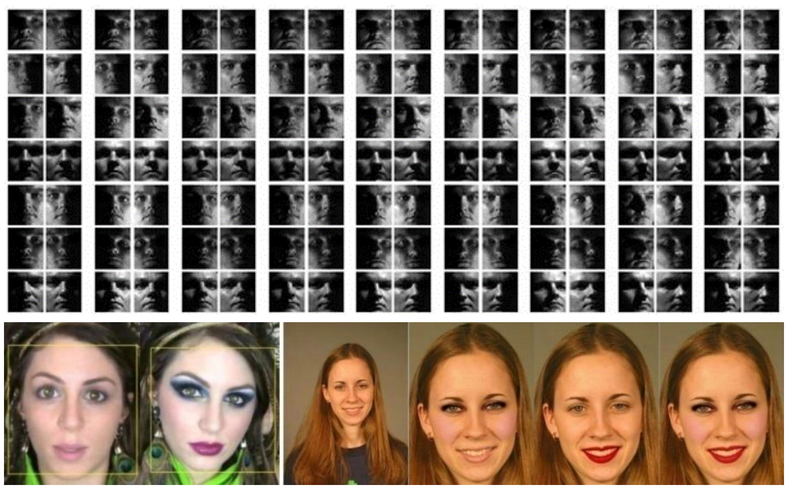
Three-dimensional (3D) facial identity and expression analysis: from handcrafted to learned features

Huibin Li
（李慧斌）
http://gr.xjtu.edu.cn/web/huibinli
数学与统计学院
西安交通大学

VALSE webinar, October 12th, 2016

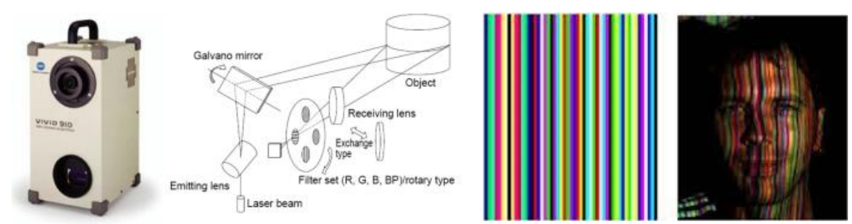
What is biometrics?

Why 3D face recognition?



2D face recognition

Illumination, pose, make up…

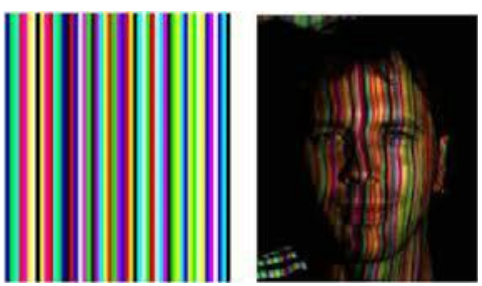


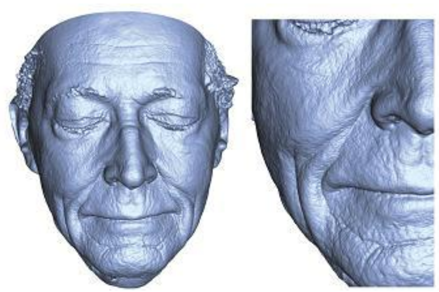
3D face recognition

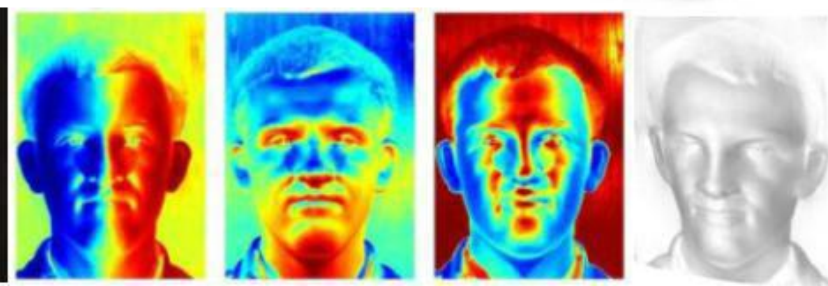
stable to lighting, pose and make up …

3D face acquisition

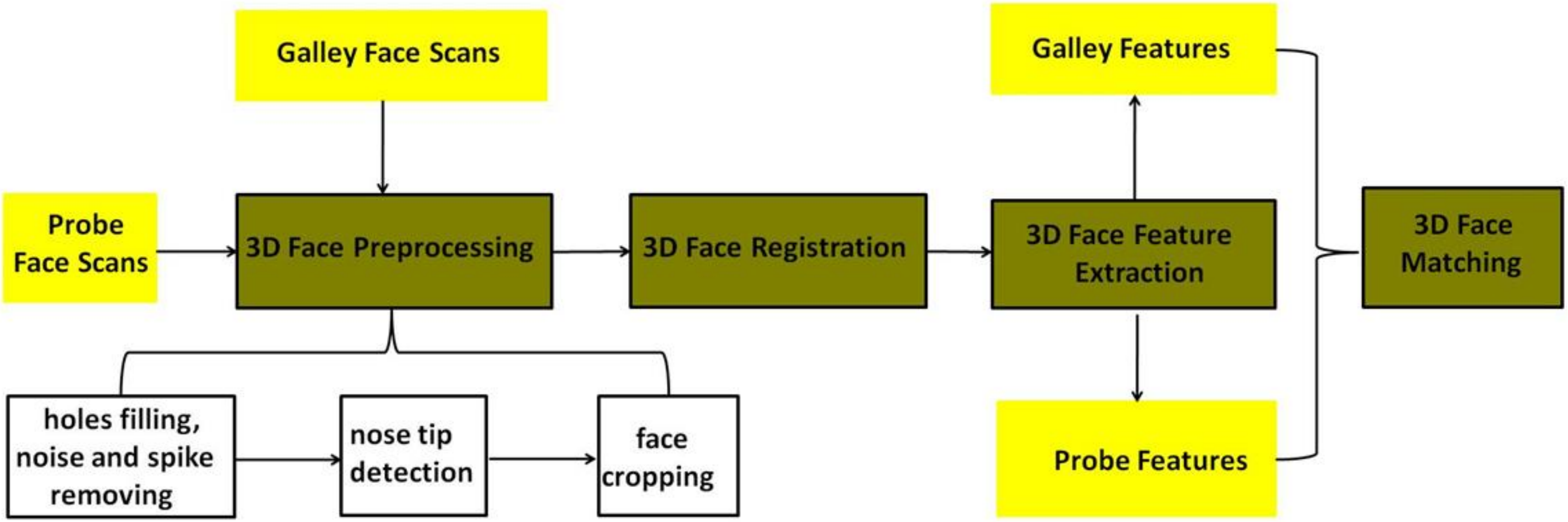
- **structure lighting**: **encoding structure light**
- **Multi-view stereo: computer stereo vision**
- **Photometric stereo: shape from shading**
- **Laser scanner**





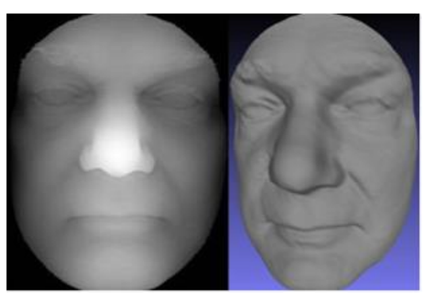


3D face recognition

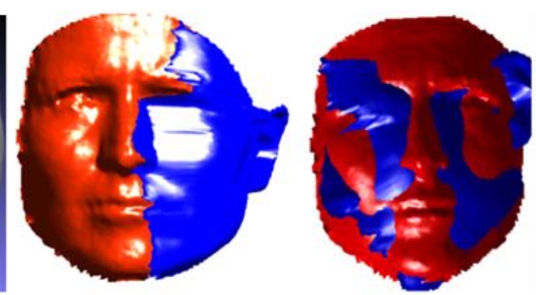


basic processing flow

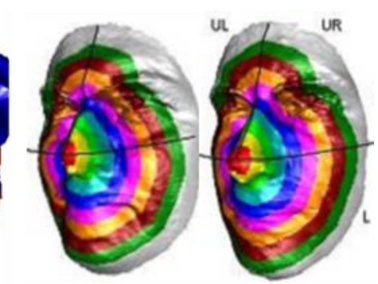
face scan



normalization



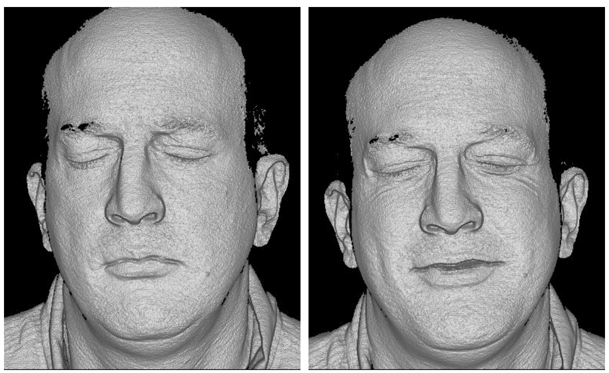
registration

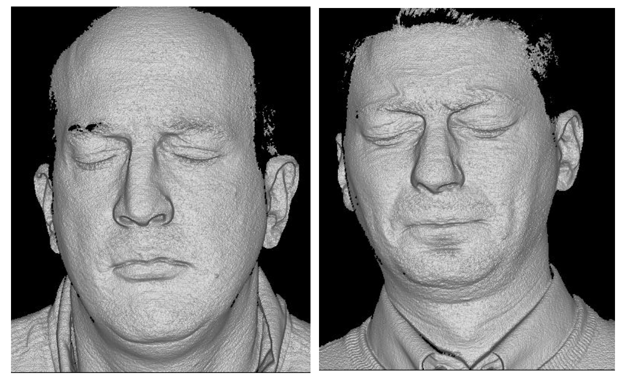


feature

3D face recognition scenario

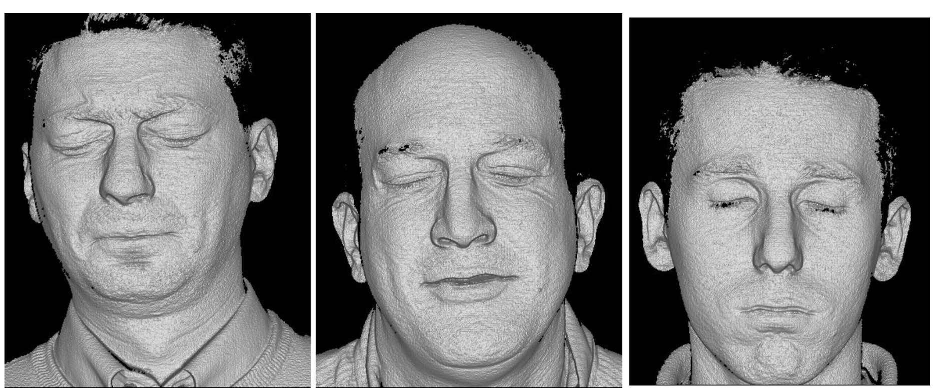
- **Verification (1:1 matching)**



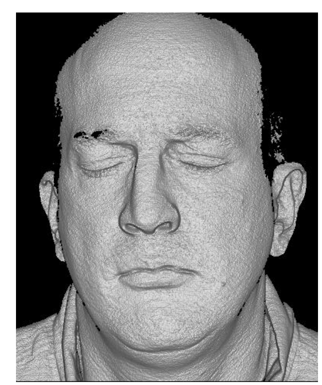


same person?

- **Identification (1:N matching)**



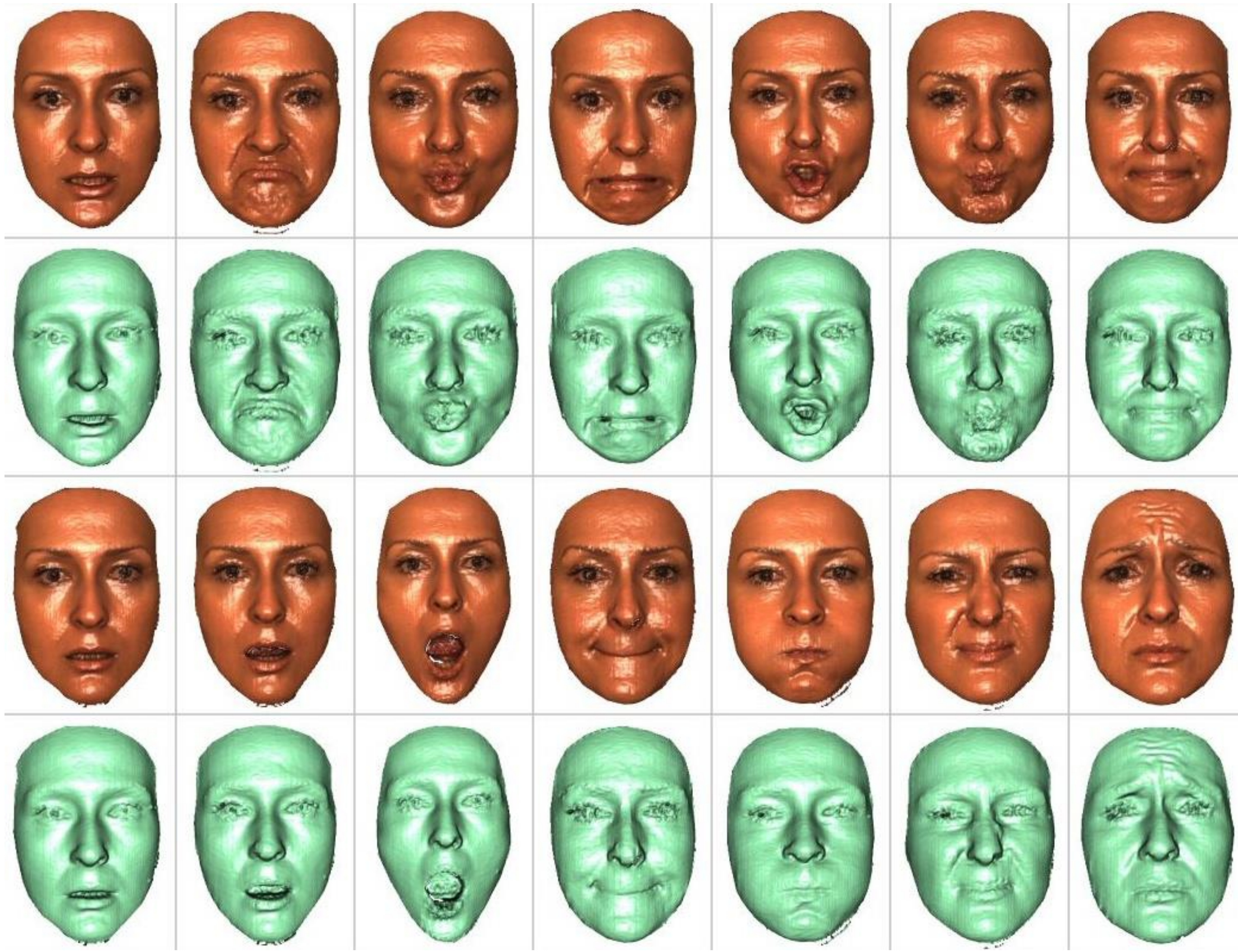
gallery subjects (training set)



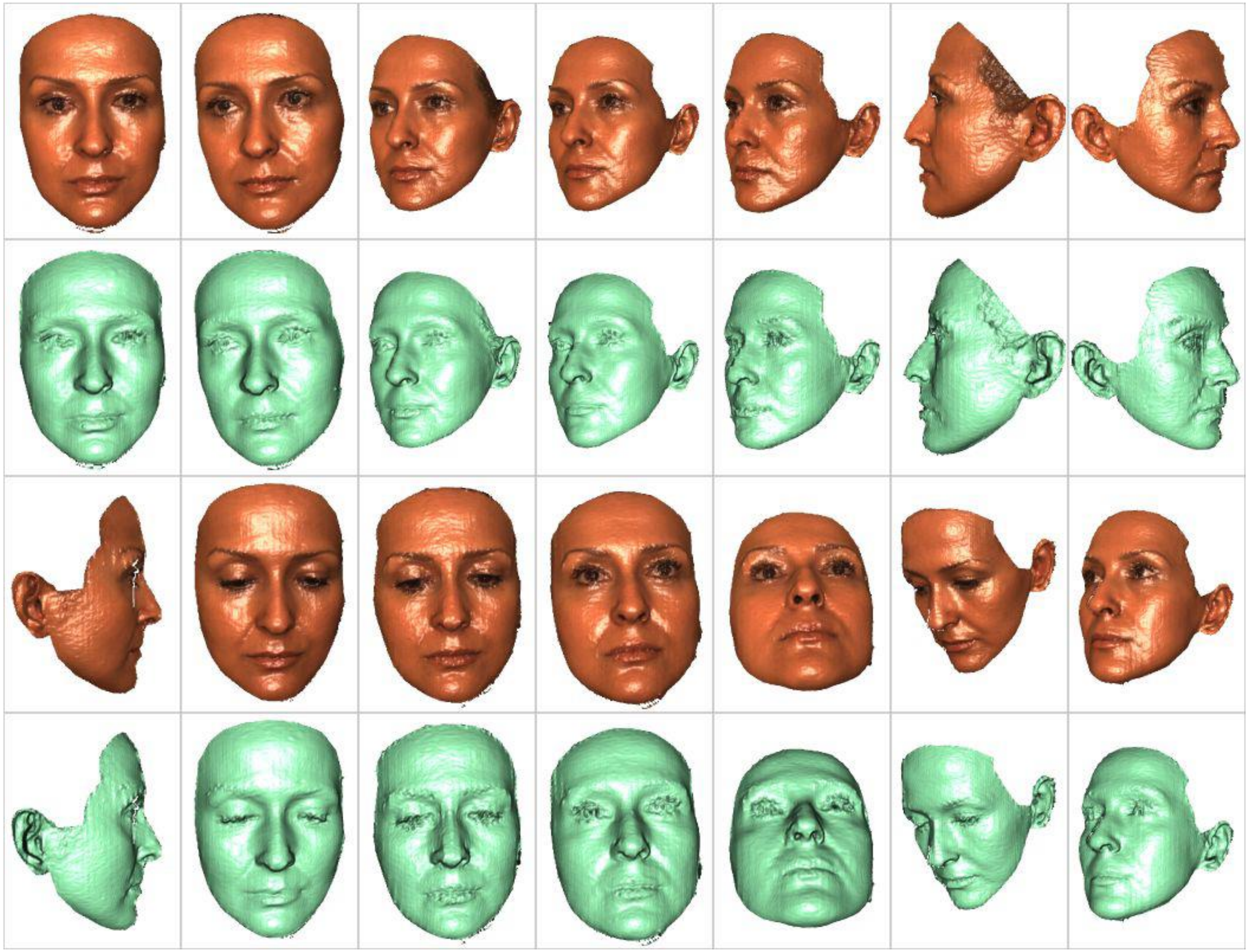
probe (user)

Who he is?

Main challenges: expression variations



Main challenges: pose variations



Main challenges: facial occlusions

Related works

Author	Journal	E	P	O	Registration
1. Samir & Daoudi et al.	PAMI-2006	√	×	×	√
2. Chang & Bowyer et al.	PAMI-2006	√	×	×	√
3. Kakadiaris et al.	PAMI-2007	√	×	×	√
4. Lu & Jian	PAMI-2008	√	×	×	√
5. Mian et al.	PAMI-2007	√	×	×	√
6. Wang et al.	PAMI-2008	√	×	×	√
7. Berretti & Pala et al.	PAMI-2010	√	×	×	√
8. Queirolo et al.	PAMI-2010	√	×	×	√
9. Kakadiaris et al.	PAMI-2011	√	√	×	√
10. Drira & Daoudi et al.	PAMI-2012	√	√	√	√

E: expression, P: pose, O: occlusion

Related works

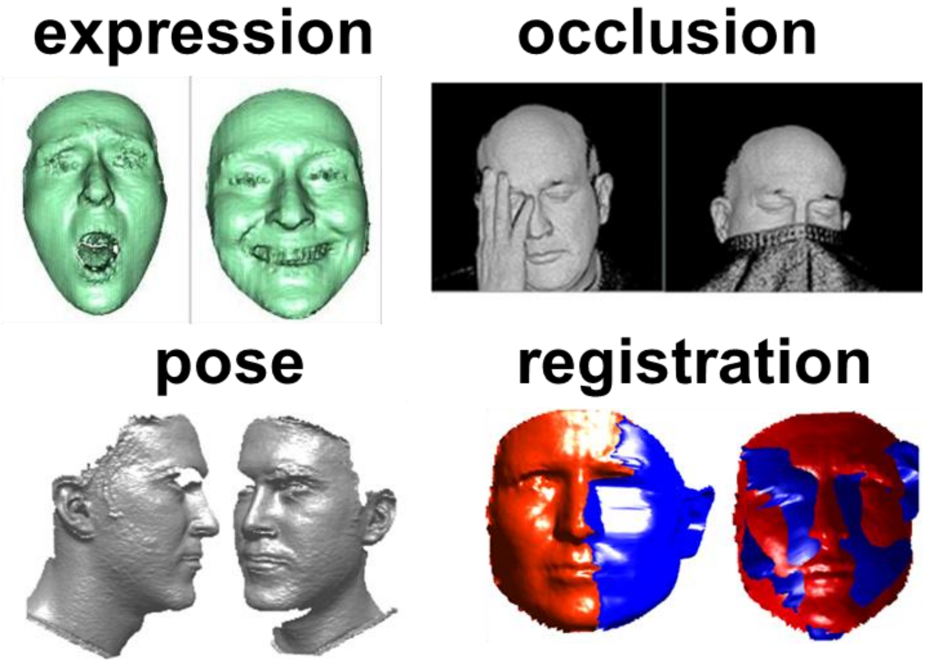
Author	Journal	E	P	O	Registration
11. Bronstein et al.	IJCV-2005	√	×	×	√
12. Mian et al.	IJCV-2008	√	×	×	√
13. Samir & Daoudi et al.	IJCV-2009	√	×	×	√
14. Al-Osaimi & Mian et al.	IJCV-2009	√	×	×	√
15. Spreeuwers	IJCV-2011	√	×	×	√
16. Faltemier et al.	TIFS-2008	√	×	×	√
17. Alyuz & Gokberk et al.	TIFS-2010	√	×	×	√
18. Huang et al.	TIFS-2012	√	√	×	× (near frontal)
19. Berretti & Pala et al.	TIFS-2013	√	√	×	√
20. Alyuz & Gokberk et al.	TIFS-2013	√	×	√	√

E: expression, P: pose, O: occlusion

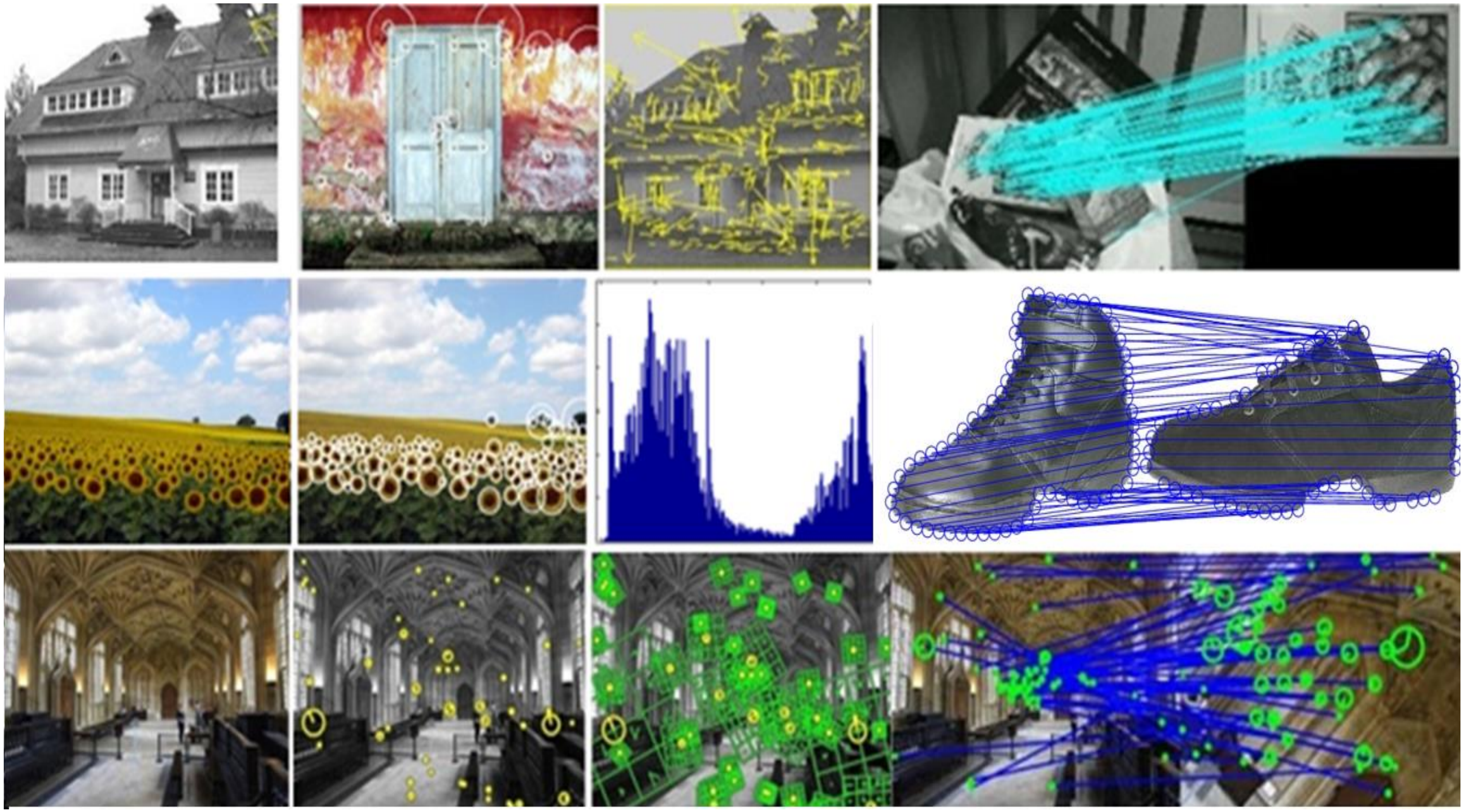
Motivation

Develop a 3D face recognition method which has potential for real biometric applications:

1. **It can deal with expression, pose and occlusion issues in a unified framework.**
2. **It can be fully automatic and totally registration needless.**



SIFT-like matching for 2D images

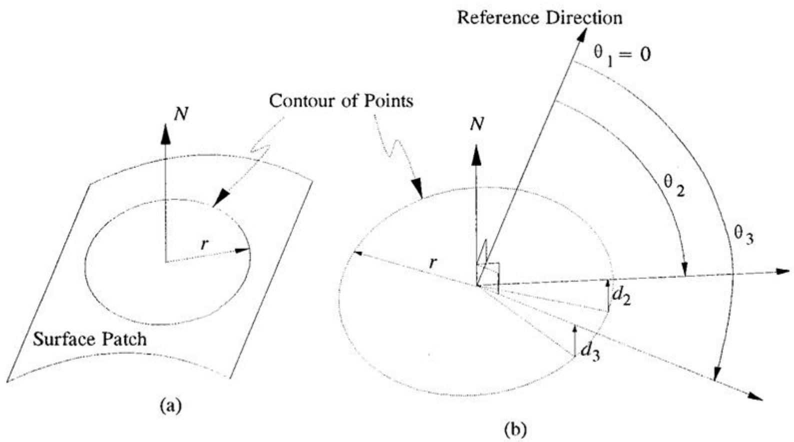


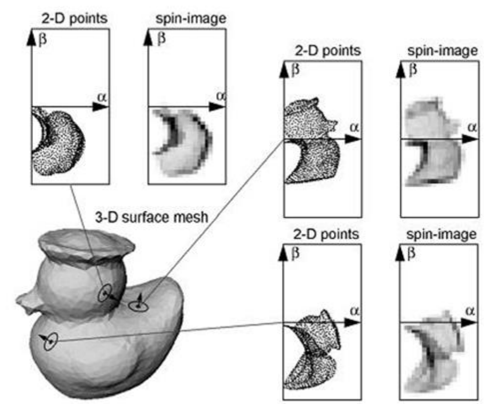
SIFT (SIFT ICCV 1999,IJCV 2004)

keypoint detection, description and matching

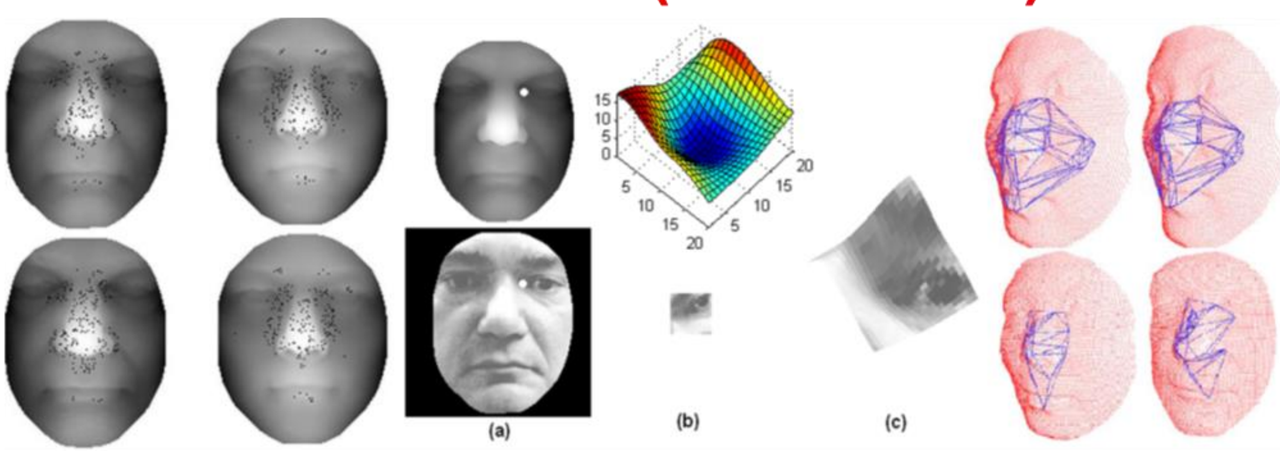
SIFT-like matching for 3D surfaces

Point signature (IJCV 1997), Spin image (PAMI 1999)



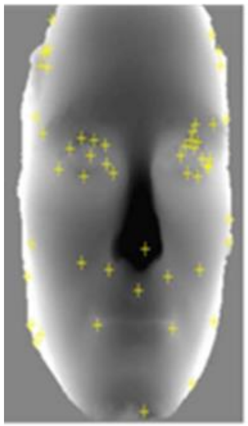


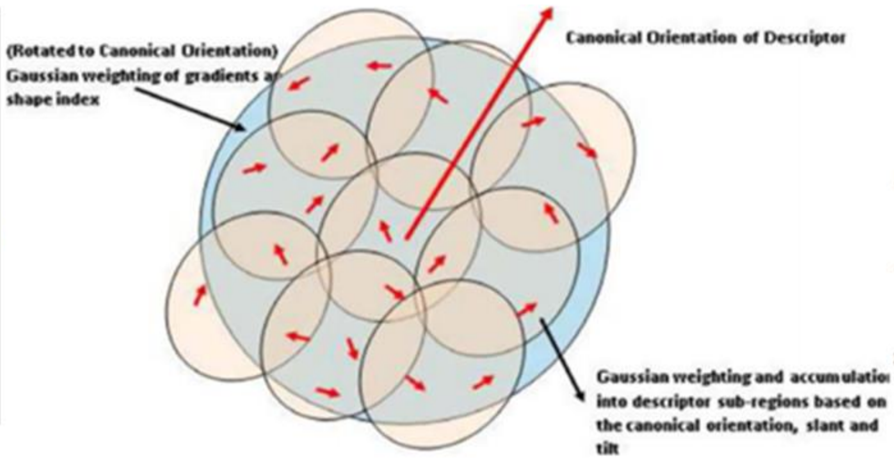
Mian et al. (IJCV 2007)

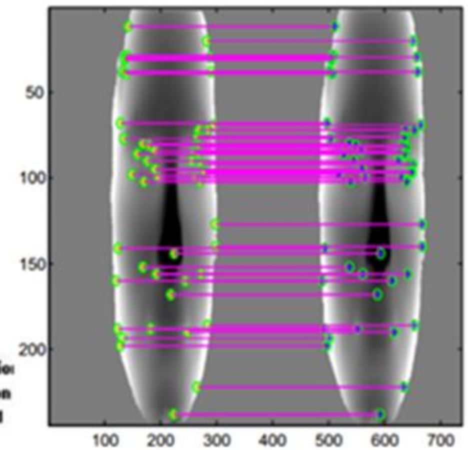


SIFT-like matching for 3D surfaces

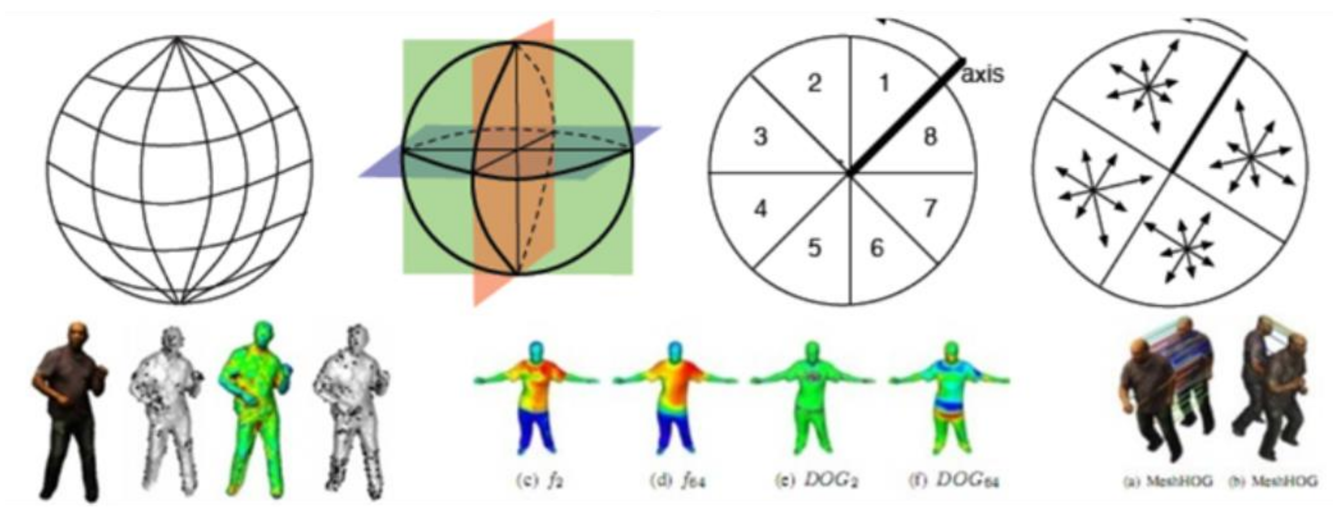
2.5D SIFT (CVIU 2009)





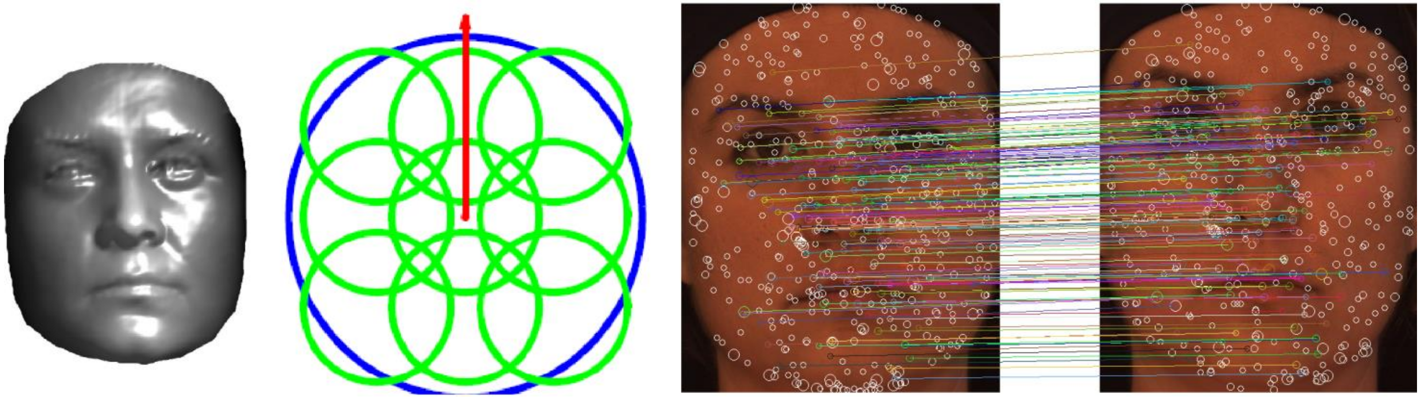


meshHOG (CVPR 2009)

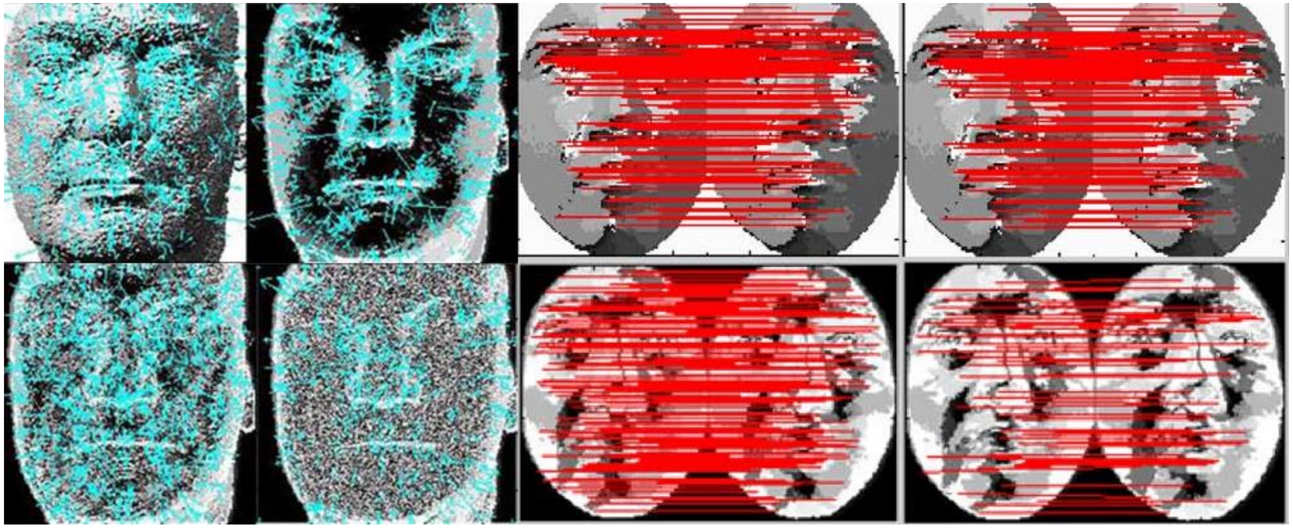


SIFT-like matching for 3D surfaces

meshSIFT (BTAS 2010, CVIU2013)

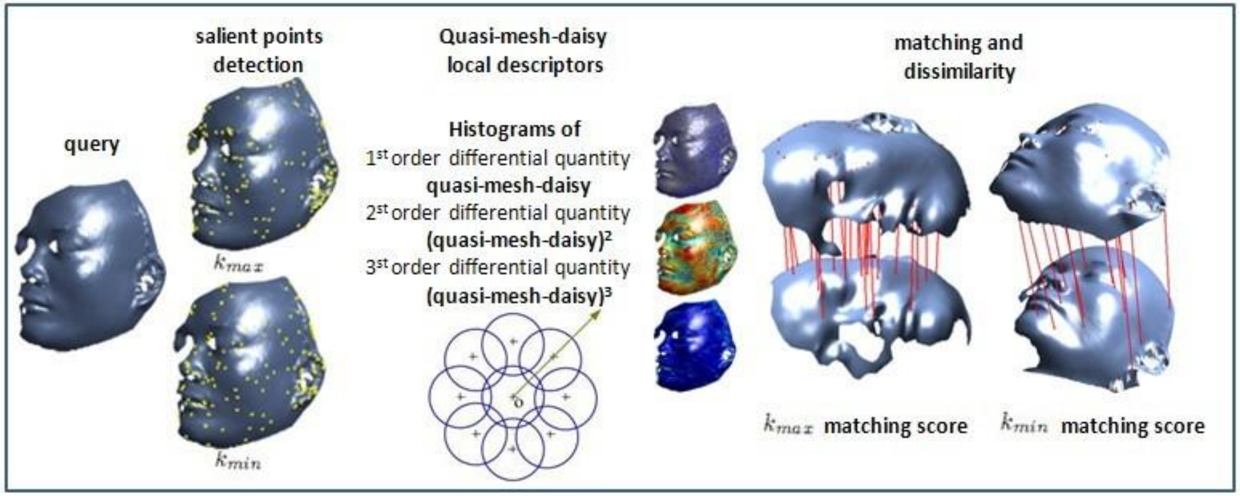


Huang et al. (BTAS 2010, TIFS2012)



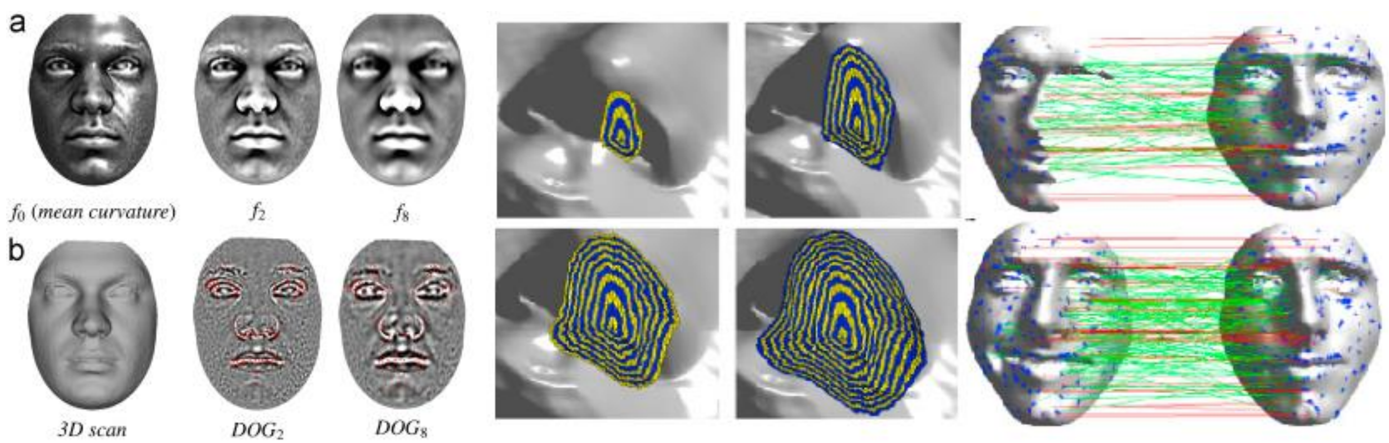
SIFT-like matching for 3D surfaces

Our work (SHREC 2011, ICIP 2011)



Quasi-mesh-daisy local shape descriptor based free-form surface matching framework

Stefano Berretti (Computer & Graphics 2013)



Overview of our approach

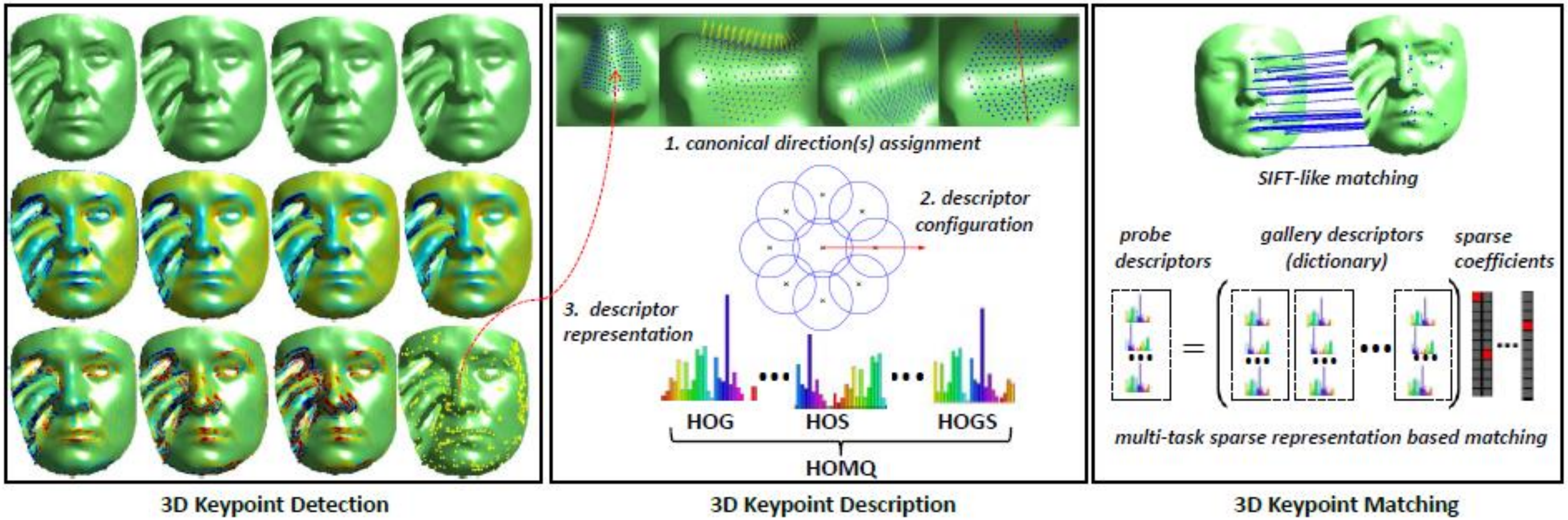


Fig. 1 Overview of the proposed method. 3D keypoint detection: from top to bottom, the original face scan and three smoothed face scans, their corresponding minimum principal curvature, Differences of Gaussian maps, and the detected 3D keypoints. The same procedures are also carried out for maximum principal curvature; 3D keypoint description: canonical direction assignment, quasi-daisy descriptor configuration, and descriptor representation by multi-order surface differential quantities; 3D keypoint matching: SIFT-like coarse-grained matching (*top*) and multi-task sparse representation-based fine-grained matching (*bottom*).

3D keypoint detection

1. Scale-space construction

$$\mathbf{p}_s = \frac{\sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p},1)} w_s(\mathbf{p},\mathbf{q}) \cdot \mathbf{q}}{\sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p},1)} w_s(\mathbf{p},\mathbf{q})}, \tag{1}$$

$$w_s(\mathbf{p},\mathbf{q}) = G_{\sigma_s}(d_e(\mathbf{p},\mathbf{q})) = \exp(-\|\mathbf{p}-\mathbf{q}\|^2/2{\sigma_s}^2). \tag{2}$$

2. Scale-space extrema

$$f(x,y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2 + Dx^3 + Ex^2y + Fxy^2 + Gy^3 \tag{3}$$

$$\rho(C_M,\mathbf{p}_s) = C_M(\mathbf{p}_s) - C_M(\mathbf{p}_{s-1}), \tag{4}$$

$$\rho(C_m,\mathbf{p}_s) = C_m(\mathbf{p}_s) - C_m(\mathbf{p}_{s-1}). \tag{5}$$

3D keypoint detection

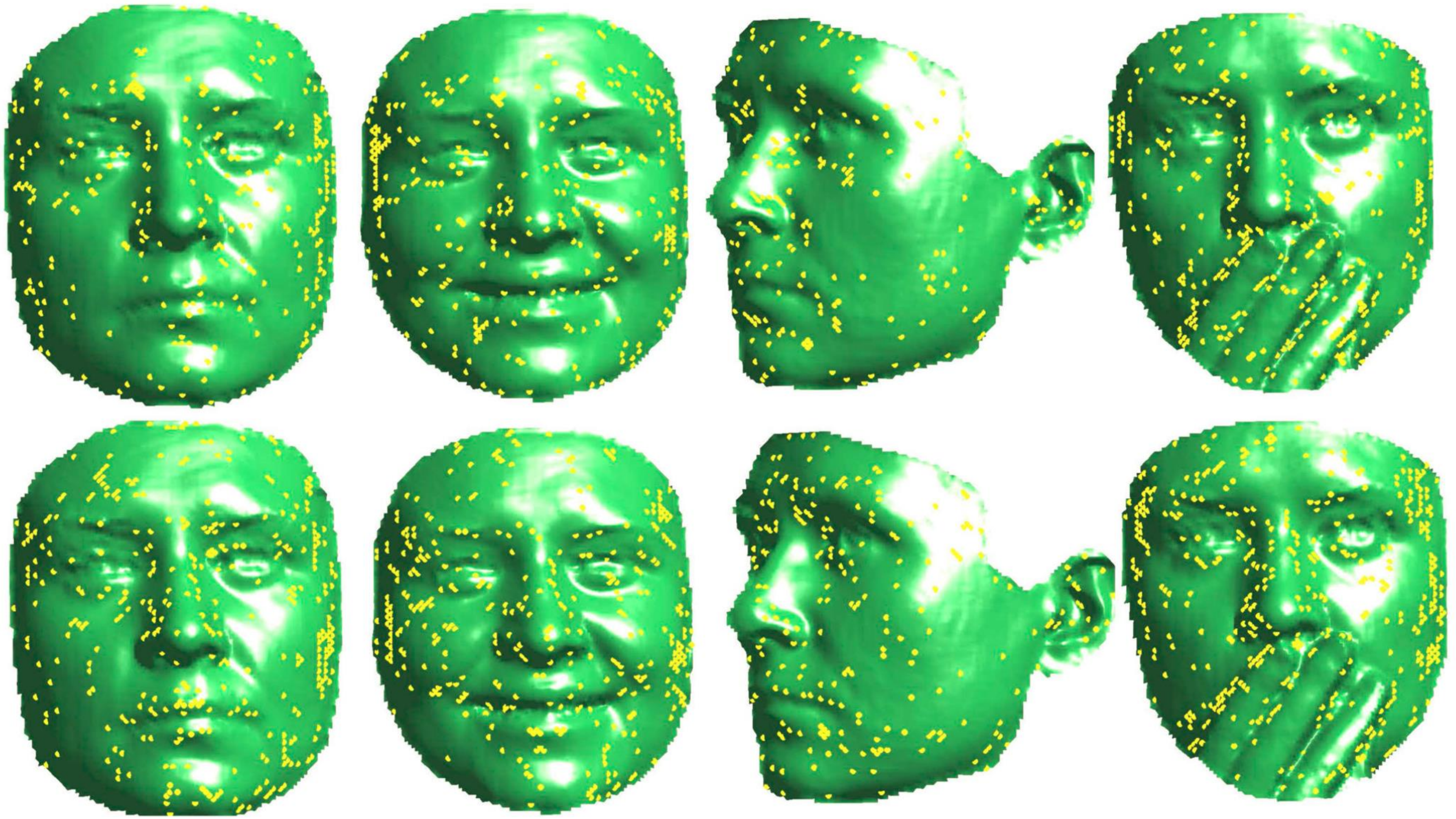
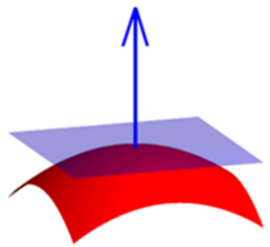
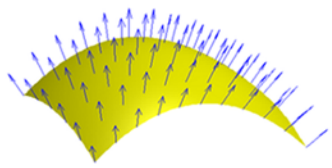


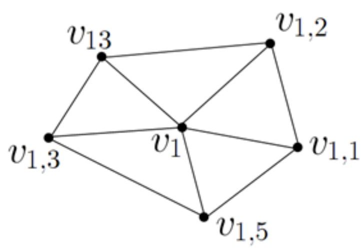
Fig. 2 The detected 3D keypoints by C_M (*top*) and C_m (*bottom*) for a neutral face, a happy face, a 45° rotated face, and a face with mouth occlusion

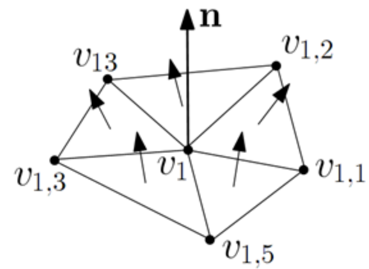
Multi-order Surface Differential Quantities

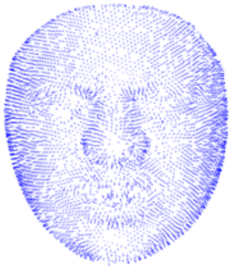
- 1-order surface normal: direction information





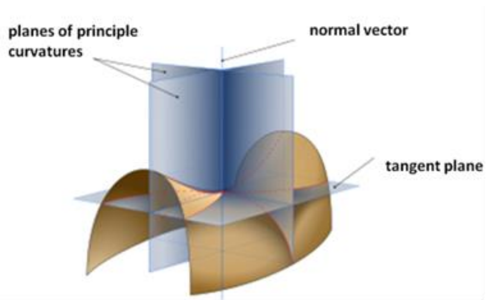


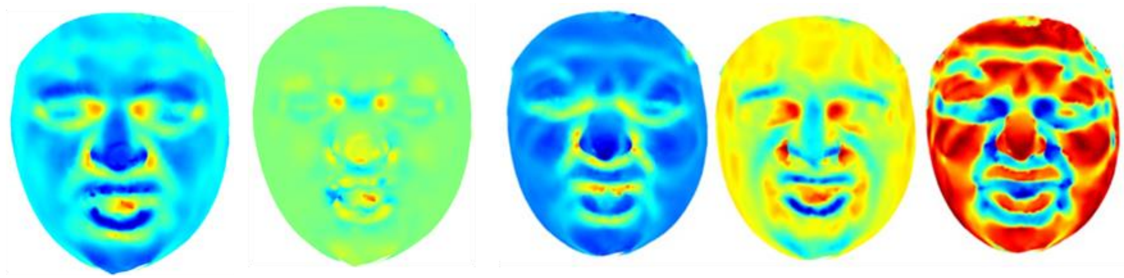




normal

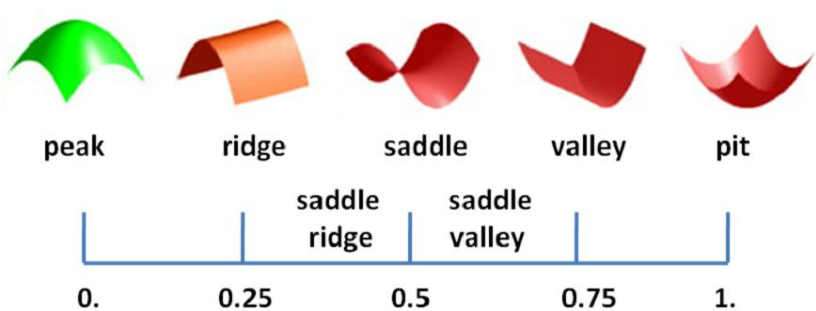
- 2-order curvatures: local shape bending information



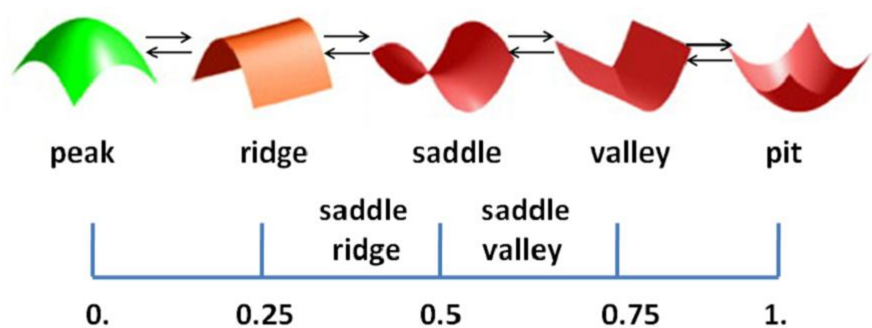


mean Gaussian maximum minimum shape index

- 2-order Shape index

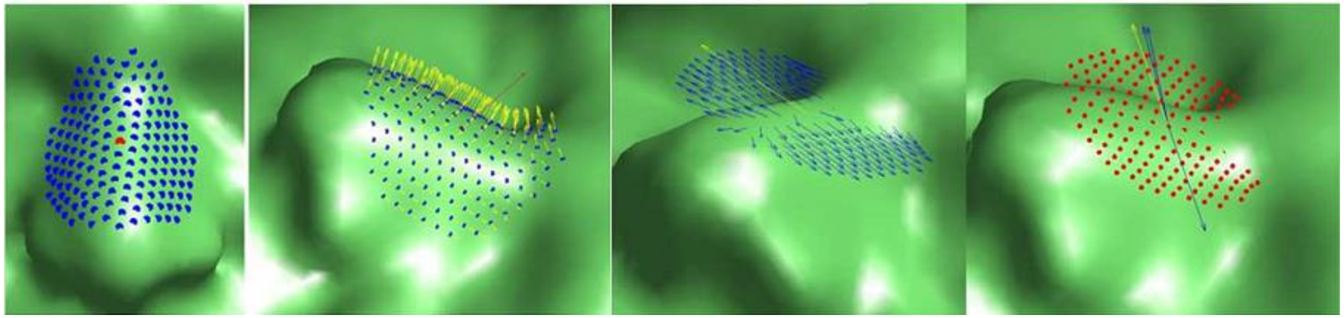


- 3-order shape variation information

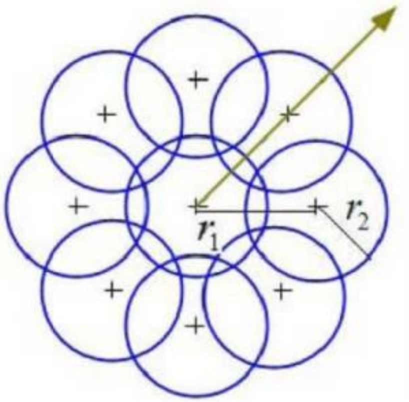


3D keypoint description

1. Canonical direction assignment



2. Spatial configuration

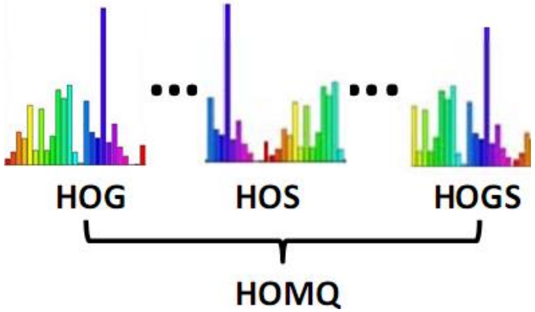


$= (hog_1, hog_2, \ldots, hog_9),$

$= (hos_1, hos_2, \ldots, hos_9),$

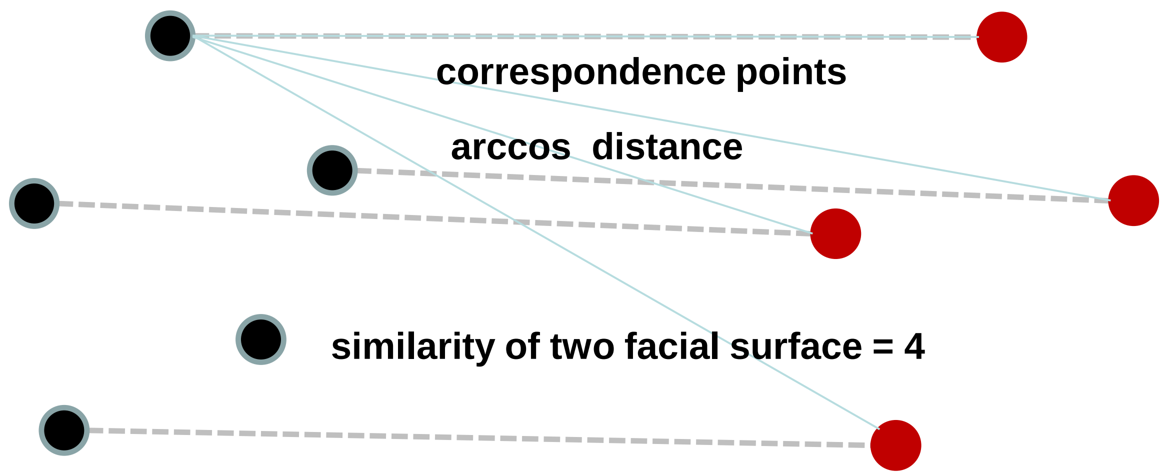
$= (hogs_1, hogs_2, \ldots, hogs_9),$

3. Differential quantities statistics

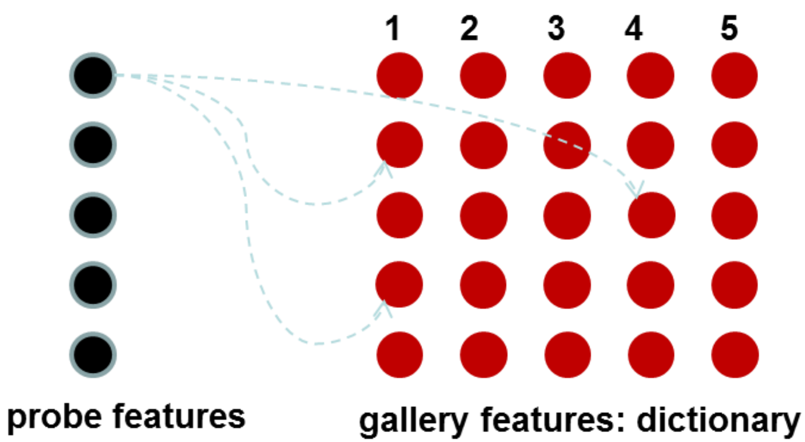


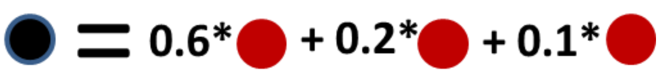
3D keypoint matching

- **Coarse Grained Matcher (CGM): SIFT-like matcher**



- **Fine Grained Matcher (FGM): SR-like matcher**





subject based reconstruction error

$$\mathbf{e}_1 = [0.12 \;\; 1.00 \;\; 1.00 \;\; 0.88 \;\; 1.00]^{\mathrm{T}}$$

$$\mathbf{e}_2 =, \cdots, \mathbf{e}_5 =$$

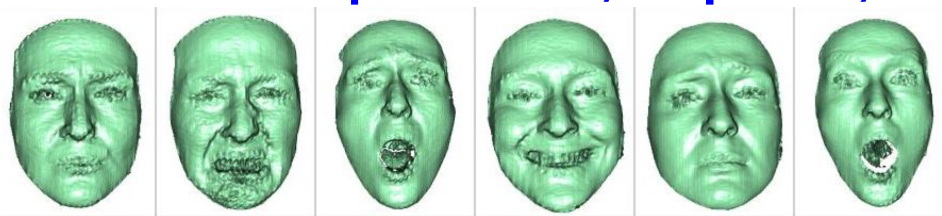
Similarity: average reconstruction error

$$\bar{\mathbf{e}} = [0.15 \;\; 0.98 \;\; 0.95 \;\; 0.80 \;\; 0.92]^{\mathrm{T}}$$

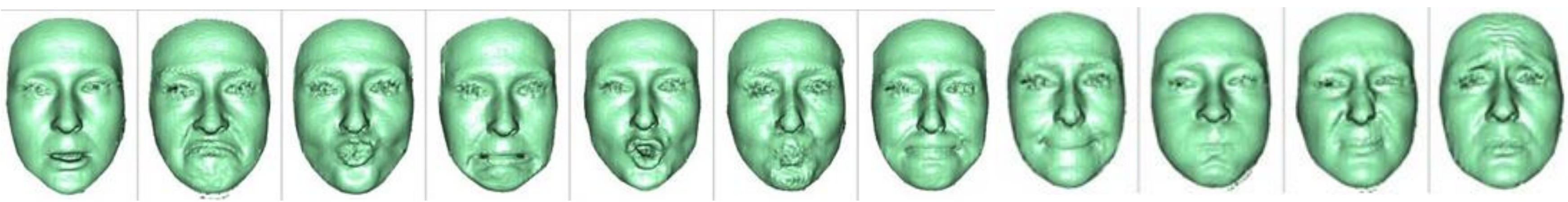
3D keypoint matching

Dataset and evaluation protocol

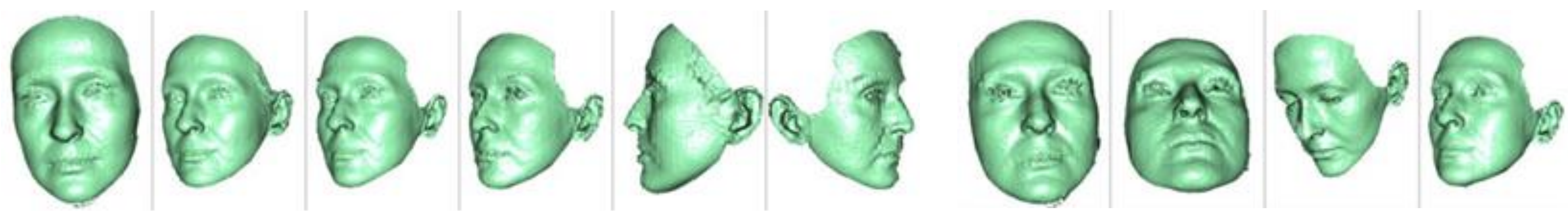
Bosphorus 3D Face Database: **4666 3D scans of 105 subjects, around 34 expressions, 13 poses, and 4 occlusions for each subject**



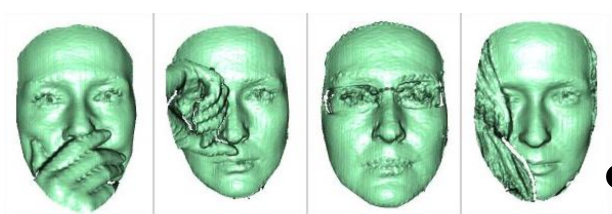
- **Basic expressions neutral, anger, disgust, fear, happy, sad, and surprise**



- **lower, upper and combined action units action units**



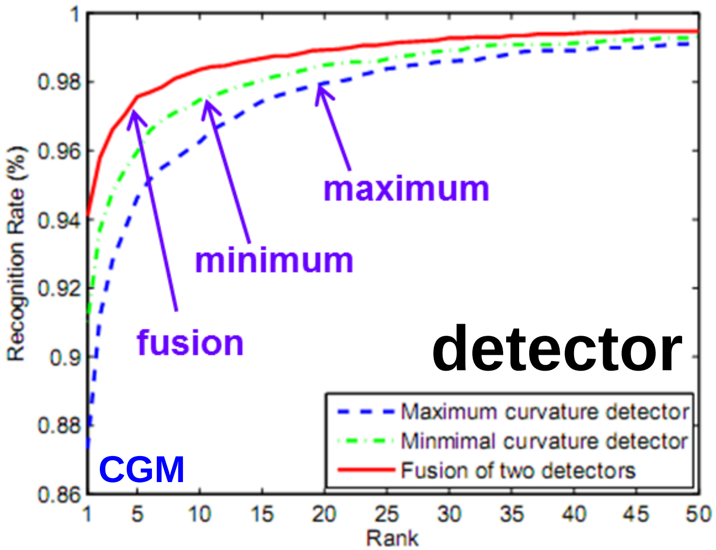
- **yaw rotations of 10, 20, 30, 45, and 90 degrees, pitch rotation, cross rotations**

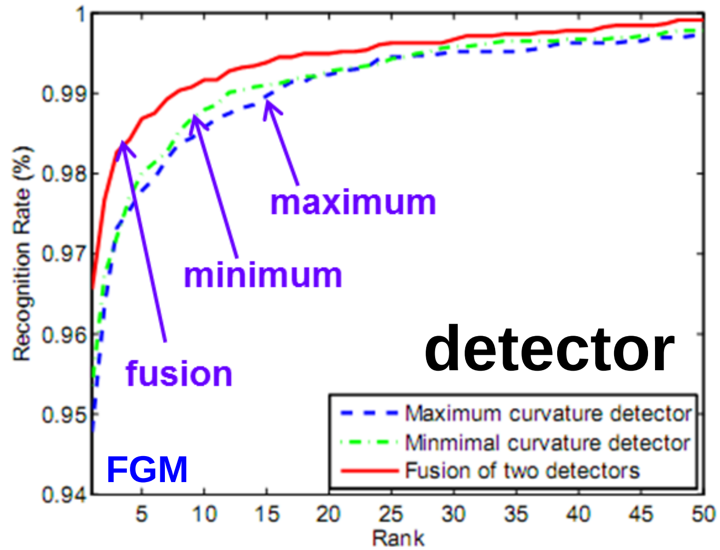


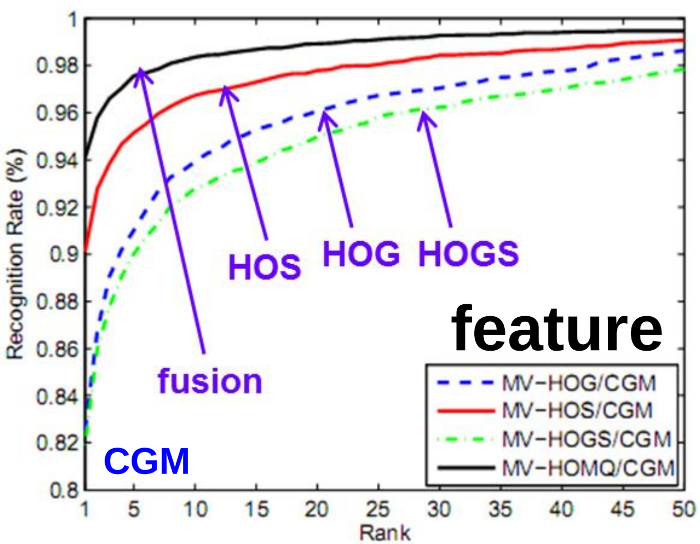
- **occlusions**

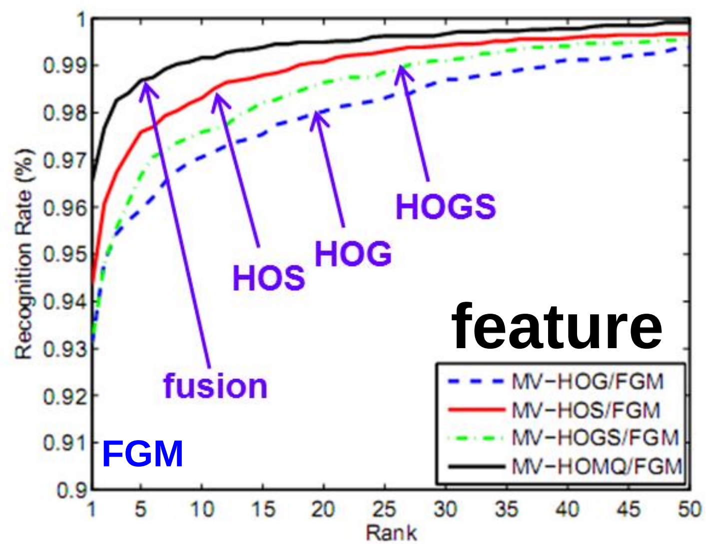
- **Gallery: first 105 neutral scans, Probe: other scans**

Experimental results: fusion









Experimental results: expression subset

Table 1 Performance in terms of rank-one recognition rates on the subsets of expressions, poses, occlusions, unlabeled, and the entire Bosphorus database

	HOG (%)		HOS (%)		HOGS (%)		HOMQ (%)	
	CGM	FGM	CGM	FGM	CGM	FGM	CGM	FGM
Neutral (105 scans) vs. expressions (2,797 scans)								
Neutral (194)	99.48	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Anger (71)	69.01	91.55	87.32	98.59	76.06	94.37	88.73	97.18
Disgust (69)	50.72	84.06	56.52	78.26	60.87	88.41	76.81	86.96
Fear (70)	71.43	92.86	88.57	92.86	84.29	94.29	92.86	98.57
Happiness (106)	79.25	91.51	85.85	96.23	75.47	98.11	95.28	98.11
Sadness (66)	80.30	95.45	90.91	96.97	86.36	96.97	95.45	100.0
Surprise (71)	81.69	100.0	97.18	100.0	84.51	97.18	98.59	98.59
LAU (1,549)	88.96	96.97	95.09	98.13	90.70	98.52	97.22	98.84
UAU (432)	94.21	99.07	98.15	100.0	95.14	99.54	99.07	100.0
CAU (169)	92.90	100.0	97.04	99.41	95.86	100.0	98.82	100.0
All (2,797)	88.09	96.96	94.32	97.96	90.24	98.32	96.89	98.82

Experimental results: pose subset

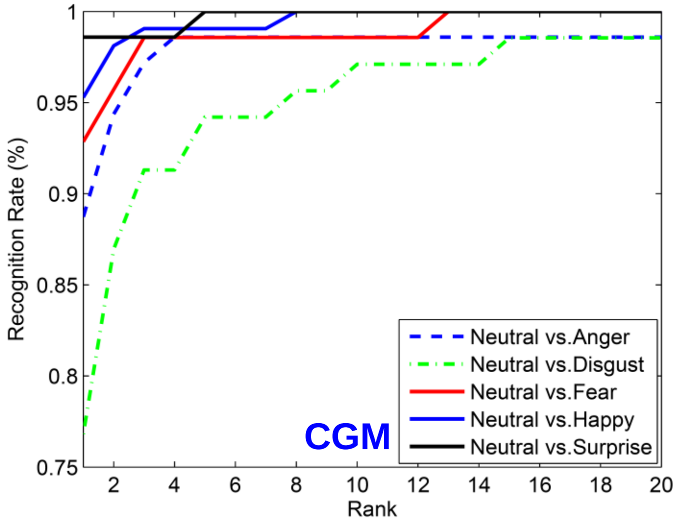
Table 1 Performance in terms of rank-one recognition rates on the subsets of expressions, poses, occlusions, unlabeled, and the entire Bosphorus database

	HOG (%)		HOS (%)		HOGS (%)		HOMQ (%)	
	CGM	FGM	CGM	FGM	CGM	FGM	CGM	FGM
Neutral (105 scans) vs. poses (1,365 scans)								
YR10 (105)	98.10	100.0	100.0	100.0	96.19	100.0	100.0	100.0
YR20 (105)	90.48	99.05	96.19	99.05	87.62	98.10	99.05	100.0
YR30 (105)	84.76	98.10	92.38	100.0	75.24	96.19	98.10	99.05
YR45 (210)	49.52	89.05	80.48	91.90	49.05	84.76	90.95	97.62
YR90 (210)	07.62	16.67	17.62	25.24	02.86	10.00	33.33	47.14
YR (735)	55.37	72.65	69.25	76.19	51.84	71.43	77.96	84.08
PR (419)	94.27	99.28	96.90	98.57	84.73	99.28	98.81	99.52
CR (211)	60.66	90.05	79.62	94.79	48.34	89.10	94.31	99.05
All (1,365)	68.13	83.52	79.34	85.93	61.39	81.47	86.89	91.14

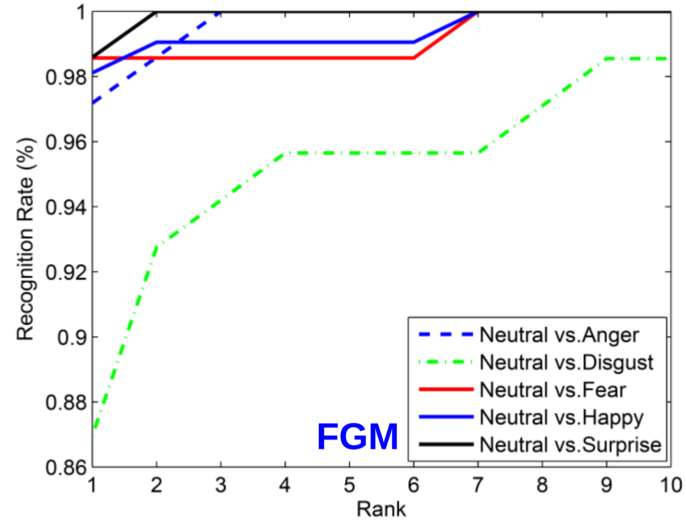





Experimental results: CMC curves

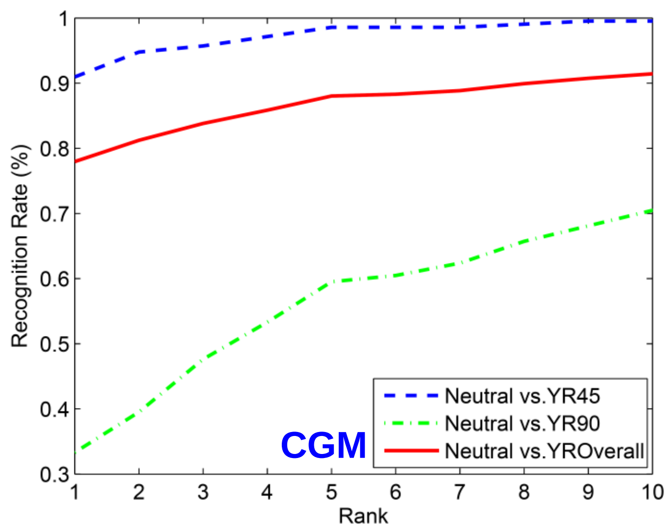


(a)

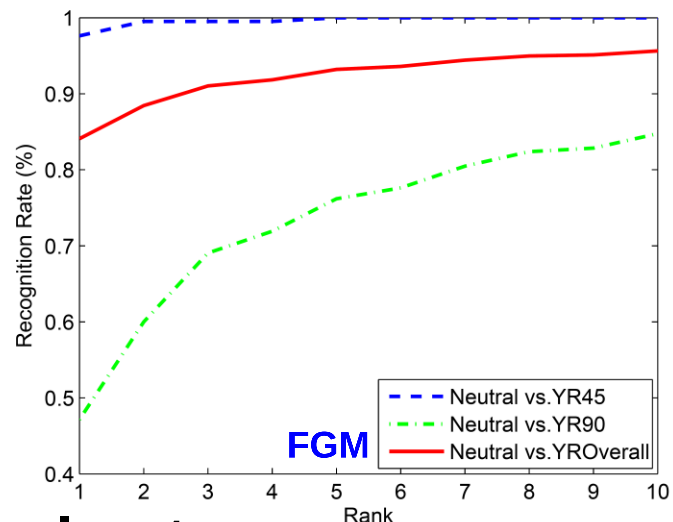


(b)

expression subset



(c)



(d)

pose subset

Experimental results: occlusion subset

Table 1 Performance in terms of rank-one recognition rates on the subsets of expressions, poses, occlusions, unlabeled, and the entire Bosphorus database

	HOG (%)		HOS (%)		HOGS (%)		HOMQ (%)	
	CGM	FGM	CGM	FGM	CGM	FGM	CGM	FGM
Neutral (105 scans) vs. occlusions (381 scans)								
E-Hand (105)	96.19	100.0	100.0	100.0	96.19	100.0	100.0	100.0
M-Hand (105)	90.48	100.0	97.14	98.10	93.33	99.05	99.05	100.0
E-Glasses (104)	95.19	100.0	97.12	97.12	97.12	97.12	100.0	100.0
F-Hair (67)	86.57	94.03	94.03	91.04	86.57	82.09	97.01	95.52
All (381)	92.65	98.95	97.38	97.90	93.96	96.59	99.21	99.21
Neutral (105 scans) vs. unlabeled scans (18 scans)								
All (18)	88.89	100.0	100.0	100.0	94.44	100.0	100.0	100.0
Neutral (105 scans) vs. all meshes except yaw 90 (4,351 scans)								
All (4,351)	86.12	96.81	93.61	97.70	85.75	97.15	97.04	98.94
Neutral (105 scans) vs. all scans (4,561 scans)								
All (4,561)	82.50	93.12	90.11	94.37	81.93	93.14	94.10	96.56

Four proposed keypoint descriptors (HOG, HOS, HOGS, and HOMQ) combined with two keypoint matchers (CGM and FGM) are compared with each other. The names and the number of probe scans for each subset are listed in the left-hand column

whole dataset

Experimental results: comparisons

Table 2 Comparison of rank-one recognition rates on the subsets of expressions, poses, occlusions, and the entire Bosphorus database.

best rate!

	Expressions	Poses	YR	YR90°	PR	CR	Occlusions	E-Hand	M-Hand	E-Glasses	F-Hair	All scans
Alyüz et al. (2008)	-	-	-	-	-	-	93.6%	93.6%	93.6%	97.8%	89.6%	-
Alyüz et al. (2010)	98.2%	-	-	-	-	-	-	-	-	-	-	-
Colombo et al. (2011)	-	-	-	-	-	-	87.6%	91.1%	74.7%	94.2%	90.4%	-
Ocegueda et al. (2011)	98.2%	-	-	-	-	-	-	-	-	-	-	-
Drira et al. (2013)	-	-	-	-	-	-	87.0%	97.1%	78.0%	94.2%	81.0%	-
Smeets et al. (2013)	97.7%	84.2%	-	24.3%	-	-	-	-	-	-	-	93.7%
Berretti et al. (2013)	95.7%	88.6%	81.6%	45.7%	98.3%	93.4%	93.2%	-	-	-	-	93.4%
This paper	98.8%	91.1%	84.1%	47.1%	99.5%	99.1%	99.2%	100%	100%	100%	95.5%	96.6%

Experimental results: FRGC v2.0 database

Table 14 Performance comparison on the whole FRGC v2.0 database.

Method	Year	Data format	Face alignment	Rank-one score
Chang *et al.* [14]	2006	Range	Yes	91.9%
Kakadiaris *et al.* [29]	2007	Range	Yes	91.9%
Mian *et al.* [38]	2007	Range	Yes	96.2%
Mian *et al.* [39]	2008	Range	Yes	93.5%
Faltemier *et al.* [19]	2008	Range	Yes	**97.2%**
Alosaimi *et al.* [2]	2009	Range	Yes	**96.5%**
Alyuz *et al.* [3]	2010	Range	Yes	**97.5%**
Queirolo *et al.* [46]	2010	Range	Yes	**98.4%**
Wang *et al.* [61]	2010	Range	Yes	**98.3%**
Huang *et al.* [25]	2011	Range	Yes	**97.2%**
Spreeuwers [54]	2011	Range	Yes	**99.0%**
Li *et al.* [30]	2011	Range	Yes	96.3%
Ballihi *et al.* [6]	2012	Mesh	Yes	98.2%
Smeets *et al.* [53]	2013	Mesh	**No**	89.6%
MV-HOMQ/FGM	2013	Mesh	**No**	96.3%

Discussion and future work

1. 3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, Jianwei Wan. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI): 2014, 36(11), 2270-2287
2. A Comprehensive Performance Evaluation of 3D Local Feature Descriptors, Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, Jianwei Wan, Ngai Ming Kwok. International Journal of Computer Vision (IJCV):2016 ,116(1),66-89
3. Rotational Projection Statistics for 3D Local Surface Description and Object Recognition, Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Min Lu, Jianwei Wan. International Journal of Computer Vision (IJCV): 2013,105(1) ,63-86
4. Performance Evaluation of 3D Keypoint Detectors, Tombari, Federicoand Salti, Samueleand Di Stefano, Luigi, International Journal of Computer Vision (IJCV): 2013, 102:198

Deep learning on manifolds and non-Euclidean domains

5. Geodesic convolutional neural networks on Riemannian manifolds, Jonathan Masci,Davide Boscaini,Michael M. Bronstein,Pierre Vandergheynst, ICCV, 2015.

References and code

1. Huibin Li, Di Huang, Jean-Marie Morvan, Yunhong Wang, Liming Chen, Towards 3D Face Recognition, A Registration-Free Approach using Fine-Grained Matching of 3D Keypoint Descriptors, International Journal of Computer Vision (**IJCV**), 113(2): 128-142, 2015.

2. Huibin Li, Di Huang, Pierre Lemaire, Jean-Marie Morvan, Liming Chen: Expression-robust 3D Face Recognition via Mesh-based Histograms of Multiple-order Surface Differential Quantities, IEEE International Conference on Image Processing (**ICIP**), pp. 3053-3056, Brussels, Belgium, 2011.

3. Remco C.Veltkamp, Stefan van Jole, Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, Huibin Li, Liming Chen, Peter Claes, Dirk Smeets, Jeroen Hermans, Dirk Vandermeulen, Paul Suetens. SHREC’11 Track: 3D Face Model Retrieval. Euro-graphics Workshop on 3D Object Retrieval (3DOR), page: 89-95, Llandudno, UK, 2011.

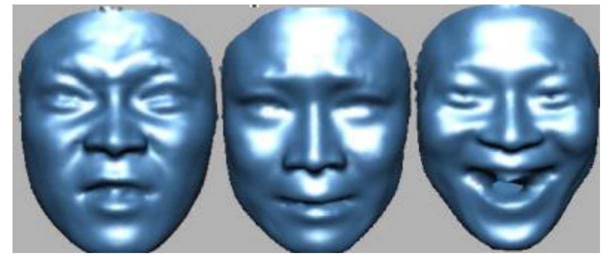
Code and demo: http://gr.xjtu.edu.cn/web/huibinli/code/toolbox-FGM-3DKD.rar

Facial expression recognition (FRE)

- **Data modality:**

visible image

infrared image



3D face scan

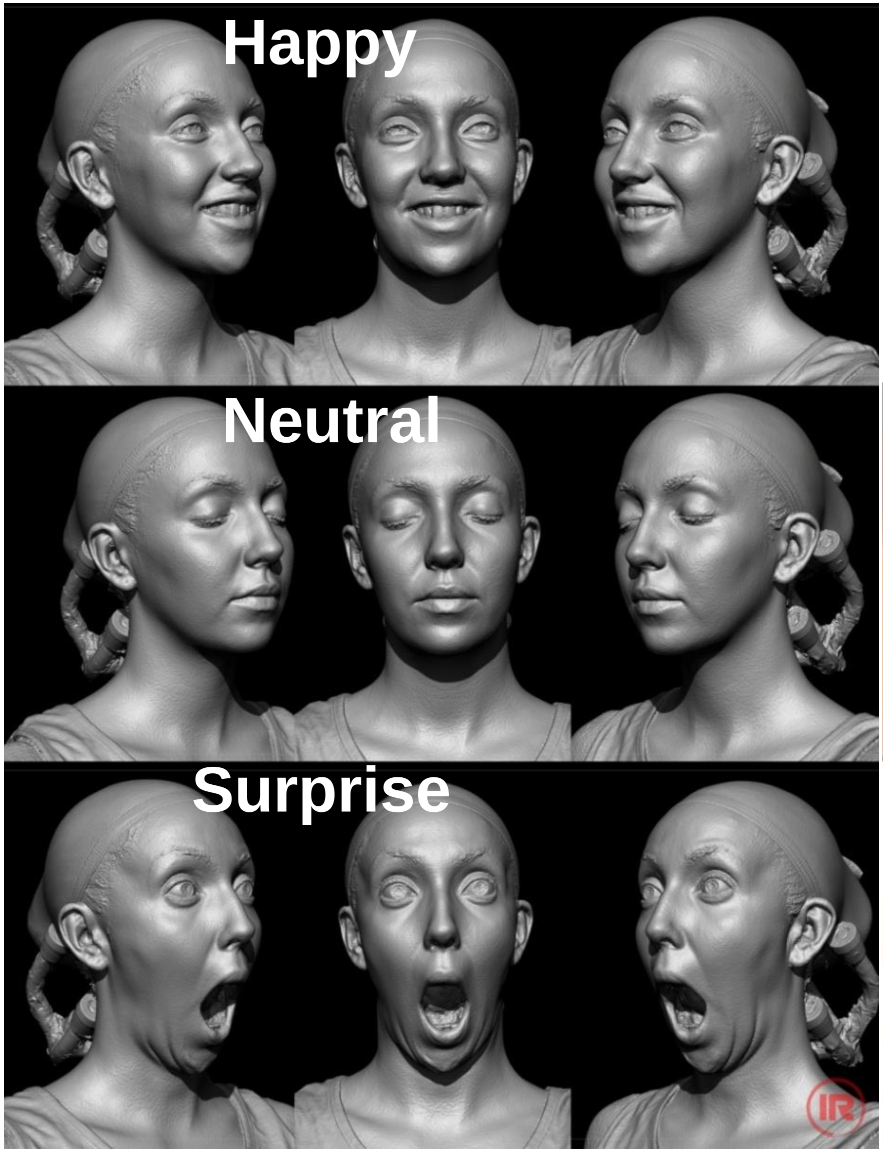
- **Emotion granularity:**

Action unit detection

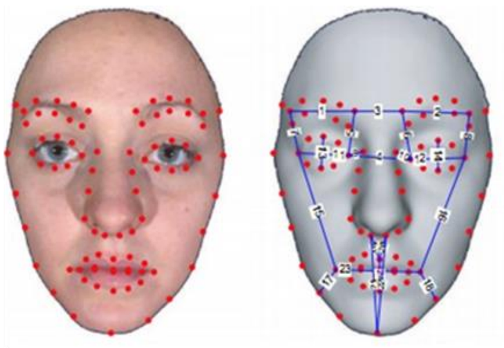
Basic emotion classification

- **Spontaneity:** posed and un-posed (spontaneous) expressions
- **Expression Intensity:** micro-expression, intensity estimation
- **Temporal dynamics:** video-based, frame-based

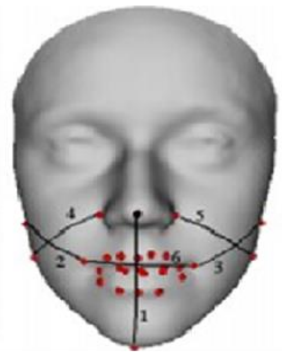
This paper: 2D+3D FER, basic emotions, static data



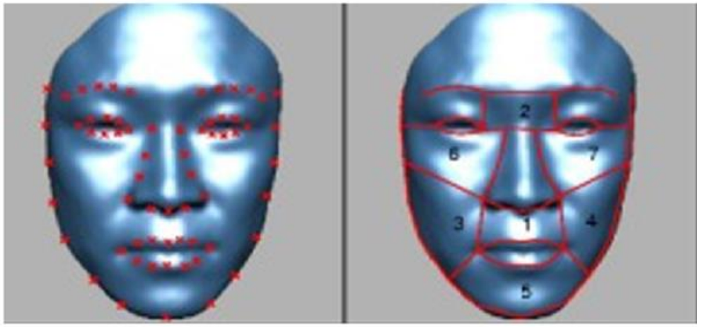
Motivation: hand-crafted v.s. learning-based FER



83 facial points
[Tang et al. CVPR-2008]



85 facial points
[Berretti et al. ICPR-2010]



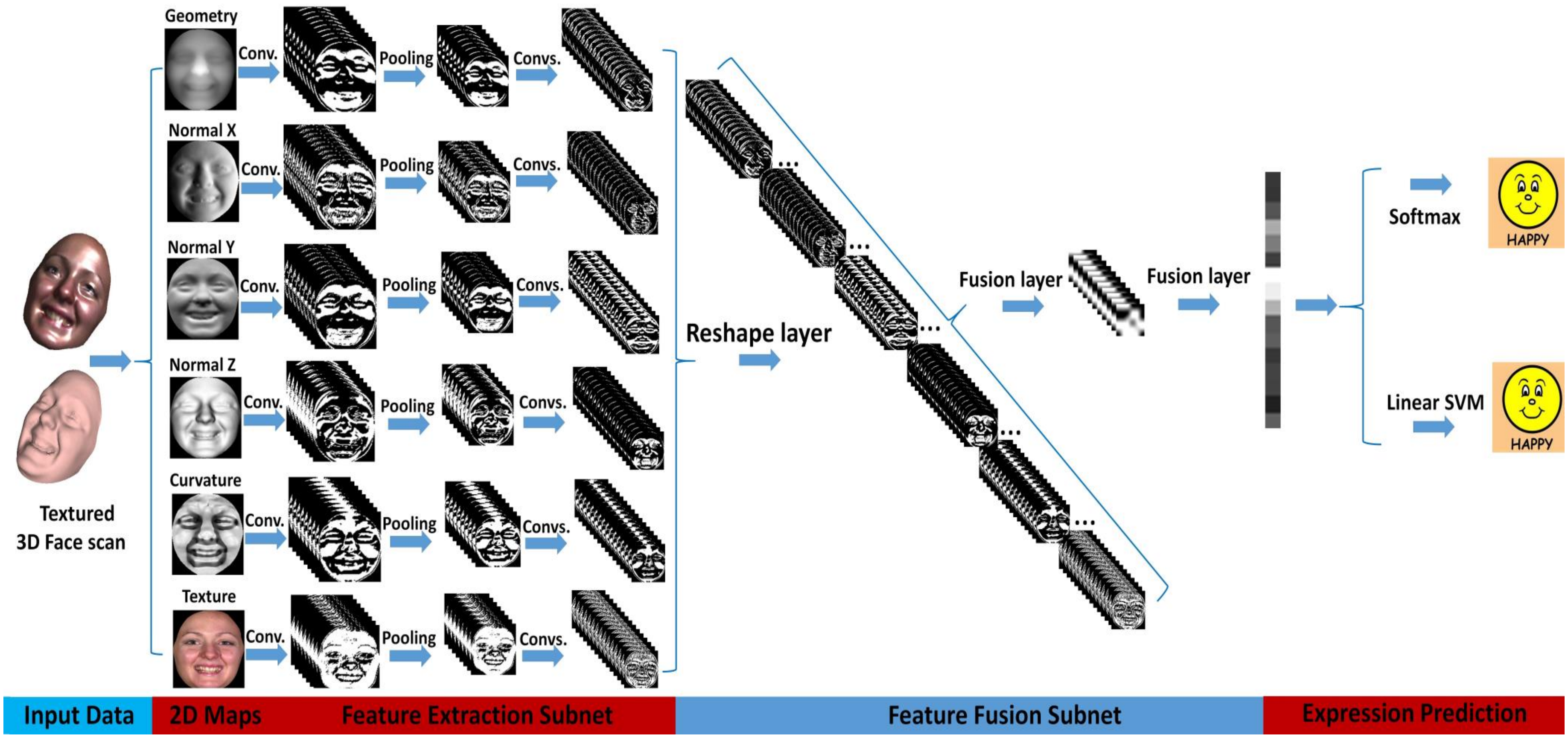
64 facial points and 7 regions
[Wang et al. CVPR-2006]

	Handcrafted features	Learned Features
2D FER	HOG: Hu et al. [7] LBP: Zhao et al. [64] Gabor: Zhang et al. [63]	Deep CNN: Khorrami et al. [20] DBN: Liu et al. [30] Auto-Encoder: Rifai et al. [38]
3D FER	Depth-SIFT: Berretti et al. [1] Normal-LBP: Li et al. [23] Curvature-HOG: Lemaire et al. [22]	Learned feature for 3D FER?
2D+3D FER	Handcrafted feature-level fusion Handcrafted score-level fusion Savran et al. [41], Li et al. [24]	Learning-based fusion for 2D+3D FER?

There are very limited numbers of 3D faces with expression labels.

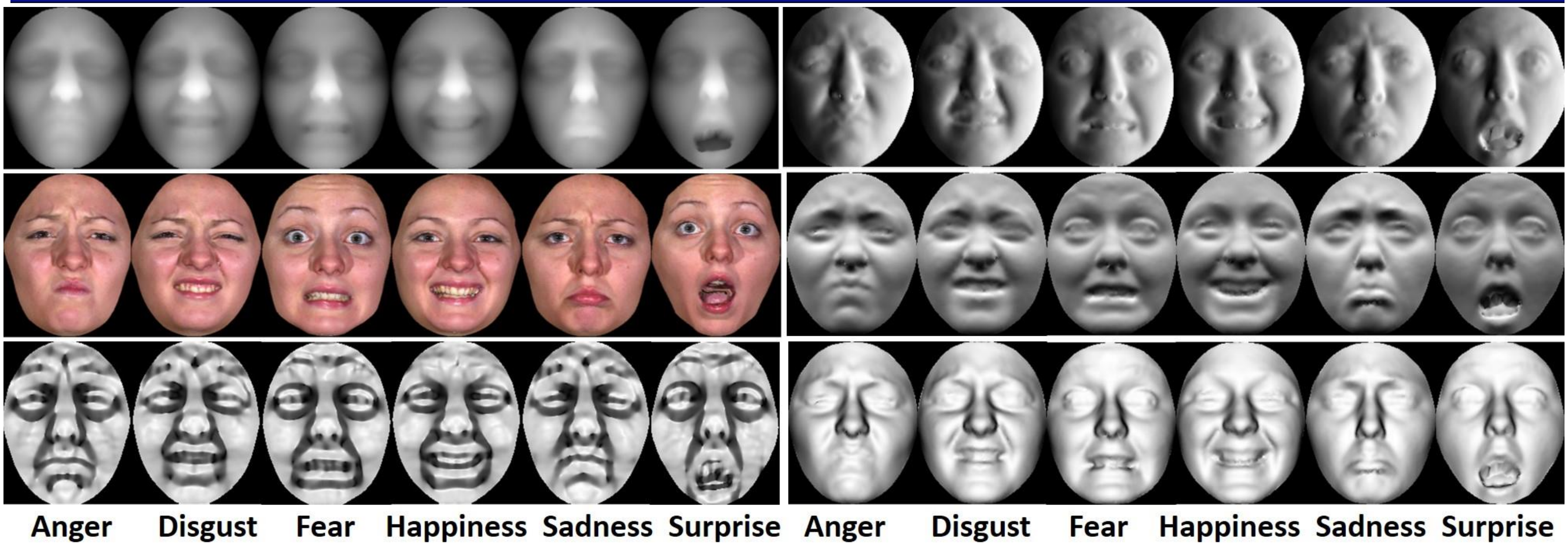
Solution: Deep fusion CNN (DF-CNN)

- **DF-CNN is an end-to-end training framework for both feature learning and fusion learning.**

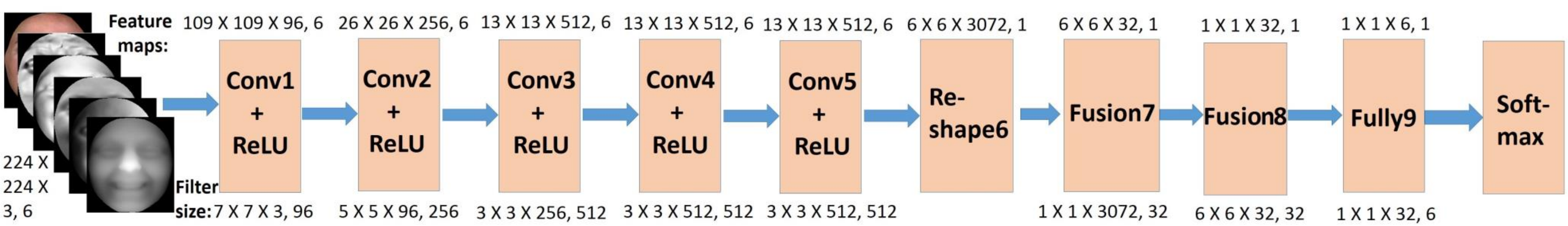


Approach overview: DF-CNN

Facial attribute maps: depth, texture, curvature and normal

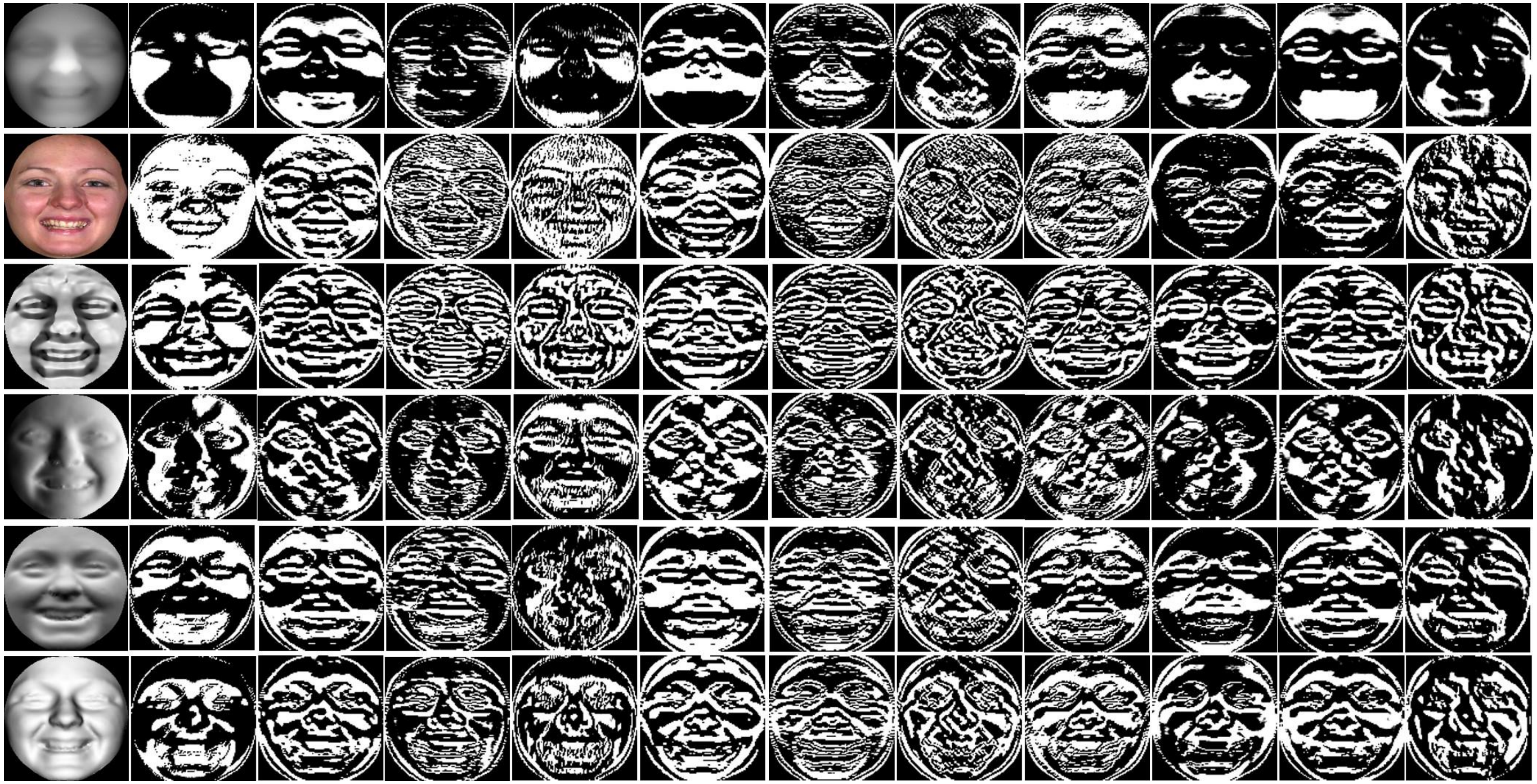


Architecture of DF-CNN:

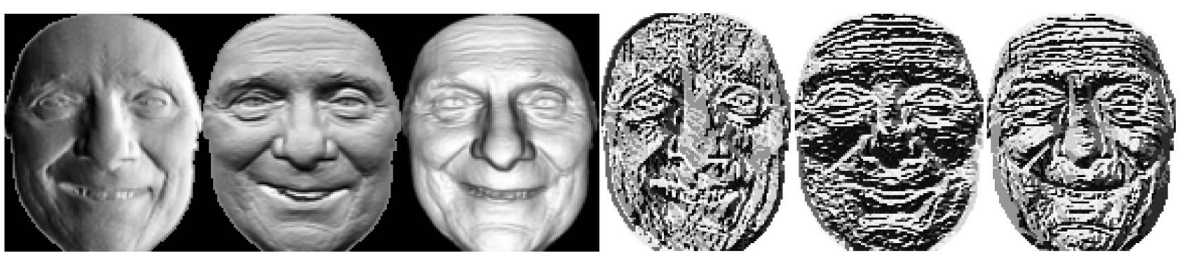


- **convolutional layers:** pre-trained deep model (e.g., vgg-m-net)
- **other layers:** randomly initialized

Visualization of feature maps: 1st conv. layer of DF-CNN

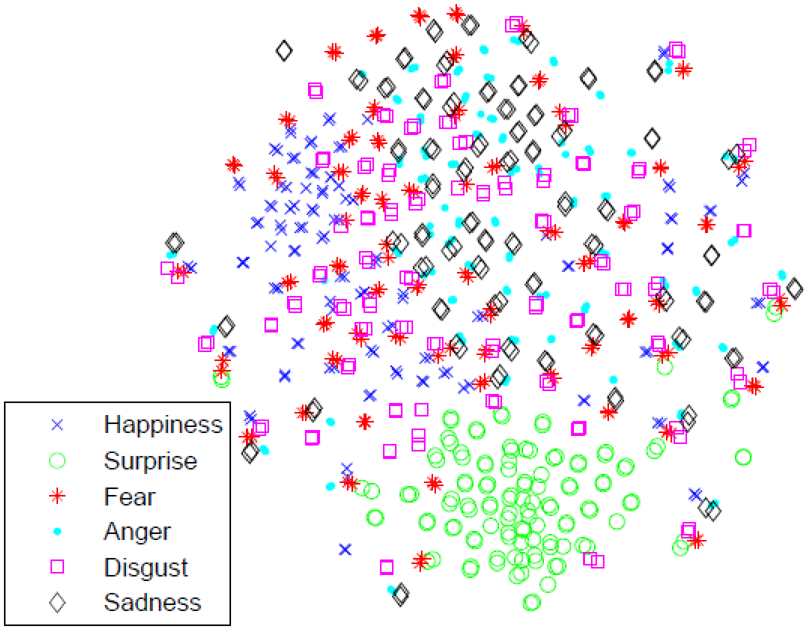


Similar to gradient-like facial maps: e.g., normal-LBP facial maps



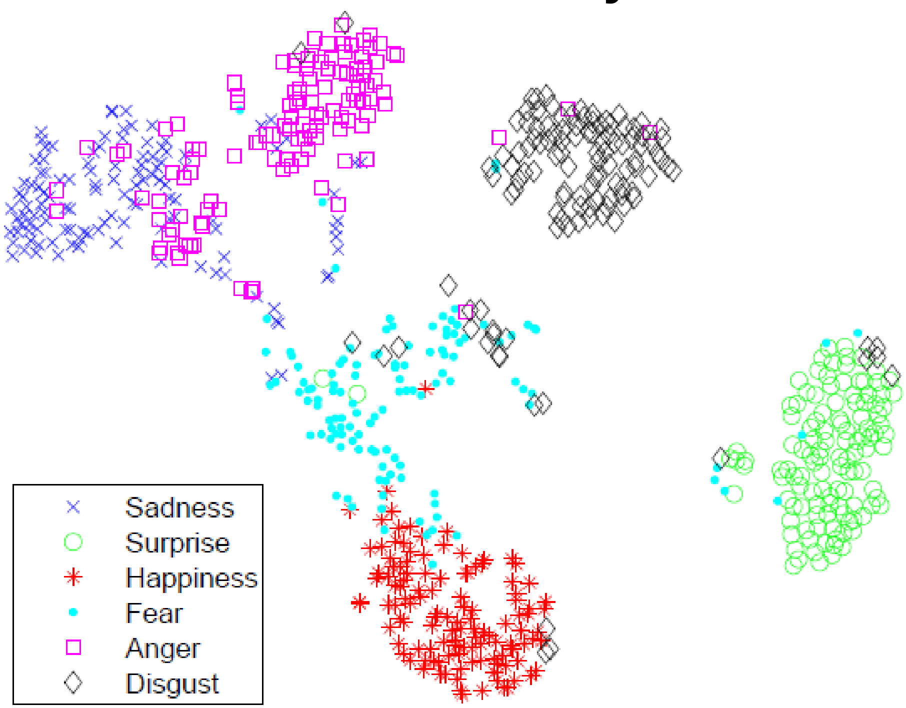
Visualization of handcrafted and learned features:

(t-SNE-based embedding)



Gabor features

learned features by DF-CNN



Visualization of facial expression saliency maps:

anger

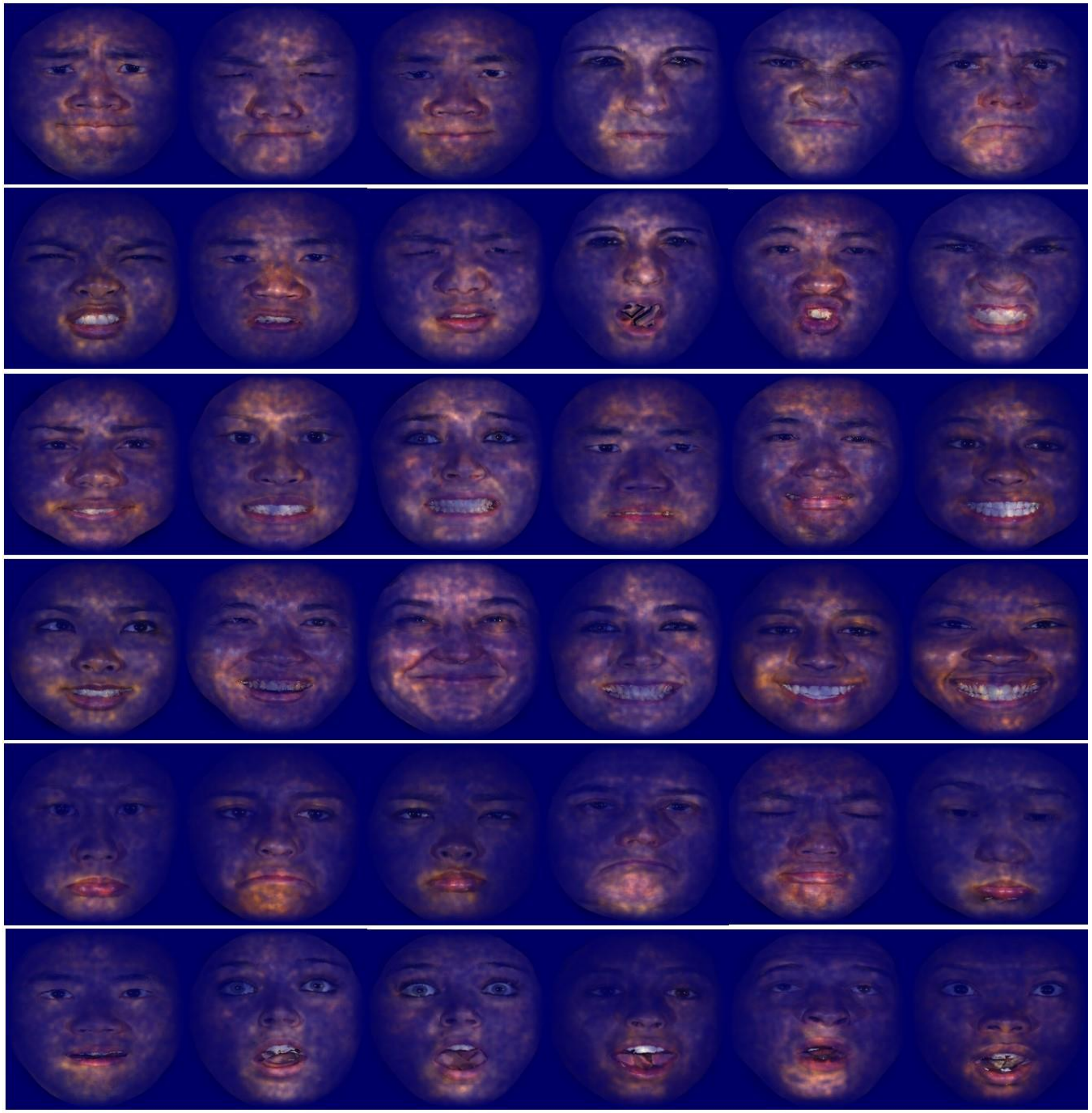
disgust

fear

happiness

sadness

surprise



The saliency maps indicates the pixel-level importance for FER, where blue color means less important pixels.

Different facial deformations correspond different patterns saliency maps

Datasets and Experimental Protocols

- **BU-3DFE database I: (standard settings)**
  60 subjects, 2 high levels of intensity, 6 expressions, 100 times 10-fold cross-validation, DF-CNN training: remaining 40 subjects
- **BU-3DFE database II: 2400 samples, 10-fold cross-validation**
- **Bosphorus database: 60 subjects, 6 expressions, 10-fold cross-validation**

Experimental results: BU-3DFE database I

- **Comparisons with hand-crafted features**

Method	I_g	I_n^x	I_n^y	I_n^z	I_c	I_t	All
MS-LBP	76.47	76.77	77.87	76.41	77.70	71.65	81.74
dense-SIFT	80.29	79.97	**82.35**	80.95	80.28	75.56	83.16
HOG	**81.89**	**82.09**	80.58	**81.81**	77.95	78.11	83.74
Gabor	77.95	78.80	81.97	81.10	**81.65**	**80.36**	**84.72**
DF-CNN$_{svm}$	-	-	-	-	-	-	**86.86**
DF-CNN$_{softmax}$	-	-	-	-	-	-	86.20

- **Comparisons with pre-trained deep features**

Method	I_g	I_n^x	I_n^y	I_n^z	I_c	I_t	All
caffe-alex-conv5	77.53	78.87	81.50	78.71	**80.83**	81.40	83.74
vgg-net-m-conv5	80.38	**80.37**	81.68	81.23	79.23	**82.14**	**84.22**
vgg-net-16-conv5-3	**81.72**	78.55	**83.06**	**81.25**	76.95	78.46	83.78
caffe-alex-full7	68.64	73.43	76.64	75.72	**74.52**	**74.45**	**82.56**
vgg-net-m-full7	73.34	**74.99**	**77.51**	**76.77**	68.81	70.93	81.56
vgg-net-16-full7	**76.71**	72.22	73.87	74.61	64.35	67.03	82.45
DF-CNN$_{svm}$	-	-	-	-	-	-	**86.86**
DF-CNN$_{softmax}$	-	-	-	-	-	-	86.20

Experimental results: BU-3DFE database I

Comparisons with fine-tuned deep features

Method	I_g	I_n^x	I_n^y	I_n^z	I_c	I_t	All
caffe-alex-ft-full7$_{svm}$	79.44	79.84	80.51	79.50	79.46	80.83	84.05
vgg-net-m-ft-full7$_{svm}$	79.68	**82.85**	**82.15**	80.30	**82.01**	81.62	84.85
vgg-net-16-ft-full7$_{svm}$	**80.21**	82.30	82.04	**80.43**	80.87	**84.10**	**86.01**
caffe-alex-ft$_{softmax}$	78.19	80.96	81.94	78.75	78.89	80.83	83.61
vgg-net-m-ft$_{softmax}$	**78.33**	**83.06**	**82.78**	**81.11**	**81.11**	80.42	85.00
vgg-net-16-ft$_{softmax}$	78.33	82.08	80.69	79.19	79.31	**84.17**	**85.14**
DF-CNN$_{svm}$	-	-	-	-	-	-	**86.86**
DF-CNN$_{softmax}$	-	-	-	-	-	-	86.20

- Approach based on fine-tuned deep features of a pre-trained deep model:
  (1) Separately fine-tuning the pre-trained deep model by the training data of different types of facial attribute maps;
  (2) Separately extracting deep features from fine-tuned deep models;
  (3) Linear SVM and score-level fusion.

Experimental results: BU-3DFE database I

- **Comparisons with state-of-the-art methods**

Methods	Data	Feature	Classifier	Accuracy
Wang et al. [52]	3D	curvatures/hist.	LDA	61.79
Soyel et al. [43]	3D	points/distance	NN	67.52
Soyel et al. [44]	3D	points/distance	NN	-
Tang et al. [45]	3D	points/distance	LDA	74.51
Tang et al. [46]	3D	slopes, distance	SVM	-
Mpiperis et al. [34]	3D	deformable model	ML	-
Gong et al. [15]	3D	depth/PAC	SVM	76.22
Berretti et al. [1]	3D	depth/SIFT	SVM	77.54
Maalej et al. [32]	3D	facial curves	SVM	-
Li et al. [26]	3D	normals, curv./hist.	SVM	82.01
Li et al. [23]	3D	normals/LBP	MKL	80.14
Lemaire et al. [22]	3D	curvature/HOG	SVM	76.61
Ocegueda et al. [35]	3D	coordinates, normals curvatures/DWT	Logistic Reg.	-
Zeng et al. [62]	3D	curvatures/LBP	SRC	70.93
Zhen et al. [66]	3D	coordinates, normals, shape index	SVM	84.50
Yang et al. [56]	3D	depth, normals, curv./scattering	SVM	84.80
Zhao et al. [65]	2D+3D	intensity,coordinates, shape index/LBP	BBN	-
Li et al. [24]	2D+3D	meshHOG/SIFT meshHOS/HSOG	SVM	86.32
DF-CNN$_{svm}$	2D+3D	32-D deep feature	SVM	**86.86**
DF-CNN$_{softmax}$	2D+3D	6-D deep feature	Softmax	86.20

Comparison results:

- data modality:
  2D+3D multimodality
- expression features:
  hand-crafted v.s. learned;
  high-dimension v.s. low-dim;
- classifiers:
  non-linear SVM, MKL… v.s. linear SVM, net
- Accuracy: best one

Experimental results: other databases

Method	I_g	I_n^x	I_n^y	I_n^z	I_c	I_t	All
MS-LBP	73.50	74.58	73.54	73.21	73.37	66.08	77.75
dense-SIFT	**76.25**	75.79	77.42	76.58	75.88	71.79	79.42
HOG	76.25	**76.88**	76.29	**77.75**	76.29	72.04	79.71
Gabor	73.04	75.00	**78.29**	76.42	**76.33**	**75.86**	**80.00**
vgg-net-m-conv5	**76.17**	75.04	76.92	**76.54**	75.54	76.42	**79.75**
vgg-net-m-full7	70.21	69.71	72.67	70.67	67.00	66.83	77.38
vgg-net-m-ft-full7$_{svm}$	75.17	**76.62**	**77.08**	75.83	**78.12**	**78.67**	**81.08**
vgg-net-m-ft$_{softmax}$	74.62	75.33	76.96	75.79	77.88	78.54	80.71
DF-CNN$_{svm}$	-	-	-	-	-	-	81.04
DF-CNN$_{softmax}$	-	-	-	-	-	-	**81.33**

BU-3DFE database II

Bosphorus database

Method	I_g	I_n^x	I_n^y	I_n^z	I_c	I_t	All
MS-LBP	71.11	69.44	70.56	66.67	62.78	62.50	73.33
dense-SIFT	70.28	73.89	72.78	73.89	72.50	65.56	76.39
HOG	**72.50**	**74.22**	73.89	**74.72**	71.94	**71.94**	77.22
Gabor	67.78	73.61	**75.83**	71.61	**75.56**	70.56	**77.50**
vgg-net-m-conv5	**71.94**	72.50	73.61	71.67	72.78	73.06	**79.72**
vgg-net-m-full7	61.11	63.33	63.89	65.83	60.56	61.94	75.56
vgg-net-m-ft-full7$_{svm}$	71.67	**72.78**	74.72	**76.11**	71.94	**73.61**	79.17
vgg-net-m-ft$_{softmax}$	71.39	72.78	**75.28**	75.00	**73.33**	73.61	**79.72**
DF-CNN$_{svm}$	-	-	-	-	-	-	**80.28**
DF-CNN$_{softmax}$	-	-	-	-	-	-	80.00

Experimental results: other databases

- **Comparisons with state-of-the-art methods**

Method	BU-3DFE Database II	Bosphorus database
Li et al. (2012) [23]	78.50	75.83
Li et al. (2015) [24]	80.42	79.72
Yang et al. (2015) [56]	80.46	77.50
DF-CNN$_{\text{svm}}$	**81.04**	**80.28**
DF-CNN$_{\text{softmax}}$	**81.33**	**80.00**

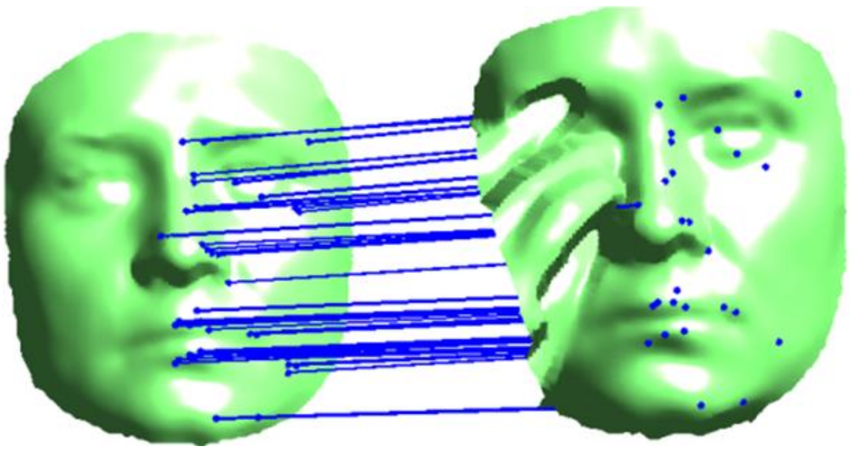
Future work & references

- **BP4D-Spontanous: 3D dynamic spontaneous facial expression database: 328 3D face videos, 368,036 frames of 41 subjects, with size about 2.6TB.**

1. Huibin Li, Jian Sun, Dong Wang, Zongben Xu, Liming Chen, Deep Representation of Facial Geometric and Photometric Attributes for Automatic 3D Facial Expression Recognition, http://arxiv.org/pdf/1511.03015.pdf, 2015.
2. Huibin LI, Jian Sun,, Zongben Xu, Liming Chen, Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network, IEEE Transactions on Multimedia, under review.

Conclusion

- A brief introduction of 3DFE and 3DFER
- An overview of SIFT-like matching framework
- Our work: registration-free approach for 3DFE
- Our work: DF-CNN for 3DFER
- Future directions: deep learning on surfaces



Thanks for your attention!