Hybrid Patch Based Diagonal Pattern Geometric Appearance Model for Facial Expression Recognition

Deepak Kumar Jain[1,2,3(✉)], Zhang Zhang[1,2,3], and Kaiqi Huang[1,2,3]

[1] Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing, China
deepak.juet@cripac.ia.ac.cn

[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[3] University of Chinese Academy of Sciences, Beijing, China

Abstract. Automatic Facial Expression Recognition (FER) is an imperative process in next generation Human-Machine Interaction (HMI) for clinical applications. The detailed information analysis and maximization of labeled database are the major concerns in FER approaches. This paper proposes a novel Patch-Based Diagonal Pattern (PBDP) method on Geometric Appearance Models (GAM) that extracts the features in a multi-direction for detailed information analysis. Besides, this paper adopts the co-training to learn the complementary information from RGB-D images. Finally, the Relevance Vector Machine (RVM) classifier is used to recognize the facial expression. In experiments, we validate the proposed methods on two RGB-D facial expression datasets, i.e., EURECOMM dataset and biographer dataset. Compared to other methods, the comparative analysis regarding the recognition and error rate prove the effectiveness of the proposed PBDP-GAM in FER applications.

1 Introduction

Numerous advancements in computer technology makes an automatic Facial Expression Recognition (FER) as an attractive research area for HMI. The knowledge about the facial parametric model is the prior requirement to recognize the expression status. The Facial Action Coding Systems (FACS) and the Facial Animation Parameters (FAP) [1] define the muscle actions and animations for standardized face parameterization. The utilization of 3D and 4D recordings [2] improves the ability of exploitation of facial information. The Local Diagonal Number (LDN) pattern [3] is the derived method from LDP that encodes the structural and intensity variations for specific face texture. The RGB-D images captured by low-cost sensors (Kinect) [4,5] extends the FER systems applicability into Human-Robot Interaction (HRI). Based on the properties of the image, the feature extraction is split up into two categories such as static image-based and the image sequence-based methods. Among them, the static methods utilize the fewer data to achieve the fast recognition whereas the sequence-based

Z. Zhang and K. Huang (Eds.): IVS 2016, CCIS 664, pp. 107–113, 2016.
DOI: 10.1007/978-981-10-3476-3_13

methods requires more data. The Gabor Wavelet Transform (GWT) [11] that extracts the features in two domains namely, spatial and frequency from the static image that leads to high dimension. The existence of facial variations, pose, illumination and the cultural change causes the performance degradation in FER systems. Hence, there is a need of large-scale data to overcome the problems in FER. The introduction section addresses the major issues in the traditional FER methods such as the large scale data and the detailed information analysis. The technical contributions of proposed PBDP-GAM are listed as follows:

1. The Patch-Based Diagonal Pattern (PBDP) proposed in this paper supports the reliable detection and tracking of facial points that increases the size of labeled pool.
2. The incorporation of PBDP on Geometric Appearance Models (GAM) and the co-training extract the facial features in the multi-direction and the complementary information learning.
3. The multi-directional feature extraction and the maximization of labeled database by PBDP-GAM supports the detailed information analysis.

2 Related Work

The capture of facial surface deformations is the necessary stage in FER systems and it suffers from illumination variations. Ghosh and Ari [6] utilized the Gray World (GW) algorithm to overcome the illumination variation from grayscale and color images. An accurate prediction of scene illumination variation depends on the hand-crafted features that degraded the performance. Convolutional Neural Network (CNN) and Differential Mean Curvature Map (DMCM) multithreading cascade of rotation-invariant HOG (McRiHOG) and Dynamic Bayesian Network (DBN) captured the facial interactions in different levels such as bottom-top and top-bottom [7]. The two-way facial feature tracking algorithms have the great influence on expression/Action Unit recognition performance. Principal Component Analysis (PCA), Gray Level Co-occurrence Matrix (GLCM) and Fuzzy-logic based Image Dimensionality Reduction using the Shape Primitives (FIDRSP) reduced the gray level with efficient recognition. The existence of redundant and irrelevant features increased the complexity and the computational cost in classification algorithms. Feature selection methods, time-series classification methods, Relevance Vector Machine (RVM) [8], Output-Associative RVM (OA-RVM) and Continuous Conditional Neural Fields (CCNF) [10] predicted the multi-dimensional output vectors for the specific features and the spatial-temporal dependencies inclusion affected the robustness adversely. Hence, there is a need of large scale database to analyze the expressions. The evolution of Co-training methods [13] improved the recognition performance with the large size templates utilization.

3 Patch-Based Diagonal Pattern on Geometric Appearance Model

Figure 1 shows the working flow of PBDP-GAM. Initially, the preprocessing stage comprises noise removal and skin pixel detection from input RGB images in the KinectDB. The Gaussian filter removes the noise in the images. The Viola-Jones [9] method detects the face from the input RGB images.

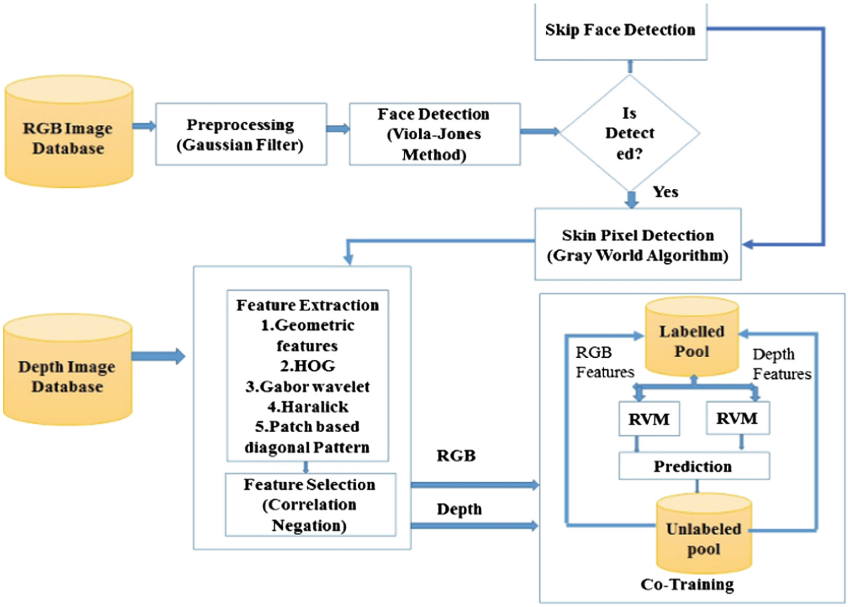

Fig. 1. Architecture of proposed method.

The GAM accepts the segmented skin pixels and converts them into binary to extract the top of the head pixel. Then, GAM predicts the nose, mouth, chin, and forehead from the top pixel by using distance measurement. The HOG, Gabor followed by GAM algorithm counts the gradient orientation occurrences for localized images and represents the image variations respectively. Then, we utilize the Haralick feature extraction for every 30-degree orientation that provides clear texture analysis. Based on diagonal pixel values, PBDP extracts image patterns. Initially, the image is divided into 3 * 3 matrix. The average of diagonal pixels is calculated. If the average pixel value is greater than the neighboring pixel, then the cell is filled with the value '1'. Otherwise, it is '0'. The count of ones and zeros decides the necessary patterns for RGB-D image. The likelihood estimation function in RVM classifier [8] identifies the facial expressions as neutral and smile for RGB-D images respectively. Figure 2(a)–(d) show the facial points detection. Figure 2(e)–(h) show the state of expression.

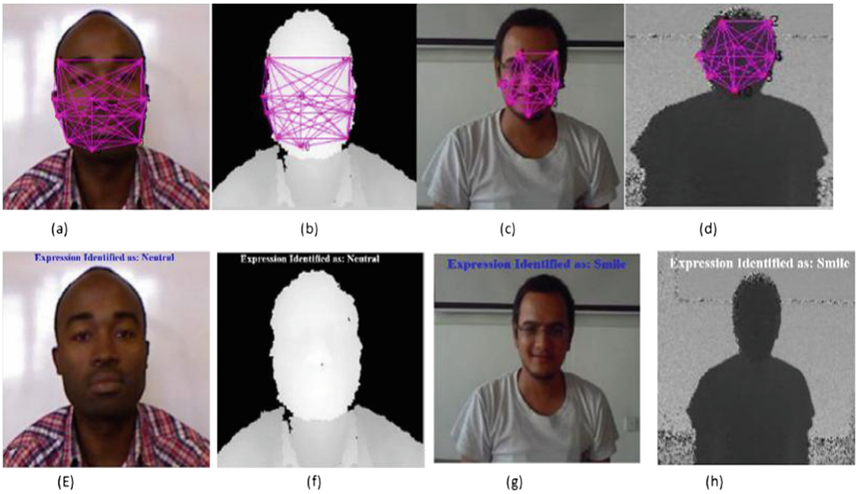

Fig. 2. Facial points detection and expression (neutral/smile state) recognition.

The integration of large unlabeled data with the small labeled data improves the RGB-D object recognition by a semi-supervised approach refers co-training [13]. The training of RGB and Depth classifier (C_{RGB}, C_{Depth}) with an independent feature set improves the large size unlabeled data learning. The probability scores of RGB-D classifier determines the input instance category with cross validation co-efficient (α) follows:

$$c = arg_{c_i \epsilon x} Max(\alpha P^{c_i}_{C_{RGB}} + (1-\alpha) P^{c_i}_{C_{Depth}}) \tag{1}$$

If the RVM classified result has the sufficient confidence then the corresponding images are added to the database.

4 Performance Analysis

The EURECOMM and biographer database validates the effectiveness of proposed algorithm. The EURECOM dataset [13] contains facial images of 52 peoples (14 females, 38 males) captured by Kinect sensor in two sessions at different time periods. The Biographer-RGB-D database contains facial images of 13 peoples (Chinese males with an age of fewer than 30 years with 25 images of each) captured by the softKinetic camera with 15 fps frame rate.

4.1 Performance Metrics

The comparative analysis of proposed PBDP-GAM without co-training, with co-training and SVM classifier regarding True Positive (TP), True Negative (TN), False Positive (FP), False negative (FN), accuracy, sensitivity, specificity, precision, recall (True Positive Rate (TPR), Jaccard coefficient, Dice overlap, Kappa

Table 1. Parametric equations

Parameters	Descriptions
True Positive (TP)	Cases of correct predictions of expressions
True Negative (TN)	Cases of correct predictions of no expressions
False Positive (FP)	Cases of incorrect predictions of expressions
False Negative (FN)	Cases of incorrect predictions of no expression.
Accuracy	(TP + TN)/(TP + TN + FP + FN)
Sensitivity	TP/(TP + FN)
Specificity	TN/(FP + TN)
Precision	TP/(TP + FP)
Recall	TP/(TP + FN)
Dice coefficient	Measure of similarity between the expression sets
Jaccard coefficient	Size of intersection of expression sets/size of union of expression sets
Kappa coefficient	$k = \frac{P_0 - P_e}{1 - P_e}$

Table 2. Performance analysis

Performance metrics	PBDP-SVM	PBDP-RVM without co-training	PBDP-RVM with co-training
Sensitivity	86.2069	94.8276	98.2759
Specificity	99.2908	98.5816	99.2908
Precision	96.1538	93.2203	96.6102
Recall	86.2069	94.8276	98.2759
Jaccard coeff	97.0588	97.9412	99.1176
Dice overlap	98.5075	98.9599	99.5569
Kappa coeff	0.8916	0.9277	0.969
Accuracy	97.0588	97.94	99.12

coefficient on Biographer RGB-D database shows the effectiveness of proposed work. The performance parameters specify how the proposed algorithm recognizes the facial expression of the images. Table 1 describes the equations used to evaluate the performance parameters. Table 2 shows the estimated values for each performance metrics. The RVM combined with the PBDP approach effectively reduces the computational cost due to the generalized form.

Figure 3 shows the recognition and error rate analysis for various expression states over the existing methods. The PBDP-GAM offers 2.17, 14.92, 2.57, 6.65 and 3.89% improvement in recognition rate for smile, OM, SG, OH and OP respectively compared to SANN method [12]. Similarly, PBDP-GAM reduces the error rate by 36.87, 25.12, 74.78, 25.21, 50.97 and 45.87% compared to SANN respectively.

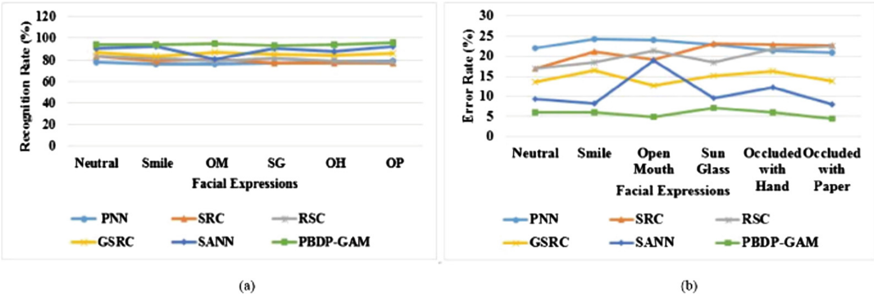

Fig. 3. (a) Recognition rate and (b) error rate analysis.

5 Conclusion

This paper addressed the limitation of real-time HMI-based applications such that the detailed information analysis that requires large scale labeled pool. The novel Patch-Based Diagonal Pattern (PBDP) on Geometric Appearance Models (GAM) are proposed that extracted the features on multiple direction for detailed information analysis. The co-training utilization created an effective large scale database with better identification and acceptance rate values. Finally, the application of Relevance Vector Machine (RVM) to the extracted features effectively classified the facial expression status. The experimental results of proposed PBDP-GAM show the efficient recognition performance over the existing methods regarding the recognition and the error rate values. Extension of this work includes the deep learning methods for 3D face FER system.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grants 61473290, the National High Technology Research and Development Program of China (863 Program) under Grant 2015AA042307, and the international partnership program of Chinese Academy of Sciences, Grant No. 173211KYSB20160008.

References

1. Bettadapura, V.: Face expression recognition, analysis: the state of the art. arXiv preprint arXiv:1203.6722 (2012)
2. Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L.: Static and dynamic 3D facial expression recognition: a comprehensive survey. Image Vis. Comput. **30**, 683–697 (2012)
3. Ramirez Rivera, A., Rojas, C., Oksam, C.: Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**, 1740–1752 (2013)
4. Goswami, G., Bharadwaj, S., Vatsa, M., Singh, R.: On RGB-D face recognition using Kinect. In: IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–6 (2013)

5. Romero, P., Cid, F., Núnez, P.: A novel real time facial expression recognition system based on Candide-3 reconstruction model. In: Proceedings of the XIV Workshop on Physical Agents (WAF 2013), Madrid, Spain, pp. 18–19 (2013)
6. Ghosh, D., Ari, S.: Static Hand Gesture Recognition Using Mixture of Features and SVM Classifier (2015)
7. Bianco, S., Cusano, C., Schettini, R.: Color constancy using CNNs. arXiv preprint arXiv:1504.04548 (2015)
8. Lemaire, P., Ardabilian, M., Liming, C., Daoudi, M.: Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–7 (2013)
9. El Maghraby, A., Abdalla, M., Enany, O., El Nahas, M.Y.: Detect and analyze face parts information using Viola-Jones and geometric approaches. Int. J. Comput. Appl. **101**, 23–28 (2014)
10. Baltrušaitis, T., Robinson, P., Morency, L.-P.: Continuous conditional neural fields for structured regression. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 593–608. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10593-2_39
11. Ruan, J., Yin, J., Chen, Q., Chen, G.: Facial expression recognition based on gabor wavelet transform and relevance vector machine. J. Inf. Comput. Sci. **11**, 295–302 (2014)
12. Wong, Y., Harandi, M.T., Sanderson, C.: On robust face recognition via sparse coding: the good the bad and the ugly. IET Biom. **3**, 176–189 (2014)
13. Rui, M., Kose, N., Dugelay, J.L.: KinectFaceDB: a kinect database for face recognition. IEEE Trans. Syst. Man Cybern.: Syst. **44**, 1534–1548 (2014)