Univ Access Inf Soc
DOI 10.1007/s10209-013-0312-5

LONG PAPER

Emotion Detection via Discriminant Laplacian Embedding

Hua Wang · Heng Huang · Fillia Makedon

Abstract Human emotion detection is of substantial importance in a variety of 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. This work proposes to place the emotion detection problem under the framework of Discriminant Laplacian Embedding (DLE) to integrate these two types of facial expression data in a shared subspace, such that the advantages of both of them are exploited. Due to the reinforcement between the two types of facial features, the new data representation is more discriminative and easier to classify. Encouraging experimental results in empirical studies demonstrate the practical usage of the proposed DLE method for emotion detection.

H. Wang
Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401, USA
e-mail: huawangcs@gmail.com

H. Huang (✉) · F. Makedon
Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
e-mail: heng@uta.edu

F. Makedon
e-mail: makedon@uta.edu

Keywords Emotion detection · Facial expression recognition · Discriminative learning · Kernel · Facial feature

1 Introduction

In recent years, pervasive and ubiquitous systems, which can assist with the in-home care of the elderly and people with the chronic cognitive disabilities or traumas, such as Alzheimer's disease and Parkinson's disease, have become increasingly useful in everyday life. Enabling such persons to remain productive and active is of high socioeconomic importance to every country's workforce and health care system. Assistive environments is a field that addresses the fact that as people grow older or become disabled, they increasingly rely on computer technology to be able to live in and function in their homes [6, 21]. As an important component of the intelligent cognitive systems for assistive environments, automatic human emotion detection systems play a significant role in pervasive applications in assistive environments [21, 22], such as those intended for handicapped persons or persons with cognitive disabilities who are not able to correctly recognize emotions [2, 12, 38], or for developing intelligent graphic user interface (GUI) environments [31].

Among a number of body languages conveying human emotion, facial expression is generally recognized as the most straightforward and effective one, which thereby has attracted a lot of attention in scientific search [5, 20, 29] and stimulated a plethora of facial expression-based automatic human emotion detection methods for a large range of practical applications [18, 23, 24, 26, 37], from psychology [27, 30] to the applied science such as pervasive health care. Existing facial expression recognition methods

Published online: 27 July 2013

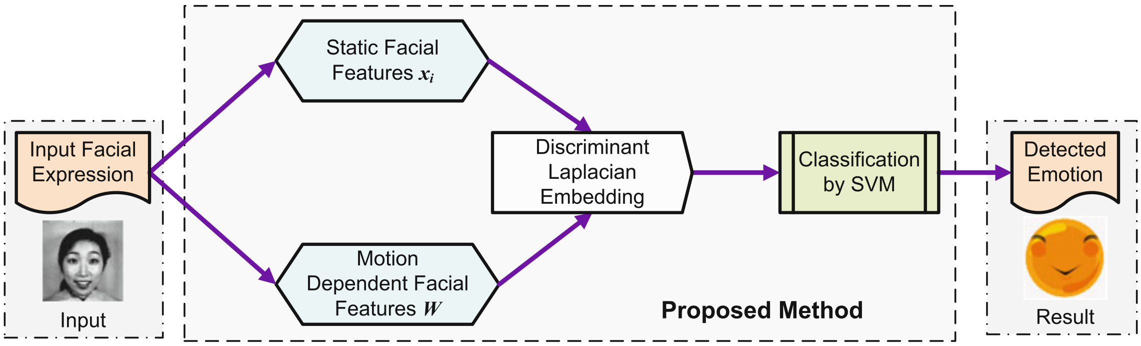


Fig. 1 Diagram of the proposed Discriminant Laplacian Embedding (DLE) method

can usually be categorized into two classes: *static* methods or *motion-dependent* methods. In the former, recognition of a facial expression is performed using one single image of a face, which examines the geometrical relations established among facial features or facial organs, such as eyes, eyebrows, and mouth, and the variation of location of features [18, 28, 32, 33]. A limitation of static methods lies in their dependencies on the precise alignment of faces, which, sometimes, is difficult to achieve for real-world data due to head rotations 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 [10, 23, 36]. 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 order to tackle the difficulties of the two classes of existing methods and better exploit available data, previous work of the authors [35] has made a moderate attempt to integrate static data and motion-dependent data together. In [35], a Discriminant Laplacian Embedding (DLE) method for emotion detection was proposed, 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) [11], respectively. In the resulting subspaces, data points representing different emotions are separated apart far away from one another, while those representing the same emotion are mapped close together in a more compact manner. Then, the two transformed feature vectors are integrated via a hybrid kernel [4], such that the performance of subsequent classifications carried out on it is doubly enhanced for the sake of taking advantage of emotion information contained in both types of facial expression data.

Although the discriminative kernel (DK) method [35] has demonstrated its effectiveness, a defect prevents it from exploring the full potential of the available data. Because the DK method learns the subspaces from the static facial features and motion-dependent facial features separately via two different LDA processes, their interrelatedness is not exploited. To address this, in this work, instead of learning the new data representation for the two types of facial features separately, the emotion detection problem is placed under the framework of DLE proposed in [34] to learn a shared subspace from the two original data spaces. As a result, the new data representation enjoys the advantages of the both types: the discriminativity from the static facial features and locality consistency from the motion-dependent facial features. The work flow of the proposed DLE method for emotion detection is illustrated in Fig. 1. When an input facial expression (left picture) comes in, the task is to perform the proposed DLE method on it to decide the conveyed human emotion (right cartoon) among a list of prescribed emotions.

In this work, six human emotions are considered: 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 Fig. 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 Fig. 3. These facial feature points will be used to construct facial expression features in the sequel, including both static and motion-dependent expression features.

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

Fig. 2 Sample facial expressions for neutral condition and six emotions from JAFFE database. **a** Neutral. **b** Anger. **c** Disgust. **d** Fear. **e** Joy. **f** Sorrow. **g** Surprise

Fig. 3 Typical facial feature points (landmarks) shown by *green circles*

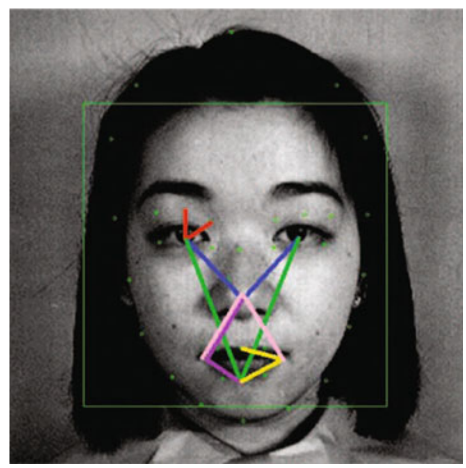


Fig. 4 Static facial expression features: angels. *A*0 (*rose color*), *A*1 (*green color*), *A*2 (*red color*), *A*3 (*blue color*), *A*4 (*yellow color*), *A*5 (*purple color*)

2.1 Static facial expression features

Static facial expression features are extracted from one single image of a face to assess the 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, which is necessary for distance-based features [9]. Furthermore, typical angles show a large coincidence between different persons, whereas typical distances vary considerably between different persons [9]. To these ends, six angle features are used in this work, which are shown in Fig. 4.

The angles are separated into two groups: A_2 and A_3 belong to the upper part of the face; 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 [8, 9], 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 and 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 Fig. 3, motion-dependent facial expression features may be defined. The simplest motion-dependent facial features can be defined as the displacements (Euclidean distance) of these facial feature points between a neutral facial expression and the "peak" of a particular emotive expression [25], as illustrated in Fig. 5. As a result, every input facial expression is quantified as a motion-dependent facial expression feature vector as follows:

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

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

Although the fixed length feature vectors as described above are easy to handle in many machine learning algorithms, it suffers from a critical problem, i.e., baseline neutral face expression selection. The correct neutral face expression is often not easy to select due to many reasons, such as camera speed, facial trait variations between persons, etc. Because the neutral face expression works as the reference and facial feature vectors for other expressions are computed from it, an inaccurate selection of neutral face expressions will detriment the subsequent classification performance. Therefore, it is desirable to design a type of motion-dependent facial expression features that are independent of neutral face expression selection. To this end, instead of using the facial landmarks displacement vectors as in the previous conference publication in [35], in this paper, the pairwise similarity between peak facial expressions is used. Specifically, given the current facial landmark position vector $\mathbf{x}_i^{\mathrm{M}}$ and that of another expression $\mathbf{x}_j^{\mathrm{M}}$, the similarity between the two facial expressions can be computed as follows:

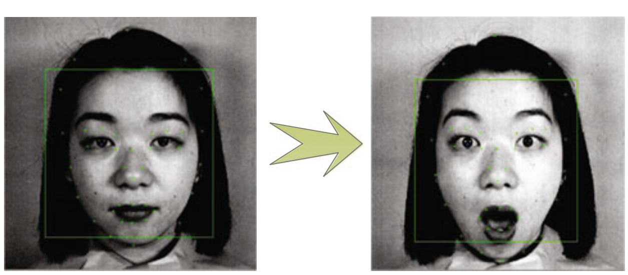


Fig. 5 Motion-dependent facial expression features: facial feature points displacements

$$W_{ij} = \exp\left(\frac{\|\mathbf{x}_i^{\mathrm{M}} - \mathbf{x}_j^{\mathrm{M}}\|^2}{\sigma^2}\right), \tag{3}$$

where σ is a hyperparameter. Obviously, in this formulation, the impact of the baseline neutral facial expression is implicitly removed. Pairwise similarity naturally leads to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{\mathbf{x}_i^{\mathrm{M}}\}$ is the vertex set consisting of all the concerned facial expressions, and $\mathcal{E} = \mathcal{V} \times \mathcal{V}$. Give an edge $e_{ij} \in \mathcal{E}, W_{ij}$ measures the similarity between the vertices, i.e., facial expressions, connected by e_{ij}.

3 Enriched facial expression features via Discriminant Laplacian Embedding

Given the static facial expression features $\mathbf{x}^{\mathrm{S}}$ or motion-dependent features W as in Sect. 2, one can conduct emotion detection on them using statistical classification techniques. Typical characteristic emotion patterns, such as those introduced in [8, 25], can be used to build classifiers. However, directly using these original facial expression features suffer from three critical problems: inefficiency, ineffectiveness, and data utilization.

First, computational complexity is one of the major considerations in designing a practical emotion detection system. Therefore, one 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 with 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 effective.

Last, but not least, in order to better exploit the available data, the aim is at utilizing both static- and motion-dependent facial expression features to build a unified framework to make use of these two types of features at the same time to boost the emotion detection rate.

In order to tackle the difficulties and achieve better emotion classification accuracy, a discriminative kernel method has been proposed in the previous conference publication of the authors [35], which reduces the dimensionality of the two types of features into the same number of dimensions and builds a hybrid kernel upon the reduced features. As discussed earlier, although this method utilizes both of the two types' features, it performs dimension reduction separately and does not have the capability to exploit the inter-relatedness between them. To address this,

this work proposes to build the data representation under the framework of DLE [34], which seamlessly integrates the two types of the features, i.e., static facial features in vector form and motion-dependent features in graph form.

3.1 Formulation of discriminant Laplacian analysis

For a classification task with n data points and K classes, each data point $\mathbf{x}_i \in \mathbb{R}^p$ is associated with a subset of class labels represented by a binary vector $\mathbf{y}_i \in \{0, 1\}^K$ such that $\mathbf{y}_i(k) = 1$ if $\mathbf{x}_i$ belongs to the kth class, and 0 otherwise. In single-label classification, each data point belongs to only one class, $\sum_k \mathbf{y}_i(k) = 1$, while in multi-label classification, each data point may belong to multiple classes at the same time, $\sum_k \mathbf{y}_i(k) \geq 1$. Meanwhile, there are also pairwise similarities $W \in \mathbb{R}^{n\times n}$ among the n data points with W_{ij} indicating how close $\mathbf{x}_i$ and $\mathbf{x}_j$ are related. Suppose the number of labeled data points is $l(<n)$, the goal is to predict labels $\{\mathbf{y}_i\}_{i=l+1}^n$ for the unlabeled data points $\{\mathbf{x}_i\}_{i=l+1}^n$. Thus, $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$. In the context of emotion detection, $\mathbf{x}_i$ is obtained from static facial expression features as in Sect. 2.1, while W is obtained from motion-dependent facial expression features as in Sect. 2.2

Traditional embedding algorithms learn a projection only from one type of data, either attribute data X, such as principal component analysis (PCA) [16] and linear discriminant analysis (LDA) [11], or pairwise similarity data W, such as Laplacian Eigenmap [1] and locality preserving projection (LPP) [14]. Besides, though useful, the label information from training data $\{\mathbf{y}_i\}_{i=1}^l$ is not always used, because many embedding algorithms are devised closely in conjunction with clustering algorithms which are unsupervised by nature. Therefore, it is proposed to use DLE [34] approach to realize these two expectations, data integration and making use of supervision information, which produces a transformation $U \in \mathbb{R}^{p\times r}$ by solving the following optimization objective [34]:

$$\arg\max_U \mathbf{tr}\left(U^T\left(A_+^{-\frac{1}{2}}S_w^{-\frac{1}{2}}S_bS_w^{-\frac{1}{2}}A_+^{-\frac{1}{2}}\right)U\right), \tag{4}$$

where r ($\ll p$) is the dimensionality of the projected subspace, S_b and S_w are the between-class and within-class scatter matrices defined just as in standard LDA, and $A = X(D-W)X^T$ with $D = \mathrm{diag}(d_1, \ldots, d_n), d_i = \sum_j W_{ij}$. Thus, $L = D - W$ is the graph Laplacian [3]. The solution to Eq. (4) is well established in mathematics by solving the following eigenvalue problem:

$$\left(A_+^{-\frac{1}{2}}S_w^{-\frac{1}{2}}S_bS_w^{-\frac{1}{2}}A_+^{-\frac{1}{2}}\right)\mathbf{u}_k = \lambda_k\mathbf{u}_k, \tag{5}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the resulting eigenvalues and $\mathbf{u}_k$ are the corresponding eigenvectors. Hence, $U = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$, and r is empirically selected as $K-1$.

Then, $\mathbf{x}_i$ may be projected into the embedding space as:

$$\mathbf{q}_i = U^T\mathbf{x}_i, \tag{6}$$

and subsequent classification is carried out using $\mathbf{q}_i$.

3.2 DLE backgrounds

Given training data $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^l$, LDA targets on projecting the original data from a high p-dimensional space to a much lower r-dimensional subspace to separate different classes as much as possible (maximize between-class scatter S_b), while condense each class as much as possible (minimize within-class scatters S_w). Computing S_b and S_w by the standard definitions in LDA [11], the standard LDA optimization objective to maximize is achieved:

$$J_{\mathrm{LDA}} = \mathbf{tr}\left(\frac{G^TS_bG}{G^TS_wG}\right). \tag{7}$$

The transformation $G \in \mathbb{R}^{p\times r}$ is usually obtained by applying the eigen-decomposition on S_w^{-1} S_b (or sometimes by solving the generalized eigenvalue problem $S_b\,\mathbf{u}_k = \lambda_k\,S_w\,\mathbf{u}_k$), when S_w is nonsingular. This method, however, cannot guarantee the orthonormality of G. To this end, motivated by the way frequently used in spectral clustering, letting $F = S_w^{\frac{1}{2}}G$, the following symmetric objective is maximized:

$$J_{\mathrm{LDA}} = \mathbf{tr}\left(F^TS_w^{-\frac{1}{2}}S_bS_w^{-\frac{1}{2}}F\right), \tag{8}$$

where the constraint $F^T F = I$ is automatically satisfied and thereby removed from the formulation.

Given the pairwise similarity W, Laplacian embedding preserves the same relationships and maximize the smoothness with respect to the intrinsic manifold of the data set in the embedding space by minimizing [13]

$$\sum_{i,j}\|\mathbf{q}_i - \mathbf{q}_j\|^2W_{ij} = \mathbf{q}_i^T(D-W)\mathbf{q}_i. \tag{9}$$

For multi-dimensional embedding, Eq. (9) becomes $\mathbf{tr}(Q^T(D-W)Q)$, using linear embedding $Q^T = [\mathbf{q}_1, \ldots, \mathbf{q}_n] = F^TX$. Hence, Eq. (9) can be written as follows:

$$J_{\mathrm{Lap}} = \mathbf{tr}\left(F^TX(D-W)X^TF\right). \tag{10}$$

3.3 Motivations and formulation of DLE

As an embedding to leverage both attributes and pairwise similarity of a same data set is sought, given two individual optimization objectives as in Eqs. (8) and (10), an additive embedding objective may be constructed to maximize:

$$J_{\mathrm{DLE}} = \alpha J_{\mathrm{LDA}} - (1-\alpha)J_{\mathrm{Lap}}, \tag{11}$$

where $0 < \alpha < 1$ is a tradeoff parameter. In practice, however, it is hard to choose an optimal α. Therefore, instead of using the trace of *difference*, the objective is formulated as the trace of *quotient* so that α is removed as follows:

$$J_{\mathrm{DLE}} = \mathbf{tr}\left(\frac{F^T S_w^{-\frac{1}{2}} S_b S_w^{-\frac{1}{2}} F}{F^T X(D-W)X^T F}\right). \tag{12}$$

Let $A = X(D-W)X^T$ and $U = (A)_+^{-\frac{1}{2}} F$, the symmetric optimization objective of the proposed DLE is obtained as in Eq. (4), which yields orthonormal U as the solution.

Considering the fact that A is usually rank deficient, let

$$A = V \begin{bmatrix} \Sigma & \\ & \mathbf{0} \end{bmatrix} V^T \tag{13}$$

be the eigen-decomposition of A with diagonal line of Σ being the positive eigenvalues of A, where:

$$A_+^{\frac{1}{2}} = V_1 \Sigma^{\frac{1}{2}} V_1^T \quad \text{and} \quad A_+^{-\frac{1}{2}} = V_1 \Sigma^{-\frac{1}{2}} V_1^T, \tag{14}$$

where V_1 is composed of the eigenvectors of A corresponding to its positive eigenvalues. Note that, A is semi-positive definite and thereby has no negative nonzero eigenvalues.

Because in many real applications such as computer vision and document content analysis, the number of features of a data set is often much larger than that of data points, i.e., $p > n$, S_w could also be rank deficient. In this case, $S_w^{-\frac{1}{2}}$ in Eq. (4) is replaced by $S_{w+}^{-\frac{1}{2}}$, which is computed in a same way as in Eqs. (13–14).

Finally, through an integral optimization objective, together with the supervision information contained in training data, data attributes and pairwise similarities on all the data points including those unlabeled are integrated by DLE. Therefore, DLE provides a framework for semi-supervised learning, which is summarized in Table 1.

3.4 Classification procedures

In this work, the support vector machine (SVM) [4] is used to perform classification tasks on the learned new data representations $\mathbf{q}_i$. The traditional binary-class SVM is extended to multi-class classification for emotion detection following [7], where the "one-against-one" approach [19] is used because of its advantages as analyzed in [15]. 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.

Table 1 Semi-supervised classification using DLE

Input:

$\{\mathbf{x}_i\}_{i=1}^n$: Attribute vectors for all the data points

$\{\mathbf{y}_i\}_{i=1}^l$: labels for the training data points

W: pairwise relationship among all the data points

Steps:

(1) Compute S_b and S_w as in standard LDA algorithm

(2) Compute $A = X(D-W)X^T$

(3) Resolve the eigenvalue problem in Eq. (5). Construct the projection matrix U by the resulting eigenvectors corresponding to the $(K-1)$ largest eigenvalues

(4) Compute the projected vectors $\{\mathbf{q}_i\}_{i=1}^n$ for all the data points including those unlabeled as in Eq. (6)

(5) Use projected training data $\{(\mathbf{q}_i, \mathbf{y}_i)\}_{i=1}^l$ to classify the testing data points $\{\mathbf{q}_i\}_{i=l+1}^n$ via 1NN method

Output:

Class labels for testing data points $\{\mathbf{y}_i\}_{i=l+1}^n$

4 Implementation details

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

For facial feature extraction preparation component, OpenCV[2] is used 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 Fig. 6.

The data processing is described in detail in Sects. 2 and 3. Upon the projected feature vectors $\mathbf{q}_i$ in 6, the LIBSVM library[4] is used to implement SVM for classification. LIBSVM supports multi-class classification, and it outputs probability of the output class membership decisions.

The user interface of the automatic emotion detection system using the proposed DLE method is shown in Fig. 7. The facial expressions are captured as a rate of 20 frames per minute, and the emotion is output as in the top left panel in Fig. 7.

5 Empirical studies

In this section, an evaluation of the classification performance for emotion detection of the proposed DLE method is presented. The proposed DLE method is run using

[2] http://opencv.willowgarage.com/wiki/.

[3] http://www.luxand.com/facesdk/.

[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

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

Fig. 6 Facial feature points generated by Luxand faceSDK, including face contour points (*blue points* in **a**) and main landmark points (*white crosses* in **b**)

Fig. 7 User interface of the automatic emotion detection system using the proposed DLE method

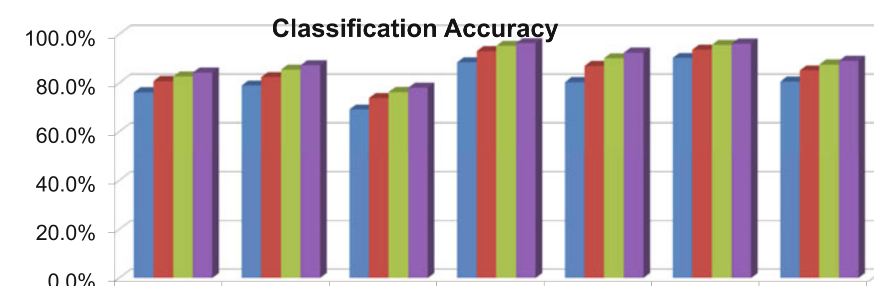
standard fivefold cross-validation and compared with two recent automatic emotion detection methods using facial expressions: fuzzy mode (FM) method [9], which uses only static facial features, and displacement computation (DC) method [25], which use only motion-dependent facial features. In addition, the proposed method is also compared with the discriminative kernel (DK) method presented in a previous publication of the authors [35]. Two benchmark data sets are used for evaluation. The first set is the **JAFFE** database, 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. The second data set is the **Cohn–Kanade** database [17], which consists of 100 university students aged from 18- to 30-years-old, of which 65 % were female, 15 % were African-American, and 3 % were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of prototypic emotions.

The classification accuracy of every emotion and the overall classification accuracy over all the six emotions are evaluated, and the results are reported in Figs. 8 and 9. From the results, it can be seen that the proposed DLE method consistently outperforms the other three compared methods. This concretely confirms the effectiveness of the proposed method and demonstrates that using both static- and motion-dependent facial expressions is superior to using only one of them.

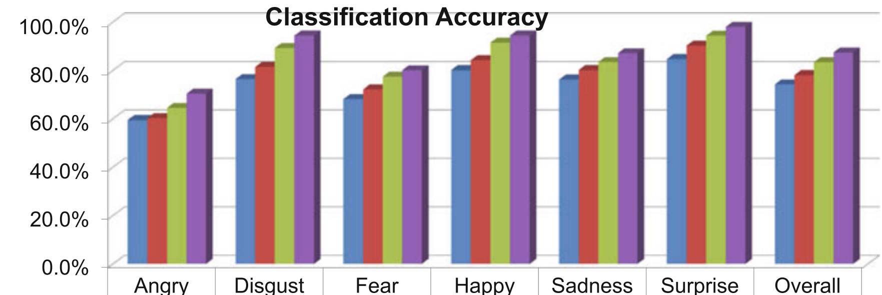
6 Conclusion and future works

In this paper, a novel automatic human emotion detection technique using DLE method has been proposed. Different from previous related works that use only static facial expression data or motion-dependent facial expression data, the proposed DLE method learns a shared subspace from the original spaces of the two different types of facial features. In the learned subspace, the new data

Fig. 8 Classification accuracies of three compared methods for emotion detection on JAFFE database



	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%
DK method	82.6%	85.4%	76.3%	95.2%	90.1%	95.6%	87.5%
Our method	84.2%	87.2%	78.1%	96.3%	92.4%	96.1%	89.0%



	Angry	Disgust	Fear	Happy	Sadness	Surprise	Overall
FM method	59.4%	76.4%	68.1%	80.1%	76.2%	84.7%	74.2%
DC method	60.2%	81.4%	72.2%	84.2%	80.1%	90.2%	78.0%
DK method	64.5%	89.2%	77.4%	91.5%	83.5%	94.3%	83.4%
Our method	70.4%	94.3%	80.0%	94.4%	87.0%	98.0%	87.4%

Fig. 9 Classification accuracies of three compared methods for emotion detection on Cohn–Kanade database

representations are more discriminative as they take taking advantage of the inter-relatedness between the two types of facial features, which makes the data points easier to classify. Promising experimental results demonstrated the clear advantages of the proposed DLE method.

In the future, it is expected that the method will be further developed mainly in the following two aspects: first, instead of using only two facial expression images of a face as in the current work, the use of multiple consecutive frames will be considered to model motion-dependent facial features using Markov random field. Second, the proposed DLE method indeed presents a general framework to integrate heterogeneous data for emotion detection. Besides facial expressions, verbal or nonverbal information, such as speech signals, will be incorporated. These emotion data from other sources are to be first pre-processed as in Sects. 2 and 3 to generate discriminative features and then integrated into the current model using a data-specific kernel.

Acknowledgments 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.

References

1. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. NIPS'02 (2002)
2. Chellappa, R., Wilson, C., Sirohey, S., et al.: Human and machine recognition of faces: a survey. Proc. IEEE **83**(5), 705–740 (1995)
3. Chung, F.: Spectral Graph Theory. American Mathematical Society, Providence (1997)
4. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines: and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge (2000)
5. Dailey, M., Joyce, C., Lyons, M., Kamachi, M., Ishi, H., Gyoba, J., Cottrell, G.: Evidence and a computational explanation of cultural differences in facial expression recognition. Emotion **10**(6), 874 (2010)
6. Doukas, C., Maglogiannis, I.: Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components. IEEE Trans. Inf. Technol. Biomed. **15**(2), 277–289 (2011)
7. Duan, K., Keerthi, S.: Which is the best multiclass SVM method? An empirical study. Multiple Classif. Syst. **3541**, 278–285 (2005)
8. Ekman, P., Friesen, W.: Facial Action Coding System: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, California (1978)
9. Esau, N., Wetzel, E., Kleinjohann, L., Keinjohann, B.: Real-time facial expression recognition using a fuzzy emotion model. In: IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007, pp. 1–6 (2007)
10. Fellenz, W., Taylor, J., Cowie, R., Douglas-Cowie, E., Piat, F., Kollias, S., Orovas, C., Apolloni, B.: On emotion recognition of faces and of speech using neuralnetworks, fuzzy logic and the ASSESS system. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000 (IJCNN 2000), vol. 2 (2000)
11. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, London (1990)
12. Graham, D., Allinson, N.: Face recognition: from theory to applications. NATO ASI Series F. Comput. Syst. Sci. **163**, 446–456 (1998)
13. Hall, K.: An r-dimensional quadratic placement algorithm. Manag. Sci. **17**(3), 219–229 (1970)
14. He, X., Niyogi, P.: Locality preserving projections. In: NIPS'03 (2003)
15. Hsu, C., Lin, C.: A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. **13**(2), 415–425 (2002)
16. Jolliffe, I.: Principal Component Analysis. Springer, Berlin (2002)
17. Kanade, T., Cohn, J., Tian, Y.: Comprehensive database for facial expression analysis. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings, pp. 46–53. IEEE (2000)
18. Kimura, S., Yachida, M.: Facial expression recognition and its degree estimation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 295–300 (1997)
19. Kressel, U.: Pairwise classification and support vector machines. Adv. Kernel Methods Support Vector Learn, MIT Press, Cambridge, pp 255–268 (1999)
20. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M.: The computer expression recognition toolbox (cert). In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), pp. 298–305. IEEE (2011)
21. Maglogiannis, I., Betke, M., Pantziou, G., Makedon, F.: Assistive environments for the disabled and the senior citizens: theme issue of petra 2010 and 2011 conferences. Pers. Ubiquit. Comput. 1–3, November (2012)
22. Maglogiannis, I., Makedon, F., Pantziou, G., Baillie, L.: Pervasive technologies for assistive environments: special issue of

petra 2008 conference. Pers. Ubiquit. Comput. **14**(6), 469–471 (2010)
23. Maglogiannis, I., Vouyioukas, D., Aggelopoulos, C.: Face detection and recognition of natural human emotion using Markov random fields. Pers. Ubiquit. Comput. **13**(1), 95–101 (2009)
24. Metallinou, A.S.L.S.N.: Audio-visual emotion recognition using gaussian mixture models for face and voice. In: Proceedings of the Tenth IEEE Symposium on Multimedia (2008)
25. Michel, P.: Support Vector Machines in Automated Emotion Classification. Churchill College, London (2003)
26. Rizon, M., Karthigayan, M., Yaacob, S., Nagarajan, R.: Japanese face emotions classification using lip features. In: Geometric Modeling and Imaging, 2007 (GMAI'07), pp. 140–144 (2007)
27. Rottenberg, J., Gross, J.: Emotion and emotion regulation: a map for psychotherapy researchers. Clin. Psychol. Sci. Pract. **14**(4), 323–328 (2007)
28. Shackleton, M., Welsh, W.: Classification of facial features for recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991 (Proceedings CVPR'91), pp. 573–579 (1991)
29. Shih, F., Chuang, C., Wang, P.: Performance comparisons of facial expression recognition in jaffe database. Int. J. Pattern Recogn. Artif. Intell. **22**(3), 445–460 (2008)
30. Sloan, D., Kring, A.: Measuring changes in emotion during psychotherapy: conceptual and methodological issues. Clin. Psychol. Sci. Pract. **14**(4), 307–322 (2007)
31. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
32. Suwa, M., Sugie, N., Fujimora, K.: A preliminary note on pattern recognition of human emotional expression. In: International Joint Conference on Pattern Recognition, pp. 408–410 (1978)
33. Terzopoulos, D., Waters, K.: Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Trans. Pattern Anal. Mach. Intell. **15**(6), 569–579 (1993)
34. Wang, H., Huang, H., Ding, C.: Discriminant Laplacian Embedding. In: Twenty-Fourth AAAI Conference on Artificial Intelligence (2010)
35. Wang, H., Huang, H., Hu, Y., Anderson, M., Rollins, P., Makedon, F.: Emotion detection via discriminative kernel method. In: Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, pp. 1–7. ACM (2010)
36. Yacoob, Y., Davis, L.: Computing spatio-temporal representations of human faces. In: IEEE Computer Society Conference On Computer Vision And Pattern Recognition, pp. 70–70. Citeseer (1994)
37. Zeng, Z., Hu, Y., Fu, Y., Huang, T., Roisman, G., Wen, Z.: Audio-visual emotion recognition in adult attachment interview. In: Proceedings of the 8th International Conference on Multimodal Interfaces, p. 145. ACM (2006)
38. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. (CSUR) **35**(4), 399–458 (2003)