Building the Facial Expressions Recognition System Based on RGB-D Images in High Performance

Trung Truong[1(✉)] and Ngoc Ly[2]

[1] Faculty of Information Technology, University of Science, VNU-HCMC, Ho Chi Minh City, Vietnam
sirquangtrung@gmail.com

[2] Faculty of Information Technology, Computer Vision and Robotics, University of Science, VNU-HCMC, Ho Chi Minh City, Vietnam
lqngoc@fit.hcmus.edu.vn

Abstract. In this paper, we propose a novel idea for automatic facial expression analysis with the aim of resolving the existing challenges in 2D images. The subtle combination of the geometry-based method with the appearance-based features in depth and color images contributes to increasing in distinguishable features among various facial expressions. Particular functions are utilised to calculate the correlation between expressions in order to determine the exact facial expression. Our approach consists of a sequence of steps including estimating the normal vector of facial surface, then extracting the geometric features such as the orientation of normal vector in the point cloud. The useful color information is known as LBP. According to the result of the experiment, we demonstrate that the effective fusion scheme of texture and shape feature on color and depth images. In comparison with the non fusion scheme, our fusion scheme has resulted in the increase of recognition under low and high illuminated light, about 19.84 % and 1.59 %, respectively.

Keywords: Facial expression recognition · Local binary pattern · Normal vector · Covariance matrix · RGBD datasets

1 Introduction

Facial expression analysis plays an important role in daily communication, which contributes to the success in conveying semantic information of the speaker's face when chatting. Therefore, the development of automatic facial expression recognition will be extremely useful. For instance, it will facilitate the human-machine interaction, security and health care, etc. However, there are still many challenging problems in recognizing facial expressions are not yet completely solved such as illumination, pose, occlusion. Therefore, in this paper, we apply new algorithm to solve part of the problems.

Two popular approaches for facial expression in static image can be classiffied as geometric feature-based method and appearance feature-based method.

N.T. Nguyen et al. (Eds.): ACIIDS 2016, Part II, LNAI 9622, pp. 377–387, 2016.
DOI: 10.1007/978-3-662-49390-8_37

Specifically, geometry-based method is the shape feature, which involves calculating the distance, the angle of the particular vector as normal vector on the surface of the face. Recently, geometric feature-based method has been utilized as ASM proposed by Shbib et al. [17]. Besides, the appearance-based method represents the texture feature of another change of facial expression. They are not lighting and direction invariant features, but increase discriminative characteristics in analyzing the expression in the case of good lighting conditions. Appearance-based method includes LBP like S. Guo et al. [6]. Features on dynamic images or video have tended to exact feature which typically rely on motion feature or moving space-time information such as Liu et al. [10].

Limitations of the above methods only evaluate facial expression feature on particular lighting condition, except when various challenges in case of lighting changes or pose will be no longer in the effective area. The data are collected according to traditional methods caused in the loss of the 3D surface information. Therefore it is difficult for many computer vision tasks such as object recognition, detection, identification, tracking, scene reconstruction, and motion analysis. The techniques used for exploiting depth map are geometric feature-based method and appearance feature-based method. Geometry-based method has also been studied by Malawski et al. [11]. Nonetheless, it has faced with the noise and low resolution from dataset induced to fail approach proposed. In terms of a specific application of MS-LNP on 3D models as Hengliang Tang et al. [8], they exploited multiple global histograms of local normal patterns from multiple normal components and multiple binary encoding scales for the classification. However, the 3D model is relatively less noisy, more smooth compared with the point cloud set derived from depth map.

In this work, we introduce a novel algorithm for automatic human facial expression recognition. There are limitations in our approach since it has not been done yet to conduct the experiments in the conditions of occlusion and pose. The main contributions of this paper are summarized as follows:

- We empirically evaluate facial representation based on statistical features such as LBP and normal information derived 3D facial surface. We focus on exploring the characteristic of normal vector to address problems that geometric feature does not affect to the challenges such as color channels. This will help to distinguish the changing expression shape.
- We address many challenges: diverse age, gender, skins, lighting changes even dramatically illumination. Our experimental result proves the truth that in good illumination the texture feature performs the recognition of human facial expression best. However, we strongly show under different or bad lighting conditions that we fuse both features of color and depth image, the features derived from the depth image will contribute to the ineffectively recognized color features to increase the facial expression recognition result.
- Utilizing covariance matrix for feature representation will bring promising results to contribute to increasing the discrimination and emerging the different characteristic in numeric data that is strong feature for analyzing the facial expression.

The remainder of the paper is organized as follows. Section 2 reviews related works. Our approach will be presented in details in Sect. 3. Section 4 will be devoted to experimental results that show the high performance of our proposed method. In Sect. 5, we draw conclusions of our work and indicate our future works.

2 Related Works

Facial expression recognition has been an active research topic during the last few decades. However, there are still scientific challenges and potential applications. In pursuit of using color images, several feature extraction methods used in the face image are given such as Gabor filter Y. Pang et al. [14], LBP [6]. Among these features, [14] proposed Gabor filter for feature extraction in face recognition. It produces a large number of features from extracting multi-scale and multi-orientation coefficients. However, the computational complexity is expensive due to face images with multi-banks of Gabor filter. Some researchers have improved LBP features in order to reduce the dimension computational time and to achieve higher accuracy than Gabor feature such as [6].

However, robust recognition with conventional 2D cameras is still not possible in real conditions, in the presence of unexpected illumination, occlusion and pose variations. To deal with these problems, some researchers proposed a histogram of mesh gradient (HoG) and histogram of shape index (HoS) [9], the Scale-invariant feature transform (SIFT) [1], Local normal pattern (LNP) [8] following the 3D face models with the support of expensive specialized sensor. These challenges are completely solved because the approach to 3D model is not affected by these problems in comparison with those of the 2D face image. In addition, [9] suggested an idea to calculate the weighted statistical distributions of surface differential quantities, including HoG and HoS based on curvature estimation method. However, the fusion schemes are quite simple and a set of landmarks is selected by manual landmarks. Therefore, it is quite expensive and manually selected landmarks in the database that leads to higher accuracy of the experiment. [8] proposed a 3D face model-only method that represents multiple global histograms of local normal patterns from multiple normal components and binary encoding scales. These features are on 3D facial geometry, thus achieving high accuracy. 3D scanner remains a limited application of expression recognition in real life because the sensor is not cheap and has slow acquisition speed.

On the contrary, some new 3D sensors such as the Kinect sensor have low cost but gain high speed. Some researchers proposed a new method using low quality 3D data (RGB-D) such as the approach of surface curvature measurements [7,16]. Billy Y.L. Li et al. [7] proposed shape and texture features for face recognition and in the experiment, fusing depth and texture images to improves the performance more considerably than both depth image-only method and color image-only method. Also, [16] proposed a mean curvature based on low quality of 3D face data. Moreover, the accuracy of fusion of 3D and 2D method is higher than that of 3D-only method and 2D only-method. Introduced as a proposal of Duc Fehr et al. [4], several features using covariance descriptors provide a low dimensional representation of the data for object recognition. On the other hand, representation of local feature based on covariance matrix, which has succeeded in the case of not only discrimination ability, but also facial expression recognition, is offered from [6,14].

3 Our Approach

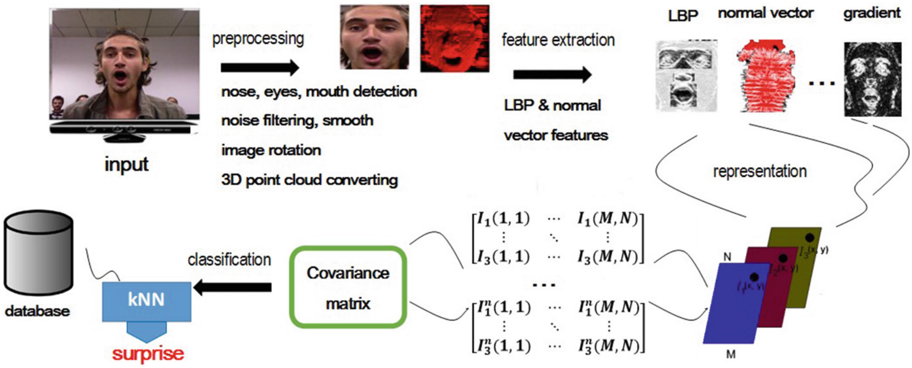

Fig. 1. Conceptual diagram of the proposed system

3.1 Pre-processing

We apply an adequate filter in order to normalize intensity of color image and approximate surface of point cloud, which affects the accuracy of the system due to both noise, holes in depth images and highly contrasted intensity of pixels in color images. We make algorithms for point cloud preprocessing with the aim of removing outliers known as an abnormal point near to hair, ear..., filling holes as well as robust smooth surfaces. To extract many vision information of images, holes existing on depth images are firstly filled by applying morphological closing. Besides, the histogram equalization can also be used to normalize the color images. After that, the face images are aligned at the same position. Then, some sub-regions of eyes, nose, and mouth have cropped the color face image to provide the most crucial facial expression feature. Finally, the depth images are converted into 3D point clouds (Fig. 1).

3.2 Feature Description

Local Binary Pattern on Color Image. The LBP operator is designed for texture descriptor and also, has been proved to be highly discriminative. It was introduced by Ojala et al. [13]. Each pixel in LBP image could be constructed in the way to threshold its neighborhood in the original image and then having the result as a binary number. By doing this, the pixel of LBP image is presented a decimal number, which is converting its the binary number. Formally, the LBP operator takes the form:

$$LBP_{8,1}(x_c, y_c) = \sum_{i=0}^{7} s(I_i - I_c) * 2^i \tag{1}$$

Where I_c, I_i are the grey level values at c, i. In this case, i is the label of parts around the center pixel location (x_c, y_c) and $s(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0. \end{cases}$

By doing this, the LBP operator become the technique of simplicity. To enhance the texture descriptor, we present some extensions of LBP operator including using neighborhoods of different sizes extended LBP and multi scale LBPs.

Using Neighborhoods of Different Sizes Extended LBP. [6] proposed $LBCM_{8,2}$, $LBCM_{16,2}$ to use neighborhoods of different sizes for each pixel in the whole facial image aim to recognize facial expression. Additionally, in their experiment, the LBP method has acquired the recognition precision rate higher than other methods. In this paper, we present the extended LBP operator to use 8 neighborhoods of size as 2, to capture dominant features from visually salient facial region including eyes, nose, and mouth. Notation $LBP_{8,2}$ is symbolic for our approach.

Multi-scale LBPs. Multi-scale LBPs considers the quantity of the neighborhood and the scale coefficient, which affect more the captured detail at various scales. Furthermore, it has expected to outperform other standard LBP operator because it encodes the micro structures of the facial expression but also provides a more extensive description than the standard LBP operator. By doing this, it enhances the discriminating ability for strong suitability in facial expression recognition. The approach can increase the computational cost of each LBP image. Otherwise, we propose a different approach to multi scale analysis using LBPs, which is defined as $LBP_1^{ms}(8, R)$ with R = 1..8. And $LBP_2^{ms}(8, R)$, LBP (8, R) with R=1, 2 and LBP(16,R) with R=2, 3, 4.

Normal Estimation on Point Clouds. In fact, the normal description which brings good results in processing 3D face model [8] is also successful in exploring in terms of 3D point cloud with many noises e.g. [15]. As a result, it has been expected to not only improve the discriminating power but capture more informative geometric information as well. Furthermore, all points in point cloud have already exacted the normal vector such variations in 3D point cloud [15]. To set up surface normal in a point cloud has the pseudo code known as Algorithm 1.

As a result, it has been expected to not only improve the discriminating power and robustness in facial expression but also capture more informative geometric information as well. Furthermore, all points in point cloud have already exacted the normal vector such variations in 3D point cloud.

Algorithm 1. Setting up surface normal in a point cloud

```
1: Input: 3D point cloud
2: Output: Normal vector description
3: procedure EstimateSurfaceNormal
4:     P // 3D point cloud
5:     k = 20 // number of nearest neighbors of point
6:     for p in P do
7:         Create the surface sampled around p: P_p = {p_1, p_2, .., p_k}
8:         Estimate the approximated normal vector n to P_p: n = f(P_p)
9:     end for
10:    Select the correct scale of all normal vector n
11: end procedure
```

3.3 Covariance Descriptor

The covariance descriptor was proposed by Duc Fehr et al. [4] for object detection and classification. Furthermore, Yanwei Pang et al. [14] also proposed face recognition using the covariance descriptor. We propose calculating covariance matrix of the given point cloud include RGB-XYZ channels, which have been constructing the data representation and denotes $F(x, y) = \phi(I, D, x, y) = z_i$.

For texture classification, we define the mapping function F(x,y)=(I,x,y) in color image as:

$$F_3 = \left[x\ y\ I\ I_x\ I_y\ \frac{I_x}{I_y}\ LBP(8,2) \right]. \tag{2}$$

Otherwise, for shape classification, we define the mapping function $F(x, y) = \phi(D, x, y)$ in point cloud as:

$$F_{12} = [x\ y\ n_x\ n_y\ n_z] \tag{3}$$

Finally, we propose to combine color and depth image the mapping function $F(x, y) = \phi(I, D, x, y)$ as

$$F_{Fusion} = \left[x\ y\ I\ I_x\ I_y\ \frac{I_x}{I_y}\ LBP(8,2)\ n_x\ n_y\ n_z \right]. \tag{4}$$

Förstner and Moonen [5] claimed that covariance matrices do not adhere to the Euclidean geometry, but span a Riemannian space with an associated Riemannian distance metric. Therefore, we present the distance function for calculating the difference between classes. One metric that can be used, the geodesic distance, has been introduced by [5,14] and is defined as follows:

$$\rho(C^G, C^P) = \sum_{i=1}^{12} \rho(C_i^G, C_i^P) - max[\rho(C_i^G, C_i^P)], \tag{5}$$

where $\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{d} ln^2 \lambda_i(C_1, C_2)}$; C_i^G, C_i^P are sub-regions, located in the same regions such as eyes, nose and mouth regions. $\{\lambda_i(C_1, C_2) | i = 1, 2, .., d\}$ are generalized eigenvalues of C_1 and C_2 in $\lambda_1 * C_1 * u_1 = C_2 * u_2$.

4 Experiment

4.1 Database

In our experiments, we have used the face Warehouse database which was introduced by Cao et al. in [2]. Recall that Face Warehouse database, the facial expression database, is a sort of RGBD data based on Kinect camera. In addition, on the color face database, we have created 3 subsets, including the original subset, the distorting for dark and light of the original subset.

4.2 Experimental Results

For our experimental system, 4 main steps were used to recognize the facial expression, including preprocessing, feature extraction, feature representation and classification. In addition, to perform the preprocessing for both color and depth images, the first step is that we approach the STASM library automatically proposed by Tim Cootes [3] to display the landmarks of a face in a color image, i.e. eyes, nose, mouth, thus cropped 3 regions from the face image providing crucial facial expression information for expression recognition. Above all, it is linked with reducing the number of features. The extracted face image is accounted for the large size of features and then it is represented feature to adopt the formula of $d^2 * n^2 * 3$ where, d is a quantity of kinds of feature combined with our method, n denotes the number of sub-regions cropped from eyes, nose, mouth regions. We have done two trends of conducting an experiment included in early fusion and late fusion of technique of 2D and 3D processing. But the result shows that the early fusion of 2D and 3D images achieve much higher than what observed on the later fusion. In our proposed method, the number of features on face color image, depth image and fusion of color and depth image in the datasets are 588, 300, 1400, respectively. Finally, in order to classify facial expression, it is based on the k-nearest neighbor. We perform a number of experiments to choose the parameter of k that is appropriate for accurate performance. To put it another way, the range value of k nearest neighbor is 1–20, thus the maximum accuracy is achieved whenever k = 9 to 14 in comparison to the others. All experiments are approached with 3-fold, except for the Table 3 which approached with leave-one-out cross-validation.

Experiment 1: Normalized Illumination. We show several types of features and also compare the accuracy of these types of feature. This allows us to find the feature descriptor which improves the facial expression recognition. All of the proposed 2D-only methods, F_4 feature achieves 84.13 % average recognition rate to be highest accuracy as shown in Table 1. In addition, F_4 feature consists of 588 dimensions. Classification rates achieved by F_1,F_2,F_3 are 81.75 %, 80.16 %, 81.75 %, respectively. Although, the number of feature of F_2 is 2352 features more than the number feature of F_3 account for 1252 but its classification rate is less rate than F_3. In the same way, the number of feature of F_1 (192 features) is less in comparison with F_3 (1452 features), but the accuracy rates are in contrast to the quantity of number of features. Comparing on classification rate on F_4, it shows that the size of feature vector is middle size range but its features is effectively improvement. Maximum curvature values: P_{max}, Minimum curvature values: P_{min}, Mean curvature:H, Gaussian curvature values:K, Shape index value: S. $F_1 = [x\ y\ I\ LBP(8,2)]$, $F_2 = \left[x\ y\ I\ I_x\ I_y\ \frac{I_x}{I_y}\ LBP_1^{ms}\right]$, $F_3 = \left[x\ y\ I\ I_x\ I_y\ \frac{I_x}{I_y}\ LBP_2^{ms}\right]$, $F_4 = \left[x\ y\ I\ I_x\ I_y\ \frac{I_x}{I_y}\ LBP(8,2)\right]$, $F_6 = [x\ y\ S]$, $F_5 = [H\ K\ S\ P_{min}\ P_{max}]$, $F_7 = [x\ y\ n_x\ n_y\ n_z\ P_{min}\ P_{max}]$, $F_8 = [x\ y\ n_x\ n_y\ n_z\ K]$, $F_9 = [n_x\ n_y\ n_z]$, $F_{10} = [x\ y\ n_x\ n_y\ n_z\ H]$, $F_{11} = [x\ y\ n_x\ n_y\ n_z\ S]$, $F_{12} = [x\ y\ n_x\ n_y\ n_z]$.

Also, Table 3 shows the several results of many 3D point cloud method, including $F_{5..11}$ on the corresponding range of 47.62 % to 69.05 %. It combines shape index, maximum curvature, minimum curvature, mean curvature, Gaussian curvature values with normal vector and location. Besides, the important result is the fact that F_{12} achieves 72.22 % on the average and has 300 features, which has a higher accuracy than the curvature features in Table 2. To put another way, the number of features of F_{12} is less size than most of the feature of $F_{5..11}$. Here, both F_{12} feature and other curvature features are known as the 3D point cloud-only method.

Table 1. Recognition rate (%) for original of FaceWarehouse dataset with color images-only under good illumination conditions). The 2D features utilize the LBP feature.

Features	F_1	F_2	F_3	F_4
Accuracy	80.95	80.16	81.75	84.13

Table 2. Recognition rate (%) for original of FaceWarehouse dataset with depth images-only under good illumination conditions

Features	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
Accuracy	47.62	62.70	66.67	68.25	62.70	62.70	69.05	72.22

Table 3. Recognition rate (%) for recognition facial expression in 3-Fold and leave-one-out cross-validation.

Datasets	Accuracy	
	3-fold	leave-one-out cross-validation
RGB dataset	84.13	87.3
Depth dataset	72.22	75.40
RGBD dataset	78.57	80.16

Table 4. Recognition rate (%) for recognition facial expression in 3-Fold and leave-one-out cross-validation

Datasets	Accuracy	
	Low illumination	High illumination
RGB dataset	53.97	74.98
Depth dataset	72.22	72.22
RGBD dataset	73.81	76.57

Experiment 2: Low, High Illumination. In these experiments, the color images are only distorted the light channel in order to closely resemble the real conditions, whereas the depth images are not distorted. As before, we do this experiment by fusing the color image and depth image in order to show the improvement of recognition performance in the real conditions when using our approach.

The most important result is the fact that, as shown in Table 4, the low and high illumination increase 19.84 % and 1.59 % average recognition rates, respectively, in the case of the fusion of color image and depth image method in comparison with the color image-only method. Indeed, this shows the fact that when using our recognition proposal, the fusion of color images and depth images gives a better result than the infusions under unexpected illumination, and namely the result approached the color image only method. Hence, it brings the performance of results what it is serious challenges in previous time.

Table 5. Recognition rate (%) for performance comparison of different methods with depth images- only method

Approach	Method	No. Express	Accuracy(%)
Mao [12]	AUs, FPPs, FPs	4	<72
Proposed method	**Normal descriptor, kNN**	**6**	**72.22**

Comparative Study. Table 5 compares two results approached different methods based on depth images of Facewarehouse database. The criteria for comparison of performing consist of the number of expressions, the recognition rate. Although the amount of emotion is different, experiments show that the result of our proposed approach is higher than its Mao [12]. Thus, it figures out our potential actually approach compete in the facial expression recognition.

5 Conclusion and Future Works

In summary, we have presented a novel descriptor and framework for automatic facial expression recognition in real conditions based on the fusion of both color image and depth image. In addition, the data representation is relied on the covariance matrix which depends on the number of features without the size of the region. We constructed the covariance matrices with LBP in color image and normal vector in facial point cloud. Therefore, our final descriptor brings both texture feature and shape feature contributing to the increase in the discrimination of emotional facial expressions. In specific, we indicated that the fusion of depth with color images method improves the accuracy in comparison with the depth image-only and color image only performances under low and high illumination condition. It is true that under bad illumination condition, the accuracy of the depth image-only method is not affected, whereas the accuracy of color image-only method is decreasing. The fusion of depth in cooperation with color images which has the less feature recognition in real conditions (low and high illumination), has the mutual support to increase the precision. This implies that this method is better than the color image-only method or depth image-only method.

We leave as future research the study of experiments which are based on an extension of our proposed framework in the challenge of head pose. Furthermore, the proposed framework used in this paper will be investigated for a sequence of images.

Acknowledgments. This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number B2014-18-02.

References

1. Berretti, S., Amor, B.B., Daoudi, M., Del Bimbo, A.: 3D facial expression recognition using sift descriptors of automatically detected keypoints. Int. J. Vis. Comput. **27**(11), 1021–1036 (2011)
2. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: a 3D facial expression database for visual computing. IEEE Trans. J. Vis. Comput. Graph. **20**(3), 413–425 (2014)
3. Tim, C.: Model-based methods in analysis of biomedical images. In: Baldock, R., Graham, J. (eds.) Image Processing and Analysis, pp. 223-247. Oxford University Press (2000)
4. Fehr, D.A.: Covariance based point cloud descriptors for object detection and classification. University of Minnosita, August 2013
5. Förstner, W., Moonen, B.: A metric for covariance matrices. In: Grafarend, E.W., Krumm, F.W., Schwarze, V.S. (eds.) Journal of Geodesy-The Challenge of the 3rd Millennium, pp. 299–309. Springer, Heidelberg (2003)
6. Guo, S., Ruan, Q.: Facial expression recognition using local binary covariance matrices. In: Proceedings of 4th IET International Conference on Wireless, Mobile and Multimedia Networks (ICWMMN 2011) (2011)
7. Li, B.Y., Mian, A.S., Liu, W., Krishna, A.: Using kinect for face recognition under varying poses, expressions, illumination and disguise. In: 2013 IEEE Workshop on Proceedings of Applications of Computer Vision (WACV), pp. 186–192. IEEE (2013)
8. Li, H., Chen, L., Huang, D., Wang, Y., Morvan, J.: 3D facial expression recognition via multiple kernel learning of multi-scale local normal patterns. In: 2012 21st International Conference on Proceedings of Pattern Recognition (ICPR), pp. 2577–2580. IEEE (2012)

9. Li, H., Morvan, J.-M., Chen, L.: 3D facial expression recognition based on histograms of surface differential quantities. In: Blanc-Talon, J., Kleihorst, R., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2011. LNCS, vol. 6915, pp. 483–494. Springer, Heidelberg (2011)
10. Liu, M., Shan, S., Wang, R., Chen, X.: Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In: IEEE Conference on Proceedings of Computer Vision and Pattern Recognition, pp. 1749–1756. IEEE (2014)
11. Malawski, F., Kwolek, B., Sako, S.: Using kinect for facial expression recognition under varying poses and illumination. In: Ślęzak, D., Schaefer, G., Vuong, S.T., Kim, Y.-S. (eds.) AMT 2014. LNCS, vol. 8610, pp. 395–406. Springer, Heidelberg (2014)
12. Mao, Q., Pan, X., Zhan, Y., Shen, X.: Using kinect for real-time emotion recognition via facial expressions. J. Front. Inf. Tech. Electr. Eng. **16**, 272–282 (2015)
13. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. J. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
14. Pang, Y., Yuan, Y., Li, X.: Gabor-based region covariance matrices for face recognition. J. IEEE Trans. Circuits Syst. Video Technol. **18**(7), 989–993 (2008)
15. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: IEEE International Conference on Proceedings of Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011, pp. 1–4. IEEE (2011)
16. Savran, A., Gur, R., Verma, R.: Automatic detection of emotion valence on faces using consumer depth cameras. In: 2013 IEEE International Conference on Proceedings of Computer Vision Workshops (ICCVW), pp. 75–82. IEEE (2013)
17. Shbib, R., Zhou, S.: Facial expression analysis using active shape model. J. Signal Process. Image Process. Pattern Recogn. **8**(1), 9–22 (2015)