Morphable Displacement Field Based Image Matching for Face Recognition across Pose

Speaker: Iacopo Masi

Authors:

Shaoxin Li	Xin Liu	Xiujuan Chai
Haihong Zhang	Shihong Lao	Shiguang Shan

Work presented as poster @ ECCV 2012, Florence

Face Recognition: definition

Definition

Face recognition means order a gallery of face imagery given a probe face used as query in order to get the true positive at first rank. Usually the gallery is ordered throughout a score or error function.

(a) Query

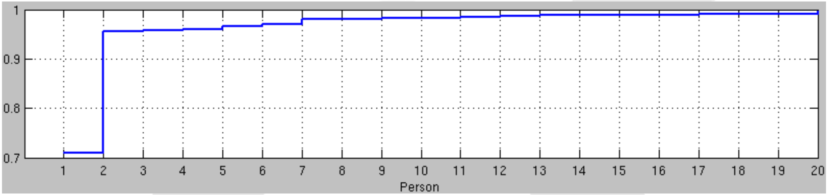
(b) Sorted Gallery

Face Recognition: definition

Definition

Face recognition means order a gallery of face imagery given a probe face used as query in order to get the true positive at first rank. Usually the gallery is ordered throughout a score or error function.

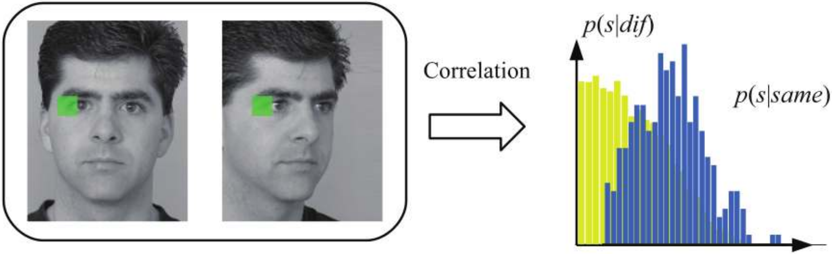
(c) Sorted Gallery



(d) Sorted Gallery

The problem: Face Recognition across Pose

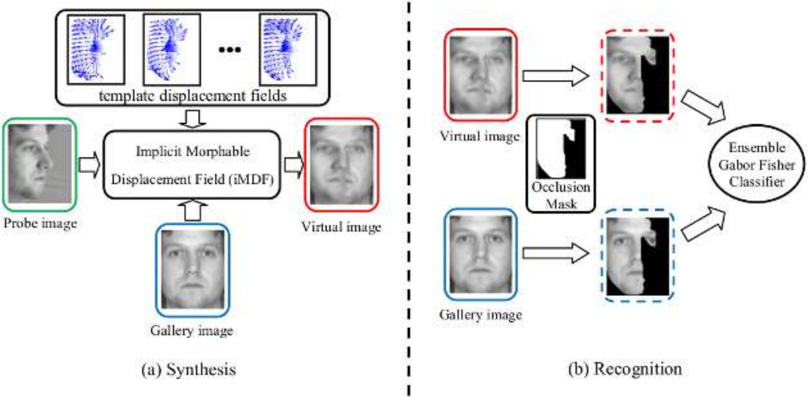
- Face Recognition is a well-know problem in Computer Vision and the 2D frontal face recognition problem is now well experimented with technique such as Nearest Subspace or descriptor like Local Binary Pattern.
- One of the many challenges in Face Recognition, it is the face pose assumed by the imaged subject.
- Basically the pose variation wastes the face representation that it used to recognize. An example can be viewed in the following figure in which brings the due distributions $p(s|same)$ and $p(s|diff)$ to be overlapped.



Overview

These are two building block of the paper:

1. Virtual View Synthesis *(More important)*
2. Face Recognition *(Less Important, considering that any method can be used)*



Displacement Field

Definition

Displacement Field is a 2D point-wise displacement matrix computed using a 3D model.

Given a 3D model S_i, the associated Displacement Field T_i is given by:

$$T_i = L^{frontal} \cdot S_i - L^{pose} \cdot S_i \quad (1)$$

where $S_i \in \mathbb{R}^{P \times 3}$ and $T_i \in \mathbb{R}^{P \times 2}$.

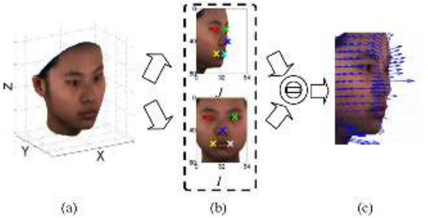


Figure: Displacement field computation

Displacement Field

$$T_i = L^{frontal} \cdot S_i - L^{pose} \cdot S_i$$

The displacement field has the following properties:

- That is to say the linear operator applied to 3D face shape model
- It is pose-dependent but person-independent

In order to get a displacement field each 3D model should be normalized using two eyes locations (f.e.). In this way the displacement matrix has the same dimensions for all the subjects.

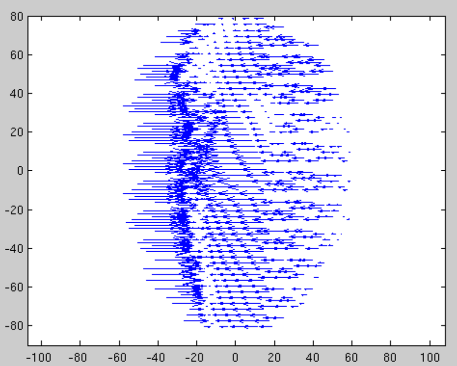


Figure: A displacement field computed with a rotation of -45 degree.

Virtual View Synthesis

The basic idea of virtual view synthesis is the of finding the optimal displacement of the pixels from the profile face to the frontal face in order to minimize some cost function (e.f. norm ℓ_2 of pixel difference).

Probe image   Virtual image

$$\mathbf{T}^{\star} = \arg\min_{\mathbf{T}} \sum_{z} \left\| I(z) - J(\ z + \mathbf{T}(z)\ ) \right\|_2 \ , \quad \forall I \in \mathcal{G} \qquad (2)$$

Of course this displacement depends of both:

1. the pose of the face.
2. the complex 3D structure of the subject.

Though simple and plausible, the expression (2) suffers from severe over-fitting due to the high dimensionality of displacement field $\mathbf{T}(z)$ (i.e. $|\mathbf{T}(z)| = P \times 2$).

Virtual View Synthesis

The basic idea of virtual view synthesis is the of finding the optimal displacement of the pixels from the profile face to the frontal face in order to minimize some cost function (e.f. norm ℓ_2 of pixel difference).

Probe image   Virtual image

$$\mathbf{T}^{\star} = \arg\min_{\mathbf{T}} \sum_{z} || \underbrace{I(z)}_{gallery\ image} - \underbrace{J(\ z + \mathbf{T}(z)\ )}_{synthesized\ frontal\ image} ||_2\ , \quad \forall I \in \underbrace{\mathcal{G}}_{gallery\ set} \tag{3}$$

Of course this displacement depends of both:

1. the pose of the face.
2. the complex 3D structure of the subject.

Though simple and plausible, the expression (1) suffers from severe over-fitting due to the high dimensionality of displacement field $\mathbf{T}(z)$ (i.e. $|\mathbf{T}(z)| = P \times 2$).

Linear Combination of Displacement Fields

According to literature [1], assuming 3D face shape S_i vectors approximately consist a linear object class

$$S = \sum_{i}^{N} \alpha_i S_i \quad s.t. \quad \sum_{i}^{N} \alpha_i = 1, \alpha_i \geqslant 0$$

Authors here suppose they can model the displacement field building a statistic of previously acquired field and combine them in a linear combination like Blanz and Vetter [1, 2] did for Face Morphable Model.

$$\begin{aligned} \mathbf{T}^{\star} = L^{frontal} \cdot S_i - L^{pose} \cdot S_i = \\ = \sum_{i}^{N} \alpha_i \left[L^{frontal} \cdot S_i - L^{pose} \cdot S_i \right] = \\ \approx \sum_{i}^{N} \alpha_i \mathbf{T}_i . \end{aligned} \tag{4}$$

Morphable Displacement Field (MDF)

A linear combination of Displacement Field is called a Morphable Displacement Field (MDF).

Benefits

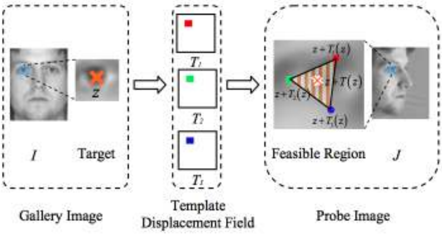
- Modeling the target displacement field as a convex combination of template displacement fields ensures that the obtained displacement field falls in a rational parameter space (globally conforming)
- MDF also quarantee Local consistency indicates that spatial relationship of neighborhood pixels stays unchanged.

The optimization becomes:

$$\alpha^{\star} = \arg\min_{\alpha} \sum_{z} ||I(z) - J(\ z + \sum_{i}^{N} \alpha_i \mathbf{T}_i\ )||_2 \tag{5}$$

But $J \sim f(\alpha_i)$ is still function of the combination of the displacement field.

Implicit Morphable Displacement Field (iMDF)



When N = 3, for each target point z in image I, the feasible region of matching point is actually a triangle. Considering that, if the feasible region is sufficiently small, then grayscale of any pixels in it can be approximately interpolated by grayscale of the convex hulls vertices:

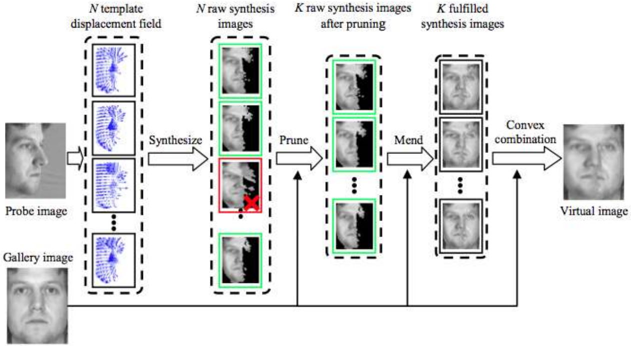
$$J(\ z + \sum_{i}^{N} \alpha_i \mathbf{T}_i\ ) \approx \sum_{i}^{N} \alpha_i J(\ z + \mathbf{T}_i\ ) \qquad (6)$$

But this time, $J \neq f(\alpha_i)$, it shouldn't be computed in the optimization.

Overall Optimization

$$\alpha^\star = \arg\min_{\alpha_{N_i}} \sum_z ||I(z) - \sum_i^K \alpha_{N_i} J(\ z + \mathbf{T}_{N_i}\ )||_2 \quad \forall I \in \mathcal{G} \tag{7}$$

But $J \neq f(\alpha_i)$, it shouldn't be computed in the optimization, but just pre-computed offline. (N_i is the $i-th$ nearest neighbor with $K \ll N$).

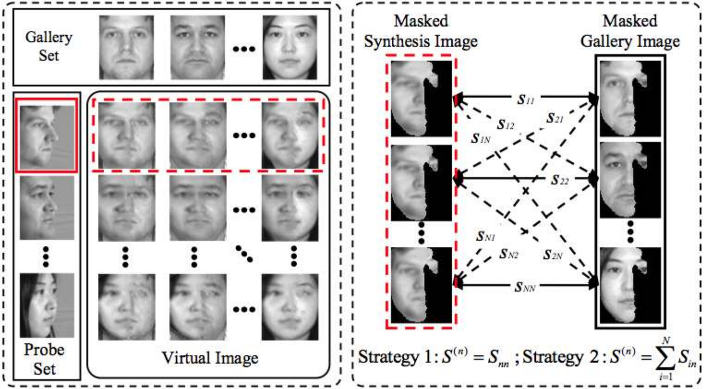


Face Recognition

They designed two kinds of recognition strategies:

- Image-to-Image recognition $S^{(n)} = S_{nn}$,
- Image-to-Stack recognition $S^{(n)} = \sum_{i=1}^{N} S_{in}$.

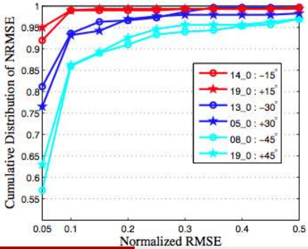
They train a hierarchical ensemble of Gabor fisher classifier (EGFC).

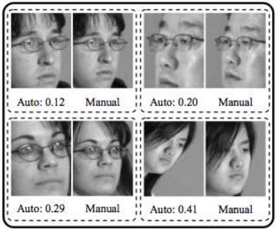


(a) Virtual image synthesized by iMDF (b) Recognition strategy with virtual image

Experimental Evaluation

Multi-PIE Fully-Auto Evaluation							
Method	08_0	13_0	14_0	05_0	04_1	19_0	Avg
	-45°	-30°	-15°	+15°	+30°	+45°	
LGBP[5]	37.7	62.5	77.0	83.0	59.2	36.1	59.3
Asthana11[3]	74.1	91.0	95.7	95.7	89.5	74.8	86.8
EGFC[15]	15.4	56.6	**99.7**	**99.3**	63.7	16.6	58.5
EGFC-S1	78.7	94.0	99.0	98.7	92.2	81.8	90.7
EGFC-S2	**84.7**	**95.0**	99.3	99.0	**92.9**	**85.2**	**92.7**





References I

Volker Blanz and Thomas Vetter.
A morphable model for the synthesis of 3d faces.
In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

Volker Blanz and Thomas Vetter.
Face recognition based on fitting a 3d morphable model.
IEEE Trans. Pattern Anal. Mach. Intell., 25(9):1063–1074, September 2003.