CAPTURE EXPRESSION-DEPENDENT AU RELATIONS FOR EXPRESSION RECOGNITION

Jun Wang[1], *Shangfei Wang*[1*], *Mengdi Huai*[1], *Chongliang Wu*[1], *Zhen Gao*[1], *Yue Liu*[1], *Qiang Ji*[2]

[1]Key Lab of Computing and Communication Software of Anhui Province, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, P.R.China, 230027.
junwong@mail.ustc.edu.cn, sfwang@ustc.edu.cn, mdhuai@mail.ustc.edu.cn, clwzkd@mail.ustc.edu.cn, gzgqllxh@mail.ustc.edu.cn, uknowly@mail.ustc.edu.cn
[2]Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA, 12180.
qji@ecse.rpi.edu

ABSTRACT

To date, there is only limited research that explicitly exploits the relationships among Action Units and expressions for facial expression recognition. In this paper, we propose an facial expression recognition method through modeling the expression-dependent AU relations. First, the incremental association Markov blanket algorithm is adopted to select crucial action units for a certain expression. Second, a Bayesian Network (BN) is constructed to capture the relationships between a certain expression and its crucial action units. Given the learned BNs and measurements of AUs and expression, we can then perform expression recognition within the BN through a probabilistic inference. Experimental results on the CK+ and MMI databases demonstrate the effectiveness and generalization ability of our method.

***Index Terms*—** expression recognition, BN structure, Markov blanket, AU

1. INTRODUCTION

Facial expression recognition has attracted increasing attention due to its wide applications in human-computer interaction [1]. There are two kinds of descriptors of expressions: expression category and Facial Action Units (AUs) [2]. The former describes facial behavior globally, and the latter represents facial muscle actions locally. Therefore, there exist close relations between AUs and expressions. For example, AU23 and AU24 must be present in the AU combination for anger expression [3]; AU4 is a component of negative expression, and AU4 must not appear in happy expression. This expression-dependent AU relation is important information for expression recognition. However, current research mainly recognize facial action units and expression individually, ignoring such relations.

Till now, there exist only a few works considering the relation between expressions and AUs to help expression recognition, or to jointly recognize AUs and expressions [4]. For example, Pantic and Rothkrantz [5] summarized the production rules of expressions from AUs using the AUs-coded descriptions of the six basic emotional expressions given by Ekman [2]. Since automatic AU recognition is error prone process, such rule based expression recognition method is very sensitive to false positives and misses among the AUs. Zhang and Ji [6] proposed dynamic Bayesian networks to model the relation of facial expressions to the complex combination of facial AUs, and temporal behaviors of facial expressions. Li et al [4] introduced a dynamic model to capture the relation among AUs, expressions, and facial feature points, and use the model to perform simultaneous AU and expression recognition, and facial feature tracking. In these two works, the links between expression nodes and AU nodes of DBN are manually defined according to the AUs-code expression descriptions.

Different from the related works, we first select discriminative AUs for a certain expression using incremental association Markov blanket (IAMB) algorithm [7]. Then, we construct a Bayesian network to systematically capture the dependencies between expression specific AUs and expression. The nodes of the BN represent the AUs and expressions. The links and their parameters capture the probabilistic relations among AUs and expressions. After that, we design an expression recognition method with the help of the extracted Markov blanket AU labels as hidden knowledge. We train expression recognition algorithm by leveraging on the relationships among the selected AUs and the expression. We refer the selected AU labels as the hidden knowledge since they are only available and used during training, and they will not be

*Shangfei Wang is the corresponding author. This work has been supported by National Program 863 (2008AA01Z122), the National Science Foundation of China (Grant No. 61175037, 61228304), project from Anhui Science and Technology Agency (1106c0805008) and the Fundamental Research Funds for the Central Universities.

available during testing. Given the trained BN, we can infer the expression by combining the BN and the measurement during testing, and the measurement nodes can be instantiated with the AUs and expressions' estimates obtained from a traditional image-based method. The experimental results on CK+ database show the superior performance of our model to the image-driven method. The experiment results on MMI database demonstrates the generalization ability of our model.

2. METHOD

The goal of this work is to construct an expression classifier which can learn and infer from facial images with the help of AU knowledge that is available during training. Our approach consists of three modules: discriminative AU selection (i.e. extraction of the Markov Blanket AUs of each expression), AU and expression measurement extraction, and relations between selected AUs and expression modeling by BN.

The training phase of our approach includes obtaining the Markov blanket AUs of each expression, training the traditional image-based classifiers for AU and expression measurement extraction and training the BN to capture the semantic relationships among Markov blanket AUs and expressions. For Markov blanket AUs extraction of each expression, IAMB algorithm is used. For measurement extraction, a current image-based algorithm is used. Given the measurements, we infer the final labels of samples through the most probable explanation (MPE) inference with the BN model.

2.1. Expression Dependent AU selection

Let $D = \{X_i, (\lambda_{1i}, ..., \lambda_{li}, \lambda_{l+1i})\}_{i=1}^{m}$ be the training data, where $X_i \in R^d$ is the facial image features, $(\lambda_{1i}, ..., \lambda_{li})$ are the multiple AU labels, which are only available during training, l is the total number of AU labels; λ_{l+1i} is the expression label, and m is the number of training samples.

The objective of AU selection is to seek a number of significant AUs for a certain expression to facilitate the recognition of such expression. Since the Markov blanket [8] of a target variable is the only knowledge needed to predict the target variable, we want to find the Markov blanket AUs of a certain expression. Given the the Markov blanket AUs of a certain expression, the distribution of this expression is conditionally independent of all the other AUs. Here, we adopt incremental association Markov blanket (IAMB) [7] algorithm, which consists of two phases: the growing phase and the shrinking phase. The growing phase starts with an empty set for the MB(λ_{l+1}) and then gradually adds AUs, λ_i, that maximizes a heuristic function [9, 10] as follows:

$$MI(\lambda_i, \lambda_{l+1}|MB(\lambda_{l+1})) = \sum_{\lambda_j \in MB(\lambda_{l+1})}^{\lambda_j} P(\lambda_j)\left[p(\lambda_i, \lambda_{l+1}|\lambda_j)\log\frac{p(\lambda_i, \lambda_{l+1}|\lambda_j)}{p(\lambda_i|\lambda_j)p(\lambda_{l+1}|\lambda_j))}\right] \tag{1}$$

where $\lambda_i \in \{D - MB(\lambda_{l+1}) - \{\lambda_{l+1}\}\}$.

During the growing phases, the computation of conditional dependency depends on the current formed Markov Blanket, which may cause false positives. Thus, the shrinking phase tests the conditional independence and remove the AUs that do not belong to the MB(λ_{l+1}) by testing whether a node λ_j from MB(λ_{l+1}) is independent of λ_{l+1} given the remaining MB(λ_{l+1}). Finally, we obtain the Markov blanket AUs $\lambda_j, j \in [1, n]$ for each expression λ_{l+1}, as shown in Fig.1 (in which the contempt picture is downloaded from internet because of the copyright limitation of the CK+ database).

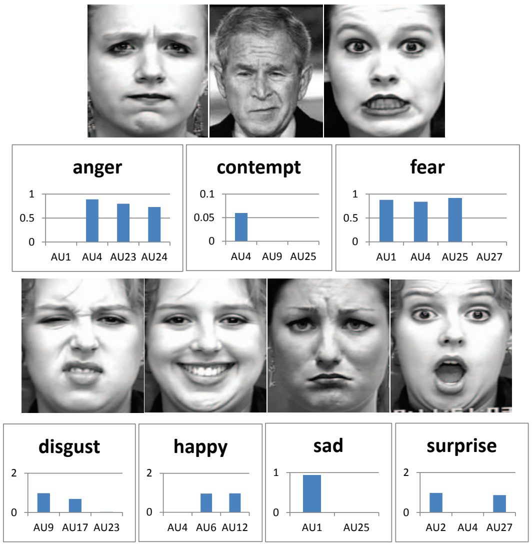


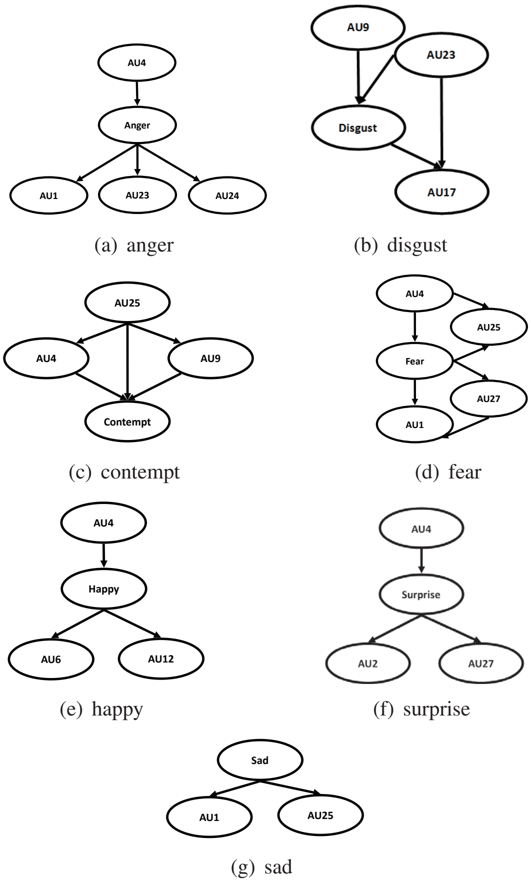
Fig. 1. The Markov blanket AUs and the co-existence probabilities of each expression on the CK+ database.

2.2. Measurement extraction

Let $DE = \{X_i, (\lambda_{1i}, ..., \lambda_{ni}, \lambda_{n+1i})\}_{i=1}^{m}$ be the training data for a certain expression λ_{n+1i}, where $X_i \in R^d$ is the facial image features, $(\lambda_{1i}, ..., \lambda_{ni})$ is the multiple Markov blanket AU labels, n is the number of Markov blanket AU labels of the expression; λ_{n+1i} is the expression label, and m is the number of training samples. The measurements $m\lambda$ are the preliminary estimations of the Markov blanket AUs and expression labels using an existing image-driven recognition method based on training data. In this work, the movements of the feature point between the neural and apex images are used as the image features, and Support Vector Machine (SVMs) are used as the classifier to obtain the measurements.

2.3. Modeling Dependencies between Expression and AUs by Bayesian Network

In order to model the semantic relationships among expression and Markov blanket AUs, a BN model is utilized in this work. As a probabilistic graphical model, BN can effectively capture the dependencies among variables in data. In our work, each node of the BN is an AU or expression label, and the links and their conditional probabilities capture the probabilistic dependencies among AUs and expression. Fig.2 shows the BN models of the 7 expressions on CK+ database.



(a) anger (b) disgust (c) contempt (d) fear (e) happy (f) surprise (g) sad

Fig. 2. BN models of each expression with its Markov blanket AUs on CK+ database

2.3.1. *BN Structure and Parameters Learning*

A BN is a directed acyclic graph (DAG) $G = (\Lambda, E)$, where $\Lambda = \{\lambda_i\}_{i=1}^{n+1}$ represents a collection of $n+1$ nodes and E denotes a collection of arcs.

Given the dataset of multiple target labels $TD = \{\lambda_{ij}\}$, where $i = 1, 2, ..., n, n+1$ is an index to the number of nodes, and $j = 1, 2, ..., m$ is index to the number samples. The structure and parameter learning is to find a structure G that maximizes a score function. In this work, we employ the Bayesian Information Criterion (BIC) [11] score function which is defined as Eq. 2

$$\begin{aligned} &Q^{BIC}(G, \theta : G^*, \theta^*) = \\ &E_{G^*,\theta^*}[logP(D|G,\theta)] - \frac{Dim(G)}{2}\log N \end{aligned} \tag{2}$$

where the first term is the log-likelihood function of structure G with respect to data D, representing how well G fits the data. The second term is a penalty relating to the complexity of the network, where Dim_G is the number of independent parameters and N is the number of samples.

To learn the structure, we propose to employ the BN structure learning algorithm [12]. By exploiting the decomposition property of the BIC score function, this method allows learning an optimal BN structure efficiently and it guarantees to find the global optimum structure, independent of the initial structure. Furthermore, the algorithm provides an anytime valid solution, i.e., the algorithm can be stopped at any-time with a best current solution found so far and an upper bound to the global optimum. Representing state of the art method in BN structure learning, this method allows automatically capturing the relationships among emotions. Details of this algorithm can be found in [12].

After the BN structure is constructed, parameters can be learned from the the ground truth labels and their measurements of the training data. Learning the parameters in a BN means finding the most probable values $\hat{\theta}$ for θ that can best explain the training data. Here, let λ_i denotes a variable of BN. Let θ_{ijk} denote a probability parameter for node λ_i^kBN, then,

$$\theta_{ijk} = P\left(\lambda_i^k \mid pa^j\left(\lambda_i\right)\right) \tag{3}$$

where $i \in \{1, ..., n\}$, $j \in \{1, ..., r_i\}$ and $k \in \{1, ..., s_i\}$.

$pa(\lambda_i)$ is a collection of parent instantiations for variable λ_i, r_i represents the number of the possible parent instantiations for variable λ_i, and s_i indicates the number of the state instantiations for λ_i, λ_i^k denotes the k_{th} state of variable λ_i.

In this work, the "fitness" of parameters θ and training data TD is quantified by the log likelihood function $log(P(TD|\theta))$, denoted as $L(\theta)$. Assuming the training data are independent, based on the conditional independence assumptions in BN, the log likelihood function is shown in Eq. 4, where n_{ijk} indicates the number of elements in TD containing both λ_i^k and $pa^j(\lambda_i)$.

$$L\left(\theta\right) = log\left(\prod_{i=1}^{n}\prod_{j=1}^{r_i}\prod_{k=1}^{s_i}\theta_{ijk}^{n_{ijk}}\right) \tag{4}$$

Maximum Likelihood Estimation (MLE) method can be described as a constrained optimization problem, which is shown in Eq. 5.

$$\begin{aligned} &MAX \quad L(\theta) \\ &S.T \quad g_{ij}\left(\theta\right) = \sum_{k=1}^{s_i}\theta_{ijk} - 1 = 0 \end{aligned} \tag{5}$$

where g_{ij} imposes the constraint that the parameters of each node sums to 1 over all the states of that node. Solving the above equations, we can get $\theta_{ijk} = \frac{n_{ijk}}{\sum_k n_{ijk}}$.

2.3.2. *BN Inference*

A complete BN model is obtained after parameter and structure learning. Given the expression and Markov blanket AUs measurements obtained in the former procedure, the true expression category of the input sample is estimated through BN inference. During the BN inference, the posterior probability of categories can be estimated by combining the likelihood from measurement with the prior model. Let λ_i and $m\lambda_i,\ i \in \{1, ..., n, n+1\}$, denote the label variable and the corresponding measurement obtained a multi-label learning method respectively. Then, most probable explanation (MPE) [9] inference is used to estimate the joint probability of AUs and expression, then the label of expression is inferred according to Eq.6.

$$Y^{\star} = \arg\max_{\lambda_{n+1}} P(\lambda_{n+1} \mid m\lambda_1, ..., m\lambda_n, m\lambda_{n+1})$$
$$= \arg\max_{\lambda_{n+1}} \sum_{\lambda_1,\lambda_2,...\lambda_n} (\prod_{i=1}^{n+1} P(m\lambda_i \mid \lambda_i) \prod_{i=1}^{n+1} P(\lambda_i \mid pa(\lambda_i))) \tag{6}$$

The first part of the equation is the likelihood of λ_j given the measurements and the second part is the product of the conditional probabilities of each category node λ_j given its parents $pa(\lambda_j)$, which are BN model parameters that have been learned. In this work, the inferred label gets the expression value with the highest probability given $m\lambda_1, ..., m\lambda_n, m\lambda_{n+1}$.

3. EXPERIMENTS

3.1. Experimental condition

The Extended Cohn-Kanade Dataset (CK+) [3], in which 7 expression categories (i.e. Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise) and 30 AU labels are provided for parts of the samples, is used to validate our method. Finally, 327 samples with both expression category and FACS labels are selected, and 13 AUs whose frequencies of all the selected samples are more than 10% are considered, which are: AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU12, AU17, AU23, AU24, AU25, and AU27. For each expression category, the experiment is a binary classification. The expression labels of all the samples are assigned to be 0 (i.e. absence) or 1 (i.e. presence). Therefore, for each binary classification problem, the instances of positive class are much less than negative class.

To validate the supplementary role of the selected AUs in assisting expression recognition, two experiments are conducted: the image-driven expression recognition and our model that recognize expression with the help of AUs. The image-driven expression recognition is the same as the initial expression measurement estimation discussed in section 2.2.

Furthermore, in order to evaluate the generalization ability of the proposed method, the cross-database expression recognition experiments are conducted. The Markov blanket AUs of each expression are extracted from the CK+ database, as shown in Fig.1, then the BN model of each expression are trained using the extracted Markov AUs and the measurements of MMI database [13]. The MMI database consists of over 2900 videos and high-resolution still images of 75 subjects. It is fully annotated for the presence of AUs in videos (event coding), and partially coded on frame-level, indicating for each frame whether an AU is in either the neutral, onset, apex or offset phase. 102 videos clips with both AU and emotion labels are adopted in this work.

In this work, F_1 score of positive instance and accuracy of all the samples are considered as metrics to evaluate our method. And 10-fold cross validation is adopted in these expreiments.

3.2. Dependencies between AU and expression on CK+ database

We quantify the dependence between different AUs and expressions using a conditional probability of $P(\lambda_j|\lambda_i)$, as shown in Table 1, which measures the probability of label λ_j happens, given label λ_i happens. From Table 1, we can find that there exist two kinds of relationships between AUs and expressions: coexistent and mutual exclusion. For example, $P(AU25|surprise)$ and $P(AU9|disgust)$ are higher than 0.980, which shows AU25 is always coexistent with surprise expression and AU9 is always coexistent with disgust expression. $P(AU1|anger)$ and $P(AU2|happy)$ are 0.00, which means anger expression never coexists with AU1, and AU2 is inactive when happy expression happens. To conclude, the AUs are important information for expressions. Comparing Table 1 with Fig.1, we can see that our method almost select the most coexistent and mutual exclusion AUs for each expression. To be specific, for anger, AU1 is one of the most mutual exclusion AUs, and the other three AUs are all among the first four coexistent AUs; for contempt, AU9 and AU25 are among the most mutual exclusion AUs; for disgust, AU9 and AU17 are the two most coexistent AUs, and AU23 is a mutual exclusion AU; for fear, AU1, AU4 and AU25 are the three most coexistent AUs, and AU27 is the most mutual exclusion AU; for happy, AU6 and AU25 are among the three most coexistent AUs, and AU4 is one of the most mutual exclusion AUs; for sad, AU25 is one of the most mutual exclusion AUs and AU1 is the second coexistent AU; for surprise, AU2 and AU27 are among the first four coexistent AUs, and AU4 is one of the most mutual exclusion AUs. Therefore, in most cases, our method capture the most coexistent and mutual exclusion AUs of each expression.

Table 1. Dependencies between AUs and expressions (each entry a_{ij} represents $P(\lambda_j = 1|\lambda_i = 1)$).

λ_i \ λ_j	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU12	AU17	AU23	AU24	AU25	AU27
anger	**0.000**	0.000	**0.889**	0.133	0.178	0.711	0.067	0.022	0.867	**0.800**	**0.733**	0.000	0.000
contempt	0.056	0.056	**0.056**	0.000	0.000	0.000	**0.000**	0.278	0.278	0.056	0.111	**0.000**	0.000
disgust	0.000	0.000	0.610	0.000	0.305	0.559	**0.983**	0.034	**0.695**	**0.034**	0.119	0.153	0.000
fear	**0.880**	0.400	**0.840**	0.640	0.120	0.240	0.000	0.080	0.120	0.000	0.000	**0.920**	**0.000**
happy	0.000	0.000	**0.000**	0.000	**0.957**	0.101	0.000	**0.971**	0.000	0.000	0.000	0.971	0.000
sad	**0.929**	0.250	0.821	0.000	0.000	0.036	0.000	0.000	0.964	0.107	0.036	**0.000**	0.000
surprise	0.976	**0.976**	**0.012**	0.843	0.000	0.000	0.000	0.036	0.000	0.012	0.000	0.988	**0.867**

Table 2. Experimental results on CK+ database.

method	parameter	anger	contempt	disgust	fear	happy	sadness	surprise	average
Image-driven	$F_1 score$	0.832	0.813	0.818	0.744	0.935	0.692	0.963	0.828
	Accuracy	0.954	0.982	0.929	0.966	0.973	0.951	0.982	0.962
Our's	$F_1 score$	0.832	0.813	**0.862**	0.744	**0.958**	**0.720**	0.963	**0.842**
	Accuracy	0.954	0.982	**0.951**	0.966	**0.982**	**0.957**	0.982	**0.968**

3.3. Experimental results on CK+ database

The experimental results of expression recognition are shown in Table 2. From this table, we can conclude that, our method outperforms the image-driven method for three of the seven expressions, including disgust, happy and sadness. Since both the F_1 score and accuracy of these three expressions with our method are higher than those with the image-driven method. That means considering both recall and precision, our method achieves the better performance than the image-driven method on the positive class. And our method correctly predicts much more instances than image-driven method for both the positive class and the negative class. The image-driven method directly predicts each expression from the image features, ignoring the AU knowledge. However, in fact, the expressions are not totally independent with AUs. For the other four of the seven expressions, the performance of our method are same as those of the image-driven method. So, for all of the 7 expressions and the average value, both the F_1 score and accuracy of our method are not less than the image-driven method. Therefore, our method works effectively.

3.4. Cross-database experiments

Table 3 shows the cross-database expression recognition results, where the "Image-driven" lines show the F_1 score and accuracy of the directly image-driven method and the "Our's" lines show the cross database recognition results. From this table, we can obtain that the proposed method works well when trained on CK+ database and tested on MMI database, since for three of the five expressions, both the F_1 score and accuracy are improved by our method. The average F_1 score is improved from 77.3% to 80.0% and the accuracy is improved from 90.8% to 92.2% .It demonstrates the generalization ability of our method.

3.5. Comparison with related works

As mentioned in the introduction, a few works consider the relations between expressions and AUs to help expression recognition, or to jointly recognize AUs and expressions. Among all the works, Pantic and Rothkrantz [5] evaluate their facial expression method on a set of 265 dual-view images; Zhang and Ji [6] presented sequences acquired from their IR illumination-based system to verify the effectiveness of the proposed dynamic Bayesian networks; Li et al [4] performed experiments on CK+ database and MMI database. Therefore, we compare our work with Li et al. Experimental results on CK+ database of Li et al [4] are listed in Table 4. From this table, we can see, for three of the six expressions, our method works more effective than their's, and for the other three expressions, their method outperforms our's. Considering the average F_1 score their method slightly outperforms our's, and considering the accuracy, our method slightly better than their's. However, our method use only the still image, while their method models dynamic relations using the whole image sequence. Therefore, in general, our method models capture more effective information than their method and the AUs we selected observably supplement expression recognition. Li et al [4] do experiments on MMI database to evaluate the generalization ability of their model and they achieves an average expression recognition rate of 82.4%. While our method achieve average expression recognition rate of 92.2%, which is much higher than their's, although the different samples of the same database are used in the our experiment and their experiment. It verifies the the improvement of our method.

4. CONCLUSION

In this paper, we propose an expression classifier, consisting of two steps: first, the Markov blanket AUs of certain expression are extracted, second, construct BN modeling the dependencies among the Markov blanket AUs and each certain ex-

Table 3. Cross database experimental results on MMI database.

method	parameter	anger	disgust	happy	sadness	surprise	average
Image-driven	$F_1 score$	0.629	0.596	0.913	0.833	0.894	0.773
	Accuracy	0.873	0.814	0.961	0.941	0.951	0.908
Our's	$F_1 score$	**0.706**	**0.636**	0.913	0.833	**0.913**	**0.800**
	Accuracy	**0.902**	**0.843**	0.961	0.941	**0.961**	**0.922**

Table 4. Comparison with related works on CK+ database.

method	parameter	anger	disgust	fear	happy	sadness	surprise	average
Yongqiang's	$F_1 score$	0.759	**0.923**	**0.833**	**0.986**	0.647	0.952	**0.850**
	Accuracy	0.939	**0.971**	**0.974**	**0.994**	0.922	0.974	0.962
Our's	$F_1 score$	**0.832**	0.862	0.744	0.958	**0.720**	**0.963**	0.842
	Accuracy	**0.954**	0.951	0.966	0.982	**0.957**	**0.982**	**0.968**

pression. With this classifier, the Markov blanket AUs are available during training but unavailable during testing. This BN structure not only model the dependencies between AU and expression but also the dependencies among AUs. Furthermore, the relations between AU and expressions are only dependent on the ground truth label data, independent on the facial images. Therefore, the method of capturing such relations can be applied to not only posed, frontal, and pre-segmented sequences, but also spontaneous, non-frontal, non-pre-segmented sequences. Markov blanket AUs and expression recognition results using image-driven method are adopted as measurements. Experimental results on CK+ database demonstrates that our method further improve the expression recognition performance of image-driven method. Cross database experiments shows the generalization ability of the proposed method.

References

[1] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[2] Paul Ekman and Wallace V Friesen, "Facial action coding system: A technique for the measurement of facial movement. palo alto," *CA: Consulting Psychologists Press. Ellsworth, PC, & Smith, CA (1988). From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion*, vol. 12, pp. 271–302, 1978.

[3] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[4] Yongqiang Li, Shangfei Wang, Yongping Zhao, and Qiang Ji, "Simultaneous facial feature tracking and facial expression recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2559–2573, 2013.

[5] Maja Pantic and Leon JM Rothkrantz, "Expert system for automatic analysis of facial expressions," *Image and Vision Computing*, vol. 18, no. 11, pp. 881–905, 2000.

[6] Yongmian Zhang and Qiang Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 5, pp. 699–714, 2005.

[7] Pabitra Mitra, CA Murthy, and Sankar K. Pal, "Unsupervised feature selection using feature similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 301–312, 2002.

[8] Daphne Koller and Mehran Sahami, "Toward optimal feature selection," 1996.

[9] Judea Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, 1988.

[10] Ioannis Tsamardinos, Constantin F Aliferis, Alexander R Statnikov, and Er Statnikov, "Algorithms for large scale markov blanket discovery.," in *FLAIRS Conference*, 2003, vol. 2003, pp. 376–381.

[11] Gideon Schwarz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[12] Cassio Polpo de Campos and Qiang Ji, "Efficient structure learning of bayesian networks using constraints.," *Journal of Machine Learning Research*, vol. 12, no. 3, pp. 663–689, 2011.

[13] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 5–pp.