Analyses of a Multimodal Spontaneous Facial Expression Database

Shangfei Wang, *Member*, *IEEE*, Zhilei Liu, Zhaoyu Wang,
Guobing Wu, Peijia Shen, Shan He, and Xufa Wang

Abstract—Creating a large and natural facial expression database is a prerequisite for facial expression analysis and classification. It is, however, not only time consuming but also difficult to capture an adequately large number of spontaneous facial expression images and their meanings because no standard, uniform, and exact measurements are available for database collection and annotation. Thus, comprehensive first-hand data analyses of a spontaneous expression database may provide insight for future research on database construction, expression recognition, and emotion inference. This paper presents our analyses of a multimodal spontaneous facial expression database of natural visible and infrared facial expressions (NVIE). First, the effectiveness of emotion-eliciting videos in the database collection is analyzed with the mean and variance of the subjects' self-reported data. Second, an interrater reliability analysis of raters' subjective evaluations for apex expression images and sequences is conducted using Kappa and Kendall's coefficients. Third, we propose a matching rate matrix to explore the agreements between displayed spontaneous expressions and felt affective states. Lastly, the thermal differences between the posed and spontaneous facial expressions are analyzed using a paired-samples t-test. The results of these analyses demonstrate the effectiveness of our emotion-inducing experimental design, the gender difference in emotional responses, and the coexistence of multiple emotions/expressions. Facial image sequences are more informative than apex images for both expression and emotion recognition. Labeling an expression image or sequence with multiple categories together with their intensities could be a better approach than labeling the expression image or sequence with one dominant category. The results also demonstrate both the importance of facial expressions as a means of communication to convey affective states and the diversity of the displayed manifestations of felt emotions. There are indeed some significant differences between the temperature difference data of most posed and spontaneous facial expressions, many of which are found in the forehead and cheek regions.

Index Terms—Spontaneous facial expression, database construction, analysis

1 Introduction

With advances in the facial expression recognition field, more and more researchers have found that there is a great difference between the analysis of spontaneous and deliberately posed facial expressions. The former seem to have a much more profound theoretical and practical significance. Moreover, because most facial expression analysis is based on expression databases, constructing a database of spontaneous expressions is a key requirement in this field. Consequently, great progress has been made in recent years toward creating larger and more natural expression databases [1]. However, several problems still need to be solved before a spontaneous expression database can be constructed successfully. First, the emotion elicitation step deals with the problem of effectively evoking emotions from subjects who participate in the corresponding experiment. Second, segmenting and labeling facial images are very time-consuming and no standard measurements for facial expressions are available [2]. Furthermore, the data analyses may provide guidance/insights for expression and emotion recognition. The relationship between expression and affective state may help researchers infer emotions from facial expressions. An analysis of the differences between spontaneous and posed expressions may be helpful to both construct a spontaneous expression database and to recognize spontaneous expressions, as it may provide clues to data segmentation and labeling as well as discriminative feature selection for recognition [3].

This paper presents several first-hand analyses of a multimodal spontaneous expression database, the natural visible and infrared facial expressions (NVIE) database [4]. The NVIE database includes two subdatabases: a spontaneous expression database, which consists of spontaneous expressional image sequences from onset up to and including apex, and a posed expression database, which consists of both neutral and posed apex expression images. Each subdatabase includes visible and infrared images that were recorded simultaneously by two cameras under three different illumination conditions: left, front, and right illumination. The subjects' actual emotions experienced in the emotion-inducing experiments were recorded using two dimensions, valence and arousal, and labeled according to six basic basic affective states: happiness, disgust, fear, surprise, sadness, and anger. The same was done for the labels of the apex expression frames (Exp_Apex) and

- *The authors are with the School of Computer Science and Technology, University of Science and Technology of China (West Campus), No. 96, JinZhai Road Baohe District, Hefei, Anhui 230027, P.R. China.*
E-mail: {sfwang, xfwang}@ustc.edu.cn, (leivo, wazhy, guobing, speijia, shanhe)@mail.ustc.edu.cn.

Manuscript received 15 June 2012; revised 9 Sept. 2012; accepted 18 Sept. 2012; published online 28 Sept. 2012.
Recommended for acceptance by Q. Ji.
For information on obtaining reprints of this article, please send e-mail to: taffc@computer.org, and reference IEEECS Log Number TAFFC-2012-06-0046.
Digital Object Identifier no. 10.1109/T-AFFC.2012.32.

expression image sequences (Exp_Seq). After a brief introduction to the data collection and annotation procedure of NVIE, four kinds of analyses are presented. First, the analysis on the effectiveness of emotion solicitation of the videos selected in our experiments is conducted using the subjects' self-reported data. Second, an interrater reliability analysis of the raters' subjective evaluation data with respect to apex expression images and image sequences is conducted using Fleiss's Kappa coefficient and Kendall's coefficient of concordance [5], [6], [7]. Third, the agreements between spontaneous expressions and affective states are explored through a matching rate matrix (MRM) defined in this research. Finally, a paired-samples t-test is conducted to investigate the differences between posed and spontaneous facial expressions using four statistical parameters computed from the temperature difference between the neutral and apex infrared frame. These analyses have produced conclusions that can benefit both developers and users of facial expression databases.

The remainder of this paper is organized as follows: Section 2 presents a brief review of the existing analyses of natural facial expression database construction and the differences between posed and spontaneous facial expressions. The data collection and annotation procedure of our NVIE database are briefly introduced in Section 3. In Section 4, the effectiveness of the elicitation videos used in our experiments is analyzed based on the subjects' self-reported data. Analyses of the interrater reliability of raters' subjective evaluation data are presented in Section 5. Section 6 explores the agreements between displayed expressions and felt affective states. Section 7 investigates the temperature differences between the posed and spontaneous facial expressions. Finally, discussions and conclusions are presented in Sections 8 and 9.

2 Related Work

2.1 Analyses of Spontaneous Expression Database Construction

There are many existing databases of facial expressions, exhaustive surveys of which are detailed in [8] and [9]. Constructing such a database consists of two main steps: database collection, and annotation. In the first step, elicitation methods, elicitor selection, and elicitor evaluation are considered. In the second step, expression description, video segmentation, labeling, and interrater reliability are considered.

Currently, researchers primarily use one of three possible approaches to elicit spontaneous affective behaviors: human-human conversation [10], [11], [12], [13], [14], [15], [16], [17], human-computer interaction [18], [19], or emotion-inducing videos [8], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Because the NVIE database focuses only on facial expressions instead of on speech or language, the use of emotion-inducing videos is a suitable approach, particularly because the data sets do not include any facial changes caused by speech. Recent research has also noted the necessity of studying affect patterns in situ [2], for example, using intelligent tutoring systems [18], because this approach provides more meaningful interpretations. Although data collection using emotion-inducing videos provides limited experimental control, it has potential applications for the implicit tagging of videos [30], [31].

Few database constructors have attempted to analyze the effectiveness of their elicitation methods, although it is the first and the most fundamental step in creating a database. Psychologists have proposed using self-reported measures [32] to evaluate the efficacy of the elicitor. Coan and Allen, and Gross and Levenson have developed an emotion stimulus video set and adopted the mean rating of subjects' self-reported data for each emotion state to select videos and validate their efficacy by discriminability, discreteness, and similarity [26], [32]. Petrantonakis and Hadjileontiadis [33] use electroencephalogram (EEG)-based emotion recognition as an emotion elicitation evaluation measure inspired by the frontal brain asymmetry concept. We believe the effectiveness of the emotion elicitor relies on subjects' elicited emotions, including a psychological response (e.g., subjects' self-reported data), a physiological response (such as EEG), and some visible clues (such as facial expressions). We do believe that self-reported data are a good starting point and provide a foundation for determining whether such elicitors also produce behavioral and physiological signs of the target emotion [32]. Thus, in this paper, we analyze the effectiveness of emotion-eliciting videos using the mean and variance of the experiment participants' self-reported data.

When constructing a spontaneous facial expression database, it can be expensive to manually segment the apex expression images or expression image sequences from the videos recorded during the emotion-eliciting experiments. There are three main methods for segmentation: using only the apex expression image [23], using expression image sequences of a constant time length [15], [19], [22], and using expression image sequences of variable time length [17], [18], [20], [21], [24], [28] that portray one or more neutral-expressive-neutral facial behavior patterns. Because expressions vary in length, with some occurring over a few frames and others lasting many seconds, sequences of variable length are preferred for this research. In this paper, both the apex expression images and the expression image sequences with variable time length from the neutral frame to the apex expression frame are segmented from subjects' video records.

After segmentation, raters must give expression labels to the sequences or apex images according to a set of categories, dimensions, or action units [29], [34]. Based on the theory of emotion, the widely used expression descriptors adhere to the categorical approach [20], [21], [27], [35], [14], the dimensional approach [13], [24], [36], [37], and an appraisal-based approach [38]. Six basic expression categories (happiness, disgust, fear, sadness, surprise, and anger) as well as the two most-used dimensions (valence and arousal) are used for the NVIE database in this paper.

Both the segmentation and labeling of emotions or expressions are normally performed using post hoc self-reports [14], [18], [23], [28], [39], [40] of the subjects or the subjective reports of multiple raters [10], [13], [14], [15], [16], [18], [19], [20], [21], [39], [41]. If these multiple raters' evaluation categories are different, a consistency strategy, such as a majority rule [41], must be used to determine the final category. The Kappa coefficient and the Kendall's coefficient are widely used to analyze the interrater reliability of multiple observers' evaluations [12], [41], [42]. In this paper, self-reports are used to indicate each subjects' truly induced feelings immediately after watching the emotion-eliciting videos. An observer first segments the expression videos to isolate clips containing affective-cognitive states of interest. The segmented clips are then evaluated by several raters. Then, the interrater reliability of raters' evaluations is examined using Kappa and Kendall's coefficients.

Some psychologists believe that facial expressions have a primarily communicative function in conveying information about affective states [43]. Therefore, the study of the relationship between naturally displayed facial expressions and felt affective states is significant for research on emotional reasoning and the analysis of facial expressions. This paper provides the first quantitative analysis of the degree of agreement between subjects' internal feelings and their displayed facial expressions using our proposed MRM.

2.2 Analyses of the Differences between Posed and Spontaneous Facial Expressions

It would be helpful to distinguish between spontaneous and posed facial expressions for real-life human-robot communications. It could also provide valuable references for spontaneous or posed facial expression elicitation or recognition. Present research has shown that spontaneous facial expressions differ from posed ones in appearance, timing, and accompanying head movements. Most studies have been performed based on the visible image domain, in which some geometric features in certain special regions, such as the brow or eyes [44], [45], [46], [47], temporal information [48], or facial action units are applied to distinguish between expressions, especially the smile or facial expressions indicating happiness [46], [47], [48], [49]. Because physiological changes evoked by the autonomic nervous system may also provide hints to distinguish between spontaneous and posed expressions, the temperature variances reflected through noninvasive infrared thermal images may be useful as well. Khan et al. [50] have conducted experiments on the recognition of posed and evoked facial expressions from thermal images. However, until this point, no research has reported an investigation on the differences between spontaneous and posed facial expressions through infrared thermal images [50]. In this paper, a paired-samples t-test is conducted to explore the differences between posed and spontaneous facial expressions in different facial regions as reflected by the temperature difference between the neutral frames and the apex expression frames.

Compared to the related research mentioned above, our contributions in this paper can be summarized as follows: We use our self-built large-scale NVIE database to perform a variety of comprehensive first-hand data analyses. These are the first reported analyses of spontaneous facial expression database construction to the best of our knowledge, and may provide references and insights for future spontaneous facial expression database construction and expression analysis. Despite the importance of evaluating the effectiveness of affect elicitors, it is not frequently studied. We propose to analyze the effectiveness of emotion-eliciting videos using the mean and variance of subjects' self-reported data. We are the first to introduce a matching relationship matrix to explore the relationship between the inner emotional state according to the subjects' self-reported data and the outer expression categories reflected by the raters' subjective evaluation data.

We are also the first to analyze the differences between spontaneous and posed facial expressions by measuring variations in temperature in each facial subregion using infrared thermal images.

Several interesting conclusions are reached from these analyses, and lessons and suggestions are proposed for future database construction and expression/emotion recognition.

3 NVIE Construction Procedure

3.1 Subjects

Two hundred and fifteen healthy subjects were recruited to participate in the experiment, including 157 males and 58 females ranging in age from 17 to 31. Each participant signed an informed consent form before the experiment and received compensation for their participation after completion. Each subject participated in three emotion-inducing experiments under different illumination conditions: front, left, and right illumination. Then, each subject was asked to display six expressions: happiness, sadness, surprise, fear, anger, and disgust. Some samples were omitted from the final released database because fewer expressions or emotions were elicited successfully or because of experimental device errors. Samples gathered from approximately 100 subjects, approximately 71 percent of whom wore glasses, were obtained under each of the three illumination conditions.

3.2 Stimuli

According to our own cultural environment and in consultation with our in-house psychologist, we selected different kinds of emotional videos from the Internet, including 13 intended to elicit happiness, 8 anger, 45 disgust, 7 fear, 7 sadness, 7 surprise, and 32 emotionally neutral videos, as judged by the authors. Each emotional video was approximately 3 to 4 minutes long. These videos were collated into several playlists, each of which contained six segments corresponding to the six types of emotions. To reduce the interaction between different emotions induced by the emotional video clips, neutral clips approximately 1 to 2 minutes in length were shown between emotional segments. For each experiment, the order of the playlist for a particular subject was randomly shuffled to avoid any impact on feelings caused by the playlist order. Thus, the subject did not know the content of the video in advance.

3.3 Database Collection

Before the experiment, subjects were informed about the experimental process, the meanings of arousal and valence, and how to assess the emotions they experienced throughout the experiment. Two cameras were used to record subjects' facial images: a DZ-GX25M visible-light camera capturing 30 frames per second with a resolution of 704×480 pixels and a SAT-HY 6850 infrared camera capturing

25 frames per second with a resolution of 320×240 pixels, sensitivity $0.08°\mathrm{C}$ and wave band 8 to 14 $\mu\mathrm{m}$. After all of the experimental devices had been set up, the subjects made themselves comfortable in chairs that were provided. The LCD screen was set to play the prepared playlist chosen by the authors. After watching a video clip, subjects were asked to report the emotion they actually experienced using emotional valence and arousal values ranging from -2 to 2, representing negative to positive valence and calm to exciting arousal, respectively [35], [51], [52]. Subjects also rated the intensity of the six basic emotion categories, which ranged from 0 to 4, where 0 indicates no particular feeling and 4 indicates a strong feeling. The subject's report is taken as his/her emotion as a result of watching the stimulus video. Only the emotion category with the highest evaluation value was selected as the main emotion category. More details regarding the image acquisition setup and the data acquisition procedure can be found in [4].

3.4 Database Annotation

After the spontaneous facial expression videos of the subjects were recorded, an observer manually identified the onset frame and apex frame with the most exaggerated expression in the visible facial video, which could feature multipeak expressions. Then, both the visible and infrared thermal videos during these onset and apex expression frames were segmented into image frames. From each posed facial expression video, the observer manually selected one onset and one apex frame. Thus, our NVIE database includes two subdatabases: a spontaneous database consisting of image sequences from onset to apex and a posed database consisting of neutral and apex expression images, with each containing both the visible and infrared images that were recorded simultaneously under three different illumination conditions (left, front, and right). The posed database also includes expression images with and without glasses for each subject.

Following the practice used by other database developers [13], [18], [22], [53], we employed experimenters for expression annotation. Specifically, 12 raters, who are familiar with facial expression recognition and analysis as well as the experiment's design and purpose, participated in the expression labeling. The presentation order of the visible expressional apex images or expressional sequences was randomized to avoid any ordering effect on the raters. Each apex image was rated by five raters. All raters evaluated each apex image using an intensity value for the six expression categories (happiness, sadness, surprise, fear, anger, and disgust), as well as arousal and valence values registered on a three-point scale. The intensity scale for the expression categories featured the values 0, 1, and 2, and that for arousal and valence featured the values $-1, 0$, and 1; these scales were much narrower than the scale used in the subjects' self-reported data described in Section 3.3 because raters generally reported that they were more uncertain about their evaluation and were more exhausted when using a five-point scale versus a three-point scale. One month later, three raters evaluated the visible facial image sequences using the same scale. Finally, the expression label of each apex expression image or expressional image sequence was determined based on the highest average evaluated value for the six expression categories, and the average arousal and valence values were used as labels to represent the other aspects of the expression.

Just like the expression categories, the affective states of these samples were labeled based on the highest evaluated value among the six basic affective states based on the subjects' self-reported data.

The NVIE database is freely available for research purposes only at http://nvie.ustc.edu.cn, more details about database annotation can be found in [4].

4 Analysis on the Effectiveness of Eliciting Videos

Given the introduction to the NVIE database and the methods used for its construction, we are now ready to discuss the analyses we conducted on the database. Effective emotion elicitation is important for database construction. In this section, we evaluate the emotion elicitation method. Specifically, based on the subjects' self-report of their emotions, we evaluate the effectiveness of emotion-elicitation videos.

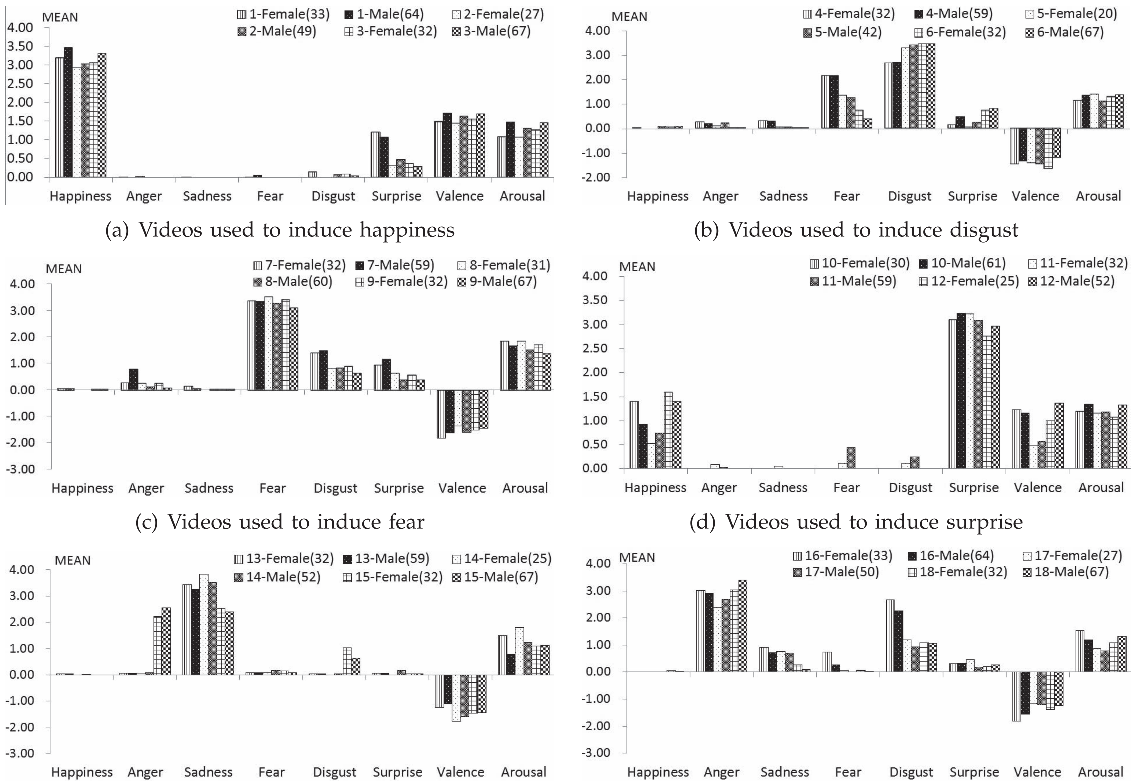
4.1 Methodology

In the NVIE database, the subjects' self-reported data for the six basic emotional states as well as arousal and valence values for each emotion-eliciting video can be regarded as a measure of the videos' ability to induce emotion. First, the self-reported data of the subjects included in the final NVIE database for the 18 most-viewed emotion-eliciting videos are selected. Second, two kinds of statistical parameters for the self-evaluated data of each video are computed. One is the mean evaluation value of each emotion state as well as the valence and arousal, which reflect the overall evaluation results for each emotion-eliciting video. The other is the variance, which reflects the degree of fluctuation in the subjects' self-reported data for each emotion-eliciting video.

4.2 Results and Analyses

The mean and variance of subjects' self-reported data for each emotion-eliciting video are shown in Figs. 1 and 2 for both genders. The following conclusions can be drawn from the results:

- Fig. 1 shows that overall the major emotional states expressed are the same as those intended to be elicited by the videos, which demonstrates the effectiveness of our selected elicitors. This claim is further supported by the means of the valences, since only those videos meant to elicit positive emotions such as happiness and surprise have positive valences, and those eliciting negative emotions are negative.
- Regarding the mean value of the arousal, we find that all these videos' arousal values are positive because to arouse the interest of the subject and to induce expressions and emotions successfully, the videos selected for the experiments should have higher arousal values. This is a common phenomenon in most emotion-eliciting experiments performed using emotional videos [52], [54].



(a) Videos used to induce happiness

(b) Videos used to induce disgust

(c) Videos used to induce fear

(d) Videos used to induce surprise

(e) Videos used to induce sadness

(f) Videos used to induce anger

Fig. 1. The mean value of the subjects' self-reported data for eliciting videos. (1. Happy_1.flv; 2. Happy_3.flv; 3. Happy_6.flv; 4. Disgust_3.flv; 5. Disgust_7.avi; 6. Disgust_8.flv; 7. Fear_3.flv; 8. Fear_5.flv; 9. Fear_6.flv; 10. Surprise_2.flv; 11. Surprise_3.flv; 12. Surprise_5.flv; 13. Sad_2.flv; 14. Sad_4.flv; 15. Sad_5.flv; 16. Anger_1.flv; 17. Anger_3.flv; 18. Anger_5.flv. the contents of these videos can be found in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2012.32. The numbers in parentheses are the number of subjects used for these analyses.)

- Fig. 1 shows that emotional videos often induce multiple emotions. For instance, the videos that induce happiness may also indicate some degree of surprise. The videos that induce disgust may also induce some degree of fear. The videos that induce fear may also induce some degree of disgust and surprise. The videos that induce surprise may also induce some degree of happiness. The videos that induce anger may also induce sadness and disgust. This is consistent with previous study results described in [32]. Moreover, it should be noted that the co-occurrence of emotion pairs is not symmetric; for example, when we try to elicit anger, disgust might be elicited simultaneously. However, when we try to elicit disgust, the major co-occurring emotion is fear instead of anger. This phenomenon of multiemotion co-occurrence reflected from the participants' self-evaluated data could also affect emotion or expression recognition because the emotion or expression labels of the samples are determined by these self-reported values. This finding suggests that for emotion and expression recognition, it may be more useful to label each video/image with multiple emotions of different intensities instead of labeling each video/image only with the dominant emotion/expression as most of the existing annotations do. The emotion/expression descriptors are not limited to six basic categories. They can be new categories. Action units are also appropriate descriptors. Multiemotion labeling allows us to capture the statistical dependencies among emotions. These captured dependencies may be used as prior information for emotion recognition.
- Fig. 2 shows that the variance values of most videos used to elicit happiness and sadness are smaller than those of videos used to elicit the other emotions. This may be because the explicitness of these videos' emotion-eliciting purpose is greater than that of other videos. One exception is the video labeled Sad 5.flv: This video's ability to elicit sadness is not as good as that of the other sad videos. This ineffectiveness is also reflected in this video's mean values for these six emotion categories. Although the main emotion that the video tries to induce is sadness, the means of other emotions, such as anger and disgust, are still very high, as described in Fig. 1.

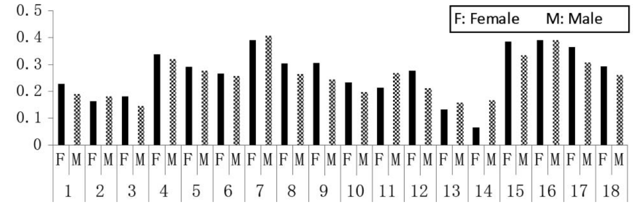


Fig. 2. The variance of the subjects' self-reported data for each emotion-eliciting video. (The meaning of 1 to 18 is the same as that in Fig. 1.)

- During the data collection procedure, each subject participated in three experiments under different lighting conditions. The order of the three selected emotion-eliciting videos for each emotion is the same as the order of the three experiments. For most emotions, the variances of the selected emotion-eliciting videos decrease as the order changes, as shown in Fig. 2. This suggests that the latter emotion-eliciting videos are much better than the former ones, such as those for anger, fear, and disgust. There are two possible reasons for this. First, the authors selected more effective emotion-eliciting videos after observing the former experiments. It is also possible that, after attending one or two experiments, the participants became aware of the purpose of the experiment and hence might provide more precise evaluation data.
- The variance of induced emotion for females is greater than that for males in 13 out of 18 cases, which indicates that female subjects display more emotion than male subjects when they watch the same video. In particular, it is much easier to induce sadness and fear in females than in males, as evidenced by the fact that the mean data evaluated for sadness and fear in female subjects are greater than those evaluated for sadness and fear in male subjects. This may be because female subjects are more emotional and can be more easily affected by those emotions. On the other hand, the mean values evaluated for happiness in males are greater than those in females, indicating that it is easier to induce happiness in male subjects. This implies that gender differences may affect the induction of emotions, as has been seen in a past neurophysiological study [55]. The gender difference in response to emotional stimuli may help improve the study of emotion recognition. For example, it may be used as prior information during inference or as side information [56] during training to produce a better emotion classifier.
- From the analytical results of this section, we can define the most effective emotion-eliciting video as one for which the mean evaluation value of the intended emotion is as high as possible and for which the variance is as low as possible. Furthermore, to elicit more natural emotions, emotion-eliciting videos with much broader arousal ranges should be selected. These results should provide helpful for video selection in future studies.

5 Analysis of Interrater Reliability

This section focuses on analyzing the emotion annotation method. Specifically, we study the interrater reliability for expression annotation. Two measures, Kendall's coefficient of concordance [5], [7] and Fleiss's Kappa coefficient [6], are used to evaluate interrater variability.

5.1 Methodology

We used Kendall's coefficient of concordance [5], [7] and Fleiss's Kappa coefficient [6] to examine interrater reliability because more than two observers labeled each apex facial image and sequence. The former is used when raters give ordinal ratings, while the latter is adopted when raters assign categorical ratings or classifications to a number of items. Here, Kendall's coefficient of concordance is used to analyze the original rating data for each expression category, while Fleiss's Kappa coefficient is used to analyze the maximum label, which is the category label with the maximum intensity value for the corresponding image or sequence.

Assume that there are K raters and N apex images or image sequences. For each expression category, the Kendall's coefficient of concordance can be calculated as follows [7]:

$$W = 12 \times \frac{\sum_{i=1}^{N} R_i^2 - \frac{\left(\sum_{i=1}^{N} R_i\right)^2}{N}}{K^2(N^3 - N) - K\sum_{m=1}^{K}\sum_{j=1}^{g_m}\left(n_{jm}^3 - n_{jm}\right)}, \quad (1)$$

where R_i is the sum of score ranks for the ith apex image or image sequence as rated by all raters. Images or sequences with the same score rated by the same rater are classified as images with a tied rank. g_m is the number of groups of ties for rater m. n_{jm} is the number of images or sequences with the jth group of tied ranks rated by rater m. The range of Kendall's coefficients is between 0 and 1. A value close to 1 indicates more agreement [7].

Fleiss's Kappa coefficient for multiple categories and multiple raters can be calculated as follows [6]:

$$kappa = \frac{P_a - P_{e|\pi}}{1 - P_{e|\pi}}, \quad (2)$$

where

$$\begin{cases} P_a = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{q}\frac{K_{ij}(K_{ij}-1)}{K(K-1)} \\ P_{e|\pi} = \displaystyle\sum_{j=1}^{q}\pi_j^2, \text{and } \pi_j = \frac{1}{N}\sum_{i=1}^{N}\frac{K_{ij}}{K}. \end{cases} \quad (3)$$

q represents the number of categories into which assignments are made; in our analysis, we set seven categories (six basic expressions and one other, when the scores of six expressions are equal). K_{ij} represents the number of raters who assign the ith image or image sequence to the jth category. P_a represents the observed consistency ratio, and $P_{e|\pi}$ denotes the expected consistency ratio. The value of Fleiss's Kappa ranges from -1 to 1. A value greater than 0.75 indicates high consistency. A value between 0.4 and 0.75 reflects general consistency. A value less than 0.4 indicates poor consistency [57].

5.2 Results and Analyses

The current analysis focuses on the authors' published database [4]. Table 1 presents the number of subjects and their images or sequences under different illumination conditions.

Table 2 lists the Kendall's coefficients of concordance for each expressional category under the three illumination conditions, while Table 3 lists the Fleiss's Kappa coefficient values for multiple raters under the three illumination conditions. The following conclusions can be drawn from the results:

- From Table 2, it is apparent that the Kendall's coefficient of concordance ranges from 0.546 to 0.955,

TABLE 1
Selected Samples for Analysis of Interrater Reliability

Lighting	Num of apex images and image sequences	Num of subjects
Front	576	102
Left	544	99
Right	546	103

TABLE 3
Fleiss's Kappa Coefficients under Three Illumination Conditions

Front		Right		Left	
Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq
0.6483	0.8167	0.5222	0.7047	0.5812	0.6855

indicating a fairly high level of consistency for each expression. Table 3 shows that Fleiss's Kappa coefficient ranges from 0.5222 to 0.8167, indicating that the raters' assessments of the data have a certain degree of consistency.

- Both the Kendall's and Kappa coefficients for expression sequences are larger than those for apex expression images. This may have occurred because sequences include more facial and head movements, which help observers more accurately ascertain the subjects' expressions. Furthermore, with image sequences, the uncertainty in the labels may be reduced and hence it is better to analyze expressions from image sequences than from single images.
- Both the Kendall's and Kappa coefficients for the three different illumination conditions are similar. This means that the different illumination conditions used in our experiments [4] seem to have no effect on the raters' assessment, although illumination conditions can often cause significant changes in facial appearance, posing challenges to expression annotation and recognition.
- From Table 2, we can see that happiness shows the highest level of agreement in most cases, which reflects that it is easier than other expressions for raters to label.
- A comparison of Table 2 and Table 3 shows that, in general, the Kendall's coefficients are greater than the Kappa coefficients, although they do not reflect the same parameters. This may indicate that labeling an expression image or image sequence with multiple categories together with their intensities is a better approach than labeling the expression image or image sequence into one category based on the maximum evaluation.

6 Agreement between Spontaneous Expressions and Affective States

In everyday communication, human affective states can be expressed through facial expressions, spoken tones, body gestures, and various physiological changes, such as blood flow, temperature, and heart rate. Among these emotional indicators, facial expressions play the most important role and are much more intuitively perceived than other nonverbal channels [43], [51], [58], [59], [60]. Therefore, an analysis of the relationship between naturally displayed facial expressions and felt affective states is important for emotion recognition from facial expressions. In this section, we use the data from our database to analyze this relationship.

6.1 Methodology

Because of the difficulty in evaluating or monitoring expression variations during the entire video-recording process, we use the expression label of the most exaggerated apex expression image or sequence as segmented in Section 3 to represent the major expression category of the subject while he/she is watching the emotion-eliciting video. The major inner affective state is determined by the highest emotion state of reported by the participant provided after watching the emotion-eliciting video. The relationships between the outer spontaneous facial expressions and inner affective states of the subjects are displayed using our proposed MRM, inspired by the confusion matrix in the field of pattern recognition.

Assuming that there are N samples, the kth sample $S_k = <S_k.Exp_k, S_k.Emt_k>$, where $S_k.Exp_k$ and $S_k.Emt_k$ represent the sample's expression label and affective label, respectively. The MRM is calculated according to

$$MRM(i,j) = \frac{C(i,j)}{SNum(j)}, \tag{4}$$

where

$$C(i,j) = \sum_{k=1}^{N} \{(S_k.Exp_k == i) \text{ and } (S_k.Emt_k == j)\}. \tag{5}$$

$C(i,j)$ is the total number of samples with expression label i, as determined by the highest average evaluated value for the six basic expressions in the expression-labeling step, and with affective label j, as determined based on the largest evaluated value for the six basic affective states in the self-reported data for this sample. $SNum(j)$ is the total number of samples with affective label j. $MRM(i,j)$ represents the matching rate of the sample with expression label i and affective label j.

6.2 Results and Analyses

Evaluated data for 1,658 apex images and image sequences with their corresponding self-reported data are used in

TABLE 2
Kendall's Coefficients of Concordance for Each Expression under Three Illumination Conditions

	Happiness		Disgust		Fear		Surprise		Anger		Sadness	
	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq	Exp_Apex	Exp_Seq
Front	0.945	0.955	0.703	0.847	0.74	0.92	0.701	0.867	0.658	0.841	0.637	0.856
Left	0.89	0.921	0.634	0.781	0.666	0.847	0.749	0.877	0.546	0.765	0.566	0.734
Right	0.905	0.933	0.706	0.841	0.614	0.813	0.762	0.804	0.591	0.748	0.621	0.718

TABLE 4
Selected Samples for Analysis of Matching Relationships between Affective States and Expressions

	Happ.	Disg.	Fear	Surp.	Ang.	Sad.	Other	Subject Num
Front	104	127	94	95	52	99	0	102
Left	117	103	87	82	56	88	8	99
Right	111	110	80	95	117	33	0	103

this section. The detailed sample information is shown in Table 4.

The MRM analysis of the relationships between affective states and Exp_Apex and between affective states and Exp_Seq are shown in Tables 5 and 6, respectively. We can draw the following conclusions from these results:

- The situation in which an affective state and its expression are both the same is referred to as "agreement." In this analysis, the ranking of the affective states based on the agreement rate (which is the same as $MRM(i, i)$) for Exp_Apex is happiness (0.892), sadness (0.755), surprise (0.724), fear (0.651), disgust (0.641), and anger (0.636), and for Exp_Seq is happiness (0.895), sadness (0.786), fear (0.717), surprise (0.680), anger (0.640), and disgust (0.626). This reflects the extent to which each displayed expression reflects a felt affective state using the apex expression images. In other words, happiness is the affective state that is the easiest to express and to elicit [8], [26], [32]. Furthermore, the agreement between an inner feeling of happiness and its outward expression is the highest among the six expressions. Happiness emotion is, therefore, easier to recognize. On the other hand, negative expressions like anger and disgust are harder to recognize since anger is hard to elicit and the expression of disgust may vary significantly by subject. This suggests that it is necessary to use extra information such as operating context, personal information, temporal context, or another sensory modality (e.g., physiological data or thermal image) to reliably recognize certain negative emotions.
- In addition to the principal expression for each affective state, certain other expressions also emerged in some cases. For example, when subjects felt pleasant surprise, their expressions exhibited a combination of happiness and surprise; when subjects felt a mixture of sorrow and anger, their expressions combined both sadness and anger. All of these phenomena are apparent in the analytical results shown in Tables 5 and 6, which reflect the ambiguities of both expressions and affective states as well as the diversity of the outer manifestations of affective states, especially for certain negative or positive affective states and expressions (sadness/anger/disgust/fear or happiness/surprise). These ambiguities may present challenges to emotion recognition using visual images alone. Multimodal emotion recognition may help reduce these ambiguities.

TABLE 5
Matching Relationships between Affective States and Exp_Apex

Emo Exp_Apex	Happ.	Disg.	Fear	Surp.	Ang.	Sad.	Other
Happ.	0.892	0.029	0.019	0.107	0.031	0.005	0.25
Disg.	0.018	0.641	0.165	0.066	0.062	0.041	0.125
Fear	0.006	0.074	0.651	0.055	0.013	0.018	0.125
Surp.	0.066	0.035	0.069	0.724	0.009	0.023	0.125
Ang.	0.006	0.165	0.050	0.007	0.636	0.155	0.25
Sad.	0.012	0.056	0.046	0.040	0.249	0.755	0.125
Other	0	0	0	0	0	0.005	0

TABLE 6
Matching Relationships between Affective States and Exp_Seq

Emo Exp_Seq	Happ.	Disg.	Fear	Surp.	Ang.	Sad.	Other
Happ.	0.895	0.024	0.008	0.162	0.022	0	0.375
Disg.	0.021	0.626	0.172	0.070	0.093	0.064	0.125
Fear	0	0.103	0.716	0.044	0.004	0.023	0.125
Surp.	0.069	0.021	0.034	0.680	0.009	0.005	0
Ang.	0.006	0.174	0.031	0.007	0.640	0.114	0.250
Sad.	0.009	0.041	0.027	0.029	0.2	0.786	0.125
Other	0	0.012	0.011	0.007	0.031	0.009	0

- The observed disparities between felt affective states and their displayed expressions could also be caused by individual differences both among the experimental subjects and among those who performed the expression labeling, which is to some extent inevitable.
- Furthermore, most of the agreement rates shown in Table 6 are higher than those in Table 5, which indicates that more emotional information is available from expression image sequences (e.g., muscular movements and head motions) than from single expression images, especially with respect to fear. This means that an image sequence is better for emotion recognition than a single apex image in most cases.

7 Analysis of the Difference between Spontaneous and Posed Facial Expressions in Infrared Thermal Images

Since our database consists of both posed and spontaneous facial expressions, we conducted an investigation into their differences. Such an analysis benefits not only the database developer but also the developers of algorithms for spontaneous expression recognition. Furthermore, the analysis is unique in that it is the first time that such an analysis has been performed in the IR image domain. The analysis will demonstrate the importance of thermal images in distinguishing between posed and spontaneous expressions. In the section below, we summarize the results of this analysis.

7.1 Methodology

First, 40 subject samples who exhibited six expressions under three lighting conditions both in the spontaneous and posed expression database were selected. Please note that samples from the same subject but under different lighting conditions are treated as different subject samples. Fig. 3 shows visible/thermal sample images, and the thermal

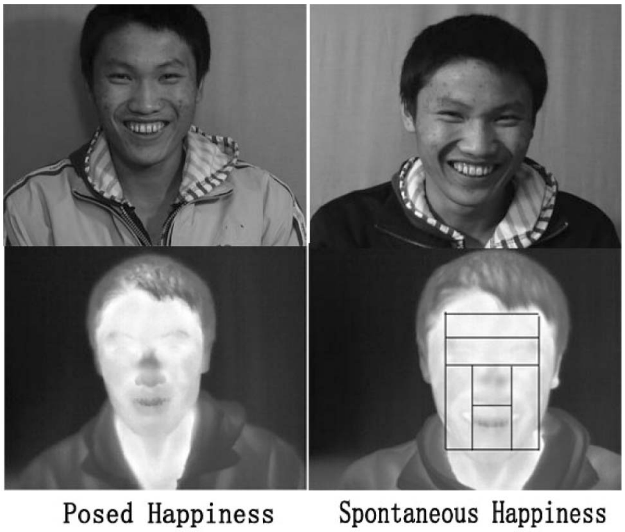


Fig. 3. Posed/spontaneous samples (happiness) and face segmentation.

images only are used in this section. To effectively obtain the temperature data gathered from their faces, subjects with eyeglasses and with long hair masking the facial region are excluded from this section. The posed and spontaneous pair numbers of the selected samples with different expressions can be found in Table 7, and more detailed subject information can be found in Appendix C, available in the online supplemental material. The expression labels of samples from the spontaneous database were determined based on the category with the highest mean subjective evaluation value for the visible apex expression images, as described in Section 3.4.

Second, to minimize the influences of the environmental temperature changes, the temperature shift of infrared thermal cameras, and the differences in individual vascular structure [61], the temperature difference data between the neutral and expressional frames are used. To obtain the temperature difference data for the facial region, we manually segmented the original infrared facial region as well as the corresponding temperature matrix into five subregions, namely forehead (F), eyes (E), nose (N), mouth (M), and cheeks (C), as shown in Fig. 3. We also ensure that the facial segmentation size and ratio of each sample's neutral and expressional frames are consistent.

To effectively quantify the temperature differences between the posed and spontaneous expressions, we employ four commonly used statistical parameters: the mean (MEAN), the variance (VAR), the mean of the absolute values (ABS), and the variance of the absolute value (ABSVAR) of the temperature difference matrix.

Finally, we conduct paired-samples t-test [62] to analyze the significant difference between the types of spontaneous and posed expression in each facial subregion.

TABLE 7
Selected Samples for Analysis of the Difference between Spontaneous and Posed Facial Expressions

Samples	Happ.	Disg.	Fear	Surp.	Ang.	Sad.
Number	37	25	13	29	28	31

TABLE 8
Significant Differences between Spontaneous and Posed Facial Expressions

Sig.	Parameter	F	E	N	C	M
Happ.	Mean	0.116	0.480	0.151	0.768	0.442
	ABS	**0.012***	**0.028***	0.127	0.068	0.137
	VAR	**0.046***	0.135	0.189	0.140	0.197
	ABSVAR	**0.023***	0.112	0.126	0.074	0.248
Disg.	Mean	0.356	0.217	0.657	0.471	0.367
	ABS	**0.008***	**0.009***	**0.050***	**0.001***	0.117
	VAR	**0.046***	0.052	0.118	**0.001***	0.144
	ABSVAR	0.064	0.071	0.134	**0.001***	0.190
Fear	Mean	0.215	0.521	0.680	0.288	0.719
	ABS	**0.050***	**0.033***	**0.033***	**0.009***	0.105
	VAR	0.069	0.076	0.229	**0.020***	0.469
	ABSVAR	0.063	0.068	0.329	**0.038***	0.863
Surp.	Mean	0.622	0.846	0.202	0.385	0.908
	ABS	**0.010***	**0.009***	**0.028***	**0.000***	**0.020***
	VAR	**0.033***	0.117	0.151	**0.002***	0.183
	ABSVAR	**0.031***	0.073	0.145	**0.003***	0.521
Ang.	Mean	0.426	0.367	0.202	0.161	0.417
	ABS	**0.008***	**0.014***	**0.030***	**0.002***	**0.032***
	VAR	**0.026***	**0.016***	0.131	**0.004***	**0.030***
	ABSVAR	**0.030***	**0.029***	0.143	**0.002***	**0.021***
Sad.	Mean	0.822	0.287	0.161	0.178	0.306
	ABS	**0.005***	**0.004***	**0.017***	**0.000***	**0.011***
	VAR	**0.036***	**0.034***	0.143	**0.000***	**0.008***
	ABSVAR	**0.032***	**0.035***	**0.040***	**0.000***	**0.005***

It means there exists difference at a 0.05 significant level.

7.2 Results and Analyses

The analytical results are gathered in Table 8. From Table 8, we can conclude the following:

- With respect to the facial regions, we find that the temperature patterns of the cheek and forehead are the most significant and important for distinguishing between spontaneous and posed facial expressions, as they have the highest significant difference (15 cases each). The importance of the temperature of the forehead in distinguishing spontaneous and posed expression corresponds with the research results of Merla and Romani and Jenkins et al. [63], [64]. This is vastly different from the results obtained from visible images, in which changes in the mouth and eyes are more important. This shows that the thermal spectrum can be a useful and necessary complement to the visible spectrum in distinguishing between spontaneous-posed facial expressions.
- With respect to facial expressions, we find that the significant differences between spontaneous and posed expressions are sorted as follows: sadness (14 cases), anger (13 cases), surprise (9 cases), disgust (7 cases), fear (6 cases), and happiness (4 cases). It is surprising that the smallest difference is observed between spontaneous and posed happiness, which is the most studied expression in the visible spectrum [44], [45], [46], [47], [48], [49]. However, we believe that there exist morphological and dynamic differences between posed and spontaneous happiness that are not clear in thermal images as they may be in visible images. One of the main reasons may be that the temporal information and head-motion features [44], [48], [50] between the neutral and apex expression frames are not considered in this analysis. Another reasonable

explanation may be that happiness, through such expressions as social and amusing smiles, is the most-expressed emotion in our daily communication. We express happiness or smile so much that when we are asked to give a posed happiness expression, the degree of exaggeration exhibited by facial muscle movement as well as the related blood-flow pattern of the posed expression may be very similar to those of the spontaneous one. This observation, however, is tentative and further investigation is needed to study its generalizability to other databases.

- With respect to the parameters explored in this study, we find that the most discriminative parameter in distinguishing posed and spontaneous expressions is ABS (25 cases), followed by ABSVAR (15 cases), VAR (14 cases), and MEAN (0 cases). This may indicate that absolute rather than simple temperature differences can better distinguish between spontaneous and posed expressions.

8 Discussions

The construction of a spontaneous facial expression/emotion database is a time-consuming, complex, and meticulous process [21]. There are many open-ended issues, such as the design of the natural experiment environment, the enrollment of the subjects, the selection of emotion-eliciting videos, and the labeling method of the participants' emotions or expression states [20], [21], [26]. Below, we have summarized some experiences and lessons that we have learned from our analyses that may be useful for both database and algorithm developers.

8.1 Observations for Database Developers

1. Diversity of subjects: According to the analysis in Section 4, there are gender differences in response to emotional stimuli. The same phenomenon may also apply to subjects of different ages, personalities, and occupations. Thus, multiethnic, multiage, multipersonality, and multioccupation subjects should be considered during the subject-enrollment process. This can help reduce database bias [39].
2. Measures of emotion elicitation effectiveness: Evaluating the effectiveness of elicitors is very important for database construction. In Section 4, we propose the mean and variance of self-report as elicitation effectiveness measures. Other methods and parameters to evaluate emotion elicitation effectiveness should also be investigated.
3. Diversity of elicitor: The video-based emotion elicitation method was adopted in our research and shows its effectiveness according to the analyses in Section 4. However, it has some inherent problems. For example, the subject's familiarity with the video clips may affect his/her emotion elicitation effectiveness. Also, the video-based emotion elicitation may not be effective for certain negative emotions [32]. Therefore, other complementary emotion elicitation methods such as reality shows should also considered if conditions permitting.
4. Requirements of segmentation: In this paper, we first chose the maximum apex image and then segmented the videos into image sequences of varying time lengths, including the onset frame and maximum apex expression frame. During the segmentation process, we observe multiple peaks in the spontaneous expression video. Thus, the spontaneous expressional image sequences should include the onset frame and all the peak frames, including the apex expression image. Moreover, because the purpose of constructing our database is to establish an emotional expression database, some samples with microexpressions are not included in the final database and are not taken into consideration in the analyses of this paper. This may omit some useful information, since it is common for people to experience emotions without expressing them. Therefore, the neutral images should also be included in the final database.
5. Requirement of multiple annotations: How to model, label, and interpret the subtlety, complexity, and continuity of affective behaviors in terms of dimensions or discrete emotional categories remains an open-ended issue in this field. Our analyses in Section 5 show that listing multiple categories together with their intensities may be better than one category. The coexistence of multiple emotions and expressions discussed in Section 6 also suggests that it is better to label videos with multiple emotions or expressions with intensities. Regardless, some standard methods should be established for emotional database builders to evaluate the effectiveness of their annotation.
6. Requirements of annotation over time: During our database construction, the subjects only provided their emotion labels once after watching the videos, and raters also labeled the expression image sequences once. Time-series labels from subjects and raters may be necessary considering the temporal and multipeak characteristics of spontaneous expression and emotion. When emotions change rapidly, subjects may have trouble describing this variation in posttreatment self-reports [65]. However, collecting emotion self-reports directly from subjects during the emotion-inducing process may interrupt the process. Therefore, it is important to develop a method in which subjects provide their self-reported data during the emotion-eliciting process without any interruption of the experiment.
7. Requirements of multimodality: The importance of the cheek and forehead regions discussed in Section 7 demonstrates the supplementary function of infrared thermal images to visible images in emotion analyses. The same situations may exist in other modalities due to the complexity and diversity of inner emotion manifestation. Thus, a multimodal emotional database [66], [67] capturing both internal physiological signals, such as EEG, electrocardiography (ECG), electromyography (EMG), and galvanic skin resistance (GSR), as well as external signals, such as facial expressions and body gestures, in both 2D and 3D [68], and so on, should be constructed to meet this demand.

8.2 Observations for Algorithm Developers

Our analyses also provide some suggestions for expression recognition and emotion inference as follows:

1. Expression/emotion recognition considering their co-occurrence: The coexistence of multiple emotions and expressions discussed in Sections 4 and 6 indicate that some emotions/expressions appear together with a high probability, while others do not. Such co-occurrence or mutually exclusiveness relationships may be exploited for emotion/expression recognition.
2. Emotion/expression modeling by exploiting gender differences. The gender differences in response to emotional stimuli mentioned in Section 4 may provide prior knowledge in emotion modeling. For example, the prior probability of emotions for male and female subjects may be different in the same situation. This observation is consistent with previous studies [55], [69]. The neurophysiological study by Lithari et al. [55] shows the existence gender differences in response to emotional stimulus. Furthermore, Town and Saatci's work [69] shows that there are gender-specific differences in the appearance of facial expressions, which can be exploited to improve the expression recognition performance. Thus, despite the challenges in gender classification, gender differences, if available, probably could be exploited as prior information during testing to improve classification or be exploited during training as side information to produce a better emotion/expression classifier.
3. The relationship between inner emotions and outer expressions: Based on the analysis results in Section 6, outer facial expressions can largely reflect inner emotional states. However, the relationships change depending on the emotion type. Positive inner emotional states correspond highly to their external expressions, but this is not true for certain negative emotions. This suggests that facial expression alone is not enough to recognize negative emotions. It should be augmented with additional contextual information or another sensory modality.
4. IR Thermal images: The observations in Section 7 demonstrate the supplementary role of infrared thermal images to visible images. The thermal images may benefit both expression recognition and help to distinguish between posed and spontaneous expressions.

9 Conclusion

This paper presents comprehensive analyses of a natural facial expression database. We performed four analyses on the effectiveness of emotion elicitation, the interrater reliability for expression annotation, the relationship between spontaneous expressions and affective states, and the differences between posed and spontaneous expressions.

The results of the first analysis reveal the effectiveness of our selected emotion-eliciting videos, the gender differences in emotional response, and the coexistence of multiple emotions.

The results of the second analysis indicate that the raters' assessments of the data have a certain degree of consistency. They also show that the facial image sequences are more informative than the apex images for expression recognition. Furthermore, labeling an expression image or sequence with multiple categories together with their intensities is a better approach than labeling them with one category.

The third analysis shows that the displayed apex expression images or expression image sequences are closely related to felt emotions. In fact, some emotions are highly correlated with the corresponding expressions, but the degree of these correlations varies with emotion type. The analysis also reveals the diversity of the manifestations of felt emotions.

The fourth analysis verifies that there exist differences between posed and spontaneous facial expressions in the thermal infrared spectrum, and most of these significant differences appear in the cheek and forehead regions.

We realize the limitations of our database and its inherent biases. As a result, some of the conclusions or observations we have derived may be preliminary and require further validation. Nevertheless, some of the observations from our analyses are apparently generalizable because they concur with the existing theories and findings. With all of the open questions in spontaneous expression/emotion database construction and expression/emotion analyses, we expect that the analyses described in this paper can serve as a reference for future researchers. It is our hope that this research can inspire others to perform similar analyses so that more generalizable and accurate conclusions can be derived to guide future database construction and facial expression/emotion recognition.

Acknowledgments

The authors would like to thank all of the subjects who participated in the experiments. This work was supported by National Program 863 (2008AA01Z122), the National Science Foundation of China (Grant No. 61175037), Special Innovation Project on Speech of Anhui Province (11010202192), the project from Anhui Science and Technology Agency (1106c0805008), and the Youth Creative Project of University of Science and Technology of China (USTC).

References

[1] R.A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Trans. Affective Computing,* vol. 1, no. 1, pp. 18-37, Jan.-June 2010.

[2] R. Cowie, "Building the Databases Needed to Understand Rich, Spontaneous Human Behaviour," *Proc. IEEE Int'l Eighth Conf. Automatic Face Gesture Recognition (FG '08),* pp. 1-6, Sept. 2008.

[3] M. Hoque, D.J. McDuff, L.-P. Morency, and R.W. Picard, "Machine Learning for Affective Computing," *Proc. Fourth Int'l Conf. Affective Computing and Intelligent Interaction,* pp. 567-567, 2011.

[4] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference," *IEEE Trans. Multimedia,* vol. 12, no. 7, pp. 682-691, Nov. 2010.

[5] C. Yu, *SPSS and Statistic Analysis.* Publishing House of Electronics Industry, 2007.

[6] K.L. Gwet, "Computing Inter-Rater Reliability and Its Variance in the Presence of High Agreement," *British J. Math. and Statistical Psychology,* vol. 61, no. 1, pp. 29-48, 2008.

[7] P. Legendre, "Species Associations: The Kendall Coefficient of Concordance Revisited," *Am. Statistical Assoc. and the Int'l Biometric Soc. J. Agricultural, Biological, and Environmental Statistics,* vol. 10, no. 2, pp. 226-245, 2005.

[8] 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 Analysis and Machine Intelligence,* vol. 31, no. 1, pp. 39-58, Jan. 2009.

[9] H. Association, "A Collection of Emotional Databases," http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/a m_tools_doc/index.html, 2012.

[10] G.I. Roisma, J.L. Tsai, and K.-H. S. Chiang, "The Emotional Integration of Childhood Experience: Physiological, Facial Expressive, and Self-Reported Emotional Response during the Adult Attachment Interview," *Developmental Psychology,* vol. 40, no. 5, pp. 776-789, 2004.

[11] S. Kollias, "Emotionally Rich Man-Machine Intelligent System," http://www.image.ntua.gr/ermis, 2012.

[12] Z. Zen, Y. Fu, G.I. Roisman, Z. Wen, Y. Hu, and T.S. Huang, "Spontaneous Emotional Facial Expression Detection," *J. Multimedia,* vol. 1, no. 5, pp. 1-8, 2006.

[13] G. McKeown, M.F. Valstar, R. Cowie, and M. Pantic, "The SEMAINE Corpus of Emotionally Coloured Character Interactions," *Proc. IEEE Int'l Conf. Multimedia and Expo (ICME '10),* pp. 1079-1084, July 2010.

[14] A. Zara, V. Maffiol, J.C. Marti, and L. Deviller, "Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics," *Proc. Second Int'l Conf. Affective Computing and Intelligent Interaction (ACII '07),* pp. 464-475, 2007.

[15] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *Proc. IEEE Fourth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 46-53, 2000.

[16] B. Schuller, R. Müller, B. Hörnler, A. Höthker, H. Konosu, and G. Rigoll, "Audiovisual Recognition of Spontaneous Interest within Conversations," *Proc. Ninth Int'l Conf. Multimodal Interfaces (ICMI '07),* pp. 30-37, 2007.

[17] M. Hoque, L.-P. Morency, and R.W. Picard, "Are You Friendly or Just Polite?—Analysis of Smiles in Spontaneous Face-to-Face Interactions," *Proc. Fourth Int'l Conf. Affective Computing and Intelligent Interaction,* Part I, pp. 135-144, 2011.

[18] S. Afzal and P. Robinson, "Natural Affect Data—Collection and Annotation in a Learning Context," *Proc. Third Int'l Conf. Affective Computing and Intelligent Interaction and Workshops (ACII '09),* pp. 1-7, Sept. 2009.

[19] M.E. Hoque, R.E. Kaliouby, and R.W. Picard, "When Human Coders (and Machines) Disagree on the Meaning of Facial Affect in Spontaneous Videos," *Proc. Ninth Int'l Conf. Intelligent Virtual Agents,* pp. 337-343, 2009.

[20] M. Pantic and M.S. Bartlett, "Machine Analysis of Facial Expressions," *Face Recognition,* pp. 377-416, I-Tech Education and Publishing, July 2007.

[21] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-Based Database for Facial Expression Analysis," *Proc. IEEE Int'l Conf. Multimedia and Expo (ICME '05),* pp. 317-321, July 2005.

[22] 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 Analysis and Machine Intelligence,* vol. 27, no. 5, pp. 812-816, May 2005.

[23] N. Sebe, M.S. Lew, I. Cohen, Y. Sun, T. Gevers, and T.S. Huang, "Authentic Facial Expression Analysis," *Proc. IEEE Sixth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 517-522, May 2004.

[24] http://emotion-research.net/download/vam, 2012.

[25] E. Douglas-Cowie, "D5i: Final Report on Wp5," Technical Report IST FP6 Contract no. 507422, 2008.

[26] J.A. Coan and J.J.B. Allen, *Handbook of Emotion Elicitation and Assessment.* Oxford Univ. Press, 2007.

[27] N. Aifanti, C. Papachristou, and A. Delopoulos, "The Mug Facial Expression Database," *Proc. 11th Int'l Workshop Image Analysis for Multimedia Interactive Services (WIAMIS),* pp. 1-4, Apr. 2010.

[28] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising Spontaneous Facial Micro-Expressions," *Proc. IEEE Int'l Conf. Computer Vision (ICCV),* pp. 1449-1456, Nov. 2011.

[29] http://www.engr.du.edu/mmahoor/DU-iFACS.htm, 2012.

[30] H. Joho, J.M. Jose, R. Valenti, and N. Sebe, "Exploiting Facial Expressions for Affective Video Summarisation," *Proc. ACM Int'l Conf. Image and Video Retrieval (CIVR '09),* pp. 31:1-31:8, 2009.

[31] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J.M. Jose, "Integrating Facial Expressions into User Profiling for the Improvement of a Multimodal Recommender System," *Proc. IEEE Int'l Conf. Multimedia and Expo (ICME '09),* pp. 1440-1443, 2009.

[32] J.J. Gross and R.W. Levenson, "Emotion Elicitation Using Films," *J. Cognition and Emotion,* vol. 9, no. 1, pp. 87-108, 1995.

[33] P.C. Petrantonakis and L.J. Hadjileontiadis, "A Novel Emotion Elicitation Index Using Frontal Brain Asymmetry for Enhanced EEG-Based Emotion Recognition," *IEEE Trans. Information Technology in Biomedicine,* vol. 15, no. 5, pp. 737-746, Sept. 2011.

[34] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978.

[35] P. Ekman and W.V. Friesen, "Constants across Cultures in the Face and Emotion," *J. Personality and Social Psychology,* vol. 17, no. 2, pp. 124-129, 1971.

[36] J.A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology,* vol. 39, no. 6, pp. 1161-1178, Dec. 1980.

[37] L. Shen, E. Leon, V. Callaghan, and R. Shen, "Exploratory Research on an Affective E-Learning Model," *Proc. Workshop Blended Learning (WBL '07),* pp. 267-278, Aug. 2007.

[38] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," *Int'l J. Synthetic Emotions,* vol. 1, no. 1, pp. 68-99, Jan. 2010.

[39] I. Sneddon, M. McRorie, G. McKeown, and J. Hanratty, "The Belfast Induced Natural Emotion Database," *IEEE Trans. Affective Computing,* vol. 3, no. 1, pp. 32-41, Jan.-Mar. 2012.

[40] R. Gajsek, V. Struc, F. Mihelic, A. Podlesek, L. Komidar, G. Socan, and B. Bajec, "Multi-Modal Emotional Database: Avid," *Informatica (Slovenia),* vol. 33, pp. 101-106, 2009.

[41] R. el Kaliouby and A. Teeters, "Eliciting, Capturing and Tagging Spontaneous Facialaffect in Autism Spectrum Disorder," *Proc. Ninth Int'l Conf. Multimodal Interfaces (ICMI '07),* pp. 46-53, Nov. 2007.

[42] V.K. Hinson, E. Cubo, C.L. Comella, C.G. Goetz, and S. Leurgans, "Rating Scale for Psychogenic Movement Disorders: Scale Development and Clinimetric Testing," *Movement Disorders,* vol. 20, no. 12, pp. 1592-1597, 2005.

[43] "Emotion and Facial Expression," http://face-and-emotion.com/dataface/emotion/expression.jsp, 2012.

[44] M.F. Valstar, M. Pantic, Z. Ambadar, and J.F. Cohn, "Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions," *Proc. Eighth Int'l Conf. Multimodal Interfaces (ICMI '06),* pp. 162-170, 2006.

[45] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating Spontaneous from Posed Facial Expressions within a Generic Facial Expression Recognition Framework," *Proc. IEEE Int'l Conf. Computer Vision Workshops,* Nov. 2011.

[46] M.F. Valstar, H. Gunes, and M. Pantic, "How to Distinguish Posed from Spontaneous Smiles Using Geometric Features," *Proc. Ninth Int'l Conf. Multimodal Interfaces (ICMI '07),* pp. 38-45, 2007.

[47] H. Dibeklioglu, R. Valenti, A.A. Salah, and T. Gevers, "Eyes Do Not Lie: Spontaneous versus Posed Smiles," *Proc. ACM Int'l Conf. Multimedia,* pp. 703-706, 2010.

[48] J. Cohn and K. Schmidt, "The Timing of Facial Motion in Posed and Spontaneous Smiles," *J. Wavelets, Multi-resolution and Information Processing,* vol. 2, pp. 1-12, 2004.

[49] H. Dibeklioglu, R. Valenti, A.A. Salah, and T. Gevers, "Classification of Spontaneous and Posed Smiles," *Proc. IEEE 19th Conf. Signal Processing and Comm. Applications (SIU),* pp. 1165-1168, Apr. 2011.

[50] 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," *ACM Trans. Applied Perception,* vol. 6, no. 1, pp. 6:1-6:22, Feb. 2009.

[51] M.P. Hatice Gunes and T. Jan, "Face and Body Gesture Analysis for Multimodal HCI," *Proc. Computer Human Interaction,* M. Masoodian S. Jones, and B. Rogers, eds. pp. 583-588, 2004.

[52] C. Breazeal, "Emotion and Sociable Humanoid Robots," *Int'l J. Human-Computer Studies,* vol. 59, no. 1/2, pp. 119-155, July 2003.

[53] E. Douglas-Cowiea, N. Campbellb, R. Cowiea, and P. Roach, "Emotional Speech: Towards a New Generation of Databases," *Speech Comm.,* vol. 40, no. 1, pp. 33-60, 2003.

[54] X. Jin and Z. Wang, "An Emotion Space Model for Recognition of Emotions in Spoken Chinese," *Proc. First Int'l Conf. Affective Computing and Intelligent Interaction (ACII '05),* pp. 397-402, 2005.

[55] C. Lithari, C. Frantzidis, C. Papadelis, A. Vivas, M. Klados, C. Kourtidou-Papadeli, C. Pappas, A. Ioannides, and P. Bamidis, "Are Females More Responsive to Emotional Stimuli? A Neurophysiological Study across Arousal and Valence Dimensions," *Brain Topography,* vol. 23, pp. 27-40, 2010.

[56] V. Vapnik and A. Vashist, "A New Learning Paradigm: Learning Using Privileged Information," *Neural Networks,* vol. 22, no. 5/6, pp. 544-557, 2009.

[57] J.L. Fleiss, *Statistical Methods for Rates and Proportions,* second ed. John Wiley, 1981.

[58] Z. Kasiran, S. Yahya, and Z. Ibrahim, "Facial Expression Recognition as an Implicit Customers' Feedback," *Advances in Human Computer Interaction,* Shane Pinder ed., InTech, 2008.

[59] N. Sebe and M.S. Lew, "Facial Expression Recognition," *Robust Computer Vision: Theory and Applications (Computational Imaging and Vision),* chapter 7, Kluwer Academic, 2003.

[60] C.D. Mortensen and A. Mehrabian, "Communicate without Words," *Comm. Theory,* second ed., chapter 13, transaction, Dec. 2007.

[61] P. Buddharaju, I. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-Based Face Recognition in the Thermal Infrared Spectrum," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 613-626, Apr. 2007.

[62] "Paired-Samples T-Test," http://www.statisticssolutions.com/resources/directory-of-statistical-a nalyses/paired-sample-t-test, 2012.

[63] A. Merla and G.L. Romani, "Thermal Signatures of Emotional Arousal: A Functional Infrared Imaging Study," *Proc. IEEE 29th Ann. Int'l Conf. Eng. in Medicine and Biology Soc. (EMBS '07),* pp. 247-249, Aug. 2007.

[64] S. Jenkins, R. Brown, and N. Rutterford, "Comparing Thermographic, EEG, and Subjective Measures of Affective Experience During Simulated Product Interactions," *Int'l J. Design,* vol. 3, pp. 53-65, 2009.

[65] C. Conati and H. Maclaren, "Empirically Building and Evaluating a Probabilistic Model of User Affect," *User Modeling and User-Adapted Interaction,* vol. 19, no. 3, pp. 267-303, Aug. 2009.

[66] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis; Using Physiological Signals," *IEEE Trans. Affective Computing,* vol. 3, no. 1, pp. 18-31, Jan.-Mar. 2012.

[67] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multi-Modal Database for Affect Recognition and Implicit Tagging," *IEEE Trans. Affective Computing,* vol. 3, no. 1, pp. 42-55, Jan.-Mar. 2012.

[68] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato, "A 3D Facial Expression Database for Facial Behavior Research," *Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognition (FGR '06),* pp. 211-216, 2006.

[69] Y. Saatci and C. Town, "Cascaded Classification of Gender and Facial Expression Using Active Appearance Models," *Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognition (FGR '06),* pp. 393-398, Apr. 2006.

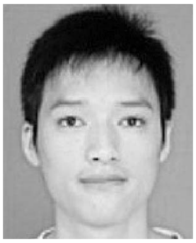
Shangfei Wang received the MS degree in circuits and systems and the PhD degree in signal and information processing from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1999 and 2002. From 2004 to 2005, she was a postdoctoral research fellow at Kyushu University, Japan. She is currently an associate professor in the School of Computer Science and Technology, USTC. Her research interests include computation intelligence, affective computing, multimedia computing, information retrieval, and artificial environment design. She has authored/coauthored more than 50 publications. She is a member of the IEEE.

Zhilei Liu received the BS degree from the School of Mathematics and Information Science, Shandong University of Technology, Zibo, Shandong, China, in 2008, and is currently working toward the PhD degree at the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China. His primary research interest is affective computing.

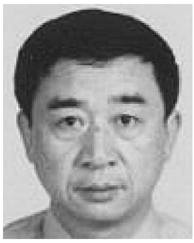
Zhaoyu Wang received the BS degree in computer science and technology from Anhui University of Technology, Ma Anshan, Anhui, China, and is currently working toward the MS degree at the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China. His primary research interest is affective computing.

Guobing Wu received the BS degree from the Department of Computer Science and Technology, Anhui University, Hefei, Anhui, China, and the MS degree from the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China, in 2012.

Peijia Shen received the BS degree from the Department of Computer Science and Technology, Hefei University of Technology, Hefei, Anhui, China, and the MS degree from the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China, in 2012.



Shan He received the BS degree in computer science from Anhui Agriculture University, Hefei, China, in 2010, and is currently working toward the MS degree in computer science from the University of Science and Technology of China, Hefei. His primary research interest is affective computing.



Xufa Wang received the BS degree in radio electronics from the University of Science and Technology of China (USTC), Hefei, China, in 1970. He is currently a professor in the School of Computer Science and Technology, USTC, and the director of the Key Lab of Computing and Communicating Software of Anhui Province. He has published five books and more than 100 technical articles in journals and proceedings in the areas of computation intelligence, pattern recognition, signal processing and computer networks. He is an editorial board member of the *Chinese Journal of Electronic*, the *Journal of Chinese Computer Systems*, and the *International Journal of Information Acquisition.*