Pose Normalization via Learned 2D Warping for Fully Automatic Face Recognition

Akshay Asthana[1,2]
aasthana@rsise.anu.edu.au

Michael J. Jones[1] and Tim K. Marks[1]
{mjones,tmarks}@merl.com

Kinh H. Tieu[1,3]
kinh.tieu@gmail.com

Roland Goecke[2,4]
roland.goecke@ieee.org

[1]Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA
[2]CECS, Australian National University (ANU), Canberra, ACT, Australia
[3]now at Heartland Robotics, Boston, MA, USA
[4]Faculty of ISE, University of Canberra, Canberra, ACT, Australia

We propose a novel 2D pose normalization method for face recognition that: (1) is fully automatic, (2) can handle continuous pose variation, and (3) is not database specific. It uses multidimensional Gaussian process regression [4] to learn a nonlinear mapping function from the 2D shapes of non-frontal faces at any pose to the corresponding 2D frontal face shapes. Our system takes an input image of a new face at an arbitrary pose, predicts its frontal shape using the learned mapping function, and warps the texture from the non-frontal input shape to this predicted frontal shape to generate a synthetic frontal face that is used for recognition. Our fully automatic system includes an automatic method for extracting 2D facial feature points and estimating 3D head pose. This information is used as input to the 2D pose-normalization algorithm.

Continuous-Pose Regression (CPR)

Suppose we have training data at n discrete poses P_i, where $i = \{1,..,n\}$. For each pose P_i, the training data consist of the face shapes of m training subjects, $\mathbf{s}^i_j$, where $j = \{1,\ldots,m\}$. The shape vector $\mathbf{s}^i_j$ contains the (x, y) coordinates of ℓ landmark points on the face. The goal is to learn a mapping function $\mathcal{F}$ from the shape vector of any face at any training pose to the corresponding shape vector at the frontal pose, P_1. Moreover, this function $\mathcal{F}$ should be continuous so that it can also map from any intermediate pose (not in the discrete set of training poses) to P_1.

We use Gaussian process regression [4] to learn this mapping function as a multidimensional regression function $\mathcal{F}$ that applies to all poses, by including the 3D pose of the face as an input to the regression:

$$\mathcal{F} : \{\mathbf{s}^i_j, P_i\} \rightarrow \mathbf{s}^1_j \,. \tag{1}$$

We call this method *Continuous-Pose Regression* (CPR). We learn $\mathcal{F}$ as a collection of independent functions F_h, where $h = \{1,\ldots,2\ell\}$, each mapping a single landmark point's x or y coordinate from all poses to the frontal pose.

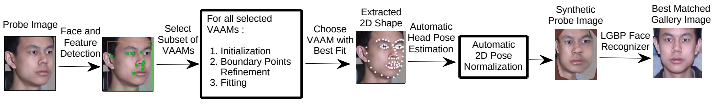

Figure 1: Overview of our fully automatic pose-invariant face recognition system.

Fully Automatic Face Recognition System

Our fully automatic face recognition system is summarized in Figure 1. First, we run a multi-view face detector to find all faces in the input image. For each face, a set of facial feature detectors locates several facial landmark points, which are used to initialize a set of view-specific Active Appearance Models (VAAMs). If the pose class from the multi-view face detector is left or right half-profile, the initial VAAM boundary point locations are refined to ensure a better fit of the VAAM to the face boundary. Next, each VAAM model is iteratively fit to the input image using the Efficient Approximation to the Simultaneous Inverse Compositional algorithm [1]. The VAAM with the smallest fitting error is selected, yielding a set of 68 facial points. Using these points, we normalize the roll angle of the face and then use a regression function to estimate the yaw and pitch angles. This 3D pose information and the best-fitting VAAM points are used as input to the 2D pose-normalization algorithm, which outputs a synthetic frontal face image. This synthetic image is used for recognition by the Local Gabor Binary Pattern (LGBP) face recognizer [6]. The entire process takes about 6 seconds on a modern Pentium processor.

Experiments

We conducted face recognition experiments using the USF Human ID 3D, CMU PIE, and Multi-PIE databases. Failure to acquire (FTA) occurs if the face is not detected or fewer than 3 facial features are located, in which case no pose-normalized face image is output. We remove any FTA cases from the test set so that we can clearly distinguish between detection failures and recognition failures. Figure 2 compares our results to previous methods.

CMU PIE

Method	Alignment	Trained on PIE	Gallery Size	Poses Handled	c11	c29	c07	c09	c05	c37	**Avg**
Kanade03	manual	yes	34	discrete	96.8	100.0	100.0	100.0	100.0	100.0	99.5
Chai07	manual	no	68	discrete	89.8	100.0	98.7	98.7	98.5	82.6	94.7
Sarfraz10	automatic	yes	34	continuous	87.9	89.2	99.8	92.8	91.5	87.9	91.5
Sarfraz10	automatic	no	68	continuous	84.0	87.0	–	–	94.0	90.0	88.8
LGBP	automatic	no	67	N/A	71.6	87.9	78.8	93.9	86.4	74.6	82.2
Ours	automatic	no	67	continuous	88.1	100.0	98.5	98.5	95.5	89.4	95.0

USF

Pitch Range (°)	-15 to +15		-30 to -20 and +20 to +30	
Yaw Range (°)	LGBP	**Ours**	LGBP	**Ours**
-15 to +15	97.1	99.8	84.4	98.7
-30 to -20 and +20 to +30	88.8	98.8	67.2	96.3
-45 to -35 and +35 to +45	78.3	95.2	–	–

Multi-PIE

Method	080 -45°	130 -30°	140 -15°	051 0°	050 +15°	041 +30°	190 +45°	**Avg**
LGBP	37.7	62.7	77.0	92.6	83.0	58.7	35.9	62.0
Ours	43.8	83.3	94.0	96.3	94.7	70.0	41.2	74.8

Figure 2: Pose-wise rank-1 recognition rates (in %) for CMU PIE, USF 3D, and Multi-PIE databases. The numbers for Kanade03[3], Chai07[2] and Sarfraz10[5] were estimated from plots. For LGBP results, we used code from the authors of [6]

USF 3D: We rendered 199 different poses up to ±45° yaw and ±30° pitch for each of the 94 unique subjects. The FTA rate was 3.37%, which is higher than on the other data sets due to the combined large yaw and pitch angles. Overall recognition rate was **97.8%**.

CMU PIE: The gallery consisted of the frontal image (Pose ID c27) and the probe set consisted of 6 non-frontal poses. Test set consisted of 67 subjects (gallery image for 1 subject was a FTA). The FTA rate was 1.1%. Overall recognition rate was **95.0%**.

Multi-PIE: We tested on 137 subjects (Subject ID 201–346) with neutral expression at 7 different poses from all 4 sessions. Our VAAM model was trained on 200 subjects (ID 001–200), who were not used in our recognition test. The FTA rate was 1.2%. Overall recognition rate was **74.8%**.

Figure 3: Examples from USF 3D, CMU PIE, and Multi-PIE. *Top:* input face images. *Bottom:* pose-normalized face images.

[1] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical report, Robotics Institute, Carnegie Mellon University, USA, 2003. 1

[2] X. Chai, S. Shan, X. Chen, and W. Gao. Locally linear regression for pose-invariant face recognition. *IEEE Transactions in Image Processing (TIP)*, 16 (7):1716 –1725, 2007. 1

[3] T. Kanade and A. Yamada. Multi-Subregion Based Probabilistic Approach Toward Pose-Invariant Face Recognition. In *IEEE CIRA*, 2003. 1

[4] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. 1

[5] M. S. Sarfraz. *Towards Automatic Face Recognition in Unconstrained Scenarios*. PhD thesis, Technische Universität Berlin, 2008. 1

[6] 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. *IEEE ICCV*, 2005. 1