

**DEEP LEARNING BASED THREE DIMENSIONAL FACE EXPRESSION
RECOGNITION USING GEOMETRY IMAGES FROM
THREE DIMENSIONAL FACE MODELS**

M.Sc. THESIS

Neşe GÜNEŞ

Department of Computer Engineering

Computer Engineering Programme

SEPTEMBER 2019

**DEEP LEARNING BASED THREE DIMENSIONAL FACE EXPRESSION
RECOGNITION USING GEOMETRY IMAGES FROM
THREE DIMENSIONAL FACE MODELS**

M.Sc. THESIS

**Neşe GÜNEŞ
(504161544)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Uluğ BAYAZIT

SEPTEMBER 2019

**ÜÇ BOYUTLU YÜZ MODELLERİNDEN ELDE EDİLEN
GEOMETRİ GÖRÜNTÜLERİ KULLANILAN DERİN ÖĞRENME TABANLI
ÜÇ BOYUTLU YÜZ İFADELERİNİ TANIMA**

YÜKSEK LİSANS TEZİ

**Neşe GÜNEŞ
(504161544)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Uluğ BAYAZIT

EYLÜL 2019

Neşe GÜNEŞ, a M.Sc. student of ITU Graduate School of Science Engineering and Technology 504161544 successfully defended the thesis entitled “DEEP LEARNING BASED THREE DIMENSIONAL FACE EXPRESSION RECOGNITION USING GEOMETRY IMAGES FROM THREE DIMENSIONAL FACE MODELS”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Uluğ BAYAZIT**
Istanbul Technical University

Jury Members : **Prof. Dr. Mustafa Ersel KAMAŞAK**
Istanbul Technical University

Prof. Dr. Hasan Fehmi ATEŞ
Istanbul Medipol Univesity

.....

Date of Submission : **4 September 2019**

Date of Defense : **6 September 2019**





In memory of my grandfather,



FOREWORD

In pursuit of giving an answer to the following question, we study on 3D face expression recognition task: “Can we employ geometry images to have a geometry based solution to the task of 3D face expression recognition using the power of neural networks as well?” The dissertation has been put in writing to satisfy the requirements for graduation at Computer Engineering programme in Istanbul Technical University. The question of conducted research was constructed cooperatively by my supervisor, Prof. Dr. Uluğ Bayazıt. It would be great if I could express my special gratitudes to my thesis mentor considering his moral support throughout the experimenting procedures and during the semester. I also have a need to express my thanks to my professors Prof. Dr. Gözde Ünal and Prof. Dr. Elif Sertel, without the experience I obtained at ITU-CSCRS (ITU-UHUZAM) I would not have been able to conduct this analysis. All in all it’s just another brick in the wall, I hope you enjoy reading.

September 2019

Neşe GÜNEŞ
Computer Engineer

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xix
ÖZET	xxiii
1. INTRODUCTION	1
2. MESH PROCESSING PIPELINE	5
2.1 Geometry Image Based Approach.....	5
2.2 Overview of the Pipeline	7
2.3 3D Registration.....	8
2.4 Surface Parameterization Methods.....	11
2.4.1 Conformal mapping.....	12
2.4.2 Computational conformal geometry.....	15
2.5 Color Interpolation using Barycentric Coordinates.....	16
3. DEEP LEARNING IN RECOGNITION SYSTEMS	21
3.1 Introduction	21
3.2 Convolutional Neural Networks	23
3.2.1 Architecture of a traditional convolutional neural network.....	24
4. APPLICATIONS	27
4.1 The BU-3DFE Facial Expression Database	27
4.2 Experiments.....	30
4.2.1 Feature extractor: vggnet architecture.....	30
4.2.2 Binary classifier: one-versus-one classification	33
4.2.3 Binary classifier: one-versus-all classification	33
4.3 Experimental Results.....	34
4.3.1 3d fer using gim database: one-versus-one classification with 7.2m learnable parameters.....	34
4.3.2 3d fer using gim database: one-versus-one classification with 16.8m learnable parameters.....	37
4.3.3 3d fer using gim database: one-versus-all classification with 7.2m learnable parameters.....	39
5. CONCLUSION	41
REFERENCES	43
APPENDICES	47



ABBREVIATIONS

2D	: 2 Dimensional
3D	: 3 Dimensional
Acc	: Accuracy
AE	: Autoencoder
ANN	: Artificial Neural Network
AUs	: Action Units
BU-3DFE	: Binghamton University 3D Facial Expression Database
CNNs	: Convolutional Neural Networks
DL	: Deep Learning
F2D	: 2D Facial Expression Image
F3D	: 3D Facial Expression Model
FACS	: Facial Action Coding System
FDP	: Facial Definition Parameters
FER	: Face Expression Recognition
GIM	: Geometry Image
GPU	: Graphics Processing Unit
ILSVRC	: ImageNet Large-Scale Visual Recognition Challenge
L2	: Euclidean Distance
ML	: Machine Learning
MPEG	: Moving Pictures Expert Group
NN	: Neural Network
PCA	: Principal Component Analysis
ReLU	: Rectified Linear Unit
SVM	: Support Vector Machines
VGG	: Visual Geometry Group



LIST OF TABLES

	<u>Page</u>
Table 1.1 : Expression classes considering closest emotions.....	3
Table 4.1 : Principal details of commonly utilized 3D face expression datasets which contain non-neutral faces.	27
Table 4.2 : 3D dynamic (plus time dimension) and 3D static face databases for 3D face expression analysis.	28
Table 4.3 : One-versus-one recognition Results: 79.5% accuracy in average on test data. One binary classifier for each expression performs a one-versus-one classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, the last four convolutional layers of the VGG16 model are trainable, in total 7.2M parameters are learnable.....	36
Table 4.4 : One-versus-one recognition results: 81.4% accuracy in average on test data with leave-10-person-out data split method. One binary classifier for each expression performs a one-versus-one classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, all of the sixteen layers of the VGG16 model are trainable, in total 16.8M parameters are learnable.....	38
Table 4.5 : One-versus-all recognition results: 72.8% accuracy in average on test data with leave-6-person-out data split method. One binary classifier for each expression performs a one-versus-all classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, the last four convolutional layers of the VGG16 model are trainable, in total 16.8M parameters are learnable.....	40
Table 1 : Angry expression recognition with female models: loss and accuracy values.	48
Table 2 : Angry expression recognition with male models: loss and accuracy values.	48
Table 3 : Disgust expression recognition with female models: loss and accuracy values.	49
Table 4 : Disgust expression recognition with male models: loss and accuracy values.	49
Table 5 : Fear expression recognition with female models: loss and accuracy values.	50
Table 6 : Fear expression recognition with male models: loss and accuracy values.	50
Table 7 : Happy expression recognition with female models: loss and accuracy values.	51

Table 8 : Happy expression recognition with male models: loss and accuracy values. 51

Table 9 : Sad expression recognition with female models: loss and accuracy values. 52

Table 10 : Sad expression recognition with male models: loss and accuracy values. 52

Table 11 : Surprise expression recognition with female models: loss and accuracy values. 53

Table 12 : Surprise expression recognition with male models: loss and accuracy values. 53



LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : Rendering of a geometry image. (a) The initial face scan. (b) Geometry image is visualized as [r, g, b], 12 bits/channel. (c) Reconstruction of geometry, obtained completely from (b).	6
Figure 2.2 : Proposed mesh processing pipeline.....	8
Figure 2.3 : Matching bounding boxes. (a) Before matching. (b) After matching.	9
Figure 2.4 : Depiction of 3D registration. (a) The original face scans with different poses, side view. (b) The original face scans with different poses, frontal view. (c) The resulting transformation after alignment, side view. (d) The resulting transformation after alignment, frontal view.	10
Figure 2.5 : Fixed Boundary Conformal Mapping on a circle. (L) Original mesh displayed with Riemann Mapper [40]. (M) The mesh is unfolded in UV space, also known as parameter space. (R) Parameterized mesh, textured mesh with checkers.....	12
Figure 2.6 : Planar Parameterization of Triangulated Surface Meshes. (L) Input mesh (M) Planar mesh (R) Textured mesh with checkers.....	13
Figure 2.7 : Conformal mapping of a mesh. Surface parameterization is the process to map a surface to a region of the plane.	14
Figure 2.8 : The .m mesh file format. (u, v) are the texture coordinates of the vertex, (n_x, n_y, n_z) are the normal vector on the vertex, (x, y, z) are the coordinates of a vertex; (id_1, id_2, id_3) are the vertex ids, arranged counter-clock-wisely, for every face.	15
Figure 2.9 : Alignment to the ground in parameter space. (L) Conformal map wireframed and displayed in Meshlab. The boundary edges are shown in green. (R) Aligned planar mesh.	16
Figure 2.10 : Barycentric coordinates. (a) The fulcrum which shows the proportion of weights placed to the rod, is positioned at the barycentric coordinates. (b) The barycentric coordinates of the point P, are represented by the ratio of weights.	17
Figure 2.11 : Gouraud Shading: Color interpolation inside a triangle.....	18
Figure 2.12 : Barycentric coordinates can be seen as the area of sub-triangles CAP (for u), ABP (for v) and BCP (for w) over the area of the triangle ABC which is why they are also called areal coordinates.....	19
Figure 2.13 : Interpolating inside a planar mesh. (L) Flattened mesh, output of the previous steps. (R) The interpolated data: Geometry pixels.....	20
Figure 2.14 : Different geometry images of the same face model. (L) No alignment [v1]. (M) Alignment in Cartesian coordinate system [v2]. (R) Alignment in parameter space [v3].....	20

Figure 3.1 : Illustrating general biometrics authentication pipeline, which consists of five stages..... 22

Figure 3.2 : Pictorial representation of a perceptron..... 23

Figure 3.3 : Diagrammatic representation of a CNN having input as 24x24..... 25

Figure 4.1 : BU-3DFE dataset: textured 3D facial scan of a participant performing happy face expression at the each stage of different intensities (low to highest) also neutral expression at most left. 28

Figure 4.2 : The BU-3DFE dataset: (a) the 83 facial landmark points presented using a 3D textured facial model having neutral state; (b) the amount of non-automatically defined landmark points considering various areas of the facial scans. 29

Figure 4.3 : Illustration of the VGGNet model. 30

Figure 4.4 : Fine-tuning a VGG16 network pretrained on ImageNet. 31

Figure 4.5 : A customized VGG16 model fine-tuned with our database. 32

Figure 4.6 : One-vs-all classification. 33



DEEP LEARNING BASED THREE DIMENSIONAL FACE EXPRESSION RECOGNITION USING GEOMETRY IMAGES FROM THREE DIMENSIONAL FACE MODELS

SUMMARY

The thesis studies the recognition of 3D facial expressions when a subject is attempting to show her/his emotions. These expressions make the detection and recognition problem difficult for inexperienced people. A motivation of 3D face expression recognition in computer vision is an automatic recognition of human behaviour. A large variety of disciplines may benefit from revealing the phenomenon, e.g. human robot interaction, security services, psychologists, teachers, etc. We propose a pipeline to recognize 3D facial expressions based on modeling surface geometry. 3D geometry of a face model is stored as color image also known as geometry images. The proposed method uses neural networks to perform expression recognition.

The initial research on expression recognition field has existed since the seventies. According to pioneers of the field, during performing an emotion, several facial expressions might occur on human faces. These facial appearances might be classified into six category, that is to say, sadness, happiness, disgust, anger, surprise, fear. The neutral face appearance is considered as the seventh category. This classification of face expressions has been also demonstrated to be compatible over several nationalities and societies. Therefore, these emotions are in certain perception, in all cases, recognizable.

The mesh processing pipeline focuses on 3D geometry of the human face and uses a geometry image representation for expression recognition. We devise this pipeline based on two fundamental complementary components: (1) facial mesh models pose correction (2) surface parameterization and geometry image creation. The first component uses bounding box normalization method to scale and align the mesh models and this results in placing all the meshes in one bounding box. The second component uses a Riemann Mapper for mesh parameterization and geometry image pixel coordinates are obtained using areal barycentric coordinates. Moreover, our approach on face expression recognition shows that normalization is a fundamental step especially in geometric approaches. Actually, an error in normalization might not be fixed in the next steps of our method such as after geometry image creation applying normalization might not be possible or requires extremely hard work. We apply a normalization technique which provides accurate and robust alignment even with the facial expressions being present. The traditional idea is that we apply pose correction to each mesh using the same face model as a reference and as primary step in mesh processing pipeline. The normalization evaluates a rigid transformation which involves rotation and translation. Our normalization method finds maximum and minimum vertex coordinates for X, Y and Z separately then it subtracts the minimum value from each coordinate and divides them by difference of maximum and minimum values. The algorithm deals with the normalization task through scaling size of the

bounding boxes, hence scaling the size of each mesh model, into an interval $[0, 1]$ which physically means that scaling into one bounding box.

Traditionally, research in feature extraction focused largely on handcrafted features such as Gabor and Haralick features. Many such hand-crafted features encode the pixel variations in the images to generate robust feature vectors for performing classification. Building on these, more complex hand-crafted features are also proposed that encode rotation and scale variations in the feature vectors as well. With the availability of training data, researchers have started focusing on learning-based techniques, resulting in several representation learning-based algorithms. Moreover, because the premise is to train the machines for tasks performed with utmost ease by humans, it seemed fitting to understand and imitate the functioning of the human brain. This led researchers to reproduce similar structures to automate complex tasks, which gave rise to the domain of deep learning. Research in deep learning began with the single unit of a perceptron, which was able to mimic the behavior of a single brain neuron. The perceptron generates an output based on the input as follows: where w_i corresponds to the weight for the i_{th} element of the input. The behavior of the perceptron is said to be analogous to that of a neuron, since, depending on a fixed threshold, the output would become 1 or 0. Thus, behaving like a neuron receiving electrical signal (input), and using the synapse (weight) to fire its output. Treating the perceptron as a building block, several complex architectures have further been suggested. Recently, the domain of deep learning has seen steep development. It is being used to address a multitude of problems with applications in biometrics, object recognition, speech, and natural language processing. Deep learning architectures can broadly be categorized into three paradigms: restricted Boltzmann machines, autoencoders, and convolutional neural networks. Auto-encoders and Restricted Boltzmann machines are traditionally unsupervised models used for learning meaningful representations of the given data. Convolutional neural networks, on the other hand, are traditionally supervised models with the objective of improving the overall classification performance.

The BU-3DFE database is used for the study. An entire of 100 people present in the data set, separated among male (44 people) as well as female (56 people). The performers are effectively divided up among various national categories and ethnic descents, containing Middle-East Asian, Latino-Americans, East-Asian, Black, White and the others. Every subject shows the six simple face expression, specifically, surprise (SU), sadness (SA), anger (AN), happiness (HA), disgust (DI), fear (FE) and also the neutral state (NE). Every face emotion has four stages of intensities — ground, medium, strong and top — excluding neutral faces which only have one level of intensity. In this study, our objective is to present the effectiveness of employing geometric information in 3D recognition and neural networks are only used as an option to handcrafted feature extraction. At first, transfer learning is applied and fine-tuning is not preferred not to let neural networks surpass the impact of geometry based approach. In our work, neural networks perform not only the feature extraction step, but also the feature classification step consecutively unlike the traditional methods. In recognition step, one binary classifier for each expression performs a one-versus-rest or one-versus-one classification. The mean accuracy is computed using the average of the binary classifier accuracy per expression. We present several average recognition rates for different experimental setups.

The importance of experimenting in 3D domain is that 3D modeling improves the 2D drawbacks such as illumination, pose variations, etc. Therefore, having a good recognition method might be useful for human computer interaction, criminal investigation, airport security or psychological examination. We propose a pipeline to recognize 3D facial expressions based on modeling surface geometry. The framework was designed to spot 3D expressions from geometry images obtained using 3D face models. The proposed method was evaluated on just GIM databases. VGG deep neural network model and different binary classifiers were used to obtain classification results. We observed that the expressions tend to give similar scores to other recognition frameworks using 3D models from the BU-3DFE database. Our experimental results show that VGG16 model has a best classification accuracy of 72.8% on geometry image database using one-versus-rest classification method. In conclusion, the results reveal that a VGG16 network is capable of handling complicated information from geometry pixels of 3D expression models. Our results also produce precious intuition into the application of neural networks on 3D domain along with the facial expression recognition task.





ÜÇ BOYUTLU YÜZ MODELLERİNDEN ELDE EDİLEN GEOMETRİ GÖRÜNTÜLERİ KULLANILAN DERİN ÖĞRENME TABANLI ÜÇ BOYUTLU YÜZ İFADELERİNİ TANIMA

ÖZET

Tez, bir kişi duygularını göstermeye çalışırken 3B yüz ifadelerinin tanınmasını incelemektedir. Bu ifadeler deneyimsiz insanlar için tespit ve tanıma problemini zorlaştırır. Bilgisayar görüşünde 3B yüz ifadesi tanıma motivasyonu, insan davranışının otomatik olarak tanınmasıdır. Çok çeşitli disiplinler fenomenin ortaya çıkmasından, örneğin; insan robot etkileşimi, güvenlik hizmetleri, psikologlar, öğretmenler vb. Yüzey geometrisini modellemeye dayanan 3B yüz ifadelerini tanımak için bir boru hattı öneriyoruz. Bir yüz modelinin 3D geometrisi, geometri görüntüleri olarak da bilinen renkli görüntü olarak saklanır. Önerilen yöntem, ifade tanıma gerçekleştirmek için sinir ağlarını kullanır.

İfade tanıma alanındaki ilk araştırma yetmişli yıllara dayanmaktadır. Alanın öncülerine göre, bir duygunun oluştuğu sırada, insan yüzünde birkaç yüz ifadesi görülebilmektedir. Bu yüz görünümleri, üzüntü, mutluluk, iğrenme, öfke, sürpriz, korku gibi altı kategoriye ayrılmaktadır. İfadesiz yüz görünümü yedinci kategori olarak kabul edilmektedir. Yüz ifadelerinin bu sınıflandırmasının birkaç millet ve toplum üzerinde de uyumlu olduğu gösterilmiştir. Bu nedenle, bu duygular kesin olarak algılanabilir ve her durumda tanınabilir olarak kabul görmüştür.

Yöntem, insan yüzünün 3B geometrisine odaklanmakta ve ifade tanıma için bir geometri görüntüsü gösterimi kullanmaktadır. Bu boru hattını iki temel tamamlayıcı bileşene dayanarak tasarlıyoruz: (1) yüz modellerinin hizalanması, duruşlarının düzeltilmesi (2) yüzey parametreleştirme yöntemi ve geometri görüntüsü oluşturmaktır. İlk bileşen, yüz modellerini ölçeklemek ve hizalamak için normalleştirme yöntemini kullanmaktadır ve bu, tüm yüz modellerini bir sınırlayıcı kutuya yerleştirerek sonuçlanmaktadır. İkinci bileşen, yüzey parametreleştirme yöntemi için bir Riemann dönüştürücüsü kullanmaktadır ve geometri görüntü piksel koordinatları, her bir üçgenin ağırlık merkezi koordinatları kullanılarak elde edilmektedir. Dahası, yüz ifadesi tanıma konusundaki yaklaşımımız normalleşmenin özellikle geometrik yaklaşımlarda temel bir adım olduğunu göstermektedir. Aslında, normalizasyonda adımında bir hata yapılması yöntemin sonraki adımlarında düzeltilemeyebilir, ve örneğin normalizasyonu geometri görüntüsü oluşturma işleminden sonra oluşan görüntüler üzerinde uygulamak mümkün olmayabilir veya çok zor bir çalışma olabilir. Doğru ve sağlam bir normalizasyon tekniği uygulandığını kullanılan yüz ifadeleri modelleri göstermektedir. Geleneksel yöntem, referans olarak aynı yüz modelini kullanarak yüz modeli işleme boru hattında birincil adım olarak her yüze poz düzeltilmesi uygulamaktır. Normalleştirme, dönme ve çevirmeyi içeren doğrusal bir dönüşümü oluşturmaktadır. Normalleştirme yöntemimiz, X, Y ve Z için ayrı ayrı maksimum ve minimum köşe koordinatlarını bulmaktadır, ardından her bir koordinattan minimum değeri çıkarmakta ve bunları maksimum

ve minimum değerlerin farkına bölmektedir. Algoritma, sınırlayıcı kutularını ölçeklendirme vasıtasıyla normalleştirme adımını gerçekleştirmektedir, bu sayede tüm yüz modellerinin boyutunu, fiziksel olarak tek bir sınırlama kutusuna ölçeklendirme anlamına gelen bir aralık olan $[0, 1]$ 'e ölçeklendirmektedir.

Geleneksel olarak, özellik çıkarımı araştırmaları büyük ölçüde Gabor ve Haralick özellikleri gibi el yapımı özelliklere odaklanmıştır. Bu el yapımı özelliklerin birçoğu, sınıflandırma yapmak için sağlam özellik vektörleri üretmek üzere görüntülerdeki piksel varyasyonlarını kodlar. Bunlara dayanarak, özellik vektörlerindeki rotasyon ve ölçek değişikliklerini kodlayan daha karmaşık el yapımı özellikler de önerilmiştir. Eğitim verilerinin kullanılabilirliği ile araştırmacılar öğrenmeye dayalı tekniklere odaklanmaya başlamış ve birkaç temsili öğrenmeye dayalı algoritmalar ortaya çıkmıştır. Dahası, makinelerin insanlar tarafından en kolay şekilde gerçekleştirilen işler için eğitilmesi söz konusu olduğunda, insan beyninin işleyişini anlamak ve taklit etmek uygun görülmüştür. Bu, araştırmacıların, derin öğrenme alanında, karmaşık görevleri otomatikleştirmek için benzer yapıları yeniden üretmeye çalışmalarına neden olmuştur. Derin öğrenme araştırmaları, tek bir beyin nöronunun davranışını taklit edebilen bir algılayıcının tek birimi ile başlamıştır. Algılayıcı birim, girdiyi temel alan bir çıktı üretmektedir: burada w_i , girişin i_{th} elemanının ağırlığına karşılık gelmektedir. Algılayıcının davranışının bir nöronunkine benzer olduğu söylenebilir, çünkü sabit bir eşikğe bağlı olarak, çıkış 1 veya 0 olmaktadır. Algılayıcının bir yapı taşı olarak ele alınmasıyla, çeşitli karmaşık mimariler ileri sürülmüştür. Son zamanlarda, derin öğrenme alanında ahızlı bir gelişme görülmüştür. Derin öğrenme yöntemleri, biyometrik, nesne tanıma, konuşma ve doğal dil işleme alanındaki uygulamalarla ilgili birçok sorunu çözmek için kullanılmaktadırlar. Derin öğrenme mimarileri genel olarak üç paradigmaya ayrılabilir: kısıtlı Boltzmann makineleri, otomatik kodlayıcılar ve evrişimli sinir ağları. Otomatik kodlayıcılar ve Sınırlı Boltzmann makineleri, verilen verilerin anlamlı temsillerini öğrenmek için kullanılan geleneksel olarak denetimsiz modellerdir. Konvolüsyonel sinir ağları, genel sınıflandırma performansını iyileştirmek amacıyla geleneksel olarak denetlenen modellerdir.

BU-3DFE veri tabanı araştırma için kullanılmıştır. Veri seti 100 kişiden oluşmaktadır, sırayla erkek ve kadın kişi sayıları 44 ve 56 olmak üzere. Veri seti, Orta Doğu, Asya, Latin Amerika, Doğu Asya, siyahi, beyaz tenli ve birçoklarını içeren çeşitli ulusal kategoriler ve etnik kökenler arasında etkili bir şekilde bölünmüştür. Her kişi, altı basit yüz ifadesini, sürpriz (SU), üzüntü (SA), öfke (AN), mutluluk (HA), öğrenme (DI), korku (FE) ifadelerini gerçekleştirmektedir ve ayrıca ifadesiz yüz durumu (NE) gerçekleştirilmektedir. Her yüz ifadesi dört şiddet aşamasına sahiptir - zemin, orta, güçlü ve üst. İfadesiz yüzler, yalnızca tek bir şiddet seviyesine sahiptir. Bu çalışmada amacımız, 3B tanımada geometrik bilgilerin kullanılmasının etkinliğini sunmaktır ve sinir ağları sadece el yapımı özellik çıkarımı için bir seçenek olarak kullanılmaktadır. İlk önce, aktarımsal öğrenme yöntemi uygulanmıştır ve ince ayar yapılması tercih edilmemiştir, çünkü sinir ağlarının geometriye dayalı yaklaşımın önüne geçmesine izin verilmek istenmemiştir. Çalışmamızda sinir ağları sadece özellik çıkarma adımını değil, aynı zamanda geleneksel yöntemlerden farklı olarak özellik sınıflandırma adımını da yerine getirmektedir. Tanıma adımında, her bir ifade için ikili bir sınıflandırıcı bire-karşı-bir ya da bire-karşı-hepsi sınıflandırma yöntemini gerçekleştirmektedir. Ortalama doğruluk, ifade başına ikili sınıflandırıcı doğruluğunun ortalaması kullanılarak hesaplanmıştır. Tez, farklı deney kurulumları için birden fazla ortalama tanıma oranı sunmaktadır.

3B alanda yüz tanımanın önemi, 3B modellemenin aydınlatma, poz farklılıkları ortaya koyma vb. gibi 2B alanındaki dezavantajları iyileştirmesidir. Bu nedenle, iyi bir tanıma yöntemine sahip olmak, insan-bilgisayar etkileşimi, ceza soruşturması, havaalanı güvenliği veya psikolojik incelemelerde için yararlı olabilmesi açısından önem arz etmektedir. Yüzey geometrisini modellemeye dayanan 3B yüz ifadelerini tanımak için bir boru hattı önerilmektedir. Tanıma sistemi, 3B yüz modelleri kullanılarak elde edilen geometri görüntülerinden 3B yüz ifadelerini tespit etmek için tasarlanmıştır. Önerilen yöntem sadece GIM veritabanlarında değerlendirilmiştir. Sınıflandırma sonuçlarını elde etmek için VGG derin sinir ağı modeli ve farklı ikili sınıflandırıcılar kullanılmıştır. BU-3DFE veri tabanındaki 3B modeller kullanılarak ifadelerin sınıflandırılması, önerilen sistemin literatürdeki diğer tanıma sistemlerine benzer doğruluk değerleri verme eğiliminde olduğu göstermiştir. Deneysel sonuçlar, VGG16 modelinin bire-karşı-hepsi sınıflandırma yöntemini kullanarak geometri görüntüsü veritabanında en iyi %72.8'lik bir sınıflandırma doğruluğuna sahip olduğunu göstermektedir. Sonuç olarak, elde edilen değerler bir VGG16 ağının 3B ifade modellerinin geometri piksellerinden karmaşık bilgileri işleme yeteneğine sahip olduğunu ortaya koymaktadır. Sonuçlarımız ayrıca, sinir ağlarının yüz ifadesi tanınmasının yanı sıra 3B alanında uygulanmasıyla da araştırma çalışmalarına değerli bir yaklaşım oluşturmaktadır.



1. INTRODUCTION

As mechanical machines begin to be incorporated in daily lives and participate not only in social lives but also business lives, smarter devices became a requirement to the perceive the feelings and sentiments of human beings. Having these intelligent machines which have the ability to analyze any person's feelings and state of mind is exactly what the human-machine interaction field is aiming in the perceptual computing along with human-computer interaction researchers. The subsequent list is a few compelling fields where automated facial expression analysis frameworks have practical implementations:

- **Human-computer interaction:** Face appearance is a manner of non-verbal interaction among a lot of different ways (e.g., hand gesture). Expression recognition is inherent for human beings, however it is an extremely tough job for mechanical machines; hence, the aim of an expression analysis framework is to examine feelings in a manner with which human-computer interaction might be improved and to create machinery and automaton considerably similar to human kind.
- **Health purposes and services area:** Face emotions are the straight method to recognize when particular psychological changes (e.g., hurt, suffering, melancholy) happen.
- **Psychology area:** Emotion recognition is extremely practical for the investigation of the human psychological behaviour.
- **Surveillance area:** Decrypting the micro-emotions is critical in many cases for instance, during questioning, any double-dealing behavior from suspected people. For the reason that micro-emotion is a short-lived reflexive face expression that human unknowingly show when preventing a feeling from being apparent.
- **Academic area:** Students' face emotions notify the instructor to modify the lecture.

The initial research on facial expression recognition [1] [2] has existed since the seventies. Many studies showed that during performing an emotion, several facial expressions occur. These facial appearances might be classified into six category plus neutral appearance, that is to say, sadness, happiness, disgust, anger, surprise, fear. This classification of face expressions has been also demonstrated to be compatible over several nationalities and societies. Therefore, these emotions are in certain perception, in all cases, recognizable.

At a recent time, there has been a increasing shift from 2-Dimensional to 3-Dimensional in face recognition, mostly inspired by the strength of the 3-Dimensional models to lighting variations, posture, and size changes. A small number studies of have used the benefit of the 3D surface geometry to conduct facial expression analysis. Previously, the initial results to accomplish facial expression analysis in an automatic way using 3D facial scans were suggested employing extremely small-scale data sets and classifying just a certain number of facial emotions. The accessibility of novel facial expression data sets, such as one established in Turkey at Boğaziçi University [3], Bosphorus database [4] [5], and another one the BU-3DFE database [6] from Binghamton University. These databases currently moved the study one step forward. Specifically, the BU-3DFE data set has begun to be the quality measurement for comparison of different facial expression recognition methods.

In the beginning of nineteenth century, Charles Darwin gave birth to the face expression research field. Darwin demonstrated the basis of emotions and the definition of them in 1872. Darwin conducted expression recognition experiments involving animals and people. His emotion analysis furthermore classified face expressions into different classes. His discoveries regarding face expression classes were shyness, anger, joy, disgust, sulkiness, surprise and anxiety.

In 1970s Paul Ekman and his teammates [7] put a significant milestone in face expression field. Their discoveries were concerning the categorization of face emotions in seven principal classes: neutral, surprise, sadness, happiness, anger, disgust and fear. Later on, Ekman and one of his colleagues Friesen [7] described FACS: Facial Action Coding System to describe face movements in which action units are first mentioned and defined. Facial Action Coding system is associated with muscle movements. The system contains the face muscles which undergoes a change when an emotion becomes

Table 1.1 : Expression classes considering closest emotions.

Basic Expression	Related Emotions
Anger	Rage, outrage, fury, wrath, hostility, ferocity, bitterness,
Disgust	Revulsion, contempt.
Fear	Alarm, shock, fright, horror, terror, panic, hysteria.
Happiness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness.
Sadness	Depression, despair, unhappiness, melancholy.
Surprise	Amazement, astonishment.

apparent in the face. The movements of these facial parts also known as Action Units. This research is crucial for the literature in a manner which numerous scientists pursued the evolution and enabled the growth of ongoing recognition frameworks.

Human facial emotion related expressions might vary across different ethnicities and societies. Ekman and his colleagues [7] discoveries considering six principal facial expressions have been declared that those are the clearest expression categories to be recognized between all the nations on earth. Apart from this, besides the six principal expressions, human being's face has the ability to express several different emotions. Table 1.1 sums up associated emotions assigned to six principal classic expressions.

In 1999, Moving Pictures Expert Group, MPEG, suggested a framework for facial recognition research and the framework became a milestone in the literature. Their work systematize face animation and facial features considering MPEG-4 format. The framework established a face shape model using 83 feature points which are called Facial Definition Parameters, FDPs, also defining a human face in the neutral pose. MPEG-4 format described 68 Facial Animation Parameters, FAPs, that are employed for the animation of the face via the muscle movements of the defined feature points. Facial animation parameters might be employed to animate the human faces as well as to form principal face expressions. Apart from this, FAPs can also be employed for representing facial expressions using a general face shape model. MPEG-4 standard FAPs are extensively employed via numerous researchers for academic purposes concerning facial expression analysis and recognition systems.

Majority of the research about 3 dimensional face emotion analysis might be classified using especially the ones built on the general face shape model and feature categorization. In the initial class, a generic 3D face shape model, deformable model, is prepared using preliminary information, for instance local geometry labels, shape

and texture variations, or feature points. A strong correlation within points of 3D face models is often necessary to construct the deformable modelling. For instance, it is accomplished that a correlation within expressive faces and the associated neutral match via optimizing energy based functions. MEM is a morphable expression model which is built via employing the Principle Component Analysis to various emotions, therefore novel emotions might be represented as points with a reduced dimensionality space built by via eigen-faces acquired using a morphable expression model.

In an earlier review, it appears that the big amount of current studies on 3D face emotion analysis depends on the existence of landmarks precisely described on the facial model. Approaches on the basis of general face model utilize landmarks to accomplish correlations among face models in formation of a deformable facial model. Generally, the approaches are challenging in a computational way because of the deformation procedure. Results on the basis of classification of extracted features in most occasions calculate distances among landmark points and assess how come the measurements vary between expressive scans and neutral face appearance. These landmark points are not recognizable in an automatic way. Moreover, various results require a neutral face model considering every person to be able to assess the dissimilarities produced from a 3D face scan via face emotions regarding neutral referenced scans. Finally, the approaches limits the ability to be applicable in real life scenarios.

The thesis is structured as follows. Chapter 2 gives the detailed information about proposed mesh processing pipeline. Chapter 3 describes the convolutional neural networks and briefly introduces architecture of a traditional convolutional neural network. Chapter 4 presents the experimental results for 3D facial expression recognition. Chapter 5 concludes the thesis.

2. MESH PROCESSING PIPELINE

2.1 Geometry Image Based Approach

The standard geometric representation used in today's graphics hardware is the irregular mesh. A mesh consists of an array of triangles, and an array of vertices. Each triangle refers to three vertex indices. To represent more detail over meshes, it is common to use texture mapping. The mesh vertices are assigned texture coordinates, defining a parametrization of the mesh onto an UV domain. An image is placed on this domain, which is then mapped back to the surface. In this case, the texture image represents surface normals, which can be used in per-pixel shading. Previous studies in remeshing, has gone part of the way in creating more regular geometric representations. But, such methods still use an irregular base mesh, which leads to a sampling that is only semi-regular. The approach [8] is to employ an arbitrary surface using a completely regular grid of samples on a square domain which is called a geometry image. In Figure 2.1 (b), as shown, it looks just like an ordinary image. It has 257 by 257 pixels. The only difference is that it has 12 bits per channel instead of the usual 8. It is called a geometry image because its RGB colors encode XYZ positions, and in fact entirely describe the 3D face model shown in Figure 2.1 (c).

The outline of the fundamental steps involved in creating a geometry image is as follows: First, we open up the mesh through a proper cut path sets, thus we create a surface having the topology of a disk. Second, parametrization is performed to this disk surface to map it into the square domain of the geometry image. Next, we overspread a regular 2D sampling grid over the domain and map these samples through the parametrization back onto the surface. We should note that it is important to create a good cut and a good parametrization, so that the samples are evenly distributed over the surface. Then, we compute the XYZ positions of the surface samples, as well as other surface attributes and store these in the geometry image. Therefore we visualize

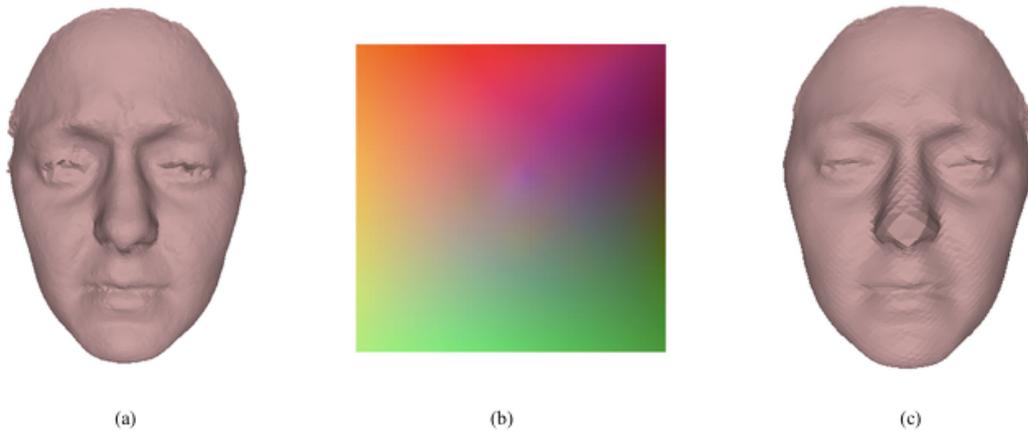


Figure 2.1 : Rendering of a geometry image. (a) The initial face scan. (b) Geometry image is visualized as $[r, g, b]$, 12 bits/channel. (c) Reconstruction of geometry, obtained completely from (b).

the XYZ data as RGB colors. Given a geometry image, it is straightforward to render it as shaded 3D geometry.

Consequently, geometry images [8] [9] [10] are a powerful, completely regular representation of triangular meshes, describing each vertex and its attributes as a pixel in one or more rectangular images and thus defining an implicit connectivity through neighbouring pixels. It can be understood as a next step after color and normal mapping, representing the whole model as a set of images.

As expected, random surface models might not usually be reverted immediately into a square-like grid, as the topological domain might vary from model to model. Illustrating surface models using geometry images might introduce a few issues [8]:

- A proper cut must be found, to open the 3D mesh into a topology of a disk. The cut allows a fine parameterization of the 3D surface inside the disk.
- Undersampling might cause to geometrical blur.
- Simple compression of geometry images might represent holes or splits through the cut paths of the surface.

Geometry based approach has the subsequent restrictions [8]:

- Geometry images can only characterize manifold geometry.
- Opening up a complete 3D mesh in a single chart might cause significant deformation in parameterization.

2.2 Overview of the Pipeline

This section gives the detailed information about proposed mesh processing pipeline. It focuses on 3D geometry of the human face and uses a geometry image representation for expression recognition. We devise this pipeline based on two fundamental complementary components: (1) facial mesh models pose correction (2) surface parameterization and geometry image creation. The first component uses bounding box normalization method to scale and align the mesh models and this results in placing all the meshes in one bounding box. The second component uses a Riemann Mapper for mesh parameterization and geometry image pixel coordinates are obtained using areal barycentric coordinates. Moreover, our approach on face expression recognition shows that normalization is a fundamental step especially in geometric approaches. Actually, an error in normalization might not be fixed in the next steps of our method such as after geometry image creation applying normalization might not be possible or requires extremely hard work. We apply a normalization technique which provides accurate and robust alignment even with the facial expressions being present. The traditional idea is that we apply pose correction to each mesh using the same face model as a reference and as primary step in mesh processing pipeline. The normalization evaluates a rigid transformation which involves rotation and translation. Our normalization method finds maximum and minimum vertex coordinates for X, Y and Z separately then it subtracts the minimum value from each coordinate and divides them by difference of maximum and minimum values. The algorithm deals with the normalization task through scaling size of the bounding boxes, hence scaling the size of each mesh model, into an interval $[0, 1]$ which physically means that scaling into one bounding box. After applying bounding box normalization, the rest of the steps of the geometric approach are listed below.

- Using the Riemann Mapper [11], we obtain (u, v) parameters for each vertex.
- We establish to find the closest points on the circle using the (u, v) coordinates, which are the feature points near the edge of the face disk.

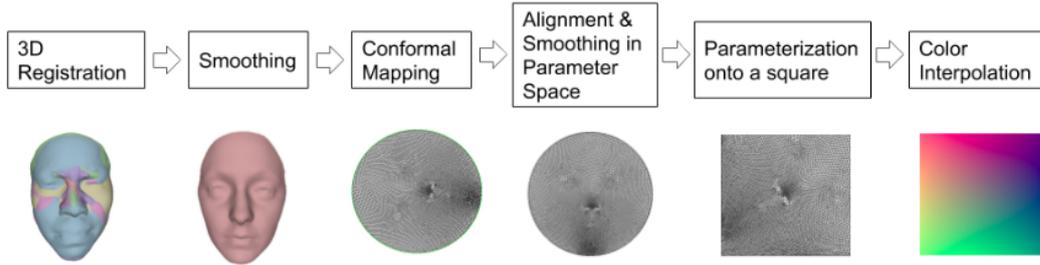


Figure 2.2 : Proposed mesh processing pipeline.

- The angle of rotation of one face disk relative to the other is calculated using the closest points on the circle and the coordinates of all the nodes (u, v) of the face disk is rotated with this angle.
- Smoothing is applied with the coordinates of the face nodes (u, v) . At first, we find the average of invariable features -or nodes- and the average features are used to deform other nodes for all faces.
- The geometry image pixel coordinates is calculated using Barycentric interpolation, as the weighted average of the triangular nodes in which they are located.

The proposed framework, in Figure 2.1, for facial expression recognition focuses 3D geometry of the human face and uses a geometry image representation for expression recognition. We can divide it into two fundamental parts: (1) facial mesh models correction and geometry image creation and (2) facial expression representation and classification. The first part uses a Riemann Mapper for parameterization and geometry image pixel coordinates are obtained using barycentric coordinates. The second part is responsible for obtaining a deep representation and classification of the facial expression. We employ a softmax loss function that classifies one expression at a time, with binary classifiers. The second part will be examined in Chapter 4 in detail.

2.3 3D Registration

The models we have examined are not aligned with the center considering bounding boxes. This can make learning quite difficult. It is also best to align the fixed feature points on the face when aligning the bounding boxes. We match the bounding boxes of the original meshes. To do this, we subtract the minimum value for each coordinate

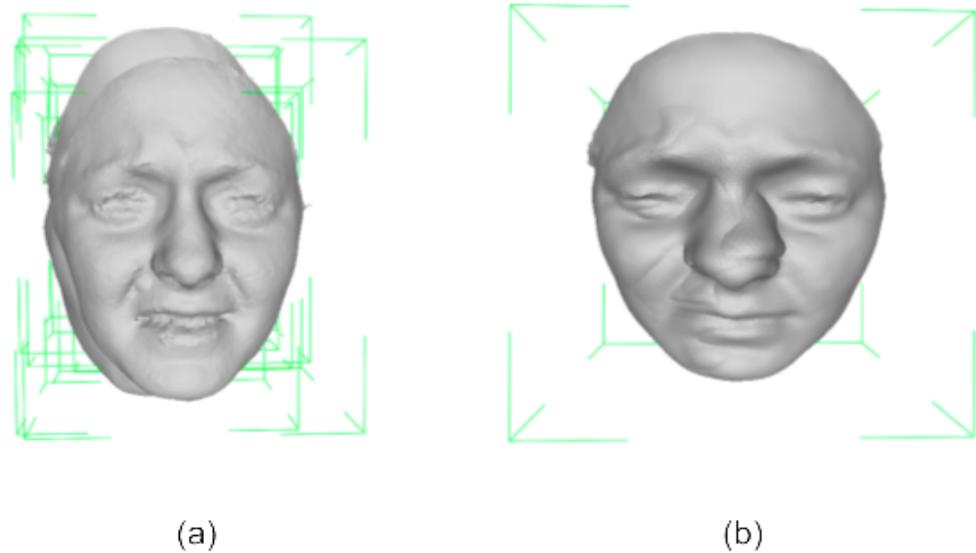


Figure 2.3 : Matching bounding boxes. (a) Before matching. (b) After matching.

and divide it into the difference of maximum and minimum values. It's a kind of normalization, and it's geometrically very meaningful and the resulting scaling is shown in Figure 2.2.

Our approach for face expression recognition shows that 3D registration -alignment- is a fundamental step especially in geometric approaches. Actually, an error in alignment might not be fixed in the further steps. The traditional idea is that we apply pose correction to each mesh using the same face model as a reference and as primary step in mesh processing pipeline. The alignment evaluates a rigid transformation which involves rotation and translation. Our alignment method employs the Iterative Closest Point algorithm. The algorithm deals with the alignment task through reducing the distance between two different oriented triangle meshes. The resulting transformation is shown in Figure 2.3.

The 3D data registration step, widely known as data alignment, is implemented using MeshLab and it supplies a well-established tool to transform the independent meshes onto a general reference system, it also could handle large-scale range-maps. 3D registration is mostly implemented as follows: a tweaked ICP one-to-one alignment step, and a worldwide bundle adjustment to reduce error-distribution is also applied. The alignment could be performed on point clouds and meshes originated from different

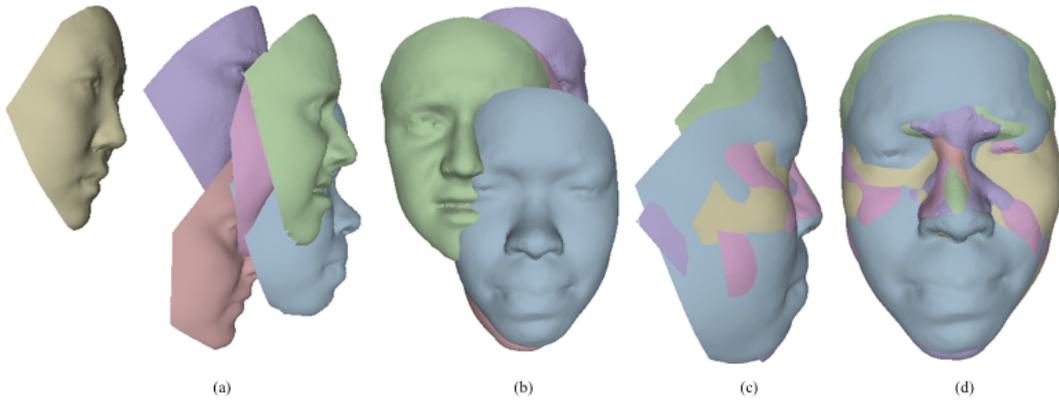


Figure 2.4 : Depiction of 3D registration. (a) The original face scans with different poses, side view. (b) The original face scans with different poses, frontal view. (c) The resulting transformation after alignment, side view. (d) The resulting transformation after alignment, frontal view.

sources, consisting of 3D-from-image tools and active scanners (both long-range and short-range).

Rigid Registration. Rigid registration produces a rigid mapping which transforms one cloud of points to the other, given two point sets. A rigid transformation or mapping is described as a transformation which does not alter the distance between any two points. Regularly, a transformation such that includes rotation and translation. Exceptionally, the point set might be reflected.

Non-Rigid Registration. The registration produces a non-rigid mapping which transforms one cloud of points to the other, given two point sets. Non-rigid transformations consists of affine transformations such like shear mapping and scaling. Nevertheless, in the concept of point set data alignment, non-rigid registration usually includes non-linear transformations. The nonlinear transformation may be parametrized by the eigenvalues, if the eigen modes of variation of the point set are known. A non-linear transformation might additionally be parametrized as a thin plate spline or TPS in short.

Iterative Closest Point. The ICP (Iterative Closest Point) algorithm is widely-known and used for minimizing the dissimilarity between two point sets. ICP is generally employed to construct 3D or 2D surfaces from various scanned data, for achieving optimal path planning (mostly while wheel odometry is not reliable because of slippery terrain) and co-registering bone models, localization of robots, etc.

The ICP algorithm steps are as follows:

- Match the closest point in the referenced cloud of points (or a picked group of vertices), for every point in the source cloud of points (from the entire group of vertices usually introduced as a selection of pairs of vertices).
- Determine the combination of translation and rotation employing a root mean square point-to-point metric method for distance minimization that will ideally apply alignment to every source point set to its counterpart which was found in the earlier stage. The process might additionally include rejecting outliers and weighting points before alignment.
- Map the source cloud of points employing the obtained mapping.
- Repeat (re-connect the points, and so on so forth).

2.4 Surface Parameterization Methods

The procedure of mapping a mesh on to areas of the planar unit is described as parameterization. The mapping permits operational tasks on a surface to be carried out such like when it is flattened. For many applications, parameterization is necessary such as construction of geometry images, remeshing, texture synthesis, texture mapping.

There exist various techniques for parameterizing a surface, such as Orbifold Tutte Embeddings, Floater Mean Value Coordinates, Least Squares Conformal Maps, Discrete Authalic Parameterization, Discrete Conformal Maps, Tutte Barycentric Mapping or As Rigid As Possible Parameterization. All of these techniques principally differentiate by the deformation they try to minimize (areas vs. angles), and one-to-one correspondence assures the transformation on to the planar disk (free border vs. convex polygon).

Based on the required kind of border parameterization, various parameterization techniques could be categorized in three groups: borderless, free border and fixed border. Fixed border surface parameterizations are identified as having a restricted border in parameter space.

Even though principal purpose at the back of initial parameterization techniques was the utilization to texture mapping, currently parameterization is often employed to map

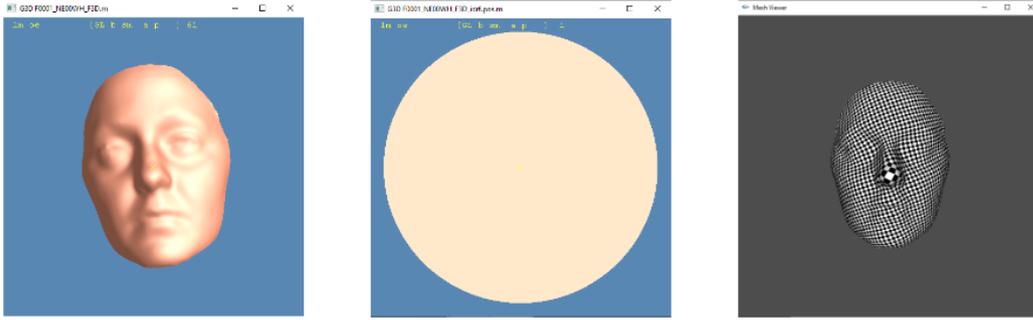


Figure 2.5 : Fixed Boundary Conformal Mapping on a circle. (L) Original mesh displayed with Riemann Mapper [40]. (M) The mesh is unfolded in UV space, also known as parameter space. (R) Parameterized mesh, textured mesh with checkers.

other applications such as remeshing, approximating surfaces, repairing CAD models, re-parameterizing spline surfaces, and fitting scattered data.

Surface parameterization is equivalent to find a injective mapping from an acceptable domain to the mesh. A high quality transformation perform minimization for either area distortions (equiareal mapping) or angle distortions (conformal mapping) in a certain way. In the thesis, we pay attention to parameterize homeomorphic triangular meshes which can be transformed to a sphere or a disk, and on piece-wise linear transformations on to a unit flattened disk.

2.4.1 Conformal mapping

Mesh parameterization describes the procedure for mapping a three-dimensional triangular mesh over a two-dimensional (flattened) domain, the majority of mesh parameterization algorithms use the concept differential geometry as a basis. A conformal parameterization represents a three-dimensional surface on to a planar two-dimensional domain, in such a way that the parameterization is angle-preserving, or identically, conformal mapping [11] [12] directly maps extremely small circles on the 3D surface to the extremely small circles on the 2D plane. It can be seen in the subsequent Figure 2.4, the three-dimensional human face model is mapped on to the unit disk which is flattened so that planar in 2D domain using a conformal parameterization algorithm. In Figure 2.4, we place the texture, checker board, on the flattened mesh, and draw it back on to the three-dimensional mesh surface, so that we preserve each of the right corner angle of the checker board texture. In a similar fashion, we place another texture, a circle packing, on the planar mesh, afterwards

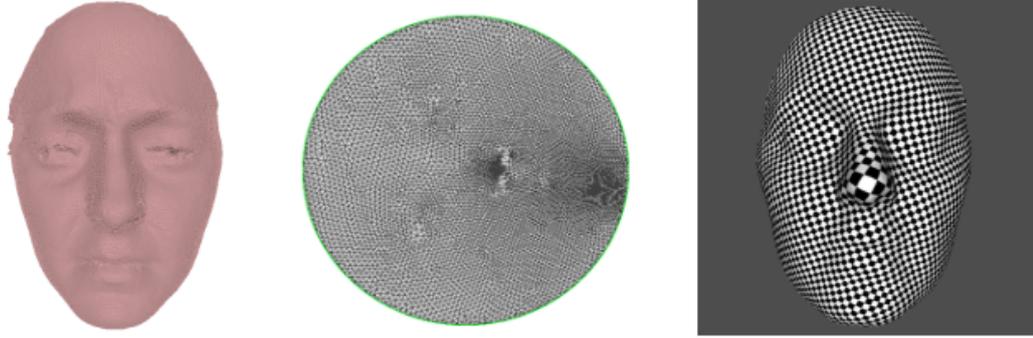


Figure 2.6 : Planar Parameterization of Triangulated Surface Meshes. (L) Input mesh (M) Planar mesh (R) Textured mesh with checkers.

draw it back on to the three dimensional mesh surface, eventually we preserve all the small circles. As specified by uniformization theory, every surface in actual life could be parameterized conformally to one of three recognized shapes, the hyperbolic space, the plane, the sphere.

The conformal parameterization evaluates an injective correspondence from a three dimensional triangle mesh to a straightforward two dimensional planar disk, see Figure 2.5. On the triangular surface, the transformation is piece-wise linear. A set of (u,v) parameter coordinates associated with every vertex of the input surface mesh is the outcome of the transformation. Based on the selection of the mapping algorithm, an injective correspondence might be assured or not.

The conformal mapping aims to reduce the angle distortion and it attempts to minimize a discrete version of the Dirichlet energy. A one-to-one correspondence is assured when the two subsequent constraints are satisfied: the border is convex and the barycentric mapping condition which is the condition each vertex being a convex combination of its nearest vertices in parameter space.

A conformal transformation is inherent to the 3D geometry of a surface, is not dependent of the resolution of the 3D surface, which maintains the stability of the positioning.

Conformal parametrization, as shown in Figure 2.6, has nice properties which makes it suitable for many practical applications. Conformal parameterization has been proposed for visualization, geometry remeshing and texture mapping due to its angle preserving property. Conformal parameterization might be used to match two

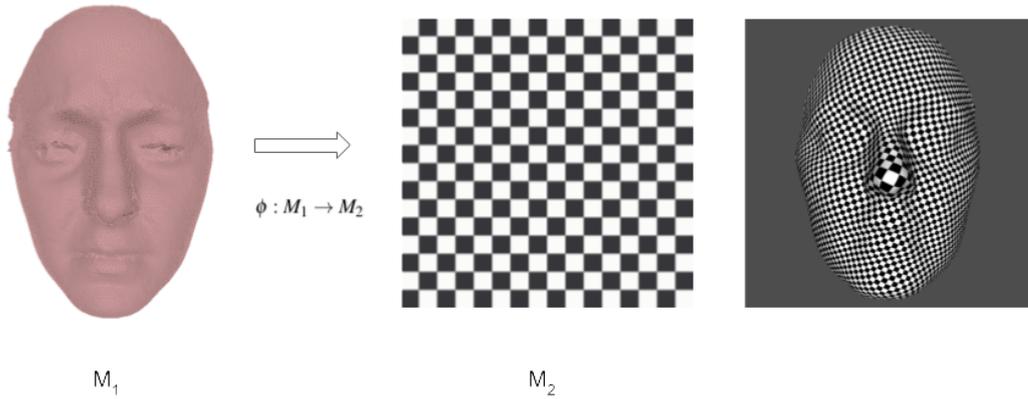


Figure 2.7 : Conformal mapping of a mesh. Surface parametiation is the process to map a surface to a region of the plane.

comparable surfaces as it is based on the metrics of the surface. Moreover, by conformal invariants, all surfaces can be simply classified.

Various methods have been proposed for evaluating conformal mappings, however nearly every methods just handle surfaces with genus zero and division of the surfaces into segments becomes a must. Many techniques break meshes apart onto topological units, afterwards map every part separately. Decomposition distrupts continuity along the boundaries of patches so that conformality might not be maintained in every spot. Global conformal parametrization which maintains conformality in almost all of the points (excluding a few points) is extremely beneficial to stay away from the problems related with discontinuity of the boundaries.

Conformal parameterization for closed arbitrary genus surfaces is addressed in many studies. Riemann Mapper [11] uses global conformal mapping for nonzero genus surfaces with boundaries and we can utilize its application on constructing geometry images.

Riemann Mapper. Riemann Mapper [11] is a widespread purpose tool for the aim of mesh parameterization. The toolkit is implemented in general C++ using OpenGL, as well as glut interface. It is avalaible on both Linux and Windows platforms. It could also manage surfaces including many kinds of topology. It is a strong and efficient tool.



Vertex 1	x_1	y_1	z_1	{normal=(n_{x1} n_{y1} n_{z1}) uv=(u v)}	Face 1	id_1 id_2 id_3
Vertex 2	x_2	y_2	z_2	{normal=(n_{x2} n_{y2} n_{z2}) uv=(u v)}	Face 2	id_1 id_2 id_3
Vertex 3	x_3	y_3	z_3	{normal=(n_{x3} n_{y3} n_{z3}) uv=(u v)}	Face 3	id_1 id_2 id_3
Vertex 4	x_4	y_4	z_4	{normal=(n_{x4} n_{y4} n_{z4}) uv=(u v)}	Face 4	id_1 id_2 id_3
...			
Vertex n	x_n	y_n	z_n	{normal=(n_{xn} n_{yn} n_{zn}) uv=(u v)}	Face m	id_1 id_2 id_3

Figure 2.8 : The .m mesh file format. (u, v) are the texture coordinates of the vertex, (n_x, n_y, n_z) are the normal vector on the vertex, (x, y, z) are the coordinates of a vertex; (id_1, id_2, id_3) are the vertex ids, arranged counter-clock-wisely, for every face.

Input File Format. The tool maintains .m file and .obj file formatings. The tool package consists of numerous sample 3D mesh models, that are formatted in ASCII. They could be opened employing a word processor and analyse.

Mesh parameterization or mesh flattening is the process of unfolding -unwrapping- a mesh and Riemann Mapper [40] is used to perform surface parameterization on a circle. In Figure 2.7, mesh file format used in the mapper tool is presented in detail.

2.4.2 Computational conformal geometry

Let S_1 and S_2 be two surfaces with Riemannian metrics g_1 and g_2 , $\phi : (S_1, g_1) \rightarrow (S_2, g_2)$ be a diffeomorphism between them. We say ϕ is conformal if it preserves angles. Infinitesimally, a conformal mapping is a scaling transformation; it preserves local shapes. For example, it maps infinitesimal checkerboard to infinitesimal checkerboard. As shown in Figure 2.6, the face surface is mapped to the plane via a conformal mapping ϕ . If a checkerboard packing is given on the plane and pulled back by ϕ , it produces a checkerboard packing on the face surface.

Conformal Structure is a natural geometric structure on surfaces, which governs many physics phenomena, such as heat diffusion, electric-magnetic fields, etc. Conformal field theory plays fundamental role in string theory. In mathematics, conformal means angle preserving. Conformal structure is a special atlas of the surface, such that angles among tangent vectors can be coherently defined on different local coordinate systems. Furthermore, concepts in complex analysis can be defined on the surface via conformal

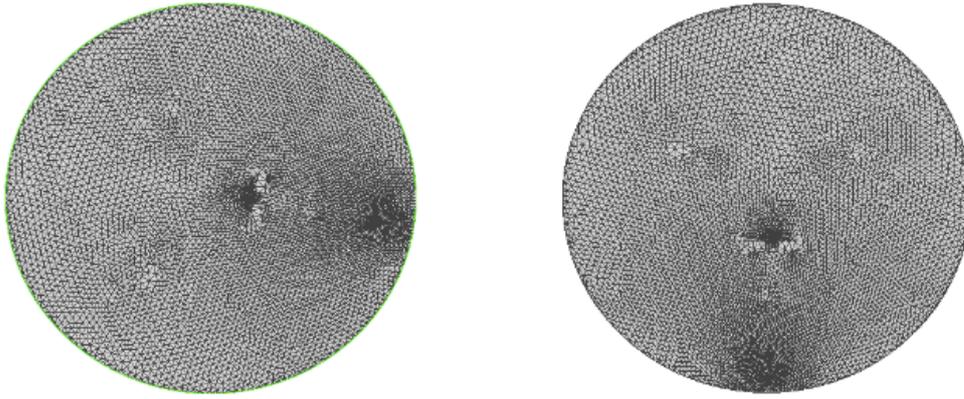


Figure 2.9 : Alignment to the ground in parameter space. (L) Conformal map wireframed and displayed in Meshlab. The boundary edges are shown in green. (R) Aligned planar mesh.

structure. Conformal geometry is the intersection of algebraic geometry, differential geometry, complex analysis and algebraic topology.

In engineering, conformal structure is between topological structure and geometric structure, which is more rigid than topology and more flexible than geometry. Therefore, conformal structure leads to canonical non-rigid deformation, which is important for engineering applications, especially for shape analysis, classification and registration.

Alignment in Parameter Space. After applying the mapping, we get UV parameters and a planar disk. In UV space, we first compute the boundary positions of the conformal maps. In Figure 2.9, the boundary is shown in green. Then we rotate the coordinates of all the vertices along with the boundary positions. The rotation is not relative to each other but relative to the ground and the transformation performs alignment to the ground in parameter space.

2.5 Color Interpolation using Barycentric Coordinates

In 3D Modelling, triangles are stored as a sequence of three vertices. We generally know information about the vertices, for instance the color. We should note that barycentric coordinates can be thought as an internal coordinate system for a triangle. Hence, we use barycentric coordinates to define the color of an arbitrary point inside the triangle. Barycentric coordinates allow us to interpolate over the whole triangle.

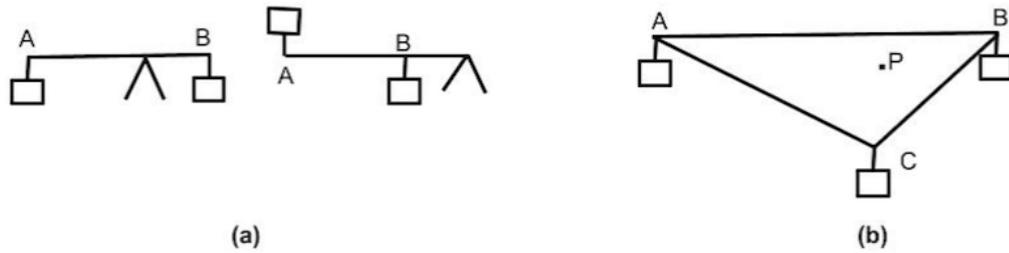


Figure 2.10 : Barycentric coordinates. (a) The fulcrum which shows the proportion of weights placed to the rod, is positioned at the barycentric coordinates. (b) The barycentric coordinates of the point P, are represented by the ratio of weights.

For each triangle, we first compute the mean value coordinates, and we use them to interpolate the data from the boundary of the unit disk to its interior.

Barycentric Coordinates. In geometry processing, the barycentric coordinate system represents the position of a data point as a simplex such as a tetrahedron or a triangle etc. The system is defined so as to be appear to be the centre of mass, essentially barycenter, of generally different masses located in its vertice coordinates. System coordinates may furthermore reach out of outer surface of the triangle, in which one or more coordinates take negative values.

Barycentric coordinates were initially described by August Ferdinand Mobius between the years 1790 - 1816. The barycentric calculus, Mobius' book, released in 1827. Mobius started with the scheme of weights attached at two distinct points on a weightless rod, see Figure 2.9, and was curious about finding the center of gravity of the rod. That is to say, he desired to locate the point to balance the rod, at which a fulcrum could be put. In his calculations, he moreover set weights as negative values. Even though this might look like contrary to common sense, a negative weight might be thought as an object which performs a force to the top, for instance a balloon. In this occasion, we should note that the center of gravity is outside of the two tied up weights.

Mobius established to find the coordinates of the center of gravity or in other words at which barycentre was positioned. The barycentre shows the proportion of weights hanged to the rod. We indicate the barycentric coordinates of the point in Figure 2.9 part (a) at which the fulcrum is placed. The specific position of the fulcrum will not be investigated as it is not the scope of the thesis.

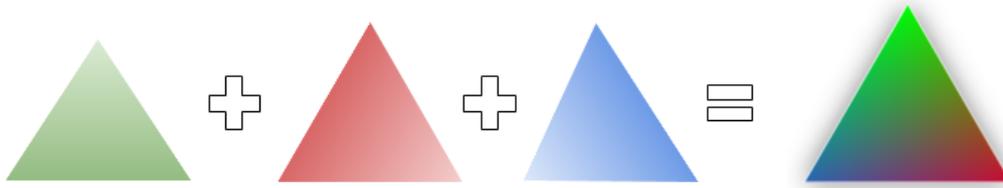


Figure 2.11 : Gouraud Shading: Color interpolation inside a triangle.

At that time Mobius enlarged his design to a system of three weighted points, which forms a triangle, see Figure 2.9 (b). Assuming that the weights A , B and C are located at the associated vertices of the triangle ABC . Thus the location of its centre of gravity namely the barycentric coordinates of the point P , are presented by the proportion of weights attached in the vertices of the triangle.

In Figure 2.10, Gouraud shading calculates the illumination for each pixel. Using triangle shading as an example, suppose each corner is assigned a pure color. Using normalized barycentric coordinates from each point in the triangle, the ratio of the color at each vertex applies to each pixel. The result is a smooth color transition across the triangle. The combination of green, red and blue, into a triangle is presented.

Areal Definition of Barycentric Coordinates. Barycentric coordinates can be used to express the position of any point located on the triangle with three scalars. The location of this point includes any position inside the triangle, any position on any of the three edges of the triangles, or any one of the three triangle's vertices themselves. To compute the position of this point using barycentric coordinates we use the following equation (1): $P = uA + vB + wC$ where A , B and C are the vertices of a triangle and u , v , and w (the barycentric coordinates), three real numbers (scalars) such that $u + v + w = 1$ (barycentric coordinates are normalized). Note that from two of the coordinates we can find the third one: $w = 1 - u - v$ from which we can establish that $u + v \leq 1$ (we will use this simple but important property later on). Equation 1 defines the position of a point P on the plane of the triangle formed by the vertices A , B and C . The point is within the triangle (A , B , C) if $0 \leq u, v, w \leq 1$. If any one of the coordinates is less than zero or greater than one, the point is outside the triangle. If any of them is zero, P is on one of the lines joining the vertices of the triangle.

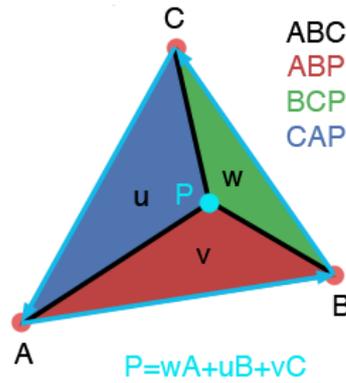


Figure 2.12 : Barycentric coordinates can be seen as the area of sub-triangles CAP (for u), ABP (for v) and BCP (for w) over the area of the triangle ABC which is why they are also called areal coordinates.

Barycentric coordinates are also known as areal coordinates. Although not very commonly used, this term indicates that the coordinates u , v and w are proportional to the area of the three sub-triangles defined by P , the point located on the triangle, and the triangle's vertices (A , B , C). These three sub-triangles are denoted ABP , BCP , CAP (see figure 2.11). Which leads us to the formulas used to compute the barycentric coordinates:

$$u = \frac{\text{TriangleCAP}_{Area}}{\text{TriangleABC}_{Area}} \quad \text{and} \quad v = \frac{\text{TriangleABP}_{Area}}{\text{TriangleABC}_{Area}} \quad \text{and} \quad w = \frac{\text{TriangleBCP}_{Area}}{\text{TriangleABC}_{Area}}$$

Color Interpolation Inside a Planar Mesh. A smooth transition between two colors can be created by linearly interpolating the color coordinates. When we have a triangle and each vertex of that triangle corresponds to a certain value, for example a color. Given an arbitrary point, inside of the triangle, how do we define its value? The solution is to use barycentric coordinates. The coordinates are triplets of numbers (x, y, z) associated with values located in the vertices of a referenced triangle.

When a data point takes its place inside of the triangle, each of its barycentric coordinates (x, y, z) place in the interval $(0,1)$. When a data point is not inside the triangle, it's still possible to compute barycentric coordinates but at least one of the coordinates will be < 0 .

GIM Database. In total, we currently have almost 12.5K geometry images to feed into the neural networks. The difference between the versions is the way we apply alignment, in parameter space or in Cartesian coordinate system. In the first version of the GIMs, there is no alignment applied. In Figure 2.13, three different versions of

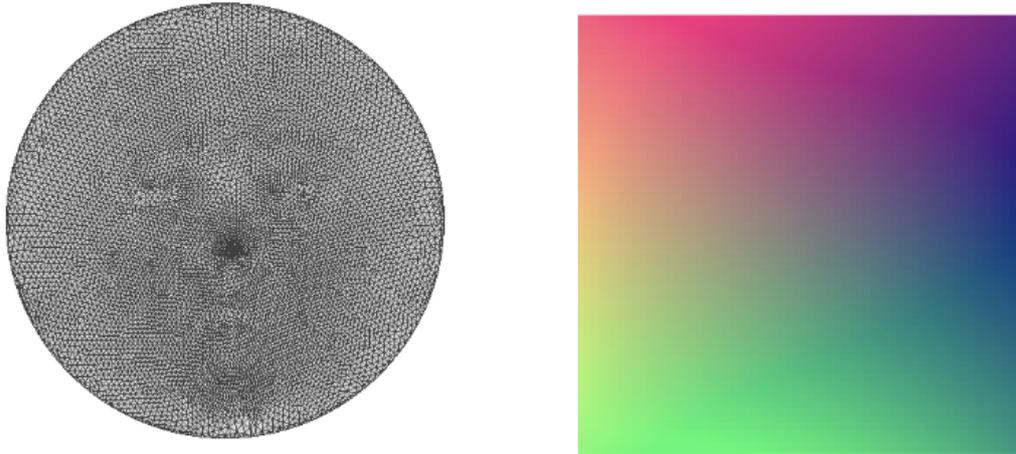


Figure 2.13 : Interpolating inside a planar mesh. (L) Flattened mesh, output of the previous steps. (R) The interpolated data: Geometry pixels.

the GIMs are presented. During experiments, we use the final, the fifth version, of the GIMs. In Chapter 4, experimental results will be published using the fifth version of the geometry imagery.

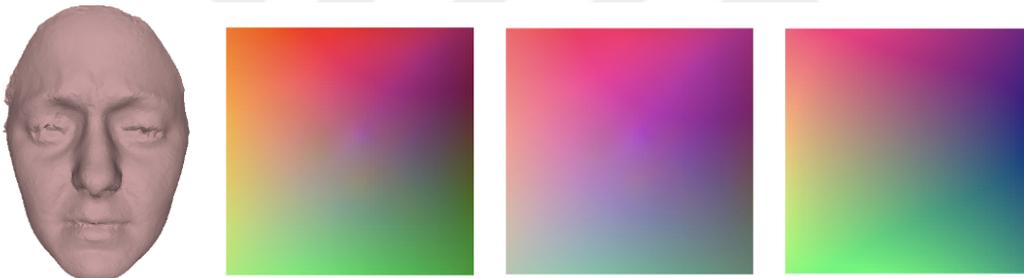


Figure 2.14 : Different geometry images of the same face model. (L) No alignment [v1]. (M) Alignment in Cartesian coordinate system [v2]. (R) Alignment in parameter space [v3].

3. DEEP LEARNING IN RECOGNITION SYSTEMS

3.1 Introduction

The science of uniquely identifying a person based on his or her physiological or behavioral characteristics is termed bio statistics. Physiological characteristics include face, iris, fingerprint, and DNA, whereas behavioral modalities include handwriting, gait, and keystroke dynamics. Jain et al. lists seven factors that are essential for any trait (formally termed modality) to be used for biometric authentication. An automated biometric system aims to either correctly predict the identity of the instance of a modality or verify whether the given sample is the same as the existing sample stored in the database. Figure 3.1 presents a traditional pipeline of a biometric authentication system.

Input data corresponds to the raw data obtained directly from the sensor(s). Segmentation, or detection, refers to the process of extracting the region of interest from the given input. Once the required region of interest has been extracted, it is preprocessed to remove the noise, enhance the image, and normalize the data for subsequent ease of processing. After segmentation and preprocessing, the next step in the pipeline is feature extraction. Feature extraction concerns the procedure of extracting unique as well as discriminatory information from the given database. These features are then used for performing classification. Classification concerns to the process of creating a model, which given a seen/unseen input feature vector is able to provide its correct label. For example, in case of a face recognition pipeline, the aim is to identify the individual in the given input sample.

Here, the input data consists of images captured from the camera, containing at least one face image along with background or other objects. Segmentation, or detection, corresponds to detecting the face in the given input image.

Several techniques can be applied for this step; the most common being the Viola Jones face detector. Once the faces are detected, they are normalized with respect



Figure 3.1 : Illustrating general biometrics authentication pipeline, which consists of five stages.

to their geometry and intensity. For feature extraction, hand-crafted features such as Gabor filterbank, histogram of oriented gradients, and local binary patterns and more recently, representation learning approaches have been used. The extracted features are then provided to a classifier such as a support vector machine or random decision forest for classification.

Automated biometric authentication systems [13] [14] [15] [16] have been used for several real-world applications, ranging from fingerprint sensors on mobile phones to border control applications at airports. One of the large-scale applications of automated biometric authentication is the ongoing project of Unique Identification Authority of India, pronounced “Aadhaar.” Initiated by the Indian Government, the project aims to provide a special recognition digit for every local who lives in India and capture his or her biometric modalities—face, fingerprint, and irises. This is done in an attempt to facilitate digital authentication anytime, anywhere, using the collected biometric data. Currently, the project has enrolled more than 1.1 billion individuals. Such large-scale projects often result in data having large intraclass variations, low interclass variations, and unconstrained environments. In following figure, sample images showcasing the large intraclass and low interclass variations that can be observed for the problem of face recognition.

Traditionally, research in feature extraction focused largely on handcrafted features such as Gabor and Haralick features. Many such hand-crafted features encode the pixel variations in the images to generate robust feature vectors for performing classification. Building on these, more complex hand-crafted features are also proposed that encode rotation and scale variations in the feature vectors as well. With the availability of training data, researchers have started focusing on learning-based techniques [17] [18] [19] [20], resulting in several representation learning-based algorithms. Moreover, because the premise is to train the machines for tasks performed with utmost ease by humans, it seemed fitting to understand and imitate the functioning of the human brain. This led researchers to reproduce similar structures to automate complex tasks, which

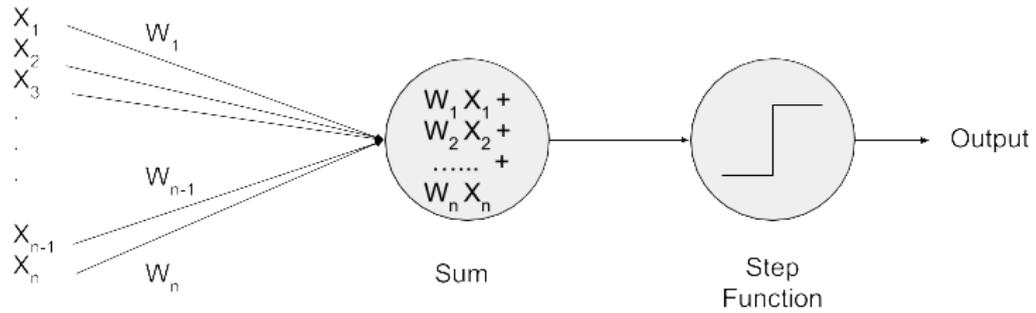


Figure 3.2 : Pictorial representation of a perceptron.

gave rise to the domain of deep learning [21] [22] [23]. Research in deep learning began with the single unit of a perceptron, which was able to mimic the behavior of a single brain neuron. Figure 3.2 illustrates a perceptron for an input vector of dimensionality $n \times 1$. The perceptron generates an output based on the input as follows: where w_i corresponds to the weight for the i th element of the input. The behavior of the perceptron is said to be analogous to that of a neuron, since, depending on a fixed threshold, the output would become 1 or 0. Thus, behaving like a neuron receiving electrical signal (input), and using the synapse (weight) to fire its output. Treating the perceptron as a building block, several complex architectures have further been suggested. Recently, the domain of deep learning has seen steep development. It is being used to address a multitude of problems with applications in biometrics, object recognition, speech, and natural language processing. Deep learning architectures can broadly be categorized into three paradigms: restricted Boltzmann machines (RBMs), autoencoders, and convolutional neural networks (CNNs). Autoencoders and Restricted Boltzmann machines are traditionally unsupervised models used for learning meaningful representations of the given data. CNNs, on the other hand, are traditionally supervised models with the objective of improving the overall classification performance.

3.2 Convolutional Neural Networks

A neural network (NN) is a supervised learning model used to perform classification [24] [25] [26]. The CNN architecture is inspired from the arrangement of neurons in the visual cortex of animals. CNNs are used to learn an efficient feature representation for a given set of images by performing spatial convolution on a two-dimensional

input followed by pooling to ensure translational invariance. The CNN architecture is inspired from the arrangement of neurons in the visual cortex of animals. CNNs are used to learn an efficient feature representation for a given set of images by performing spatial convolution on a two-dimensional input followed by pooling to ensure translational invariance. During each forward pass of a given CNN, the model learns filters or kernels, which are used for performing convolution. Deep CNNs are hierarchical in nature, that is, they learn earlier stage attributes in a shallow layer, such as edge detection, which are combined for learning higher levels of abstraction in the deeper layers of the network.

3.2.1 Architecture of a traditional convolutional neural network

A CNN is made up of several different types of layers, each performing a specific function. A traditional CNN is made up of convolution and a pooling layer alternatively, pursued with a fully-connected layer to perform classification. An explanation of each layer follows. Convolutional layer : As the name suggests, this is the building block of a CNN and is of utmost importance. Several filters are used to perform convolutions on the input vector by sliding it over the image. These filters are learned as part of the training process. The feature vector obtained after convolving an image with a filter is referred to as an activation map or filter map . The number of activation maps obtained is equal to the number of filters learned over the input. This operation encodes the fine details as well as the spatial information of the input in the feature maps. Given an input image of size $n \times n \times d$, the convolutional layer of the model learns m kernels of size $k \times k$, with a stride of size s . Thus, the output of the given layer is of the following dimension:

Rectified linear units layer (ReLU): This is used to introduce nonlinearity in the network to learn more discriminative features. It is usually applied after each convolutional layer. In the past, activation functions such as sigmoid and tanh have been used to introduce nonlinearity, however it has been observed that ReLU is faster and reduces the training time significantly. It also eliminates the vanishing gradient problem by converting all negative values to 0. The function applied in this layer is given as follows:

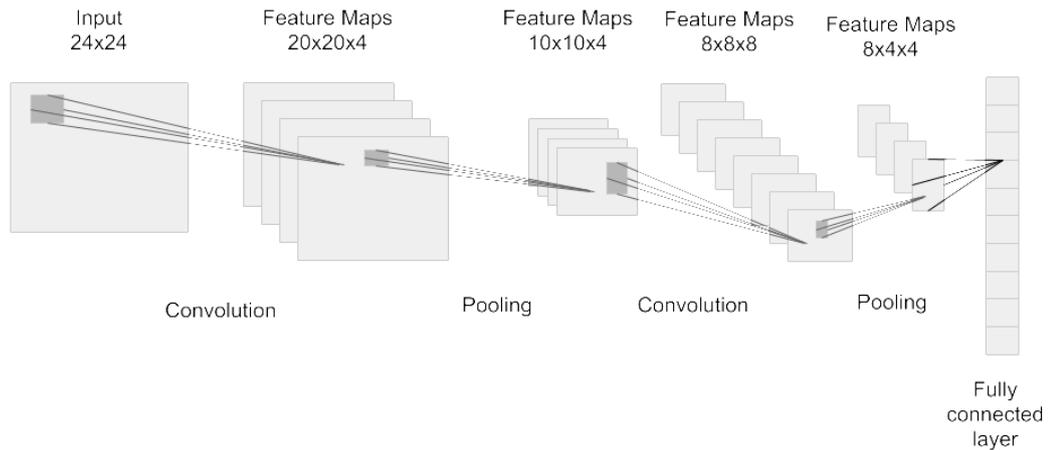


Figure 3.3 : Diagrammatic representation of a CNN having input as 24x24.

Pooling layer : Pooling layers or downsampling layers are used for dimensionality reduction of the feature maps after the convolution and ReLU layers. Generally, a filter size is chosen and an operation such as max or average is applied on the input space, which results in a single output for the given subregion. For example, if the operation defined is max-pooling for a filter size of 2x2, the max of all values in the subregion is the output of the filter. This is done for the entire feature map by sliding the filter over it. The aim of this operation is to encode the most representative information, while preserving the relative spatial details. This step not only enables dimensionality reduction, but also prevents over-fitting. Fully connected layer : After the convolutional and pooling layers, fully connected layers are attached in the network. These layers function like a traditional neural network, where each element is considered an independent node of the neural network. The output dimension of the final layer is equal to the number of classes, and each value of the output vector is the probability value associated with a class. This type of layer is used to encode supervision in the feature-learning process of the CNNs because the last layer is used for classification. In CNNs, there is no fixed order in which its constituent layers are stacked. However, typically a convolutional layer is followed by a pooling layer forming a convolutional-pooling block. This block is repeated, depending on the desired size of the network. These layers are followed by fully connected layers and the final layer is responsible for classification. ReLU is often attached after each convolutional and fully connected layer to incorporate nonlinearity in the feature-learning process. An example of a traditional CNN model consisting of five layers: two convolutional and two pooling, stacked alternatively, and the final

layer being the fully connected layer. Owing to the flexibility in the architecture, researchers have developed different models for performing feature extraction and classification tasks. Some recent developments involving CNNs are discussed in the next subsections.



4. APPLICATIONS

4.1 The BU-3DFE Facial Expression Database

Various 3D face data sets have been shared recently to be accessible for the research community to test algorithmic solutions which utilize 3D databases to carry out 3D face modelling, recognition, and analysis. The data sets consists of facial scan of people which perform expressive face models along with different pose variations. Table 4.1 provides generic databases which are presently accessible considering the problem of 3D facial recognition. Every database gives details regarding the sensors utilized during data collection, the entire amount of participated people and scanned models are presented, along with the performed expressions.

Table 4.1 : Principal details of commonly utilized 3D face expression datasets which contain non-neutral faces.

Dataset	Subjects	Scans	Expressions
FRGC v2.0	466	4007	not categorized - 4 expressions
GAVAB	61	549	smile, laugh, random
BU-3DFE	100	2500	6 expressions
Bosphorus	105	4666	action units, 6 expressions

The details of a number of popular and common 3D face data sets is summarized in Table 4.2. The data sets contain people performing neutral appearance and expressive emotions. A number of databases have special attributes to realize classification of face expressions. In a way, there is a relevance which the amount of emotions annotated in the database and which represent the true labels considering the aim of expression analysis. Across existing databases, the ones established at the Binghamton University and Boğaziçi University, have played an important role to lead the research area of classification and recognition of 3D face expressions. Nowadays, a few 4D data sets which contain active sequence of 3D face scans within time, have been made publicly available for academic purposes. The 4D databases are principally proposed to study face expression recognition that are inherently associated with the changes of the face



Figure 4.1 : BU-3DFE dataset: textured 3D facial scan of a participant performing happy face expression at the each stage of different intensities (low to highest) also neutral expression at most left.

scan in time dimension. Detailed information about one of the databases are presented in the this section of the thesis, namely the BU-3DFE, a static face expression database.

Table 4.2 : 3D dynamic (plus time dimension) and 3D static face databases for 3D face expression analysis.

Dataset	Format	Resolution	Year
BU-3DFE	static	1040 x 1329	2006
Bosphorus	static	1600 x 1200	2008
ZJU-3DFE	static	-	2006
ADSIP	dynamic	601 x 549	2009
BU-4DFE	dynamic	1024 x 681	2008
Hi4D-ADSIP	dynamic	1024 x 1728	2011

At a recent time, the BU-3DFE database [27] was established at Binghamton University. It was planned to supply 3D face scans of a community of various people, every one of them performing a group of classical emotion expressions at different stages of intensity. An entire of 100 people present in the data set, separated among male (44 people) as well as female (56 people). The subjects are effectively divided up among various national categories and ethnic descents, containing Middle-East Asian, Latino-Americans, East-Asian, Black, White and the others. Throughout the collection, every person was demanded to show the six simple face expression described via Ekman [38], specifically, surprise (SU), sadness (SA), anger (AN), happiness (HA), disgust (DI), fear (FE) and also the neutral state (NE). Every face emotion has four stages of intensities — ground, medium, strong and top — excluding neutral faces.

There exists 25 3D face expressions for every person, producing a total of 2500 3D face expressions in the data set. For instance, Figure 4.1 presents the 3D facial textured scans of a participant for happy emotion with four stages of intensities along



(a)

face region	#points
right eye	8
left eye	8
right eyebrow	10
left eyebrow	10
nose	12
mouth	20
face boundary	15

(b)

Figure 4.2 : The BU-3DFE dataset: (a) the 83 facial landmark points presented using a 3D textured facial model having neutral state; (b) the amount of non-automatically defined landmark points considering various areas of the facial scans.

with the neutral scan. Every one of the 3D face scans is furthermore identified using an unprocessed 3D mesh, a cut mesh, a group of 83 non-automatically noted face landmark points, along with face pose vectors. The database supply an entire 3D identification of a facial scan performing one of the vaious face expressions. Textured and cropped 3D facial scans, as well as the 83 face landmark points are presented in Figure 4.2. As a summary of Figure 4.2 , the landmark points are shared within correlation to the greatest determining characteristics of the facial scans, which is, mouth, nose, eyebrows, eyes as well as a few landmark points onto the facial boundaries.

A description in more details of the BU-3DFE data set might be obtained from the research paper [41]. Eventually, we discover that the data set are obtained using an stereo camera setting with an ad-hoc network which constructs the 3D scan of the participant with two separate right and left directions. By making so, stereo cameras furthermore obtain 2D colorful depictions of the facial scan from a right and left

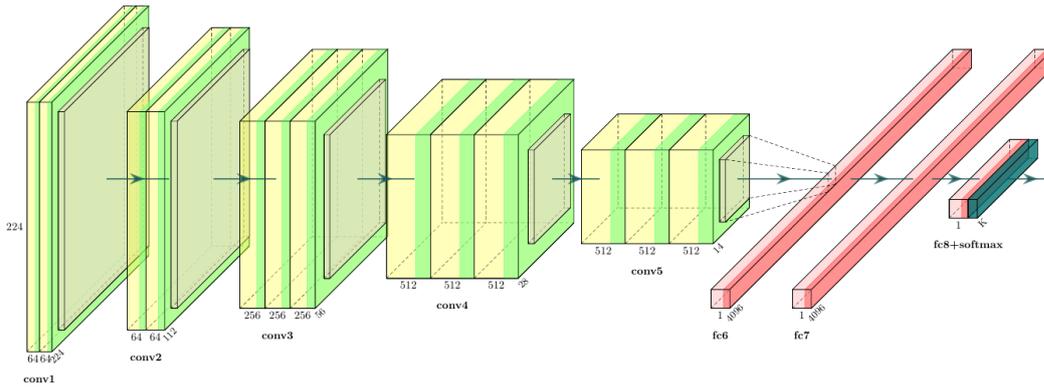


Figure 4.3 : Illustration of the VGGNet model.

perspective of the facial model. The main attributes of the BU-3DFE database might be summed up in a subsequent manner:

Advantages — classified face expression scans; various stages of intensities for every expressive scan; prepared facial scans; models having annotated face landmark points

Disadvantages — average amount of participants; only front view poses; 2D colorful images supplied however with right and left views.

4.2 Experiments

4.2.1 Feature extractor: vggnet architecture

ImageNet is officially a research work focused on, in a manual way, tagging and classifying images in approximately 22,000 diverse object classes considering the aim of computer vision analysis. On the other hand, we are familiar with the phrase ImageNet in the concept of deep neural networks as well as machine learning, it is apparently concerning the ImageNet Large Scale Visual Recognition Challenge. Purpose of the object categorization competition is for training a learn-able model which is capable of accurately categorizing a given image into one of 1,000 different object classes. Learn-able weights are modelled on approximately 1.2M imagery in the training images along with 50K as validation set as well as 100K images in the test set. The 1,000 object classes show image categories which we come across in our daily lives, for instance vehicle types, various household objects, cats, species of dogs, and more and more. The complete table of entity classes is available in the ILSVRC challenge [42]. Whenever it is about classifying images, the ImageNet is

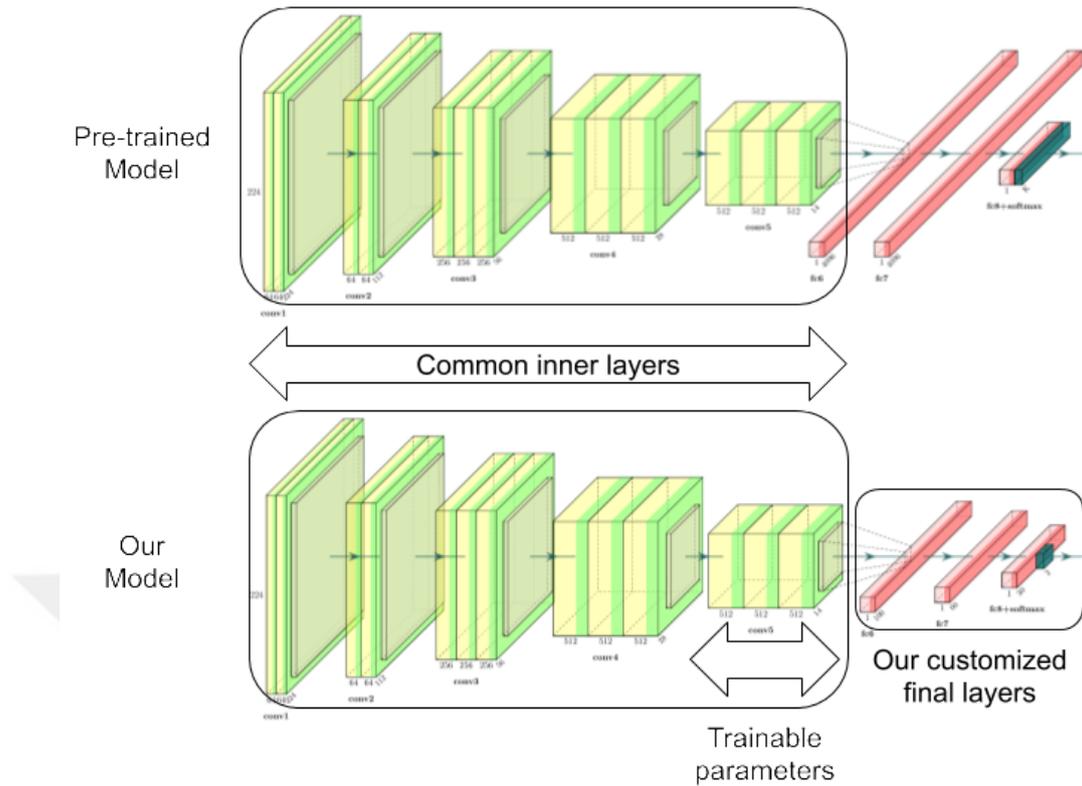


Figure 4.4 : Fine-tuning a VGG16 network pretrained on ImageNet.

the ultimate standard for classification algorithms in computer vision domain — and the scoreboard of the competition has been in the control of deep neural networks and convolutional methods since 2012. The most recent, pre-trained neural networks involved in the core Keras deep learning library show one of the best achieving results. Those networks furthermore represent a powerful capability to make generalization to the objects which are not in the ImageNet data by applying transfer learning, for instance extracting features followed by a fine tuning and training only a classifier on top the convolutional base.

The VGG-Net model, see Figure 4.3, was announced in a research paper [43] by Simonyan and Zisserman in 2014. The neural network is distinguished considering the simpleness of its architecture, utilizing just 3x3 convolution layers which are stacking one after the other resulting in an increase of the depth. Diminishing size of the volume is overcome using maximum pooling. Finally, two entirely connected layers, every one of them with 4,096 neurons are afterwards pursued via a soft-max classification function. Back in 2014, 19 and 16 layered neural networks took into consideration as extremely deep, in spite of the fact that we currently have the Res-Net that might

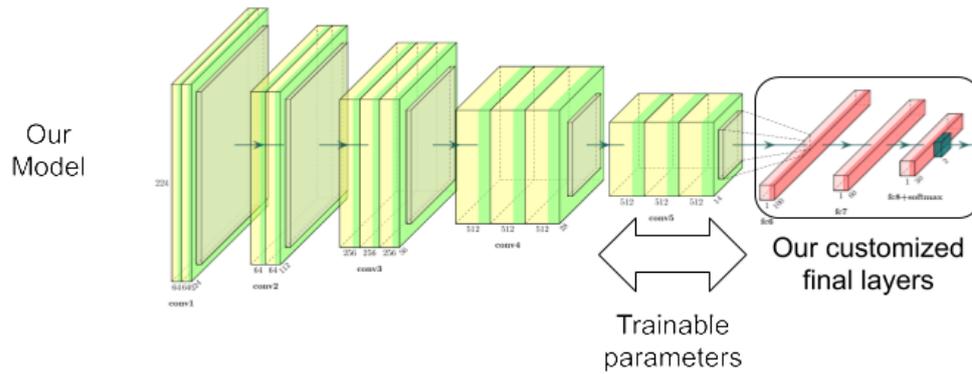


Figure 4.5 : A customized VGG16 model fine-tuned with our database.

be modeled successfully from 50 to 200 layers using ImageNet as well as more than 1,000 stacked layers considering CIFAR-10, a small-scale data set. Simonyan and Zisserman [43] discovered that to train VGG-19 or VGG-16 are competitive especially considering convergency while using more deeper neural networks, therefore to obtain an easier training time, they initially trained small varieties of VGG-Net along with smaller number of learn-able weights. The networks with less layers converges quickly so that they are afterwards utilized like initializer of the deeper, bigger neural networks — the procedure is also known as pre-training, see Figure 4.4. When thinking logically in a sensible way, pre-training is an extremely time wasting, monotonous process, demanding a complete neural network to be modeled and obtained learn-able weights prior to serving like an initializer in a more deep neural network. We do not prefer to utilize pre-training in majority of the occasions and as an alternative we propose Xavier/Glorot initializer and MSRA initializer, generally known as He initializer. The significance of how to initialize weights as well as the convergency of neural networks are both entirely important research topics and it is not the scope of this thesis. It is unfortunate that, there exists a few crucial disadvantages using VGG model architecture:

- It is extremely tiresome while training.
- The learn-able weights of the network are absolutely big with respect to disk and bandwidth.

According to the number of fully connected neurons and the deepness, VGG-Net takes up to the 574 MB for VGG-19 model and up to 533 MB for VGG-16 model. It

causes to deploy VGG-Net as a tedious process. We utilize VGG-Net in a number of various image classification tasks in computer vision as well as other deep neural networks applications. On the other hand, networks with smaller architectures are generally more demandable for instance Squeeze-Net, GoogLeNet, etc. Consequently, in our experiments we will use VGG-Net model and transfer learning such as feature extraction followed by a fine tuning. Different scenarios will be utilized, see Figure 4.5, during the experiments ranging from training only a classifier on top of the VGG base, to making all of the VGG parameters learn-able.

4.2.2 Binary classifier: one-versus-one classification

A support vector machine is only a binary classifier: that is, it can only classify two classes at a time. Therefore, in order to classify multiple classes, i.e., more than two, it has to train two or more binary classifiers by selecting groups of classes to belong to one or the other class in a pair. This is known as multi-class classification.

One scheme for doing this is one-vs-one. Two pairs of classes are selected at a time and a binary classifier trained for them. This is done for every possible pair of classes thus there are $n(n - 1)/2$ of them where n is the total number of classes. During the classification phases, all the binary classifiers are tested. For each of them, a win for one class is a vote for that class. The class with the most votes wins.

4.2.3 Binary classifier: one-versus-all classification

Logistic regression is a simple classification method which is widely used in the field of machine learning. Logistic regression is a linear model, which means that the decision boundary has to be a straight line. Suppose we have a classifier for sorting out input data into 3 categories: Class 1 (+), Class 2 (-), Class 3 (x).

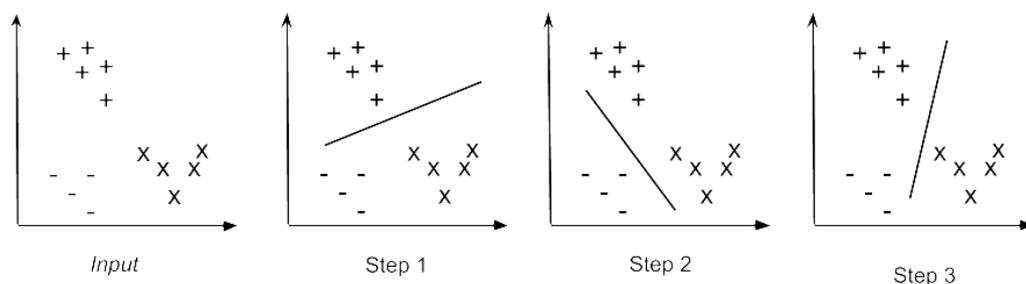


Figure 4.6 : One-vs-all classification.

We may turn this problem into 3 binary classification problems (i.e. where we predict only if y is an element of class 1 or not) to be able to use classifiers such as Logistic Regression. We take values of one class and turn them into positive examples, and the rest of classes into negatives. In Figure 4.6, in step 1, plus signs are positive, and the rest are negative and we run a classifier on them. Then we calculate $h_{\theta}^{(1)}(x)$ for it. In the next step, we do same with negative sign: make them positive, and the rest negative then we calculate $h_{\theta}^{(2)}(x)$ for it. In step 3, finally, we make cross sign as positive and the rest as negative and calculate $h_{\theta}^{(3)}(x)$ for it. So we have fit 3 classifiers:

- $h_{\theta}^{(i)}(x) = P(y = i|x; \theta), i = 1, 2, 3$
- Now, having calculated the vector: $h_{\theta}(x) = [h_{\theta}^{(1)}(x), h_{\theta}^{(2)}(x), h_{\theta}^{(3)}(x)]$ We just pick up the maximal value.
- Finally, we choose $\max_i h_{\theta}^{(i)}(x)$

4.3 Experimental Results

4.3.1 3d fer using gim database: one-versus-one classification with 7.2m learnable parameters

Automatic recognition of facial expression is a major challenge for the pattern analysis research community because the expressions are generated by non-rigid object deformations, which vary from person to person. Expression recognition from 3D models, in particular recognition from a static image, is more intriguing due to various information during performing an expression. In this thesis, facial expression recognition technique is presented using 3D face models from the BU-3DFE database. The whole recognition process is fully automatic. Results of our study reveal that facial expressions can be identified and recognized from 3D mesh models and 3D face models capable of capturing the characteristic of the subject well.

An original facial expression recognition system for geometry images was developed using the proposed method. Face expression models from BU-3DFE face database from Binghamton University were used to assess the framework's performance assessment. The original mesh models were normalized (in orientation and scale), and the two eyes were normalized to align approximately to the same location in the

final 3D models. A few examples of normalized faces are shown in Chapter 2. Readers can refer to [41] for details on the collection of models in the BU-3DFE face database. In this experiment, a total of 2500 models from 100 people (56 females and 44 males) were used. The number of models corresponding to angry, disgust, fear, happy, sad and surprised expressions is, 400 models, the same.

Experimental results in facial expressions is summarized in Table 4.3. The system correctly recognizes 77.9% of women's facial expressions, while 81.1% of men's expressions are recognized correctly. Also recognition rate of the surprise face is the lowest one, especially for women. This may be due to the fact that surprise expressions are relatively different from human to human. We use one-versus-one binary classifiers and the last four convolutional layers is trainable resulting in a total of 7.2 million learnable parameters.

Due to differences in testing our data, it is difficult to directly compare our results with the reported techniques. However, the average recognition rate of 79.5% is a very promising result. Our results are obtained using our geometry image database, which is publicly available. Therefore, it can be used by other researchers as a criterion for direct performance comparison.

Table 4.3 : One-versus-one recognition Results: 79.5% accuracy in average on test data. One binary classifier for each expression performs a one-versus-one classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, the last four convolutional layers of the VGG16 model are trainable, in total 7.2M parameters are learnable.

	Angry	Disgust	Fear	Happy	Sad	Surprise	Average
Female	78.3	70.4	84.2	80.7	82.6	71.3	77.9
Male	82.1	84.4	77.3	74.7	86.3	81.5	81.1
Average	80.2	77.4	80.8	77.7	84.5	76.4	79.5

4.3.2 3d fer using gim database: one-versus-one classification with 16.8m learnable parameters

In this experimental scenario, we increase the number of trainable layers and actually we train all the layers of the VGG-16 model. This results in a total of 16.8 million trainable parameters to tweak for our expression recognition task. We use one-versus-one binary classifiers and all of the convolutional layers of VGG-16 is trainable resulting in a total of 16.8 million learnable parameters. One-versus-one binary classifier performs a classification between two classes for instance, happy and neutral expression.

In Table 4.4, we show that there is a 1.9% increase in average recognition rate after making all the parameters trainable. Our recognition rate is now 81.4%.

Experimental results after making all the VGG layers trainable, is summarized in Table 4.4. The system correctly recognizes 82.1% of women's facial expressions, while 80.7% of men's expressions are recognized correctly. Also recognition rate of the surprise face is the lowest one, especially for women. This may be due to the fact that surprise expressions are relatively different from human to human. Due to differences in testing our data, the average recognition rate of 81.4% is a promising result.

Table 4.4 : One-versus-one recognition results: 81.4% accuracy in average on test data with leave-10-person-out data split method. One binary classifier for each expression performs a one-versus-one classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, all of the sixteen layers of the VGG16 model are trainable, in total 16.8M parameters are learnable.

	Angry	Disgust	Fear	Happy	Sad	Surprise	Average
Female	80.0	82.3	85.2	88.7	78.2	77.8	82.1
Male	79.8	79.9	80.1	81.2	83.8	79.2	80.7
Average	79.9	81.1	82.7	84.9	81.0	78.5	81.4

4.3.3 3d fer using gim database: one-versus-all classification with 7.2m learnable parameters

Our renewed results are obtained under the circumstances listed below.

- One binary classifier for each expression performs a one-versus-all classification.
- The accuracy is computed using the average of the binary classifier accuracy per expression.
- The number of trainable parameters is 7.2 million which means we only train the last four convoluional layers in this experimental setup.

The system correctly recognizes 82.7% of sad expressions, while 64.7% of fear expressions are recognized correctly. We should note that previous experiments, 79.5% and 81.4% recognition rates, were obtained using one-versus-one classification. Our recognition rate is now 72.8% using one-versus-all classification and the average recognition rate of 72.8% is a very promising result.

Table 4.5 : One-versus-all recognition results: 72.8% accuracy in average on test data with leave-6-person-out data split method. One binary classifier for each expression performs a one-versus-all classification. The accuracy is computed using the average of the binary classifier accuracy per expression. During the experiments, the last four convolutional layers of the VGG16 model are trainable, in total 16.8M parameters are learnable.

Database	Angry	Disgust	Fear	Happy	Sad	Surprise	Average
BU-3DFE	72.9	66.7	64.7	72.2	82.7	77.3	72.8

5. CONCLUSION

The importance of experimenting in 3D domain is that 3D modeling improves the 2D drawbacks such as illumination, pose variations, etc. Therefore, having a good recognition method might be useful for human computer interaction, criminal investigation, airport security or psychological examination.

Geometry based approach mainly describes the way to process geometric pixels. 3D geometry of a face model can be stored as a 2D color image which is so-called a geometry image. Geometry images [5] are a powerful, completely regular representation of triangular meshes, describing each vertex and its attributes as a pixel in one or more rectangular images and thus defining an implicit connectivity through neighbouring pixels.

We propose a pipeline to recognize 3D facial expressions based on modeling surface geometry. The framework was designed to spot 3D expressions from geometry images obtained using 3D face models. In 3D domain, The BU-3DFE database is one of the most widely used databases. We use our own geometry image database which is created using 3D expression models from BU-3DFE. There exists 2500 facial models in the original database and associated 2500 geometry images in the GIM database. The proposed pipeline was evaluated on only using different versions of GIM databases. VGG deep neural network model and different binary classifiers were used to obtain classification results. We observed that the expressions tend to give similar scores [28] [29] [30] [31] [32] [33] [34] like other recognition frameworks in the literature, which are employing 3D models from the BU-3DFE database.

Our experimental results show that VGG-16 model has a best classification accuracy of 72.8% on geometry image database using one-versus-rest classification method.

In conclusion, the results reveal that a VGG16 network is capable of handling complicated information from geometry pixels of 3D expression models.



REFERENCES

- [1] **Revina, I.M. and Emmanuel, W.S.** (2018). A survey on human face expression recognition techniques, *Journal of King Saud University-Computer and Information Sciences*.
- [2] **Ou, J.** (2012). Classification Algorithms Research on Facial Expression Recognition, *Physics Procedia*, 25, 1241–1244.
- [3] **Savran, A. and Sankur, B.** (2008). Non-rigid registration of 3D surfaces by deformable 2D triangular meshes, *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, pp.1–6.
- [4] **Savran, A., Alyüz, N., Dibekliöglu, H., Çeliktutan, O., Gökberk, B., Sankur, B. and Akarun, L.** (2008). Bosphorus database for 3D face analysis, *European Workshop on Biometrics and Identity Management*, Springer, pp.47–56.
- [5] **Savran, A., Alyüz, N., Dibekliöglu, H., Çeliktutan, O., Gökberk, B., Sankur, B. and Akarun, L.** (2008). 3d face recognition benchmarks on the bosphorus database with focus on facial expressions, *Proceedings of Workshop on Biometrics and Identity Management (WBIM)*.
- [6] **Yin, L., Wei, X., Sun, Y., Wang, J. and Rosato, M.J.** (2006). A 3D facial expression database for facial behavior research, *7th international conference on automatic face and gesture recognition (FGR06)*, IEEE, pp.211–216.
- [7] **Ekman, P.** (2004). Emotional and conversational nonverbal signals, Language, knowledge, and representation, Springer, pp.39–50.
- [8] **Gu, X., Gortler, S.J. and Hoppe, H.** (2002). Geometry images, *ACM Transactions on Graphics (TOG)*, volume 21, ACM, pp.355–361.
- [9] **Losasso, F., Hoppe, H., Schaefer, S. and Warren, J.** (2003). Smooth geometry images, *Symposium on Geometry Processing*, volume 43, Citeseer, pp.138–145.
- [10] **Sander, P.V., Wood, Z.J., Gortler, S., Snyder, J. and Hoppe, H.** (2003). Multi-chart geometry images.
- [11] **Gu, X. and Yau, S.T.** (2003). Global Conformal Surface Parameterization, *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP '03*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, pp.127–137, <http://dl.acm.org/citation.cfm?id=882370.882388>.

- [12] **Wei, X., Li, H. and Gu, X.D.** (2017). Three Dimensional Face Recognition via Surface Harmonic Mapping and Deep Learning, *Chinese Conference on Biometric Recognition*, Springer, pp.66–76.
- [13] **Byeon, Y.H. and Kwak, K.C.** (2014). Facial expression recognition using 3d convolutional neural network, *International journal of advanced computer science and applications*, 5(12).
- [14] **Amin, D., Chase, P. and Sinha, K.** (2017). Touchy Feely: An Emotion Recognition Challenge, *Palo alto: Stanford*.
- [15] **Zulqarnain Gilani, S. and Mian, A.** (2018). Learning from millions of 3D scans for large-scale 3D face recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1896–1905.
- [16] **Li, C. and Soares, A.** (2011). Automatic facial expression recognition using 3D faces, *Int. J. Eng. Res. Innov*, 3, 30–34.
- [17] **Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J. and Worek, W.** (2005). Overview of the face recognition grand challenge, *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, IEEE, pp.947–954.
- [18] **Cheng, S., Kotsia, I., Pantic, M. and Zafeiriou, S.** (2018). 4dfab: A large scale 4d database for facial expression analysis and biometric applications, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.5117–5126.
- [19] **Khorrarni, P., Paine, T. and Huang, T.** (2015). Do deep neural networks learn facial action units when doing expression recognition?, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp.19–27.
- [20] **Oyedotun, O.K., Demisse, G., El Rahman Shabayek, A., Aouada, D. and Ottersten, B.** (2017). Facial expression recognition via joint deep learning of rgb-depth map latent representations, *Proceedings of the IEEE International Conference on Computer Vision*, pp.3161–3168.
- [21] **Mollahosseini, A., Hasani, B., Salvador, M.J., Abdollahi, H., Chan, D. and Mahoor, M.H.** (2016). Facial expression recognition from world wild web, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.58–65.
- [22] **Mayya, V., Pai, R.M. and Pai, M.M.** (2016). Automatic facial expression recognition using DCNN, *Procedia Computer Science*, 93, 453–461.
- [23] **Kim, D., Hernandez, M., Choi, J. and Medioni, G.** (2017). Deep 3D face identification, *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, pp.133–142.
- [24] **Jan, A., Ding, H., Meng, H., Chen, L. and Li, H.** (2018). Accurate facial parts localization and deep learning for 3d facial expression recognition, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, pp.466–472.

- [25] **Huynh, X.P., Tran, T.D. and Kim, Y.G.**, (2016). Convolutional neural network models for facial expression recognition using bu-3dfe database, *Information Science and Applications (ICISA) 2016*, Springer, pp.441–450.
- [26] **Hussain, N., Ujir, H., Hipiny, I. and Minoi, J.L.** (2017). 3D Facial Action Units Recognition for Emotional Expression, *arXiv preprint arXiv:1712.00195*.
- [27] **Zhang, Z., Girard, J.M., Wu, Y., Zhang, X., Liu, P., Ciftci, U., Canavan, S., Reale, M., Horowitz, A., Yang, H. et al.** (2016). Multimodal spontaneous emotion corpus for human behavior analysis, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.3438–3446.
- [28] **Berretti, S., Del Bimbo, A., Pala, P., Amor, B.B. and Daoudi, M.** (2010). A set of selected SIFT features for 3D facial expression recognition, *2010 20th International Conference on Pattern Recognition*, IEEE, pp.4125–4128.
- [29] **Gong, B., Wang, Y., Liu, J. and Tang, X.** (2009). Automatic facial expression recognition on a single 3D face by exploring shape deformation, *Proceedings of the 17th ACM international conference on Multimedia*, ACM, pp.569–572.
- [30] **Wang, J., Yin, L., Wei, X. and Sun, Y.** (2006). 3D facial expression recognition based on primitive surface feature distribution, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, IEEE, pp.1399–1406.
- [31] **Soyel, H. and Demirel, H.** (2007). Facial expression recognition using 3D facial feature distances, *International Conference Image Analysis and Recognition*, Springer, pp.831–838.
- [32] **Tang, H. and Huang, T.S.** (2008). 3D facial expression recognition based on automatically selected features, *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, IEEE, pp.1–8.
- [33] **Lemaire, P., Ben Amor, B., Ardabilian, M., Chen, L. and Daoudi, M.** (2011). Fully automatic 3D facial expression recognition using a region-based approach, *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, ACM, pp.53–58.
- [34] **Lemaire, P., Ardabilian, M., Chen, L. and Daoudi, M.** (2013). Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients, *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, pp.1–7.



APPENDICES

APPENDIX A.1 : Angry expression recognition: loss and accuracy tables

APPENDIX A.2 : Disgust expression recognition: loss and accuracy tables

APPENDIX A.3 : Fear expression recognition: loss and accuracy tables

APPENDIX A.4 : Happy expression recognition: loss and accuracy tables

APPENDIX A.5 : Sad expression recognition: loss and accuracy tables

APPENDIX A.6 : Surprise expression recognition: loss and accuracy tables



APPENDIX A.1

Angry expression recognition

The anger and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 500 to 4000 were tested on the models. Experimental results in anger expression is summarized in Table 1 for female models and Table 2 for male models in detail.

Table 1 : Angry expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
500 Epochs	0.337	0.857	0.780	0.773
1000 Epochs	0.294	0.880	0.695	0.780
1500 Epochs	0.245	0.889	0.874	0.755
2500 Epochs	0.230	0.902	1.108	0.772
4000 Epochs	0.210	0.918	1.396	0.783

Table 2 : Angry expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.448	0.800	0.605	0.817
1500 Epochs	0.415	0.812	0.759	0.761
2500 Epochs	0.407	0.793	0.846	0.821
3500 Epochs	0.374	0.798	1.253	0.793

APPENDIX A.2

Disgust expression recognition

The disgust and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 1000 to 3500 were tested on the models. Experimental results in anger expression is summarized in Table 3 for female models and Table 4 for male models in detail.

Table 3 : Disgust expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.242	0.906	1.127	0.673
1500 Epochs	0.185	0.992	1.188	0.691
2500 Epochs	0.146	0.940	1.414	0.704
3500 Epochs	0.066	0.971	2.794	0.689

Table 4 : Disgust expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.449	0.797	0.617	0.819
1500 Epochs	0.378	0.828	0.888	0.738
2500 Epochs	0.377	0.852	0.831	0.756
3500 Epochs	0.410	0.803	0.972	0.844

APPENDIX A.3

Fear expression recognition

The fear and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 1000 to 3500 were tested on the models. Experimental results in anger expression is summarized in Table 5 for female models and Table 6 for male models in detail.

Table 5 : Fear expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.288	0.875	0.596	0.842
1500 Epochs	0.161	0.930	1.025	0.811
2500 Epochs	0.222	0.883	1.067	0.793
3500 Epochs	0.190	0.917	0.740	0.813

Table 6 : Fear expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.465	0.799	0.533	0.771
1500 Epochs	0.411	0.818	0.587	0.751
2500 Epochs	0.418	0.809	0.598	0.764
3500 Epochs	0.409	0.799	0.646	0.773

APPENDIX A.4

Happy expression recognition

The happy and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 1000 to 3500 were tested on the models. Experimental results in anger expression is summarized in Table 7 for female models and Table 8 for male models in detail.

Table 7 : Happy expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.247	0.902	0.554	0.775
1500 Epochs	0.180	0.930	0.640	0.778
2500 Epochs	0.149	0.939	0.818	0.807
3500 Epochs	0.163	0.922	0.916	0.779

Table 8 : Happy expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.280	0.878	0.670	0.747
1500 Epochs	0.178	0.927	0.955	0.663
2500 Epochs	0.109	0.966	1.241	0.561
3500 Epochs	0.134	0.960	1.471	0.624

APPENDIX A.5

Sad expression recognition

The sad and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 1000 to 3500 were tested on the models. Experimental results in anger expression is summarized in Table 9 for female models and Table 10 for male models in detail.

Table 9 : Sad expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.309	0.864	0.611	0.824
1500 Epochs	0.297	0.872	0.801	0.826
2500 Epochs	0.337	0.872	0.617	0.819
3500 Epochs	0.512	0.792	0.473	0.821

Table 10 : Sad expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.481	0.793	0.421	0.863
1500 Epochs	0.429	0.810	0.441	0.861
2500 Epochs	0.427	0.792	0.550	0.863
3500 Epochs	0.456	0.790	0.556	0.862

APPENDIX A.6

Surprise expression recognition

The surprise and neutral faces in the database were used in this experiment, a total of 500 models from 100 people (56 females and 44 males). The number of models corresponding to angry and neutral expressions are 400 and 100, respectively. Effects of one of the most important hyper-parameters, number of epochs, ranging from 1000 to 3500 were tested on the models. Experimental results in anger expression is summarized in Table 11 for female models and Table 12 for male models in detail.

Table 11 : Surprise expression recognition with female models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.199	0.918	1.129	0.698
1500 Epochs	0.150	0.927	1.410	0.713
2500 Epochs	0.171	0.925	1.408	0.698
3500 Epochs	0.363	0.793	1.093	0.693

Table 12 : Surprise expression recognition with male models: loss and accuracy values.

Number of Epochs	Training Time		Test Time	
	Loss	Accuracy	Loss	Accuracy
1000 Epochs	0.289	0.869	0.472	0.815
1500 Epochs	0.293	0.817	0.384	0.790
2500 Epochs	0.339	0.814	0.385	0.813
3500 Epochs	0.278	0.830	0.416	0.797



CURRICULUM VITAE



Name Surname: Neşe Güneş

Place and Date of Birth: Akhisar, 1986

E-Mail: nese.gunes@itu.edu.tr

EDUCATION:

- **B.Sc.:** 2015, Galatasaray University, Faculty of Engineering and Technology, Department of Computer Engineering
- **M.Sc.:** 2019, Istanbul Technical University, Faculty of Computer and Informatics, Department of Computer Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2018-Present Research Assistant at Istanbul Technical University
- 2017-2018 Istanbul Technical University at Center for Satellite Communications and Remote Sensing (CSCRS).
- 2015 Completed B.E. degree at Galatasaray University

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Gunes N., 2019. Geometry Based Three Dimensional Face Expression Recognition. *International Conference on Natural and Engineering Science*, August 27-30, 2019 Istanbul, Turkey.
- Gunes N., 2019. Geometry Based Three Dimensional Face Expression Recognition, *International Journal of Natural and Engineering Sciences*, ISSN: 1307-1149, E-ISSN: 2146-0086, 2019.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Gunes N., 2019. Flappy Bird is NP-hard. *International Mediterranean Science and Engineering Congress*, April 25-27, 2019 Antalya, Turkey.

