Emotion Detection via Discriminative Kernel Method

Hua Wang
Department of Computer Science and Engineering
University of Texas at Arlington, TX, USA
huawang2007@mavs.uta.edu

Heng Huang
Department of Computer Science and Engineering
University of Texas at Arlington, TX, USA
heng@uta.edu

Yanzi Hu
School of Humanities and Arts
Xidian University
Xi'an, Shaanxi, China
xd.yanhu@gmail.com

Mindi Anderson
College of Nursing
University of Texas at Arlington, TX, USA
manderso@uta.edu

Pamela Rollins
Callier Center for Communication Disorders
University of Texas at Dallas, TX, USA
rollins@utdallas.edu

Fillia Makedon
Department of Computer Science and Engineering
University of Texas at Arlington, TX, USA
makedon@uta.edu

ABSTRACT

Human emotion detection is of substantial importance in diverse pervasive applications in assistive environments. Because facial expressions provide a key mechanism for understanding and conveying emotion, automatic emotion detection through facial expression recognition has attracted increased attention in both scientific research and practical applications in recent years. Traditional facial expression recognition methods normally use only one type of facial expression data, either static data extracted from one single face image or motion dependent data obtained from dynamic face image sequences, but seldom employ both. In this work, we propose a novel Discriminative Kernel Facial Emotion Recognition (DKFER) method to integrate these two types of facial expression data using a hybrid kernel, such that the advantages of both of them are exploited. In addition, by using Linear Discriminant Analysis (LDA) to transform the two types of original facial expression data into two more discriminative lower-dimensional subspaces, the succeeding classification for emotion detection can be carried out in a more efficient and effective way. Encouraging experimental results in empirical studies demonstrate the practical usage of the proposed DKFER method for emotion detection.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition Applications**]: Computer Vision, Signal Processing; H.1.2 [**Information System**]: Models and Principals—*User/Machine System*; I.2 [**Artificial Intelligence**]: Learning

General Terms

Design, Algorithms, Experimentation

PETRA'10, June 23–25, 2010, Samos, Greece.

Keywords

Emotion Detection, Facial Expression Recognition, Discriminative Learning, Kernel, Facial Feature

1. INTRODUCTION

Human emotion detection plays a significant role in pervasive applications in assistive environments, such as those intended for handicapped persons or persons with cognitive disabilities who are not able to recognize emotions [1, 8, 24], or for developing intelligent graphic user interface (GUI) environments [19]. Among a number of body languages conveying human emotion, facial expression is generally considered the most straightforward and effective, which thereby has attracted a lot of attention in scientific search and stimulated a slew of facial expression based automatic human emotion detection methods in a large range of practical applications [10, 23, 15, 13, 12], from psychology [18, 16] to the applied science such as pervasive health care.

Existing facial expression recognition methods are usually categorized into two classes: *static* methods or *motion dependent* methods. In the former, recognition of a facial expression is performed using a single image of a face, which examines geometrical relations established among facial features or facial organs, such as eyes, eyebrows and mouth, and the variation of location of features [20, 21, 17, 10]. A general limitation of static methods is the dependency on the precise alignment of faces, which is difficult to achieve for real world data due to head rotation of the subject and different face geometries. In the latter, temporal information is extracted using at least two instances of a face, representing the face in its neutral condition and the face at the peak of one expression [22, 6, 12]. Motion dependent methods enjoy the benefit that the facial feature points can be easily normalized to a reference coordinate system, where the deformation vectors can be calculated directly. However, the performance of this kind of methods often suffers from the reliability problem of optical flow estimation in image sequences.

In this work, we propose a novel Discriminative Kernel Facial Emotion Recognition (DKFER) method for emotion de-

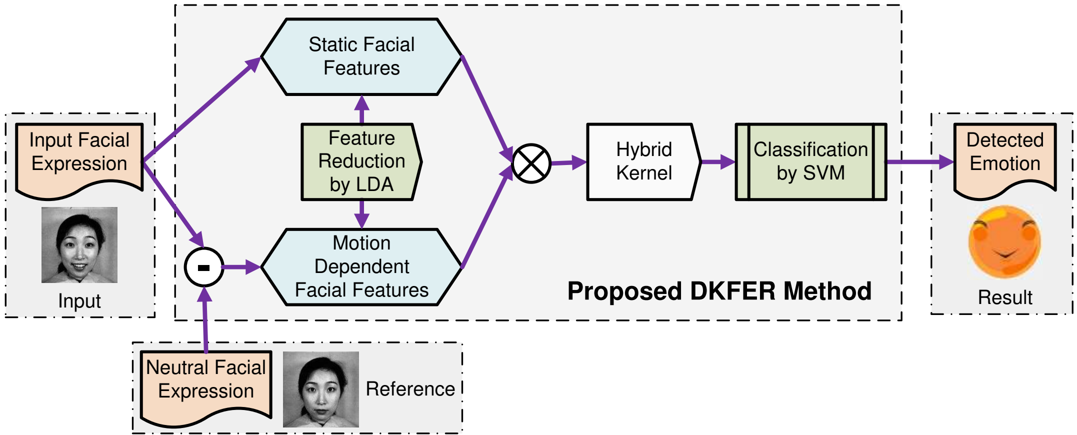
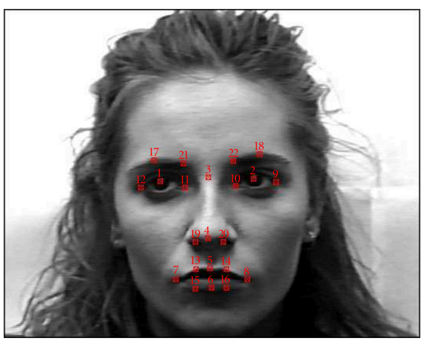


Figure 1: Diagram of the proposed Discriminative Kernel Facial Emotion Recognition (DKFER) method.

(a) Neutral. (b) Anger. (c) Disgust. (d) Fear. (e) Joy. (f) Sorrow. (g) Surprise.

Figure 2: Sample facial expressions for neutral condition and six emotions from JAFFE database.



1. Right pupil
2. Left pupil
3. Nose root
4. Nose tip
5. Upper lip center
6. Lower lip center
7. Right mouth corner
8. Left mouth corner
9. Left outer eye corner
10. Left inner eye corner
11. Right inner eye corner
12. Right outer eye corner
13. Right upper lip
14. Left upper lip
15. Right lower lip
16. Left lower lip
17. Right eyebrow center
18. Left eyebrow center
19. Right nostril
20. Left nostril
21. Right inner eyebrow
22. Left inner eyebrow

Figure 3: Typical facial feature points (landmarks).

tection, which makes use of both static and motion dependent facial features. The two types of facial features are first transformed into two ***discriminative*** subspaces with close dimensionalities by Linear Discriminant Analysis (LDA) [7] respectively. In the resulted subspaces, data points representing different emotions are separated apart far away from one another, while those representing a same emotion are mapped close together in a more compact manner. Then the two transformed feature vectors are integrated via a hybrid ***kernel*** [2], such that the performance of subsequent classification carried out on it is doubly enhanced due to taking advantage of emotion information contained in the both types of facial expression data. In this work, Support Vector Machine (SVM) [2], a state-of-the-art classification technique, is used for the final emotion classification.

The work flow of proposed DKFER method for emotion detection is illustrated in Figure 1. A neutral facial expression is first given (left bottom picture) as a reference. When an input facial expression (left picture) comes in, our task is to perform the proposed DKFER method on it to decide the conveyed human emotion (right cartoon) among a set of possible emotions. In this work, we consider six human emotions: anger, disgust, fear, joy, sorrow and surprise. Sample facial expressions for these six emotions and that for neutral condition from the Japanese Female Facial Expression (JAFFE) Database[1] are listed in Figure 2.

2. FACIAL EXPRESSION FEATURES

In facial expression recognition, a face is typically characterized by a set of facial feature points, also called as "landmarks", such as eyes, eye corners, eyebrows, mouth corners, nose tip, *etc.*, as depicted in Figure 3. We will use these facial feature points to construct facial expression features in the sequel, including both static and motion dependent expression features.

2.1 Static Facial Expression Features

Static facial expression features are extracted from a single image of a face to assess geometrical relations established among facial feature points. In general, facial deformations caused by an emotion can be measured in terms of either angles or distances between certain facial feature points. In order to support a size invariant representation of facial data, angle metrics are preferable because they save the effort for face normalization that is necessary for distance based features [5]. Furthermore, typical angles show a large coincidence between different persons whereas typical distances

[1] `http://kasrl.org/jaffe.html`

vary considerably between different persons [5]. Six angle features used in this work are shown in Figure 4.

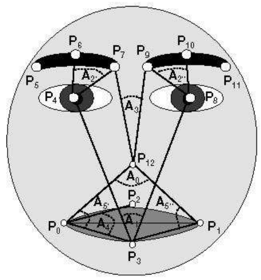


Figure 4: Static facial expression features: angels.

The angles are separated into two groups: A_2, A_3 belong to the upper part of the face and A_0, A_4 and A_5 belong to the lower part of the face, A_1 is not clearly mapped to either group. According to [4, 5], the lower-part angles A_0, A_4 and A_5 are involved for expressing joy (corners of the mouth are raised — A_0), sorrow (corners of the mouth are lowered — A_0) or fear (mouth is opened widely — A_4 and A_5). The upper-part angles A_2, A_3 are deformed when expressing anger (A_2 is larger, A_3 is smaller) or fear (A_2 is smaller). The angle A_1 is considerably decreased when the mouth is open, which may indicate fear or joy. These angles build a six dimensional feature vector as follows:

$$\mathbf{x}^{\mathrm{S}} = [A_0, \ldots, A_5]^T \in \mathbb{R}^5. \tag{1}$$

2.2 Motion Dependent Facial Expression Features

Equipped with the facial feature points as in Figure 3, we define motion dependent facial expression features as the displacements (Euclidean distance) of these facial feature points between a neutral facial expression and the "peak" of a particular emotive expression [14], as illustrated in Figure 5. As a result, every input facial expression is quantified as a motion dependent facial expression feature vector as:

$$\mathbf{x}^{\mathrm{M}} = [d_1, d_2, \ldots, d_p]^T \in \mathbb{R}^p, \tag{2}$$

where p denotes the total number of facial feature points, and $d_i (1 \leq i \leq p)$ denotes the displacement of the i-th facial feature point.

3. DISCRIMINATIVE FACIAL EXPRESSION FEATURES VIA LDA

When static facial expression features $\mathbf{x}^{\mathrm{S}}$ or motion dependent features $\mathbf{x}^{\mathrm{M}}$ is obtained as in Section 2, one can conduct emotion detection on them using statistical classification techniques. Typical characteristic emotion patterns by motion dependent facial expression features are shown in Figure 6, by which classification algorithms can be devised.

Figure 5: Motion dependent facial expression features: facial feature points displacements.

Directly using these original facial expression features, however, suffers from two critical problems: inefficiency and ineffectiveness. First, computational complexity is one of the major consideration in designing a practical emotion detection system. Therefore, we would expect a low-dimensional feature space, such that classification carried out on it can be more efficient. Second, but more important, similar to many other classification problems, class memberships only correlate to some patterns with much lower dimensionality hidden in the original data, and many of the features are irrelevant and sometimes even harmful. Consequently, dimensionality reduction is expected to prune the irrelevant patterns, such that classification on the intrinsic feature subspace can be more effectively.

In addition, because this work aims at utilizing both static and motion dependent facial expression features, we would expect the two feature vectors, $\mathbf{x}^{\mathrm{S}}$ and $\mathbf{x}^{\mathrm{M}}$, have similar numbers of dimensions, such that these two heterogeneous data can be integrated in a balanced manner. In this work, we use LDA [7] to perform dimensionality reduction.

Given a data set with n training samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$ and K classes, where $\mathbf{x}_i \in \mathbb{R}^d$ are feature vectors and $\mathbf{y}_i \in \{0,1\}^K$. $\mathbf{y}_i(k) = 1$ if $\mathbf{x}_i$ belongs to the k-th class, and 0 otherwise. In the context of emotion detection, the feature vectors could be the static or motion dependent facial expression features as in Eq. (1) or Eq. (2); the classes are the six emotions of interest in this work. Let input data be partitioned into K groups as $\{\pi_k\}_{k=1}^K$, where π_k denotes the sample set of the k-th class with n_k data points. LDA seeks a linear transformation $G = \mathbb{R}^{d \times r}$ that maps $\mathbf{x}_i$ in the high d-dimensional space to a vector $\mathbf{x}_i \in \mathbb{R}^r$ in a lower $r(< d)$-dimensional space by $\mathbf{q}_i = G^T \mathbf{x}_i$. In LDA, the *between-class*, *within-class*, and *total* scatter matrices are defined as follows [7]:

$$S_b = \sum_{k=1}^{K} n_k \left(\mathbf{m}_k - \mathbf{m}\right) \left(\mathbf{m}_k - \mathbf{m}\right)^T , \tag{3}$$

$$S_w = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \pi_k} \left(\mathbf{x}_i - \mathbf{m}_k\right) \left(\mathbf{x}_i - \mathbf{m}_k\right)^T , \tag{4}$$

$$S_t = \sum_{i=1}^{n} \left(\mathbf{x}_i - \mathbf{m}\right) \left(\mathbf{x}_i - \mathbf{m}\right)^T , \tag{5}$$

where $\mathbf{m}_k = \frac{1}{n_k} \sum_{\mathbf{x}_i \in \pi_k} \mathbf{x}_i$ is the class mean (class centroid) of the k-th class, $\mathbf{m} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ is the global mean (global centroid), and $S_t = S_b + S_w$. The optimal G is chosen such that the between-class distance is maximize whilst the within-class distance is minimized in the low-dimensional projected space, which leads to the standard LDA optimization objective [7] as follows:

$$\arg\max_{G} \mathbf{tr} \left(\frac{G^T S_b G}{G^T S_w G} \right). \tag{6}$$

The solution of the above optimization problem in Eq. (6) is well established in mathematics as following. Solving the

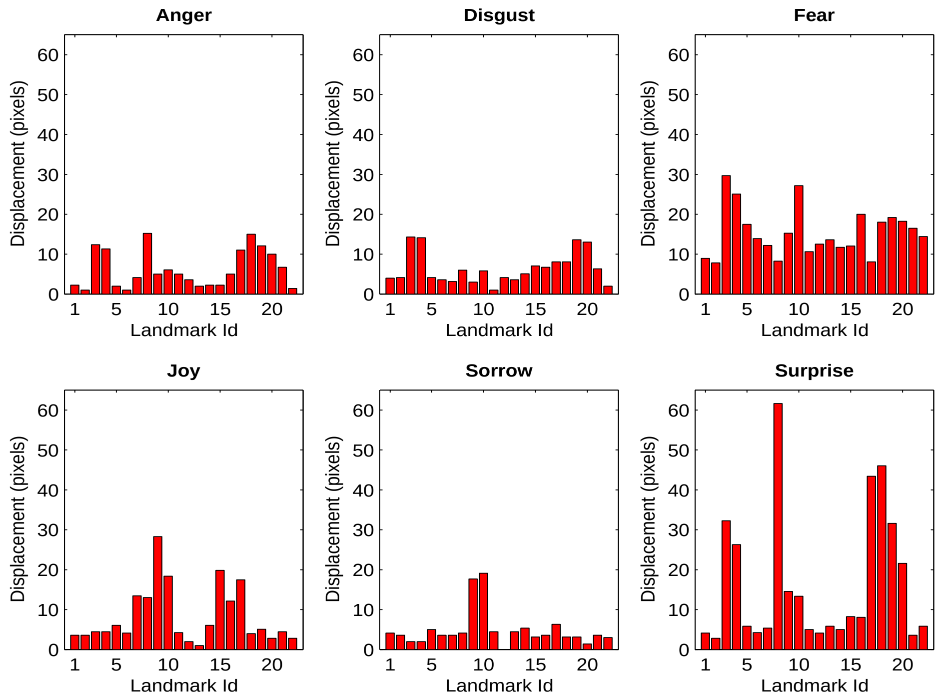


Figure 6: Typical characteristic emotion patterns by motion dependent facial expression features (landmarks).

eigenvalue problem

$$S_w^+ S_b \mathbf{v}_k = \lambda \mathbf{v}_k, \tag{7}$$

where S_w^+ is the pseudo-inverse of S_w in case S_w is rank deficient. G is thus constructed by taking the eigenvectors corresponding to the r largest eigenvalues. In this work, we follow the standard way and let $r = K - 1 = 5$.

For static facial expression features, we can obtain a transformation matrix $G^{\mathrm{S}} \in \mathbb{R}^{d^{\mathrm{S}} \times r}$ by solving Eq. (6). For a new coming facial expression, its transformed static facial expression vector is

$$\mathbf{q}^{\mathrm{S}} = \left(G^{\mathrm{S}}\right)^T \mathbf{x}^{\mathrm{S}}. \tag{8}$$

Similarly, its transformed motion dependent facial expression feature vector is

$$\mathbf{q}^{\mathrm{M}} = \left(G^{\mathrm{M}}\right)^T \mathbf{x}^{\mathrm{M}}. \tag{9}$$

where $G^{\mathrm{M}} \in \mathbb{R}^{d^{\mathrm{M}} \times r}$ is the transformation matrix for motion dependent features.

Now we have two representations $\mathbf{q}^{\mathrm{S}} \in \mathbb{R}^r$ and $\mathbf{q}^{\mathrm{M}} \in \mathbb{R}^r$ with same number of dimensions for every facial expression. Therefore, in order to make use of both of them we can construct a new feature vector as following:

$$\mathbf{q} = \begin{bmatrix} \mathbf{q}^{\mathrm{S}} \\ \mathbf{q}^{\mathrm{M}} \end{bmatrix}. \tag{10}$$

$\mathbf{q}$ in Eq. (10) is the most straightforward and naive form to integrate static and motion dependent facial expression features, and we call it as *hybrid facial expression feature vector.* We will further analyze it and derive a more theoretically elegant form for feature integration later in Section 4.

4. KERNEL FACIAL EXPRESSION RECOGNITION FOR EMOTION DETECTION

Once the hybrid facial expression feature vector $\mathbf{q}$ is computed, any traditional classification method can be used for emotion detection. In this work, we use support vector machine (SVM) because of its elegant theoretical foundation and powerful classification capability. A special benefit of using SVM therefore kernel is that our hybrid facial expression feature vector can be conveniently used from kernel perspective. Specifically, let $\mathcal{K}(\mathbf{q}_i, \mathbf{q}_j)$ be a radial basis function (RBF) kernel on $\mathbf{q}_i$:

$$\begin{aligned} \mathcal{K}(\mathbf{q}_i, \mathbf{q}_j) &= \exp\left(-\gamma \|\mathbf{q}_i - \mathbf{q}_j\|^2\right) \\ &= \exp\left(-\gamma \|\mathbf{q}_i^{\mathrm{S}} - \mathbf{q}_j^{\mathrm{S}}\|^2\right) \exp\left(-\gamma \|\mathbf{q}_i^{\mathrm{M}} - \mathbf{q}_j^{\mathrm{M}}\|^2\right) \\ &= \mathcal{K}_S(\mathbf{q}_i^{\mathrm{S}}, \mathbf{q}_j^{\mathrm{S}}) \mathcal{K}_M(\mathbf{q}_i^{\mathrm{M}}, \mathbf{q}_j^{\mathrm{M}}), \end{aligned} \tag{11}$$

where $\mathcal{K}_S(\mathbf{q}_i^{\mathrm{S}}, \mathbf{q}_j^{\mathrm{S}})$ and $\mathcal{K}_M(\mathbf{q}_i^{\mathrm{M}}, \mathbf{q}_j^{\mathrm{M}})$ are the kernels on transformed static facial expression feature vectors and transformed motion dependent facial expression vectors, respectively. Therefore, the construction of $\mathbf{q}$ in Eq. (10) indeed can be seen as a multiplicative kernel, which comprises information from both static and motion dependent data. Similarly, $\mathcal{K}(\mathbf{q}_i, \mathbf{q}_j)$ can been as an additive kernel when linear kernel is used, and the same conclusions can be drawn for other popular kernels used in SVM.

Therefore, instead of devising an explicit feature construction as in Eq. (10), we may focus on the design of a discriminative kernel for improved classification accuracy. Most importantly, by introducing kernel, explicit construction of hybrid facial expression feature vector in Eq. (10) is no longer needed. Instead, we may use a more flexible and meaningful kernel for better classification.

4.1 A Brief Review of Support Vector Machine

Standard SVM algorithm deals with binary classification problem, which comprises only two classes: positive class and negative class. Therefore the label indicator is $y_i = \{-1, +1\}$, such that $y_i = +1$ if data point $\mathbf{q}_i$ belongs to the positive class, and -1 otherwise. In standard SVM algorithm, a linear classifier $\mathbf{w}$ is optimized to maximize the margin between the data points of two different classes [2], and the learning problem is thus to resolve a quadratic constrained optimization problem whose primal form is

$$\begin{aligned} \min_{\mathbf{w},\xi,b} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n}\xi_i, \\ s.t. \quad & y_i\left(\mathbf{w}^T\phi\left(\mathbf{q}_i\right) + b\right) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad 1 \leq i \leq n, \end{aligned} \tag{12}$$

and its dual form is

$$\begin{aligned} \max \quad & \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathcal{K}\left(\mathbf{q}_i, \mathbf{q}_j\right), \\ s.t. \quad & 0 \leq \alpha_i \leq C, \\ & \sum_i \alpha_i y_i = 0, \quad 1 \leq i \leq n, \end{aligned} \tag{13}$$

where b is a threshold, the ξ_i are slack variables necessary for the case when the training data points are not linearly separable, and C is the error penalty. $\mathcal{K}(\mathbf{q}_i, \mathbf{q}_j) = \langle\phi(\mathbf{q}_i), \phi(\mathbf{q}_i)\rangle$ is a kernel function, by which a data point $\mathbf{q}_i$ is mapped into a higher (maybe infinite) dimensional space by the mapping function ϕ. Thus the class membership assigned to an unseen data point $\mathbf{q}$ is given by

$$\mathrm{sign}\left(f(\mathbf{q}) = \sum_{i=1}^{n} y_i\alpha_i\mathcal{K}\left(\mathbf{q}_i, \mathbf{q}\right) + b\right), \tag{14}$$

where $f(\mathbf{q})$ is the decision function of the SVM.

Extending SVM to multi-class classification, such as in emotion detection, has been well studied and many approaches have been devised [3], among which the "one-against-one" approach [11] is widely used because of its advantages as analyzed in [9]. In this approach, $K(K-1)/2$ classifiers are constructed and one for each pair of two different classes. After that, a voting strategy is used: each binary classification is considered to be a voting, and a data point is finally designated to be in the class with maximum number of votes.

4.2 Hybrid Kernel Construction to Enhanced Utilization of Facial Expression Features

From Eq. (11), we can conclude that the construction of hybrid facial expression feature vector in Eq. (10) can be seen as a hybrid kernel. Therefore, instead of using simple vector concatenation, we may use a more meaningful kernel suitable for facial expression recognition.

Besides the RBF kernel used in Eq. (11), a number of other kernel functions broadly used in SVM are listed in Table 1. Empirically, when using RBF kernel for transformed static facial expression vector $\mathbf{q}^{\mathrm{S}}$ and second order polynomial kernel for transformed motion dependent facial expression vector $\mathbf{q}^{\mathrm{M}}$, we achieve best classification performance when using JAFFE database as training data. More specifically, in

Table 1: Popular kernels used in SVM.

Kernel type	Kernel function
Linear	$\mathbf{u}\cdot\mathbf{v}$
Polynomial	$(\gamma\mathbf{u}\cdot\mathbf{v} + c)^{\mathrm{degree}}$
RBF	$\exp\left(-\gamma\|\mathbf{u}-\mathbf{v}\|^2\right)$
Sigmoid	$\tanh\left(\gamma\mathbf{u}\cdot\mathbf{v} + c\right)$

(a) Face contour. (b) Facial landmarks.

Figure 7: Facial feature points generated by Luxand faceSDK, including face contour points (blue points in Figure 7(a)) and main landmark points (white crosses in Figure 7(b)).

this work, the hybrid kernel used in SVM for classification is:

$$\begin{aligned} \mathcal{K}_{ij} &= \mathcal{K}_S(\mathbf{q}_i^{\mathrm{S}}, \mathbf{q}_j^{\mathrm{S}}) \cdot \mathcal{K}_M(\mathbf{q}_i^{\mathrm{M}}, \mathbf{q}_j^{\mathrm{M}}), \\ &= \left[\exp\left(-\gamma_{\mathrm{S}}|\mathbf{q}_i^{\mathrm{S}} - \mathbf{q}_i^{\mathrm{S}}|^2\right)\right] \cdot \left(\gamma_{\mathrm{M}}\mathbf{q}_i^{\mathrm{M}} \cdot \mathbf{q}_i^{\mathrm{M}} + c\right)^2. \end{aligned} \tag{15}$$

where the parameters γ_{S}, γ_{M} and c are fine tuned using standard ten-fold cross-validation [7].

5. IMPLEMENTATION DETAILS

There are three major components in the proposed automatical emotion detection system using DKFER method: facial feature extraction preparation component, feature data process component and classification component. The last two components constitute the proposed DKFER method.

For facial feature extraction preparation component, we use OpenCV[2] to capture face picture and Luxand faceSDK[3] (version 1.7) to identify facial feature points. The Luxand faceSDK generates 40 facial feature points including face contours and main landmarks, as shown in Figure 7.

The data process are already described in detail in Section 2 and Section 3. Using the hybrid kernel $\mathcal{K}$ is constructed in Eq. (15) as elaborated in Section 4, we LIBSVM library[4] to implement SVM for classification. LIBSVM supports multi-class classification, it outputs probability of the output class membership decisions.

[2] `http://opencv.willowgarage.com/wiki/`
[3] `http://www.luxand.com/facesdk/`
[4] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

Figure 8: User interface of the automatical emotion detection system using the proposed DKFER method.

The user interface of the automatical emotion detection system using the proposed DKFER method is shown in Figure 8. A facial expression in neutral condition is first captured and stored, and then the new facial expressions are captured for every 3 second and the emotion is output as in the top left panel in Figure 8.

6. EMPIRICAL STUDIES

In this section, we evaluate the classification performance for emotion detection of the proposed DKFER method. We run the proposed DKFER method using standard five-fold cross-validation, and compare to two recent automatic emotion detection methods using facial expressions: 1, Fuzzy mode (FM) method [5] which uses only static facial features, and displacement computation (DC) method [14] which use only motion dependent facial features. We use JAFFE database for evaluation, which contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

We evaluate the classification accuracy of every emotion and the overall classification accuracy over all the six emotions, which are reported in Figure 9. From the results, we can see that the proposed DKFER method consistently outperforms the other two compared methods. This concretely confirms the effectiveness of our method, and demonstrates that using both static and motion dependent facial expressions is superior to using only one of them.

7. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a novel automatical human emotion detection technique using Discriminative Kernel Facial Emotion Recognition (DKFER) method. Different from previous related works which use only either static facial expression data or motion dependent facial expression data, the proposed DKFER method integrates both of them in a hybrid *kernel* so as to make use of the information from the both data sources. In addition, we used linear discriminant analysis (LDA) to transform these two types of data into two lower-dimensional subspaces with close dimensionalities, such that the data integration can be carried out in a balanced manner. Most importantly, in the reduced *discriminative* subspaces, data points representing different emotions are separated apart far away from one another, whereas those representing a same emotion are mapped close together in a more compact way. Therefore, the subsequent classification using support vector machine (SVM) is more efficient and effective. Promising experimental results demonstrated the clear advantages of our proposed DKFER method.

In the future, we will further develop our method mainly in the following two aspects. First, instead of using only two facial expression images of a face as in the current work, we will consider to use multiple consecutive frames to model motion dependent facial features using Markov Random Field. Second, the proposed DKFER method indeed presents a general framework to integrate heterogeneous data for emotion detection. Besides facial expressions, we will incorporate verbal or non-verbal information, such as speech signals. These emotion data from other sources are first preprocessed as in Section 2 and Section 3 to generate discriminative features, and then integrated into the current model using a data specific kernel.

ACKNOWLEDGEMENTS

This research is supported by NSF IIS-1015219, NSF CCF-0830780, NSF CCF-0939187, NSF CCF-0917274, NSF DMS-0915228, NSF CNS-0923494, UTA-REP. The authors would like to thank NSF Scholar Award to Attend Doctoral Consortium of Petra 2010.

8. REFERENCES

[1] R. Chellappa, C. Wilson, S. Sirohey, et al. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

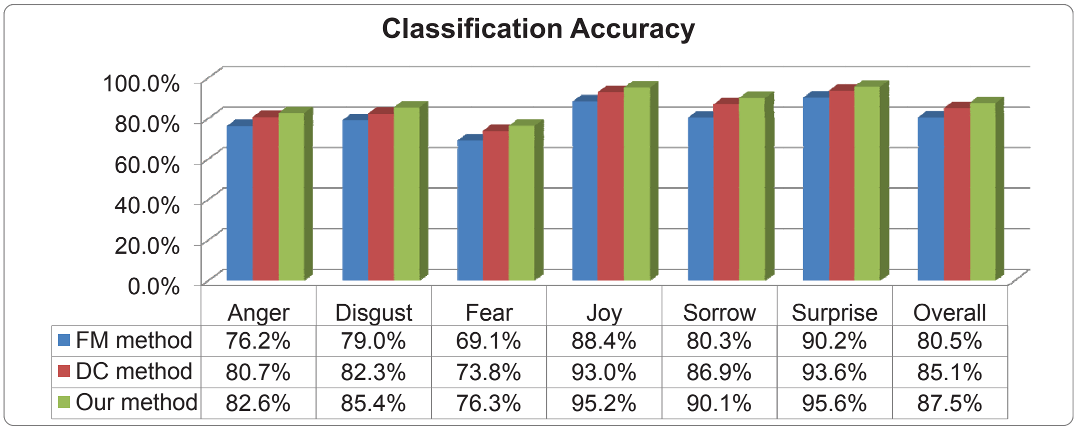
[2] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods.* Cambridge Univ Pr, 2000.

[3] K. Duan and S. Keerthi. Which is the best multiclass SVM method? An empirical study. *Multiple Classifier Systems*, 3541:278–285, 2005.

[4] P. Ekman and W. Friesen. Facial action coding system: A technique for the measurement of facial movement, 1978.

[5] N. Esau, E. Wetzel, L. Kleinjohann, and B. Keinjohann. Real-time facial expression recognition using a fuzzy emotion model. In *IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007*, pages 1–6, 2007.

[6] W. Fellenz, J. Taylor, R. Cowie, E. Douglas-Cowie, F. Piat, S. Kollias, C. Orovas, and B. Apolloni. On emotion recognition of faces and of speech using neuralnetworks, fuzzy logic and the ASSESS system. In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, volume 2, 2000.



	Anger	Disgust	Fear	Joy	Sorrow	Surprise	Overall
FM method	76.2%	79.0%	69.1%	88.4%	80.3%	90.2%	80.5%
DC method	80.7%	82.3%	73.8%	93.0%	86.9%	93.6%	85.1%
Our method	82.6%	85.4%	76.3%	95.2%	90.1%	95.6%	87.5%

Figure 9: Classification accuracies of three compared methods for emotion detection on JAFFE database.

[7] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, 1990.

[8] D. Graham and N. Allinson. Face recognition: From theory to applications. *NATO ASI Series F, Computer and Systems Sciences*, 163:446–456, 1998.

[9] C. Hsu and C. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.

[10] S. Kimura and M. Yachida. Facial expression recognition and its degree estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 295–300, 1997.

[11] U. Kressel. Pairwise classification and support vector machines. *Advances in kernel methods: support vector learning*, pages 255–268, 1999.

[12] I. Maglogiannis, D. Vouyioukas, and C. Aggelopoulos. Face detection and recognition of natural human emotion using Markov random fields. *Personal and Ubiquitous Computing*, 13(1):95–101, 2009.

[13] A. Metallinou, L. S., and N. S. Audio-visual emotion recognition using gaussian mixture models for face and voice. *Proceedings of the tenth IEEE Symposium on Multimedia*, 2008.

[14] P. Michel. Support vector machines in automated emotion classification. *Churchill College, June*, 2003.

[15] M. Rizon, M. Karthigayan, S. Yaacob, and R. Nagarajan. Japanese face emotions classification using lip features. *Geometric Modeling and Imaging, 2007. GMAI'07*, pages 140–144, 2007.

[16] J. Rottenberg and J. Gross. Emotion and emotion regulation: A map for psychotherapy researchers. *Clinical Psychology: Science and Practice*, 14(4):323–328, 2007.

[17] M. Shackleton and W. Welsh. Classification of facial features for recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91.*, pages 573–579, 1991.

[18] D. Sloan and A. Kring. Measuring changes in emotion during psychotherapy: Conceptual and methodological issues. *Clinical Psychology: Science and Practice*, 14(4):307–322, 2007.

[19] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380, 2000.

[20] M. Suwa, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression. In *International Joint Conference on Pattern Recognition*, pages 408–410, 1978.

[21] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.

[22] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. In *IEEE Computer Society Conference On Computer Vision And Pattern Recognition*, pages 70–70. Citeseer, 1994.

[23] Z. Zeng, Y. Hu, Y. Fu, T. Huang, G. Roisman, and Z. Wen. Audio-visual emotion recognition in adult attachment interview. In *Proceedings of the 8th international conference on Multimodal interfaces*, page 145. ACM, 2006.

[24] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003.