Fully Automatic 3D Facial Expression Recognition using Local Depth Features

Mingliang Xue
Department of Computing, Curtin University
Kent Street, Bentley WA 6102, Australia
mingliang.xue@postgrad.curtin.edu.au

Ajmal Mian
Computer Science and Software Engineering
The University of Western Australia
35 Stirling Highway, Crawley WA 6009, Australia
ajmal@csse.uwa.edu.au

Wanquan Liu
Department of Computing, Curtin University
Kent Street, Bentley WA 6102, Australia
W.Liu@curtin.edu.au

Ling Li
Department of Computing, Curtin University
Kent Street, Bentley WA 6102, Australia
L.Li@curtin.edu.au

Abstract

Facial expressions form a significant part of our nonverbal communications and understanding them is essential for effective human computer interaction. Due to the diversity of facial geometry and expressions, automatic expression recognition is a challenging task. This paper deals with the problem of person-independent facial expression recognition from a single 3D scan. We consider only the 3D shape because facial expressions are mostly encoded in facial geometry deformations rather than textures. Unlike the majority of existing works, our method is fully automatic including the detection of landmarks. We detect the four eye corners and nose tip in real time on the depth image and its gradients using Haar-like features and AdaBoost classifier. From these five points, another 25 heuristic points are defined to extract local depth features for representing facial expressions. The depth features are projected to a lower dimensional linear subspace where feature selection is performed by maximizing their relevance and minimizing their redundancy. The selected features are then used to train a multi-class SVM for the final classification. Experiments on the benchmark BU-3DFE database show that the proposed method outperforms existing automatic techniques, and is comparable even to the approaches using manual landmarks.

1. Introduction

Human computer, or man machine, interaction based on verbal commands and hand gestures have become very common in modern televisions and gaming consoles such as the X-Box One. However, facial expression recognition is still a challenging task for computers. Future human computer interfaces are envisioned to operate in human-centered intelligent ways where computers are able to respond spontaneously to human interactive behavior rather than waiting for rigid commands. Facial expressions constitute an important part of human behavior and social interaction. They are nonverbal means of communication and reveal information about a person's emotions, preferences and intensions. However, due to the subtlety, complexity and diversity of facial expressions and inter person facial differences, automatic facial expression recognition is a challenging task.

Although there could be more categories, six universal expressions are commonly defined in the literature [4], namely anger, disgust, fear, sadness, happiness and surprise. Prior research has mainly focused on 2D images or videos [10] but recently 3D facial expression recognition [12] has gained popularity due to its invariance to pose and illumination. This paper focuses on static 3D facial expressions. Although current sensors can acquire 3D videos in real time, we do not consider 3D videos because we are interested to maximize performance on a single frame. Temporal information will be considered in our future work.

Feature extraction from face data plays a crucial role in recognition systems. Many methods analyze 3D faces by extracting local patches around manually labeled landmarks to achieve good expression recognition performance [18][8]. In [18], local patches around landmarks are approximated with polynomial surfaces and principle curvatures are extracted by eigenvalue decomposition. They report an average accuracy of 83.6%. More recently, [8] proposed to define a patch in terms of concentric geodesic rings that follow the curvature of the facial surface and calculated the Riemannian distance between corresponding rings of patches on the test face against a reference scan face. They report impressive results of 98.8% average accuracy. Surprisingly,

这是什么？同心圆？
concentric geodesic rings

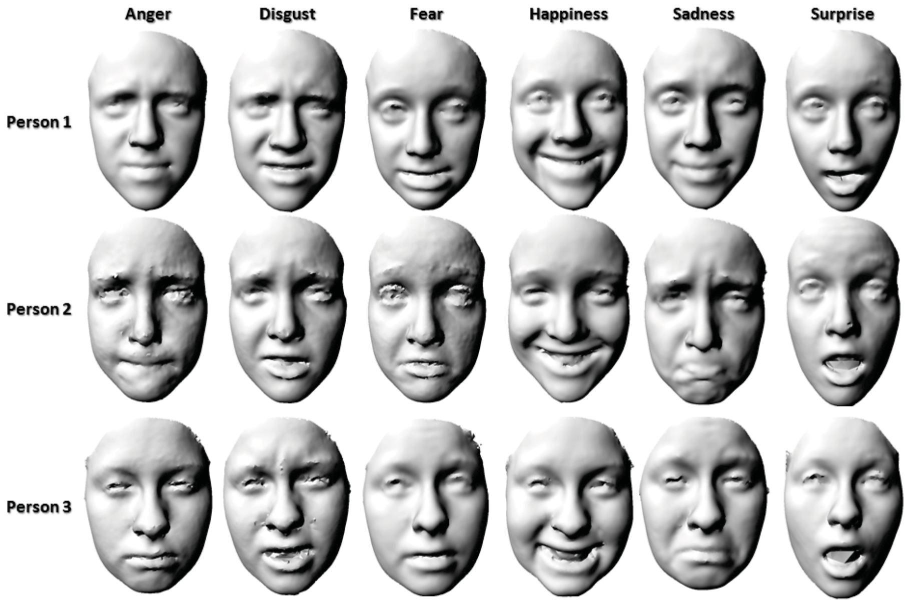


Figure 1. Facial expression examples from the BU-3DFE database.

this performance is even better than the human performance [20] on the same BU-3DFE database.

For all practical applications, facial expression recognition must be fully automatic. It is easier to manually label the expression, which is the required final outcome of the process, than to manually label multiple landmarks on a face. However, the above methods define patches around *manually* labeled landmarks that can be consistently located across faces and expressions. It is still an open problem to automatically recognize expressions without manual reference landmarks. The authors in [7] attempt automatic facial expression recognition by extracting whole-face DM-CM features that capture facial surface deformations caused by expressions without using facial landmarks. They report an average recognition rate of 78.1%. Since they used the entire face, their features include face regions that are not relevant to particular expressions. In contrast, patch-based approaches work well as they can be located specifically on important landmarks such as the mouth, cheeks and eyes and different set of features can be defined for each patch.

We propose a fully automatic facial expression recognition algorithm based on depth features extracted from local patches. In order to define local patches without human intervention, we detect the nose tip and four eye corners automatically as five fiducial landmark points. From these, another 25 heuristic landmarks are generated and local depth features are extracted from patches around all the 30 landmarks. Then, mutually exclusive features which jointly have the largest characterizing power are selected from the extracted depth features using mRMR (maximum Relevance Minimum Redundancy) [11]. Feature selection is a critical step as depth features contribute differently to each type of expressrion. Moreover, it also takes care of errors in landmark locations. We seek to use a similar approach to that of [8] due to their very high recognition rates, however their concentric geodesic rings cannot be segmented to facilitate feature selection. Hence we instead utilize a discrete sampling of the depth patch as our features. Finally, the selected features are fed to a SVM classifier for expression classification.

Depth Features对不同的表情
产生的作用不同

2. Pre-processing

The raw 3D faces in BU-3DFE [20] are noisy and have minor pose variations as shown in Figure 1. As illustrated in Figure 2, we preprocess the faces before feature extraction. At first 5 fiducial points are automatically detected, followed by registration of facial point clouds. Heuristic points are generated for feature extraction. In fiducial point detection, the nose tip and four eye corners are located by a Haar detector and AdaBoost classifier [17] which enables the proposed method to be fully automatic. Then the 3D facial point clouds are aligned and registered according to a T-area located on each face using the five fiducial points. We use only the T-area since it is not very sensitive to noise. On each of the registered faces, 30 heuristic points are generated based on the 5 fiducial points to extract depth features.

注册点
云，这个
怎样理
解？

2.1. Realtime fiducial points detection

Automatic landmark detection on 3D faces is still an open problem due to the significant topology changes caused by expressions, such as opening mouth in surprise. We notice that features vary very little around some points when expression changes, such as the four eye corners and

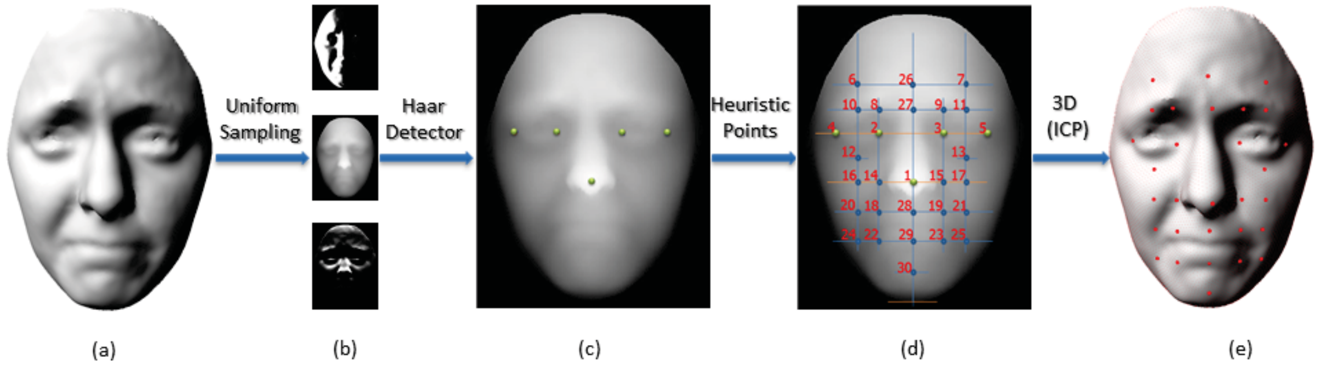


Figure 2. Pre-processing of a 3D face. (a) Original 3D face; (b) Range image and its x and y gradients rendered from 3D face; (c) Detected 5 fiducial points; (d) Generating heuristic points on range image; (e) Locating heuristic points on 3D face.

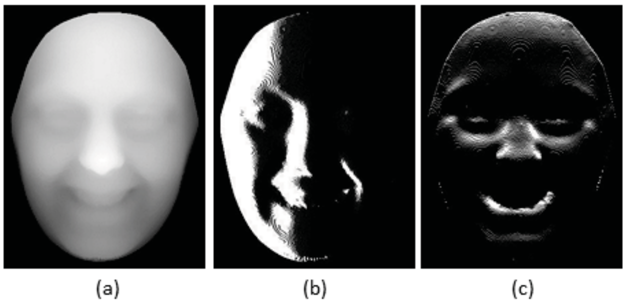


Figure 3. Example range image and its x and y gradients.

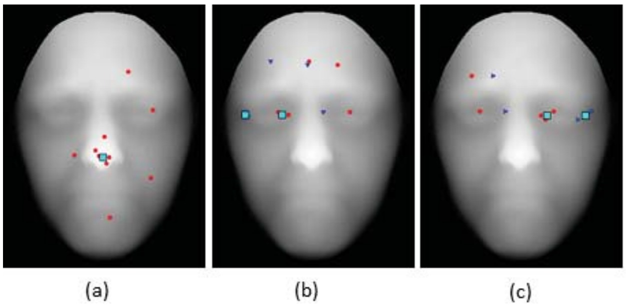


Figure 4. Demonstration of fiducial point detection. Small dots are candidates and large dots represent the final detections.

这两个方向的梯度是怎样得到的？

nose tip. Our realtime detection method detects these five points on a 3D facial surface. These five fiducial points and their relative distances are used to generate another 25 heuristic points on the face.

We use the Haar-cascade classifiers [9], which are based on the AdaBoost algorithm used for face detection [17]. Given a 3D face as shown in Figure 2(a), the surface is uniformly sampled by a grid in the x, y-plane, and the depth information (z-direction) is encoded in a range image. The resulting range image and its x and y gradients (see Figure 3) are used to train the Haar cascade classifiers. For each point, the detector returns several candidate locations. The facial structure and relative location relationships between eyes and nose tip are utilized to remove the outliers and identify the correct eyes and nose clusters. Note that the detector in [9] is in fact trained to find the eye centers and horizontal scale. The scale is then used to localize the outer eye corner. We extend the same idea to additionally localize the inner eye corners as well and thus detect five points rather than three as reported in [9]. The process is illustrated in Figure 4.

什么是range image？

如何利用这两个方向的梯度来训练分类器？

这里面对scale的计算适用于所有脸型吗？

We run our detection on all the 2500 faces in BU-3DFE database and the average detection time is recorded in Table 1. The total detection time for one face is less than 130 ms. Furthermore, since the BU-3DFE database provides manually labeled ground truth locations for the four eye corners, the distance from the detected location to the corresponding ground truth is calculated and illustrated in Figure 5, which shows that 90% of the detection errors are less than 4mm.

这是什么意思?

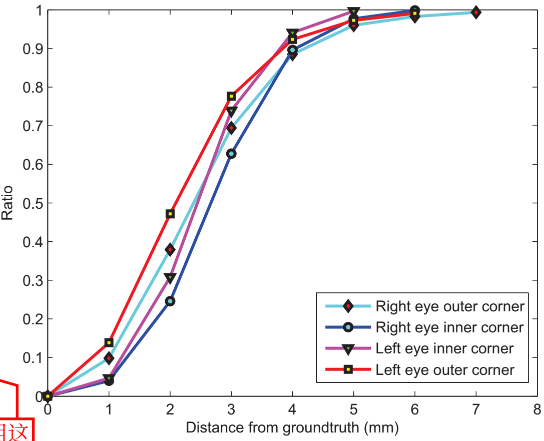
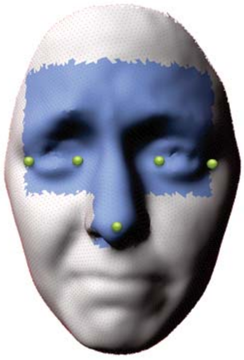


Figure 5. Detection error of the eye corners.

Table 1. Detection time of fiducial points.

Feature Point	Detection time
Nose	69.27 ms
Left eye	27.09 ms
Right eye	27.38 ms
Total	123.74 ms



参考：三维图像配准

Figure 6. T-area for registration.

为什么要注册？

迭代最近点算法 Iterative Closest Points

为什么是two faces？是不是一个face为standard face，另一个face为当前注册的face？这两个face之间有什么关联？

2.2. Registration

Minor pose variation exists in the BU-3DFE. Thus, it is necessary to register the faces against a 'standard' face. In the proposed method, all the faces are registered to the first female's neutral face in the BU-3DFE database by the Iterative Closest Points (ICP) [1] algorithm. The full faces are not suitable for rigid registration due to the nonrigid deformations caused by expressions. Therefore, we crop out a T-area from the face surface using a binary mask generated with the five fiducial points (see Figure 6). The T-area point clouds of two faces are fed to the ICP algorithm to calculate rotation matrix and translation vector, which are then used to register the corresponding two faces.

2.3. Heuristic point generation

The nose tip and eye corners are suitable to serve as fiducial points, but not representative enough to extract expression features. Thus, we generate another 25 heuristic points for expression feature extraction, as illustrated in Figure 7. The orange horizontal lines are the location of the eyes, nose and chin (the bottom point of the face). The eye-nose separation and nose-chin separation are denoted by h and d respectively. They are taken as the *'length unit'* to measure the face along the vertical direction, and render positions to draw horizontal baselines. Similarly, in the vertical direction, the location of the four eye corners and the eye centers are selected to draw vertical baselines which intersect the horizontal baselines. The heuristic points are then selected from the intersections of these baselines.

According to the research done in [5], eyebrows and mouth area convey the most important information of facial expressions. Thus, the majority of heuristic points are selected around the eyebrows (points 8-11, 27) and mouth area (points 18-25, 28 and 29). This is a flexible scheme to generate heuristic points, in which the locations of some heuristic points can adjust according to different expressions. For example, d would be longer on a surprised face because of the opening mouth, so points 18-25, 28 and 29

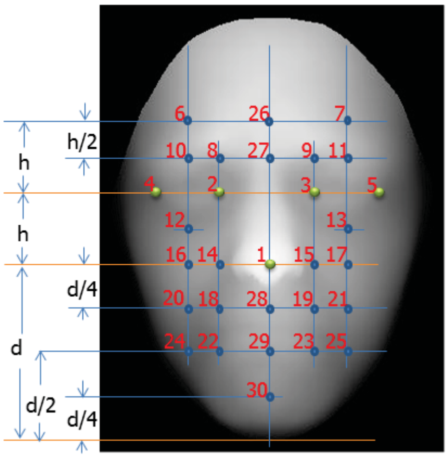


Figure 7. Schema of generating heuristic points.

will consequently be lowered to cover mouth area. Once the x, y-coordinates of the heuristic points are obtained from range images, the x, y, z-coordinates can be easily determined by finding the corresponding vertex on the uniformly sampled 3D point cloud.

为什么要用patch？因为所有的启发式点都是由直线连接的，并不是特别准确？

3. Feature extraction

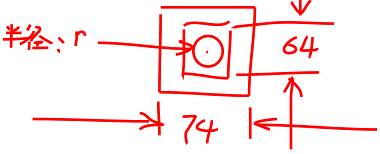
Although our heuristic point generation based on the distance ratios of the fiducial points adjusts according to the changes in facial shape, the heuristic points are not as accurate as manually labelled ones. Thus we cannot assume perfect alignment of patches surrounding these heuristic points. To overcome this, we select a subset of features within a patch that are useful in expression recognition despite errors in landmark location. To facilitate this sub-patch feature selection, we choose to extract depth features sampled by a discrete grid on the patch. Our features can be essentially viewed as a discrete approximation of the rings used in [8] but also offer feature selection to choose arbitrary sub-patches – a process that could not be accomplished with the concentric geodesic ring features used by [8] .

借助了一个第三方工具包来获取一个cubic patch

3.1. Patch-based depth feature extraction

As shown in Figure 8, once a heuristic point has been located, we use a sphere with radius r centered at this point to crop a cluster of points. Then, a cubic patch is fitted to the cropped points using the code from [3]. The fitted patch is then sampled on a uniform 74×74 grid, but only the central 64×64 samples covering the points in r region are kept as the patch-based depth feature, in order to avoid the artifacts at boundaries. All sampled patches end up with equal resolution which is necessary for the classification. Figure 9 shows the same patch on the mouth corner of three different subjects under the six expressions.

The 64×64 depth feature matrix of each patch is then reshaped into a 4096-dimension row vector. Thus, each



这里边的a cluster of points是不是前面生成的30个点（5个基准点+25个启发点）？

这里边存放的深度特征信息是什么？是一个单独的数字值？

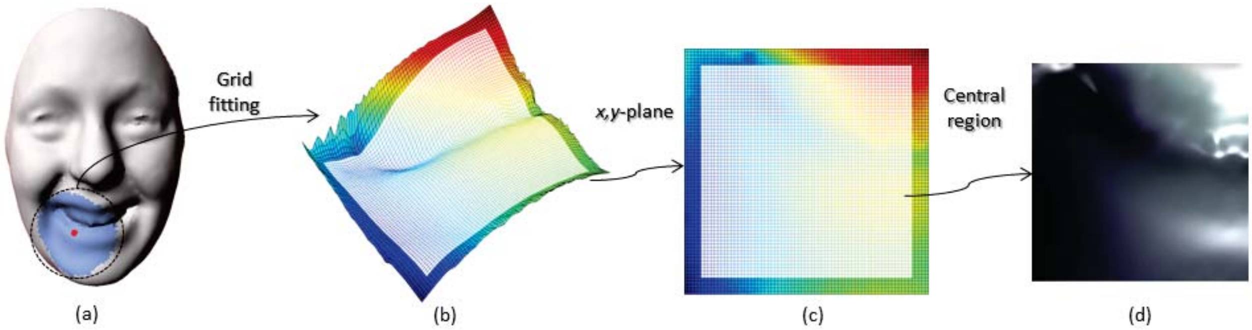


Figure 8. Patch-based depth feature extraction on 3D face surface.

3D face is represented by a 30×4096 matrix, as there are 30 patches. A dimension of 4096 is quite large for a feature vector that only describes a local patch, and there are overlaps between adjacent patches. Fortunately, it is possible to compress these vectors by projecting them into a linear subspace defined by 2DPCA. The goal is to discard the redundant information in preparation for feature selection. Assuming that there are N 3D face samples in the training set, the ith training sample is denoted by an m×n (our case 30×4096) matrix $A_i (i = 1, 2, ..., N)$, and the average of all training samples is denoted by $\overline{A}$. Then, the scatter matrix C can be calculated by

$$C = \frac{1}{N} \sum_{i=1}^{N} (A_i - \overline{A})^T (A_i - \overline{A}). \tag{1}$$

According to [19], the criterion of 2DPCA can be expressed by

$$J(X) = X^T CX, \tag{2}$$

where X is a unitary column vector. The optimal projection vector that maximizes the criterion is the eigenvectors of C corresponding to the largest eigenvalues. Normally, we select a set of the projection vectors, $X_1, ..., X_d$, subjected to the orthonormal constraints. This can be achieved by applying the Singular Value Decomposition (SVD) on the scatter matrix C as

$$C = USV^T, \tag{3}$$

where U is a 4096×4096 matrix of the eigenvectors and S is a diagonal matrix of eigenvalues, both sorted in descending order. The first d columns of U are the optimal projection vectors. To determine d, the ratio of the first d eigenvalues over the total eigenvalues is calculated by

$$\eta = \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{4096} \lambda_i}, \tag{4}$$

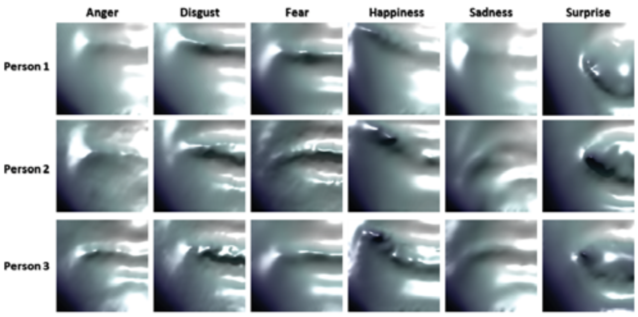


Figure 9. Comparison of 3D facial patch (mouth corner) under different expressions. The images are for the same three persons in Fig. 1

Table 2. Recognition rates of different parameters (patch radius r and fitting grid size) in feature selection.

Rates(%)	20×20	32×32	64×64
Radius=25mm	84.3	84.4	85.0
Radius=30mm	84.0	84.7	**85.4**
Radius=35mm	85.0	84.7	84.4

where λ_i is the ith eigenvalue. In our experiment, the ratio η always reaches 0.99 swiftly at only $d = 50$. Thus, the first $d = 50$ eigenvectors are kept as the optimal projection matrix U_d, and used to compress the samples as

$$F = (A - \overline{A})U_d, \tag{5}$$

where F is a 30×50 matrix.

To optimize the parameter r and grid size, we tested three different radiuses (25mm, 30mm and 35mm), and fitted into 20×20, 32×32 and 64×64 grids. The recognition rates of these settings are given in Table 2. It can be seen that the patch-based features around the 30 heuristic points are not very sensitive to the radius of the cropping sphere and the fitting grid size. r=30mm and 64×64 grid are hence used for all our remaining experiments.

3.2. Feature selection

Identifying the most characterizing features of the observed data is crucial to minimize the classification error. The idea of feature selection is that a simple combination

Binghamton University 3D Facial Expression
http://www.cs.binghamton.edu/~lijun/
Research/3DFE/3DFE_Analysis.html

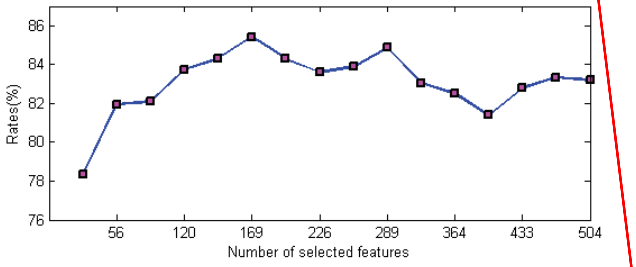


Figure 10. Recognition rates of different size of selected features.

of individually good features does not necessarily lead to good classification performance. That is to say, “the m best features are not the best m features” [11]. We adopt the framework of the *minimal-redundancy-maximal-relevance* (mRMR) [11] to select the best features for recognition. This involves a two-stage selection algorithm.

First, the mRMR criterion is used to select mutually exclusive features $S = \{x_1, ..., x_m\}$ that jointly have the largest characterizing power on each of the six prototypic expressions class c:

$$\begin{cases} max\Phi(D,R) = D(S,c) - R(S), \\ D(S,c) = \frac{1}{|S|}\sum_{x_i \in S} I(x_i;c), \\ R(S) = \frac{1}{|S|^2}\sum_{x_i,x_j \in S} I(x_i,x_j) \end{cases} \quad (6)$$

where $I(x_i;c)$ is the mutual information value between individual feature and class, $I(x_i,x_j)$ is the mutual information value between two features.

When candidate features are selected, the next task is to determine the optimal number of features m. A wrapper that tests features with an SVM classifier is utilized to decide the size of the feature set, with the direct goal of minimizing the recognition error of the specific classifier on the training set.

We performed 10-fold cross validation to select discriminant features. Each time, we selected a feature set $S_i, i = 1, ..., 10$ enclosing 800 features by incremental search [11], in which the features are arranged in descending order of characterizing power. The common features $\bar{S} = \bigcap_{i=1}^{10} S_i$ are taken as the feature candidates. With the purpose of determining the optimal number of features m, we give the common features of the first $50k$ features in $S_i, i = 1, ..., 10$ to the classifier, where k is the iteration number. The common features that yields the best recognition performance is considered as the optimal feature set. As shown in Figure 10, there are 169 common features in the first 300 feature candidates (k=6), and these 169 features are adopted as *”the best m features”* since they achieved the best recognition rates.

4. Classification

Support vector machines (SVMs) have proven to be powerful for facial expression classification. SVM achieves the best performance according to a comprehensive study [14]. Therefore, we adopt SVMs as the classifiers for facial expression recognition. SVMs attempt to find the hyperplane that maximizes the margin between the positive and negative observations for a specified class. Given a training set of labelled examples $\{(x_i, y_i), i = 1, ..., k\}$ where $y_i \in \{-1, 1\}$, a testing example x is labelled by the following function:

$$\boxed{f(x) = sgn(\sum_{i=1}^{k} \alpha_i y_i K(x_i, x) + b)} \quad (7)$$

where α_i are Lagrange multipliers of a dual optimization problem that determine the classification hyperplane, $K(\cdot, \cdot)$ is a kernel function, and b is the threshold parameter of the hyperplane.

SVMs make binary decisions. However, there are six classes in facial expression recognition. Therefore, we use LIBSVM [2] for the training and testing of SVMs, which achieves multi-classes classification according to the one-against-rest technique. With regard to the parameter selection, we carry out coarse-to-fine grid search in a 10-fold cross-validation on the training dataset.

5. Experimental results

BU-3DFE database is one of the very few publicly available databases of annotated 3D facial expressions. It consists of 100 subjects (56 females and 44 males) from different ethnic ancestries and ages. Each subject has 25 facial scans, including one neutral face and 24 faces of 6 prototypic expressions with 4 levels of intensity. The 3D locations of 83 facial landmarks are provided for each 3D face. These manually labeled landmarks are widely used by most existing analysis algorithms.

The experiment is performed on 54-vs-6 setup, which is a commonly used protocol by most methods [8][16][15][18][13]. The samples of 60 subjects (30 females and 30 males) with two high-intensities for each expression (03 and 04), which are randomly selected from the 100 subjects in BU-3DFE. In order to conduct person-independent facial expression recognition, we randomly split these 60 subjects into 10 folds, take 9 folds (54 subject, 648 samples) as training data, and the remaining fold (6 subjects, 72 samples) as the testing data.

Following the process of other methods [8][13] that used the BU-3DFE database, we select 60 subjects to form a 54-vs-6 setup. However, one issue is that precisely which 60 subjects are selected is never clearly specified by previous methods. This is an issue for performance comparison since different random samples of 60 subjects can give very

Table 3. Confusion matrix of recognition on BU-3DFE database.

	AN	DI	FE	SA	HA	SU
AN	**80.9**±3.7	3.8	4.0	10.4	0.8	0.1
DI	8.0	**81.5**±2.7	5.3	1.6	2.7	0.9
FE	4.1	7.1	**70.8**±3.1	4.0	9.9	4.3
SA	13.0	1.7	5.3	**79.6**±3.1	0.4	0.0
HA	0.2	0.9	7.3	0.0	**91.1**±2.0	0.6
SU	0.3	1.5	3.0	0.5	0.7	**94.0**±1.7

Table 4. Comparison between the proposed method and other 3D facial expression recognition approaches. The type of "manual" means the landmarks used in the corresponding method are manually labeled, while "auto" means points are automatically detected or not necessary.

Method	Type	AN	DI	FE	SA	HA	SU	Avg
L. Yin [20]	Human expert	*94.9*	*95.4*	*96.4*	*96.2*	*99.4*	*99.0*	*96.8*
A. Maalej [8]	Manual	97.9	99.2	99.7	99.3	98.6	98.2	98.8
U. Tekguc [16]	Manual	86.0	87.3	85.3	82.9	93.4	94.7	88.2
H. Tang [15]	Manual	86.7	84.2	74.2	82.5	95.8	99.2	87.1
J. Wang [18]	Manual	80.0	80.4	75.0	80.4	95.0	90.8	83.6
T. Sha [13]	Manual	78.7	83.9	69.8	84.8	88.5	95.4	83.5
The proposed	Auto	**80.9**	**81.5**	**70.8**	**79.6**	**91.1**	**94.0**	**83.0**
P. Lemaire [7]	Auto	74.1	74.9	64.6	74.5	89.8	90.9	78.1
P. Lemaire [6]	Auto	69.4	78.2	42.8	82.9	88.8	92.5	75.8

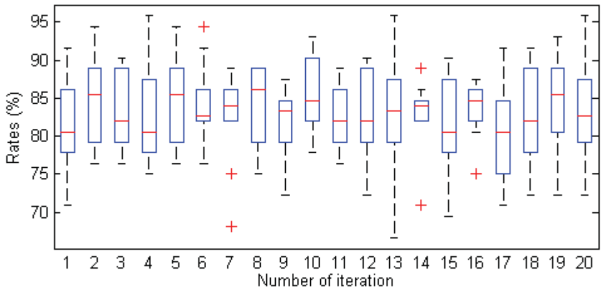


Figure 11. Boxplot of 20 times repeated 10-fold cross validation results of the proposed method.

different results and the selection of 60 "*easy*" faces can give very high accuracy. To ensure unbiased experimental results, we perform 20 random selections and conduct 10-fold cross validation on each of the 20 sets. Thus our total experiments are 20×10 = 200. The recognition results for each of the 20 times are shown in the box plot of Figure 11. Note the significant variations in expression recognition between different sets of 60 subjects.

The recognition rates and stand derivations across all 20 random selections of 10-fold experiments are averaged and reported in Table 3. The proposed method achieved a 83% average recognition rate for the six prototypic expressions. The major confusions are 13.0% (sadness is misclassified as anger), 10.4% (anger is misclassified as sadness) and 9.9% (fear is misclassified as happiness).

Table 4 compares our method to existing manual and automatic 3D facial expression recognition techniques. The first row in the table reports the results of the performance of human experts on the same BU-3DFE database using the same two expression intensity levels i.e. 03 and 04. These experiments were performed by two psychologists who are the experts in human facial expression research [20]. It is surprising to see that not even humans have perfect accuracy on this database and even more surprisingly, the method in [8] performs better than humans. Our method has the best performance among automatic methods and compares well with other manual methods except [8]. It is worth noting that our results are averaged over 20 random picks of 60 subjects multiplied by 10-fold experiments for each pick, whereas the results reported by others are based on a single random pick of 60 subjects.

5.1. Analysis and discussion

Although landmark detection on 3D face models remains an open problem, it is inevitable in designing a fully automatic facial expression analysis system. The experimental results reveal that our method outperforms existing automatic techniques [6][7], with better accuracy for every single expression (except sadness). In addition, clearly the errors in our heuristic points makes the recognition task much more difficult than methods based on manual landmarks. However, as shown in Table 4, the depth feature extracted around 30 heuristic points still achieved comparable results to many manual techniques [8][16][15][18][13] which use the manually labeled 83 landmarks.

The local depth feature utilized in proposed method facilitates an effective feature selection. That is why we can achieve comparable performance with those methods using manual landmarks. The discrete depth features are projected to a subspace by 2DPCA for dimension reduction. By discarding the redundant dimensions, the resulting feature vectors conserve most of the essential information with large variance. However, the variances in the resulting feature vector are not purely caused by facial expressions. It also contains the changes caused by the facial differences in subjects and the inaccuracy of the heuristic points. We performed facial expressions on the *'contaminated features'*, and only achieved an average recognition rate of 75.4%. This shows that feature selection is vital to our performance, increasing the accuracy significantly to 83%.

6. Conclusion

This paper presented a fully automatic 3D static facial expression recognition method using local patch-based depth features. We extracted depth feature around 30 heuristic points, generated from 5 fiducial points, to represent facial expressions. A multi-class SVM was trained to classify the expressions based on the extracted feature after mRMR feature selection. The proposed method outperformed existing fully automatic methods by a significant margin.

References

[1] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[2] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[3] J. R. D'Errico. Understanding gridfit. *Information available at: http://www. mathworks. com/matlabcentral/fileexchange/loadFile. do*, 2006.

[4] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[5] I. Kotsia, I. Buciu, and I. Pitas. An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing*, 26(7):1052–1067, 2008.

[6] P. Lemaire, B. Ben Amor, M. Ardabilian, L. Chen, and M. Daoudi. Fully automatic 3d facial expression recognition using a region-based approach. In *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, pages 53–58. ACM, 2011.

[7] P. Lemaire, L. Chen, M. Ardabilian, M. Daoudi, et al. Fully automatic 3d facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In *Workshop 3D Face Biometrics,*, 2013.

[8] A. Maalej, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berretti. Shape analysis of local facial patches for 3d facial expression recognition. *Pattern Recognition*, 44(8):1581–1589, 2011.

[9] A. Mian. Robust realtime feature detection in raw 3d face images. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 220–226. IEEE, 2011.

[10] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1424–1445, 2000.

[11] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[12] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 2012.

[13] T. Sha, M. Song, J. Bu, C. Chen, and D. Tao. Feature level analysis for 3d facial expression recognition. *Neurocomputing*, 74(12):2135–2141, 2011.

[14] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.

[15] H. Tang and T. S. Huang. 3d facial expression recognition based on properties of line segments connecting facial feature points. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[16] U. Tekguc, H. Soyel, and H. Demirel. Feature selection for person-independent 3d facial expression recognition using nsga-ii. In *Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on*, pages 35–38. IEEE, 2009.

[17] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[18] J. Wang, L. Yin, X. Wei, and Y. Sun. 3d facial expression recognition based on primitive surface feature distribution. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1399–1406. IEEE, 2006.

[19] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):131–137, 2004.

[20] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3d facial expression database for facial behavior research. In *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pages 211–216. IEEE, 2006.