

Automatic 3D Facial Expression Recognition using Geometric and Textured Feature Fusion

Asim Jan and Hongying Meng

Department of Electronic and Computer Engineering, Brunel University London, UK

Abstract—3D facial expression recognition has gained more and more interests from affective computing society due to issues such as pose variations and illumination changes caused by 2D imaging having been eliminated. There are many applications that can benefit from this research, such as medical applications involving the detection of pain and psychological effects in patients, in human-computer interaction tasks that intelligent systems use in today's world. In this paper, we look into 3D Facial Expression Recognition, by investigating many feature extraction methods used on the 2D textured images and 3D geometric data, fusing the 2 domains to increase the overall performance. A One Vs All Multi-class SVM Classifier has been adopted to recognize the expressions Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise from the BU-3DFE and Bosphorus databases. The proposed approach displays an increase in performance when the features are fused together.

I. INTRODUCTION

Emotion is an integrated part of a person that represents the mental states of a person, it can give some clues for unpredictable events caused by these emotions. Capturing the emotion of a person from facial expression will help us adapt our words in a conversation, interact in a proper way and prevent bad things from happening. Automatic facial expression recognition has been an intensive research area from images or videos in last few years evidenced by several international challenge competitions and workshops held (e.g. FERA 2011 [20], AVEC2011 [15], AVEC2012 [16], AVEC2013 [22], AVEC2014 [21]). However, due to issues such as pose variations and illumination changes in 2D imaging, facial expression from 2D images or videos is still a challenging task with the overall performance far from satisfied in real world applications.

With the fast development on 3D imaging systems, the use of 3D facial data for facial expression has attracted increasing interest as 3D data are theoretically pose invariant and robust to illumination changes. Furthermore, they also capture accurate geometry information closely sensitive to expression variations. In recent years, research has been done in this area [5], and significant progress has been made [10]. Existing methods for facial expression recognition based on static 3D data can be categorized into two streams, i.e. feature based or model based. The first category claims that the distributions of facial surface geometric information such as gradient and curvature distances between pairs of interest landmarks and local shapes near landmarks are closely related to expression categories. The second category tries to simulate the physical process of generating expression and explores a generic elastically deformable face model, which

can generate universal expressions by adjusting parameters. The first category is mostly used because the second category means extensive computing cost.

In the early stage, some methods for 2D facial expression techniques have been adopted on 3D images such as Gabor wavelets and Facial action coding units (FACS). Gabor wavelet takes is a set of multi-scale and multi-orientation coefficients are calculated to describe the appearance variations in the facial region. FACS is a method of assigning each facial feature of the face a value, when the emotion is expressed (movement of facial features), a combination of these values are recorded to describe the expression. Now that 3D imaging provides more detailed observations such as depth and geometric data of the facial features, it can be used to give a more accurate detection of facial muscle movements and changes.

There are a few 3D facial expression databases that contributes to the fast development in this area such as BU-3DFE Database [24], Bosphorus database [14] and so on. In the past few years there have been many researchers that have used the BU-3DFE database to apply their algorithms to achieve high recognition rate of the different expressions.

P.Lemaire [8] uses Textured data based on morphing facial structure and applying Histogram of oriented gradients (HOG). T.Yun [25] uses a 3D Gabor Library to extract color/density information with the help of geometric data. They use Principal Component Analysis (PCA) for feature dimensionality reduction. Wang et al. [23] has provided the baseline results using Primitive Surface Feature Distribution (PSFD) along with Linear Discriminant Analysis (LDA) to recognize 6 expressions (Angry, Disgust, Fear, Happiness, Sadness and Surprise). Rabiou et al. [13] has used the geometric data to obtain 16 feature distances based on the FACS principle, along with 27 Angles using maximum relevance minimum redundancy (mRMR) to reduce the features and then a Support Vector Machine (SVM) for classification. Yurtkan et al. [26] recently propose a feature selection procedure for improved facial expression recognition utilizing 3-Dimensional (3D) geometrical facial feature point positions. It classifies expressions in six basic emotional categories which are anger, disgust, fear, happiness, sadness and surprise.

Almost all the work available uses only one type of method to extract features. This paper will address 3D facial expression recognition problem using feature based method. Apart from geometric features, different texture features will also be investigated. Furthermore, feature fusion was used

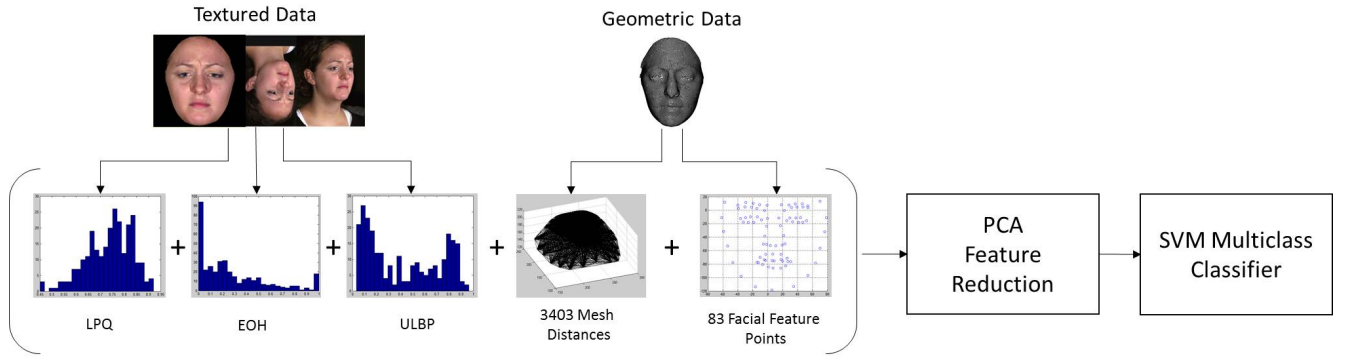


Fig. 1. Overall framework of the proposed approach, combining the different features extracted using the geometric and textured data.

for different combinations of the all features. The proposed approach was evaluated in two public available datasets and clear performance improvement was achieved from feature fusion.

In this paper there are two databases on 3D Facial Expressions, similar to each other, that have been chosen which are the Binghamton University 3D Facial Expression (BU-3DFE) Database [24] and the Bosphorus Database [14]. The main contribution will be based on the BU-3DFE Database because of its consistency per subject.

The rest of the paper is organised as the following. The proposed method is described in details in section II. The experimental results are given in section III and section IV concludes the paper.

II. METHODOLOGY

Facial expressions are used to visually describe the human state of emotion, this paper proposes a novel method that comprehensively models the variations in visual clues by fusing key features obtained from the Geometric and Textured domains, to investigate how the overall performance is affected. This section describes how the features are extracted, fused together and reduced using feature dimensionality reduction methods, and classified using machine learning techniques.

A. System Overview

Feature Extraction is required for many classification applications, which is to obtain key information from the available data that is used to assist intelligent systems. From the many existing feature extraction methods available today, we have chosen a few to use that would extract relevant information about the facial expressions.

Figure 1 illustrates the process of how the features are extracted from each of the facial models, reduced in dimensionality, and used to classify expressions with machine learning. The following section will discuss how these processes are computed.

B. Textured Feature Extraction

The Textured features are taken from the cropped facial images and uncropped side-view images of the subjects provided by the BU-3DFE database. From each image we

extract multiple features which are described in the following sections.

1) *Local Phase Quantization (LPQ)*: local phase quantization (LPQ) is proposed for texture analysis [12] and applied to blurred face recognition [7] [2]. The process involves creating blocks across the whole image which can be any size. We then compute the local Fourier frequency coefficients for all the pixels in each block, and use a scalar quantizer to transform the coefficients into an 8-bit binary code from which produce a histogram of 256 bins.

We take the same two approaches as we did with the ULBP method, by applying LPQ on the whole facial image and then on separate windows of the image. However when we window the image, we produce a 4 x 4 grid to make 16 equally sized images; rather than 8 x 8 grid used in ULBP; because LPQ creates its own M x M blocks on top of each image. All the resulting histograms are then concatenated to produce the LPQ Feature Vector.

2) *Edge Oriented Histogram (EOH)*: Edge Oriented Histogram (EOH) is made on the same principles as Histogram of Oriented Gradients (HOG), and it is more efficient and powerful operator that will capture an edge or the local shape information of an image. This is applicable to various applications such as: computer vision such as hand gesture recognition [6], human detection [4] and facial expression recognition [8]. Edges can be detected in an image using edge operators such as Sobel to detect to horizontal edges and vertical edges strengths. The angle interval is divided into N bins and the strengths in the same bin are summed to build a histogram. The whole image is divided into cells and each cell into blocks. The histogram relative to each block are linked to generate the EOH feature, we use this algorithm to obtain the angle gradients from the directional edges across each expression from each of the textured facial images.

3) *Uniform Local Binary Patterns (ULBP)*: Local Binary Patterns (LBP) is a very common 2D image feature extraction method. Its main purpose is to describe the local texture structure of an image using binary patterns which is obtained from its surroundings. It is very computationally efficient and effective, and has been used for face recognition [1], and many other applications [11].

The LBP operator compares each pixel with its surrounding 8 pixels; depending on the radius size; using a threshold to compare if the gray-scale value is higher or lower. This creates a string of 8 (8 surrounding pixels) bits (1 for higher, 0 for lower) and converted to decimal format from which a histogram of 256 bins is created. This is the standard LBP operator method, however we apply Uniform LBP. This involves looking for a minimum of 2 bit-wise transitions in the binary pattern [11], e.g. 0(01)11(10)0. From this we produce a histogram of 59 bins to describe the occurrences of the different patterns.

We take two approaches when applying this method, the first is to apply standard LBP on the whole facial image, producing a vector of 256 components. The second approach is to window the image with an 8 x 8 grid, producing 64 equally sized images. Uniform LBP is then ran across each image separately and all the resulting histograms (64 x 59) will be concatenated together.

C. Geometric Feature Extraction

For the Geometric features, we obtain two sets, first to modify the 83 facial feature points given with the database and also calculate the 3D distances of a mesh generated using the 83 points. Both features are described further in the following sections.

1) *83 Facial Feature Points*: The BU-3DFE database provides 83 key feature points (X,Y,Z coordinates) annotated from the cropped face image for each expression as shown in Fig. 2. These include points from the eyebrows, eyes, nose, mouth and around the face. These points are then normalized so that each face is aligned correctly and the values range from 0 to 1. This process is shown in equation 1.

$$N_i = \frac{P_i - \min(P_i)}{\max(P_i) - \min(P_i)} \quad (1)$$

Where each of the 83 points X,Y and Z coordinates is represented by $P_i = (x_i, y_i, z_i)$, i is the points index from 1 to 83 and N_i is the normalized X,Y and Z coordinates for each point.

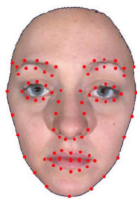


Fig. 2. 83 Facial Feature Points annotated on a cropped image.

2) *Facial Mesh Distances*: Using the 83 Facial feature points, we have taken the approach of calculating the 3403 possible combinations between each of the 83 points, similar to [19], to get the 3D distance between the combinations of the coordinates $P_i = (x_i, y_i, z_i)$. Equation 2 shows how each distance is calculated. Each of the possible distances $d_{(i,j)}$ for each combination is calculated by obtaining the Euclidean distance between the points.

$$\|d_{i,j}\| = \sum_i \sum_{j < i} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (2)$$

These distances will provide more detail of how each facial feature is moved further or closer by the expression, e.g. the distance between the eyes and mouth as they go further apart when the expression is Surprise.

D. Feature Dimensionality Reduction

Fusing the many features produced by the algorithms can result in a large feature vector which can slow down the training period of the system. The main purpose of using feature dimensionality reduction techniques is to reduce the size of the feature vector whilst retaining its quality. Reducing the dimensions generally means increasing the speed of the learning process making it less computationally expensive.

We have chosen Principal Component Analysis (PCA) to reduce the dimensionality of our feature vector. PCA is a method of reducing a very large feature vectors down to a manageable size. [17] shows a step by step guide to explain the concept using an example of the physics data generated by a ball on a spring. We use PCA to take all the relevant information from the various feature sets to produce a smaller yet relevant and accurate feature vector for facial expression recognition.

E. Facial Expression Classification

There are many different existing Machine Learning models available to use for multi-class classification. We have chosen to use the Support Vector Machine (SVM). This machine learning method creates a hyper-plane when being trained to separate the differently classed data the best it can. The SVM Classifier has been designed using the LibSVM Library [3] for MATLAB. This has been modified and optimized to work with Multi-class classification.

For the SVM model, the approach we used was One Vs All. This meant that the classification will be done against all the classes, as opposed to One Vs One where the classification would be done between 2 labels and would require multiple SVM models. The optimization process for all networks have been based specifically for this application, each method has a parameter that can be adjusted to give the best overall performance. For the LibSVM model, we have chosen to use a polynomial kernel from which the Degree value has been optimized through a loop which changes the respective parameter in step sizes, each parameter value is tested across 10 sets of data and the average is measured. The value with the best average is then selected for the experimental testing.

III. EXPERIMENTAL RESULTS

The main objective of the upcoming experiments is to show how fusing features from Texture and Geometric domains can improve the accuracy of a system.

A. BU-3DFE Database

1) *Data*: The BU-3DFE database is developed by Li et al. [24]. from Binghamton University for researching purposes. This Database contains Facial models of 100 subjects (44 Males, 56 Females) of various ethnic backgrounds. The data for each subject contains 6 prototypical expressions which are Angry, Disgust, Fear, Happy, Sad and Surprise, along with the basic Neutral expression. Each of these expressions apart from Neutral contain 4 levels of intensity, with the first being weak and last being strong, which totals to 2500 facial models. For each model, the database provides a cropped image containing just the face; an uncropped image that contains both side views of the face; 83 manually annotated landmarks of the main facial features and a single 3D Face Mesh Model containing 3D coordinates with a resolution of about 25K to 35K polygons [24].

2) *Testing Protocol*: There are 2 main experiments that have taken place on the BU-3DFE database, classifying for 7 expressions and then for 6 expressions which excludes the Neutral Expression. Each experiment is split up multiple ways, the 3 main testing protocols are Geometric features only, Texture features only and the Geometric and Texture features fused. Further tests are done of combining different combinations of the textured and geometric feature sets. Finally a quick validation experiment is executed on the Bosphorus database using similar protocols as the BU-3DFE database.

Following majority of the related work, each experiment consists of 100 tests, each test from the BU-3DFE database randomly selects 60 out of the 100 subjects. Each subject contains 4 intensities of each expression (except Neutral), from which the strongest 2 are selected. From the 60 subjects, we use 54 for training and 6 for testing, applying 10-fold cross validation to ensure that each subject is tested. The results are calculated by accumulating all the confusion matrices across each cross-fold from every test (100x10), to then calculate the overall average performance across the 100 tests. From the features sets, we denote 83 Facial Feature Points as (83P), Facial Mesh Distances as (FD).

3) *Individual Feature Experiment*: This section shows the individual performances of the different 2D feature sets (EOH, ULBP, LPQ), and 3D feature sets (83 Facial feature points and Facial Mesh Distances). Each feature set has been reduced in dimensionality using PCA, apart from EOH and the 83 feature points as they have few features to start with.

From Table I it can be seen that the 83 Facial Feature Points has the best accuracy from both domains, ULBP is the best from the Texture Domain which is 0.2% lower accuracy than the 83 Facial Feature Points. Overall each individual performance is in the high 70s apart from EOH. We can test the effectiveness of the EOH feature by removing it when fusing with the other feature sets to see if it is beneficial or not to include it. The feature highlights the gradients in the facial images, which may be weak on its own but could improve other features when fused together.

4) *Feature Fusion Experiment*: This section shows the performance when combining the Textured features (EOH,

TABLE I
PERFORMANCE OF FACIAL EXPRESSION RECOGNITION USING INDIVIDUAL SET OF FEATURES, TAKING AVERAGE RECOGNITION RATE ACROSS 100 TESTS. CLASSIFYING 7 EXPRESSIONS AND 6 EXPRESSIONS (EXCLUDING NEUTRAL)

| Domain | Feature | 7 Expressions | 6 Expressions |
|-------------|---------|---------------|---------------|
| Texture2D | EOH | 72.34% | 75.77% |
| Texture2D | ULBP | 81.10% | 83.53 |
| Texture2D | LPQ | 79.89% | 82.01 |
| Geometric3D | 83P | 81.30% | 83.35% |
| Geometric3D | FD | 79.89% | 81.18% |

ULBP and LPQ); when combining the Geometric features (83 Facial feature points and Facial Mesh Distances); and when both domains are fused. We also fuse the best features from the Texture and Geometric domain and try a few other combinations of the feature sets.

TABLE II
PERFORMANCE OF FACIAL EXPRESSION RECOGNITION FUSING VARIOUS COMBINATIONS OF THE INDIVIDUAL SETS OF FEATURES, TAKING AVERAGE RECOGNITION RATE ACROSS 100 TESTS. CLASSIFYING 7 EXPRESSIONS AND 6 EXPRESSIONS (EXCLUDING NEUTRAL)

| Feature | 7 Expressions | 6 Expressions |
|---------------------|---------------|---------------|
| ULBP+LPQ+EOH | 83.67% | 85.06% |
| 83P+FD | 80.33% | 81.25% |
| ULBP+83P | 80.80% | 83.78% |
| ULBP+LPQ+83P+FD | 87.68% | 89.75% |
| ULBP+LPQ+EOH+FD | 88.13% | 89.84% |
| ULBP+LPQ+EOH+83P | 83.79% | 86.19% |
| ULBP+LPQ+EOH+83P+FD | 88.32% | 90.04% |

Table II compares the accuracy when fusing different feature sets together in different combinations. When classifying 7 expressions, we can see that the highest accuracy occurs when all the feature sets are fused together, producing 88.32%. With the individual domains fused we get 83.67% for Texture; which is higher than any single Texture Feature set from Table I, and 80.33% for the Geometric Domain; which has decreased from the individual contribution of the 83 facial feature points. It can be seen that there is a significant jump of 4.65% when the Texture and Geometric domains are fused, even when using 6 Expressions there is a jump of 4.98%, proving that having multiple feature sets from different perspectives of 3D modeling contributes beneficially towards the FER application performance.

The Facial Mesh Distances (FD) feature set has shown big improvements to the overall accuracy when combined with the Texture Domain (ULBP+LPQ+EOH+FD increasing from 83.67% to 88.13%). It has a bigger impact than adding the 83P feature set (ULBP+LPQ+EOH+83P increasing from 83.67% to 83.79%) because it contains more detailed information and more features (3403) giving it a stronger weight when fused with the texture domain. Even though the EOH feature produced the lowest accuracy, it still improved the overall performance from 87.68% (ULBP+LPQ+83P+FD) to

88.32%. This shows that having each feature set provide its own translation of the model, it can contribute to a better recognition rate.

Having only 6 Expressions to classify increases the accuracy of the system whilst retaining the similar pattern of results when classifying 7 Expressions. These results supports the idea of fusing the Texture and Geometric (90.04%) domain by showing an increase of 4.98% from the Texture domain (85.06%) only , and an increase of 8.79% from the Geometric domain (81.25%).

TABLE III
CONFUSION MATRIX OF FUSING ALL FEATURE SETS
(ULBP+LPQ+EOH+83P+FD) ACHIEVING A TOTAL OF 88.32%
OVERALL ACCURACY

| | AN% | DI% | FE% | HA% | NE% | SA% | SU% |
|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| AN | 88.24 | 2.5 | 1.3 | 0.5 | 2.4 | 4.9 | 0.0 |
| DI | 2.4 | 90.60 | 2.6 | 0.9 | 0.3 | 1.3 | 1.6 |
| FE | 1.6 | 3.2 | 86.56 | 2.1 | 1.1 | 3.2 | 2.0 |
| HA | 0.0 | 0.6 | 7.1 | 91.34 | 0.0 | 0.0 | 0.7 |
| NE | 3.7 | 0.8 | 3.3 | 0.2 | 76.21 | 14.3 | 1.3 |
| SA | 4.5 | 1.0 | 4.7 | 0.0 | 4.3 | 85.32 | 0.0 |
| SU | 0.3 | 1.3 | 3.6 | 0.5 | 0.3 | 0.0 | 93.81 |

Table III is the resulting Confusion Matrix of fusing all the feature sets (ULBP+LPQ+EOH+83P+FD) together. The total average accuracy for the 7 expressions has been calculated based on the number of samples across 100 tests because Neutral has half the number of samples when compared to the other expressions. From the 100 tests, the sum of each correct sample has been divided by the total number of samples used, from this we obtain the overall accuracy.

The results show a lot of confusion between the Neutral and Sad expressions, as they are very similar; having slight movement of facial muscles; majority of the samples for Neutral were predicted as Sad. Surprise has very high accuracy because the expression creates a lot of muscle movement significantly changing the appearance of the subject, this can easily be picked up using the geometric feature extraction methods.

Table IV list the state-of-the-art performance on the BU-3DFE Database. From this list, it can be seen that only one type of feature was used in their method. Although some of the methods achieved good performance, their performance can be improved further if two types of features were used.

TABLE IV
RELATED EXPERIMENTS OF FACIAL EXPRESSION RECOGNITION ON
BU-3DFE DATABASE

| Method | Mean RR | Feature |
|---------------------|---------|-----------|
| Wang et al. [23] | 83.6% | Geometric |
| Rabiu et al. [13] | 92.2% | Geometric |
| Soyel et al. [18] | 91.3% | Geometric |
| Xioli et al. [9] | 90.2% | Geometric |
| Tekguc et al. [19] | 88.1% | Geometric |
| T.Yun [25] | 85.39% | Texture |
| Lemaire et al. [8] | 78.13% | Texture |
| Yurtkan et al. [26] | 88.2% | Geometric |

B. Bosphorus Database

1) *Data*: The Bosphorus database [14] developed by A.Savran et al. also used for researching purposes. The database includes a total of 4,666 scans collected from 105 subjects, from which 61 are male and 44 female. There are multiple facial expressions included which are represented in 2 ways, first being the basic expressions of Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise which is shown in Fig. 3. The other are expressions based on Action Units. Ideally each subject would contain a single frontal face image and 3D landmark file; which is annotated in Fig. 3; for each expression (except Neutral which contains 4 per subject), however the database is not very consistent with some subjects missing certain expressions yet having the others.

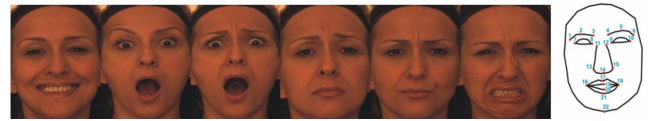


Fig. 3. Image 6 expressions and 22 facial landmarks taken from the Bosphorus Database [14].

2) *Performance*: We have used a similar protocol when using the Bosphorus database. As it contains less data and is less consistent with the subjects expressions than the BU-3DFE database, it is difficult to run accurate experiments. The landmarks of the facial features are only 22 - 25 points, compared to 83 of the BU-3DFE database. We have taken the common 22 feature points (22P) of all the expressions and normalized them as well as creating 231 Facial Mesh Distances (FD). EOH, ULBP and LPQ are applied to only the frontal face image for each subjects expressions, as the database lacks the side-view images. The features are also reduced in dimensionality using PCA and classified with SVM using the same parameters as the previous experiment. 3 main experiments have been executed that are similar to the previous experiment, which are using geometric features only, textured features only and a fusion of both. We have decided to combine all the relevant expressions from every subject and to randomly shuffle them around throughout the 100 tests per experiment.

TABLE V
VALIDATION TEST USING BOSPHORUS DATABASE, TESTING
INDIVIDUAL AND FUSED DOMAINS FOR 7 EXPRESSION.

| Domain | Feature | 7 Expressions |
|-------------|---------------------|---------------|
| Geometric3D | 22P+FD | 75.68% |
| Texture2D | ULBP+LPQ+EOH | 74.43% |
| Fusion | ULBP+LPQ+EOH+22P+FD | 79.46% |

Table V shows the readout of the experiments on the Bosphorus dataset. The Geometric domain (75.68%) has marginally outperformed the Textured domain (74.43%), by 1.25%, whereas the fusion of both domains has produced the best results of 79.46%, which is an increase of 3.78%. This

experiment also shows that fusing the 2 domains together does increase the overall performance of a system.

IV. CONCLUSION

In this paper, a different approach was proposed for 3D facial expression recognition by fusing multiple feature extraction methods used on the textured and geometric data. This includes an approach of taking all the possible distances across the 83 facial feature points provided, and a host of existing methods such as EOH, LBP, and LPQ. This feature dimensionality is then reduced using Principal Component Analysis and a Multi-class SVM was selected to be trained and to classify a combinations of experiments. From the experiments on the BU-3DFE database, we can see a total of 4.8% increase in overall accuracy when fusing all feature sets from both geometric and textured domains, with the test on 7 Expressions producing 88.32% and 90.04% for 6 Expressions. A validation test using the untidy Bosphorus database has also confirmed the effects of fusing both domains, giving an increase of 3.78% in overall accuracy.

In comparison with other state-of-the-art methods, our method didn't achieved the best performance. Some of the methods might some advanced classification method, some of them might use some specific, optimized, and selected features. In the future work, we will select the best geometric and texture features together with best feature selection method to improve the performance further. For example, we will use the geometrical data to highlight the main regions of the face that vary per expression, and to then apply the textured algorithms; such as ULBP and LPQ; only on those regions, effectively removing any noise and similarity across the features generated by the other regions.

The main contribution of the paper is to show the effects of fusing feature sets from different domains together with significant performance improvement. It might encourage other researcher to apply it in their work for better performance.

REFERENCES

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–41, Dec. 2006.
- [2] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä. Recognition of blurred faces using local phase quantization. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 2005.
- [5] T. Fang, X. Zhao, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris. 3D facial expression recognition: A perspective on promises and challenges. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 603–610, 2011.
- [6] W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. Technical report, Mitsubishi Electric Research Labs., 201, 213.
- [7] Z. Lei and S. Z. Li. Fast multi-scale local phase quantization histogram for face recognition. *Pattern Recognition Letters*, 33(13):1761–1767, Oct. 2012.
- [8] P. Lemaire, M. Ardabilian, L. Chen, and M. Daoudi. Fully automatic 3d facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–7, April 2013.
- [9] X. Li, Q. Ruan, and Y. Ming. 3d facial expression recognition based on basic geometric features. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 1366–1369, Oct 2010.
- [10] I. Mpiperis, S. Malassiotis, and M. Strintzis. Bilinear Models for 3-D Face and Facial Expression Recognition. *IEEE Transactions on Information Forensics and Security*, 3, 2008.
- [11] T. Ojala, M. Pietikainen, and T. Maenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, Jul 2002.
- [12] V. Ojansivu, E. Rahtu, and J. Heikkilä. Rotation invariant local phase quantization for blur insensitive texture analysis. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.
- [13] H. Rabiou and M. Saripan. 3D facial expression recognition using maximum relevance minimum redundancy geometrical features. *EURASIP Journal on Advances in Signal Processing* 2012, pages 1–8, 2012.
- [14] A. Savran, N. Alyuz, and H. Dibekliolu. Bosphorus database for 3D face analysis. *Biometrics and Identity Management*, pages 47–56, 2008.
- [15] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic. AVEC 2011: The first international audio/visual emotion challenge. In *International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, 2011.
- [16] B. Schuller, M. F. Valstar, R. Cowie, and M. Pantic. AVEC 2012: the continuous audio/visual emotion challenge - an introduction. In *ICMI*, pages 361–362, 2012.
- [17] J. Shlens. A tutorial on principal component analysis. *CoRR*, abs/1404.1100, 2014.
- [18] H. Soyel and H. Demirel. Facial expression recognition using 3D facial feature distances. *Image Analysis and Recognition*, pages 1–13, 2007.
- [19] 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. ISICIS 2009. 24th International Symposium on*, pages 35–38, Sept 2009.
- [20] M. F. Valstar, B. Jiang, M. Mhu, M. Pantic, and K. Scherer. The first facial expression recognition and analysis challenge. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2011.
- [21] M. F. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. AVEC 2014 - 3d dimensional affect and depression recognition challenge. In *International Conference on ACM Multimedia - Audio/Visual Emotion Challenge and Workshop*, 2014.
- [22] M. F. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic. AVEC 2013 - the continuous audio/visual emotion and depression recognition challenge. In *International Conference on ACM Multimedia - Audio/Visual Emotion Challenge and Workshop*, 2013.
- [23] 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, 2006.
- [24] L. Yin, X. Wei, and Y. Sun. A 3D facial expression database for facial behavior research. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:1399–1406, 2006.
- [25] T. Yun and L. Guan. Human emotion recognition using real 3d visual features from gabor library. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 505–510, Oct 2010.
- [26] K. Yurtkan and H. Demirel. Feature selection for improved 3d facial expression recognition. *Pattern Recognition Letters*, 38(0):26 – 33, 2014.