Pose-Invariant 3D Face Alignment

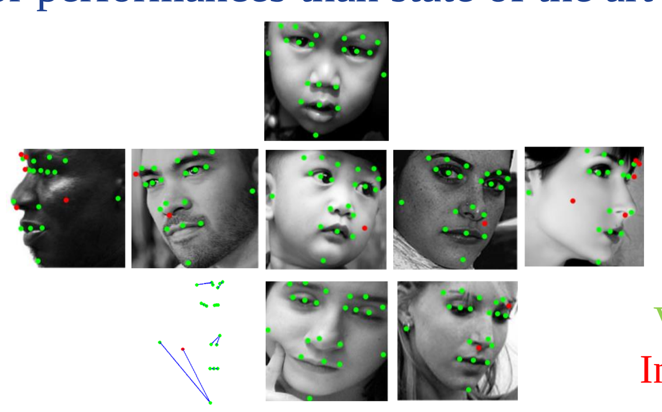
Amin Jourabloo, Xiaoming Liu

Computer Vision Lab, Department of Computer Science, Michigan State University

Introduction

We proposed a method which

- estimates 2D landmarks and their visibility for a face with **arbitrary** pose.
- estimates both the projection matrix and 3D landmarks.
- achieves superior performances than state of the art methods.



Visible Landmarks

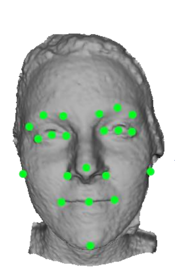
Invisible Landmarks

Prior Work

Method	3D landmark	Visibility	Pose-related database	Pose range	Landmark #	Estimation Error
RCPR (ICCV2013)	NO	Yes	COFW	frontal w. occlu	19	8.5
CoR (ECCV2014)	NO	Yes	COFW; LFPW-O; Helen-O	frontal w. occlu	19; 49; 49	8.5
TSPM (CVPR2012)	NO	NO	AFW	all poses	6	11.1
CDM (ICCV2013)	NO	NO	AFW	all poses	6	9.1
OSRD (CVPR2014)	NO	NO	MVFW	$< \pm 40^\circ$	68	N/A
TCDCN (ECCV2014)	NO	NO	AFLW, AFW	$< \pm 60^\circ$	5	8.0, 8.2
PIFA	**Yes**	**Yes**	**AFLW, AFW**	**all poses**	**21, 6**	**6.5, 8.6**

3D Face Modeling

3D scans w. labels



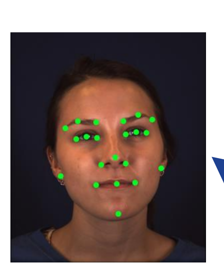
$$S = \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \\ 1 & 1 & \dots & 1 \end{pmatrix} = S_0 + \sum_{i=1}^{N_S} p_i S_i$$

$$p = (p_1, p_2, \dots, p_{N_S})$$

$$M = \begin{pmatrix} sR & \begin{matrix} t_x \\ t_y \end{matrix} \end{pmatrix}$$

a 2×4 matrix with 7 degrees of freedom (pitch, yaw, roll, 2 scales and 2 translation).

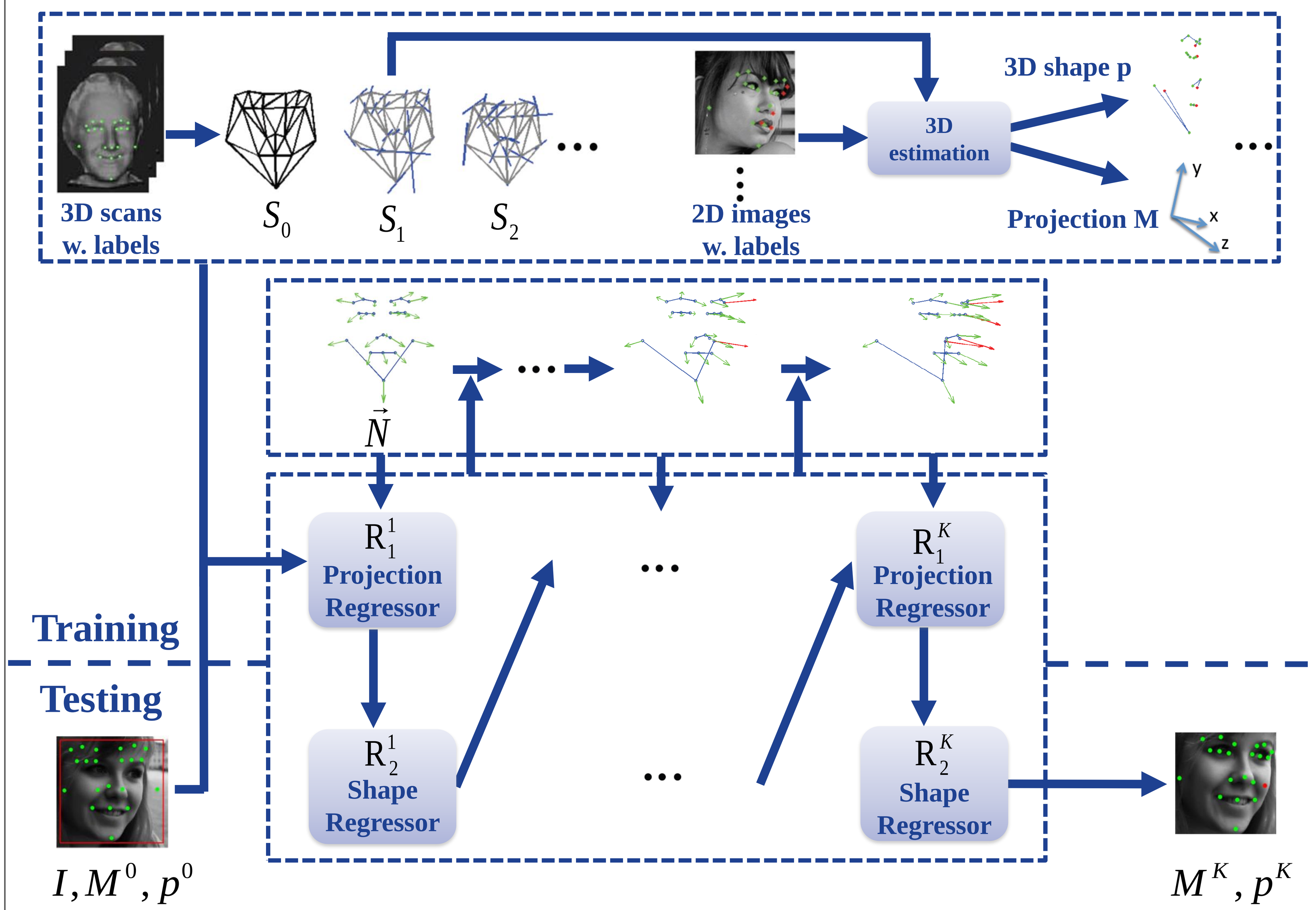
2D image w. landmarks



$$U = \begin{pmatrix} u_1 & u_2 & \dots & u_N \\ v_1 & v_2 & \dots & v_N \end{pmatrix} = MS$$

We represent 2D landmarks U as pair of projection matrix M and 3D shape parameter p.

PIFA Approach



Cascaded Couple-Regressor

K-th Projection Matrix Regressor

$$U_i = M_i^{k-1}\left(S_0 + \sum_{i=1}^{N_S} p_i^{k-1} S_i\right)$$

$$\theta_1^k = \arg\min_{\theta_1^k} \sum_{i=1}^{N_d} \left\| \Delta M_i^k - R_1^k\left(I_i, U_i, v_i^{k-1}; \theta_1^k\right) \right\|^2$$

$$M_i^k = M_i^{k-1} + \Delta M_i^k$$

K-th 3D Shape Parameter Regressor

$$U_i = M_i^k\left(S_0 + \sum_{i=1}^{N_S} p_i^{k-1} S_i\right)$$

$$\theta_2^k = \arg\min_{\theta_2^k} \sum_{i=1}^{N_d} \left\| \Delta p_i^k - R_2^k\left(I_i, U_i, v_i^k; \theta_2^k\right) \right\|^2$$

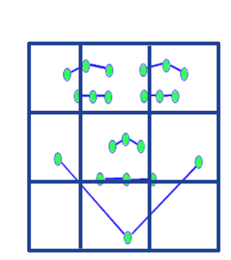
$$p_i^k = p_i^{k-1} + \Delta p_i^k$$

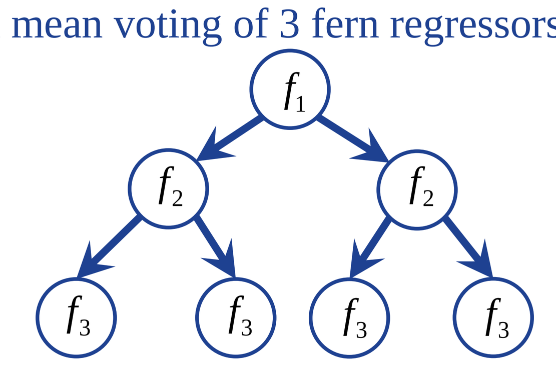
- Any regressor might be used for R. We used
  - Linear regressor

$$R(.) = \theta^T \begin{bmatrix} v_1 & 0 & \dots & \dots & 0 & 0 \\ 0 & v_1 & & & & 0 \\ \vdots & & \ddots & & & \vdots \\ & & & \ddots & & \\ 0 & & & & v_N & 0 \\ 0 & 0 & \dots & \dots & 0 & v_N \end{bmatrix} f(I_i, U_i)$$

$f(I, U)$ is a function to extract 32 N-dim HOG feature vector.

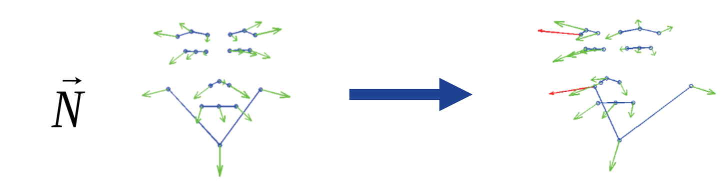
  - Fern regressor
    - 3 out of 9 zones with least occlusion are selected.
    - For each selected zone, a depth 5 random fern regressor is learned.
    - The final regressor is a weighted mean voting of 3 fern regressors.





3D Surface-Enable Visibility

- We rotate the 3D normal surface vectors according to the rotation angle indicated by projection matrix.



- The sign of z coordinate indicates 2D landmark visibility.

$$v = \vec{N}^T . \left(\frac{m_1}{\|m_1\|} \times \frac{m_2}{\|m_2\|} \right)$$

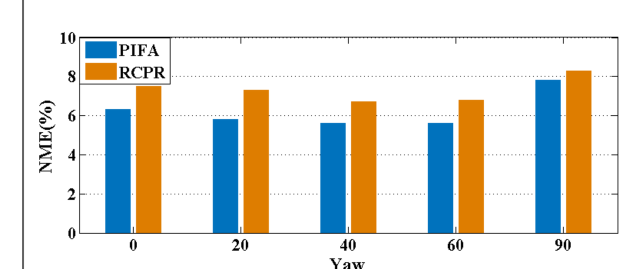
$$\begin{matrix} m_1 \\ m_2 \end{matrix} \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix} M$$

Experimental Results

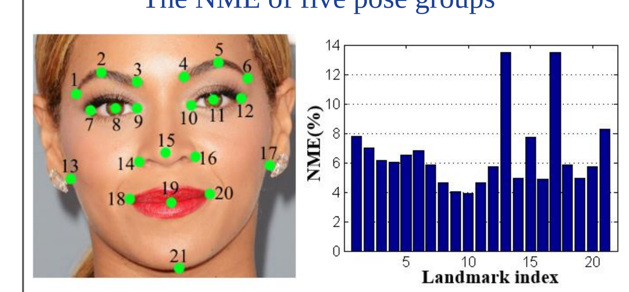
- AFLW dataset experiments
  - 5200 images selected evenly within $[0^\circ, 30^\circ], [30^\circ, 60^\circ], [60^\circ, 90^\circ]$ yaw angles.
  - Randomly partitioned into 3901 training and 1299 testing images.

Visible Landmarks

Invisible Landmarks

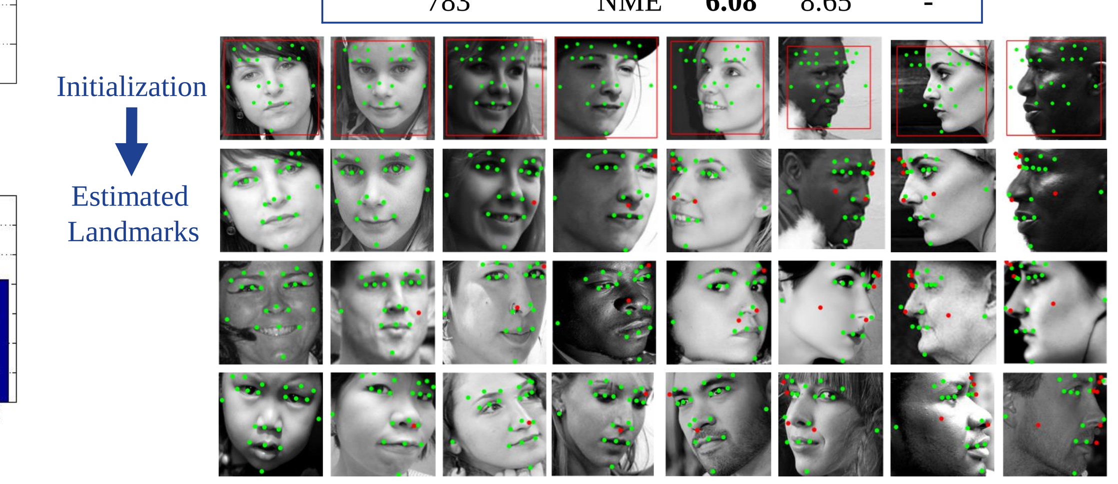


The NME of five pose groups



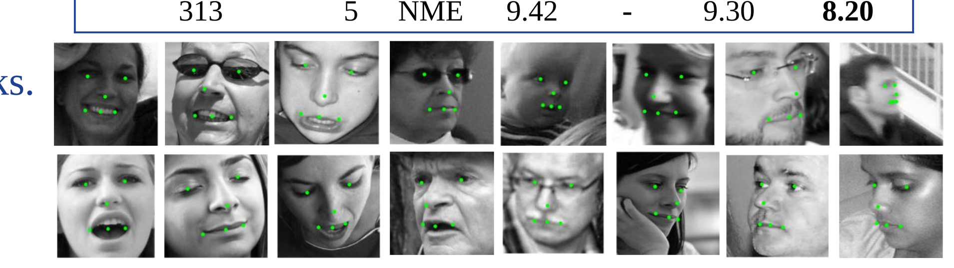
The average NME of each landmark

Number of images	Metric	PIFA	CDM	RCPR
1299	NME	**6.52**	-	7.15
783	NME	**6.08**	8.65	-



- AFW dataset experiments
  - 468 faces in 205 images with poses ± 90.
  - Labeled with up to 6 visible landmarks.

Number of images	N	Metric	PIFA	CDM	RCPR	TCDCN
468	6	MAPE	**8.61**	9.13	-	-
313	5	NME	9.42	-	9.30	**8.20**



- BP4D-S database experiments
  - Includes pairs of 2D images and 3D scans of 41 subjects.
  - Half of selected 1100 images for training and rest for testing.
  - The mean 3D shape is used as a baseline (after global transformation).
  - The MAPE of baseline is 5.02, while PIFA is **4.75**.