AN AUTOMATIC FACIAL EXPRESSION RECOGNITION APPROACH BASED ON CONFUSION-CROSSED SUPPORT VECTOR MACHINE TREE

Qinzhen Xu[1], Pinzheng Zhang[2], Wenjiang Pei[1], Luxi Yang[1], Zhenya He[1]*

[1]School of Information Science and Engineering, Southeast University, Nanjing, 210096, China
[*]summer@seu.edu.cn
[2]School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China
luckzpz@seu.edu.cn

ABSTRACT

Automatic facial expression recognition is the kernel part of emotional information processing. This paper dedicates to develop an automatic facial expression recognition approach based on confusion-crossed support vector machine tree (CSVMT) to improve recognition accuracy and robustness. After the Pseudo-Zernike moment features were extracted, they were used to train a CSVMT for automatic recognition. The structure of CSVMT enables the model to divide the facial recognition problem into sub-problems according to the teacher signals, so that it can solve the sub-problems in decreased complexity in different tree levels. In the training phase, those sub-samples assigned to two internal sibling nodes perform decreasing confusion cross, thus, the generalization ability of CSVMT for recognition of facial expression is enhanced. The compared results on Cohn-Kanade facial expression database also show that the proposed approach appeared higher recognition accuracy and robustness than other approaches.

***Index Terms*—**Artificial intelligence, Image recognition, Feature extraction, Image classification, Face recognition.

1. INTRODUCTION

Human facial expression contains important individual emotional and psychic information for human computer interface. Automatic Facial expression recognition has received great attention in many research fields such as emotion analysis, psychology research, image understanding, image retrieval, with the development of computer technique and its popular application [1]. Ekman and Friesen defined six basic emotions: happiness, sadness, fear, disgust, surprise, and anger (See fig.1) [2]. Most of the current automatic facial expression recognition systems are founded on the psychologic hypothesis of the six basic facial expressions.

Support vector machine (SVM) based approaches have been widely applied in pattern recognition and function fitting. The predominant performance of SVM has been studied and validated in both theory and experiment [3]. Currently, we addressed the problems associated with complex pattern recognition and presented a confusion-crossed Support Vector Machine tree (CSVMT) [4]. A CSVMT is a binary decision tree with SVMs embedded in internal nodes. Those patterns assigned to two internal sibling nodes perform confusion cross. It is developed to achieve a better performance for complex distribution problems with lower dependence on the two parameters of SVM and better robustness on unbalanced classification problems. The experimental results demonstrated with some typical complex classification problems showed that the performance of the proposed method is distinctly enhanced compared with SVMT introduced in [5] and single binary SVM. In this paper we introduce CSVMT in facial expression recognition phase.

(a)Surprise (b) Sadness (c) Fear

(d)Anger (e) Disgust (f) Happiness

Fig. 1. Six basic emotions

The rest of the paper is organized as follows. The Pseudo-Zernike moments based facial feature extraction approach is illustrated in section 2. CSVMT is introduced in section 3. Experimental results and discussions are described in section 4. Eventualy, conclusions are made in section 5.

2. PSEUDO-ZERNIKE MOMENTS BASED FACIAL FEATURE EXTRACTION

Facial expression recognition deals with the classification of facial motion and facial feature deformation into abstract classes [6], and thus facial feature extraction and learning of these features are of great importance for automatic facial expression. A number of developed feature extraction approaches are grouped into two types: shape-based features and image-based features. The shape-based approaches describe the facial features on the variance of parameters of face models [7]. These approaches usually require robust face tracking, hence high computation cost [1]. The imaged-based approaches express features on the pixel intensities of the whole face image or certain regions of the face image. These approaches are popular researched and applied. Here, we introduce Pseudo-Zernike moments to extract imaged-based features for its denoising ability. Moments are used to depict the distribution of random variables in statistics. The images can be treated as two-dimensional or three-dimensional density distribution functions. In this manner, moments are introduced in image analysis.

Pseudo-Zernike moment is defined by the orthogonal complex-value polynomials in the unit circle. The Pseudo-Zernike moment of two-dimensional image $f(\rho,\theta)$ is defined as [8]:

$$C_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\theta)W_{nm}^*(\rho,\theta)\rho\, d\rho\, d\theta \qquad (1)$$

where $|m| \le n$ are the orders of Pseudo-Zernike polynomial. Pseudo-Zernike moment at small values of m and n represents the global image information. Correspondingly, Pseudo-Zernike moment denotes the detailed image information when m and n are large. $W_{nm}^*(\rho,\theta)$ is the conjugate of basis function of Pseudo-Zernike moment. It is defined as:

$$W_{nm}(\rho,\theta) = S_{nm}(\rho)e^{jm\theta} \qquad (2)$$

where $S_{nm}(\rho)$ is Pseudo-Zernike polynomial:

$$S_{nm}(\rho) = \sum_{s=0}^{n-|m|}\frac{(-1)^s(2n-s+1)!\rho^{n-s}}{s!(n-|m|-s)!(n-|m|-s+1)!} \qquad (3)$$

To improve the computation efficiency, the recursive formulas of Pseudo-Zernike moment are applied for fast computation of discretized Pseudo-Zernike moment values [9].

3. CONFUSION-CROSSED SUPPORT VECTOR MACHINE TREE

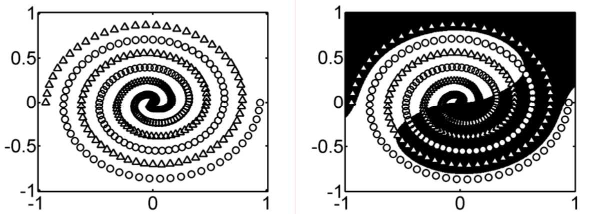
The construction of a CSVMT model is actually a process of applying the divide-and-conquer idea to solve tough problems. Let SVM_j and SVM_{j+1} be the two internal sibling nodes derived from the common parent node SVM_p, and S_p be the training set assigned to node p. Consider the spacing variable on the trained SVM at node $p: \gamma_p(x) = \sum_{i=l_1}^{l_N}\alpha_i y_i K(x_i,x)+b$, where α_i are the Lagrange multipliers, x_i , $i \in \{l_1,\cdots l_N\} \subseteq \{1,\cdots,l\}$ are the support vectors. The confused set is defined as those examples, which are close to the decision hyperplane and accordingly more likely to be misclassified as depicted by $S_C = \{x \mid x \in S_p, |\gamma_p(x)| \le C_0\overline{\gamma}_p\}$, where $\overline{\gamma}_p = (1/|S_p|)$, $\sum_{x_i \in S_p}\gamma_p(x_i)$ and $0 < C_0 < 1$ is a small valued positive number. Instead of partitioning the training examples to node j and $j+1$ by the symbol function $\mathrm{sign}(\gamma_p)$, the confusion cross process is presented to implement decreasing cross between the two training subsets partitioned by $\mathrm{sign}(\gamma_p)$ to keep those confused patterns in both training subsets of the two internal sibling nodes. The reassignment process is controlled by confusion cross factor

$$C_{\gamma_p,m} = \rho_0 \exp(-\lambda\cdot m)\overline{\gamma}_p \qquad (4)$$

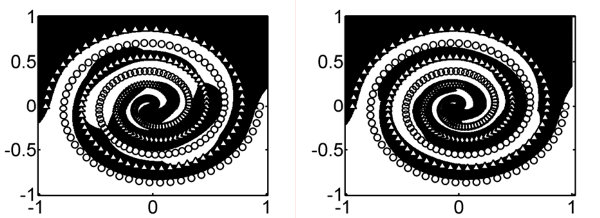
where m is the node level of internal nodes j and $j+1$, $\rho_0 \in (0,1)$ is the initial confusion cross rate, and λ controls the convergence of the cross and terminates the cross process at the deep tree levels. When confusion cross is performed, the subset of training examples S_p reassigned to those two child nodes j and $j+1$ are $S_j = \{x \mid x \in S_p, \gamma_p(x) \ge -C_{\gamma_p,m}\}$ and $S_{j+1} = \{x \mid x \in S_p, \gamma_p(x) \le C_{\gamma_p,m}\}$. In this way, the set of crossed examples $S_{c_m} = S_j \cap S_{j+1} = \{x \mid x \in S_p, |\gamma_p(x)| \le C_{\gamma_p,m}\}$, which are close to the decision hyperplane and accordingly more likely to be misclassified, are kept in both two child nodes for further construction of decision hyperplane at a fine node level, i.e., those confused examples are made a validation of their contribution to the fine decision hyperplane.

The property of the tree-structure approach allows the models to divide the problems in different levels according to the teacher signals constructed by the heuristic approach and then conquer the sub-problems with decreased complexity. Further, The accuracy of classification by an SVM with Guassian kernel $k(x_i,x_j) = \exp[-\|x_i - x_j\| /(2\sigma^2)]$ is dependent on the kernel width σ and the penalty parameter. Inappropriately selected values of these two parameters may cause overfitting or underfitting problems. Some approaches, such as cross-validation approach [10], were developed to solve this problem. CSVMT can achieve better performance for complex distribution problems with lower dependence on the two parameters.

Fig. 2. is an example on double-spiral problem composed of 352 patterns,176 in class 1 and 176 in class 2. Note that we didn't optimize the kernel width σ and the penalty parameter of the SVMs with Gaussian kernel in this experiment. The SVM parameter values were randomly set in the experiment, e.g., $\sigma = 1$ and penalty parameter $C = 3000$. The compared results showed that the CSVMT model achieved a competitive recognition rate with a much

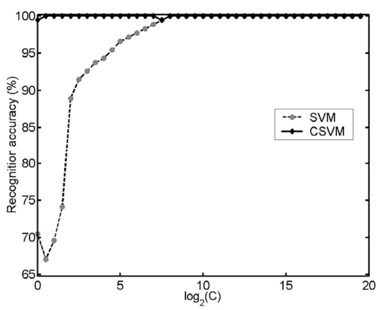


(a)Double-spiral problem (b) Decision surface of SVM

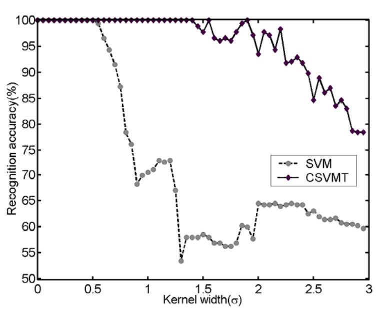


(c) Decision surface of SVMT (d) Decision surface of CSVMT

Fig. 2. The compared decision surface of double-spiral problem



(a) Dependency on the penalty parameter *C* under fixed kernel width *σ*



(b) Dependency on the kernel width *σ*under fixed penalty parameter *C*

Fig. 3. The compared results between SVM and CSVMT on dependency of the two parameters

simpler separation hyperplane than the other compared methods. Fig. 3 shows CSVMT has lower dependency on the two parameters than single SVM model.

4. EXPERIMENTAL RESULTS

In this section we reported the experimental results on Cohn-Kanade facial expression database[11]. It includes facial expressions of age 18 to 30 of different race. The data has been referenced many times in different facial expression research work. 66 facial features were extracted for the order $n \leq 11$ of Pseudo-Zernike polynomial. 1427 facial expression patterns were included in the experiment, i.e., 460 for surprise, 464 for happiness, 156 for sadness, 127 for anger, 126 for disgust and 94 for fear. 1/2 of the patterns were for training and the rest for testing. The initial confusion cross rate $\rho_0 = 0.8$, and $\lambda = 0.3$. The numerical results were the average of 8 runs of the recognition process. The test recognition accuracy of different recognition approaches trained on Pseudo-Zernike moment facial features were listed in tab. 1. The compared approaches included three SVM modes studied in Hsu and Lin's work[12] (i.e. DAGSVM, 1-V-1 SVM and 1-V-R SVM), SVMT, linear discriminant analysis, and k-nearest neighbors.

It was observed in the experiments that the training recognition accuracy of DAGSVM, 1-V-1 SVM, 1-V-R SVM, SVMT, and CSVMT reached 100.0%. Tab. 1 showed that the CSVMT based facial expression recognition approach achieved better test recognition accuracy than other approaches in total. Further, the proposed approach trained on the unbalanced training samples reached not only high recognition accuracy for large sample facial expressions, but also better accuracy than other approaches for small sample facial expressions, such as sadness, anger, disgust and fear. It indicated that the proposed approach might achieve better generalization ability and higher recognition robustness on unbalanced facial expression recognition problems.

We also compared the experimental results with that of other approaches adopted the Cohn-Kanade facial expression database on six basic emotions: AdaBoost approach was introduced in the feature extraction phase and a set of binary SVMs were used to recognize 6 basic facial expressions in [13], and its best recognition accuracy was 92.9%; A multistream hidden Markov based recognition model was presented in [1], which reached recognition accuracy of 93.66%; The Kanade-Tucas-Tomasi feature track approach was applied to compute the distances and angels between feature points in extraction phase and the expressions were classified by a rough-set based approach in [14], and the recognition accuracy reached 79.00%. The compared results showed that the proposed approach appeared higher recognition accuracy than the other approaches.

5. CONCLUSIONS

This paper dedicates to develop an automatic facial expression recognition approach based on confusion-crossed

Tab. 1. Facial expression recognition accuracy based on CSCMT

Methods / Accuracy	DAG SVM	1-V-1 SVM	1-V-R SVM	SVMT	LDA	KNN	CSVMT
Surprise(%)	98.86	99.42	97.68	97.39	90.82	92.46	97.39
Happiness(%)	99.65	100.00	98.85	100.00	91.77	95.61	99.86
Sadness(%)	90.00	88.32	64.10	90.17	75.06	72.27	**92.31**
Anger(%)	89.52	86.77	76.04	92.59	70.52	75.01	**94.18**
Disgust(%)	84.44	87.89	75.26	83.60	66.11	60.99	**92.06**
Fear(%)	82.97	85.82	71.01	82.98	76.02	66.45	**88.65**
Total(%)	95.01	95.36	88.76	94.86	84.45	85.21	**96.31**

support vector machine tree (CSVMT) to improve recognition accuracy and robustness. After the Pseudo-Zernike moment features were extracted, they were used to train a CSVMT for automatic recognition. The structure of CSVMT enables the model to divide the facial recognition problem into sub-problems according to the teacher signals, so that it can solve the sub-problems in decreased complexity in different tree levels. In the training phase, those sub-samples assigned to two internal sibling nodes perform decreasing confusion cross, thus, the generalization ability of CSVMT for recognition of facial expression is enhanced. The experiments are conducted on Cohn-Kanade facial expression database. Competitive recognition accuracy 96.31% is achieved by the CSVMT based approach. The compared results on Cohn-Kanade facial expression database also show that the proposed approach appeared higher recognition accuracy and robustness than other approaches.

6. ACKNOWLEDGEMENT

This work was partly supported by the Major Project of NSFC(60496310, 60672093, 60672095), a Foundation of Excellent Doctoral Dissertation of Southeast University (YBJJ0412) and a Scientific Research Starting Foundation of Southeast University.

7. REFERENCES

[1] P.S. Aleksic, and A.K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Transactions on Information Forensics and Security*, vol. 1, pp. 3-11, 2006.

[2] P. Ekman, and W.V. Friesen, *Facial Action Coding System: A technique for the measurement of facial movement*, Palo Alto, CA: Consulting Psychologists Press, 1978.

[3] V.N. Vapnik, "An Overview of Statistical Learning Theory", *IEEE Transactions on Neural Networks*, vol. 10, pp.988-999, 1999.

[4] Q.Z. Xu, A.G. Song, W.J. Pei, L.X. Yang, and Z.Y. He, "Tree-structured support vector machine with confusion cross for complex pattern recognition problems," *In: Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology*, 2005, pp.195-198

[5] S.N. Pang, D.J. Kim, and S.Y. Bang, "Face membership authentication using SVM classification tree generated by membership based LLE data partition," *IEEE transactions on Neural networks*, vol. 16, pp. 436-446, 2005.

[6] B. Fasel, and J. Luettin, "Automatic facial expression analysis:A survey," *Pattern Recognition*, vol. 36, pp.259-275, 2003.

[7] M. Pardàs, and A. Bonafonte, "Facial animation parameters extraction and expression detection using HMM," *Signal Processing: Image Communication*, vol. 17, pp. 675–688, 2002.

[8] C.H. Teh, and R.T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 496-513, 1988.

[9] C.W. Chong, P. Raveendran, and R. Mukundan, "An Efficient Algorithm for Fast Computation of Peseudo-Zernike Moments," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, pp. 1011-1023, 2003.

[10] G. M. Foody, and A. Mathur, "A Relative Evaluation of Multiclass Image Classification by Vector Machines," *IEEE Transactions on Geoscience and remote Sensing,* vol. 42, pp.. 1335-1343, 2004.

[11] T. Kanade, J.F. Cohn, and Y.I. Tian, "Comprehensive database for facial expression analysis," *Proc of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000.

[12] Hsu C.W., Lin C.J.. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Trans. Neural Networks*, vol. 13, pp. 415-525, 2002.

[13] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J.R. Movellan, Dynamics of facial expression extracted automatically from video, *Proc. Of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video,* Orlando, FL, 5, 2004, p.80

[14] C.C. Chibelushi, and F. Bourel, "Hierarchical multistream recognition of facial expressions," *IEE Proceedings on Vision, Image and Signal Processing,* vol. 151, pp. 307-313, 2004.