FACIAL ACTION UNIT PREDICTION UNDER PARTIAL OCCLUSION BASED ON ERROR WEIGHTED CROSS-CORRELATION MODEL

Jen-Chun Lin, Chung-Hsien Wu, and Wen-Li Wei

Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan, Taiwan
Email: {jenchunlin, chunghsienwu, lilijinjin} @gmail.com

ABSTRACT

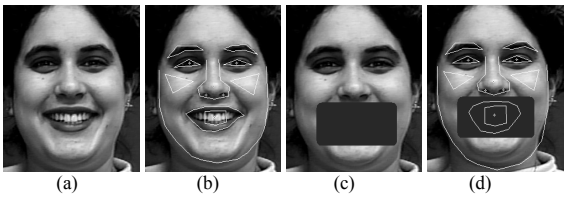
Occlusive effect is a crucial issue that may dramatically degrade performance on facial expression recognition. As emotion recognition from facial expression is based on the entire facial feature, occlusive effect remains a challenging problem to be solved. To manage this problem, an Error Weighted Cross-Correlation Model (EWCCM) is proposed to effectively predict the facial Action Unit (AU) under partial facial occlusion from non-occluded facial regions for providing the correct AU information for emotion recognition. The Gaussian Mixture Model (GMM)-based Cross-Correlation Model (CCM) in EWCCM is first proposed not only modeling the extracted facial features but also constructing the statistical dependency among features from paired facial regions for AU prediction. The Bayesian classifier weighting scheme is then adopted to explore the contributions of the GMM-based CCMs to enhance the prediction accuracy. Experiments show that a promising result of the proposed approach can be obtained.

Index Terms— Occlusive effect, facial expression recognition, action unit, Gaussian mixture model

1. INTRODUCTION

With the growing and varied uses of human-computer interactions, emotion recognition technology has been used to provide harmonious interactions or communication between computers and humans [1-8]. Facial expression undoubtedly plays an important role in social interaction, perception, memory, and emotional lives. Understanding latent meaning of facial expression is indispensable for day-to-day functioning of humans. Hence, constructing a high-performance emotion perception and recognition system from facial expression is highly desirable.

Although most researches for automatic facial expression (i.e., facial affect (emotion) or facial muscle action (Action Unit (AU))) recognition have been successfully developed [9-17], the recordings were made under particular type of facial images, in which the facial pose is constrained to be frontal or near-frontal view and without considering partial facial occlusion. To increase the system's value in real life applications, an important issue on facial expression recognition is the effect of partial occlusion for the impact on the recognition accuracy. It is well known that when one or more regions of the face are occluded, some information will be lost and lead to inaccurate estimate of the facial features. Therefore, the recognition result may be unsatisfactory. This problem is significant, because human face is often occluded by adornments or hand gesture in real communication. Fig. 1 shows an example of Active Appearance Model (AAM) [18] alignment results for non-occluded and occluded images in the Cohn-Kanade (CK) database [19]. According to our observations, Fig. 1 shows a striking effect of partial occlusion on performance in the AAM alignment. Compared with Fig. 1 (b), Fig. 1 (d) shows a poorer alignment result on the lower facial regions including mouth and facial contours. Such results may lead to inaccurate facial feature extraction, and then influence the recognition performance.

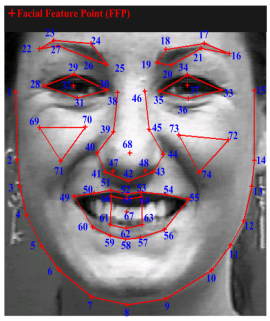


(a) (b) (c) (d)

Fig. 1. Examples of AAM alignment results for non-occluded and occluded face images (a) original image (adopted from [19]) (b) aligned result of the original image (c) occluded original image (d) aligned result of the occluded original image.

Several studies have seen increasing attention being given to occlusive effect on facial expression recognition. Bourel et al. [20], used the decision level fusion approach combining the six local interpretations of the face model into a general classification score in the presence of partial facial occlusion, which try to utilize the advantage of data fusion strategy for alleviating the occlusive effect. However, from the analysis of experimental results, we found that the occlusive region often contains critical information and may dominate the recognition performance. To analyze six prototypical facial expressions under partial occlusion, two types of facial features (i.e., texture and shape features) were investigated by Kotsia et al. [21], to determine which one was robust for occlusive effect. They also provided an extended analysis to explore which part of the face regions contains the most discriminative information for facial expression recognition, as well as to define the pairs of facial expressions that were usually confused with each other. Although this study provides a rich analysis for understanding how partial occlusion affects facial expression recognition, it does not propose any method to deal with the occlusive effect. Miyakoshi et al. [22], proposed an emotion detection system considering partial occlusion of the face using causal relations between explanatory variables. There are two learning phases presented by the proposed Bayesian network classifier: an internal phase and an external phase, which are applied to construct the causal structure among facial features first and then characterize the relationships between the structure and

TABLE 1. LIST AND DESCRIPTION OF FACIAL FEATURES WITH RESPECT TO THE EXTRACTED FDPS AND RELATED FFPS

	Facial Features	FDPs Num.	FFPs Displacement Between Test Frame and Neutral Frame
	Eyebrows	EB_1, EB_2 EB_3, EB_4 EB_5, EB_6 EB_7, EB_8	$D_{vertical}$(FFP16_Neutral–FFP16), $D_{vertical}$(FFP17_Neutral–FFP17) $D_{horizontal}$(FFP19_Neutral–FFP19), $D_{vertical}$(FFP19_Neutral–FFP19) $D_{vertical}$(FFP22_Neutral–FFP22), $D_{vertical}$ (FFP23_Neutral–FFP23) $D_{horizontal}$ (FFP25_Neutral–FFP25), $D_{vertical}$(FFP25_Neutral–FFP25)
	Eyes	EY_1, EY_2 EY_3, EY_4	$D_{vertical}$(FFP34_Neutral–FFP34), $D_{vertical}$(FFP36_Neutral–FFP36) $D_{vertical}$(FFP29_Neutral–FFP29), $D_{vertical}$(FFP31_Neutral–FFP31)
	Nose	NO_1, NO_2 NO_3	$D_{horizontal}$(FFP44_Neutral–FFP44), $D_{vertical}$(FFP68_Neutral–FFP68) $D_{horizontal}$(FFP40_Neutral–FFP40)
	Cheeks	CH_1, CH_2 CH_3, CH_4	$D_{vertical}$(FFP74_Neutral–FFP74), $D_{horizontal}$ (FFP74_Neutral–FFP74) $D_{vertical}$(FFP71_Neutral–FFP71), $D_{horizontal}$ (FFP71_Neutral–FFP71)
	Mouth	MO_1, MO_2 MO_3, MO_4 MO_5, MO_6	$D_{vertical}$(FFP55_Neutral–FFP55), $D_{horizontal}$(FFP55_Neutral–FFP55) $D_{vertical}$(FFP52_Neutral–FFP52), $D_{vertical}$(FFP49_Neutral–FFP49) $D_{horizontal}$(FFP49_Neutral–FFP49), $D_{vertical}$(FFP58_Neutral–FFP58)
	Jaw	JA_1	$D_{vertical}$(FFP8_Neutral–FFP8)

the emotion. Although the performance of the proposed Bayesian network structure was better than that of conventional one, it was still far from satisfactory under occlusive effect especially when brow or mouth region was occluded.

In summary, previous work did not provide an effective method to deal with the problems for missing or error features produced from the occluded region, while entire facial information was considered for emotion recognition. Hence the performance improvement is limited. To solve the occlusive effect on emotion recognition from facial expression, different from the previous studies, we turn our attention to the problem of facial expressions in particular prediction of AU or AU combination in the occluded region using the proposed Error Weighted Cross-Correlation Model (EWCCM), which try to provide AU information in the occluded region for emotion recognition.

The rest of the paper is organized as follows. Section 2 briefly outlines the procedure for feature extraction. Section 3 details the derivation of error weighted cross-correlation model. Section 4 shows the experimental results. Section 5 offers a conclusion.

2. FEATURE EXTRACTION

For providing initial facial position and reducing the time for error convergence in Facial Feature Points (FFPs) alignment, the Adaboost-cascade face detector [23] is performed first. The AAM [18] is then applied to effectively localize human facial features on a 2D visual image and is employed to extract the 74 labeled FFPs from six facial features including eyebrows, eyes, nose, cheeks, mouth, and jaw for later Facial Deformation Parameters (FDPs) calculation. According to the difference of recording environment, the recorded data may contain different facial locations and scales. To eliminate such variation, the normalization technique is applied to normalize each detected face into the fixed location and scale.

In terms of FDPs, measurement of FDPs requires a basis frame, which is manually selected from the database at the beginning. The subject's expression in the basis frame is assumed to be neutral. Thus, the FDPs from each subject are calculated by the FFPs displacements between the test frame and the neutral frame. Table 1 summarizes the information about each facial feature with respect to the extracted FDPs and related FFPs.

3. ERROR WEIGHTED CROSS-CORRELATION MODEL

Respecting that the facial features are interplayed to show meaningful facial expressions [24], in this study, the Gaussian Mixture Model (GMM)-based CCM predictors are first proposed in modeling the relationships among facial features, and then integrating with Bayesian classifier weighting scheme (i.e., Error Weighted Classifier Combination (EWC) method [25-27]) to form a novel statistical model EWCCM for obtaining an optimal prediction result of AU or AU combination. According to the analysis in [21], facial expression is seen to be relevant to upper, middle, and lower regions of the face, with similar facial muscle contractions in each region. Hence, when the occluded region is detected (e.g., lower facial region), the GMM-based CCM predictors are first used for occluded region AU or AU combination prediction, based on the paired facial features from non-occluded regions, such as eyebrows-nose features in upper and middle regions. The Bayesian classifier weighting scheme, is then adopted to integrate all the GMM-based CCM predictors, each for one paired facial features, to give the final decision. The block diagram of the proposed EWCCM is depicted in Fig. 2 and the formula is derived as follows.

Given a prediction task with K classes, the goal is to utilize non-occluded facial feature pair (x^m, x^n) from the m-th and the n-th facial regions (e.g., upper and middle facial regions) to predict the occluded region (e.g., lower facial region) into an AU or AU combination class w by combining the decisions determined by the predictors from C and D facial features referring to Fig. 2. Two sets of features, $x^m = \{x^m_{i=1}, x^m_{i=2} \ldots x^m_{i=C}\}$ and $x^n = \{x^n_{j=1}, x^n_{j=2} \ldots x^n_{j=D}\}$, represent multiple independent facial observation channels, respectively, where x^m_i and x^n_j are the i-th and the j-th features of x^m and x^n for GMM-based CCM predictor with model parameters λ^m_i and λ^n_j, respectively. For each feature pair (x^m, x^n), based on C and D facial features, this study tries to identify a true AU or AU combination class w out of the K classes. Assuming that subsets x^m_i are disjoint and conversely the subsets x^n_j are also disjoint, $P(w \mid x^m, x^n, \lambda^m_i, \lambda^n_j) \approx P(w \mid x^m_i, x^n_j, \lambda^m_i, \lambda^n_j)$ is deduced for all predictors $(\lambda^m_i, \lambda^n_j)$. Therefore, we can obtain

$$\begin{aligned} P(w \mid x^m, x^n) &= \sum_{i=1}^{C}\sum_{j=1}^{D} P(w, \lambda^m_i, \lambda^n_j \mid x^m, x^n) \\ &= \sum_{i=1}^{C}\sum_{j=1}^{D} P(w \mid \lambda^m_i, \lambda^n_j, x^m, x^n) P(\lambda^m_i, \lambda^n_j \mid x^m, x^n) \end{aligned} \quad (1)$$

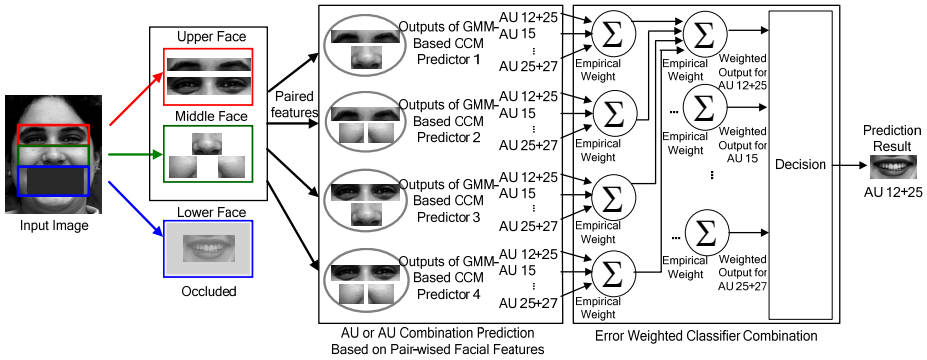


Fig. 2. The architecture of the proposed error weighted cross-correlation model.

where $P(\lambda_i^m, \lambda_j^n \mid x^m, x^n)$ is an empirical weight assigned to each predictor representing the confidence of the decision of the predictor [27]. This weight can be computed from the confusion matrix of predictor $(\lambda_i^m, \lambda_j^n)$.

Since exploring the true AU or AU combination class w is based on combining the outputs of the all predicted AU or AU combination class $\tilde{w}_k$ from individual GMM-based CCM predictors, the $\tilde{w}_k$ is further included. The identification of the true AU or AU combination class w can be described as follows:

$$\begin{aligned} P(w \mid \lambda_i^m, \lambda_j^n, x^m, x^n) &= \sum_{k=1}^{K} P(w, \tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n) \\ &= \sum_{k=1}^{K} P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n, x^m, x^n) P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n) \end{aligned} \quad (2)$$

where the conditional error distribution $P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n, x^m, x^n)$ is further approximated by its projection $P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n)$, which can be simply obtained from the confusion matrix of the corresponding predictor, and regarded as an empirical weight. Based on this assumption, (2) can be approximated as follows:

$$P(w \mid \lambda_i^m, \lambda_j^n, x^m, x^n) \approx \sum_{k=1}^{K} P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n) P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n) \quad (3)$$

where $P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n)$ denotes the prediction probability of AU or AU combination class $\tilde{w}_k$ given the facial feature pair (x^m, x^n) and the corresponding predictor $(\lambda_i^m, \lambda_j^n)$. Substituting (3) into (1) we arrive at (4)

$$\begin{aligned} P(w \mid x^m, x^n) \approx \sum_{i=1}^{C} \sum_{j=1}^{D} \sum_{k=1}^{K} & P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n) \\ & \times P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n) P(\lambda_i^m, \lambda_j^n \mid x^m, x^n). \end{aligned} \quad (4)$$

In this study, each facial region contains two facial features (i.e., C and D are equal to 2). Each predictor consists of five (K=5) GMM-based CCMs, one for each AU or AU combination class. For the GMM-based CCM, the probability $P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n)$ can be further decomposed using Bayes rule as

$$\begin{aligned} P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n, x^m, x^n) &= \frac{P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n)}{P(x^m, x^n \mid \lambda_i^m, \lambda_j^n)} \\ &\approx P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n) \end{aligned} \quad (5)$$

where the probability of $P(x^m, x^n \mid \lambda_i^m, \lambda_j^n)$ are the same for all predicted AU or AU combination classes $\tilde{w}_k$ and can thus be omitted. $P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n)$ denotes the probability that a predictor $(\lambda_i^m, \lambda_j^n)$ assigns a predicted AU or AU combination class $\tilde{w}_k$. This probability can also be approximated from the confusion matrix of $(\lambda_i^m, \lambda_j^n)$ as an empirical weight. $P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k)$ in the proposed GMM-based CCM can be approximated using a co-occurrence dependency measure which not only models each of extracted facial features but also explores the statistical dependency among features from paired facial regions for AU or AU combination prediction. The formulation of the proposed GMM-based CCM is described as follows.

Assuming that two component GMMs are modeled separately and their observations are related to each other for the same predicted AU or AU combination class $\tilde{w}_k$, $P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k)$ can be approximated by (6)

$$P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) \approx P(x^m, x^n \mid \lambda_i^m, \tilde{w}_k) P(x^m, x^n \mid \lambda_j^n, \tilde{w}_k) \quad (6)$$

where $P(x^m, x^n \mid \lambda_i^m, \tilde{w}_k)$ can be divided into two parts using Bayes rule which consists of the prediction output of GMM $P(x^m \mid \lambda_i^m, \tilde{w}_k)$ and the probability of co-occurrence dependency of facial features $P(x^n \mid x^m, \lambda_i^m, \tilde{w}_k)$ from the m-th facial region to the n-th facial region. Conversely, the $P(x^m, x^n \mid \lambda_j^n, \tilde{w}_k)$ can also be inferred by the Bayes rule. Therefore, (6) can be re-written as

$$\begin{aligned} P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) \approx & P(x^m \mid \lambda_i^m, \tilde{w}_k) \\ & \times P(x^n \mid x^m, \lambda_i^m, \tilde{w}_k) P(x^m \mid x^n, \lambda_j^n, \tilde{w}_k) P(x^n \mid \lambda_j^n, \tilde{w}_k) \end{aligned} \quad (7)$$

where the probabilities $P(x^n \mid x^m, \lambda_i^m, \tilde{w}_k)$ and $P(x^m \mid x^n, \lambda_j^n, \tilde{w}_k)$ are difficult to be obtained, because the collected training data are unlikely to sufficiently cover all ranges of observation features x^m and x^n for co-occurrence dependency construction, especially when the scale of observation values is large. Accordingly, the quantization technique is employed to quantize the observed features x^m and x^n into types of β^m and α^n in (8) by simplifying the co-occurrence dependency condition:

$$\begin{aligned} P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) \approx & P(x^m \mid \lambda_i^m, \tilde{w}_k) \\ & \times P(\alpha^n \mid \beta^m, \lambda_i^m, \tilde{w}_k) P(\beta^m \mid \alpha^n, \lambda_j^n, \tilde{w}_k) P(x^n \mid \lambda_j^n, \tilde{w}_k) \end{aligned} \quad (8)$$

Because each facial feature contains different number of feature parameters (i.e., FDPs), the probabilities of co-occurrence dependency $P(\alpha^n \mid \beta^m, \lambda_i^m, \tilde{w}_k)$ and $P(\beta^m \mid \alpha^n, \lambda_j^n, \tilde{w}_k)$ in (8) can be further represented by (9) and then form the GMM-based CCM.

$$P(x^m, x^n \mid \lambda_i^m, \lambda_j^n, \tilde{w}_k) \approx P(x^m \mid \lambda_i^m, \tilde{w}_k) \times \prod_{p=1}^{P} \prod_{q=1}^{Q} P(\alpha_p^n \mid \beta_q^m, \lambda_i^m, \tilde{w}_k) P(\beta_q^m \mid \alpha_p^n, \lambda_j^n, \tilde{w}_k) P(x^n \mid \lambda_j^n, \tilde{w}_k) \quad (9)$$

where both the horizontal and vertical FDPs of β_q^m and α_p^n are quantized into five meaningful types. For example the vertical FDP of β_q^m or α_p^n will be quantized into one type of strongly raised, raised, invariable, descending, and strongly descending feature through threshold setting. The threshold is determined through the training set analysis. Thus, the probabilities of $P(\alpha_p^n \mid \beta_q^m, \lambda_i^m, \tilde{w}_k)$ and $P(\beta_q^m \mid \alpha_p^n, \lambda_j^n, \tilde{w}_k)$ can be estimated by the co-occurrence probabilities, respectively. Finally, combining (9) and (5) into (4) yields (10) for AU or AU combination prediction using EWCCM:

$$P(w \mid x^m, x^n) \approx \sum_{i=1}^{C} \sum_{j=1}^{D} \left\{ \sum_{k=1}^{K} P(w \mid \tilde{w}_k, \lambda_i^m, \lambda_j^n) \Big[P(x^m \mid \lambda_i^m, \tilde{w}_k) \prod_{p=1}^{P} \prod_{q=1}^{Q} P(\alpha_p^n \mid \beta_q^m, \lambda_i^m, \tilde{w}_k) P(\beta_q^m \mid \alpha_p^n, \lambda_j^n, \tilde{w}_k) P(x^n \mid \lambda_j^n, \tilde{w}_k) \Big] P(\tilde{w}_k \mid \lambda_i^m, \lambda_j^n) \right\} P(\lambda_i^m, \lambda_j^n \mid x^m, x^n). \quad (10)$$

4. EXPERIMENTAL RESULTS

This study evaluated the performance of the proposed method based on the Cohn-Kanade facial expression database [19]. The database contains more than 100 subjects covering different races, ages, and genders exhibiting single AU and AU combination. The image sequences begin with a neutral face and end with maximum intensity facial expression. For preliminary experiments, total of 176 images were selected from 90 subjects of CK database to generate the data of partial occlusion of the face where the lower facial region being occluded. The gray box was used to simulate the occlusive effect as shown in Fig. 1. Five types of the AUs (contains AU and AU combination), including AU12+25, AU15, AU20+25, AU 23+24, and AU25+27 were considered. The GMM with three mixtures was used to model the color distribution of mouth region for occlusive facial region detection. Three types of the statistical models were used for occluded region AU or AU combination prediction including traditional GMM and the proposed CCM and EWCCM. The experiments were performed on the leave-one-subject-out cross validation for each target AU. In the experiments, the GMM with eight mixtures was used in the approaches for the mentioned three AU prediction models.

In the experiments, the detection rate of occlusive region achieved 96.59% accuracy (170 occlusive data are successfully detected) that can provide reliability for ensuing AU prediction. Tables 2, 3, and 4 show the confusion matrixes obtained as the results of AU prediction based on the three statistical models. Among three models, traditional GMM achieved worst prediction accuracy. A reasonable explanation is that high dimensional feature set may easily suffer from the data sparseness problem and lead to over-fitting effect; because the traditional GMM is directly applied in modeling the facial features jointly. Compared with traditional GMM, the proposed CCM improved average prediction accuracy by approximately 20% (i.e., traditional GMM achieved 57.06% average prediction accuracy, and the CCM achieved 76.47%). The findings are in accordance with literature analyses

TABLE 2. CONFUSION MATRIX FOR AU AND AU COMBINATION PREDICTION USING TRADITIONAL GMM

AUs	12+25	15	20+25	23+24	25+27	Images
12+25	30	15	6	6	2	**59**
15	5	6	1	1	0	**13**
20+25	6	0	5	2	3	**16**
23+24	1	3	0	17	0	**21**
25+27	0	0	20	2	39	**61**

TABLE 3. CONFUSION MATRIX FOR AU AND AU COMBINATION PREDICTION USING CCM

AUs	12+25	15	20+25	23+24	25+27	Images
12+25	51	0	1	1	6	**59**
15	0	10	0	2	1	**13**
20+25	6	0	2	1	7	**16**
23+24	1	1	0	14	5	**21**
25+27	6	0	0	2	53	**61**

TABLE 4. CONFUSION MATRIX FOR AU AND AU COMBINATION PREDICTION USING EWCCM

AUs	12+25	15	20+25	23+24	25+27	Images
12+25	51	0	1	1	6	**59**
15	0	12	0	1	0	**13**
20+25	5	0	7	1	3	**16**
23+24	1	1	0	18	1	**21**
25+27	5	0	0	2	54	**61**

[24]; exploring information with mutual correlations among various facial features is significant on facial expression recognition. In addition, since each facial feature such as eyebrows, eyes, cheeks, and nose in the CCM is modeled by the corresponding GMM individually, it may not increase the dimensionality. Hence it leads to a better understanding of why the proposed CCM outperforms the traditional GMM method. In order to enhance the prediction accuracy, CCM is further combined with EWC to form a novel statistical model EWCCM. Accordingly, among the mentioned AU prediction models, the average prediction accuracy of the proposed EWCCM is the best (i.e., achieved 83.53% accuracy). The result confirmed that considering the relationships among facial features and to gain insights into the role of each paired facial features is a great help for AU prediction. Finally, in considering the error propagation from detection result of occlusive region in which six error detection results were also regarded as error for AU prediction, the proposed EWCCM still obtains satisfactory performance (i.e., achieved 80.68% accuracy). Based on these analyses, the experiments demonstrated that the proposed method is effective to provide correct facial AU information in occluded region for ensuing emotion recognition.

5. CONCLUSION

This paper presented a novel statistical model to predict the AU or AU combination under partial facial occlusion for providing the correct facial information for emotion recognition. Experimental results show that considering the relationships among facial features and exploring their contributions to different AUs or AU combinations in the proposed EWCCM is helpful for enhancing the prediction accuracy. The proposed approach can be further applied to provide useful information in helping reconstruct human face under occlusive effect for face identification. Furthermore, the proposed EWCCM is highly flexible and can be easily applied to other computer vision and multimodal classification problems. Future research to predict an expanded set of AUs and explore different types of occlusions is envisioned.

6. REFERENCES

[1] R.W. Picard, *Affective Computing*. MIT Press, 1997.

[2] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 31, no. 1, pp. 39–58, 2009.

[3] C.H. Wu, Z.J. Chuang, and Y.C. Lin, "Emotion recognition from text using semantic label and separable mixture model," *ACM Trans. on Asian Language Information Processing*, vol. 5, no. 2, pp. 165–182, Jun. 2006.

[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in Human-Computer Interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[5] C.H. Wu and W.B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels," *IEEE Trans. Affective Computing*, vol. 2, no.1, pp. 1–12, 2011.

[6] M. Pantic and M. Bartlett, "Machine analysis of facial expressions," in *Face Recognition*, K. Delac and M. Grgic, Eds. Vienna, Austria: I-Tech Educ. Publishing, pp. 377–416, 2007.

[7] M. Pantic, L.J.M. Rothkrantz, and H. Koppelaar, "Automation of non-verbal communication of facial expressions," in *Proc. Conf. Euromedia*, pp. 86–93, 1998.

[8] J.C. Lin, C.H. Wu, and W.L. Wei, "Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition," *IEEE Trans. Multimedia*, vol. 14, no.1, pp. 142–156, Feb. 2012.

[9] W.F. Liu, J.L. Lu, Z.F. Wang, and H.J. Song, "An expression space model for facial expression analysis," in *Proc. CISP*, vol. 2, pp. 680–684, May 2008.

[10] K. Anderson and P.W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Trans. Syst., Man, Cybern. B, Cybern*., vol. 36, no. 1, pp. 96–105, 2006.

[11] S. Park, J. Shin, and D. Kim, "Facial expression analysis with facial expression deformation," in *Proc. 19th ICPR*, pp. 1–4, Dec. 2008.

[12] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features" *IEEE Trans. Affective Computing*, vol. 2, no. 4, pp. 219–229, 2011.

[13] S. Koelstra, M. Pantic, and I. Patras, "A dynamic texture-based approach to recognition of facial actions and their temporal models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1940–1954, 2010.

[14] M.F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Trans. Syst., Man, Cybern. B, Cybern*., vol. 42, no.1, pp. 28–43, 2012.

[15] B. Jiang, M.F. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog*., pp. 314–321, 2011.

[16] M. Pantic and I. Patras, "Detecting facial actions and their temporal segments in nearly frontal-view face image sequences," in *Proc. IEEE Int. Conf. Syst., Man, Cybern*., pp. 3358–3363, 2005.

[17] M.H. Mahoor, M. Zhou, K.L. Veon, S.M. Mavadati, and J.F. Cohn, "Facial action unit recognition with sparse representation," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog*., pp. 336–342, 2011.

[18] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, no. 6, pp. 681–685, 2001.

[19] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog*., pp. 46–53, 2000.

[20] F. Bourel, C.C. Chibelushi, and A.A. Low, "Recognition of facial expressions in the presence of occlusion," *British Machine Vision Conference*, Manchester, UK, pp. 213–222, 2001.

[21] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Jnl. Image and Vision Computing*, vol. 26, no. 7, pp. 1052–1067, 2008.

[22] Y. Miyakoshi and S. Kato, "Facial emotion detection considering partial occlusion of face using Bayesian network," *IEEE Symposium Computers & Informatics*, pp. 96–101, 2011.

[23] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. Int'l Conf. Computer Vision Pattern Recognition*, vol. 1, pp. 511–518, 2001.

[24] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 29, no. 10, pp. 1683–1699, 2007.

[25] A. Kapoor, R.W. Picard, and Y. Ivanov, "Probabilistic combination of multiple modalities to detect interest," in *Proc. Int'l Conf. Pattern Recognition*, vol. 3, pp. 969–972, 2004.

[26] Y. Ivanov, T. Serre, and J. Bouvrie, "Error weighted classifier combination for multi-modal human identification," Technical Report CBCL paper 258, Massachusetts Institute of Technology, Cambridge, MA, 2005.

[27] A. Metallinou, S. Lee, and S. Narayanan, "Audio-visual emotion recognition using Gaussian mixture models for face and voice," in *Proc. Int'l Symposium on Multimedia (ISM' 08)*, pp. 250–257, 2008.