [2], [3] by proposing a face pose estimation that selects visible and stable landmarks.
- Similarly to [3], a 3D Morphable Model (3DMM) is efficiently fit on a image using a Ridge regression solution that globally preserves the face shape while locally minimizing the landmark reprojection error.
- By exploiting the previous contribution, instead of computing LBP on a uniform grid [4], we localize the LBP histograms on the deformed vertices. This gives more precision to the method, obtaining features vectors of the same dimension irrespective of the image size.

The paper is organized as follows: in Sect. II we review the most recent papers about face recognition across pose and in Sect. III we describe our approach to learn a 3DMM. Then in Sect. IV we address the problem of fitting this model by minimizing the reprojection error on detected landmarks, while preserving the face shape. Once the model is fit on a generic non frontal image, in Sect. V we design our face recognition scheme by computing LBP histograms on the deformed vertices. Finally in Sect. VI we perform extensive evaluation experiments respect to regular baselines such as rectification with 2D similarity, 3D average model and other recents approaches [5], [6], [2], [7], [8].

II. Related work

Usually face recognition performance is satisfactory for near frontal faces [9] but drops drastically when the face is not showing a pose similar to the one in the gallery. Currently computer vision community has proposed several innovative methods to recognize faces across pose.
Authors in [10] propose to extend the patch-based approach of Kanade and Yamada [11] by adding a data-driven extension in which it is not only modeled how a face patch varies in appearance, but also how it is deformed geometrically as the viewpoint varies. In their case the deformation is encoded locally in an affine matrix that warps the patch without preserving the face integrity.
Instead of just using a local affine warp, the authors in [1] are the first to introduce a 3D generic face model to improve the patch-based alignment problem, rather than relying on 3D cylindrical or 3D ellipsoid model. They compare faces in different viewpoints using a similarity score that is measured by correlations in a media subspace between different poses

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.766

on patch level. The media subspace is learned by Canonical Correlation Analysis (CCA) in order to maximize the intra-individual correlations. A powerful tool to model non-rigid transformation, firstly proposed in computer graphics literature [12] and then applied to face recognition [13], is the 3DMM which has been introduced by Blanz and Vetter to model the deformation of a face. The model is learned analyzing the principal components on statistics of aligned 3D faces. In particular, in [13], Blanz and Vetter propose to recognize faces by fitting a 3DMM in the gallery and in the probe and using the retrieved coefficients as discriminative feature vector. In contrast to 3DMM, recently, authors in [14], [15] proposed an efficient way to estimate a 3D model from a single frontal image using their Generic Elastic Model (GEM). The GEM assumption is that the depth variation is not containing enough information if we consider a human face and the GEM model considers just a deformation on the XY plane as sufficient to obtain quite realistic 3D models. The method has been further improved considering diverse average values of depth per ethnic group [15]. Recently, 3D data has been used also in matching renderings of 3D faces with 2D face imagery in the wild accounting for small pose variations [16]. Among all these papers, very recently, the attention has been moving on trying to normalize the pose of the profile face to a canonical frontal view. The paper in [2] is the first paper that reports the problem of landmark ambiguity. The approach synthesizes a frontal view from a profile one, exploiting a collection of set of manually labeled landmarks. They manually label one set of landmarks per pose to deal with self-occlusion. Face are normalized using a weak perspective pose estimation method and through a refining algorithm that extracts the boundary of the face. Boundary extraction is prone to fail when the background is not uniform, condition that typically occurs in video-surveillance imagery. Normalized faces are finally recognized using LGBP (Local Gabor Binary Pattern). Inspired by 3DMM, authors in [7] encode the pose variation of a test image in a linear combination of displacement fields, that they call Morphable Displacement Field. The approach is demonstrated to be robust and needs just eye-based alignment to process an image but the displacement field optimization is run for each image in the gallery, which is a strategy that does not scale for large galleries. Recently, in contrast to face normalization to a frontal view, authors in [3] tried also to modify a bank of Gabor Filters by localizing the filters in a precise manner optimizing a 3DMM on a probe image. In [8], Sharma *at al.* propose the Discriminant Multiple Coupled Latent Subspace framework. Similar to [1], the approach finds sets of projection directions for different poses such that the projected images of the same subject are maximally correlated in the latent space. Discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors due to an incorrect pose estimate. Finally, also data driven approaches [17] contributed to face recognition across pose (tackling also issues like expression and illumination changes).

III. Learning 3D Deformation Components via 3DMM

The authors in [13] showed how to build a Gaussian model from a set of 3D aligned faces considering both the shape and the texture by using Principal Component Analysis (PCA) to obtain the principal components. We apply this approach on a dataset of 200 aligned 3D untextured faces taken from a commercial software that can produce virtual human faces. Considering that the faces have a small number of vertices, we augment the mesh dimension performing a two-pass of loop-subdivision, that is able to maintain the alignment for each vertex. We stack all the linearized vertices in a matrix $\mathbf{S}$ where each row corresponds to a subject. We then proceed to label manually a set of indexes $\mathbf{I}$ on the average model $\mathbf{m} = \frac{\mathbf{1}}{\mathbf{S}} \sum_{\mathbf{i}}^{\mathbf{S}} \mathbf{S_i}$ that represents 3D reference landmarks and S is the number of models.

PCA factories $\mathbf{S}$ as follows:

$$\mathbf{S} = \mathbf{W}\ \mathbf{C}, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{S \times K}$ represents the new dataset expressed in the new subspace and $\mathbf{C} \in \mathbb{R}^{K \times 3N}$ are the K components that correspond to the eigenvectors with $K \leq S - 1$. The benefit of using PCA is that each eigenvector has an eigenvalue which is a scalar, quantifying the amount of variance in that direction. In general the vector of eigenvalues is defined as $\boldsymbol{\sigma} = [\sigma_1 \dots \sigma_K]$ from the largest to the smallest one. This vector is important because gives a way to estimate of the probability density within a face space as $p(\boldsymbol{\alpha}) \sim \mathcal{N}\big(\mathbf{0}, \mathrm{diag}(\boldsymbol{\sigma}^2)\big)$.

Considering that N is very large, instead of estimating PCA using eigendecomposition of the covariance matrix $\mathbf{S}\,\mathbf{S}^\top$, we rather use a more efficient way by performing Singular Value Decomposition (SVD) on $\mathbf{S}$. This procedure returns the matrix $\mathbf{C}$ and the corresponding eigenvalues as $\boldsymbol{\sigma}$.

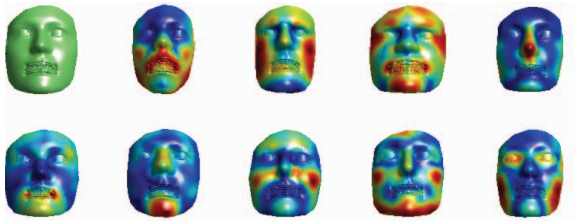


Fig. 1. Deformation Components: the superimposed heat map represents the amount of deformation on the face. The average surface is shown in the left-most box.

In Fig. 1 are shown the first nine components estimated using the previous method. Once the deformation components are learned, similarly to [13], we can easily synthesize a new face from:

$$\mathbf{S}' = \mathbf{m} + \sum_{k=1}^{K} \alpha_k\ \mathbf{C}_k, \tag{2}$$

where

$$p(\boldsymbol{\alpha}) = p(\alpha_1, \dots, \alpha_K) \sim \exp\Big(\sum_{k=1}^{K} (\alpha_k / \sigma_k^2)\Big). \tag{3}$$

IV. Fitting a Regularized 3DMM

In order to fit a 3DMM given a test image, we need firstly to get an estimate of the 3D pose in the image (rigid transformation) and then to fit the non-rigid transformation to the face (face model deformation). The cost function that

we use it is simply the reprojection error of landmarks: as argued by [13], the use of this function could cause overfitting in the deformation process, leading to a surface which is not even close to a face[1]. To overcome this, we propose to use a regularization similar to [3]: we formally address this problem as a Ridge regression in which the energy of 3DMM coefficients is proportionally limited by the prior given by the eigenvalues learned in Sec. III. We proceed to explain firstly how the pose is estimated and then how the model is deformed.

A. 3D Face Pose Estimation

In order to deal with face pose variation, it is necessary to establish correspondences between the labeled landmarks $\mathbf{m}(\mathbf{I})$ in the generic model and the face framed in the image. To this end, we employ the landmark detector [18] that provides good recall in the face detection task and decent precision in the localization. We get an estimate of the pose using an affine camera model that maps each vertex of the model to the image. Under the affine camera model, the relation between the annotated points on the model $\mathbf{L} \doteq \mathbf{m}(\mathbf{I}) \in \mathbb{R}^{3\times|\mathbf{I}|}$ and the detected points $\mathbf{l} \in \mathbb{R}^{2\times|\mathbf{I}|}$ is the following:

$$\mathbf{l} = \mathbf{A}\ \mathbf{L} + \mathbf{t} \tag{4}$$

where $\mathbf{A}$ contains the affine camera parameters and $\mathbf{t} \in \mathbb{R}^{2\times|\mathbf{I}|}$ is a translation on the image. To recover these parameters, firstly we subtract the mean from each points set, then we recover the 2×3 affine matrix in a least square sense as $\mathbf{A} = \mathbf{l}\cdot\mathbf{L}^{\dagger}$, where $\mathbf{L}^{\dagger}$ is the pseudo-inverse matrix of $\mathbf{L}$. Secondly, we estimate the translation as $t = \mathbf{l} - \mathbf{A}\ \mathbf{L}$. Furthermore, the matrix $\mathbf{A}$ can be decomposed with QR decomposition in two matrices: a matrix $\mathbf{S} \in \mathbb{R}^{2\times3}$ that expresses the scale parameters along with the shear and another one, $\mathbf{R} \in \mathbb{R}^{3\times3}$, that contains the 3D rotation parameters of the model with respect to the image. The final affine camera models is thus defined as:

$$\mathbf{l} = \mathbf{S}\ \mathbf{R}\ \mathbf{L} + \mathbf{t}. \tag{5}$$

Considering Eq. (5), it is possible to get an estimate of the pose $\mathbf{P}$ as $[\mathbf{S}\ \mathbf{R}, \mathbf{t}]$ and to map each vertex of the generic model on the image.

B. Landmark Ambiguity Resolution

One of the problem reported in literature is the ambiguity in the landmark locations detected on an image w.r.t. the 3D reference landmarks in the model. The ambiguity does not hold if we consider a frontal face, but when the face undergoes large pose variations, the locations of detected landmarks become unstable especially if we consider fiducial landmarks on the boundaries like the jaws. This issue was reported in [2] and [3]: while the authors of [2] use a look-up table to access different 3D reference landmarks for each pose, the authors of [3] do not use at all the boundary landmarks.

In our approach we overcome this issue in this way: given a test image, we automatically select stable 3D landmarks from the ones present in the labeled set $\mathbf{I}$. As the face undergoes self-occlusion caused by the pose, some 3D reference landmarks will not be visible. Our approach automatically selects the visible landmarks to use as reference using an Hidden Point Removal Operator (HPR) [19], and refines the pose using just a subset of indices $\mathbf{I_v} \subseteq \mathbf{I}$. The algorithm is reported in Alg. 1. This process gives a way to select stable landmarks and handles continuous pose variation and this information of visibility will be used in the recognition stage to select from the gallery the visible part of the face. We use all the landmarks for the frontal gallery faces, omitting the boundary ones at recognition time.

[1]This happens considering that there are a lot of surfaces that can project on the image minimizing the reprojection error.

Algorithm 1: Landmark Ambiguity Resolution

Input: 3D landmarks $\mathbf{L}$, detected landmarks $\mathbf{l}$
Output: $\mathbf{S}$, $\mathbf{R}$, $\mathbf{t}$, $\mathbf{I_v} \subseteq \mathbf{I}$

1. Estimate pose $\mathbf{A}$ with Eq. (4) using all the indices $\mathbf{I}$.
2. Decompose $\mathbf{A}$ in order to get rotation matrix $\mathbf{R}$.
3. Estimate visibile landmarks as $\mathbf{I_v} = \text{HPR}(\mathbf{m}(\mathbf{I}), \mathbf{R})$.
4. Re-Estimate the pose $\mathbf{A}$ with Eq. (4) using $\mathbf{I_v}$.
5. Decompose $\mathbf{A}$ and return $\mathbf{S}$, $\mathbf{R}$, $\mathbf{t}$, $\mathbf{I_v} \subseteq \mathbf{I}$.

C. Optimization

Given a test image we want to find the rigid transformation in term of face pose $\mathbf{P}$ and a non-rigid transformation in term of deformation components $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$. Formally we want to optimize the following:

$$\min_{\mathbf{P},\boldsymbol{\alpha}} \left\| \mathbf{l} - \mathbf{P}\Big(\mathbf{m}(\mathbf{I_v}) + \sum_{k=1}^{K} \alpha_k \mathbf{C_k}(\mathbf{I_v})\Big) \right\|_2 + \lambda\ \|\boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\alpha}\|_2 \tag{6}$$

where $\boldsymbol{\sigma}$ is defined as in Sect. III, $\mathbf{I_v}$ expresses the indices of the visible landmarks, $(\cdot)$ means the element-wise multiplication and λ is a scaling regularization parameter and has been set to 25. This values is set accordingly to the magnitude of eigenvectors.

We solve this problem by alternating between pose estimation and model coefficient estimation. We proceed to estimate the pose as report in Sect. IV-A and then given the estimated pose $\mathbf{P}$, we solve for the coefficients and the problem becomes:

$$\min_{\boldsymbol{\alpha}} \left\| \mathbf{l} - \mathbf{P}\mathbf{m}(\mathbf{I_v}) - \sum_{k=1}^{K} \mathbf{P}\alpha_k \mathbf{C_k}(\mathbf{I_v}) \right\|_2 + \lambda\ \|\boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\alpha}\|_2. \tag{7}$$

By defining $\mathbf{X} \doteq \mathbf{l} - \mathbf{P}\ \mathbf{m}(\mathbf{I_v})$, $\mathbf{Y} \doteq \mathbf{P}\ \mathbf{C}(\mathbf{I_v})$ we get:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \Big\| \underbrace{\mathbf{l} - \mathbf{P}\ \mathbf{m}(\mathbf{I_v})}_{X} - \sum_{k=1}^{K} \alpha_k \underbrace{\mathbf{P}\ \mathbf{C_k}(\mathbf{I_v})}_{Y} \Big\|_2 + \lambda\ \|\boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\alpha}\|_2 = \\ = \min_{\boldsymbol{\alpha}} \left\| \mathbf{X} - \sum_{k=1}^{K} \alpha_k \mathbf{Y} \right\|_2 + \lambda\ \|\boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\alpha}\|_2, \end{aligned} \tag{8}$$

and thus we can rewrite the system by linearizing the landmarks without changing the meaning, casting the problem as a regularized least squares one:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{X} - \boldsymbol{\alpha}\mathbf{Y}\|_2 + \lambda\ \|\boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\alpha}\|_2 \tag{9}$$

which is analytically solved as a Ridge regression where each component is weighted using the inverse of σ_k taken from $\boldsymbol{\sigma}$. The non-rigid coefficients are retrieved as:

$$\boldsymbol{\alpha} = (\mathbf{Y}^T\mathbf{Y} + \lambda \text{ diag}(\boldsymbol{\sigma}^{-1}))^{-1}\mathbf{Y}^T\mathbf{X}. \quad (10)$$

It is important to notice that the number of unknowns in the system is the number of components K while the constraints are the detected landmarks $2N$. So it is important that $K \leq 2N$ to maintain the problem not underdetermined. We use $K = 29$ (the number of deformation components) and the algorithm in Alg. 1 selects at most $N = 51$ landmarks. Fig. 2 shows the geometrical meaning of the optimization process along with the recovered shapes with and without regularization. Fig. 2(b) shows just a landmark for the sake of clarity: the projected deformation components in blue indicate the directions on which deform the model in order to minimize the reprojection error.

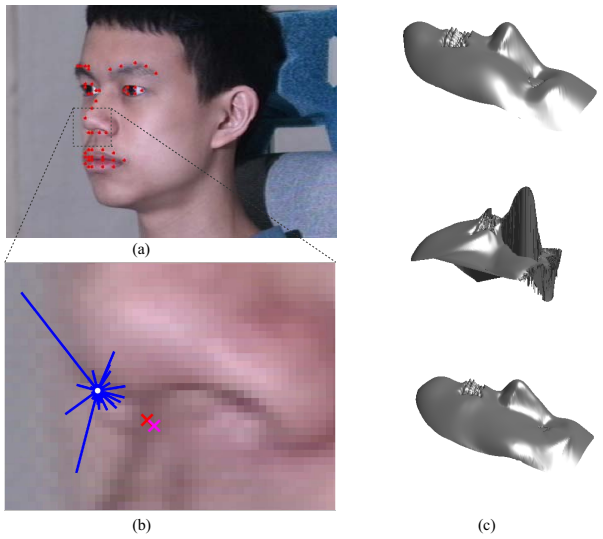


Fig. 2. (a) Detected landmarks (b) Geometric meaning of deformation components: red cross is the detected landmark while the white dot is the projected landmark from the generic model (the initial estimation). The deformation components in blue indicate the directions on which deform the model, while the coefficients are the relative magnitude. Once the model is fit, the landmark projects on the magenta cross. (c) Respectively from top to bottom: average model, estimate of the shape without regularization and finally with the proposed regularization. Note how the chin and the nose have changed.

Once the new shape is obtained using the new $\boldsymbol{\alpha}$ as detailed in Eq. (2) , we proceed to perform a final estimation of the pose $\mathbf{P}'$ with the new shape $\mathbf{S}'$ and additionally to select the visible indices of vertices of the entire shape $\mathbf{J}_v$. In the following Section, we use these two last estimates to synthesize a frontal view from a profile face accounting for rigid and non-rigid transformation.

V. Recognition by Localizing Local Binary Patterns on the Deformed Vertices

In this Section we describe our face recognition method across a range of facial poses. Considering the result achieved in the previous Sections, differently from recent approaches likes [6], [2], [3], our method can handle continuos pose variations and does not require any manual labeling of data except for the 3D landmarks in the average model. Our approach supposes to process the gallery by fitting the 3DMM for each subject and by extracting a LBP histogram on a window localized on each vertex of the deformed model. Furthermore in order to reduce the dimensionality of the descriptor, we uniformly sample the mesh vertices. From our experimental results we deduced that a 2D Gaussian low-pass filter applied on the gallery images enhances the performance. Thus we apply it with a size of 5×5 and a standard deviation of 0.9 on the frontal face rendering. Once the LBP histograms are extracted, they are stacked together to form a unique descriptor similar to [4].

Defining a LBP on the vertex has several benefits w.r.t. the state-of-the-art method [4] that divides the image with a regular uniform grid and extract LBP histogram in each cell of the grid. These benefits are the following:

- this makes the feature vector independent of the image size.
- considering that the deformed model has been optimized to fit the face, the LBP are better localized.
- it gives a straightforward way to restrict the feature vector to those parts which are not visible considering a self-occluding face.

Our method to recognize a face is the following. Given a test image, once we have an estimate of pose $\mathbf{P}'$ and a regularized shape $\mathbf{S}'$, we proceed to render a frontal view and we sample LBP on just a uniform subset of the deformed visible vertices. So in this case when the face undergoes self-occlusion, the feature dimension of the query will be less than the one in the gallery. To this end, we exploit the visible indices $\mathbf{J}_v$ and select the part of the feature vector in the gallery that corresponds to those indices. Once we a have a face descriptor $\mathbf{f} \in \mathbb{R}^{F(\mathbf{J}_v)}$ for a query image, we simply apply a Nearest Neighbor (NN) algorithm to select the closest feature from the ones in the gallery $\mathbf{G} \in \mathbb{R}^{F(\mathbf{J}_v) \times N_s}$ where N_s is the number of subjects in the gallery and $F(\mathbf{J}_v)$ is the feature dimension that arises considering each time the visible vertices. Thus our recognition rule simply is:

$$id(\mathbf{f}) = \arg\min_i \|\mathbf{f} - \mathbf{G}_i\|_2 \quad (11)$$

Each frontal face image is rendered on the XY plane considering the deformed shape, interpolating the RGB values sampled from the non-frontal face on an uniform grid with natural neighbor interpolation. This generally gives a face image size of about 120×160 pixels. On each vertex we sample a patch of size 11 pixels and each LBP histogram is quantized in 58 bins, where two bins account for non-uniform binary patterns and the remaining count the uniform binary patterns.

VI. Experimental Results

In this Section we report the experiment results obtained using our method, comparing the performance figures with the approach of [4] considering various alignment modalities such as:

- *2d-eyes*: 2D similarity alignment approach based on eyes-mouth locations. The similarity is defined by the triangle

(a) (b) (c) (d) (e) (f)

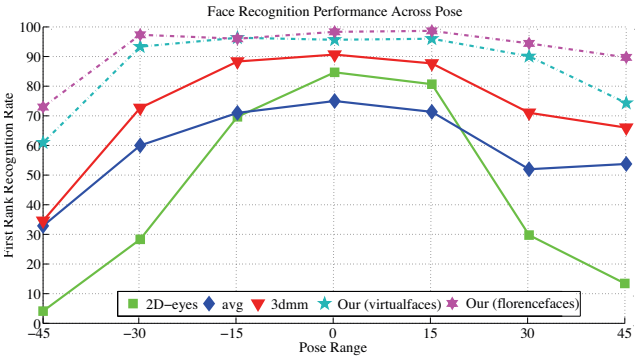
Fig. 3. Qualitative examples: *(a)* the probe image. Then the gallery image cropped with similarity alignment. *(b-f)* From left to right we show frontal renderings with the following: 2D similarity method; generic model using all the landmarks; generic model with only selected landmarks; morphed model without regularization; morphed model with regularization.

formed by the eyes and the mouth that maps into a triangle in a template image of size 200×240;

- *avg*: render a frontal image by pose normalization using a 3D average model and a final 2D similarity alignment.
- *3dmm*: similar to the previous one; obtaining a frontal image by pose normalization using a 3DMM and performing a final 2D similarity alignment.

All the approaches use the same landmark detector, that is the one used specified in Sect. IV-A. If we apply the approach of [4] on the frontal renderings, these have different dimensions and the above approach does not provide a way to extract features with a fixed length. To overcome this, we project on the rendered face the 3D landmarks and by selecting the triangle connecting the eyes and the mouth, we perform a final 2D similarity alignment that gives an image of the same size, irrespective of the rendering size. Considering these baselines, for the face poses at $\pm 45°$ we use just the left/right part of the face for matching.

In addition to these baselines, we compare with state-of-the-art results in [6], [2], [5], [7], [3] on a regular dataset used in literature: we test our approach on CMU Multi-PIE, expressing the performance by principally reporting the recognition rate at first rank under a range of facial poses. However in order to show the potentiality of our method we show also the full CMC (Cumulative Matching Characteristic) curves along with the nAUC (normalized Area Under the Curve).

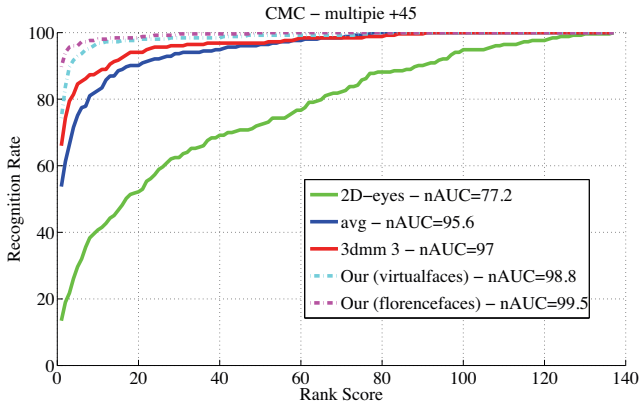


(a)

Fig. 4. The first rank recognition rate as a function of the facial pose on the CMU-MPIE dataset: note how the proposed approach leads to a performance that tends to be pose independent.

CMU Multi-PIE is the most recent of controlled face databases and our experiments on this will facilitate the comparison with future methods. Moreover this dataset is very complete because the subject are framed under every possible conditions: by varying the pose, illumination and the facial expression.

We recreate exactly the experiment settings of the approach [2] considering 137 subjects (subject from 201 to 346) with neutral expression from all 4 sessions at 7 different poses, with illumination that is frontal with respect to the face (see labels in Tab. I).



(a)

Fig. 5. CMC curves with probe faces showing a pose of $45°$ along with the nAUC.

We use the frontal image (label pose 051) from the earliest session for each subject as the gallery image (137 total) and all of the remaining per subject as the probe set. The dimension of the probe set is $1,963$ images. Note that differently from other approaches that trained PCA and LDA [7], [3] or learned the landmark detector on the first 200 subjects [2], we did not use these data nor use PCA/LDA. In Fig. 3 we show some face rectification for the subject 201 when the face undergoes a pose variation in yaw of $-45°$: on the left we report the probe profile image along with the gallery image aligned with 2D similarity. Then from left to right the rendered frontal image, respectively using the 2D similarity method (*2D-eyes*); the generic model using all the landmarks (*avg*); the generic model with only selected landmarks; the morphed models without regularization (*3Dmm*); and finally the morphed models with regularization. In Fig. 4 we show the face recognition performance across pose at first rank comparing the baselines. If we consider LBP on

Pose	-45°	-30°	-15°	0°	+15°	+30°	+45°	Mean
Label	*080_05*	*130_06*	*140_06*	*051_07*	*050_08*	*041_08*	*190_08*	–
LGBP [5]	37.7	62.5	77.0	92.6	83.0	59.2	36.1	64.0
BMVC05 [6]	43.8	83.3	94.0	96.3	94.7	70.0	41.2	74.8
ICCV11 [2]	74.1	91.0	95.7	96.9	95.7	89.5	74.8	87.7
ECCV12 s1 [7]	78.7	94.0	99.0	–	98.7	92.2	81.8	90.7
ECCV12 s2 [7]	84.7	95.0	**99.3**	–	99.0	92.9	85.2	92.7
CVIU12 [8]	**84.8**	96.6	99.2	–	**99.2**	**96.2**	89.0	**94.1**
Ours *(virtualfaces)*	61.0	93.3	96.3	95.6	96.0	90.0	74.3	86.6
Ours *(florencefaces)*	72.9	**97.3**	96.0	**98.3**	98.7	94.4	**89.7**	92.5

TABLE I. POSE-WISE FIRST-RANK RECOGNITION RATES (%). BOLD NUMBERS INDICATE THE BEST SCORE.

a regular grid [4], from our experiment evaluation, it arises that even if the face is normalized to a frontal pose, they still do not give compelling performances. They do provide better performance then 2D similarity but our approach using LBP localized with deformation components outperforms both. Moreover we experimented that the proposed approach is more discriminative if it uses 3D real face models *(florencefaces)* than virtual human faces *(virtualfaces)*: to this end we report also the performance figures of our approach when it uses the 3D models of the Florence Faces dataset [20]. This can be observed also from the CMC curve shown in Fig. 5 and in Tab. I. In the latter we present also a comparison with the state-of-the-art: it is shown a pose-wise breakdown of recognition rate at first rank against recent methods. From this comparison our approach shows comparable results against the state-of-the-art and in some case it reports better performance f.e. for poses at $\{-30, 0, +45\}$ degree.

VII. CONCLUSION

In this paper we have presented a novel approach to recognize a face from a still image irrespective of the pose, showing compelling performances up to $\pm 45°$. Given a lateral view, our approach classify a face by localizing LBP histograms on the frontal rendering, that is created sampling RGB values on the probe image. The frontal rendering is obtained by accounting for both a rigid e non-rigid registration in term of, respectively, 3D pose estimation and deformation components of a face.

REFERENCES

[1] A. Li, S. Shan, X. Chen, and W. Gao, "Maximizing intra-individual correlations for face recognition across pose differences," in *Computer Vision and Pattern Recognition*, 2009, pp. 605–611. 1, 2

[2] A. Asthana, T. Marks, M. Jones, K. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *International Conference on Computer Vision*, 2011. 1, 2, 3, 4, 5, 6

[3] D. Yi, Z. Lei, and S. Li, "Towards pose robust face recognition," in *Computer Vision and Pattern Recognition*, 2013, pp. 3539–3545. 1, 2, 3, 4, 5

[4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037 –2041, dec. 2006. 1, 4, 5, 6

[5] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *International Conference on Computer Vision*, vol. 1, oct. 2005, pp. 786 – 791 Vol. 1. 1, 5, 6

[6] T. M. K. T. Akshay Asthana, Michael Jones and R. Goecke, "Pose normalization via learned 2d warping for fully automatic face recognition," in *British Machine Vision Conference*, 2011. 1, 4, 5, 6

[7] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan, "Morphable displacement field based image matching for face recognition across pose," in *European Conference on Computer Vision*, 2012. 1, 2, 5, 6

[8] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs, "Robust pose invariant face recognition using coupled latent space discriminant analysis," *Computer Vision and Image Understanding*, vol. 116, no. 11, pp. 1095–1110, 2012. 1, 2, 6

[9] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Transactions on Pattern Analysis and Machine Intelligence*, pp. 210 –227, feb. 2009. 1

[10] A. Ashraf, S. Lucey, and T. Chen, "Learning patch correspondences for improved viewpoint invariant face recognition," in *Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 1

[11] T. Kanade and A. Yamada, "Multi-subregion based probabilistic approach toward pose-invariant face recognition," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 2003. 1

[12] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Conference on Computer Graphics and Interactive Techniques*, 1999, pp. 187–194. 2

[13] ——, "Face recognition based on fitting a 3D morphable model," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003. 2, 3

[14] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3D generic elastic models," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1952–1961. 2

[15] J. Heo and M. Savvides, "Gender and ethnicity specific generic elastic models from a single 2d image for novel 2d pose face synthesis and recognition," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2341–2350, Dec. 2

[16] I. Masi, G. Lisanti, A. Bagdanov, P. Pala, and A. Del Bimbo, "Using 3D models to recognize 2D faces in the wild," in *Computer Vision and Pattern Recognition Workshops*, 2013. 2

[17] F. Schroff, T. Treibitz, D. Kriegman, and S. Belongie, "Pose, illumination and expression invariant pairwise face-similarity measure via doppelganger list comparison," in *International Conference on Computer Vision*, Nov 2011. 2

[18] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition*, 2012. 3

[19] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," in *ACM SIGGRAPH*, 2007. 3

[20] A. D. Bagdanov, A. Del Bimbo, and I. Masi, "Florence faces: a dataset supporting 2d/3d face recognition," in *Int. Symposium on Communication Control and Signal Processing*, Rome, Italy, 2012. 6