Qi-rong Mao, Xin-yu Pan, Yong-zhao Zhan, Xiang-jun Shen, 2015. Using Kinect for real-time emotion recognition via facial expressions. *Frontiers of Information Technology & Electronic Engineering*, **16**(4):272-282. [doi:10.1631/FITEE.1400209]

Using Kinect for real-time emotion recognition via facial expressions

Key words: Kinect, Emotion recognition, Facial expression, Real-time classification, Fusion algorithm, Support vector machine (SVM)

Contact: Qi-rong Mao
E-mail: mao_qr@ujs.edu.cn
ORCID: http://orcid.org/0000-0002-5021-9057

Introduction

- Recently, sensors used in most ERFE research are RGB cameras, which can capture only 2D images. Since human faces are 3D objects, the process representing 3D faces with 2D images is deficient in essential geometrical features. Furthermore, recognition approaches based on RGB cameras are in general computationally expensive.

- In this paper, we propose a real-time ERFE approach based on animation units (AUs) and feature point positions (FPPs) tracked by a Kinect sensor to recognize six basic emotions and neutral.

Architecture for our system

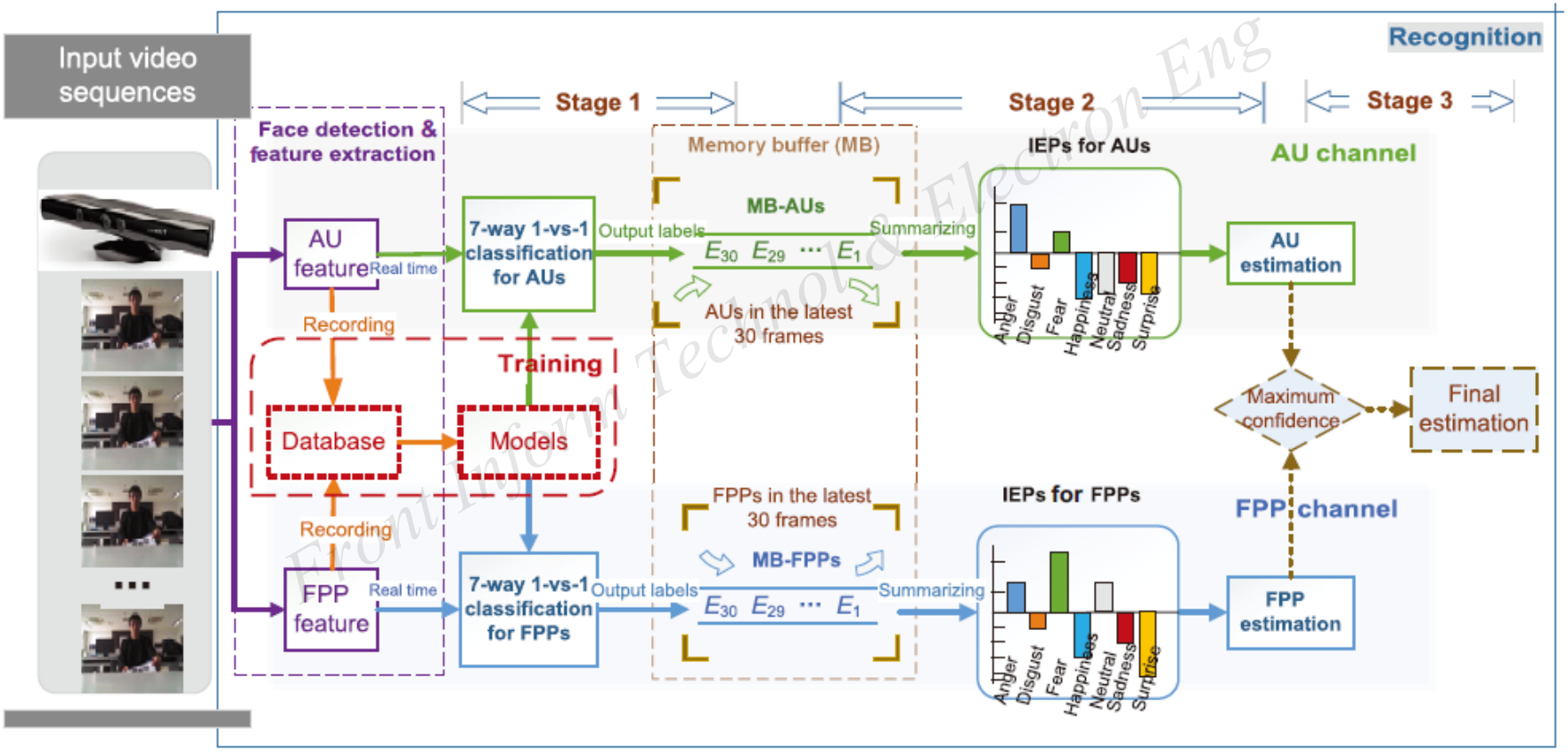

Fig. 1 Architecture for our system

Pre-recognition and fusion algorithm based on IEPs

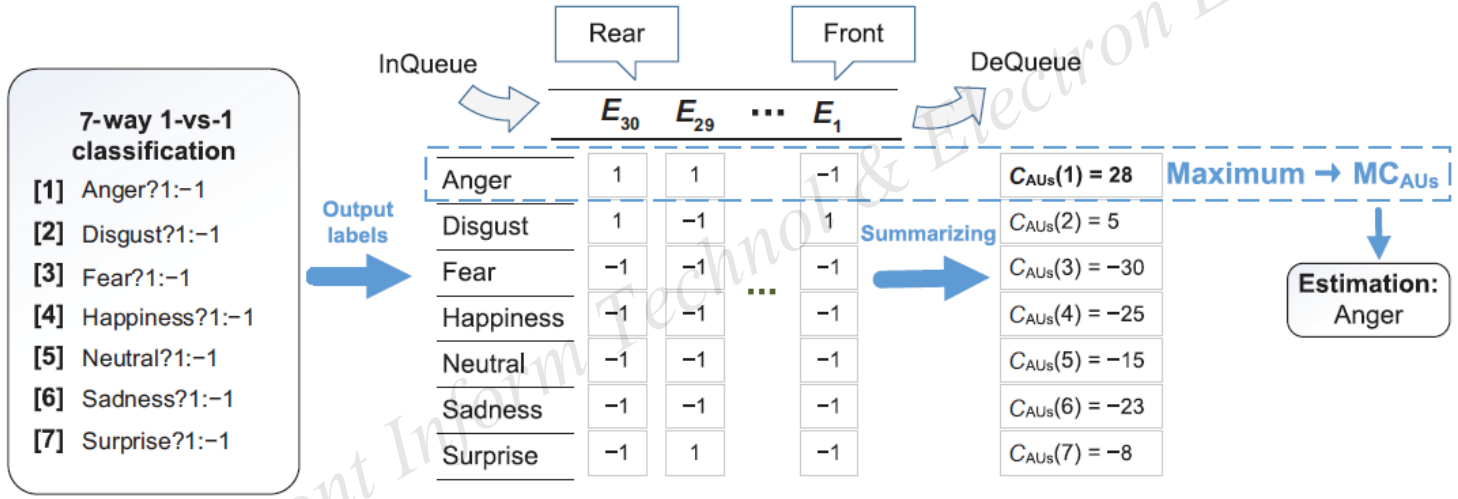

Fig. 2 Pre-recognition and fusion algorithm based on IEPs (AU channel for example). The 7-way 1-vs-1 classification in each channel contains seven sub-classifiers with the same name as corresponding emotions. Each sub-classifier has the ability of estimating whether or not input features belong to the corresponding emotion

Visualization of AUs and Kinect coordinate system

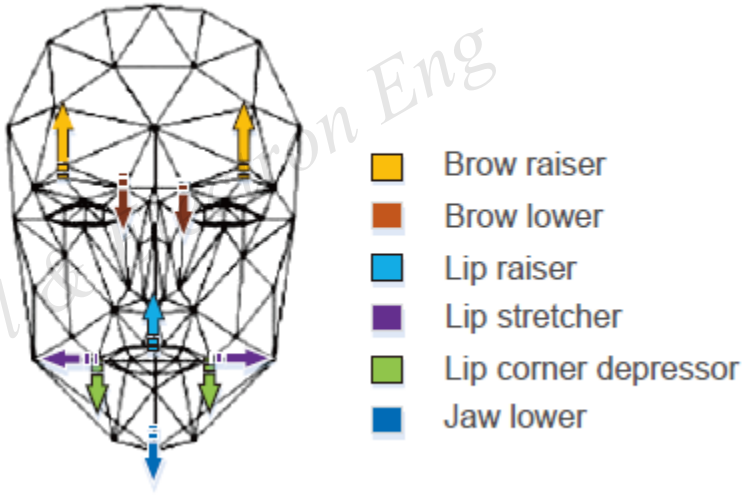

Fig. 3 Visualization of AUs

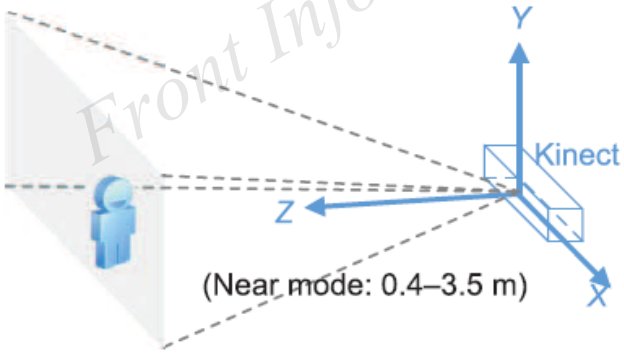

Fig. 4 Kinect coordinate system. The origin is located at the optical center of Kinect, Z-axis pointing towards a user, and Y-axis pointing up

Classifier selection

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	**79.27**	5.80	4.48	2.65	4.98	1.33	1.49
Disgust	5.19	**79.54**	1.68	3.97	4.73	4.58	0.31
Fear	2.73	3.45	**80.00**	2.91	9.64	1.10	0.00
Happiness	4.62	7.95	4.77	**75.58**	4.77	2.31	0.00
Neutral	2.41	2.56	5.57	4.37	**79.52**	5.42	0.00
Sadness	3.60	3.96	7.55	1.80	8.99	**73.74**	0.36
Surprise	1.41	2.03	0.00	0.16	0.00	0.00	**96.40**

Fig. 6 Recognition accuracies (%) of multi-class classification on each emotion

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	**85.14**	–	–	–	–	–	–
Disgust	–	**83.28**	–	–	–	–	–
Fear	–	–	**81.86**	–	–	–	–
Happiness	–	–	–	**80.67**	–	–	–
Neutral	–	–	–	–	**87.65**	–	–
Sadness	–	–	–	–	–	**80.48**	–
Surprise	–	–	–	–	–	–	**98.49**

Fig. 7 Recognition accuracies (%) of 1-vs-1 classification

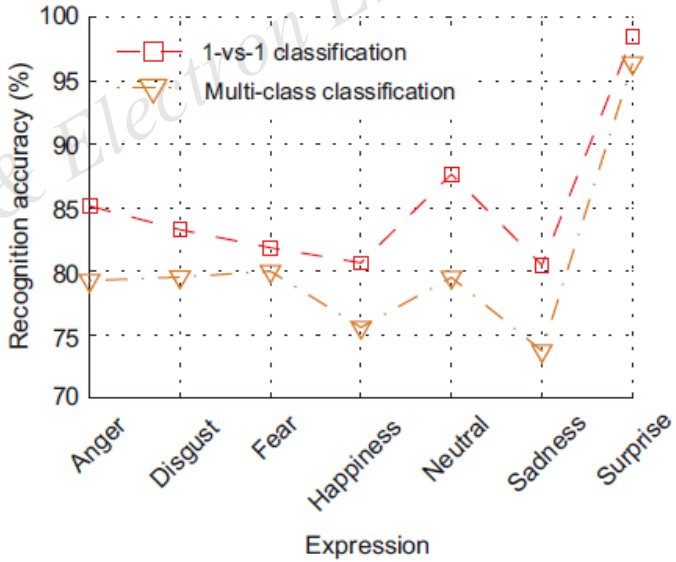

Fig. 8 Accuracy of recognition using 1-vs-1 classification and multi-class classification

Recognition accuracies using AUs and FPPs in different poses

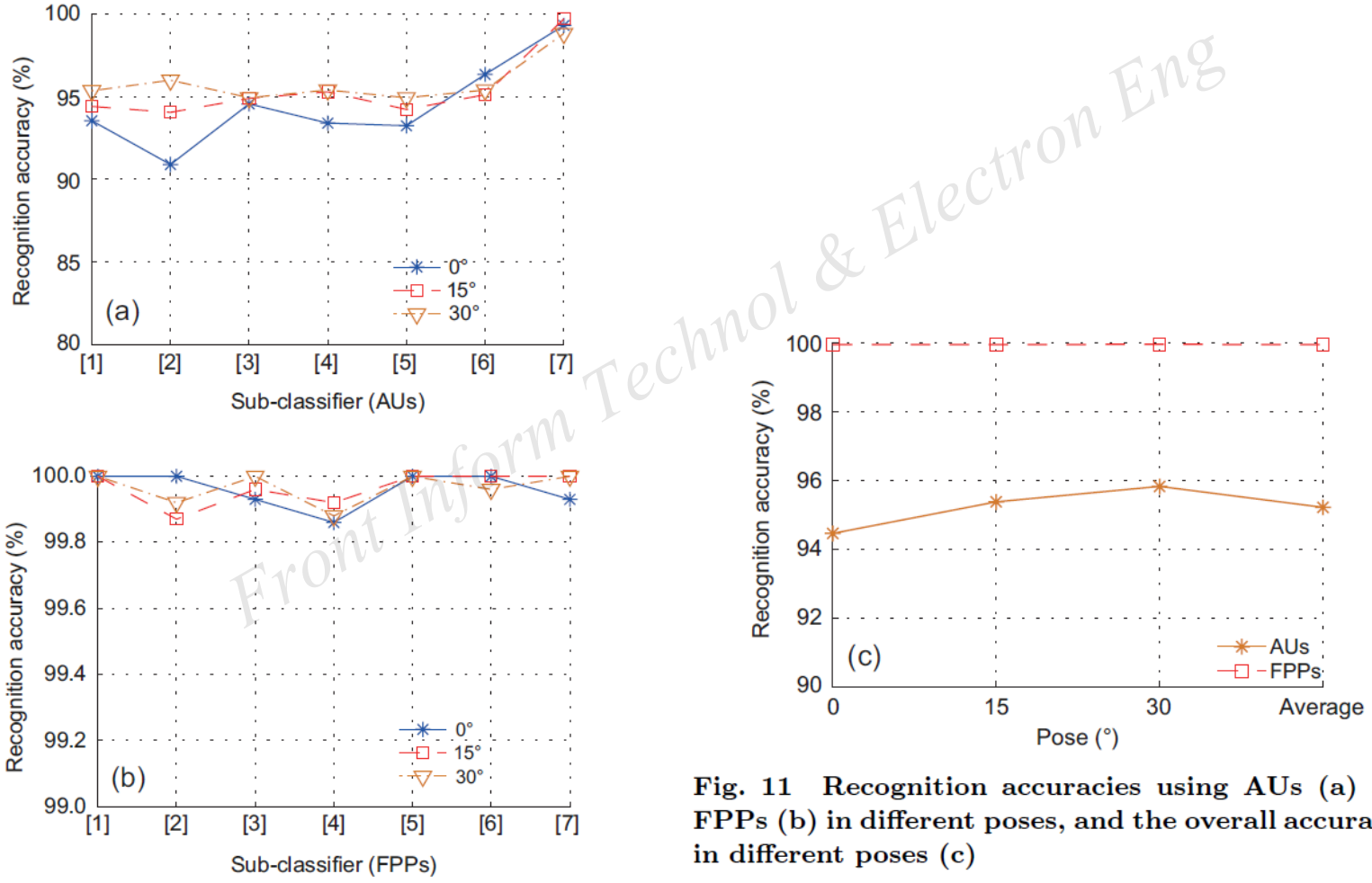

Fig. 11 Recognition accuracies using AUs (a) and FPPs (b) in different poses, and the overall accuracies in different poses (c)

Recognition accuracy of sub-classifiers using the FaceWarehouse database

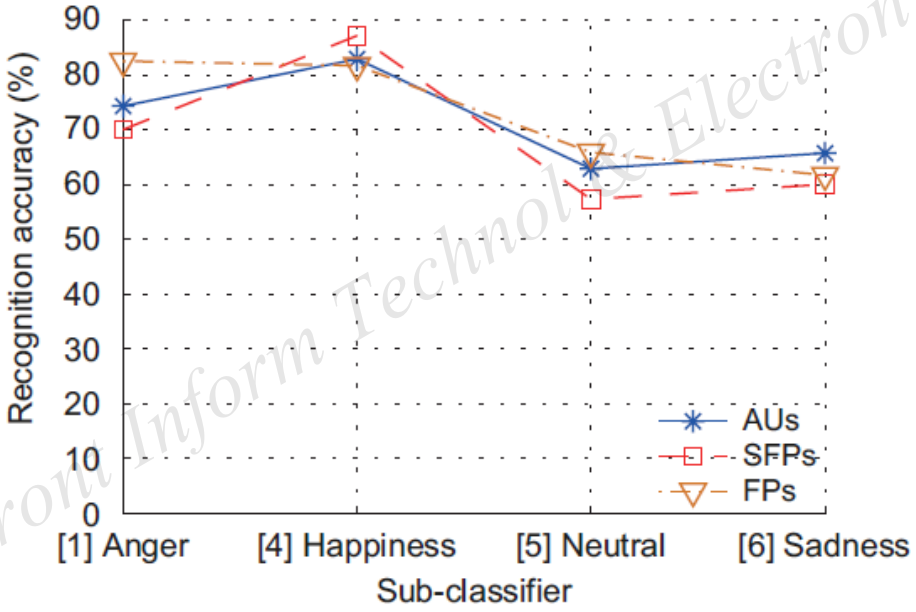

Fig. 13 Recognition accuracy of sub-classifiers using the FaceWarehouse database

Main application window

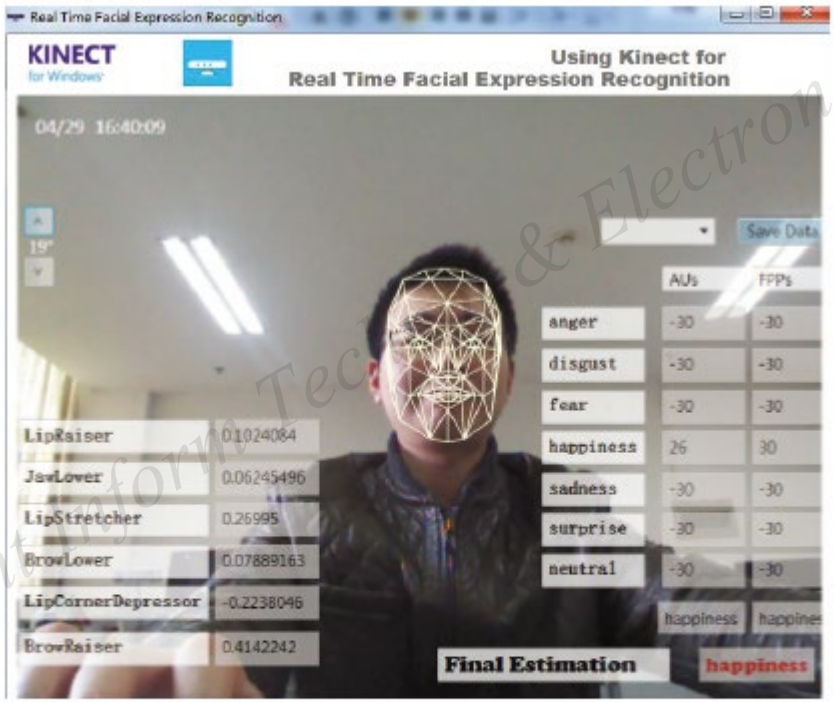

Fig. 14 Main application window. Real-time AU features are shown on the bottom left and a fusion algorithm based on IEPs and maximum confidence is displayed on the right

Conclusions

- In this paper, we proposed a real-time ERFE approach based on both 2D and 3D features captured by Kinect. A fusion algorithm based on IEPs and maximum confidence functions is the kernel of our approach. Real-time AU and FPP features, together with this fusion algorithm, enable us to recognize real-time emotions via facial expressions.