A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference

Shangfei Wang, *Member, IEEE*, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, and Xufa Wang

***Abstract*—To date, most facial expression analysis has been based on visible and posed expression databases. Visible images, however, are easily affected by illumination variations, while posed expressions differ in appearance and timing from natural ones. In this paper, we propose and establish a natural visible and infrared facial expression database, which contains both spontaneous and posed expressions of more than 100 subjects, recorded simultaneously by a visible and an infrared thermal camera, with illumination provided from three different directions. The posed database includes the apex expressional images with and without glasses. As an elementary assessment of the usability of our spontaneous database for expression recognition and emotion inference, we conduct visible facial expression recognition using four typical methods, including the eigenface approach [principle component analysis (PCA)], the fisherface approach [PCA + linear discriminant analysis (LDA)], the Active Appearance Model (AAM), and the AAM-based + LDA. We also use PCA and PCA+LDA to recognize expressions from infrared thermal images. In addition, we analyze the relationship between facial temperature and emotion through statistical analysis. Our database is available for research purposes.**

***Index Terms*—Emotion inference, expression recognition, facial expression, infrared image, spontaneous database, visible image.**

I. Introduction

FACIAL expression is a convenient way for humans to communicate emotion. As a result, research on expression recognition has become a key focus area of personalized human-computer interaction [1], [2]. Most current research focuses on visible images or videos and good performance has been achieved in this regard. Whereas varying light exposure can hinder visible expression recognition, infrared thermal images, recording the temperature distribution formed by face vein branches, are not sensitive to imaging conditions. Thus, thermal expression recognition is a useful and necessary complement to visible expression recognition [3]. Besides, a change in facial temperature is a clue that can prove helpful in emotion inference [4], [5]. Furthermore, most existing research has been based on posed expression databases, which are elicited by asking subjects to perform a series of emotional expressions in front of a camera. These artificial expressions are usually exaggerated. Spontaneous expressions, on the other hand, may be subtle and differ from posed ones both in appearance and timing. It is, therefore, most important to establish a natural database to allow research to move from artificial to natural expression recognition, ultimately leading to more practical applications thereof.

This paper proposes and establishes a natural visible and infrared facial expression database (NVIE) for expression recognition and emotion inference. First, we describe in detail the design, collection, and annotation of the NVIE database. In addition, we conduct facial expression analysis on spontaneous visible images with front lighting using several typical methods, including the eigenface approach [principle component analysis (PCA)], the fisherface approach [PCA + linear discriminant analysis (LDA)], the Active Appearance Model (AAM), and the combined AAM-based + LDA (referred to as AAM+LDA). Thereafter, we use PCA and PCA +LDA to recognize expressions from spontaneous infrared thermal images. In addition, we analyze the relationship between facial temperature and emotion through an analysis of variance (ANOVA). The evaluation results verify the effectiveness of our spontaneous database for expression recognition and emotion inference.

II. Brief Review of Existing Natural and Infrared Databases

There are many existing databases dealing with facial expressions, an exhaustive survey of which is given in [1] and [6]. Here, we only focus on natural and infrared facial expression databases. Due to the difficulty of eliciting affective displays and the time-consuming manual labeling of spontaneous expressions, only a few natural visible expression databases exist. These are listed in Table I, together with details of size, elicitation method, illumination, expression descriptions, and modality [visual (V) or audiovisual (AV)]. From Table I, we can see that researchers use one of three possible approaches to obtain spontaneous affective behavior, i.e., human-human conversation, human-computer interaction, or emotion-inducing videos [22]. Since this paper only focuses on facial expressions, and not speech or language, using emotion-inducing videos is

Manuscript received December 13, 2009; revised March 24, 2010 and June 23, 2010; accepted June 24, 2010. Date of publication July 26, 2010; date of current version October 15, 2010. This paper is supported in part by National 863 Program (2008AA01Z122), in part by Anhui Provincial Natural Science Foundation (No.070412056), and in part bySRF for ROCS, SEM. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Caifeng Shan.

The authors are with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: sfwang@ustc.edu.cn; leivo@mail.ustc.edu.cn; lsliang@mail.ustc.edu.cn; lvyp@mail.ustc.edu.cn; guobing@mail.ustc.edu.cn; dbpeng@mail.ustc.edu.cn; feichen@mail.ustc.edu.cn; xfwang@ustc.edu.cn).

Digital Object Identifier 10.1109/TMM.2010.2060716

TABLE I
DATABASES OF NATURAL FACIAL EXPRESSIONS

References	Size	Elicitation method	Lighting	Expression description	Modality
MMI[7]	29 subjects (18 adults and 11 children), 65 videos	Adults watch emotion-inducing videos. Children are told jokes or told to mimic laughing.	Variable	79 AUs and their combinations	V
UT-Dallas[8]	284 subjects, about 1540 5-second standardized clips	Subjects watch 10-minute emotion-inducing videos.	Indoor	Happiness, sadness, fear, anger, boredom, etc.	V
UA-UIUC[9]	28 subjects, one video clip for each subject	Subjects watch emotion-inducing videos.	Indoor	Neutral, joy, surprise and disgust	V
Belfast[10]	125 subjects, 298 audiovisual clips	Interactive chats	Indoor	Anger, fear, etc	AV
AAI [11]	60 subjects, one 30-60min audiovisual for each subject	Subjects were interviewed and asked to describe the childhood experience.	N/A	6 basic emotions, embarrassment, contempt, shame, general positive and negative	AV
RU-FACS [12]	100 subjects	Subjects tried to convince the interviewers they were telling the truth.	N/A	33 AU	AV
AvID [13]	15 subjects, approximately one-hour video for each subject	Subjects describe neutral photographs, play a game of Tetris, describe the game of Tetris and solve cognitive tasks.	Indoor	Neutral, relaxed, moderately aroused and highly aroused	AV
Geneva Airport Lost Luggage Study [14]	109 subjects	Unobtrusive videotaping of passengers at Geneva airport lost luggage counter followed up by interviews with passengers	outdoor	Anger, good humor, indifference, stress and sadness	AV
SALAS [16]	20 subjects	Subjects talk to artificial listener; emotional states are changed by interaction	N/A	Wide range of emotions/emotion related states but not very intense	AV
VAM [17]	104 subjects, 1421 segmented utterance videos and 1872 facial images	Television talk-show	N/A	Valence (negative vs. positive), activation (calm vs. excited) and dominance (weak vs. strong).	AV
HUMAINE [18]	48 clips, between 3secs and 2 minutes in length	Television recordings, etc.	Indoor / outdoor	A broad emotional space (positive and negative, active and passive) and all the major types of combination of emotion (consistent emotion, co-existent emotion, emotional transition over time)	AV

TABLE II
DATABASES OF INFRARED FACIAL EXPRESSIONS

Reference	Size	Wave band	Elicitation	Lighting	Thermal	Expression description
NIST Equinox[19]	More than 600 subjects, 1919 infrared images	LWIR 8-12 microns, MWIR 3-5 microns	Posed	Above, left and right	Yes	Smile, frowning and surprise
IRIS [20]	30 subjects, 4228 pairs of thermal and visible images	7-14 microns	Posed	Left, right, both lights, dark and off	Yes	Surprise, laughing and anger

a suitable approach, especially as the datasets will not include any facial changes caused by speech.

Because expression recognition in the thermal infrared domain has received relatively little attention compared with recognition in visible-light imagery, no thermal expression databases exist. There are, however, a few thermal face databases (listed in Table II) that include some posed thermal expression images. For each database, we provide information on the size, wave band, elicitation method, illumination, thermal information, and expression descriptions.

Based on Tables I and II, it is obvious that current natural expression databases focus only on visible spectrum imagery, while the existing infrared databases consist only of posed images. Therefore, we propose and establish a natural visible and infrared facial expression database, enriching the existing databases for expression recognition and emotion inference.

III. IMAGE ACQUISITION SETUP

To set up the database, we first built a photographic room. Since videos are used to induce the subjects' emotions, we chose a quiet room as the experiment environment to ensure that the effect of the screened videos was not compromised. The room was 9.6 m*4.0 m*3.0 m, with a single door and two windows. The facial expression recording system included a camera system, illumination system, glasses, and thermometer. The details of the system, depicted in Fig. 1, are described below.

A. Camera Setup

To record both visible and infrared videos, we used two cameras: a DZ-GX25M visible camera capturing 30 frames per second, with resolution 704*480, and a SAT-HY6850 infrared camera capturing 25 frames per second, with resolution 320*240 and wave band 8–14 μm. The lenses of the two cameras were placed 0.75 m in front of the subject, with a distance of 0.1 m between the two cameras. The cameras were placed at a height of 1.2 m above the ground. To begin the process, both the video and the two cameras were turned on at the same time. Then, a sun visor was placed in front of both cameras,

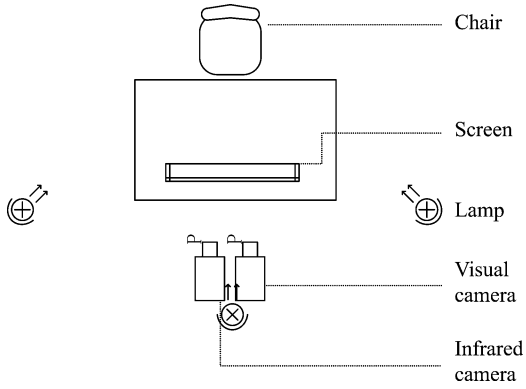


Fig. 1. Design of the photographic room.

and removed very quickly from above. Thus, the removal of the sun visor was recorded in both the visible and thermal videos. Before segmenting the emotional frames, frames containing evidence of the sun visor were carefully compared to find two frames in which the position of the sun visor was almost the same. Appropriate time stamps were calculated for each, allowing the remaining parts of the visible and infrared videos to start simultaneously.

To prevent any discomfort and to guarantee that the emotion was spontaneous, we did not require subjects to keep their heads fixed in one position. Once the screening had started, the experimenter stayed in the room to supervise the whole process.

Temperature information recorded by the infrared camera gradually becomes inaccurate after a long period of recording due to the increased heat of the sensor. Thus, timely calibration is needed. To minimize any disruption to the subject and to prevent any loss of frames, the experimenters manually invoked the auto-calibration before each experiment and while neutral videos (depicted in Fig. 2) were being screened.

B. Illumination Setup

This system controls the indoor light and the different directions of the light source. The indoor lighting was controlled by daylight lamps on the ceiling. Both the door and the curtains

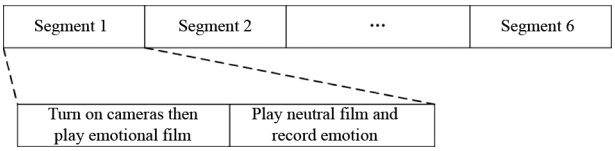


Fig. 2. Video clips.

were kept closed during the screening of the videos, keeping the lighting difference between day and night as small as possible. Another three floor lamps were used to create the different illumination conditions. The distance between the lamps and the subject was one meter. As shown in Fig. 1, the front lamp was placed in front of the subject, while the left and right lamps were at 45 degree angles to the front lamp. Bearing in mind that bright lights may cause the subjects some discomfort, 36-W lamp bulbs were used. To create front illumination, all three lamps were turned on, whereas for left or right illumination, the front lamp, together with either the left or right lamp, respectively, was turned on.

C. Glasses

When performing facial expression recognition, the wearing of glasses can disguise important information. To make the NVIE database more practical, we added facial expressions with glasses. In the spontaneous experiment, subjects were not specifically requested to wear glasses, but could do so if they chose to. In the posed experiment, all subjects were required to pose for two complete sets of expressions under each lighting condition. In the first set, subjects wore glasses, whereas in the other, they posed without glasses. Subjects could choose to wear their own glasses or the pair we provided.

D. Environmental Temperature

Although thermal emissivity from the facial surface is relatively stable under illumination variations, it is sensitive to the temperature of the environment. We, therefore, recorded the temperature of the room during the experiments. Room temperature ranged between 18 and $32^{\circ}\mathrm{C}$, with a mean of $23.29^{\circ}\mathrm{C}$ and a standard deviation of $3.39^{\circ}\mathrm{C}$.

IV. Data Acquisition

A. Subjects

A total of 215 healthy students (157 males and 58 females), ranging in age from 17 to 31, participated in our experiments. All of the subjects had normal auditory acuity and were mentally stable. Each signed an informed consent before the experiment and received compensation for participating after completing the experiment. Each subject participated in three spontaneous experiments, for each of the three illumination conditions, and one posed experiment. However, some of the subjects could only express two or fewer expressions and some thermal and visible videos were lost. Ultimately, for the spontaneous database, we obtained images of 105 subjects under front illumination, 111 subjects under left illumination, and 112 subjects under right illumination, while 108 subjects contributed to the posed database.

B. Stimuli

In our experiments, we induced the subjects' emotions by screening deliberately selected emotional videos. All the videos were obtained from the Internet, including 13 happy, 8 angry, 45 disgusted, 6 fearful, 7 sad, 7 surprised, and 32 emotionally neutral videos, as judged by the authors. Each emotional video was about 3–4 min long, while neutral videos were about 1–2 min long.

C. Experiment Procedure

Two kinds of facial expressions were recorded during our experiments: spontaneous expressions induced by the film clips and posed ones obtained by asking the subjects to perform a series of expressions in front of the cameras. Each kind of facial expression was recorded under three different illumination conditions, namely left, front, and right illumination.

Details of this procedure are given below.

First, the subjects were given an introduction to the experimental procedure, the meaning of arousal and valence, and how to self-assess their emotions.

Second, the subjects seated themselves comfortably, and then we moved their chairs forward or backwards to ensure a distance of 0.5 m between the chair and the screen.

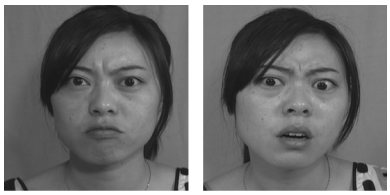
Third, a 19-inch LCD screen was used to display the video clips, each of which contained six segments (as depicted in Fig. 2), corresponding to the six types of emotional videos. In each segment, a particular type of emotional video was played and the subject's expressions recorded as synchronous videos. To reduce the interaction of different emotions induced by the different emotional video clips, neutral clips were shown between segments. Meanwhile, subjects were also asked to report the real emotion experienced by considering emotional valence, arousal, the basic emotion category, and its intensity on a five-point scale. These self-reported data were used as the emotion label for the recorded videos.

Fourth, once the video had terminated, each subject was asked to display six expressions, namely happiness, sadness, surprise, fear, anger, and disgust, both with and without glasses.

V. Design of the NVIE Database

First, we manually find the onset and apex of an expression in the visible facial videos. Then both visible and thermal videos during these periods are segmented into frames. Thus, our NVIE database includes two sub-databases: a spontaneous database consisting of image sequences from onset to apex and a posed database consisting of apex images, with each containing both the visible and infrared images that were recorded simultaneously and under three different illumination conditions, namely, illumination from the left, front, and right. The posed database also includes expression images with and without glasses.

Because it is difficult to determine automatically what kind of facial expression will be induced by a certain emotional film, five students in our lab manually labeled all the visible apex facial images in the spontaneous database according to the intensity of the six categories (happiness, sadness, surprise, fear, anger, and disgust), arousal, and valence on a three-point scale. The category with the highest average intensity was used as the



(a) (b)

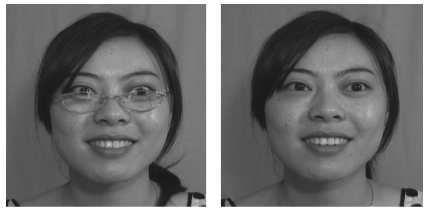
Fig. 3. Example images of a subject showing posed and spontaneous anger.

(a) (b)

Fig. 4. Visual and infrared images of a subject expressing anger.

(a) (b) (c)

Fig. 5. Example images of a subject expressing happiness with illumination from the left, front, and right.



(a) (b)

Fig. 6. Example images of a subject with and without glasses, expressing happiness.

(a) (b) (c)

(d) (e) (f)

Fig. 7. Images of a subject showing the six expressions (happiness, disgust, fear, surprise, sadness, and anger).

label for the visible and thermal apex facial images. The average arousal and valence were also adopted as labels. The kappa coefficient of the labeling was 0.65, indicating a good consistency.

As we know, expression is not emotion. For example, we may feel sad, but our expression can be neutral. The data collected from the self assessment reports, described in Section IV-C, were used to provide an emotional annotation of the corresponding image sequences.

The following sections describe the variations in our database and present some example images.

A. Posed versus Spontaneous

Both posed and spontaneous expressions were recorded in our database, allowing researchers to investigate differences further. Fig. 3 shows an example of a posed and spontaneous facial expression in the database.

B. Infrared versus Visible

Both visible and infrared images were collected. Fig. 4 shows an expression of anger in both a visible and infrared image.

C. Illumination Direction

Using the light system described in Section III-B, we captured the facial expressions with illumination from three different directions. Fig. 5 shows the images captured under the different illumination conditions.

D. Glasses

Images of a subject with and without glasses are shown in Fig. 6.

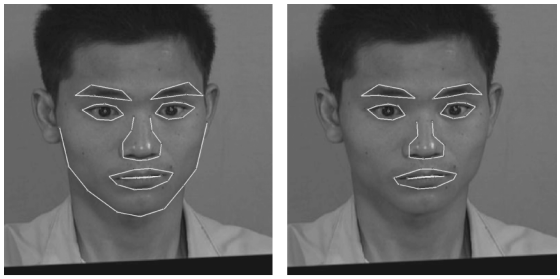
E. Six Facial Expressions

In the posed database, each subject displayed six facial expressions. In the spontaneous emotion database, the expressions were evaluated by five experimenters. Fig. 7 shows example images of the six spontaneous facial expressions.

VI. Validation and Evaluation of the Database

The main aim of this section is to present an elementary assessment of the usability of the spontaneous sub-database with front lighting for expression recognition and emotion inference, and to provide reference evaluation results for researchers using the database. We conduct visible facial expression recognition using four typical methods, namely the PCA, PCA+LDA, AAM, and AAM+LDA. We also use PCA and PCA +LDA to recognize expressions from infrared thermal images. In addition, we analyze the relationship between facial temperature and emotion through statistical analysis.

Not all subjects displayed all six types of emotion we aimed to elicit. Thus, only three expressions (i.e., disgust, fear, and happiness), which were induced successfully in most cases, are used for expression recognition. Furthermore, since we did not restrict head movement in our data acquisition experiment,



(a) (b)

Fig. 8. Face labeling (a) with and (b) without face configuration.

thereby ensuring more natural expressions, some sequences with non-frontal facial images were discarded. Ultimately, 236 apex images of 84 subjects were selected for visible and thermal expression recognition, including 83 images depicting disgust, 62 images depicting fear, and 91 images depicting happiness. In addition, the expressions were divided into four classes in the arousal-valence space: arousal positive/valence positive$(+,+)$, arousal positive/valence negative$(+,-)$, arousal negative/valence positive$(-,+)$, and arousal negative/valence negative$(-,-)$. Of the 236 images, 100 images were classified as $(+,+)$, 132 images as $(+,-)$, 0 images as $(-,+)$, and 4 images as $(-,-)$.

Although some subjects did not display sad, angry, and surprised expressions, they really did feel these, which is why all six types of emotion for 20 participants (13 male and 7 female) from the NVIE spontaneous sub-database are used for emotion inference based on facial temperature as described below.

A. Expression Recognition From Visible Images

1) Analysis Methods: Before feature extraction, preprocessing work activities including manual eye location, angle rectification, and image zoom were performed to normalize the apex images to $W * H$ rectangles. Then, four baseline algorithms are used to extract image features, namely the PCA [23], PCA+LDA [24], AAM [26], and AAM+LDA [24], [26]. Details of these algorithms can be found in [23], [24], and [26], respectively. For the PCA and PCA+LDA methods, W and H are both set as 64 to reduce the computational complexity, yet retain sufficient information of the images. Furthermore, four methods are used to normalize the grey-level values of the images, that is, histogram equalization (HE), regional histogram equalization (RHE), gamma transformation (GT), and regional gamma transformation (RGT) [25]. Then, the non-facial region was removed using an elliptic mask. For each image of a specific person in the selected 236 images, we applied PCA with the training set containing all the remaining images, except the other images of different expressions for the same person. The grey-level values of all pixels of images are the feature sets used for PCA and LDA.

For the AAM and AAM+LDA methods, W and H are both set as 400 to allow the facial features to be located relatively precisely. We consider two labeling options: with face configuration points (WFC: 61 points in total) and without (NFC: 52 points in total), as shown in Fig. 8. From the 236 images, we selected 72 images, which included 24 images for each expression, to build the appearance model. We applied this model to the 236 images to obtain their appearance parameters.

TABLE III
CONFUSION MATRIX OF CATEGORICAL EXPRESSION RECOGNITION ON VISIBLE IMAGES(%)

		PCA+KNN			PCA+LDA+KNN		
		D	F	H	D	F	H
HE	D	50.60	31.33	18.07	57.83	30.12	12.05
	F	27.42	50.00	22.58	33.87	48.39	17.74
	H	14.29	14.28	71.43	9.89	6.59	83.52
	Av.	58.47			65.25		
RHE	D	43.37	37.35	19.28	59.04	27.71	13.25
	F	20.97	51.61	27.42	32.26	51.61	16.13
	H	10.99	20.88	68.13	12.09	9.89	78.02
	Av.	55.08			64.41		
GT	D	54.22	20.48	25.30	53.01	38.55	8.43
	F	35.48	41.94	22.58	32.26	54.84	12.90
	H	20.88	12.09	67.03	14.29	5.49	80.22
	Av.	55.93			63.98		
RGT	D	54.22	21.69	24.10	53.01	36.14	10.84
	F	37.10	43.55	19.35	32.26	54.84	12.90
	H	18.68	14.29	67.03	13.19	6.59	80.22
	Av.	56.36			63.98		
		AAM+KNN			AAM+LDA+KNN		
		D	F	H	D	F	H
WFC	D	49.40	34.94	15.66	59.04	30.12	10.84
	F	22.58	59.68	17.74	37.10	45.16	17.74
	H	13.19	7.69	79.12	12.09	12.09	75.82
	Av.	63.56			61.86		
NFC	D	65.06	27.71	7.23	60.24	26.51	13.25
	F	30.65	53.23	16.13	35.48	43.55	20.97
	H	13.19	6.59	80.22	13.19	12.09	74.73
	Av.	67.80			61.44		

After feature extraction, K-nearest neighbors was used as the classifier. Euclidean distance and leave-one-subject-out cross-validation was adopted, and K was set to 1 here.

2) Experimental Results and Analysis: Tables III and IV show the performance of these algorithms with respect to our database, including confusion matrices and average recognition rates.

The following observations are evident from Tables III and IV.

- With respect to categorical expressions, happiness has a higher recognition rate than disgust and fear. According to the confusion matrices, very few instances of happiness are incorrectly recognized as disgust or fear, and vice versa. On the other hand, instances of disgust and fear are more likely to be incorrectly identified as the other. Different expressions are displayed using different facial characteristics, especially in the mouth and eye regions. In our experiments, when most subjects smiled, they lifted the corners of their mouths and kept their eyes half-closed. However, when the subjects expressed disgust or fear, some closed their mouths, while others opened their mouths slightly or even widely. Moreover, some closed their eyes, while others kept their eyes wide open. In other words, facial movements of the subjects are similar in expressing happiness, but they differ from person to person when expressing disgust or fear. This may explain why happiness is more easily recognized.

TABLE IV
CONFUSION MATRIX OF DISCRETE DIMENSIONAL EXPRESSION RECOGNITION ON VISIBLE IMAGES(%)

		PCA+KNN			PCA+LDA+KNN		
		++	+−	−−	++	+−	−−
HE	++	67.00	33.00	0.00	79.00	21.00	0.00
	+−	23.48	75.00	1.52	19.70	80.30	0.00
	−−	25.00	75.00	0.00	25.00	75.00	0.00
	Av.	70.34			78.39		
RHE	++	64.00	36.00	0.00	71.00	29.00	0.00
	+−	25.76	72.73	1.52	21.21	78.79	0.00
	−−	50.00	50.00	0.00	25.00	75.00	0.00
	Av.	67.80			74.15		
GT	++	64.00	34.00	2.00	81.00	19.00	0.00
	+−	27.27	69.70	3.03	26.52	73.48	0.00
	−−	50.00	50.00	0.00	25.00	75.00	0.00
	Av.	66.1			75.42		
RGT	++	64.00	34.00	2.00	78.00	22.00	0.00
	+−	24.24	71.97	3.79	22.73	77.27	0.00
	−−	25.00	75.00	0.00	25.00	75.00	0.00
	Av.	67.37			76.27		
		AAM+KNN			AAM+LDA+KNN		
		++	+−	−−	++	+−	−−
WFC	++	76.00	23.00	1.00	74.00	24.00	2.00
	+−	19.70	78.03	2.27	15.15	83.33	1.52
	−−	25.00	75.00	0.00	50.00	50.00	0.00
	Av.	75.85			77.97		
NFC	++	76.00	22.00	2.00	74.00	26.00	0.00
	+−	15.91	83.33	0.76	22.73	76.52	0.76
	−−	25.00	75.00	0.00	50.00	50.00	0.00
	Av.	78.81			74.15		

In the Tables III and IV, each row of the confusion matrix denotes an actual expression, while each column gives the predicted expression.

- With respect to discrete dimensional expressions, it is hard to say whether $(+,-)$ or $(+,+)$ is more recognizable. Among the incorrect cases, $(+,+)$ and $(+,-)$ are mostly recognized as one another. However, all $(-,-)$ instances are incorrectly recognized as $(+,+)$ or $(+,-)$. The main reason for this is that most instances belong to $(+,+)$ or $(+,-)$, and only four instances belong to $(-,-)$.
- With respect to the processing algorithms without LDA, AAM performs better than PCA. The reason may be that PCA uses grey-level values of each image as features, while AAM uses geometric characteristics of each image, that is, shape and texture information, which better represents facial movements. LDA improves recognition rates in all PCA cases, although it has the opposite effect in most AAM cases. This shows that the effect of LDA depends on the property of the data processed [15]. The difference in recognition rates for different feature extraction algorithms is larger than that for different preprocessing algorithms within one processing algorithm, which means that feature extraction algorithms have a greater influence on the results than preprocessing algorithms. In general, NFC+AAM+KNN is the best combination of preprocessing and processing algorithms.

B. Expression Recognition From Infrared Thermal Images

1) Analysis Methods: The preprocessing procedure for IR images is similar to that for visible images, except that the manual eye location is less accurate than in visible images, because the eyes in IR images are not as clear as those in visible images. Because some subjects wore glasses, we removed the eye region using a 64x15 rectangular mask for the infrared facial images. After preprocessing, the PCA and PCA+LDA methods are used to extract features of the images. Then, K-nearest neighbors is used as the classifier. Euclidean distance and leave-one-subject-out cross validation is adopted, and K is set to 1 here.

TABLE V
CONFUSION MATRIX OF CATEGORICAL EXPRESSION RECOGNITION ON INFRARED IMAGES (%)

		PCA+KNN			PCA+LDA+KNN		
		D	F	H	D	F	H
Apex	D	37.75	20.48	42.17	40.96	28.92	30.12
	F	19.35	43.55	37.10	35.48	30.65	33.87
	H	32.97	25.27	41.76	30.77	32.97	36.26
	Av.	40.68			36.44		
Dif.	D	39.76	22.89	37.35	42.17	24.10	33.73
	F	24.20	45.16	30.64	35.48	30.65	33.87
	H	32.97	27.47	39.56	30.77	32.97	36.26
	Av.	41.49			36.36		

TABLE VI
CONFUSION MATRIX OF DISCRETE DIMENSIONAL EXPRESSION RECOGNITION ON INFRARED IMAGES (%)

		PCA+KNN			PCA+LDA+KNN		
		++	+−	−−	++	+−	−−
Apex	++	47.00	50.00	3.00	42.00	55.00	3.00
	+−	43.18	55.30	1.52	37.12	62.88	0.00
	−−	75.00	25.00	0.00	50.00	50.00	0.00
	Av.	50.85			52.97		
Dif.	++	46.00	51.00	3.00	45.00	50.00	5.00
	+−	44.70	53.78	1.52	37.12	62.88	0.00
	−−	50.00	50.00	0.00	25.00	75.00	0.00
	Av.	49.89			53.94		

2) Experimental Results and Analysis: Both apex images and difference images (Dif.), which are created from the difference between the grey-level of the corresponding pixels of apex and neutral images, are used in the experiments. Tables V and VI show the results.

- With respect to categorical expressions, it is hard to say which one is the best discriminator in the thermal infrared domain. However, with respect to discrete dimensional expressions, $(+,-)$ is more recognizable than $(+,+)$, and all $(-,-)$ instances are incorrectly recognized. The main reason for this is that most instances belong to $(+,+)$ or $(+,-)$, and only four instances belong to $(-,-)$.
- With respect to the processing algorithms, LDA improves recognition rates in discrete dimension expression recognition, but causes the performance to deteriorate in category recognition. The reason may be similar to that given above [15]. The effect of different source images is negligible.

However, since the overall recognition rates are not very high, new methods that are suitable for thermal infrared images need to be developed.

C. Emotion Inference From Thermal Images

1) Analysis Methods: Physiological changes due to autonomic nervous system activity can impact the temperature pat-

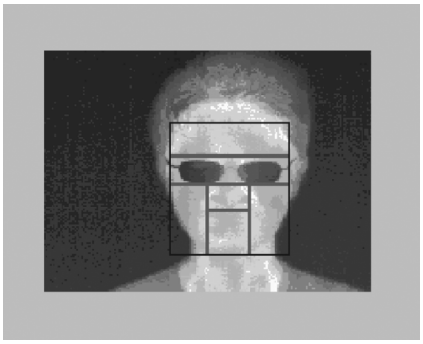


Fig. 9. Face segmentation.

terns of the face, which may be used to predict rated emotions [4], [5], [27]. In this section, we analyze the relationship between emotion and facial temperature using infrared thermal difference images. First, in order to retain the original temperature data for analysis, we manually segmented the difference infrared original images into five regions, namely forehead, eyes, nose, mouth, and cheeks, as shown in Fig. 9, and ensured that the facial segmentation size and ratio of each participant's neutral and emotional infrared images were consistent. Then, for each facial sub-region, we extracted the temperature data as the difference between the neutral and emotional infrared facial images, and calculated the five statistical parameters given below to reflect the temperature variance. As most of the participants wore glasses, thereby masking the thermal features of the eye region, the eye regions were not taken into account in this analysis.

- MEAN—the mean of each difference temperature matrix
- ABS—the mean of the absolute values of each difference temperature matrix
- ADDP—the mean of the positive values of each difference temperature matrix
- ADDN—the mean of the negative values of each difference temperature matrix
- VAR—the variance of each difference temperature matrix

Having obtained the five statistical parameters, three ANOVA analyses were conducted. First, an ANOVA was conducted to ascertain which statistical parameter is most useful for reflecting temperature change associated with changes in emotion. Second, an ANOVA was applied to the facial regions with different emotional states to ascertain in which facial regions the change in temperature due to the different emotions is the greatest. Third, an ANOVA analysis was applied to the emotional states in different facial regions to analyze which emotional state differs most in each facial sub-region.

2) Experimental Results and Analysis:

Experiment for Parameter Selection: The factors in this analysis are the six emotional states (i.e., sadness, anger, surprise, fear, happiness, and disgust) and the four facial sub-regions (forehead, nose, mouth, and cheeks). Using a multiple comparison of the post hoc test, the significant differences between the emotional states for the different statistical parameters are given in Table VII.

From Table VII, the following observations can be made.

TABLE VII
SIGNIFICANT DIFFERENCES BETWEEN DIFFERENT EMOTION STATES UNDER DIFFERENT STATIC PARAMETERS

Sig-Diff	MEAN	ABS	ADDP	ADDN	VAR
Sadness-Anger	0.92	0.55	0.50	0.29	0.85
Sadness-Surprise	0.96	0.06	0.02*	0.06	0.02*
Sadness-Fear	0.04*	0.00*	0.00*	0.07	0.00*
Sadness-Happiness	0.49	0.31	0.18	0.25	0.25
Sadness-Disgust	0.36	0.68	0.62	0.65	0.51
Anger-Surprise	0.53	0.20	0.08	0.39	0.04*
Anger-Fear	0.05	0.01*	0.00*	0.43	0.00*
Anger-Happiness	0.55	0.68	0.42	0.93	0.34
Anger-Disgust	0.41	0.32	0.86	0.13	0.64
Surprise-Fear	0.01*	0.23	0.12	0.94	0.32
Surprise-Happiness	0.22	0.39	0.27	0.44	0.27
Surprise-Disgust	0.15	0.02*	0.05	0.02*	0.11
Fear-Happiness	0.18	0.04*	0.01*	0.49	0.04*
Fear-Disgust	0.26	0.00*	0.00*	0.02*	0.01*
Happy-Disgust	0.83	0.16	0.39	0.11	0.63

* The mean difference is significant at the 0.05 level.

- When the statistical parameter VAR is used, the significance of the differences between the six emotional states is relatively higher than for all the other statistical parameters at the 0.05 level. Therefore, in the subsequent analyses, we only consider the VAR statistic.
- When VAR is used, the differential validities of the six emotional states are sorted as follows: Sadness-Fear, Anger-Fear, Fear-Disgust, Sadness-Surprise, Fear-Happiness, and Anger- Surprise. Changes in the thermal distribution on the face for different emotional states are caused by facial muscle movements, in addition to the transition of emotional states and/or other physiological changes [29]. However, for some emotional states, the thermal distribution on the face is not sufficiently unique to be distinguished from that of another state. This may be caused by ambiguity or overlap of the emotional states.

ANOVA Analysis of the Different Facial Regions With Different Emotion States Using VAR: The factors in this analysis are the facial sub-regions (forehead, nose, mouth, and cheeks), while the dependent variable is VAR. The impact is measured as the mean VAR values for different facial sub-regions, shown in Fig. 10 and Table VIII.

- From Fig. 10, we can conclude that, when using the VAR statistic, the degree of impact of the forehead and cheek regions is more significant than the other two regions for most emotional states, e.g., surprise, fear and happiness. Furthermore, the impact of the mouth region on all emotional states is the smallest compared with other regions. This means that when the emotion changes, the temperature variance in the forehead and cheek regions is much larger than that in the other regions, while the temperature variance in the mouth region is the smallest. The importance of the forehead region in the analysis of emotions using thermal images coincides with the research results of Sugimoto and Jenkins [30], [31]. The VAR values for the cheek region are relatively large for most emotional states, thereby reflecting the uniformity of this region's temperature change, while the results for the nose and mouth regions are opposite, thereby reflecting the non-uniformity

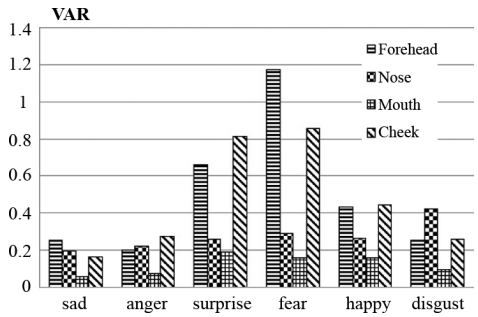


Fig. 10. Impact of different facial regions on different emotional states.

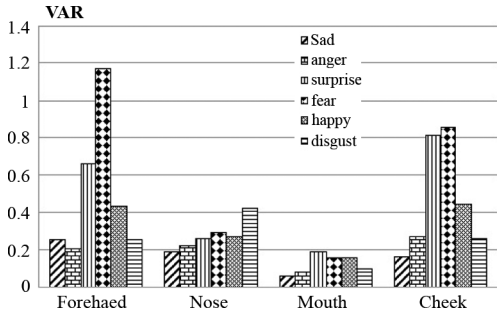


Fig. 11. Impact of the different emotional states in the different facial sub-regions.

TABLE VIII
MATRIX SHOWING SIGNIFICANT DIFFERENCES BETWEEN FACIAL REGIONS WITH DIFFERENT EMOTION STATES

Sig-Diff	Sad	Anger	Surp.	Fear	Happy	Dis.
Forehead-Nose	0.59	0.87	0.26	0.07	0.35	0.47
Forehead-Mouth	0.08	0.26	0.18	0.04*	0.13	0.52
Forehead-Cheek	0.42	0.53	0.67	0.51	0.96	0.98
Nose-Mouth	0.22	0.20	0.84	0.78	0.56	0.17
Nose-Cheek	0.79	0.64	0.12	0.24	0.33	0.49
Mouth-Cheek	0.34	0.08	0.08	0.15	0.12	0.50

* The mean difference is significant at the 0.05 level.

of the temperature change in these regions. This result is vastly different to that obtained when using visible facial expressions to reflect the various emotions, in that the change in forehead and cheek is not significant when the facial expression changes. This result is, therefore, meaningful in the research of emotion inference using changes in facial temperature.

- From Table VIII, only one facial sub-region pair, Forehead-Mouth, is significantly different for the emotion fear. We can conclude that the differences between different facial sub-regions are not significant for any of the emotional states when using the VAR statistics. This indicates that the overall change in the facial temperature occurs during the emotional state transfer. In some individual cases, for certain emotional states, a change in temperature may be caused by the combined influence of psychological factors or physiological factors related to facial muscular movements. This will be studied in depth in our future research.

ANOVA Analysis of the Different Emotion States in Different Facial Regions Using VAR: In this ANOVA analysis, the factors are the emotion states, with the dependent variable being VAR. The degree of impact is measured as the mean of the VAR values for different emotional states in each facial sub-region. The analysis results are given in Fig. 11 and Table IX.

- From Fig. 11, we can conclude that, when using the VAR statistic, the sorted degree of impact of these emotional states, i.e., Fear-Surprise-Happiness, is similar in the forehead and cheek regions, but not in the nose and mouth regions. This means that the degree of impact of the emotional states fear, surprise, and happiness in most of the facial sub-regions is more significant than that of the other emotional states. The exception of the nose and mouth regions may be due to a change in the breathing rate or a non-stationary physiological signal for different emotional states [27], [28]. Besides, muscular movement in the mouth region for different facial expressions may also be a contributing factor [29].

TABLE IX
MATRIX SHOWING SIGNIFICANT DIFFERENCES BETWEEN DIFFERENT EMOTIONAL STATES FOR EACH FACIAL SUB-REGION

Sig-Diff	Forehead	Nose	Mouth	Cheek
Sadness-Anger	0.90	0.89	0.70	0.74
Sadness-Surprise	0.31	0.73	0.02*	0.05*
Sadness-Fear	0.02*	0.62	0.06	0.04*
Sadness-Happiness	0.65	0.71	0.06	0.39
Sadness-Disgust	1.00	0.24	0.45	0.77
Anger-Surprise	0.25	0.84	0.04*	0.10
Anger-Fear	0.02*	0.72	0.14	0.08
Anger-Happiness	0.56	0.82	0.13	0.60
Anger-Disgust	0.90	0.30	0.71	0.99
Surprise-Fear	0.20	0.88	0.60	0.90
Surprise-Happiness	0.57	0.98	0.61	0.26
Surprise-Disgust	0.31	0.41	0.10	0.09
Fear-Happiness	0.07	0.90	0.99	0.21
Fear-Disgust	0.02*	0.50	0.26	0.07
Happiness-Disgust	0.65	0.42	0.25	0.58

* The mean difference is significant at the 0.05 level.

- From Table IX, we can conclude that when using the VAR statistic, three emotional state pairs are significantly different in the forehead region (Sadness-Fear, Anger-Fear, and Fear-Disgust), two emotional state pairs in the mouth region (Sadness-Surprise and Anger-Surprise), and two emotional state pairs in the cheek regions (Sadness-Surprise and Sadness-Fear) at the 0.05 level. All these differences may be caused by the combined influence of the emotional state change, the factor of the facial muscular movements, and various other psychological factors [29].

VII. CONCLUSION

The NVIE database developed in this study for expression recognition and emotion inference has four main characteristics. 1) To the best of our knowledge, it is the first natural expression database containing both visible and infrared videos. As such, it can be used for visible, infrared, or multi-spectral natural expression analysis. 2) It contains the facial temperature of subjects, thereby providing the potential for emotion recognition. 3) Both posed and spontaneous expressions by the same subject are recorded in the database, thus supplying a valuable resource for future research on their differences. 4) Both lighting and glasses variations are considered in the database. This is useful for algorithm assessment, comparison, and evaluation.

We carried out an elementary assessment of the usability of the spontaneous sub-database, using four baseline algorithms for visible expression recognition and two baseline algorithms for infrared expression recognition, and provided reference evaluation results for researchers using the database. We also analyzed the relationship between facial temperature and emotion, and provided useful clues for emotion inference from thermal images.

ACKNOWLEDGMENT

The authors would like to thank all the subjects who participated in the experiments. The authors also would like to thank the Editor and the anonymous reviewers for their insightful comments.

REFERENCES

[1] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, pp. 259–275, 2003.

[3] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *Trans. Appl. Percept.*, vol. 6, no. 1, pp. 1–22, 2009.

[4] M. M. Khan, M. Ingleby, and R. D. Ward, "Automated facial expression classification and affect interpretation using infrared measurement of facial skin temperature variations," *ACM Trans. Autonom. Adapt. Syst.*, vol. 1, pp. 91–113, 2006.

[5] A. T. Krzywicki, G. He, and B. L. O'Kane, "Analysis of facial thermal variations in response to emotion—eliciting film clips," *Proce. SPIE—Int. Soc. Optic. Eng.*, vol. 7343, no. 12, pp. 1–11, 2009.

[6] [Online]. Available: http://emotion-research.net/wiki/Databases.

[7] M. Pantic and M. Stewart Bartlett, *Machine Analysis of Facial Expressions, in Face Recognition*, K. D. a. M. Grgic, Ed. Vienna, Austria: I-Tech Education and Publishing, 2007, pp. 377–416.

[8] A. J. O'Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, and H. Abdi, "A video database of moving faces and people," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 812–816, May 2005.

[9] N. Sebe, M. S. Lew, I. Cohen, Y. Sun, T. Gevers, and T. S. Huang, "Authentic facial expression analysis," in *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004.

[10] E. Douglas-Cowie, R. Cowie, and M. Schroeder, "The description of naturally occurring emotional speech," in *Proc. 15th Int. Conf. Phonetic Sciences*, Barcelona, Spain, 2003, pp. 2877–2880.

[11] G. I. Roisman, J. L. Tsai, and K. S. Chiang, "The emotional integration of childhood experience: Physiological, facial expressive, and self-reported emotional response during the adult attachment interview," *Development. Psychol.*, vol. 40, no. 5, pp. 776–789, 2004.

[12] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR '05)*, 2005, pp. 568–573.

[13] R. Gajsek, V. Struc, F. Mihelic, A. Podlesek, L. Komidar, G. Socan, and B. Bajec, "Multi-modal emotional database: AvID," *Informatica 33*, pp. 101–106, 2009.

[14] K. R. Scherer and G. Ceschi, "Lost luggage emotion: A field study of emotion-antecedent appraisal," *Motivation and Emotion*, vol. 21, pp. 211–235, 1997.

[15] K. Delac, M. Grgic, and S. Grgic, "Independent comparative study of PCA, ICA, and LDA on the FERET data set," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 5, pp. 252–260, 2005.

[16] [Online]. Available: http://www.image.ntua.gr/ermis/.

[17] [Online]. Available: http://emotion-research.net/download/vam.

[18] D5i: Final Report on WP5, IST FP6 Contract no. 507422, 2008.

[19] [Online]. Available: http://www.equinoxsensors.com/products/HID.html.

[20] [Online]. Available: http://www.cse.ohio-state.edu/OTCBVS-BENCH/Data/02/download.html.

[21] [Online]. Available: http://www.terravic.com/research/facial.htm.

[22] J. Rottenberg, R. R. Ray, J. J. Gross, J. A. Coan, and J. J. B. Allen, *Handbook of Emotion Elicitation and Assessment*. New York: Oxford Univ. Press, 2007.

[23] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, 1991, pp. 586–591.

[24] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[25] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Proc. IEEE Int. Workshop Analysis and Modeling of Faces and Gestures*, 2003, pp. 157–164.

[26] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[27] B. R. Nhan and T. Chau, "Infrared thermal imaging as a physiological access pathway: A study of the baseline characteristics of facial skin temperatures," *Physiol. Measure.*, vol. 30, no. 4, pp. N23–N35, 2009.

[28] H. Tanaka, H. Ide, and Y. Nagashuma, "An attempt of feeling analysis by the nasal temperature change model," in *Proc. 2000 IEEE Int. Conf. Systems, Man, and Cybernetics*, 2000, vol. 1, pp. 1265–1270.

[29] Y. Sugimoto, Y. Yoshilomi, and S. Tomita, "A method for detecting transitions of emotional states using a thermal facial image based on a synthesis of facial expressions," *Robot. Autonomous Syst.*, vol. 3I, no. 3, 2000.

[30] A. Merla and G. L. Romani, "Thermal signatures of emotional arousal: A functional infrared imaging study," in *Proc. IEEE 29th Annu. Int. Conf.*, 2007, pp. 247–249.

[31] S. Jenkins, R. Brown, and N. Rutterford, "Comparing thermographic, EEG, and subjective measures of affective experience during simulated product interactions," *Int. J. Design*, vol. 3, no. 2, pp. 53–65, 2009.

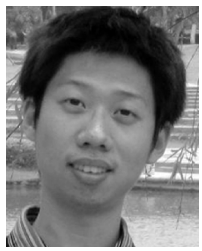
[32] S.-J. Chung, "Lexpression et la perception de l.emotion extraite de la parole spontanee: Evidences du coreen et de l'anglais," , Univ. Sorbonne Nouvelle, Paris, France, 2000.

Shangfei Wang (M'02) received the M.S. degree in circuits and systems and the Ph.D. degree in signal and information processing from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1999 and 2002.

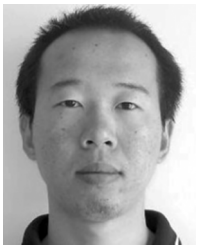
From 2004 to 2005, she was a postdoctoral research fellow with Kyushu University, Fukuoka, Japan. She is currently an Associate Professor with the School of Computer Science and Technology, USTC. Her research interests cover computation intelligence, affective computing, multimedia computing, information retrieval, and artificial environment design. She has authored or coauthored over 40 publications.

Zhilei Liu received the B.S. degree in information and computing science from Shandong University of Technology, Zibo, Shandong, China, in 2008. He is pursuing the M.S. degree in the School of Computer Science and Technology of the University of Science and Technology of China, Hefei, Anhui, China.

His research interest is affective computing.

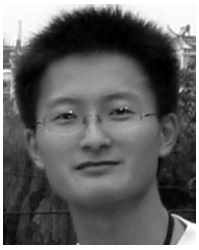


Siliang Lv received the B.S. degree in computer science and technology from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2008. He is pursuing the M.S. degree in the School of Computer Science and Technology of USTC. His major is affective computing.



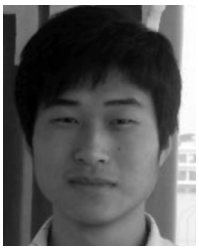
Yanpeng Lv received the B.S. degree in software engineering from Shandong University, Jinan, China, in 2008. He is currently pursuing the M.S. degree in computer science in the University of Science and Technology of China, Hefei, Anhui, China.

His research interest is affective computing.



Guobing Wu received the B.S. degree in computer science and technology from Anhui University, Hefei, Anhui, China. He is pursuing the M.S. degree in the School of Computer Science and Technology of the University of Science and Technology of China, Hefei, Anhui, China.

His research interest is affective computing.



Peng Peng received the B.S. degree in computer science and technology from the University of Science and Technology of China, Hefei, Anhui, China, in 2010. He is pursuing the M.S. degree in the School of Computing Science at Simon Frazer University, Burnaby, BC, Canada.

Fei Chen received the B.S. degree in computer science and technology from the University of Science and Technology of China, Hefei, Anhui, China, in 2010. He is pursuing the M.S. degree in the School of Computing Science at the University of Hong Kong.

Xufa Wang received the B.S. degree in radio electronics from University of Science and Technology of China, Hefei, China, in 1970.

He is currently a Professor of the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China, and the Director of the Key Lab of Computing and Communicating Software of Anhui Province. He has published five books and over 100 technical articles in journals and proceedings in the areas of computation intelligence, pattern recognition, signal processing, and computer networks.

Prof. Wang is an Editorial Board Member of the *Chinese Journal of Electronic*, the *Journal of Chinese Computer Systems*, and the *International Journal of Information Acquisition*.