PAN ARCHIVES OF ACOUSTICS
Vol. **38**, No. 4, pp. 465–470 (2013)

DOI: 10.2478/aoa-2013-0055

Speech Emotion Recognition Based on Sparse Representation

Jingjie YAN[(1)], Xiaolan WANG[(2)], Weiyi GU[(2)], LiLi MA[(2)]

[(1)] *The School of Information Science and Engineering, Southeast University*
No.2 Sipailou, Nanjing, P.R.China; e-mail: yanjingjie2012@gmail.com

[(2)] *Research Center for Learning Science, Southeast University*
No.2 Sipailou, Nanjing, P.R.China

(*received November 29, 2012; accepted April 16, 2013*)

Speech emotion recognition is deemed to be a meaningful and intractable issue among a number of domains comprising sentiment analysis, computer science, pedagogy, and so on. In this study, we investigate speech emotion recognition based on sparse partial least squares regression (SPLSR) approach in depth. We make use of the sparse partial least squares regression method to implement the feature selection and dimensionality reduction on the whole acquired speech emotion features. By the means of exploiting the SPLSR method, the component parts of those redundant and meaningless speech emotion features are lessened to zero while those serviceable and informative speech emotion features are maintained and selected to the following classification step. A number of tests on Berlin database reveal that the recognition rate of the SPLSR method can reach up to 79.23% and is superior to other compared dimensionality reduction methods.

Keywords: speech emotion recognition, sparse partial least squares regression (SPLSR), feature selection and dimensionality reduction.

1. Introduction

It is common knowledge that accurate and dependable speech emotion recognition can provide with quite momentous affect for achieving or consummating intelligent, efficient, and reliable human-computer interaction. Consequently, a great number of investigators centralize their attention and energy on this intractable and unmanageable issue, and a great quantity of tremendous advance has been gained to probe, analyze, and exploit speech emotion more informative, meaningful, convenient, and efficient during the past decades (Ayadi *et al.*, 2011; Cen *et al.*, 2008; Ser *et al.*, 2008).

How to pick up the most beneficial and serviceable speech emotion features which can accessibly and accurately represent the speech emotion information is the pivotal problem for speech emotion recognition (Chen *et al.*, 2012; Jin *et al.*, 2013; Ayadi *et al.*, 2011). Among all speech emotion features, prosodic feature and spectral feature are the most representative types of speech emotion features that are comprehensively employed in speech emotion recognition (Wu *et al.*, 2011; Jin *et al.*, 2013). The frequently adopted prosodic feature contains pitch, formants, energy, speed, and so on (Chen *et al.*, 2012; Ayadi *et al.*, 2011). The spectral feature is deemed to offer certain supplementary and various speech emotion features comparing with the prosodic feature and mel-frequency cepstral coefficients (MFCC), linear predictive cepstral coefficients (LPC) and log-frequency power coefficients (LFPC) are three most classic spectral features that are broadly adopted in a number of speech emotion recognition approaches (Wu *et al.*, 2011; Ayadi *et al.*, 2011; Jin *et al.*, 2013).

In the time of dealing with the above extracted speech emotion features, we may find out that the dimension of obtained speech emotion features can be as high as several hundred or more than one thousand, and each extracted speech emotion feature owns disparate effect on speech emotion recognition (Zhou *et al.*, 2012; Cen *et al.*, 2008; Jin *et al.*, 2013; Lai *et al.*, 2012). For instance, some features can promote and increase the recognition performance of speech emotion recognition, but some redundant and invalid features are harmful and unhelpful for final emotion classification.

Over the last several decades, plenty of investigators burgeon a mass of dimensionality reduction and feature selection approaches for working out those supervised and unsupervised studying issues (Bishop, 2006; Lai *et al.*, 2012; Jin *et al.*, 2013). Among those approaches, principle component analysis (PCA) (Turk, Pentlnd, 1991) and linear discriminant analysis (LDA) (Belhumeur, Hespanha, 1997; Chen *et al.*, 2000) are the two outstanding methods and they already have been triumphantly exploited to speech emotion recognition and other relevant research domains (Jin *et al.*, 2013; Bishop, 2006).

It is noted that PCA and LDA can only deal with and analyze one group of data, whereas canonical correlation analysis (CCA) (Hotelling, 1936) and partial least squares regression (PLSR) (Rosopal, Kamer, 2008) are investigated to deal with two groups of data at the same time (Parkhomenko *et al.*, 2009). Akin to PCA and LDA, CCA also has been addressed and applied in speech emotion recognition (Cen *et al.*, 2008). The application of PLSR is also comprehensive and efficient in various important spheres such as vehicle detection (Kembhavi *et al.*, 2011), neuroimaging (Krishnan *et al.*, 2011) and human detection (Schwartz *et al.*, 2009). But when making use of the PLSR method in speech emotion recognition, we will discover one main weakness of the PLSR method analogous to PCA, LDA, and CCA, and this weakness is that the weight vectors of PLSR are the linear combination of the whole extracted speech emotion features, therefore it is unable to clear away those redundant and insignificant speech emotion features (Cao *et al.*, 2008; Chun, Keles, 2010; McWilliams, Montana, 2010; Qiao *et al.*, 2009; Cai *et al.*, 2007; Lai *et al.*, 2012).

In the last few years, a number of sparse versions of PLSR approach have been established to resolve the above-mentioned shortcoming of PLSR method. Cao *et al.* (2008; 2009) presented a novel bilateral sparse partial least squares regression based on the SVD decomposition and sparse PCA approach (Shen, Huang, 2008) that can achieve data fusion and variable selection simultaneously when applied to biology. Similar to the work of Cao *et al.* (2008; 2009), a unilateral sparse partial least squares regression (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2011; Cao, Gall, 2011) based on unilateral sparse SVD approach has been developed. Moreover, McWilliams and Montana (2010) apply this unilateral sparse partial least squares regression method for increment learning task. Chun and Keles (2010) propose a new sparse partial least squares approach in which the objective function is rewritten as SPCA's optimization formula (Zou *et al.*, 2006). Based on the work of Chun and Keles (2010), Chung and Keles (2010) have developed two classification-based methods including SPLS discriminant analysis (SPLSDA) and sparse generalized PLS (SGPLS), and then applied them to conduct variable selection on high-dimension datasets. Moreover, Huang *et al.* (2012) utilized the SPLSR method to conduct intelligibility detection tests. In this paper, we exploit the unilateral sparse partial least squares regression method (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2011; Cao, Gall, 2011) to implement the feature selection and dimension reduction of obtained speech emotion features in speech emotion recognition.

In this paper, we investigate speech emotion recognition based on sparse partial least squares regression (SPLSR) approach. The whole of acquired speech emotion features have availably implemented the feature selection and dimensionality reduction with the sparse partial least squares regression method whose basic idea is to receive sparse projections by means of calculating a sparse singular value decomposition issue (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2008; Shen, Huang, 2008). By introducing the properties of sparsity on the whole acquired speech emotion features with SPLSR method, the component parts of redundant and meaningless speech emotion features are lessened to zero while those serviceable and informative speech emotion features are maintained for the next classification step.

We organize the remaining part of this paper as follows. We roughly provide the description of the PLSR method in Sec. 2. Section 3 introduces the SPLSR approach in detail and its specific algorithm. The speech emotion classification via SPLSR method is illustrated in Sec. 4. Section 5 introduces the adopted speech emotion databases and describes a number of experiments. Ultimately, Sec. 6 concludes the paper and gives some discussion on the future work.

2. Partial Least Squares Regression

Consider a pair of sample data matrices $X \in R^{N\times n_x}$ and $Y \in R^{N\times n_y}$ which indicate the extracted speech emotion feature matrices and its speech emotion category feature matrices respectively, where N expresses the number of samples. The fundamental purpose of PLSR method is exploring a cluster of vectors ϕ_r and ζ_r which are achieved by maximizing the under optimization formulation in the form of alleged latent vectors $X\phi_r$ and $Y\zeta_r$ (Cao *et al.*, 2008; Rosopal, Kamer, 2008; Gu, 2010; Ma, 2010; Yan *et al.*, 2013)

$$\{\phi_r;\zeta_r\} = \arg \max_{\phi^{\mathrm{T}}\phi=\zeta^{\mathrm{T}}\zeta=1} \mathrm{cov}(X\phi_r, Y\zeta_r). \tag{1}$$

As we all know, X and Y can be resolved into the following pattern (Rosopal, Kamer, 2008; Cao *et al.*, 2008)

$$\begin{aligned} X &= WP_x^{\mathrm{T}} + E_x, \\ Y &= WP_y^{\mathrm{T}} + E_y, \end{aligned} \tag{2}$$

where P_x and P_y are matrices of coefficient which are expressed as $p_x = X^{\mathrm{T}}w/w^{\mathrm{T}}w$ and $p_y = Y^{\mathrm{T}}w/w^{\mathrm{T}}w$, respectively. Besides, E_x and E_y are the corresponding residuals matrices of X and Y, respectively (Rosopal, Kamer, 2008; Cao *et al.*, 2008).

Suppose $R_{XY} = X^{\mathrm{T}}Y$, $R_{XY} = R_{YX}^{\mathrm{T}}$. Then, the formula (1) can be changed into the following optimization problem (Gu, 2010; Ma, 2010) (we seek the first cluster of vectors ϕ_r and ζ_r here):

$$\arg\max_{\phi,\zeta} \phi^{\mathrm{T}} R_{XY} \zeta, \tag{3}$$

subject to $\phi^{\mathrm{T}}\phi = 1$ and $\zeta^{\mathrm{T}}\zeta = 1$.

Then we can receive the Lagrangian of formula (3) as the following pattern (Gu, 2010; Ma, 2010; Zhou *et al.*, 2012):

$$\begin{aligned} L(\phi,\zeta,\lambda,\mu) &= \phi^{\mathrm{T}} R_{XY}\zeta \\ &\quad -\frac{\lambda}{2}(\phi^{\mathrm{T}}\phi - 1) - \frac{\mu}{2}(\zeta^{\mathrm{T}}\zeta - 1). \end{aligned} \tag{4}$$

Further, the following two equations can be obtained by calculating the partial derivatives of $L(\phi,\zeta,\lambda,\mu)$ with respect to ϕ and ζ (Ma, 2010):

$$\begin{aligned} R_{XY}\zeta &= \lambda\phi, \\ R_{YX}\phi &= \mu\zeta. \end{aligned} \tag{5}$$

Ultimately, we are capable of receiving the first desired cluster of projection vectors ϕ and ζ by figuring up the next eigenvalue equations of (6) and (7), respectively:

$$R_{XY}R_{YX}\phi = \lambda\mu\phi, \tag{6}$$

$$R_{YX}R_{XY}\zeta = \lambda\mu\zeta. \tag{7}$$

3. Sparse Partial Least Squares Regression

On the basis of the antecedent discussion and analysis, we can note that the PLSR approach is not capable of carrying out the feature selection on the whole extracted speech emotion features. Moreover, it should be noted that X and Y signify the whole extracted speech emotion features and its speech emotion category feature respectively, and the speech emotion category feature contains only category information. Therefore we only need to carry out the feature selection on the whole extracted speech emotion feature X. Consequently in the following section, we will introduce the unilateral sparse partial least squares regression method (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2011; Cao, Gall, 2011) and then employ it to carry out the feature selection on X in the form of solving sparse projection ϕ. The conventional PLSR method can be solved by means of PLSR-SVD pattern in the light of the literature of Cao *et al.* (2008). In the PLSR-SVD form, the matrix $R_{XY} = X^{\mathrm{T}}Y$ of the conventional PLSR method can be expressed in the following singular value decomposition (SVD) form (Cao *et al.*, 2008; McWilliams, Montana, 2010; Shen, Huang, 2008; Yan *et al.*, 2012):

$$R_{XY} = X^{\mathrm{T}}Y = \sum_{t=1}^{h} d_t u_t v_t^{\mathrm{T}}. \tag{8}$$

It is noted that the eigenvalues of $R_{XY}^{\mathrm{T}}R_{XY}$ and $R_{XY}R_{XY}^{\mathrm{T}}$ are $d_1, d_2, \ldots, d_h$, and the eigenvectors of $R_{XY}^{\mathrm{T}}R_{XY}$ and $R_{XY}R_{XY}^{\mathrm{T}}$ are equivalent to u_i and v_i, respectively. In accordance with the nature of the above PLSR-SVD approach, the pair of desired vectors α and β of the conventional PLSR are just equivalent to u_i and v_i in the above SVD form, and it indicates $\phi = u_i$ and $\zeta = v_i$ in other words (Cao *et al.*, 2008; McWilliams, Montana, 2010; Ma, 2010; Yan *et al.*, 2013). Consequently, the cluster of vectors ϕ and ζ that are received by the conventional PLSR solving approach can be transformed to calculate the cluster of vectors of u_i and v_i in the above PLSR-SVD form.

On the basis of the work of Shen and Huang (2008) and Cao *et al.* (2008), a unilateral sparse partial least squares regression method (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2011; Cao, Gall, 2011) is developed to obtain only one cluster of sparse projection and is different from the method of Cao *et al.* (2008) in which two sets of sparse projections can be obtained. The optimization formulation of unilateral SPLSR is adapted as the following pattern by introducing the lasso penalty on the vector u (McWilliams, Montana, 2010; Ma, 2010; Cao *et al.*, 2011; Cao, Gall, 2011):

$$\min_{u,v} \left\| R_{XY} - uv^{\mathrm{T}} \right\|_F^2 + \lambda_u \sum_{j=1}^{p} |u_j|, \tag{9}$$

where $\lambda_u \sum_{j=1}^{p} |u_j|$ stands for the lasso function, and $\lambda_u \geq 0$ signifies a positive value for deciding the sparse degree of u.

For the sake of calculating the optimization problem of (9) efficiently, McWilliams and Montana (2010) have developed an iterative algorithm with respect to solving u and v. The process of this iterative algorithm consists in that u is achieved by optimizing (9) through fixing v, and then v is achieved by optimizing (9) through fixing u (Zhou *et al.*, 2012).

By means of repeating the above iterative process, the desired optimal projection v and sparse projection u are acquired finally from the following Eqs. (10) and (11), respectively, after the iterative algorithm achieves convergence (McWilliams, Montana, 2010; Ma, 2010):

$$v^* = \frac{R_{XY}^{\mathrm{T}}u}{\left\| R_{XY}^{\mathrm{T}}u \right\|}, \tag{10}$$

$$u^* = \operatorname{sign}(R_{XY}v)(|R_{XY}v| - \lambda_u/2)_+. \tag{11}$$

For details of the above solving procedure of the unilateral SPLSR, please see the paper of McWilliams and Montana (2010). At last, we can project the whole extracted speech emotion feature X onto the above sparse projection u (u is equivalent to ϕ) received by the unilateral SPLSR to implement feature selection.

4. Emotion Recognition via Sparse Partial Least Squares Regression

Assuming that a cluster of the sparse projection matrix $\phi_X = (\phi_x^1, \phi_x^2, ...\phi_x^h)$ are received conclusively via the above unilateral SPLSR method, where h denotes the number of the above extracted sparse projection matrix ϕ_X, then on the basis of the Eq. (2) we can get the following two regression equations (Gu, 2010; Ma, 2010; McWilliams, Montana, 2010):

$$\begin{aligned} X &= w_1 P_{x1}^{\mathrm{T}} + w_2 P_{x2}^{\mathrm{T}} + \ldots + w_h P_{xh}^{\mathrm{T}} \\ &\quad + E_x^h = W P_x^{\mathrm{T}} + E_x^h, \\ Y &= w_1 P_{y1}^{\mathrm{T}} + w_2 P_{y2}^{\mathrm{T}} + \ldots + w_y P_{yh}^{\mathrm{T}} \\ &\quad + E_y^h = W P_y^{\mathrm{T}} + E_y^h, \end{aligned} \tag{12}$$

where $W = (w_1, w_2, ..., w_h)$, $p_X = (p_{X1}, p_{X2}, ...p_{Xh})$, $p_Y = (p_{Y1}, p_{Y2}, ...p_{Yh})$.

In the light of (Manne, 1987; Gu, 2010; Ma, 2010), we obtain the next equation by simple approximately calculation:

$$W = X\phi_X (p_X^{\mathrm{T}} \phi_X)^{-1}. \tag{13}$$

Form Eqs. (12) and (13), we are able to receive the following formula (Gu, 2010; Ma, 2010)

$$Y = W P_y^{\mathrm{T}} + E_y^h = X\phi_X (p_X^{\mathrm{T}} \phi_X)^{-1} P_y^{\mathrm{T}} + E_y^h. \tag{14}$$

At last, if providing a test speech emotion feature sample X_{test}, we are able to calculate the homologous speech emotion category feature Y_{test} in the light of the following formula (Gu, 2010; Ma, 2010; McWilliams, Montana, 2010):

$$Y_{\text{test}} = X_{\text{test}} \phi_X (p_X^{\mathrm{T}} \phi_X)^{-1} P_y^{\mathrm{T}}. \tag{15}$$

5. Experiments

In the following section, we will simply introduce the adopted speech emotion database, the detailed experiment design and show results of the experimental test. In this study, we assess the property of our speech emotion recognition approach based on the SPLSR algorithm by carrying out certain tests on the Berlin database (Burkhardt *et al.*, 2005; Zheng *et al.*, 2012; Jin *et al.*, 2013; Gu, 2010). In our experiment, from a number of samples representing different speech emotion categories, we select and utilize a subset of the Berlin dataset which contains 260 speech samples and five speech emotion categories including anger, boredom, fear, joy, and sadness.

In the speech emotion feature extraction procedure, two sorts of speech emotion features consisting of prosodic features and spectral features are picked up in our experiment (Gu, 2010; Wu *et al.*, 2011; Zheng *et al.*, 2012; Jin *et al.*, 2013). The details of extracted speech emotion feature in our test included pitch, formant frequency, the logarithmic form of energy, and so on (Gu, 2010; Zheng *et al.*, 2012).

Apart from the SPLSR method, we also conduct other classic dimensionality reduction methods for speech emotion recognition, i.e. the principal component analysis (PCA) method, the linear discriminant analysis (LDA) method, the gaussian mixture model (GMM) method, and the partial least squares regression (PLSR) method. For PCA and LDA, we adopt the K-nearest neighbor (KNN) classifier to classify five different emotion categories, whereas PLSR and SPLSR exploit the classification method introduced in Sec. 4. Besides, in our experiments, we employ the thirteen-fold cross-validation strategy to the above five methods (Gu, 2010; Zheng *et al.*, 2012).

The average recognition rates of the above five methods are exhibited in Table 1. It is noted that the disparate number of the reduced dimension will produce a certain influence concerning the final property, therefore we also implement the tests with the disparate dimension. Figure 1 displays the recognition rates of three methods with disparate reduced dimension.

Table 1. The average recognition rate of each method.

Method	Recognition rate
PCA	64.23
LDA	75.00
GMM	69.62
PLSR	78.46
SPLSR	79.23

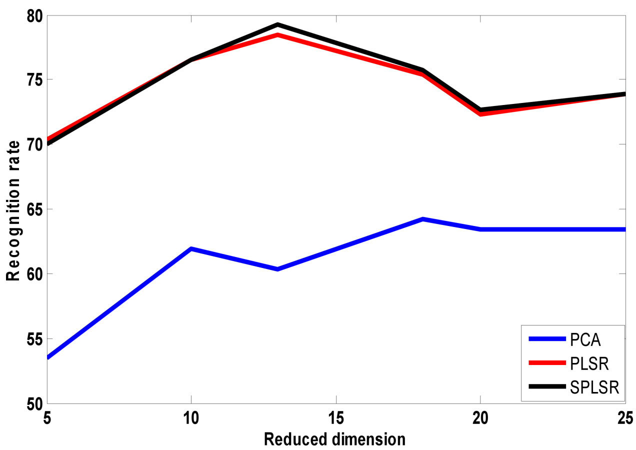

Fig. 1. Recognition rates of each method for different number of reduced dimensions.

To distinctly display the test results of different emotion category, the confusion matrices of PLSR method and SPLSR method are shown in Fig. 2 and Fig. 3, respectively. Moreover, we also implement the experiment of SPLSR method with different values of the sparse parameter. Table 2 shows the recognition rates of SPLSR method versus different values of the sparse parameter.

Fig. 2. Confusion matrix for the PLSR method.

Fig. 3. Confusion matrix for the SPLSR method.

Table 2. Recognition rates of SPLSR for the different values of sparse parameter.

Sparse parameter	Recognition rate
0.1	78.85
0.08	78.85
0.05	79.23
0.01	78.46

In accordance with the above test results, we can note that the recognition rate of the SPLSR method can reach up to 79.23% and it is superior to other compared methods such as PCA, LDA, GMM, and PLSR. This test result indicates that the SPLSR method may provide more serviceable and informative speech emotion information than those compared methods. Moreover, in line with Table 2, we can see that the sparse parameter λ can produce the effect on the recognition rate of the SPLSR approach to some degree and the best recognition rate of SPLSR is achieved when the sparse parameter λ is 0.05.

6. Conclusion and discussion

In this paper, we study speech emotion recognition based on the partial least squares regression (SPLSR) approach. We make use of the sparse partial least squares regression method to implement the feature selection and dimensionality reduction on the whole acquired speech emotion features. Moreover, a number of experimental tests on the Berlin database certify the validity and meaning of the speech emotion recognition approach based on the SPLSR method. In the future study, we can import those sparse kernel approaches onto speech emotion recognition which may allow to obtain nonlinear discriminative speech emotion information to some degree.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61231002 and Grant 61073137, the Natural Science Foundation of Jiangsu Province under Grant BK20130020, and the Program for Distinguished Talents of Six Domains in Jiangsu Province of China under Grant 2010-DZ088.

References

1. Ayadi M.E., Kamel M.S., Karray F. (2011), *Survey on speech emotion recognition: features, classification schemes, and databases*, Pattern Recognition, **44**, 572–587.
2. Belhumeur P.N., Hespanha P. (1997), *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*, IEEE Trans. Pattern Anal. Mach. Intell, **19**, 711–720.
3. Baek J.S., Kim M. (2004), *Face Recognition using paritial least squares components*, Pattern Recognition, **37**, 1303–1306.
4. Burkhardt F. *et al.* (2005), *A database of german emotional speech*, Interspeech, 1517–1520.
5. Bishop C. (2006), *Pattern recognition and machine learning*, Springer, United States of America.
6. Chen L.F., Mark H.Y., Ko M.T. (2000), *A new LDA-based face recognition system which can solve the small sample size problem*, Pattern Recognition, **33**, 1713–1726.
7. Cen L., Ser W., Liang Z. (2008), *Speech emotion recognition using canonical correlation analysis and probabilistic neural network*, International Conference on Machine Learning and Applications, 859–862.
8. Cai D., He X.F., Han J.W. (2007), *Spectral Regression: A Unified Approach for Sparse Subspace Learning*, IEEE International Conference on Data Mining, 73–82.

9. Cao K.A., Rossouwy D., Grani C.R., Besse P. (2008), *A sparse PLS for variable selection when integrating omics data*, Statistical Applications in Genetics and Molecular Biology, **7**, 35–45.

10. Cao K.A., Martin P.G.P., Robert-Granié C., Besse P. (2009), *Sparse canonical methods for biological data integration: application to a cross-platform study*, BMC Bioinformatics, **10**.

11. Cao K.A., Boitard S., C.R., Besse P. (2011), *Sparse PLS Discriminant Analysis: biologically relevant feature selection and graphical displays for multiclass problems*, Technical report, University of Queensland.

12. Cao K.A., Gall G.L. (2011), *Integration and variable selection of 'omics' data sets with PLS: a survey*, Journal de la Société Française de Statistique, **152**, 77–96.

13. Chen L.J., Mao X., Xue Y.L, Cheng L.L. (2012), *Speech emotion recognition: features and classification models*, Digital Signal Processing, **2**, 154–1160.

14. Chun H., Keles S. (2010), *Sparse partial least squares regression for simultaneous dimension reduction and variable selection*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), **72**, 3–25.

15. Chung D., Keles S. (2010), *Sparse Partial Least Squares Classification for High Dimensional Data*, Statistical Applications in Genetics and Molecular Biology, **9**.

16. Gu W.Y. (2010), *A Study on Speech Emotion Recognition*, master thesis, Research Center for Learning Science, Southeast University (in chinese).

17. Hotelling H. (1936), *Relations between two sets of variates*, Biometrika, **28**, 312–377.

18. Huang D.Y., Zhu Y.W., Wu D.J., Yu R.S. (2012), *Detecting Intelligibility by Linear Dimensionality Reduction and Normalized Voice Quality Hierarchical Features*, Interspeech, 546–549.

19. Jin Y., Zheng W.M., Zhao L., Yan J.J. (2013), *Generalized maximal margin discriminant analysis for speech emotion recognition*, Electrical review.

20. Kembhavi A., Harwood D., Davis L.S. (2011), *Vehicle detection using partial least squares*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33**, 1250–1265.

21. Krishnan A., Williams L.J., McIntosh A.R., Abdi H. (2011), *Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review*, **56**, 455–475.

22. Lai Z.H., Wan M.H., Jin Z., Yang J. (2012), *Sparse two-dimensional local discriminant projections for feature extraction*, Neurocomputing, **74**, 629–637.

23. Long N., Gianola D., Rosa G.J.M., Weigel K.A. (2011), *Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins*, Journal of Animal Breeding and Genetics, **128**, 147–257.

24. Ma L.L. (2010), *The Research and Application of Facial Expression Recognition based on Semantic Feature*, master thesis, Research Center for Learning Science, Southeast University (in chinese).

25. Manne R. (1987), *Analysis of Two Partial-Least-Squares Algorithms for Multivariate*, Calibration Chemometrics and Intelligent Laboratory Systems, **2**, 187–197.

26. McWilliams B., Montana G. (2010), *Sparse partial least squares regression for on-Line variable selection with multivariate data streams*, Statistical Analysis and Data Mining, **3**, 170–193.

27. Parkhomenko E., Tritchler D., Beyene J. (2009), *Sparse canonical correlation analysis with application to genomic data integration*, Statistical Applications in Genetics and Molecular Biology, **8**, 55–61.

28. Qiao Z., Zhou L., Huang J.Z. (2009), *Sparse linear discriminant analysis with application to high dimensional low sample size data*, International Journal of Applied Mathematics, **39**, 48–60.

29. Rosopal R., Kamer N. (2008), *Overview and recent advances in partial least squares. subspace latent structure and feature selection*, Statistical and Optimization Perspectives Workshop, **3940**, 34–51.

30. Shen H., Huang J. (2008), *Sparse principal component analysis via regularized low rank matrix approximation*, Journal of Multivariate Analysis, **99**, 1015–1034.

31. Ser W., Cen L., Yu Z.L. (2008), *A Hybrid PNN-GMM classification scheme for speech emotion recognition*, International Conference on Pattern Recognition, 1–4.

32. Schwartz W.R., Kembhavi A., Harwood D., Davis L.S. (2009), *Human detection using partial least squares analysis*, IEEE International Conference on Computer Vision, 24–31.

33. Turk M., Pentland A.P. (1991), *Face recognition using eigenfaces*, IEEE Conference on Computer Vision and Pattern Recognition, 586–591.

34. Wu S.Q., Tiago H., Falk, Chan W.Y. (2011), *Automatic speech emotion recognition using modulation spectral features*, Speech Communication, **53**, 768–785.

35. Yan J.J., Zheng W.M., Zhou X.Y., Zhao Zhi.J. (2012), *Sparse Two-dimensional Canonical Correlation Analysis via Low Rank Matrix Approximation for Feature Extraction*, IEEE Signal Processing Letters, **19**, 51–54.

36. Yan J.J., Zheng W.M., Xin M.H., Qiu Wei (2013), *Bimodal Emotion Recognition based on Body Gesture and Facial Expression*, Journal of Image and Graphics, **18**, 1101–1106 (in chinese).

37. Zou H., Hastie T., Tibshirani R. (2006), *Sparse principal component analysis,* Journal of Computational and Graphical Statistics, **15**, 265–286.

38. Zhou X.Y., Zheng W.M., Xin M.H. (2012), *Improving CCA via spectral components selection for facial expression recognition*, IEEE International Symposium on Circuits and Systems (ISCAS), 1696–1699.

39. Zheng W.M., Zhou X.Y. (2012), *Speech Emotion Recognition Based on Kernel Reduced-rank Regression*, International Conference on Pattern Recognition (ICPR 2012), 1972–1976.