M²FPA: A Multi-Yaw Multi-Pitch High-Quality Dataset and Benchmark for Facial Pose Analysis

Peipei Li[1,2] , Xiang Wu[1], Yibo Hu[1], Ran He[1,2*], Zhenan Sun[1,2]

[1]CRIPAC & NLPR & CEBSIT CASIA, [2]University of Chinese Academy of Sciences

CRIPAC Center for Research on Intelligent Perception and Computing

SCAN ME

ICCV 2019 Seoul, Korea

Background:

Facial images in surveillance or mobile scenarios often have large view-point variations in terms of pitch and yaw angles. These jointly occurred angle variations make face recognition challenging. Current public face databases mainly consider the case of yaw variations.

Database	Yaw	Pitch	Yaw-Pitch	Attributes	Illuminations	Subjects	Images	Image Size	Controllabled	Size[GB]	Paired	Year
PIE [21]	9	2	2	4	21	68	41,000+	640×486	✓	40	✓	2003
LFW [9]	No label	No label	No label	No label	No label	5,749	13,233	250×250	×	0.17	×	2007
CAS-PEAL-R1 [5]	7	2	12	5	15	1,040	30,863	640×480	✓	26.6	✓	2008
Multi-PIE [7]	13	0	2	6	19	337	755,370	640×480*	✓	305	✓	2009
IJB-A [15]	No label	No label	No label	No label	No label	500	25,809	1026×698+	×	14.5	×	2015
CelebA [17]	No label	No label	No label	No label	No label	10,177	202,599	505×606	×	9.49	×	2016
CelebA-HQ [12]	No label	No label	No label	No label	No label	No label	30,000	1024×1024	×	27.5	×	2017
FF-HQ [13]	No label	No label	No label	No label	No label	No label	70,000	1024×1024	×	89.3	×	2018
M²FPA (Ours)	**13**	**5**	**44**	**4**	**7**	**229**	**397,544**	**1920×1080**	✓	**421**	✓	**2019**

Table 1. Comparisons of existing facial pose analysis databases.

Contributions:

- We introduce a Multi-yaw Multi-pitch high-quality database for Facial Pose Analysis (M²FPA). It contains 397,544 images of 229 subjects with yaw, pitch, attribute and illumination. https://pp2li.github.io/M2FPA-dataset/
- We provide a comprehensive qualitative and quantitative benchmark of several state-of-the-art methods for face frontalization and pose-invariant face recognition, including DR-GAN, TP-GAN, CAPG-GAN on M²FPA.
- We propose a simple yet effective parsing guided discriminator to capture the local consistency during GAN optimization.

The M2FPA Database

Figure 1. An example of the yaw and pitch variations in our M²FPA database.

a) The diagram of camera positions. b) Overview of the acquisition system. c) The diagram of illumination positions.

Figure 2. Data acquisition.

The Statistics of M²FPA:

The main advantages of M²FPA:

- Large-scale.
- Accurate and diverse poses.
- High-resolution (1920 x 1080).
- Accessory (5 types of glasses).

Figure 3. Examples of four attributes.

Poses	Pitch = $+45°$	Yaw = $-90°, -45°, 0°, +45°, +90°$
	Pitch = $+30°$	Yaw = $-90°, -67.5°, -45°, -22.5°$ $0°, +22.5°, +45°, +67.5°, +90°$
	Pitch = $+15°$	Yaw = $-90°, -75°, -60°, -45°, -30°, -15°$ $0°, +15°, +30°, +45°, +60°, +75°, +90°$
	Pitch = $0°$	Yaw = $-90°, -75°, -60°, -45°, -30°, -15°$ $0°, +15°, +30°, +45°, +60°, +75°, +90°$
	Pitch = $-15°$	Yaw = $-90°, -75°, -60°, -45°, -30°, -15°$ $0°, +15°, +30°, +45°, +60°, +75°, +90°$
	Pitch = $-30°$	Yaw = $-90°, -67.5°, -45°, -22.5°$ $0°, +22.5°, +45°, +67.5°, +90°$
Attributes		Happy, Normal, Wear glasses, Surprise
Illuminations		Above, Front, Front-above, Behind Front-below, Left, Right

Table 2. The poses, attributes and illustrations in M²FPA.

Framework:

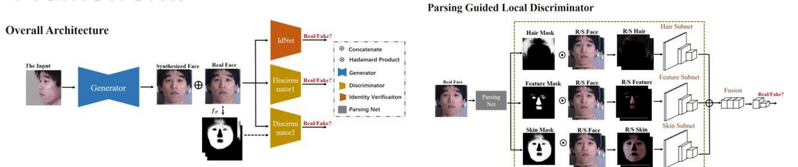

Figure 4. The overall framework of our method.

Global-Local Adversarial Loss:

$$L_{adv1} = \min_{\theta_G}\max_{\theta_{D1}} E_{Y\sim P(Y)}[\log D_{\theta_{D1}}(Y)] + E_{Y\sim P(Y)}[\log(1 - D_{\theta_{D1}}(Y))]$$

$$L_{adv2} = \min_{\theta_G}\max_{\theta_{D2}} E_{Y_h,Y_s,Y_f\sim P(Y_h,Y_s,Y_f)}[\log D_{\theta_{D2}}(Y_h,Y_s,Y_f)] + E_{Y_h,Y_s,Y_f\sim P(\hat{Y}_h,\hat{Y}_s,\hat{Y}_f)}[\log(1 - D_{\theta_{D2}}(\hat{Y}_h,\hat{Y}_s,\hat{Y}_f))]$$

Multi-Scale Pixel Loss:

$$L_{pixel} = \frac{1}{3}\sum_{i=1}^{3}\frac{1}{W_iH_iC}\sum_{w,h,c=1}^{W_i,H_i,C}\left|\hat{Y}_{i,w,h,c} - Y_{i,w,h,c}\right|$$

Identity Preserving Loss:

$$L_{id} = ||\varphi_f(\mathrm{Y}) - \varphi_f(\hat{\mathrm{Y}})||_2^2 + ||\varphi_p(\mathrm{Y}) - \varphi_p(\hat{\mathrm{Y}})||_F^2$$

Total Variation Regularization:

$$L_{tv} = \sum_{c=1}^{C}\sum_{w,h=1}^{W,H}\left|\hat{Y}_{w+1,h,c} - \hat{Y}_{w,h,c}\right| + \left|\hat{Y}_{w,h+1,c}^{b} - \hat{Y}_{w,h,c}\right|$$

Overall Loss:

$$L = \lambda_1 L_{pixel} + \lambda_2 L_{adv1} + \lambda_3 L_{adv2} + \lambda_4 L_{id} + \lambda_5 L_{tv}$$

Evaluation on Multi-PIE:

Method	±15°	±30°	±45°	±60°	±75°	±90°
FIP+LDA [31]	90.7	80.7	64.1	45.9	-	-
MVP+LDA [32]	92.8	83.7	72.9	60.1	-	-
CPF [26]	95.0	88.5	79.9	61.9	-	-
DR-GAN [24]	94.0	90.1	86.2	83.2	-	-
FF-GAN [27]	94.6	92.5	89.7	85.2	77.2	61.2
TP-GAN [10]	98.68	98.06	95.38	87.72	77.43	64.64
CAPG-GAN [8]	99.82	99.56	97.33	90.63	83.05	66.05
Ours	99.96	99.78	99.53	96.18	88.74	75.33

Table 4. Rank-1 recognition rates (%) under Setting2.

(a)Profile (b)Ours (c)CAPG-GAN (d)TP-GAN (e)FF-GAN (f) Frontal

Figure 5. Comparisons with different methods.

Face Frontalization on M²FPA:

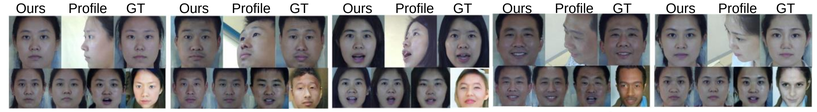

Figure 6. Frontalized results of different methods on M²FPA.

Figure 7. The 512 x 512 frontalization results on M²FPA.

Pose-invariant Face Recognition on M²FPA:

Method	Pitch	±0°	±15°	±30°	±45°	±60°	±75°	±90°
LightCNN-29 v2								
Original	+15°	100	100	100	99.8	97.5	76.5	34.3
	−15°	99.9	100	99.8	99.7	97.3	81.8	45.9
DR-GAN [24]	+15°	99.1	98.8	98.0	94.8	85.6	61.1	20.8
	−15°	98.1	98.2	96.5	93.3	83.1	62.7	31.0
TP-GAN [10]	+15°	99.8	99.8	99.7	99.5	95.7	81.6	50.9
	−15°	99.9	99.9	99.6	99.2	95.9	84.1	56.9
CAPG-GAN [8]	+15°	99.8	99.9	99.8	98.9	95.0	81.4	54.4
	−15°	99.8	99.9	99.7	98.7	95.1	85.5	65.6
Ours	+15°	**99.9**	**99.9**	**99.8**	**99.7**	**97.5**	**86.2**	**56.2**
	−15°	**99.9**	**99.9**	**99.8**	**99.7**	**97.4**	**88.1**	**66.5**
IR-50								
Original	+15°	99.8	99.9	99.6	98.7	95.7	77.1	23.4
	−15°	98.7	99.4	99.2	98.1	95.7	78.8	27.9
DR-GAN [24]	+15°	98.5	98.2	97.8	94.0	84.8	60.9	17.0
	−15°	95.8	97.2	96.2	93.3	84.8	60.3	20.8
TP-GAN [10]	+15°	99.0	99.6	99.1	98.5	94.7	79.1	40.6
	−15°	98.2	98.9	98.1	97.2	94.8	80.9	43.5
CAPG-GAN [8]	+15°	98.9	99.0	98.5	95.8	91.5	75.7	40.7
	−15°	98.5	98.5	97.9	95.3	90.3	76.0	47.8
Ours	+15°	**99.7**	**99.6**	**99.4**	**98.7**	**96.1**	**84.5**	**43.6**
	−15°	**98.6**	**99.1**	**98.7**	**98.8**	**96.5**	**83.9**	**49.7**

Table 5. Rank-1 recognition rates (%) at 15° pitch angle.

Method	Pitch	±0°	±22.5°	±45°	±67.5°	±90°
LightCNN-29 v2						
Original	+30°	99.7	99.2	96.5	71.6	24.5
	−30°	98.6	98.2	93.6	69.9	22.1
DR-GAN [24]	+30°	93.8	91.5	83.4	52.0	16.9
	−30°	91.7	90.6	79.1	46.6	16.6
TP-GAN [10]	+30°	99.7	98.8	95.8	77.2	43.4
	−30°	98.2	97.6	93.4	75.7	38.9
CAPG-GAN [8]	+30°	98.8	98.4	94.1	79.5	48.0
	−30°	98.9	98.3	93.8	75.3	49.3
Ours	+30°	**99.7**	**99.1**	**97.7**	**81.9**	**48.2**
	−30°	**98.9**	**98.7**	**95.8**	**82.2**	**49.3**
IR-50						
Original	+30°	99.2	98.1	94.7	73.5	17.6
	−30°	97.1	97.3	93.0	67.2	9.0
DR-GAN [24]	+30°	92.9	92.3	83.8	56.4	13.9
	−30°	93.0	92.0	82.1	50.3	7.5
TP-GAN [10]	+30°	98.1	97.3	94.4	76.8	34.5
	−30°	95.7	96.1	92.2	71.6	27.5
CAPG-GAN [8]	+30°	97.1	96.2	90.5	73.1	34.5
	−30°	95.8	95.4	89.2	67.6	33.0
Ours	+30°	**98.6**	**97.8**	**96.0**	**79.6**	**36.4**
	−30°	**97.2**	**97.4**	**95.1**	**76.7**	**33.1**

Table 6. Rank-1 recognition rates (%) at 30° pitch angle.

Ablation Study:

Method	±15°	±30°	±45°	±60°	±75°	±90°
LightCNN-29 v2						
w/o $L_{adv1,2}$	99.8	99.7	99.4	97.3	86.1	63.1
w/o L_{tv}	99.8	99.6	99.5	97.9	88.6	67.1
w/o L_{ip}	99.9	99.7	99.0	96.9	86.3	56.5
w/o L_{adv2}	100	100	99.7	98.4	89.3	63.5
Ours	**100**	**100**	**99.9**	**98.4**	**90.6**	**67.6**
IR-50						
w/o $L_{adv1,2}$	99.7	99.3	98.3	94.9	82.1	44.9
w/o L_{tv}	99.4	99.4	98.5	96.2	87.7	52.0
w/o L_{ip}	99.2	99.0	98.3	95.3	83.8	43.4
w/o L_{adv2}	99.7	99.3	98.3	95.7	82.4	45.9
Ours	**99.5**	**99.5**	**99.0**	**97.3**	**89.6**	**55.8**

Table 7. Rank-1 recognition rates (%) on M²FPA.

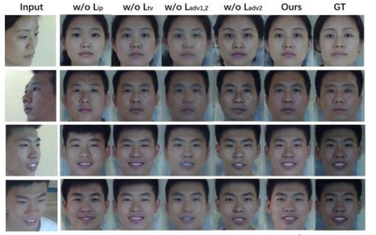

Figure 8. Model comparisons on M²FPA.