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

Asim Jan & Hongying Meng

Contents

Introduction

- 3D facial expression recognition is important to tackle issues such as **pose variations** and **illumination** changes found in 2D imaging.
- 3D imaging capture accurate geometry information closely sensitive to expression variations.
- 3D imaging provides more detailed observations such as **depth** and **geometric data** of the facial features – this is used for more accurate detection of facial muscle movements and changes
- There are many **applications** that can benefit from this research:
  - Customer service satisfaction, advising operator of the customers mood
  - Human-computer interaction tasks – recognising a persons identity through an automated system e.g. Passport Control that uses automated facial recognition
  - *Medical applications*:
    - Detecting signs of depression, aggression and pain
    - Psychological effects in patients
    - Early signs for detecting autism

Figure 1: Biometric Passport [1]

Introduction

- Static 3D data can be categorized into two streams, i.e. feature based or model based
- **Feature based** claims 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
- **Model based** tries to simulate the physical process of generating expression and explores a generic elastically deformable face model. This can generate universal expressions by adjusting the parameters
- Feature based is mostly adopted due to the expensive computing cost required by the model based approach
- There are many available 3D facial expression databases, the ones used for this research are the Binghamton University 3D Facial Expression **(BU-3DFE)** Database and **Bosphorus** Database

Related Work

- Wang et al. produced the baseline results using **Primitive Surface Feature Distribution (PSFD)** with **Linear Discriminant Analysis (LDA)** to recognize 6 expressions (*Angry, Disgust, Fear, Happiness, Sadness and Surprise),* achieving *8*3.6%.
- Rabiu et al. achieved the highest recognition rate of 92.2% when classifying 7 expressions. Using 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.

	Mean RR	Method	Feature Domain
Wang et al. [2]	83.6%	PSFD + LDA	Geometric
Tie Yun, Ling Guan [3]	85.39%	3D Gabor Features	Texture
Xiaoli et al. [4]	90.2%	28 Geometric Features	Geometric
Soyel and Hassan [5]	91.3%	6 Distance Measures	Geometric
Lemaire et al. [6]	78.43%	SIFT Features	
Tekguc et al. [7]	88.1%	NSGA-II Features	Geometric
Rabiu et al. [8]	92.2%	16 Distance Vectors, 27 Angles	Geometric
Yurtkan et al. [9]	88.2%	Entropy Analysis	Geometric

Table 1: Existing Methods on 3D FER

Related Work

- Almost all the work available uses only one type of method (**Textured** or **Geometric**) to extract features
- Various Machine learning methods have shown there effectiveness for FER, with SVM shown to be more suited for facial expression recognition
- Our research addresses 3D facial expression recognition by investigating both the **Textured** and **Geometric** domains
- Unique descriptors from different extraction methods are combined to produce a more diverse feature, collecting the benefits from the different aspects which will be described in detail in the in the upcoming slides

Methodology

Introduction

- Facial expressions are used to visually describe the human state of emotion
- Our research proposes a method that comprehensively models the variations in visual clues
- Fusing key features obtained from the Geometric and Textured domains, to investigate how this concept impacts the overall performance
- This approach is based on:
  - How the features are extracted, how information of facial expressions can represented in different ways
  - The fusion of the various features and reduced using feature dimensionality reduction
  - Classification using machine learning techniques

Methodology

System Overview

Figure 1 illustrates the process of how the features are extracted and fused from each of the facial models, reduced in dimensionality, and used to classify expressions with machine learning.

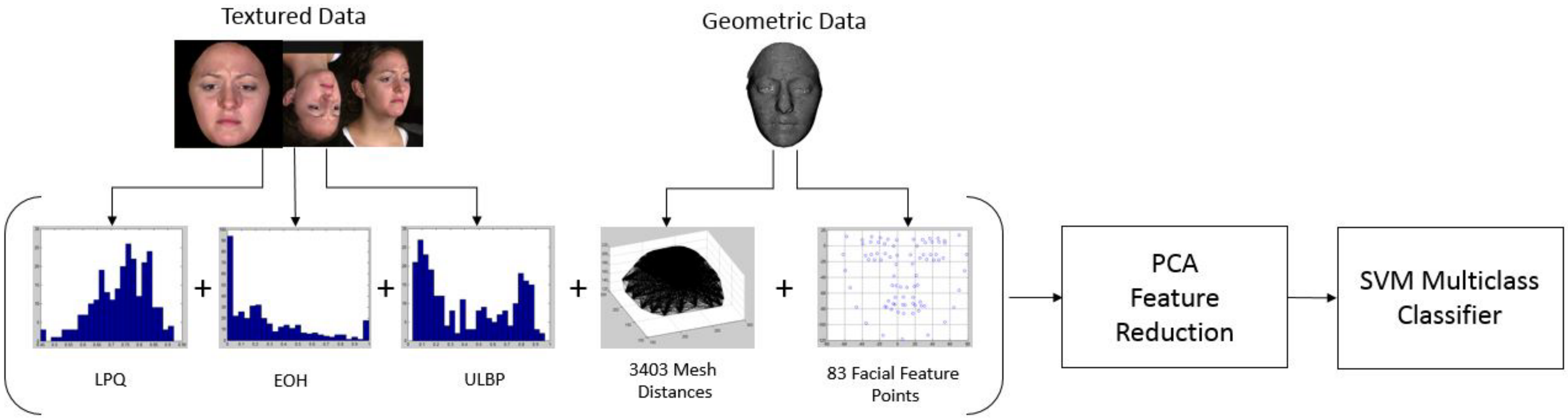


Figure 1: Overall framework of the proposed approach, combining the different features extracted using the geometric and textured data

Methodology

Textured Feature Extraction (ULBP)

- The Textured feature extraction methods are applied on facial images to provide information can can't be visibly seen
- Uniform Local Binary Patterns (ULBP) is a very common feature extraction method for 2D images, its main purpose is to describe the local texture structure of an image using binary patterns which is obtained from its surroundings
- The LBP operator compares each pixel with its surrounding 8 pixels (based on the radius size). Using a threshold, it compare if the surrounding grey-scale value is higher or lower than the centre and forms a pattern of '1's and '0's, resulting in a histogram of 256 bins
- The Uniform modification to LBP checks each pattern for bit-wise transitions (e.g. 0(01)11(10)0) to form a new pattern which produces a histogram of 59 bins
- This method is computationally efficient and effective, and has been used for face recognition [10], and many other applications [11]

Methodology

Textured Feature Extraction (EOH)

- Edge Oriented Histograms (EOH) is made on the same principles as Histogram of Oriented Gradients (HOG), and is a more efficient and powerful operator that will capture an edge or the local shape information of an image
- Edges can be detected in an image using edge operators such as ‘Sobel’ to detect to horizontal edges EH, and vertical edges EV 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
- This method is applicable to various applications in computer vision such as hand gesture recognition [12], human detection [13] and facial expression recognition [6]

Methodology

Textured Feature Extraction (LPQ)

- Local Phase Quantization (LPQ) is proposed for texture analysis [14] and applied to blurred face recognition [15] [16]
- The process involves creating blocks across the whole image; which can be any size; then to compute the local Fourier frequency coefficients for all the pixels in each block
- Before applying LPQ, each image is split into 4 equal parts
- A scalar quantizer is then applied on each part and the image as a whole to transform the coefficients into an 8-bit binary code
- Each part and the whole image produces a histogram of 256 bins which totals to 1280 bins, this produces the LPQ feature used for experimentation

Methodology

Geometric Feature Extraction (83P)

- 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

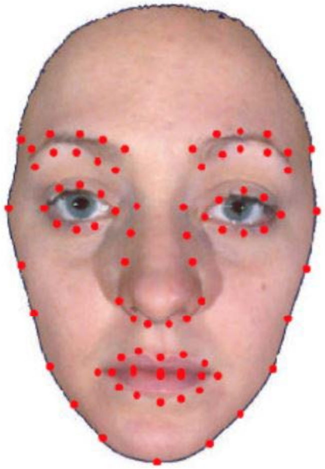


Fig. 2. 83 Facial Feature Points annotated on a cropped image.

Methodology

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.
- Principal Component Analysis (PCA) has been chosen to reduce the dimensionality of our feature vector by taking the high energy coefficients
- We use PCA to take all the relevant information from the various feature sets to produce a smaller yet significant and accurate feature vector for facial expression recognition.

Methodology

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 for MATLAB
- The approach we used was One Vs All, this meant that the classification will be done against all the classes
- The SVM has been optimised to give a good performance for our task

Experimental Results

Introduction

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.

- The Binghamton University 3D Facial Expression Database (BU3DFE) and the Bosphorus Database have been used for the experiment
- Experiments have been taken for classifying 7 expressions and 6 expressions (excluding Neutral)
- Each experiment has 3 protocols: Geometric features only, Textured features only and the Geometric and Texture features fused
- Each experiment has 100 tests and the average result is taken, each test uses 10 Fold Cross-Validation where the subjects are randomised to show robustness

Experimental Results

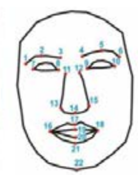
BU3DFE Dataset

- The BU-3DFE database is developed by Li et al. [17]. from Binghamton University
- This contains Facial models of 100 subjects (44 Males, 56 Females) of various ethnic backgrounds.
- Data for each subject contains 6 prototypical expressions which are Angry, Disgust, Fear, Happy, Sad, Surprise and the basic Neutral expression.
- Each expression apart from Neutral contain 4 levels of intensity: weak to strong. This totals to 2500 facial models.
- Each model contains a cropped image of the face; an uncropped image (side views of the face), 83 manually annotated landmarks and a single 3D Face Mesh Model containing 3D coordinates with a resolution of 25K to 35K polygons [17].

Experimental Results

Bosphorus Dataset

- The Bosphorus database [18] developed by A.Savran et al.
- Includes a total of 4,666 scans collected from 105 subjects - 61 male and 44 female.
- Multiple facial expressions included represented in 2 ways: the basic expressions of Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise.

- The second representation is based on Action Units.
- Each subject contains a single frontal face image and 22 - 24 3D landmarks 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.

Experimental Results

Results

- Strong suggestion that Fusing the 2D and 3D domains increase performance, results are based on classifying 7 and 6 Expressions (6 Excluding Neutral)
- Results of Bosporus Dataset also show the advantage of fusion domain features
- 83/22 Normalised Facial Feature Points (83P/22P)
- Uniform Local Binary Patterns (ULBP)
- Edge Oriented Histogram (EOH)
- Local Phase Quantization (LPQ)
- Facial Mesh Distances (FD)

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%

Table 2: Individual Feature Performance on BU3DFE Database

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

Table 4: Validation Test Using Bosphorus Database

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 3: Combined Feature Performance on BU3DFE Database

Conclusion & Discussion

- The approach was proposed for 3D facial expression recognition by fusing multiple feature extraction methods used on the textured and geometric data
- 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.
- Natural Feature Extraction methods that will best fit the application of Facial Expression Recognition can be researched, using deep learning networks such as Convolutional Neural Networks

Thank You

With the Support of Thomas Gray Gerald

References

	Reference
[1]	Bioenable, "biometric passport," 30 04 2015. [Online]. Available: http://www.bioenabletech.com/biometric-passport.
[2]	T. Ahonen, A. Hadid, and M. Pietik¨ainen. Face description with local binary patterns: application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12):2037–41, Dec. 2006.
[3]	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.
[4]	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.
[5]	H. Soyel and H. Demirel. Facial expression recognition using 3D facial feature distances. Image Analysis and Recognition, pages 1–13, 2007.
[6]	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.
[7]	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, Sept 2009.
[8]	H. Rabiu 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.
[9]	K. Yurtkan and H. Demirel. Feature selection for improved 3d facial expression recognition. Pattern Recognition Letters, 38(0):26 – 33, 2014.
[10]	T. Ahonen, A. Hadid, and M. Pietik¨ainen. Face description with local binary patterns: application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12):2037–41, Dec. 2006.
[11]	T. Ojala, M. Pietikainen, and T. Maenpaa. 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]	W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. Technical report, Mitsubishi Electric Research Labs.
[13]	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.

References

	Reference
[14]	V. Ojansivu, E. Rahtu, and J. Heikkila. 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.
[15]	Z. Lei and S. Z. Li. Fast multi-scale local phase quantization histogram for face recognition. Pattern Recognition Letters, Oct. 2012.
[16]	T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pages 1–4, Dec 2008.
[17]	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.
[18]	A. Savran, N. Alyuz, and H. Dibeklio?lu. Bosphorus database for 3D face analysis. Biometrics and Identity Management, pages 47–56, 2008.