Spontaneous Emotional Facial Expression Detection

Zhihong Zeng
University of Illinois at Urbana-Champaign, Urbana, USA
Email: zhzeng@ifp,uiuc.edu

[1]Yun Fu, [2]Glenn I. Roisman, [3]Zhen Wen, [4]Yuxiao Hu and [5]Thomas S. Huang
[1,2,4,5]University of Illinois at Urbana-Champaign
[3]IBM T.J.Watson Research Center
USA
Email: {zhzeng, yunfu2, hu3, huang}@ifp.uiuc.edu, roisman@uiuc.edu, zhenwen@us.ibm.com

***Abstract*— Change in a speaker's emotion is a fundamental component in human communication. Automatic recognition of spontaneous emotion would significantly impact human-computer interaction and emotion-related studies in education, psychology and psychiatry. In this paper, we explore methods for detecting emotional facial expressions occurring in a realistic human conversation setting—the Adult Attachment Interview (AAI). Because non-emotional facial expressions have no distinct description and are expensive to model, we treat emotional facial expression detection as a one-class classification problem, which is to describe target objects (i.e., emotional facial expressions) and distinguish them from outliers (i.e., non-emotional ones). Our preliminary experiments on AAI data suggest that one-class classification methods can reach a good balance between cost (labeling and computing) and recognition performance by avoiding non-emotional expression labeling and modeling.**

***Index Terms*—affective computing, facial expression, one-class classification, emotion recognition**

I. INTRODUCTION

Human-computer interaction has been a predominantly one-way interaction where the user needs to directly request computer responses. Changes in emotions, which are crucial in human decision-making, perception, interaction, and intelligence, are inaccessible to computing systems. Emerging technological advances are inspiring the research field of "affective computing," which aims at allowing computers to express and recognize affect [21]. The ability to detect and track a user's affective state has the potential to allow a computing system to initiate communication with a user based on the perceived needs of the user within the context of the user's actions. This enables a computing system to offer relevant information when a user needs help – not just when the user requests help. In this way, human computer interaction can become more natural, persuasive, and friendly.

Another potential application of automatic emotion recognition is to help people in emotion-related research to improve the processing of emotion data. For example, some psychological research in the area of adult attachment has explored the association between attachment dimensions and emotional expression during the Adult Attachment Interview (AAI) [23]. In such studies, the analysis of facial expressions is currently completed manually using coding systems such as the Facial Action Coding System (FACS) [8]. FACS was developed by Ekman and Friesen to code facial expressions where a set of action units are used to exhaustively describe facial movements. This coding process is very time consuming. Automatic emotion recognition can improve the efficiency and objectivity of this psychological research.

In this paper, we focus on detecting emotional facial expressions occurring in a natural conversation setting. The reason for this work is that in a realistic conversation scenario, emotional expressions of subjects are very short and subtle. Separating emotional states from non-emotional states can narrow down data of interest for emotion recognition research. After emotion detection, we can further detect negative and positive emotions which can be used as a strategy to improve the quality of interface in HCI, and as a measurement in studies conducted in the field of psychology [23].

We explore one-class classification application in analyzing spontaneous facial expressions that occurred in the Adult Attachment Interview (AAI). The AAI data in our experiment were collected by the authors in [23] to study links between adults' narratives about their childhood experiences and their facial expressive, physiological, and self-reported emotion. In this

This work is the extension of the paper titled "One-class Classification on Spontaneous Facial Expression Analysis", by Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu and T. S. Huang which appeared in the Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition 2006, Southampton, UK, April 2006. © 2006 IEEE.

interview context, participants were, most of the time, at a non-emotional state with a fraction of emotional states. An emotion recognizer must have the ability to separate these emotional facial expressions from non-emotional ones. However, non-emotion facial expressions do not have distinct descriptions in psychological study while emotional facial expressions are defined in terms of facial action units [8]. It is difficult and expensive to model non-emotional facial expressions. Thus, this emotional facial expression detection can be treated as a one-class classification problem, which is to describe target objects (i.e. emotional facial expressions) and distinguish them from outliers (i.e. non-emotional ones). In this paper, we present our preliminary experiments toward this goal.

The rest of the paper is organized as follows. In the following section, some related work toward facial expression recognition is discussed. We describe in Section III the AAI data used in our spontaneous facial expression analysis. Section IV introduces a 3D face tracker used to extract facial textures. Section V briefly describes some approaches for the one-class classification, including kernel whitening and Support Vector Data Description (SVDD). Section VI presents our preliminary experimental results to evaluate one-class classification application in our AAI data. Finally, we have concluding remarks in Section VII.

II. Related Work

In the past years, the literature on automatic facial expression recognition has grown dramatically by applying advanced techniques of image and video processing. Most studies of automatic facial expression recognition focus on six primary facial expressions or a subset of them, namely happiness, sadness, anger, fear, surprise, and disgust. The expression and recognition of these primary facial expressions were found in Ekman's extensive studies [9] to be universal in different cultures.

The studies of computer-assisted recognition of facial expressions started in 1990s. Mase [14] explored the technique of optical flow for facial expressions recognition. Lanitis et al. [15] applied a flexible shape and appearance model to recognize person identities, genders and facial expressions. Black and Yacoob [4] used local parameterized models of image motion to track non-rigid facial motion that was fed to a rule-based classifier of facial expressions. Rosenblum et al. [24] used optical flow and a radial basis function network to classify expressions. Essa and Pentland [10] also used an optical flow region-based approach to identify expressions. Otsuka and Ohya [18] used optical flow and a hidden Markov model (HMM) for facial expression recognition. Tian et al [29] explored action unit recognition by using multi-state facial component models and a neural-network-based classifier. Cohen et al. [5] introduced the structure of Bayesian network classifiers and a multi-level HMM classifier to automatically segment an arbitrary long sequence to the corresponding facial expressions. For extensive survey of facial expression analysis done in the recent years, readers are referred to the overview papers, including [19][20] written by Pantic and Rothkrantz in 2000 and 2003, [33] by Cowie et al. in 2001, and [17] by Sebe et al. in 2005.

Authentic facial expressions are difficult to collect because they are relatively rare and short lived, and filled with subtle context-based changes that make it hard to elicit emotions without influencing the results. Manual labeling of spontaneous facial expressions is very time consuming, error prone, and expensive. This state of affairs leads to a big challenge to spontaneous facial expression analysis. Due to these difficulties in facial expression recognition, most of current automatic facial expression studies were based on the artificial material of deliberately expressed emotions that were collected by asking the subjects to perform a series of facial expressions to a camera. The popular artificial facial expression databases include Ekman-Hager [1] and Kanade-Cohn facial expression data [12]. The latter one is the most comprehensive and the most commonly used database in which subjects were instructed by an experimenter to perform a series of facial expressions.

Recently, three notable exception are the studies by Sebe et al. 2004 [16], Bartlett et al. 2005 [2], and Cohn & Schmidt 2004 [6]. In [16], emotion data were collected in a video kiosk with a hidden camera, which would display segments from recent movie trailers. The subjects were attracted to watch the video and emotions were elicited through different genres of video footage. The authors explored a variety of classifiers to recognize neutral states, happiness, surprise and disgust. However, their database was not directly related to interaction. [2] presents preliminary recognition results of 17 facial action unit in a false opinion paradigm in which subjects are asked to either tell the truth or take the opposite opinion on an issue where they rated strong feelings. Although their recognition system tested the spontaneous facial action units, it was trained on the data of posed facial expression. The psychological study [23] indicates that the posed nature of the emotions may differ in appearance and timing from corresponding performances in natural settings. Especially, the study [6] indicated that posed smiles were of larger amplitude and has less consistent relation between amplitude and duration than spontaneous smiles. While the ability to recognize a large variety of facial action units is attractive, it may not be practical because the emotion data of realistic conversations is often not sufficient to learning a classifier for a variety of action units and their combination.

III. Adult Attachment Interview Data

The Adult Attachment Interview (AAI) is a semi-structured interview used to characterize individuals' current state of mind with respect to past parent-child experiences. This protocol requires participants to describe their early relationships with their parents, revisit salient separation episodes, explore instances of perceived childhood rejection, recall encounters with loss, describe aspects of their current relationship with their parents, and discuss salient changes that may have occurred from childhood to maturity.

During data collection, remotely controlled, high-resolution (720*480) color video cameras recorded the participants' and interviewer's facial behavior during the AAI. Cameras were hidden from participants' view behind a darkened glass on a bookshelf in order not to distract the participants.

Female interviewers were selected to help participants feel at easy during sensor attachment and administration of the AAI. Interviewers underwent extensive training and followed a standardized, semi-structured interview script.

As our first step to explore spontaneous emotion recognition, AAI data from two subjects (one female and one male) was used in this study. The video of the female subject lasted 39 minutes, and one of the male lasted 42 minutes. The significant amount of data allowed us to conduct personal-dependent spontaneous emotion analysis. A snapshot of the video of the male is shown in Figure 1.

Figure 1. A snapshot of a video in our experiment from the AAI database. The participant's face is displayed in the bigger window while the interviewer's face is in the smaller left-top window.

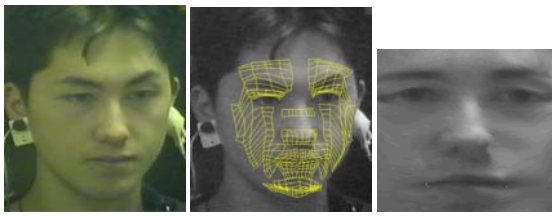
In order to objectively capture the richness and complexity of facial expressions, the Facial Action Coding System (FACS) [8] was used to code every facial event that occurred during the AAI by two certified coders. Inter-rater reliability was estimated according to the method in [8] of calculating the ratio of the number of agreements in emotional expression to the total number of agreement and disagreements, yielding for this study a mean agreement ratio of 0.85.

In this study, we explore the performance of both human perception and automatic recognition of facial expression when audio information is completely ignored. The manual coding was completed without audio in order to keep coders unaware of the content of individuals' narratives. Likewise, the later automatic recognition was done without audio. Different from our previous work in [31], we use facial texture, which captures the richness and complexity of facial expression, instead of 12 facial features which, especially features around the mouth, are sensitive to speech content.

To reduce FACS data further for analysis, we grouped combinations of AUs into two emotion categories (i.e. positive and negative emotions) on the basis of an empirically and theoretically derived Facial Action Coding System Emotion codes that was created in study [7]. In later recognition, the data of these positive and negative states are treated as the target data (emotional state).

IV. 3D FACE TRACKER

To handle the arbitrary behavior of subjects in the natural setting, the 3D face tracked is required. The face tracking in our experiments is based on a system called Piecewise Bezier Volume Deformation (PBVD) tracker that was developed in [26].



(a) (b) (c)

Figure 2. The 3D face tracker's result. (a) the video frame input; (b) tracking result where a mesh is used to visualize the geometric motions of the face; (c) extracted face texture.

This face tracker uses a 3D facial mesh model that is embedded in multiple Bezier volumes. The shape of the mesh can be changed with the movement of the control points in the Bezier volumes. This approach guarantees that the surface patches will be continuous and smooth. In the first video frame, the 3-D facial mesh model is constructed by selection of landmark facial feature points. Once the model was fitted, the tracker can track head motion and local deformations of the facial features by an optical flow method. In this study, we use rigid setting of this tracker to extract facial expression texture. The 3D ridge geometric parameters (3D rotation and 3D translation) determine the registration of each image frame to the face texture map that is obtained by wrapping the 3D face appearance. Thus, we can derive a sequence of face texture images, which capture the richness and complexity of facial expression. Figure 2 shows a snapshot of the tracking system. Figure 2(a) is the input video frame, and Figure 2(b) is the tracking result where a mesh is used to visualize the geometric motions of the face. The extracted face texture is shown in Figure 2(c).

V. ONE-CLASS CLASSIFICATION

In most pattern recognition tasks, it is assumed that a training dataset is available that represents well what can be expected in practice. Unfortunately, in our AAI data, only the emotional data that occupies a small fraction of the interview has been chosen and labeled. The rest of AAI data is regarded as non-emotional state. Thus, the non-emotional state is very diverse and expensive to model.

In order to detect this ill-represented state in the test data, we applied one-class classification methods that describe a class of target data (i.e., emotional data) and distinguish it from all other possible outlier objects (i.e., non-emotional data).

For one-class classification, several approaches have been introduced, including novelty detection [3], outlier detection [22], and concept learning [11]. One solution for outlier or novelty detection is to estimate a probability density of the target class, and judge an object as outlier or novelty if the object falls into a region where the density is lower than some threshold. This solution requires a sufficient sample size of target objects to estimate the target distribution. The other solution of one-class classification is to define the target boundary rather than estimating its density. This latter approach has better performance when the sample size of target objects is limited. In our classification application, we use both Gaussian density data description and Support Vector Data Description (SVDD) [28] that directly fits a boundary with minimal volume around the target data without density estimation. As Gaussian density data description is very popular, we only briefly introduce the technique of support vector data description in the following section.

Considering that the recognition performance of many one-class classification approaches is influenced critically by the scaling of the data and is often harmed by non-homogeneous data distributions in input space or transformed subspace. Thus, before one-class classification, we apply kernel whitening [27], which is a preprocessing to map the data to a spherical symmetrical cluster and is almost insensitive to data distributed in subspace.

A. Kernel Whitening

The study [25] indicated that nonlinear principal components from kernel PCA can lead to better recognition rates than corresponding number of linear principal component from linear PCA, and kernel PCA has more space to improve recognition performance by using more components than linear PCA. The kernel whitening [27] applies the idea of kernel PCA to map the data into a space with equal variance in each feature direction with eigenvalue larger than zero. The algorithm of kernel whitening is briefly as follows (The details can be found in [27]):

Given a set of data $x_k \in R^N$, $k = 1, \cdots, M$. Assume that by some mapping $\Phi : x_k \to y_k \in F^l$, they are mapped to a kernel space F^l in which the transformed data is centered, i.e. $\sum_{k=1}^{M} \Phi(x_k) = 0$.

1. compute the kernel matrix $K_{ij} = (k(x_i, x_j))_{ij}$. Several kernel k can be chosen. In our application, Gaussian kernel is used.
2. solve the eigenvalue problem:
$$\lambda\alpha = K\alpha \tag{1}$$
by diagonalizing K, and normalize the eigenvector expansion coefficients α^n by requiring $\lambda_n^2(\alpha^n \cdot \alpha^n) = 1$ which is to rescale the data in the kernel space to unit variance.

A new object z can be mapped onto eigenvector v^n by
$$(\hat{z})_n = \sum_{i=1}^{M} \alpha_i^n k(x_i \cdot z) \tag{2}$$
where $(\hat{z})_n$ means the n-th component of vector $\hat{z}$.

The assumption that the transformed data in the kernel space are centered can be done by slightly modifying the original kernel matrix. The dataset, transformed by using the mapping (2), can now be used by any one-class classifiers.

B. Support Vector Data Description (SVDD)

Support Vector Data Description (SVDD) [28] was used to directly fit a boundary with minimal volume around the target data – which are emotion expressions in our application. SVDD is based on the assumption that the outliers distribute evenly in the feature space. This assumption guarantees that the error minimization is corresponding to the minimal volume of the date description.

During SVDD training, training data of target class is used to define a hypersphere. During testing, all objects inside the hypersphere will be classified as target objects, and objects outside the hypersphere are labeled as outliers. The hypersphere is described by center a and radius R. Considering the possibility of outliers in the train set, the distance from data to the center a need not be strictly smaller than R^2, but larger distances should be penalized. An extra parameter ν is introduced to control the trade-off between the volume of the hypersphere and errors. An error function L of the volume of the hypersphere and the distances is minimized. Given a set of data $y_k \in F^l$ which has been transformed by kernel whitening, this yields the following function to maximize with respect to β (for details in [28]):
$$L = \sum_{k=1}^{M} \beta_k (y_k \cdot y_k) - \sum_{k,j=1}^{M} \beta_k \beta_j (y_k \cdot y_j) \tag{3}$$
with the constraint
$$0 \le \beta_k \le \nu, \sum_{k=1}^{M} \beta_k = 1 \tag{4}$$
and the center of the hypersphere is
$$a = \sum_{k=1}^{M} \beta_k y_k \tag{5}$$
Now (3) is in a standard quadratic optimization problem. After optimization, the parameters $\beta_k (k = 1, \cdots, M)$ can be zero or larger than zero. Only a few objects are called the support vectors which determine the center a and radius R of the hypersphere. R is obtained by computing the distance from the center a to the support vectors which are on the boundary.

A new object w is classified as target object if
$$\begin{aligned} f(z) &= \|w - a\|^2 \\ &= (w \cdot w) - 2\sum_{k=1}^{M} \beta_k (w \cdot y_k) + \sum_{k,j=1}^{M} \beta_k \beta_j (y_k \cdot y_j) \le R^2 \end{aligned} \tag{6}$$

VI. EXPERIMENTS

In this section, we present the experimental results of our emotion detection by using facial expression information. The AAI videos contain only a fraction of emotion expressions, and a large number of AAI facial expressions belong to non-emotional expressions. Thus, this required data preparation to narrow down data of interest for emotion recognition research.

A. Data Preparation

First, we ignored the emotion occurrences to which our two manual coders disagreed with each other. Next, we filtered out the emotion segments in which a hand occluded the face, the face turned away more than 40 degrees with respect to the optical center, or part of face moved out of camera view. This process narrowed down the inventory to potential useful emotion data for our experiment. Each emotion sequence starts from and to the emotion intensity scoring scale B (slight) or C (marked pronounced). We notice that negative emotions are more subtle and shorter than positive emotions. The possible reason is that people tend to inhibit negative emotion in this situation.

TABLE I. THE FRAME NUMBER OF FACIAL TEXTURE IMAGES

Subject	Emotional		Non-emotional	Total
	Positive	Negative		
Female	3470	1085	3302	7857
Male	1372	851	3007	5230

Because the subjects in the AAI videos most of time displayed non-emotional facial expressions, data of non-emotional state is huge and badly balanced with emotional data. Thus, we randomly choose 2-3 segments of non-emotional expression in each of 16 question-answer sections in the interview as outlier objects. We obtained 45 sequences of non-emotional states. The total number of sequences is 150 (69 positive states, 36 negative states, and 45 non-emotional states) for the female participant, and 130 for the male participant (30 positive states, 56 negative states, and 45 non-emotional states).

Next, using our 3D face tracker, we obtained 7857 frames of facial texture images for the female participant, and 5230 frames for the male participant. The detail of the numbers of frames of facial texture images are listed in Table 1.

The face tracker output 25 frames of facial texture images per video second. The facial texture images are normalized to the images with resolution of 64*64 pixel^2 and 256 grey levels per pixel.

B. Recognition

We investigated the performance of person-dependent emotion recognition through facial expression analysis. In this test, training and testing data are from the same subject. For each subject, the ten-fold cross-validation is used to validate the classification performance. The sequence data of the emotional states was divided into ten disjoint sets of almost equal size. Different from the ordinary two-class classifiers, one-class classifiers are defined on the target data without the use of outlier data.

In order to evaluate the performance of one-class classifiers in our application, we chose Kernel whitening+SVDD, conventional SVDD which applies linear PCA mapping to preprocess data, and Gaussian data description which uses linear PCA for preprocessing and a single Gaussian density to model target data. These one-class classifiers were trained and tested ten times, each time with a different emotional data set and all of non-emotional data held out as a validation set. The remaining emotional sets were used as training sets.

The experimental results are illustrated in Figure 3 and Figure 4. Figure 3 demonstrates the recognition performance of these one-class classifiers on the female data in which their ROC curves of False Emotion Fraction (FE) and True Emotion Fraction (TE) are shown. True emotion fraction is the fraction of emotional expressions (target examples) that are accepted by a one-class classifier as emotional objects, and false emotion fraction is the fraction of non-emotion expressions (outliers) that are accepted as the emotion class. The x-axis and y-axis are false emotion fraction and true emotion fraction, respectively. Figure 3 shows that with same false emotion fraction (FE), KSVDD has the largest true emotion fraction (TE), SVDD has worse TE than KSVDD but better TE than Gaussian data description, and Gaussian data description has the worst TE. Figure 4 shows the recognition performance of these one-class classifiers on the male data, also in terms of ROC curves of False Emotion Fraction (FE) and True Emotion Fraction (TE). Figure 4 demonstrates that KSVDD has the best performance, SVDD has worst performance, and Gaussian data description performs worse than KSVDD but better than SVDD.

To further clarify the performance comparison of these three one-class classifiers, we calculated the Area under FE~TE ROC curves (AUC) of Figure 3 and Figure 4, which is shown in Figure 5. The AUCs of Gaussian, SVDD, KSVDD are 0.7607, 0.8835, 0.9377, respectively, for the female subject, and are 0.7632, 0.6713, 0.8528, respectively, for the male subject. In Figure 5, the AUC of KSVDD is the largest, which means that kernel whitening + SVDD outperformed conventional SVDD and Gaussian data description. The experimental results show that the performance of SVDD is influenced by data distributions in feature space, and kernel whitening is able to improve the performance of SVDD by rescaling the data to a kernel feature space with unit variance.

The performance of facial expression recognition on the female data was better than on the male data. The possible reason is that female emotional data is larger than male emotion data, which may have resulted in more accurate estimation of the boundary of target data.

C. Comparison between One-class and Two-class Classifiers

We notice that SVDD obtains a closed boundary around the target data by minimizing volume of target data description. That is based on the assumption that the outliers distribute evenly in the feature space. This assumption guarantees that the error minimization is corresponding to the minimal volume of the data description. However, the true outlier distribution could deviate from uniform distribution in our realistic application. In addition, we can obtain some outlier samples (non-emotional expressions) for training,

although these samples are not enough to represent the outlier distribution. Intuitively, these outlier samples could be helpful for our recognition problem although they bring some bias in building the classifier at the same time. Thus we are curious at the performance of tradition two-class classifiers in our realistic application that uses both target and outlier data in training.

We conducted experiments to compare the performance of one-class classifiers and two-class classifiers. We apply two two-class classifiers, Gaussian-based classifier and tradition Support Vector Classifier (SVC) [30]. The two-class Gaussian-based classifier models both the target data and outlier data by two individual Gaussian densities, and the decision boundary between two classes is where the log of the probability ratio is zero. SVC is to define the decision boundary by using support vectors from both target and outlier classes.

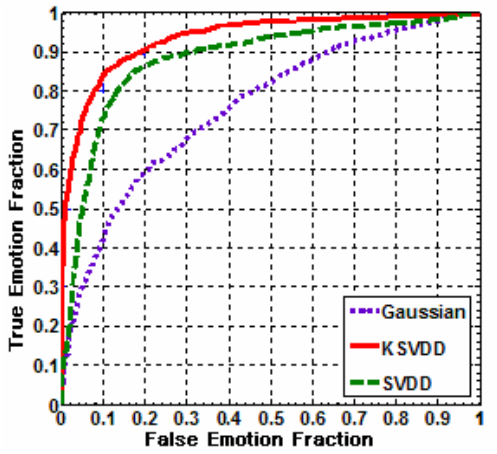


Figure 3. FE~TE ROC curves of female subject data for single Gaussian, SVDD and Kernel Whitening+SVDD.

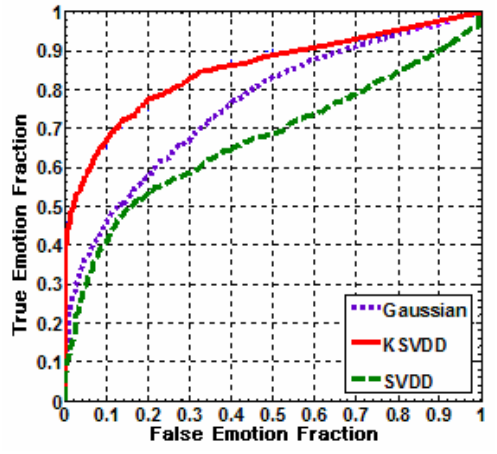


Figure 4. FE~TE ROC curves of male subject data for single Gaussian, SVDD and Kernel Whitening+ SVDD.

In the two-class classification experiments, the sequence data of the non-emotional states was also divided into ten disjoint sets of almost equal size. These two-class classifiers were trained and tested ten times, each time with a different emotional set and a different non-emotional set held out as a validation set, and the remaining emotional and non-emotional sets used as training sets.

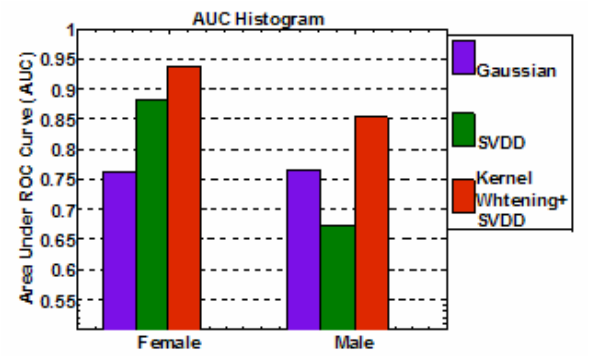


Figure 5. Area under FE~TE ROC curves (AUC) of Figure 3 and Figure 4.

The experimental results of performance comparison between one-class and two-class classification are shown in Table II. We compare results in terms of best recognition accuracy, which is the fraction of the correctly estimated samples of both emotional and non-emotional samples with optimal threshold.

Table II shows that the KSVDD has the best recognition accuracies of both the female and male data, compared with traditional two-class SVC and two-class Gaussian-based classifier. SVDD performs better than SVC on female data but worse than SVC on male data. Two-class Gaussian classifier outperforms one-class Gaussian data description on both the female and male data. This investigation suggests that although the non-emotional samples in our experiments are not enough to represent outlier distributions, they are helpful to separate emotional expressions from non-emotional ones.

TABLE II. PERFORMANCE COMPARISON BETWEEN ONE-CLASS AND TWO-CLASS CLASSIFIERS ON THE FEMALE AND MALE DATA. THEY ARE EVALUATED IN TERMS OF BEST RECOGNITION ACCURACY WHICH IS THE FRACTION OF THE CORRECTLY ESTIMATED SAMPLES OF BOTH EMOTIONAL AND NON-EMOTIONAL CLASSES WITH OPTIMAL THRESHOLD.

Methods		Female	Male
One-class	Gaussian	0.70	0.69
	SVDD	0.83	0.67
	KSVDD	0.87	0.79
Two-class	Gaussian	0.79	0.70
	SVC	0.81	0.72

The study in [28] argues that when both representative samples from the target class and outlier class are available, conventional two-class classification classifiers outperform the one-class classification classifiers. In our application, it is very expensive to collect sufficient samples of non-emotional facial expressions to represent outlier objects well. Thus, in our spontaneous facial expression analysis, one-class classification can reach a

good balance between the cost (labeling and computation) and recognition performance by avoiding non-emotional facial expression labeling and modeling.

VII. CONCLUSION

Advances in computer processing power and emerging algorithms are providing new ways of envisioning Human Computer Interaction. With an automatic emotion recognizer, a computer can respond appropriately to the user's affective state rather than simply responding to user commands. In this way, the nature of the computer interactions would become more authentic, persuasive, and meaningful. In addition, automatic emotion recognition can greatly improve the efficiency and objectivity of emotion-related studies in education, psychology and psychiatry.

In this paper, we explored detecting emotional facial expressions occurred in a realistic human conversation setting—the Adult Attachment Interview (AAI). Because non-emotional facial expressions do not have distinct description in psychological studies and are expensive to represent, we treat this emotional facial expression detection as a one-class classification problem which is to describe target data (i.e. emotional facial expressions) and distinguish them from outliers (i.e. non-emotional ones). In our work of this paper, we apply Gaussian density data description and support vector data description, and compare their performance with corresponding two-class classifiers (Bayesian Gaussian classifier and support vector classifier). Our preliminary experiments on the AAI data suggest that one-class classification methods can reach a good balance between cost and recognition performance.

Automatic analysis of facial expressions occurred in natural communication setting is a largely unexplored and challenging problem. Based on our current progress in this direction, the next stage of our work will be attempting to evaluate one-class classification application on more videos in our AAI database. After detecting emotional expressions in this conversation setting, we can further detect negative and positive emotions, which can be used as a strategy to improve the quality of interface in HCI, and as a means of efficient measurement in psychological research.

The psychological study [34] indicated that judging someone's affective states, people mainly rely on facial expressions and vocal intonations. In the communication setting, besides facial expression, audio also provides valuable information to emotion judgment. In this paper, we set aside audio information when both manual coding and automatic analysis of emotion were conducted. Thus, it is also of interest to investigate human perception and automatic analysis by using facial expression with audio information.

ACKNOWLEDGMENT

We would like to thank Ms Debra Huang for discussion of her honor thesis. This work is supported by Beckman Postdoctoral Fellowship and National Science Foundation: Information Technology Research Grant # 0085980.

REFERENCES

[1] Ekman, P., Hager, J.C., Methvin, C.H. and Irwin, W., Ekman-Hager Facial Action Exemplars, unpublished, San Francisco: Human Interaction Laboratory, University of California

[2] Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J.(2005), Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, IEEE CVPR'05

[3] Bishop, C. (1994), Novelty detection and neutral network validation, IEE Proceedings on Vision Image and Signal Processing. Special Issue on Applications of Neutral Network 141, 217-222

[4] Black, M. and Yacoob, Y.(1995), Tracking and Recognizing Rigid and Non-rigid Facial Motions Using Local Parametric Models of Image Motion, in Proc. Int. Conf. on Computer Vision, 374-381

[5] Cohen, L., Sebe, N., Garg, A., Chen, L., and Huang, T. (2003), Facial expression recognition from video sequences: Temporal and static modeling, Computer Vision and Image Understanding, 91(1-2):160-187

[6] Cohn, J.F. and Schmidt, K.L.(2004), The timing of Facial Motion in Posed and Spontaneous Smiles, International Journal of Wavelets, Multiresolution and Information Processing, 2, 1-12

[7] Huang, D. (1999), Physiological, subjective, and behavioral Responses of Chinese American and European Americans during moments of peak emotional intensity, honor Bachelor thesis, Psychology, University of Minnesota,

[8] Ekman, P. and Friesen, W.V. (1978), Facial Action Unit System: Investigator's Guide, Consulting Psychologists Press

[9] Ekman, P. (1994), Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique, Psychological Bulletin, 115(2): 268-287

[10] Essa, I. and Pentland, A. (1997), Coding, Analysis, Interpretation, and Recognition of Facial Expressions, IEEE Trans. On Pattern Analysis and Machine Intelligence, 19(7): 757-767

[11] Japkowicz, N., Myers, C., and Gluck, M. (1995), A novelty detection approach to classification. Int. Joint Conference on Artificial Intelligence, 518-523

[12] Kanade, T., Cohn, J., and Tian, Y. (2000), Comprehensive Database for Facial Expression Analysis, In Proceeding of International Conference on Face and Gesture Recognition, 46-53

[13] Moghaddam, B.; Jebara, T. and Pentland, A. (2000), Bayesian Face Recognition, *Pattern Recognition*, Vol 33, Issue 11, pp: 1771-1782, November 2000

[14] Mase, K.(1991), Recognition of Facial Expression from Optical Flow, IEICE Trans. E74(10): 3474-3483

[15] Lanitis, A., Taylor, C. and Cootes, T. (1995), A Unified Approach to Coding and Interpreting Face Images, in Proc. International Conf. on Computer Vision, 368-373

[16] Sebe, N., Lew, M.S., Cohen, I., Sun, Y., Gevers, T., and Huang, T.S.(2004), Authentic Facial Expression Analysis, Int. Conf. on Automatic Face and Gesture Recognition

[17] Sebe, N., Cohen, I., Gevers, T., and Huang, T.S. (2005), Multimodal Approaches for Emotion Recognition: A Survey, In Proc. Of SPIE-IS&T Electronic Imaging, SPIE Vol 5670: 56-67

[18] Otsuka, T. and Ohya, J. (1997), Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences, Proc. Int. Conf. on Image Processing, 546-549
[19] Pantic, M. and Rothkrantz, L. (2000), Automatic Analysis of Facial Expressions: the State of the Art, IEEE Trans. On Pattern Analysis and Machine Intelligence, 22(12): 1424-1445
[20] Pantic M., and Rothkrantz, L.J.M.(2003), Toward an affect-sensitive multimodal human-computer interaction, Proceedings of the IEEE, Vol. 91, No. 9, Sept. 2003, 1370-1390
[21] Picard, R.W. (1997), Affective Computing, MIT Press, Cambridge.
[22] Ritter, G., and Gallegos, M. (1997), Outliers in statistical pattern recognition and application to automatic chromosome classification. Pattern Recognition Letters 18, 525-539
[23] Roisman, G.I., Tsai, J.L., and Chiang, K.S.(2004), 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, 776-789
[24] Rosenblum, M., Yacoob, Y., and Davis, L. (1996), Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture, IEEE Trans. On Neural Network, 7(5):1121-1138
[25] Scholkopf, B., Smola, A., and Muller, K. (1998), Nonlinear component analysis as kernel eigenvalue problem, Neural Computing 10, 1299-1319
[26] Tao, H. and Huang, T.S. (1999), Explanation-based facial motion tracking using a piecewise Bezier volume deformation mode, IEEE CVPR'99, vol.1, pp. 611-617
[27] Tax, D., and Juszczak, P. (2003), Kernel whitening for one-class classification, International Journal of Pattern Recognition and Artificial Intelligence, vol. 17, no. 3, 333-347
[28] Tax, D. (2001), One-class classification, Ph.D. thesis Delft University of Technology
[29] Tian, Y., Kanade, T. and Cohn, J. (2001), Recognizing Action Units for Facial Expression Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2): 97-116
[30] Vapnik, V. (1998), Statistical Learning Theory, Wiley
[31] Zeng, Z., Tu, J., Liu, M., Zhang, T., Rizzolo, N., Zhang, Z., Huang, T.S., Roth, D., and Levinson, S. (2004), Bimodal HCI-related Affect Recognition, Int. Conf. on Multimodal Interfaces, 137-143
[32] Ekman, P. and Rosenberg, E. (Eds.), What the face reveals. NY: Oxford University, 1997
[33] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J.G., Emotion Recognition in Human-Computer Interaction, IEEE Signal Processing Magazine, January 2001, 32-80
[34] Mehrabian, A., Communication without words, Psychol. Today, vol.2, no.4, 53-56, 1968

Zhihong Zeng: MS degree from Tsinghua University in 1989, and Ph.D. from the Institute of Automation, Chinese Academy of Sciences in 2002. He has been working with the Image Formation and Processing group at the Beckman Institute since 2002. He is a current Beckman Postdoctoral Fellow (2005-2008). His research interests in multimodal emotion assessment for human-computer interaction includes the psychological analysis of human emotion perception, computer vision, speech processing, and machine learning.

Yun Fu B.E. and M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2001 and 2004, respectively. He is PhD candidate at Department of Electrical and Computer Engineering at University of Illinois at Urbana-Champaign. His research interests include computer vision, machine learning, computer graphics, image processing and intelligence system. Mr. Fu was the recipient of a 2002 Rockwell Automation Master of Science Award, two Edison Cups of 2002 GE Fund "Edison Cup" Technology Innovation Competition, and the 2003 HP Silver Medal and Science Scholarship for Excellent Chinese Student..

Glenn I. Roisman: Ph.D from the University of Minnesota in 2002. He is a currently an Assistant Professor of Psychology at University of Illinois at Urbana-Champaign. His interests concern the legacy of early relationship experiences as an organizing force in social and emotional development across the lifespan.

Zhen Wen: BS degree in computer science from Tsinghua University in 1998, MS and PhD degree in computer science from University of Illinois at Urbana-Champaign in 2000 and 2004. He is currently a research staff member at IBM T.J. Watson Research Center. His general research interests include machine learning, intelligent user interface, computer graphics and multimedia systems with a current focus on context-sensitive information retrieval, information visualization.

Yuxiao Hu: B.S. and M.S. degree in computer science from Tsinghua University in 1999 and 2001. He is a PhD candidate in Image Formation and Processing Group (IFP) at University of Illinois at Urbana-Champaign. His current research interests include machine learning and face related projects such as detection, tracking, pose estimation and recognition

Thomas S. Huang Sc.D. from MIT. Currently William L. Everitt Distinguished Professor of Electrical and Computer Engineering at University of Illinois at Urbana-Champaign. His professional interests are computer vision, image compression and enhancement, pattern recognition, and multimodal signal processing. His honors includes: Member, National Academy of Engineering, USA; Member, National Academy of Engineering Foreign Member, Chinese Academy of Engineering Foreign Member, Chinese Academy of Sciences; IEEE Jack S. Kilby Signal Processing Medal (2000) (co-recipient with A. Netravali); Int. Asso. of Pattern Recognition, King-Sun Fu Prize (2002); Honda Lifetime Achievement Award (2000); IEEE Third Millennium Medal (2000); Honda Lifetime Achievement Award (2000)