International Journal of Computer and Information Technology (ISSN: 2279 – 0764)
Volume 03 – Issue 05, September 2014

Surface Normals with Modular Approach and Weighted Voting Scheme in 3D Facial Expression Classification

Hamimah Ujir
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Kota Samarahan, Sarawak, Malaysia
Email: uhamimah {at} fit.unimas.my

Michael Spann
School of Electronic, Electrical and Computer Engineering
University of Birmingham
Edgbaston, Birmingham, United Kingdom

Abstract— **A crucial part for facial expression analysis is to capture a face deformation. In this work, we are interested by the employment of 3D facial surface normals (3DFSN) to classify six basic facial expressions and the proposed approach was employed on the Bosphorus database. We constructed a Principal Component Analysis (PCA) to capture variations in facial shape due to changes in expressions using 3DFSN as the feature vector. A modular approach is employed where a face is decomposed into six different regions and the expression classification for each module is carried out independently. We constructed a Weighted Voting Scheme (WVS) to infer the emotion underlying a collection of modules using a weight that is determined using the AdaBoost learning algorithm. Our results indicate that using 3DFSN as the feature vector of WVS yields a better performance than 3D facial points and 3D facial distance measurements in facial expression classification using both WVS and a Majority Voting Scheme (MVS). Our work is different with the existing works as they used the dataset with facial intensity information while we used dataset with no intensity. New insight in facial expression analysis is found particularly when no intensity information is provided. Surface normals does has a potential to be used as the feature vectors to classify six basic expressions.**

Keywords-component; Facial expression classification; 3D facial features; Principal Component Analysis; Support Vector Machines;Weighted Voting Scheme

I. INTRODUCTION

Facial expression recognition and classification is an emerging research area spanning several disciplines such as pattern recognition, computer vision and image processing. It brings benefits in human centred multimodal human-computer interaction (HCI) whereas the user's affective states motivate human action and enrich the meaning of human communication. In HCI, affective computing employs human emotion to build more flexible and natural multimodal [1]. The automatic human affect recognition system will change the ways we interact with computer systems. With efficient automated face expression classification, perhaps it will be an aid to the affect-related research community to carry out clinical psychology, psychiatry, and neurosciences research. Such systems could improve the quality of the affect-related research by improving the reliability of measurements and speeding up the currently tedious task of processing data on human affective behaviour [2].

Following the success in 3D face recognition, the face processing community is now trying to establish good 3D facial expression classification. There is a great demand for representing facial expression classification in 3D space which allows us to examine the fine structure change for universal and complex expressions [3]. 3D geometry contains ample information about human facial expression [4]. 3D scanners offer 3D geometrical data which is suitable for 3D face processing studies. 3D facial data removes the problems of illumination and pose that are inherent to 2D modality. In addition, the 3D dynamics facial data also offer out-of-plane movement that cannot be captured with 2D as well as 3D surface features which play a critical role in distinguishing subtle facial expressions.

We propose the use of 3D facial surface normals (3DFSN) to capture facial deformation caused by facial expression. Instead of using the raw 3D facial points (3DFP) which is normally provided by most of the 3D scanners, we extracted 3DFSN from the 3DFP. Surface normals are considered to be more accurate in describing facial surface changes compared to using facial points due to the fact that a surface normal depends on a facial point as well as its neighbouring facial points. We constructed a statistical model for variations in facial shape due to changes in six basic expressions Anger, Disgust, Fear, Happy, Sad and Surprise using 3DFSN as the feature vectors. In particular, we are interested in how such facial expression variations manifest themselves in terms of changes in the field of 3D facial surface normals.

Each of the basic facial expressions has levels of intensity which depend on the level of intensity of each facial feature. Intensity level of a facial expression is important as it will lead to a false impression of people's emotion if misinterpreted. For example, the smiling face with low intensity can be easily misinterpreted as a neutral facial expression [5]. A facial expression involves deformation of a collection of facial features and muscles. Classifying a facial expression from one whole face is like learning a global deformation of a face. The decomposition of a face into several modules promotes the learning of a facial local structure and therefore the most discriminative variation of the facial features in each module is emphasised. As a result, the problem of the large variation of intensity for every facial feature might be solved as the face decomposition will help to put more weight on each facial feature.

This paper reports the experiments using 3DFSN with modular approach and weighted voting scheme in 3D facial expression classification and it is organized as follows. In section II, we give the preliminaries: survey on 3D facial features used in 3D facial expression classification and we explain about existing modular-based works. Section III introduced our proposed approach which begins with 3DFSN, modular facial expression classification and our framework. Section IV described experimental setting and presents the experiments analysis. Finally we draw our conclusions in Section V.

II. PRELIMINARIES

A. 3D Facial Features

Combinations of facial features form a human facial expression. Therefore, investigating the deformation of facial features should be a suitable approach in order to determine the facial expressions shown by the subjects. Facial features can be classified as being permanent or transient [6]. Eyes, lips, eyebrows and cheeks are the permanent features. Facial lines, brow wrinkles and deepened furrows are the examples of transient features that appear with changes in expression and disappear on a neutral face.

The question is which 3D properties best describe the deformation of facial features so that a higher rate of facial expression classification can be achieved. These 3D properties chosen should be able to capture the rate of change of the facial surface efficiently. Works focusing on finding the best features to represent the facial deformation in facial expression classification are not as extensive as in the face recognition area [7].

The first study of classifying facial expressions using 3D data was carried out by Wang et al. [8]. They extracted and labelled the primitive 3D surface features and derived their statistical distributions to represent the distinct prototypical facial expressions. The expression classification is based on the distribution of the facial surface labels over the face. Thus, the same type of facial expression is expected to share a similar primitive label distribution. However, their partitioned modules do not contain the mouth and eyes, which are significant modules in determining the facial expression. Furthermore, this technique requires extensive computation of curvature features which is rather challenging.

Soyel et al. [9] used six different Euclidean distances between 3D facial landmarks to form a distance vector. They only used 11 facial features to extract the distances by utilizing facial symmetry. The distance vector is derived for every 3D model and is used to compare faces for facial expression classification. They extended their work by introducing an automatic feature selection mechanism [10]. In their extended work, all 83 facial features available in the BU-3DFE Database which was developed by Binghamton University are used to find distances between points.

Another important result was obtained by Tang et al. [11] and their work was based on the ratio of distances. A set of 96 features were devised based on properties of the line segments connecting facial feature points on a 3D face model. The features consisted of the normalized distances and slopes of the line segments connecting a subset of the 83 facial feature points. To ensure the features are person-independent, the distance features are normalized by facial animation parameter units (FAPUs).

Gong et al. [12] suggested an automatic facial expression classification approach by exploring shape deformation. The shape of an expressional 3D face is assumed as the sum of two parts, a basic facial shape component (BFSC) and an expressional shape component (ESC). The expression descriptors are computed by taking the surface changes between the original expressional face and its BFSC at selected modules.

Another approach to extract features from 3D image sequences is to map the 3D data into 2D representation. The original z-values at each x, y position were used together with the depth map of the 3D facial meshes as a 2D representation in [13]. The Scale Invariant Feature Transform (SIFT) algorithm was then applied to extract features. SIFT transforms an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion [14]. The facial landmarks were used as the key-points for the algorithms.

[15] categorized 3D facial features used in static facial expression classification into four different approaches: (1) Distance-based features [9-11], (2) Patch-based features [8, 16], (3) Morphable Models [12] and (4) 2D representations [13, 17]. A detailed discussion about 3D facial features used in facial expression classification can be found in [7, 18, 19]. The BU-3DFE database provides six basic facial expressions with four levels of intensity. Studies conducted by [8-13] who used this database only used the 2 highest intensities for every kind of expression.

In addition to the previous works on 3D facial features in facial expression classification, there are also works which based on the surface normals concept, Ceolin [20] and Sandbach et al. [21].

[20] aims to fit the statistical models of shape to 2D facial images and recover the information concerning 3D shape from these images. She used a 2.5D shape representation based on facial surface normals which is acquired from 2D intensity

images using Shape from Shading (SFS). SFS is known to recover surface shape from variations in brightness and it is more natural as it captures features of human vision system. The 2.5D surface normals (or known as facial needle maps) are then used to classify facial expression and gender.

On the other hand, Sandbach et al. [21] proposed a new feature descriptor, local normal binary patterns (LNBPs), which is exploited for detection of facial action units (AUs). AUs represent the muscular activity that produces facial appearance changes [22]. There are over 45 distinct AUs corresponding to a distinct muscle which are essentially facial phonemes that can be assembled to form facial expressions [23]. Two descriptors are formed: (1) $LNBP_{OA}$, which calculates the scalar of two normals and (2) $LNBP_{TA}$, which calculates the difference of two angles of the normals, the azimuth and the elevation. Feature vectors are then formed for each of the descriptors through the use of histogram. These histograms are concatenated into one large feature vectors.

B. Modular-based Works

A pure eigenface system can be fooled by gross variations in the input image (hats, beards, etc). [24] introduced the modular eigenspaces (or eigenfeatures) used in face recognition. According to them, the modular description allows for the incorporation of important facial features such eyes, nose and mouth. They showed that eigenfeatures alone were sufficient in achieving a 95% recognition rate in their experiment. By using a combination of eigenfeatures and an eigenface representation, a slight improvement face recognition was obtained. They also showed that a modular representation has the advantage of disambiguating false eigenface matches due to gross variations in the input image.

There are also several studies that employed face decomposition and most of them are based on a linear combination approach. [25] used a collection of Principal Component Analysis (PCA) sub-models that are independently trained but share boundaries. Their findings strengthen the hypothesis that a module-based model is better than a holistic approach and the module-based approach increases flexibility for local deformations. [26] also showed a significant result especially when there are large variations in facial expression and illumination. [27] presented a combination approach of morphable models and modular-based recognition where in their first step, 3D morphable model is used to generate 3D face models from only two input images. Then, a vast number of synthetic face images are created and used to train a component-based face recognition system. The work of [25] was based on 3D data and 2D data in [26, 27]. However, there is no facial expression classification results recorded in [25] as this work was developed for animation purposes while [26, 27] was for face recognition.

[26] discovered that if the face images are divided into very small modules the global information of the face may be lost and the accuracy of this approach is no longer acceptable. Thus, choosing the size of the modules to represent a face is also vital. Chiang et al [28] divided the face into five modules which included the left eye, the right eye, the nose, the mouth, and the bare face with each facial module identified by a facial landmark at the module centre. Fourteen modules were extracted from every face image based on the correspondence information given by the morphable model in [27]. The modules included eyebrows, eyes, area between eyebrows, bridge of nose, right lip, left lip, both cheeks, centre of mouth, entire mouth and both side of nose.

III. PROPOSED APPROACHES

A. 3D Facial Surface Normals

With the availability of raw 3D facial data, a 3D facial surface normal extraction is a straightforward task. A surface normal is a vector that is perpendicular to the tangent plane to a surface at a point. In addition, surface normals are also the features that encode the local directional gradient.

We believe that each expression has a consistent distribution of surface normals which distinguish it from other expressions. When the facial expression changes, the facial points positions also change which will cause its surface normal distribution to change. Surface normals are considered to be more accurate in describing facial surface changes compared to using facial point due to the fact that a surface normal is built from a 3D facial point as well as its neighbouring points.

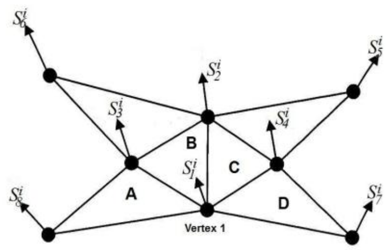

Figure 1. A triangular polygon with its surface normals.

Let $\mathbf{F_i}$ be a 3D face of the i^{th} subject. $\mathbf{F_i}$ is represented by the set of 3DFP

$$\mathbf{F_i} = \{p_1^i, p_2^i, \cdots, p_N^i\} \tag{1}$$

where the $p^i s$ are the (x, y, z) coordinate of each 3D facial point and N is the number of 3DFP in the face.

At each of the 3DFP on the facial surface, we encode the facial points using their unit surface normal vectors,

$$\mathbf{\Omega}_{\mathbf{i}} = \{s_1^i, s_{2,}^i, \cdots, s_N^i\} \quad (2)$$

where the $s_k^i s$ are the 3D unit normals $s_k^i = \{s_x, s_y, s_z\}$.

Once all of the triangular polygon normals are calculated, the normal for each vertex in the triangulated face data is computed by averaging the normals of the neighbouring polygon surface normals. Figure 1 shows an example of the triangular polygon with its vertex normals.

The difference between our work and [20] is that the surface normals are acquired from 2D intensity images using SFS. In our work, the surface normals are calculated using 3DFP. Our approach differs from [21] in the sense of calculating the normals method. In their work, the unit normal $\mathbf{n}_p$ at each point v_p which is regularly spaced at a *r* radius and *P* points around the circle is calculated. While in our work, the surface normal of a point is calculated by taking into account the surface normal of the points that are connected to that particular point on the face mesh. This means that no exact amount of points or the size of area that must be considered. Furthermore, [14] used the histograms of the surface normals to form the feature vector, whereas we used the surface normals directly as the descriptor in a statistical model.

B. Modular Facial Expression Classification

By decomposing a face into several modules, we are able to learn the local structure of each facial feature and thus the classification of the facial expression should improve. Different combinations of facial features and muscles produce different types of facial expressions. Facial expression varies from one person to another depending on their facial musculature, bone structure, facial features shapes, wrinkles, and so on. The intensity of facial features for each of the facial expression varies as well.

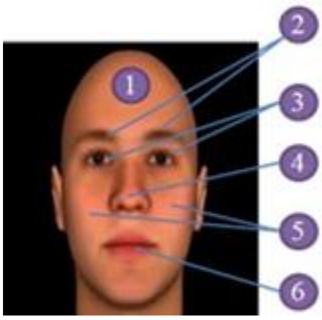

Figure 2. Face decomposition into modules.

In this work, a face is divided into six modules as illustrated in figure 2. Since the six basic facial expressions only involve symmetrical deformation for facial features that have pairs, there is no point in putting them in different modules and thus the left and right eyebrows as well as the eyes are in one module.

No facial feature in the forehead is involved in the deformation in six basic facial expressions. However we also include the forehead module in this work because we wanted to see how it influences each of the expressions.

Another interesting question is the facial features that are on the boundaries of a few modules. For example, the left and right nose saddle that belongs to the Nose and Cheeks modules. We cannot just simply put these facial features in only one module and completely ignore its effect on the other module. For that reason, we decided to include those boundary facial features in any modules that have them.

As mentioned before, in this work, the facial expression classification for each module is done independently. Therefore, each module is expected to produce a different classification from other modules. The problem arises when to decide which facial expression is being portrayed by the 3D probe in general when the results of each module classification are different. Each module has its own impact level to the six basic expressions. The simplest way to find which modules affect facial expression is to find their priority weighting.

Silapachote et al. [29] select facial features using Adaptive Boosting (AdaBoost) and successfully single out the discriminative features. Consequently the discriminative image modules without relying on a priori domain knowledge are also determined. Their experiment showed that AdaBoost has successfully picked the mouth and the eyes as being informative and it also discarded irrelevant modules. However, they used 2D images as their data and only four expressions are considered in their experiments which are Neutral, Smile, Anger and Scream.

In this work, we also used AdaBoost to determine the important facial features where we directly used the MultiBoost software package developed by Benbouzid et al [30]. Our work differs from [29] in that we used 3D facial raw points as the feature vectors in AdaBoost whereas [29] used histograms of Gabor and Gaussian derivative responses as appearance features. The important outcome of this phase in our work is the weighting priority for each of the module which [29] do not provide in their results.

TABLE I. FACE MODULES AND ITS WEIGHT [5].

Rank	Module	Weight
1	Eyebrows	0.1984
2	Mouth	0.1848
3	Eyes	0.1740
4	Cheeks	0.1719
5	Nose	0.1419
6	Forehead	0.1291

Figure 2 shows the face decomposition with the priority rank obtained from the AdaBoost result and Table 1 shows the weighting for each face module. The number on each module denotes their priority rank. In agreement with [29], the eyebrows are the most important facial module in facial expression classification, followed by the mouth and the eyes module. The Nose module only significantly deforms in the Disgust expression, which explains its rather low weighting. As expected, the forehead module has the lowest weighting and this is due to facial features that only exist at the forehead border.

C. Our Framework

Figure 3 shows the framework of our modular 3D facial expression classification. Initially, the face is decomposed into six modules and the facial expression classification for each module is done independently.

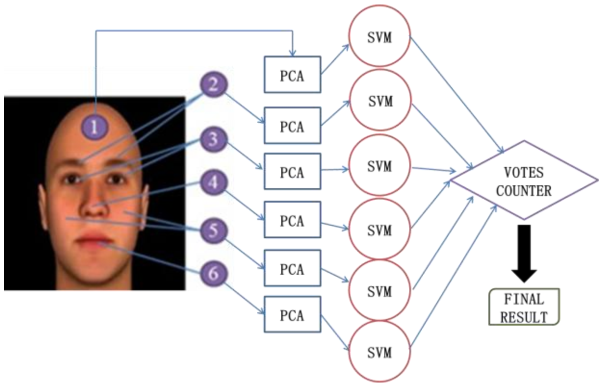

Figure 3. Our framework.

The 3D feature for each module is passed to the PCA algorithm to generate the shape weights. PCA is often used as a method which is able to get the shape representation of the face by the principal components. In this work, the goal of the PCA algorithm is to determine the principal directions of variation of the data within the data cloud. A total of 390 faces are used in this experiment, which covers both training and testing sets. Let $\mathbf{\Omega}_i$ be the training 3DFSN of the i^{th} person which has 115 3DFSN.

$$\mathbf{\Omega}_i = \left[s_{x,1}, s_{y,1}, s_{z,1}, s_{x,2}, s_{y,2}, s_{z,2}, \dots, s_{x,N}, s_{y,N}, s_{z,N}\right]^T \quad (3)$$

The 3DFSN are mean centred by subtracting the mean facial surface normal from each 3DFSN vector. Let $\bar{\mathbf{m}}$ represent the mean of the 3DFSN across the set of training images:

$$\bar{\mathbf{m}} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{\Omega}_i \quad (4)$$

Let $\mathbf{d}_i$ be defined as mean centred 3DFSN:

$$\mathbf{d}_i = \mathbf{\Omega}_i - \bar{\mathbf{m}} \quad (5)$$

The covariance matrix given by:

$$\mathbf{\Sigma} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{d}_i \mathbf{d}_i^{\mathrm{T}} \quad (6)$$

Next, we find vectors $\mathbf{u}_j$ and scalars λ_j which are the eigenvectors and eigenvalues of the covariance matrix, $\mathbf{\Sigma}$. To determine the number of principal components to use, we first rank the eigenvalues, λ_j's in decreasing order. In this work, we chose to retain 97% of the variance. At this stage, each $\mathbf{d}_i$ can be represented as a linear combination of the eigenvectors $\mathbf{u}_j$:

$$\mathbf{d}_i = \sum_{j=1}^{k} \omega_{ij}\mathbf{u}_j \quad (7)$$

The shape weights ω_{ij} are used as the feature vectors in the Support Vector Machine (SVM). Despite being inherently binary, SVM can also solve multiclass problems and in this work, we used the publicly free software package called SVMmulticlass [31] which implements a One-Against-All (OVA) multi-class approach.

Due to this independent mode, each module is expected to produce a different classification result from other modules. Each module now has been classified as showing one of the facial expressions. The class weight and result for each module are passed into the Votes Counter algorithm.

In the Votes Counter, a weighted voting system (WVS) is employed where WVS is based on the idea that not all voters are equal. In other words, one in which the preferences of some voters carry more weight than the preferences of other voters. In our work, WVS was used to determine the facial expression class for the 3D probe based on the class that has been determined for each face module. Each module (voter) has its own weight where the weight of the modules that belong to the same facial expression class is summed up. At this stage, each facial expression class has its accumulated weight and the facial expression class with the highest weight was considered as the facial expression shown by the 3D probe.

For the purpose of evaluation, we also carried out experiments using the Majority Voting Scheme (MVS). In MVS, the final classification of multiple classifications goes to the class with the majority vote. However, in the case of two or more classes having equal votes, our algorithm will classify the final expression as a false positive (FP).

IV. Experimental Setting and Analysis

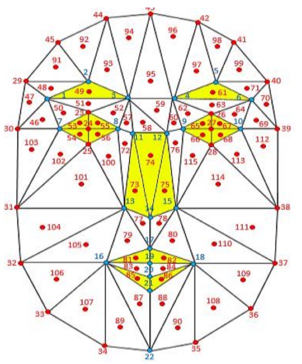

Figure 4. A 3D face using the complete set of 115 3DFP with Delaunay triangulation.

We conducted a series of experiments using the Bosphorus database [18]. The Bosphorus database provides 24 manually annotated facial landmarks, provided that they are visible in the scan. However, only 22 of the provided facial landmarks were used as the two facial landmarks (both earlobes) were not visible in the frontal scan and thus the 3D correspondence of both earlobes could not be computed. Due to the small amount of facial landmarks provided by the developer, we have developed an algorithm to elect an additional of 93 facial points. Figure 4 shows an example of a 3D face using the complete set of 115 3DFP with Delaunay triangulation. The yellow areas in figure 4 denote the eyebrows, eyes, nose and mouth. 3DFP of each face were then aligned before any value comparison between the faces takes place. This is to ensure the 3D faces are as closely aligned to each other as possible while keeping the shape unchanged.

For the purpose of evaluation, two other features were tested in order to compare them with 3D facial surface normals: (1) 3DFP and (2) 3D facial distance measurements (3DFDM) for each 3DFP to the nose tip which is defined in

$$\|d_i\| = \sqrt{(x_{13} - x_i)^2 + (y_{13} - y_i)^2 + (z_{13} - z_i)^2} \tag{8}$$

where i=1,2⋯115.

Therefore, the shape weights of the 3DFP and the 3DFDM must also be computed using PCA before the classification phase takes place. As mentioned in the previous section, we also carried out experiments using the Majority Voting Scheme (MVS) in our work. Figure 5 shows six basic facial expressions used in this work.

Figure 5. Emotion-specified facial expressions. From left: Anger. Disgust, Fear, Happy, Sad, Surprise [15].

A. *Comparison with other 3D features*

Figure 6 and 7 show the successful classification rate of WVS across the 3D features, respectively. Similar to 3DFP, 3DFSN record a 100% correct classification for happy expression. For disgust expression, 3DFSN has a 29% higher classification rate than using 3DFP. 3DFSN have an equal rate of correct classification for the Surprise expression with 3DFP. Overall, 3DFSN perform quite well compared to the two other feature vectors where the average classification of 3DFSN is 66% compared to 50% and 46% for 3DFP and 3DFDM respectively. Across voting schemes, 3DFSN performs the best. Moreover, WVS shows the highest successful rate when compared to MVS.

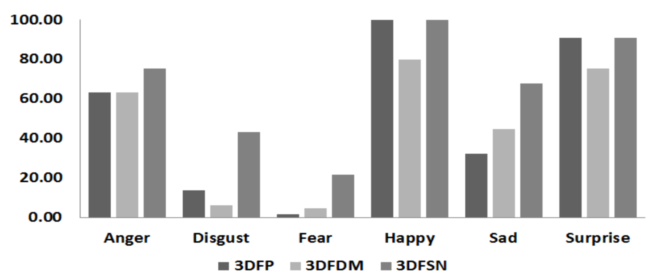

Figure 6. Successful rate of WVS.

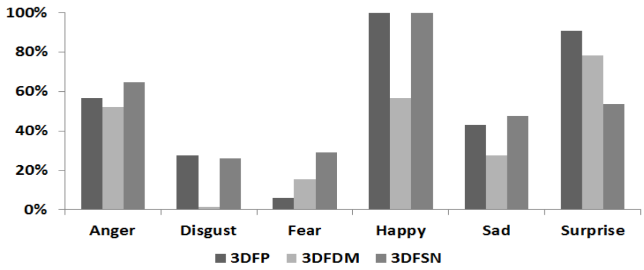

Figure 7. Successful rate of MVS.

TABLE II. F1- MEASURE FOR FACIAL EXPRESSION CLASSIFICATION RESULTS USING 3DFP, 3DFDM, 3DFSM WITH WEIGHTED VOTING SCHEME

EXPRESSION	3DFP	3DFDM	3DFSN
ANGER	53.25%	52.56%	64.05%
DISGUST	19.35%	11.43%	**56.57%**
FEAR	2.82%	7.14%	**33.73%**
HAPPY	73.03%	62.65%	**88.00%**
SAD	40.00%	42.34%	**71.54%**
SURPRISE	65.92%	58.68%	**67.43%**
AVERAGE	42.40%	39.13%	63.63%

TABLE III. F1- MEASURE FOR FACIAL EXPRESSION CLASSIFICATION RESULTS USING 3DFP, 3DFDM, 3DFSM WITH MAJORITY VOTING SCHEME

EXPRESSION	3DFP	3DFDM	3DFSN
ANGER	57.85%	58.33%	**65.19%**
DISGUST	20.51%	0.00%	47.06%
FEAR	3.77%	3.70%	24.14%
HAPPY	75.14%	68.33%	**88.00%**
SAD	42.67%	44.44%	65.00%
SURPRISE	66.67%	66.67%	66.21%
AVERAGE	44.44%	40.25%	59.41%

Table II and table III show scores computed in terms of F1-measure using WVS and MVS. Anger expression using 3DFSN with MVS shows a slightly high score compared to 3DFSN with WVS. Happy expression for both voting schemes and using 3DFSN has the same score. For the rest of the expressions, 3DFSN with WVS has the highest scores.

B. *Comparison with a non-modular approach*

For the purpose of assessment of our modular approach, we carried out a non–modular facial expression classification using 3DFSN. Table IV shows scores computed in terms of F1-measure for a non-modular. Comparing between Table II, III and IV, 3DFSN for a non-modular only shows a slightly high score in classification of fear expression.

TABLE IV. F1- MEASURE FOR FACIAL EXPRESSION CLASSIFICATION RESULTS USING 3DFP, 3DFDM, 3DFSM OF A NON-MODULAR APPROACH.

EXPRESSION	3DFP	3DFDM	3DFSN
ANGER	50.68%	42.50%	56.38%
DISGUST	37.11%	2.74%	33.66%
FEAR	10.26%	21.51%	**38.78%**
HAPPY	82.28%	51.03%	73.03%
SAD	45.53%	28.57%	50.41%
SURPRISE	66.29%	55.74%	53.44%
AVERAGE	48.69%	33.68%	50.95%

C. *Comparison with other studies*

Table V shows the score using F1-measure of the previous works. [4],[12],[13] and [32] used the BU-3DFE database in their study and they only included each of the facial expressions with the two highest levels of intensity. [4], [12] and [13] used the similar experimental setting and the classification is carried out using SVM with RBF kernel. [32] used a slightly different experimental setting and they used multiclass-SVM as their classifier. We used Bosphorus database in our work with multiclass-SVM. AS mentioned, Bosphorus database does not provide facial expressions with intensity information. Therefore, the data could be ranging from low intensity up to high intensity. We believe this is the reason for the significant difference in the classification performance as the facial expression with higher intensity means the deformation of each facial feature is apparent and easy to classify. Even with the inconsistent intensity level type of data, we managed to have a higher score than works in [12] and [13] for happy expression.

TABLE V. F1- MEASURE FOR FACIAL EXPRESSION CLASSIFICATION RESULTS OF PREVIOUS WORKS.

EXPRESSION	[7]	[19]	[27]	[30]
ANGER	69.88%	83.21%	77.82%	92.44%
DISGUST	76.04%	88.26%	77.17%	96.82%
FEAR	64.33%	77.09%	68.21%	95.37%
HAPPY	83.65%	93.51%	77.74%	98.09%
SAD	78.11%	82.83%	79.51%	91.49%
SURPRISE	85.05%	97.16%	89.83%	99.35%
AVERAGE	76.18%	87.01%	78.38%	95.59%

Figure 8 shows four different subjects from Bosphorus database portray happy expression with no intensity information in the first row and angry expression in the third row while the illustration of different intensity of happy expression in the second row and angry expression in the fourth row taken from BU-3DFE database [3]. Happy expression managed to achieve a 100% classification rate whereas none of the previous works have achieved this. This happened as each subject shows practically the same intensity level for happy expression in the Bosphorus database. We also believe the intensity of happy expression portray by the subjects from the Bosphorus Database are in the highest range when compared to BU-3DFE database [3] facial expression intensity levels (see figure 8). However, for anger expression, the deformation of facial features of Bosphorus data is not consistent as some of the subjects do not show an obvious deformation of eyebrows, eyes and mouth (see the highest intensity level of angry expression from BU-3DFE database in the fourth row for comparison).

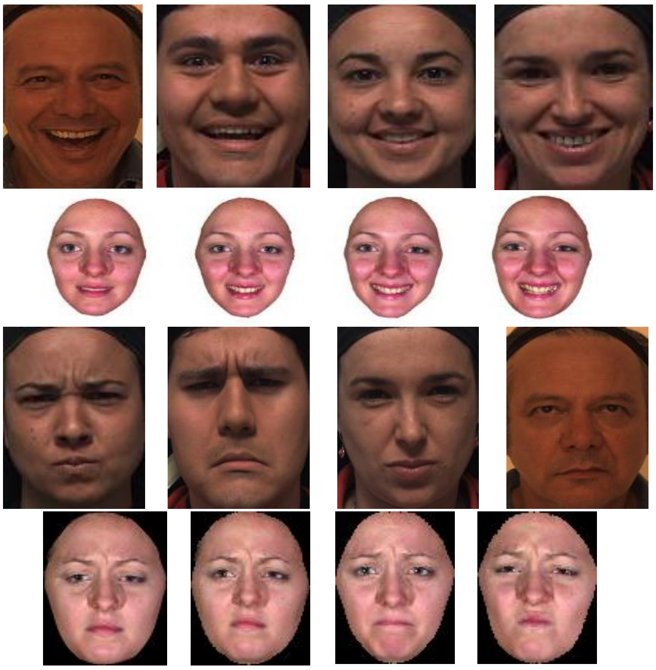

Figure 8. The four different subjects from the Bosphorus Database portray Happy expression (with no intensity information) in the first row and Angry expression in the third row. The subject in the second row show four different intensities of Happy expression and Anger expression in the fourth row, ranging from low (left) to high (right) intensity (taken from BU-3DFE database [28]).

D. Issues on Disgust and Fear Expressions

Based on the facial features rank experiment (refer Table I), the deformation of the eyebrows is significant in any facial expression. Therefore, we decided to assign eyebrows in a separate module instead of combining it with the eyes module. However it turns out that the results for the eyebrows module independently are not as good as the other modules. We also discovered that the small number of 3D facial features computed in this module is the reason for the poor result.

If we look closely at disgust and fear expressions for the three different subjects in figure 9, the differences are obvious specifically in the eyebrows, mouth and eyes modules which are the modules with the highest weight. The large differences in those modules are significant in the classification phase, where for example, for the most intense fear expression; there is an opened mouth as in the Surprise expression (see figure 9 for comparison). However, subject B and C did not show the same mouth deformation. To differentiate between fear and surprise, the eyebrows for the fear expression should be showing a different deformation in the eyebrows to the surprise expression. However, subject C is still showing the similar deformation in the eyebrows in the surprise and fear expression; hence it is obvious we cannot differentiate the surprise and fear expression because of the same deformation in Eyebrows for both expressions.

Figure 9. Three subjects A, B and C (from left to right) with Disgust (first row), Fear (second row) and Surprise (third row) expressions.

V. CONCLUSIONS AND FUTURE WORKS

Our work is different with the existing works as they used the dataset with facial intensity information while we used dataset with no intensity. A statistical modelling of 3DFSN is used along with a modular approach. The resulting shape model is used to perform facial expression classification on each module. The classifications result of each module is combined using a WVS to determine the final classification of a facial expression. We believe that each expression has a consistent distribution of surface normals which distinguish it from other expressions and therefore the facial deformation of each facial expression is easily monitored. By using the modular approach, the discriminative variations of the facial features in each module are emphasised. Our results are degraded due to the intensity differences in each subject shown in the Bosphorus database compared to the existing works that used only two highest facial expression intensity data from BU-3DFE database. However, we have shown that surface normals does has a potential to be used as the feature vectors to classify six basic expressions.

For our future work, we would like to carry out facial expression classification experiments using 3D facial data with the highest intensity information. We believe that our proposed approach will achieve a good result using this type of data. Furthermore, we would like to train our system using facial expressions with different intensity levels and as a result, we will be able to classify the intensity level of facial expression.

Dynamic information is very important in order to analyse more subtle facial expressions. We believe that with dynamic information fused with our modular approach, we could monitor the facial features deformation of each module easily and different intensities of each facial feature can be captured. In the future, we would like to use 3DFSN as the feature vector in a modular approach with 3D facial dynamic data.

References

[1] Jaimes, A. and Sebe, N. (2007) Multimodal Human – Computer Interaction: A Survey. Jurnal on Computer Vision and Image Understanding, Volume 108, Issue 1-2, pp. 116 – 134.

[2] Ekman, P. and Rosenberg, E. L. (2005) What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), second Ed., Oxford University Press.

[3] Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J. (2006) A 3D Facial Expression Database For Facial Behaviour Research, 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 211 – 216.

[4] Tang, H. and Huang, T.S. (2008) 3D Facial Expression Recognition Based on Properties of Line Segments Connecting Facial Feature Points. 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1-6.

[5] Beszédeš, M. and Culverhouse, P.F. (2007) Facial Emotions and Emotion Intensity Levels Classification and Classification Evaluation. British Machine Vision Conference, pp.1-10.

[6] Bettadapura, V., (2012) Face Expression Recognition and Analysis: The State of the Art. Computing Research Repository Journal (CoRR), Volume abs/1203.6722, pp.1-27.

[7] Vezzetti, E. and Marcolin, F. (2012) 3D Human Face Description: Landmarks Measures and Geometrical Features. Journal of Image and Vision Computing, Volume 30, Issue 10, pp. 698–712.

[8] Wang, J., Yin, L., Wei, X., and Sun, Y. (2006) 3D Facial Expression Based on Primitive Surface Feature Distribution. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1399-1406.

[9] Soyel, H. and Demirel, H. (2007) Facial Expression Recognition using 3D Facial Feature Distances. LNCS Book Series: Image and Analysis Recognition, Springer , Volume 4633, pp 831-838.

[10] Soyel, H. and Demirel, H. (2010) Optimal Feature Selection for 3D Facial Expression Recognition using Coarse-to-Fine Classification. Turkish Journal of Electrical Engineering & Computer Sciences. 18(6), pp.1031-1040.

[11] Tang, H. and Huang, T.S. (2008) 3D Facial Expression Recognition Based on Properties of Line Segments Connecting Facial Feature Points. 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1-6.

[12] Gong, B., Wang, Y., Liu, J., and Tang, X. (2009) Automatic Facial Expression Recognition on A Single 3D Face by Exploring Shape Deformation. 17th ACM International Conference on Multimedia, pp. 569-572.

[13] Berretti, S., Ben Amor, B., Daoudi, M. and Bimbo, A. (2011) 3D Facial Expression Recognition using SIFT Descriptors of Automatically Detected Keypoints, Journal of Visual Computing, Volume 27, Issue 11, pp.1021-1036.

[14] Abhishek Verma, Chengjun Liu, Jiancheng Jia. 2011. New Colour SIFT Descriptors for Image Classification with Applications to Biometrics. International Journal of Biometrics (IJBM), Volume 3 Issue 1, pp. 56-75.

[15] Sandbach, G. Zafeiriou, S. and Pantic, M. and Yin, L. (2012). Static and Dynamic 3D Facial Expression Recognition: A comprehensive survey, Journal of Image and Vision Computing, Volume 30, Issue 10, pp. 683-697.

[16] Maalej, A., Ben Amor, B., Daoudi, M., Srivastava, A. and Berreti, S. (2010) Local 3D Shape Analysis for 3D Facial Expression Recognition. International Conference on Pattern Recognition, pp. 4129 – 4132.

[17] Pinto, S.C.D., Mena-Chalco, J.P., Lopes, F.M., Velho, L. and Cesar, R.M. (2011) 3D Facial Expression Analysis by Using 2D and 3D Wavelet Transforms. 18th IEEE International Conference on Image Processing (ICIP), pp. 1281 – 1284.

[18] Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B. and Lale, A.,(2008). Bosphorus Database for 3D Face Analysis. In: Schouten, B., Juul, N.C., Drygajlo, A. and Tistarelli, M. (Eds), Biometrics and Identity Management. Springer – Verlag, Berlin, Heidelberg, pp. 47-56.

[19] Fang, T., Zhao, X., Ocegueda, O., Shah, S.K. and Kakadiaris, I.A. 2011. 3D Facial Expression Recognition: A Perspective on Promises and Challenges. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pp. 603 – 610.

[20] Ceolin. S.R., Smith, W.A.P. and Hancock, E. (2008). Facial Shape Spaces from Surface Normals. Lecture Notes in Computer Science: Image Analysis and Recognition, Volume 5112, pp. 955-965.

[21] Sandbach, G., Zafeiriou, S., and Pantic, M. 2012. Local Normal Binary Patterns for 3D Facial Action Unit Detection. Proceedings of the 19th IEEE International Conference on Image Processing (ICIP), pp. 1813- 1816.

[22] Ekman, P. and Friesen, W.V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto.

[23] Kapoor, A., Qi, Y. and Picard., R.W. 2003. Fully Automatic Upper Facial Action Recognition. IEEE International Workshop on Analysis and Modelling of Faces and Gestures, pp. 195 – 202.

[24] Pentland, A., Moghaddam, B. and Starner, T. (1994) View – based and Modular Eigenspaces for Face Recognition. IEEE Conference on Computer Vision and Pattern Recognition, pp. 84 – 91.

[25] Tena, J.R., De la Torre, F. and Matthews, I. (2011) Interactive Region-Based Linear 3D Face Model. ACM Transactions on Graphics. Volume 30 Issue 4, Article 76

[26] Gottumukkal, R. and Asari, V. K. (2003) An Improved Face Recognition Technique based on Modular PCA Approach. Pattern Recognition Letter. Volume 25 Issue 4, pp. 429 – 436.

[27] Huang, J., Blanz, V. and Heisele, B. (2002) Face Recognition using Component-Based SVM Classification and Morphable Models. Proceedings of the First. International Workshop on Pattern Recognition with Support Vector Machines, London, UK., pp.334-341.

[28] Chiang, C.C., Chen, X. and Yang, C.A. (2009) Component-Based Face Synthesizing Method, APSIPA Annual Summit and Conference, pp.24-30.

[29] Silapachote, P., Karuppiah, D. R. and Hanson, A. R. (2005). Feature Selection using AdaBoost for Face Expression Recognition. 4th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2004), pp.452-273.

[30] Benbouzid, D., Busa-Fekete, R., Casagrande, N., Collin, F.-D.and Kégl, B. (2012). MultiBoost: a Multi-Purpose Boosting Package. Journal of Machine Learning Research.13, Volume 13 Issue 1, pp. 549 – 553.

[31] Joachims, T., Finley, T., and Yu, C. (2009) Cutting-Plane Training of Structural SVMs, Journal of Machine Learning. Volume 77 Issue 1, pp. 27 – 59.

[32] Rabiu, H., Saripan, M.I., Mashohor, S. and Marhaban, M.H. 2012. 3D Facial Expression Recognition using Maximum Relevance Minimum Redundancy Geometrical Features. EURASIP Journal on Advances in Signal Processing. 2012:213.