INTERNATIONAL HELLENIC UNIVERSITY

Automated 3D facial landmarks localization for 4D dataset

Papaioannou Panagiotis

SID: 3308170015

School of Science & Technology

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

NOVEMBER 2019

THESSALONIKI – GREECE

Automated 3D facial landmarks localization for 4D dataset

Papaioannou Panagiotis

SID: 3308170015

Supervisor: Dr. Rigas Kotsakis

Supervising Committee Members:

Assoc. Prof. Name Surname

Assist. Prof. Name Surname

School of Science & Technology

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

NOVEMBER 2019

THESSALONIKI – GREECE

ABSTRACT

3D facial landmarks are crucial as a first step for various applications like alignment and registration which implies that successful localization can lead to better outcomes for further approaches. Even though the problem of landmarks localization has been examined in depth for the case of 2D images, the amount of work that has been done for the 3D case is significantly smaller and it can be said that is negligible for the 4D case (3D plus time). In this thesis, a literature review for the current approaches for 3D a localization of facial landmarks is carried out. Finally, experiments of a new method show validity and improvement in certain landmarks.

CONTENTS

I. Introduction

It is well-known that 3D facial landmarks are vital for numerous applications like alignment and registration which indicates that effective localization can have an advantage to better outcomes for the other methods. Nevertheless, the issue of landmarks localization has been analyzed in depth for the case of 2D images, the amount of work that has been done for the 3D case is considerably smaller and it can be stated that is insignificant for the 4D case (3D plus time).

In spite of the mutual methods that exploit information from 2D images and video, there is a crucial alternative, the face analysis from 3D scanned data. Three-dimensional data cannot be influenced equivalently of long-standing tasks of illumination and pose variance. Due to the fact that face recognition become a sector of huge importance, the accuracy of common 2D data systems has been enhanced by the existing shape data into structure of three-dimensional surfaces.

The majority of recognition methods use facial landmarking. For instance, the use of key landmark points benefits canonical face models to be fitted on the data. Then, statistical processes are required for a comprehensive initial marking method and face segmentation is needed in order to express analysis that is supported by the use of landmark positions. Despite being the standard procedure for many automatic 3D face systems, facial landmarking at present is completed manually or semi-automatically. However, manual landmarking is a tedious job and is susceptible to inaccuracies as well.

In the current bibliography, there are some ways to investigate our purpose. One way is the use of geometric characteristics and one other the use of a texture method as well. However, in our research we will try to find a new system without the aforementioned characteristics. We will use the existing faces from the database of Facesoft company and we will convert the faces from 2D to 3D (rendering) by finding a matrix. Rendering is the procedure intricated in the creation of a 2D or 3D picture by a pattern with the assistance of appliance platforms (commonly applied in architectural drawings, animated films, as well as simulators). These procedures and attributes that applied differ from strategy to strategy. Moreover, rendering benefits rise effectiveness as well as decrease cost for forming.

It has been past more than two decades since three-dimensional Morphable Face Models were originally introduced in “Cunningham (1999)”. Morphable models have been recommended as an ordinary facial statement as well as a moral method in analyzing images. According to “Blanz and Vetter (1999)”, they established and confronted a lot of affiliate challenges. As a result, their outcomes were judged pioneering. Nevertheless, this method does not only play an important historical role. Recently, 3DMM have been experienced again in deep learning methods and are included in plenty explanations in imaging analysis.

Three-dimensional Morphable Face Models can be used reproductively for facial formation while appearance focusing into 2 crucial concepts. To begin with, every edge of the face is in high correspondence. This correspondence generally formed at the initial steps of the process and afterwards retained during the course of any additional managing step. Due to this matching, linear patterns of faces could be specified significantly. They can generate morphological pragmatic expressions which are called morphs. The other concept requires splitting the form and the shade of the face. In this way, the above-mentioned features can be unraveled from exterior aspects like illumination. A statistical pattern of the distribution of faces might be included in the Morphable Model. The authors Blanz and Vetter used a principal component analysis (PCA) initially while in following effort they made use of additional methods.

The original investigation question mark following the concept of 3DMM was in what way a visual scheme, biological or artificial, is able to handle the high range of views that a specific class of objects know how to produce, and by what means objects are characterized to disentangle vision duties. The most important hypothesis in morphable model progress was earlier experience regarding object classes performs a crucial position in imaging. It facilitates in explanation of different ill-posed issues. A crucial role of three-dimensional Morphable Models is to obtain such preceding information, and to find out spontaneously from a set of cases. The virtualization is universal, so it might be used to various objects and duties. For many years, facial virtualization as well as facial assignment recognition were on the centered of image investigation. The methodology in Eigenfaces of "Sirovich and Kirby (1987)" as well as of "Turk and Pentland (1991)" performed an essential and very significant example in this area. In their investigation, authors discovered an obvious face illustration from models and worked completely on grey-levels in the picture field. Faces' images were handled by Eigenfaces as a vector space which after that carried out a PCA where eigenvectors indicating the major methods of fluctuation. Eigenface's main disadvantage referred to the limitation to a permanent pose and illumination, but the absence of effectiveness of shape's differences representation, as well. As soon as the factors in linear patterns of eigenvectors are altered constantly, forms can become colorless, instead of shift alongside the image plane. Consequently, the model is not able to obtain a unique factor, as an example, space in the middle of the eyes. "Atick et al. (1996)" expanded the method of eigenfaces into three-dimensional facial surfaces. They produced facial shadowing fluctuations, even by having an equal restriction.

Numerous investigation parties advanced by inserting an Eigen decomposition of 2D shape variations among different faces. This method offered equally an exact shape model. Following the distortion of images, a lined up Eigenface pattern is produced with no haziness findings. Though in earliest methodology, pictures were just lined up by a unique spot, for example the edge of the nose, novel procedures launched matching considerably extra spots. "Craw and Cameron (1991)" presented an examination of images of facial landmark distortion. In another investigation, "Cootes et al. (1995)" suggested the first statistical shape model in Active Shape Models by using nearly 200 landmarks. However, in one other study, "Cootes et al. (1998)" used not only shape models but Active

Appearance Models as well so authors recommended a mixture of form and features that proved to be effective. "Hallinan et al. (1999)" in accordance with "Jones and Poggio (1998)" processed complex pixel-wise picture matches along with optic-flow techniques in order to develop the form of face alterations. Regarding the aforementioned matching methods, pictures are skewed to an ordinary pattern. The alteration of the form is afterwards completed in an identical manner like initial one Eigenfaces, considering normalization of form illustrations. In contrast, the form pattern offers a robust and dense illustration of divergences in the form by altering pixels in the shape level. Nevertheless, in contrast to humble linear projection, study of picture assignment can be converted to a further interesting nonlinear pattern matching issue. Those two-dimensional patterns had great effectiveness in catching the form adjustment for a determined gesture, as well as, a perspective of bright illumination. This context expanded further to deviations of gesture from "Vetter and Poggio (1997)". In a similar way, "Jones and Poggio (1998)" expanded further to various classes for items including illustrations of vehicles. The outcome of such a foundation is that when splitting the form as well as the texture, pictures are able to pattern facial modification. On the antipode, the cost value for taking pose and illumination fluctuations into consideration was huge. After all, it would necessitate several distinct patterns, individually restricted to a minor variety of gestures. On the other hand, improvement of three-dimensional Graphics taking place at the decade of 90s showed effortless fluctuations of gesture in modeling, involving shading. Adjusting techniques from illustrations to facial patterning and computer vision managed to a novel visualization of face in morphable models, consequently, the concept of utilizing "analysis-by synthesis" assigned amongst 3D and 2D area. These important positions introduced by Blanz and Vetter in 1999.

Three-dimensional Morphable Models as well as two-dimensional Morphable Models depend on concentrated matching, instead of just a class of characteristic spots in face. Initially, a visual stream process for picture entry created the aforementioned dependency. Procedure of picture formation applied one basic rendering template including an outlook viewing, illumination of the surroundings, as well as Phong pattern of texture reflection containing one specular factor. Nevertheless, in analysis-by-synthesis, this method occurs at a computing cost due to the ambiguities that result to a rigorously ill-posed optimization issue because of shape-camera as proposed by “Smith [2016]” and illumination-albedo “[Egger 2017]”. Furthermore, the process of optimization contains significant cost, consequently, may be led to undesirable regional targets. Additionally, there is an enormous complexity to match an Active Appearance Model to a two-dimensional picture, against the effortless viewing required for Eigenfaces. Difficulty of 3D Morphable Model’ s matching creates further obstacles which have continued to be exciting to researchers after 20 years of progress.

Taking the finest of both cases, two-dimensional as well as three-dimensional, after utilizing three -dimensional patterns to control current pictures and utilizing two-dimensional techniques to 3D textures. In contrast with techniques that focus on meshes, initially, 3D Morphable Model applied as visual stream, multi-analyze methods as well as insertion patterns of facial parameterization of textures. Plenty of methods that used in 2D, such as the original facial machine-scanner producing textures areas on a parametrized 2D cylinder, were substituted with their 3D equivalent. It is remarkable to notify that after a progress in the direction of three-dimensional, the domain of “computer vision” returned to two-dimensional statements by applying methods of “deep learning”, and currently grows over to three-dimensional, for instance, by embedding three-dimensional Morphable Models.

During the previous decades, three-dimensional Morphable Models have been employed further than facial cases. “Allen et al. (2003)” proposed patterns in order to be applied in peoples’ exterior physique, as well as to additional individual regions such as ears like “Dai et al. (2018)” mentioned or hands introduced by “Khamis et al. (2015)” or creatures proposed by “Zuffi et al. (2018)” or still vehicles as “Shelton (2000)”.

The period in which the three-dimensional Morphable Models were expanded is a result of methods and data which were hardly discovered through scientists, as well as foundations, and it was developed like one universal facial virtualization without targeting any particular mission. Nevertheless, the pattern is undoubtedly surpassed highly precise purposes like detection of faces, as a result pattern's distinction is general among various assignments as well as applications.

The crucial element in any three-dimensional Morphable model is one characteristic class of three-dimensional forms, normally combined with equivalent feature information. Most conventional approach in building one data lake is by obtaining actual globe data.

Form Acquirement

3D form obtains definitely one crucial factor in 3D Morphable Model. Form illustration is a subject with not extensively examination into three-dimensional Morphable Models' framework. One major illustration applied frequently is indisputably the meshes that form triangle. "Atick et al. (1996) introduced occasional exclusions consist of cylindrical whereas in one other investigation "Dovgard and Basri (2004)" examined orthographic depth maps even though there is no significant permitting congruence by those visualizations. " Aldrian and Smith (2012)" investigated a surface that is normal per-vertex, while "Saito et al. (2018)" concentrated in the volume of areas positioning whereas " Park et al. (2019)" tested oriented distance functions. By applying illustrations where meshes form triangle, deep matching demands the entire sample to demonstrate similar topology while vertices must encode similar semiotic spot throughout the entire sample. Forming association among the samples is an interesting topic in itself. After that, it is important to emphasize that three-dimensional data should be acquired.

Geometric techniques

The three-dimensional coordinates of a form can be approximated straight by Geometric techniques. This outcome can be achieved by detecting similar external spot by at least two perspectives where main task requires to detect the matching spots among pictures. Otherwise it can be done by spotting a viewing model where basically assignment requires to detect the matching among recognized model and a picture of its viewing. Approaches are able to be thought as active, for instance by producing illumination or further indicators to landscape such as "Time-of-flight" probes, and passive such as "multi-viewpoint photogrammetry". As active systems can be considered "Laser scanners" as well. However, "multi-viewpoint photogrammetry" might be thought both active and passive as well method. For this reason, it is characterized as hybrid. The above-mentioned characterization depends on photogrammetry for structuredness rebuilding (passive), while "Zhang et al. (2004)" introduced the idea of enhancing the subject, including advantage characterized surface viewing restoration. Regardless shaped illumination, there is no concern regarding light's source due to the fact that displayed surface requires merely intention for enhancing the applied surface area for multi-sight corresponding.

In the beginning of three-dimensional Morphable Models, three-dimensional forms were obtained, in satisfactory condition, merely by utilizing active procedures. Blanz and Vetter's initial investigation depends on a study of "Levoy et al. (2000)" including scanning with lasers. In this study, several laser-beams rasterize the facial area. Specific spot of facial texture is lighted by laser beams as well as there is a usage of recognized procedure with lasers where the spot in three-dimensional viewing could belong in a triangle. Acquirement time constitute one major obstacle for laser scanning because simply not many models are collected in a specific moment. Nonetheless, those systems necessitate the objects to lie motionless for quite a few moments (sec) when they are scanning in highly frame speed. During experiments performed by "Geng (2011)", defined illumination scanners overwhelmed the abovementioned constraint up to limited magnitude. That fact achieved after utilizing projectors which can provide enormous number of beams. At this specific moment, one major task constitutes the detection of the unique beam that lights the subject. That can be handled if we can configure the viewed illumination into a path which permits an obviously detection of rays' source. A common method constituted by binary coding. During the procedure, black-and-white patterns are showed where every pixel is produced by using binary coding. Necessary quantity of models is yet very considerable. For instance, in order to achieve VGA analysis, nineteen distinct models are required whereas if the desirable analysis is 4K then twenty-three models are required. Therefore, the aforementioned method is declared as more appropriate in catching stationary pieces. Regarding facial illumination, further complicated techniques are suggested, like "gray codes" as well as "fringe motifs". Additionally, these patterns are able to decrease necessary quantity of frames to an extent degree, as well as to a unique frame in very rare occurrences. One other very well-known commercial device experimented by "Cao et al. (2014b)" applied for producing facial data. This defined illumination is utilized by a probe of the 1st edition of Kinect. This particular tool uses a defined model of dots that permits recreating a deep background through a unique frame giving up the analysis of spatiality. According to "Newcombe et al. (2011)", the storage of various frames can enhance the quality of the picture. Due to increased imaging capabilities provided by conventional cameras, the category of passive structures is chosen in most cases, since they are easier to accumulate and operate photogrammetry software results, both commercial and open-source solutions like Agisoft or RealityCapture and Meshroom, respectively. They are able to deliver very promising findings on human faces. These systems characteristically don't necessitate a collection of data through a time-period. Therefore, according to "Beeler et al. (2010)" these provide unique-shots acquirements. Additionally, "Beeler et al. (2011)", "Bradley et al. (2010)" and "Furukawa and Ponce (2009)" underlined the capability of obtaining sufficient results in the frame of an object.

Probable drawback emerging from the above-mentioned case could be the type element, due to the fact that it is essential that a minimum division among various contributing parts is frequently described core characteristic. A different option is time-of-flight sensors which turn out to be more and more feasible due to the thrust of the mobile manufacturing. This functions when the components have a position near the others. The 2nd edition of Kinect probe has undoubtedly its place in the above-mentioned category and several probes of a deep background are carried through new cellphones. Common difficulty which "time-of-flight" probes experience along with many active structures can be paint data have to be obtained individually. This is not fundamentally associated to three-dimensional information that can be an additional benefit of passive systems.

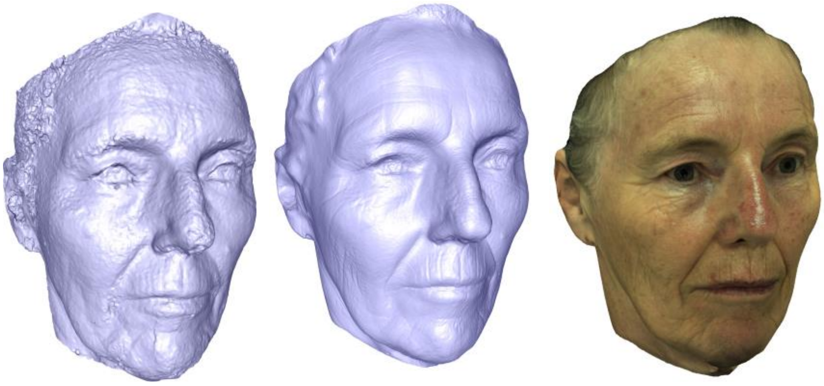
Photometric techniques

Photometric techniques usually estimate the orientation of an area, in a process that three-dimensional form is recovered through integration. A main task in this case can be the selection of patterns able to precisely attract reflectance assets while acquiring adequate dimensions and reversion of the patterns can be posed in a good manner. Comparing geometric and photometric techniques, it is proved that photometric techniques are able to provide better details in form as long as they are independent of the existence solid characteristics. Thus, they can be valid to surfaces which are slick and without many characteristics. However, they are frequently experiencing low level of bias into remodeled spots produced from pattern inaccuracies in reflection and brightness.

According to "Ackermann et al. (2015)", photometric stereo assesses the area as ordinary one at every pixel in detecting a pose of a stable view through 3 various brightness circumstances at the minimum level. This situation is "spectrally multiplexed" according to "Hernández et al. (2007)" in a process of decreasing total necessary poses. "Georghiades (2003)" revealed that universal face priors can be applied to determine the ensuing uncertainty, in case that illumination is uncalibrated, and there is a usage of a more appropriate shiny reflectance model. Naturally, the brighter an element is, the better results in its robustness and coverage can be achieved. These brightening conditions should be at least 4 according to "Zafeiriou et al. (2013)" and 9 according to "Gotardo et al. (2015)". The role of gradient in the illuminating procedure is important as long as a specific spotlight can affect the transmission of the light in an omnidirectional way. The interferences from tough light sources can be substituted by soft partial obstructions. This is the biggest benefit obtained from this procedure. Practically, the omnidirectional lightening can be achieved by a level of brightness, according to "Debevec et al. (2000)". With regard to an early investigation from "Ma et al. (2007)", 4 basic gradients are mainly used whereas many additional ones arise in continuation. In order to lessen the number of necessary requirements, modifications of temporal, spectral and polarization multiplexing have been suggested.

Hybrid Techniques

Hybrid techniques constitute a combination of geometric as well as photometric approaches. In details, these techniques decrease the low level of bias which usually appear in photometric processes. Consequently, hybrid techniques rise the details in an augmented level in comparison with geometric processes. In another investigation, "Nehab et al. (2005)" offer the procedure of combining the low level of spot data on the one hand and the augmented level of normal areas. This procedure constitutes the most effective one, including a unique outcome of a scarce system of equations characterized by linearity. This framework of three-dimensional Morphable Models is applied in the theory of "Patel and Smith (2012)". Several other mixtures of both geometric and photometric procedures were the subject of research such as "Zivanov et al. (2009)" who merge defined illumination through “photometric stereo”. In one other investigation, “Ma et al. (2007)” bring together defined illumination through basic gradient lighting. According to “Ghosh et al. (2011)”, authors merge multi-viewpoint stereo with basic gradient lighting. The Figure 1 below illustrates the result of a hybrid procedure where a combination of photometric normal regions and a multi-viewpoint stereo mesh is generated.


[1]

[1] Figure 1: (a) MVS geometry. Multi-viewpoint stereo (MVS) is utilized for rebuilding a thick mesh, (b) Hybrid geometry. A combination of region normals and the MVS mesh for creating a mesh through fine region details, (c) Rendering. In combination, those are employed for generating a synthesis of extremely pragmatic facial depictions.

Appearance Capture

Like the shape, appearance is needed as well for a lot of 3DMM assignments, like synthesizing pictures and inverse rendering. Dissimilar with shapes, which are nearly entirely signified as triangular meshes, appearance representation differs significantly. Theoretically, each vertex of the mesh might have a related appearance attribute but shapes, normally, are parameterized in two-dimensional area as well as the surface can be applicable to retain exterior attributes. Self-occlusion leads in missing data in the occluded domains, particularly when just a specific viewpoint is accessible. In the investigation of "Booth et al. (2018b)" the usage of three-dimensional Morphable Models corresponds to pictures of wildness and the method of "Principal Component Pursuit" without any values to fulfill an undetected surface. The researchers develop a pattern of display precisely into tested surfaces. However, that is one simple-minded methodology and prohibits intrinsic facial display attributes from being removed by illumination. An individual outcome arising from this kind of difficulty can be the management of lighting circumstances. An example can be the usage of various causes of illumination so as to generate ambient illumination.

Several methods have been introduced throughout the decades in a process of obtaining information of reflection suitable to parametric rendering and calculating region reflection like "Marschner et al. (1999)". A polarized circular lighting interface allows widespread albedo being depicted into one particular snapshot as well as mirror like albedo displayed in 2 views [Ma et al. 2007], as illustrated in the figurer below. Although these methods have mainly applied active settings, newly, Gotardo et al. [2018] have shown capture under passive requirements.

(a) Diffuse albedo


[2]

(b) Specular albedo

Face part precise methods

Several facial areas necessitate further aimed acquirement approaches as well as tools because there is no fitting in assumptions produced by the aforementioned methods. For instance, Bérard et al. [2014] exploit a merge of some particular processes, involving “form-from-specularity”, so as to recreate every observable part in eyes’ area. One other difficult instance constitutes human teeth’s area which shows particularly tough appearance. In addition to teeth, hair defies the ordinary hypothesis that the remodeled structure can be a silky uninterrupted region whereas necessitates dedicated methodologies which approximate fibers of the hair’s area, according to “Beeler et al. (2012)” or even encrypt hair like a region (as

[2] A photometric illumination phase “[Ma et al. 2007]” can be applied in obtaining widespread and specular albedo charts.

mentioned by “Echevarria et al. 2014”) into manufacturing. Although nearly all fiber acquirement aims into stationary recreation, many researchers catch fibers in movement as tested by “Xu et al. (2014)”. Particularly tough task is the acquisition of partially or entirely unknown attributes, like tongue’s area as imposed by “Hewer et al. (2018)”, skull’s area according to “Achenbach et al. (2018)”, “Beeler and Bradley (2014)”, or jaws’ area as introduced by “Zoss et al. (2019, 2018)”. For these occasions, expert imaging methods are needed, like Computer Tomography. Finally, skin’s region necessitates as well dedicated remedy in some domains, for instance lips’ area as imposed by “Garrido et al. (2016b)” or eyelids’ area according to “Bermano et al. (2015)”.

Dynamic capture

Generally, three-dimensional Morphable Models are mainly involved in stationary forms, either in some neutral forms by distinct persons or along including some human expressions, disregarding the facial rotation among human gestures. The majority of “dynamic capture methods” that develop three-dimensional Morphable Models could be characterized as stationary, concentrated in gaining specific forms instead of maximum acts. The time when area commences to consolidate timeless data in some forms, the demand regarding “dynamic capture methods” can increase. Active structures are judged, as geometric according to “Zhang et al. (2004)” as well as photometric in accordance with “Wilson et al. (2010)”. Nevertheless, passive approaches are the tools of selection at present, as long as these have no requirement of multiplexing as well as are able to produce forms of great value along with information of reflection as impose by “Gotardo et al. (2018)”. One positive impact emerging from these tools can be their frequent offer for forms which have already been in assignment, the elimination of the need to determine assignment in a post-managing stage and creating them appealing solutions.

II. LITERATURE REVIEW

Photography has been around for many years and the developments in digital media over the last 20 years have made digital data easily available and cheap. Therefore, it is not surprising that there has been so much research on face analysis using 2D. However, there are limiting factors with 2D data including lighting conditions, pose variations and handling expressions. These lead to major changes in the image such that matching becomes extremely difficult.

Because of the nature of 2D data it is understandable that most techniques concentrate on localizing landmarks which are mainly distinct by their texture. For example, landmarks around the eyes have been studied and localized successfully in 2D. In particular, Gabor filters are approved to be useful in modeling the surface variability around eyes' region. Other landmarks such as the mouth corners and the nostrils have also been detected with relative success. However, the nose tip which has a distinctive shape has not been the object of any studies in 2D. By using only texture cues a handful of landmarks can be detected from images. However, when this is combined with a location model a larger number of landmark locations can be estimated. The extra landmarks do not necessarily show strong texture characteristics but by constraining their relative locations, localizing can be successful. That can be the issue when utilizing ASM's as well as AAM's. Pose and lighting conditions continue to appear in the literature as unsolved problems. The performance of the systems deteriorates rapidly when in-plane rotations are applied. Several approaches have been attempted to compensate for lighting and pose variation. However, the results show that performance is reduced when using images taken under non-controlled conditions which suffer from the pose and lighting issues.

According to "Markus A. de Jong et al. (2016)", authors announced a new method for automatic three-dimensional facial landmarked by making a usage of two-dimensional Gabor wavelets. This methodology reflects face as a region and manages outline ledges in obtaining two-dimensional characteristics of raw data. The characteristics that have been obtained consist of texture, relief map, and their transformations. For this reason, authors improve an already fixed 2D landmarking approach for instantaneous assessment from those data. The aforementioned approach is proven by executing landmarking trials on two data sets by applying 21 landmarks and finally it compares the results with the implementation an active shape model. The error that produced from this method was 1.9 mm, on average, in contrast with the implementation of active shape model which resulted an error in landmarks about 2.3 mm. Another investigation examines the form of face heredity in relatives and concludes that automatic landmarking gives the same results as manual landmarking for several landmarks. Authors mentioned that this approach can be accomplished in 30 min for any dataset that constitutes any volume while permits both high-speed and vigorous three-dimensional facial landmarks.

More specifically, Gabor-wavelet’s methodology is analyzed in good way while contains a benefit in working properly through a small number of training examples in contrast, for instance, through “Active Shape Models” (ASMs) where necessitate extremely high number from data in order to be trained into precise entry. Another essential issue can be an enhanced abundance of three-dimensional data in comparison of those of two-dimensional where have been manipulated for inserting three-dimensional updates in the entry model with common approach, for instance by describing it as 2D data. Modifying landmarking is necessary for the evaluation of numerous three-dimensional data modules regarding the effect on entry precision where the elasticity's evaluation of this methodology is defining. In addition, authors assess the performance regarding heritability and contrast the suggested algorithm to an ASM technique.

As first action of this methodology can be an extraction from a “Region of Interest (ROI)” from front part of face. Then, three-dimensional dataset is modified into a 3D relief map by one display facial viewing. Afterwards, one two-dimensional image is produced from a 3D relief map. In addition, the above picture can be submitted in two-dimensional landmarking technique by “Elastic Bunch Graph Matching (EBGM)” procedure. At the end, by reversing the projection, the entered two-dimensional landmarks are displayed backwards to three-dimensional.

Same author tried to incorporate three-dimensional facial landmarks of region scans in the research of “Markus A. de Jong et al. (2018)”. Three-dimensional facial landmarks of region scans stand as essential assessment stage for medical treatments like “Genome-wide Association Study” (GWAS). Landmarks that annotated manually can be frequently used through significant expense as well as dependent volatility. Therefore, landmarks that produced automatically through negligible training are required. The authors employ statistical ensemble techniques for enhancing three-dimensional facial landmarks of region scans. Basis procedures for landmarks applying structures obtained by three-dimensional area scans can be mixed by applying both by bagging and stacking. They emphasize train with the minimal sophistication of maximum forty tests for train via model that resulted from procedures of landmarks and are effective to these functions. Furthermore, they make a correlation relationship among landmark coordinates establishing an exploration scheme directed from “Principal Components” (PCs) from landmarks that trained. The outcome of bagging contains invaluable effect, whereas stacking clearly enhances error’s accuracy at 1.7 mm throughout every twenty-one spots which leads to an upgrading of 22% in contrast to an earlier procedure. Evaluating heritability in pairs, demonstrate enhancements as well when applying landmarking distance of faces. With regard to ensemble processes, the progress is completed automatically while precise three-dimensional facial landmarks of pictures through marginal testing can be beneficial to significant group of researches such as GWAS. In condition that landmarking requires modification or more information, properties fluctuate.

Several landmarking approaches until now as “Milborrow, S. & Nicolls, F (2014])”, “de Jong M. A. et al. (2016)”, “Wiskott, L. et al (1997)” revealed encouraging outcomes for landmarking algorithm accuracy, a few deeply reliant on heuristics such as “Guo, J., Mei, X. and Tang, K. (2013)” as well as “Liang, S. (2013)”. Even now, there is no certainty if powers of particular procedures can be harmonized. For instance, if powers are merged for producing so far further precise information for landmarks. According to prior investigation, it has been demonstrated that various data transformations produce further information accessible to basic wavelet-based techniques. Nevertheless, authors mentioned two main disadvantages that they overcome with the current survey. First of all, in prior method, the performance is inadequately in regional landmarking containing small fundamental data like forehead’s and chin’s area. In addition, optimality was not adequate due to the selection of modifications that were used as an input for this algorithm consequently the methodology wasn’t systematically accepted at the maximum level. To confront this first issue, there was a notice of allocation in landmarks-spots among trained sample that might offer further data regarding barely fundamental landmarking figures. Applying PCA in the area of landmarks, the possibility to exploit this kind of information is provided, as employed by (ASMs) from “Cootes, T. F. et al. (1995}”. The second issue presents a challenge with regards to the model choice that takes for granted carefully chosen figures and modifications upon the different input information.

The landmarking algorithm merges an amount of processes of landmarks in a single whole. As base landmarks procedure is considered any procedure which is able to offer a landmark spot of new input data. The prediction of the concluding landmarks according to the above-mentioned procedure is achieved by the averages or regression. In the aforementioned execution, every fundamental algorithm is pattern-based. One slight amount of these pictures can be labeled by hand according to the patterns obtained at the stage of the training. In the phase of preliminary process, two-dimensional views of three-dimensional area information are received. Three-dimensional data are preserved inside a heightmap which relates points into two-dimensional surface, creating a conversed procedure “one to one”. An amount of characteristics is then produced through the merged two-dimensional information which operates like a source for this kind of procedures. The entire base processes can be established as per the “Gabor-wavelet” reactions. Many processes aim distinct characteristics while they operate similarly with “EBGM” process through regional exploration approaches. A methodology of large-scale exploration approach centered on “PCs” can be included in total. Landmark examination concerning base procedures can be prepared in the average of sample’s amount.

According to Xin Fan et. al. [2016], 3D facial orientation usually necessitates the precise approximation of facial landmarks. Plenty of current landmark recognition techniques make use of geometry classification while in other cases they apply regression techniques executed to vary pictures. One technique merging three-dimensional geometry knowledge as well as two-dimensional surface is hardly ever examined. During the aforementioned investigation, authors suggest one unique three-dimensional landmarking of face tracing process, established to conformal geometric plotting which is able to turn one three-dimensional model into two-dimensional by utilizing mutually geometry as well as surface knowledge. Subsequently, procedure that applies two layers regression, it can enhance constancy from landmark detection into two-dimensional geometry pictures. The aforementioned procedure is impermeable to present modifications and vigorous with regard to alterations in expression. The evaluation of the proposed method is based on widely accessible data while validate in what way usage from two-dimensional regression techniques improves strength as well as three-dimensional accuracy of face landmarking location.

As a fundamental step can be considered the “Facial landmark detection” (FLD), as “Z.-H. Feng et al. (2015)” mentioned, in lots of automatic face analysis schemes in order to create connection among three-dimensional faces through diverse angles, posing as well as emotions. “FLD” systems is separated to couple parties: procedures through characteristic recognition while the other one focus on statistical “Point Distribution Model (PDM)” as examined by “M. Song et al. (2014)”.

Regarding procedures that focus on the recognition of characteristics, regional curvatures like "Gaussian curvatures" as well as form indicator, as imposed by "P.J Besl et al. (1988)" and "C. Dorai et al. (1997)" respectively, are initially applied to identify considerable features ("H. Dibeklioglu et al. (2008)" and "X. Lu (2006)"). However, plenty techniques are able to identify just some consistent vertices including harsh curvature fluctuations from information of face like nose and chin as well as eye and mouth angles. Nevertheless, several appliances use similar facial animation as well as parameterization. The above-mentioned landmarking spots are placed on the forehead as well as on cheeks and many more particular spots. Therefore, a few other curvature-based techniques have been recommended as well to identify more landmarks propose by "S. Berretti et al. (2011)" with the accordance of "S. Jahanbin et al. (2011)". However, these methodologies necessitate three-dimensional patterns that are unevenly lined up to a front side picture as well as every spin of three-dimensional patterns is able to disturb face landmarking recognition outcomes.

Likewise, techniques focus on "PDM" are able to identify also many landmarks, for instance "ASM" as well as "Active Appearance Model" (AAM), according to "G.J. Edwards et al. (2008)". Nevertheless, on condition that three-dimensional information of face isn't lined up to an equivalent path, outcomes of aforementioned landmarks are able to imprecise as examined by "C. Qu et al. (2014)". Consequently, certain procedures betake to training the data from various observations to enhance the strength in subjective postures. Additionally, many current techniques apply just only two-dimensional and three-dimensional data in landmarking recognition. They ignore particular knowledge of characteristics on another facial angle that can produce many precise outcomes. On the other hand, some procedures take equally 2D and 3D information into account such as "Perakis et al. (2014)", whereas three-dimensional geometry data as well as two-dimensional surface can be handled individually while simply findings can be merged, so matching among three-dimensional geometry and two-dimensional surface can be gone.

III. MODELING

The concept of this chapter is the modeling of 3DMM regarding the fluctuations of digitized 3D human faces. More specifically, the most commonly forms of variations are the following three. First of all, geometric variations throughout dissimilar identities that depicted in a shape model. Frequently applied models consist of global models, which signify fluctuations of the whole facial area as well as regional patterns that characterize fluctuations from sections of the face. Secondly, geometric fluctuations among various emotions of face can be depicted into a pattern of expression. Regularly employed patterns are primarily categorized to additive as well as multiplicative patterns. Newly, expression patterns that does not contain linearity, began to be investigated. Thirdly, fluctuation into appearance as well as lightening ca be depicted into one distinct appearance pattern. A noteworthy fact is that the landmark investigation on 3DMMs, according to "Blanz and Vetter (1999)", recommended 1st patterns including 3 kinds of fluctuation which can generally be applied even nowadays.

In order to figure form, emotions, as well as appearance patterns, statistically tests can be executed throughout a facial information of a databank. Usually three-dimensional facial scanners can be employed while further newly methods realize facial patterns straightly derived by two-dimensional pictures as well. The aforementioned calculation of statistics necessitates correspondence information. This information is structurally corresponding sections of the faces necessary to be compared, and consequently recognized as explicitly or implicitly. The most well-known method refers to calculate matching information explicitly ahead of processing three-dimensional Morphable Models. Various latest procedures calculate matching data instantly while a three-dimensional Morphable Model can be created. 3DMMs are reproductive models and a key feature is the ability to compose novel faces.

Shape models

This chapter refers to modeling geometric variation throughout various subjects processed by applying classical modelling methods that make use of 3D data. In order to make use of three-dimensional images like samples-train, it is necessary to measure a distance among every couple of images while calculating one space among raw images containing many non-structured vertices. Frequently, the procedure contains a pre-processing step of the dataset by distorting a pattern mesh in every image. This mesh creates anatomical matching among images' spots. The region for the previous treated mesh is denoted as an "S". As "i-th" vertex from the area "S" can be represented as $v_i \in R^3$, and its correlated vector $c \in R^{3n}$ includes the coordinates of v_i through constant order. Every mesh is a part of similar triangulation. As "i-th" triangle ca be signified like $t_i = (t_i^1, t_i^2, t_i^3) \in \{1,\ldots,n\}^3$,where t_i^1 , t_i^2 , t_i^3 offer indices to related vertices $v_{t_i^1}$, $v_{t_i^2}$, $v_{t_i^3}$ and the complete triangulation is known by $T = (t_1, \ldots, t_m)$. The distances between shapes S_1 and S_2 are calculated the same as difference between c_1 and c_2 when strictly affiliating S_1 and S_2 in R^3.

According to "Dryden and Mardia (2002)", 3DMMs usually go along with the characterized shape "S" while including geometrical data persisting following deleted variations that produced from interpretation, spin as well as an evenly scale. Though scaling cannot be normally eradicated for people's faces, which can be frequently performed for geometrical morphometrics. As one "shape-space" can be defined a set including every configuration of "*n*" vertex defined into R^3 along through determined connectivity. Due to the fact that the main interest is to model people's faces simply by using three-dimensional Morphable Models, "shape-space" expression belongs into a boundary space, containing d-dimensions, which signifies feasible three-dimensional people's face, while $d << n$. As a result, for every single 3D face there is a related parameter vector $w \in R^d$.

Regarding three-dimensional Morphable Models, statistical study for forming can be employed like a generative pattern. Therefore, "shape-space" obtains a related probability distribution which can be known as "prior" while can be characterized from density-function "f(w)". This probability distribution calculates likelihood where a pragmatic facial three-dimensional is signified from one certain vector "w" inside "shape-space". In conjunction through small mishandling from notation, the interpretation of "c" can be a creator function as bellow:

$$c : R^d \rightarrow R^{3n} \quad (1)$$

This expression determines small level of dimensions in vector "w" into a vector from every vertex coordinate "$c(w) \in R^{3n}$". The use of $v_i(w) \in R^3$ is referred to the i-th mesh's vertex offered from "w". Whilst pattern's analysis- which consist of the total vertices- can be normally determined, gradual mesh display centered into edge breakdown oversimplification for the creator function was examined by "Patel and Smith (2011)".

That component reflects an issue in which every face from the trained information contains an almost identical (usually neutral) expression. Considering the study of "Brunton et al. (2014b)", the interest discerns global models which shape a whole facial area as well as head region derived of regional patterns where they can present statistics throughout established parts.

Global Patterns

As $\{S_i\}_i$ signify trained forms whereas $\{c_i\}_i$ signify for the related coordinate vectors. An original investigation of "Blanz and Vetter (1999)" on 3DMMs suggested one worldwide form pattern which utilizes "PCA" for the following calculating linear creator function:

$$c(\mathbf{w}) = \bar{c} + \mathrm{E}\,\mathbf{w}, \quad (2)$$

while $\bar{c}$ can be mean calculated across trained figures, "**E**" $\in R^{3nXd}$ defines a matrix which includes "d" highly leading eigenvectors including in covariance matrix calculated throughout form differences $\{c_i - \bar{c}\}_i$ as well as "**w**" implies not high dimension's form of parameter vector. A main assumption in the aforementioned pattern can be trained images which are able in linearity and the can be introduced for producing new three-dimensional poses. One other assumption of three-dimensional faces is the pursuing of "Multivariate Normal Distribution" that is straightforwardly inferred of eigenvalues relating with **E**. The above-mentioned fact suggests density function f(**w**) to be basically Mahalanobis distance of **w** to its origin by estimating the likelihood from parametric display **w** into form area.

Three-dimensional Morphable Models was initially calculated throughout 200 objects while was demonstrated in being valuable through range of purposes. This can be achieved due to the ability in producing reasonable forms as well as the straightforward fundamental pattern. One latest survey reconstructs this type of pattern of huge dataset including 9,663 three-dimensional images while repeats most excellent procedures of "Booth et al. (2016)", proving an initially recommended creator function in forming while staying on extremely related study for scientific field.

A main remark of "Blanz and Vetter (1999)" can be a shifting depiction of vector **w** elsewhere of mean that could lead to facial enhancement of individuality resulting finally to distortions in uniqueness. For shaping distinguishing characteristics of face, "Patel and Smith (2016)" recommend one alternate density function, according to subsequent examination. Let examine Mahalanobis distances in square by mean into a class by vectors with d-dimensions and go by "Multivariate Gaussian-Distribution". The above-mentioned spaces shape one χ_d^2-distribution, that contains anticipated quantity of d. Therefore, for maintaining form uniqueness associated with individuality, "Patel and Smith" limit display $\mathbf{w}$ applying Mahalanobis distance $\sqrt{d}$ by mean. According to "Lewis et al. (2014)", researchers recommend an analogous claim demonstrating that while faces follow Gaussian distribution, where is demonstrated in "Basel" information by the "Kolmogorov Smirnoff "test in shaping in which borderline distribution of form can be near into a Gaussian as imposed by "Egger et al. (2016b)", techniques which create hypothesis of regular poses situated close to mean contain no validity.

Newly, according to "Lüthi et al. (2018)", authors recommended one shape-space without linearity which creates distortions by mean likewise Gaussian procedures.

Appearance models

In this segment, there is a description of methods for shaping the facial appearance, in which it is occurred a discrimination concerning linear and nonlinear models. Albedo and illumination affect the appearance of a face. Nevertheless, the majority of three-dimensional Morphable Models don't entirely split those issues, as a result several times brightness can be burned in albedo. Therefore, this is an issue statistically obtaining the data appearance shaping. A very ordinary method in developing a pattern of appearance can be in generating statistics on appearance information by trained forms, in which the above-mentioned data can be generally signified on conditions of per-vertex rates as well as surface into *ultraviolet*-space.

Linear per-vertex shapes

Normally, color figures can be patterned in subspace of small level of dimensions which clarifies color fluctuations. The aforementioned fact indicates in a similar pattern into a shape pattern of linearity as it follows:

$$\mathrm{d}(\mathbf{w^t}) = \bar{d} + E^t \mathbf{w^t}, (3)$$

where, $\bar{d}$ and E^t allocates an equal quantity of rows like $\bar{c}$ as well as E whereas $\mathbf{w}^t$ stands as a parameter vector from a surface of small level of dimensions.

According to "Booth et al. (2017)" as well as "Booth et al. (2018b)", in order to learn a per-vertex appearance model from views, authors introduced a convex matrix factorization formulation centered on back-projection, on condition three-dimensional geometry of the face in the view can be identified. This particularly appearance model is created by applying features calculated from the images but not by employing directly the color images. The benefits of this procedure are that the features might be unchanged to illumination adjustments and that they rely on a regional area as well which could expand the range of convergence. In an analogous way, "Wang et al. (2008)" create one pattern that contains linearity based on sphere-shaped harmonious foundations. The aforementioned mutually patterns' surface and fine-scale shape, like appearance by brightness, are able to in being produced like a function by basis with linearity.

Linearity in texture-space patterns

One main drawback by per-vertex patterns can be the compatibility regarding the resolution among form as well as appearance virtualization. That can be quite unusual because normally small level of analysis in geometrical pattern is operated in combination with a high resolution 2D texture map. Some other advantages of working with a 2D texture are the likelihood of applying image processing methods to change the texture maps. Considering this fact, the aforementioned representation is receptive as well can be handled of “Convolutional Neural Networks (CNNs)”.

Regarding the built of linear appearance models based in texture space, Blanz and Vetter [1999] in their investigation, originally utilized depictions focus on texture while demonstrating faces cylindrically. In subsequent studies, depictions focus on texture have been utilized in augmented details texture such as wrinkles according to “Pascal (2010)”, otherwise as imposed by “Pierrard (2008)” in intersecting skin as well as identifying moles. As imposed by “Cosker et al. (2011)”, authors create fluctuation in appearance through ultraviolet area centered on series of pictures of faces captured by various projections. The observations of dynamic series can be in line up and they are built on one process without solidness, therefore color fluctuation is able in being patterned by utilizing a linear subspace template based on PCA. In one other study, “Dai et al. (2017)” make usage of an ultraviolet area appearance display as well and can be specified in whole head. According to “Huber et al. (2016)”, researchers utilize per-vertex appearance fluctuation pattern thorough PCA. However, they determine a familiar uv-mapping as well, therefore the model is able in being textured established on provided pictures. “Moschoglou et al. (2018)” configure one vigorous matrix factorization issue with the purpose to understand assigned ultraviolet displays of faces through one compilation of trained textures. An investigation into an impact by various ultraviolet area sticks in texture which have been introduced according to “Booth and Zafeiriou (2014)”.

Nonlinear models

Conventionally, appearance of faces can be shaped like one subspace with linearity, which most of the times a Gaussian distribution is assumed. Nevertheless, “Egger et al. (2016a)” has demonstrated empirically that the Gaussian hypothesis cannot be incredibly precise as well as might result into a partial optimal pattern.

Therefore, researchers recommended in substituting this pattern that follows a Principal Component Analysis by a “Copula Component Analysis” as introduced by “Han and Liu (2012)”. Consequently, the main concept has been expanded in communally pattern of form faces, surface as well as characteristics by “Egger et al. (2016b)”. In the study of “Alotaibi and Smith (2017)”, authors make usage of remark where the paint of skin shapes a manifold without linearity into RGB interval, nearly covered with paint of pigment’s melanin as well as hemoglobin. Authors flip side render displays from those parameters, subsequently , in the parameter space, they create a linear statistical model. The result of biophysical three-dimensional Morphable Model is ensured to generate reasonable skin paints. Further to worldwide appearance of face patterns, methods can be existed assuming patterns of regional skin fluctuations as well. In a particular paradigm, “Dessein et al. (2015)” apply one texture pattern that was built with slight overlying patches which can be removed by facial databases. In addition to this fact, “Schneider et al. (2018)” introduced one stochastic model which can be capable in synthesizing freckles.

Lately, various methodologies in appearance models focus into deep learning are recommended, in which plenty of these approaches are created as well in an analysis-by-synthesis framework. Briefly the major notions of this framework are the growth of deep learning techniques enabled to realize per-vertex appearance patterns right away of pictures, according to “Tewari et al. (2018)”. Author realize per-vertex albedo pattern displacements as a result in improving generalization capability from a current pattern that uses Principal Component Analysis. Likewise, “Tewari et al. (2019)” discover one per-vertex albedo template according to video information. “Zhou et al. (2019)” test one decoder of meshes which mutually forms surface as well as structure into a per-vertex basis. Nevertheless, this model depends on the accessibility from three-dimensional form as well as appearance information. Numerous deep learning methods can be existed which as well examine an appearance pattern focus on surface. Devoid of necessity from three-dimensional figures, “Tran and Liu (2018a)” find out an appearance of faces pattern, without linearity, signified in ultraviolet area built into Convolutional Neural Networks, but it doesn't clearly examine illumination. On a later study, researchers examined one further detailed pattern in which albedo as well as illumination are individually shaped as imposed by "Tran et al. (2019)" and "Tran and Liu (2018b)".

Joint form and appearance models

According to "Blanz and Vetter (1999)", they initially suggested creating split, separate models for shape and texture. It is of great interest, two-dimensional "Active Appearance Model" (AAM), as imposed by "Cootes et al. (1998)", has been initially recommended through mixed form as well as appearance pattern. A main benefit from this mutual pattern can be correlations among form and texture which are able to in being studied and developed like one constraint throughout matching through less parameters. In antipode, individual patterns can be adaptable. Subsequently, shape as well as texture parameters may being adjustable individually, and consecutive processes are capable of fitting the two models separately. Nevertheless, 3DMMs that equally form shape and texture are consequently judged. A "Schumacher and Blanz (2015)" introduced, they apply **a** canonic correlation analysis in researching form as well as surface correlations in addition to correlations among facial parts as well. According to "Egger et al. (2016a)", authors utilize CCA which is able to examine various ranges from form as well as texture information. In another research, "Zhou et al. (2019)" suggest one deep convolutional painted mesh autoencoder which realizes commonly patterns without linearity from form and texture.

Correspondence

An aforementioned examined pattern usually necessitates information through "point to point" correspondence among every form. Registration is a method referred in forming like dense correspondence among images.

Plenty methods occur to confirm “point to point” correspondence into typical categories by subjects, still the facial area of deformities can be firmly limited. Frequently applied facial entry techniques stick to the notion by disfiguring one pattern mesh in every image inside datasets. The above-mentioned entry method normally begins through hard lining up (frequently utilizing thin correspondence) as well as directs into dense correspondences at final stage.

Sparse correspondence calculation

A number of techniques occur to create a thin correspondence meant for a set from three-dimensional images in forecasting landmarks, such as one universal class of prominent spots, for every image. The aforementioned sparse correspondence subsequently operates like an automated initializing technique in dense correspondence approaches.

Many procedures utilize regional descriptors even a them as well as connectivity data among descriptors into forecast obvious spots. Although localized landmarks through scans can be extensively studied, as imposed by “Bulat and Tzimiropoulos (2017)”, the main emphasis can be in techniques which determine sparse correspondence among three-dimensional pictures.

Current procedures apply patterns from various geometrical descriptors. According to “Passalis et al. (2011)”, authors make usage of form index as well as rotation picture characteristics. “Berretti et al. (2011)” apply convexity in addition to “Scale-Invariant Feature Transform (SIFT)” attributes. In addition, “Creusot et al. (2013)” examine mixtures in regional attributes like Gaussian curvature as well as mean. Authors discover how the abovementioned descriptors distributed statistically through every landmark.

Moreover, current techniques apply geometrical interactions or interactions among landmarks as well as geometrical characteristic descriptors. According to

“Guo et al. (2013)”, researchers design a scan into a picture while they forecast landmarking through two-dimensional Principal Component Analysis template as well as geometrical interactions by extra surface restrictions. A study of “Salazar et al. (2014)”, in the same way as “Creusot et al. (2013)”, discover how the abovementioned descriptors distributed statistically of regional texture through supplementary “Markov network” into further examination relationships among landmarks. In another research, “Bolkart and Wuhrer (2015a)” broaden the above-mentioned study beyond into sequences while examining in addition chronological edges inside Markov’s grid.

Dense Correspondence calculation

Processes able to distort the pattern so as to create matching mainly vary in parameterization of the deformity. The current methodologies are classified based on the nature of scanning information that they usually capture. The discrimination is among static procedures, such as techniques which target in examining static three-dimensional scanning and dynamic procedures able to capture three-dimensional sequence of movements. “Blanz and Vetter (1999)” utilized the “bootstrapping methodology” to repeatedly match a three-dimensional Morphable Model. In their framework, they followed the process of scanning and refining the matching, drawing the pattern and scanning with a flow field and refining the model. “Blanz et al. (2003)” afterwards expanded the abovementioned methodology introducing the scanning of expression. “Amberg et al. (2008)” examined the scanning of expression using bent ICP[3], whereas “Hutton et al.(2001)” produced a “thin plate spline” widely known as TPS mapmaking to distort every scanned surface, as well as introduced matching by close area’s examination.

[3] ICP refers to “Iterative Closest Point” which is a process for minimization of difference among two clouds of spots. It is usually applied for reconstruction of two-dimensional to three-dimensional surfaces of various scans.

"Passalis et al. (2011)" captured scanning from distorting a commented facial model. As studied by "Kakadiaris et al. (2005)", ordinary three-dimensional facial patterns can be fragmented in various commended regions through resolving a differential equation of second order. In one other research, "Mpiperis et al. (2008)" primarily matched a subdivision surface to a scan given that the distortion of the basis mesh (such as the grid of the lowest possible level of subdivision) is driven from a spare landmark's matching. Since the set of training is entered, authors configure the deformity of the base grid with a PCA template over the training information. "Salazar et al. (2014)" utilized a common term of a mixture shape form for matching the expression of scanning, supported from bent ICP to diligently match the scanning surface. "Gerig et al. (2018)" created solid association of the Gaussian method disfigurement form with a regionally ranging core.

Numerous techniques occur to successively register sequences of motions. In the study of "Weise et al. (2009)", they applied a uniqueness Principal Component Analysis model for entering a neutral scan and thus monitoring the sequences of motions through enhancing thin and dense visual flow between successive frameworks. "Fang et al. (2012)" and "Li et al. (2017)" prepared the amelioration through registration of the prior frame for the development of temporal information. "Fang et al. (2012)" employed an AFM[4], whereas "Li et al. (2017) employed a bent ICP normalized by FLAME[5]. "Fernández Abrevaya et al. (2018)" utilized a "spatiotemporal technique" to capture the whole sequences of motions through processing the registration of whole sequences and clearly codifying temporal knowledge with a "Discrete Cosine Transform".

Additional customization without template was established on various procedures of registration. "Sun et al. (2010)" used a compact mapping to parameterize both meshes and create thick matching among planar planes causing

---

[4] AFM stands for "Atomic Force Microscope".

[5] FLAME: a program offered by Autodesk for machine learning purposes especially in computer vision and photogrammetry.

thin landmark association. “Ferrari et al. (2015)” intersected facial scanning to sections without overlapping, fractionated by geodesic curves among chosen landmark’s parts as well as constantly re-test every point.

Jointly solving for correspondence and statistical model

According to “Li et al. (2013)” and “Bouaziz et al. (2013)” mutually renovated human-particular blendshape patterns, whereas they followed sequences of motions into an actual-time methodology for the face’s simulation. In the context of their research, authors applied the pattern of “principal component analysis” which can conflate the human particular blendshapes and extra remedial base carriers. “Bouaziz et al. (2013)” contributed to the modification of remedial deformity parts.

On the other hand, “Bolkart and Wuhrer (2015b)” and “Zhang et al. (2016b)” improved the relation in a set of information of various individuals in manifold expression following a group manner. “Bolkart and Wuhrer (2015b)” mutually reconsidered point-matching within the eye surface, paring an objective function down which is able to calculate the stability of a multi-line facial pattern. “Zhang et al. (2016b)” ameliorated maps of functionality throughout a whole set of information.

Synthesis of novel model instances

Three-dimensional Morphable Models are able to create three-dimensional facial surfaces which are realistic and dissimilar from the already detected information upon trainings. The synthesis is possible to be accomplished when modifying some coefficients in parameter space like the form. Ordinary procedures involve interpolation or extrapolation among coefficients of the samples. Additionally, this kind of model is able to be applied to straightly produce original three-dimensional facial surfaces through extracting random samples in parameter space based on this previous distribution. By the model term, this sampling-process permits the synthesis and alteration of form, expression, or outlook of a static three-dimensional facial surface.

Synthesis's results have been strongly applied in recreational domains. Thus, the formation of static three-dimensional facial surfaces particularly involves created facial caricatures when relocating the identity coefficient linearly from the mean according to "Blanz and Vetter (1999)".

As long as three-dimensional Morphable Models predetermine and dissociate expressional material, it's quite simple to produce dynamic sequences through determining some coefficients of identity when altering the coefficients of countenance. Many researches try to compose dynamic three-dimensional facial videos of a stable identity incorporating three dimensional Morphable Models. They consist of operations composing four-dimensional videos from a static three-dimensional mesh combined with semantic label data according to "Bolkart and Wuhrer (2015a)", and from a static three-dimensional mesh and audio data according to "Cudeiro et al. (2019)".

IV. IMAGE FORMATION

A three-dimensional Morphable Model offers the virtualization in facial geometry and appearance based on parameters. A crucial procedure for a model like the abovementioned is synthesis. Synthesis contains two main steps. Firstly, the process of creating one model example through sampling by parameter's area and alternatively interacting with the parameters by hand. Secondly, the output of the produced model to a 2-dimensional picture through simulating the picture's formation, a similar procedure such as a graphics-pipeline of a conventional PC. The methodology shapes one essential part of three dimensional Morphable Model centralized on facial examination as well. First of all, via conventional analysis-by-synthesis and secondly via based-on pattern's decoder contained by deep-learning architecture. In this part of analysis, the main emphasis is upon modeling the picture creation procedure. It has been taken into account the whole rendering literature; thus, the main consideration is referred to methods and models employed in three dimensional Morphable Models.

Rendering and visibility

Three dimensional Morphable Models fitting procedures diverge into their synthesis of one distinguished picture in the imaging-area (therefore a specific paint per pixel) and their performance in the object-area (therefore a specific paint per vertex and alternatively per triangle). This former analysis retains a benefit of being uncomplicated in encompassing such a kind of model that was created into a high-resolution ultraviolet area. In addition, its result is usually a normal pixel grid able of being transmitted to Convolutional Neural Networks. Procedures operating into object areas calculate the appearance-error. In this process, the peaks of the model are projected onto the picture and the picture intensities are obtained from the observable peaks. Visual range is able to get calculated into an object area and an image area as well. In comparison of those two, image area is generally more efficient.

The initial study of "Blanz and Vetter (1999)" was based on object area rendering where an only color is processed for every central part of a triangle with picture space z-buffering employed for test of visibility. Several consequent techniques functioned through the object area as. The typical process contains colors per peaks calculated by applying some reflection patterns together with per peak material normals. However, recently, such a procedure commenced to shift as soon as more contractual rasterization pipelines can be contained in three dimensional Morphable Models synthesis. Rasterization connects every single pixel (x, y) ∈ I, where I = {1, . . . ,w} × {1, . . . ,h}, with a triangle index or a Null value in case the pixel isn't included by a triangle:

$$\mathbf{raster}_{C,T,w^s,w^e} : I \rightarrow \{1, \ldots, m, \text{Null}\}, (4)$$

where w^s , w^e are considered the parameters of shape and expression whereas T is the triangulation of the mesh accordingly. As long as it constitutes a distinct function, smoothness is absent as well as differentiability. Moreover, there is a computation of 3 weights regarding each singe pixel. These weights concern the peaks of rasterized triangle, such as: a_{C,T,w^s,w^e}, (x, y) ∈ $R^3 \geq 0$. The abovementioned weights depend upon the projected spots of peaks:

$$v_{t^i_{raster_{C,T,w^s,w^e(x,y)}}}, i \in \{1,2,3\}, (5)$$

Frequently, the aforementioned weights exist as barycentric coordinates of central part of pixels inside the triangle. The weights represent an orderly operation of the peak positions, therefore the figure and camera parameters as well. Consequently, rendering is different up to a rasterization's shift, meaning that, when triangle index is related to every pixel doesn't alter. "Tran and Liu (2018a)", encompassed this contractual rasterization pipeline to an in-network differentiable renderer.

When gathering all parameters concerning camera, brightness, facial geometry as well as texture, then Θ = (C, L, w^s ,w^e ,w^t), the rendered appearance can be written into subject area of vertex *j* as $I_{model}^{x,y}(\Theta)$. For a specific picture area rendering the model's appearance is denoted at pixel (x, y) by $I_{model}^{x,y}(\Theta)$. In the humblest occasion, their picture area rendering can be calculated immediately through subject area rendering applying Gouraud interpolation shading:

$$I_{model}^{x,y}(\Theta) = a_{C,T,w^s,w^e}(x,y)^T \begin{bmatrix} I_{model}^{t_j^1}(\Theta) \\ I_{model}^{t_j^2}(\Theta) \\ I_{model}^{t_j^3}(\Theta) \end{bmatrix}, (5)$$

where j = $raster_{C,T,w^s,w^e}(x,y)$. More rasterization schemes might be more complicated. At This Point, peak normals and paints can be rasterized as well as interposed, so reflectance computations can be accomplished into picture area.

It is noteworthy to mention that surpassing the non-differentiable sort of rasterization remains still unclear issue. According to "Hiroharu Kato and Harada (2018)", authors pose a nearly differentiable renderer built through the process of rasterization. In another investigation, "Liu et al. (2019a)" recommend rasterization process where triangles create smooth, therefore differentiable, impact upon the picture. In an optimistic methodology, differentiable rendering utilizing pipelines proposing differentiable path tracing was also studied by "Li et al. (2018)".

In recent times, the specific fixed models utilized in common rendering can be constantly both enhanced and substituted through learning mechanisms, known as neural rendering. One study that confirms this is the one by "Kim et al. (2018)". In their framework, the authors compile an image in the image grid that converts low-quality three-dimensional Morphable Models into a photorealistic video frame.

Analysis through a procedure of synthesis

Three dimensional Morphable Models were commonly utilized in the reformation of images. Reconstructing three-dimensional facial surface of a noticed image includes assessing the Three-dimensional Morphable Models factors. An antithesis in the process of images' synthesis contained in prior part of the analysis is underlined in this point.

Analysis through a procedure of synthesis implies an optimization-level challenges which resolves them in reducing the distance among a detected picture as well as a synthesis of a projected three-dimensional facial surface. The optimization's issue is undoubtedly unclear and with many uncertainties. Hence, is constitutes an extensively examined question, with a variety of answers discovering various input modalities, energy functions, and optimization strategies.

Analysis through a procedure of synthesis is newly utilized as a blend of learning-architectures in an attempt to reform learning-concentrated methodologies.

V. Method and Experiments

The chapter below experiments various methods on three-dimensional landmark localization. To this end, a dataset of 3D faces will be used for training and testing. I will present qualitative as well as quantitative results that indicate that my method improves quality of automatic process of 3D localization and can be used in combination with existed algorithms.

Dataset

The dataset where I conducted the experiments consists of 100 people mimicking 6 basic expressions such as Anger, Disgust, Fear, Happiness, Sadness and Surprise as shown in Fig.3. The dataset consists of 700 meshes in total (6 basic expressions and one neutral). This dataset is a subset of a bigger dataset that contains approximately 5,000 individuals captured while performing a list of expressions through a particular exhibition into Science Museum, in London, for four months period by a group of people of IBug Group. A 3dMDTM facial capture procedure has been utilized, producing a three-dimensional triangular surface for every object formed of nearly 60,000 vertices merged to almost 120,000 triangles, alongside through a high-level of resolution texture map. Each of the 5,000 objects also offered metadata regarding themselves, involving gender, age as well as ethnicity. This specific dataset includes a broad diversity of age, gender and ethnicity.

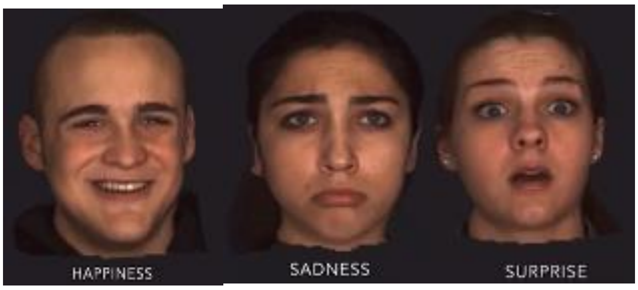


Figure 3: The basic 6 expressions

Preprocessing

To be able to use the dataset, some preprocessing steps are required. One of the most important steps is to create the ground truth landmarks for our dataset. To do this, I set up a landmarker server using the publicly available Landmarker server of Menpo Project Alabort-i-Medina (2014, November). In addition, I choose to use the 68 facial landmarks as in Figure 5. In Figure 6 we can observe the use of landmark tool into a face with texture and without texture.

Figure 5: The 68-points mark-up used for our annotations.

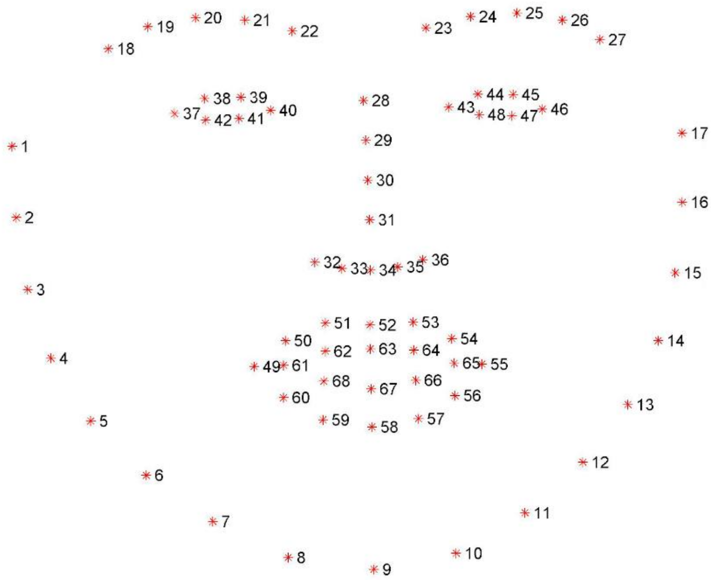
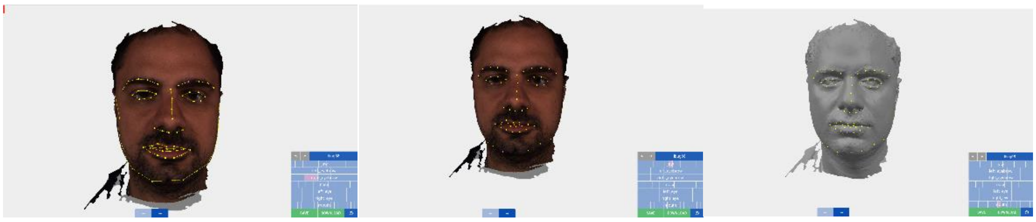


Figure 6: Landmarking a 3D Face



Method

Methods to compare

My method will improve the 3D localization process of methods that are based on the texture of the shapes. One approach is to apply a face detection and alignment framework proposed by "Zhang et al. (2016)" on the 2D texture image and establish a class of 68 thin annotations (landmarks) into 2D space. Exploiting correspondence between the texture image and the 3D mesh, we get the corresponding 3D positions of the landmarks. This approach which is very common on 3D localization doesn't consider the 3D statistics. I propose to leverage the 3D statistics by learning a 3D personal and expression agnostic linear model. I will first describe the proposed model and then I will show how it was built using the training set. Finally, I will illustrate how my approach improves the performance of the other method.

3D personal and expression agnostic linear model

For creating a spatial shape model, we make usage of our training set of landmarks. In our case, since the data are in 3D space, the humblest as well as further uncomplicated technique for representing the i-th landmark by a three-dimensional form can be a vector of its Cartesian coordinates in real domain as $\mathbf{l}_i = [x, y, z]^T$. This form vector is subsequently reclaimed by a concatenation of vectors from every 68 landmarks on a particular vector like $\mathbf{s} = [\mathbf{l}^T_1,...,\mathbf{l}^T_{68}]^T = [x_1, y_1, z_1,...,x_{68}, y_{68}, z_{68}]^T$, $x_i, y_i, z_i \in R$. Therefore, forming area is specified within the real plane is signified as $S \subset R^{3x68}$. It is worth mentioned that the aforementioned method can be implemented from a large number of deformable patterns through computer vision literature as mentioned by "Papaioannou (2017)". Each shape vector was then aligned to a predefined template of landmarks using Procrustes Analysis to remove scale, rotation and translation. At final stage of the above-mentioned procedure, we contain a set of aligned landmarks. Then we applied a dimensionality reduction method as PCA. The outcome of PCA is a mean shape $\underline{\boldsymbol{s}}$ and an orthogonal basis U which encompasses the initial N eigenvectors containing greatest variance. Then, each shape is described like a linear combination of mean shape as well as eigenvectors like it follows:

$$\boldsymbol{s} = \underline{s} + \boldsymbol{U}_p \text{ , (6)}$$

I choose to keep the first 30 components of it. This linear model has captured the variance of each landmark and we can remove noise in the landmarks by projecting a set of landmarks to this subspace and reprojecting back to the original space. The new shapes are then:

$$\boldsymbol{s}_{new} = U^T U \ \boldsymbol{s}_{old}, (7)$$

Metrics

Precision and absolute error will be used to evaluate the proposed approach. As Precision rate can be define a proportion from predicted landmarks inside a particular range from the ground truth landmarks, while absolute error refers to the Euclidean distance among forecasted landmark and ground truth.

Experiments

In the section that follows, I will present how my method can improve the 3D localization of landmarks that is built it on the 2D images.

Our dataset was split in a training set with 80 objects as well as a validation class of 20 subjects. I use the training set for creating a subspace of three-dimensional landmarks. For a specified validation face, three-dimensional landmarks have been processed employing the proposed procedure as in "Zhang et al. (2016)" as it illustrated in Figure 7 and associated by the ground truth landmarks. Then, we recalculate these landmarks using the subspace that has been found from the training set. I compare both sets through ground truth landmarks. Landmark localization error can be calculated like a Euclidean distance among an annotated landmark and an expected landmark. We account the resulted mean as well as standard deviation of landmarking localization error per landmark measuring in millimeters as witnessed in Table 1 and Figure 3.

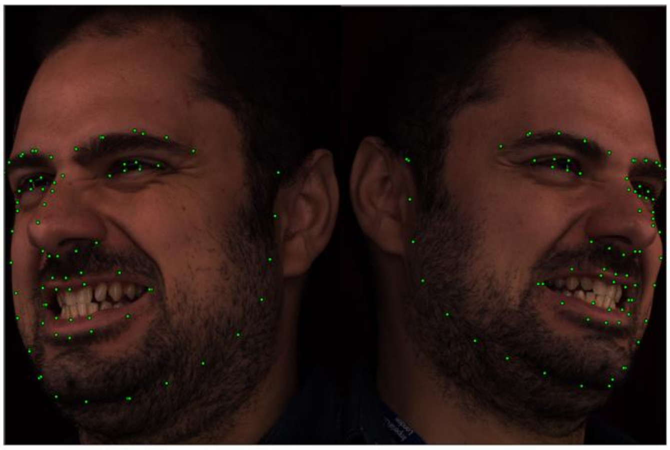


Figure 7. Landmarks predicted in 2D texture image

VI. EMPIRICAL RESULTS

Qualitative results

Throughout this chapter, I will introduce some qualitative outcomes. Figure 8 shows landmarks in a subject in various expressions. The ground truth, the predicted with Zhang et al. [2016] joint method and the predicted with the proposed method landmarks can be signified as red, green as well as blue dots, respectively. As it is clearly witnessed in the Figure 8, our method improves the prediction of the landmarks especially in the case of inner region of face (mouth, eyes). However, we have to notice that our method does not improve the landmarks of the outer region of face (jaw and eyebrows).

Figure 8: Qualitative results in various expression. Red, green and blue dots denote the ground truth, the predicted by the initial method and the predicted by our method landmarks, respectively.

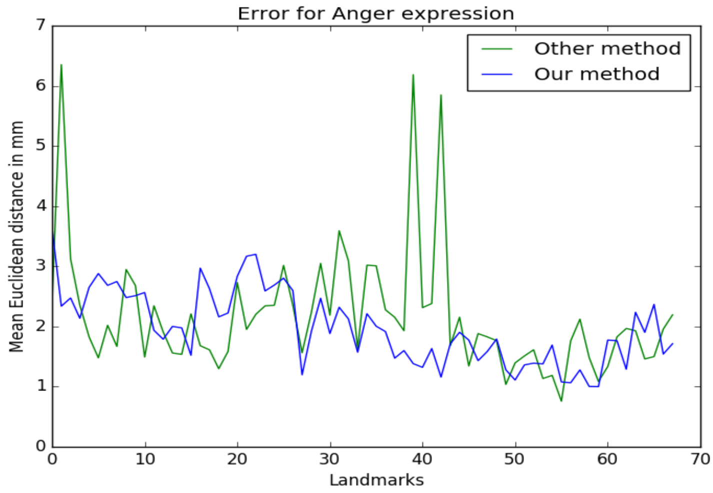
Quantitative results

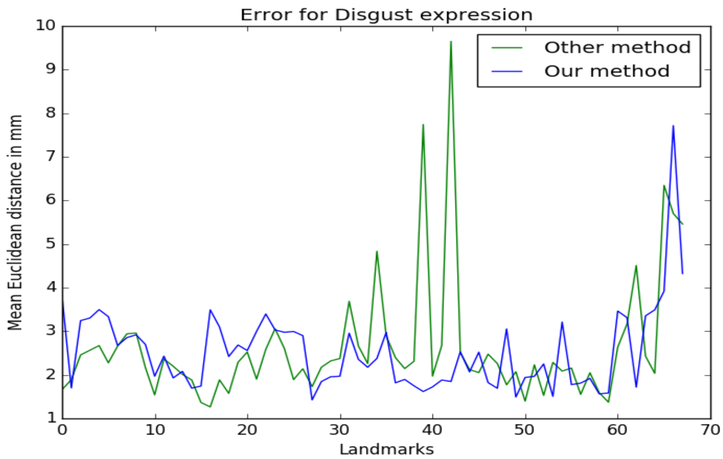
Throughout this chapter, I will display some quantitative outcomes. An absolute error will be used to evaluate the proposed approach. Absolute error refers into Euclidean distance among predicted landmark and ground truth. Table 1 shows the mean Euclidean.

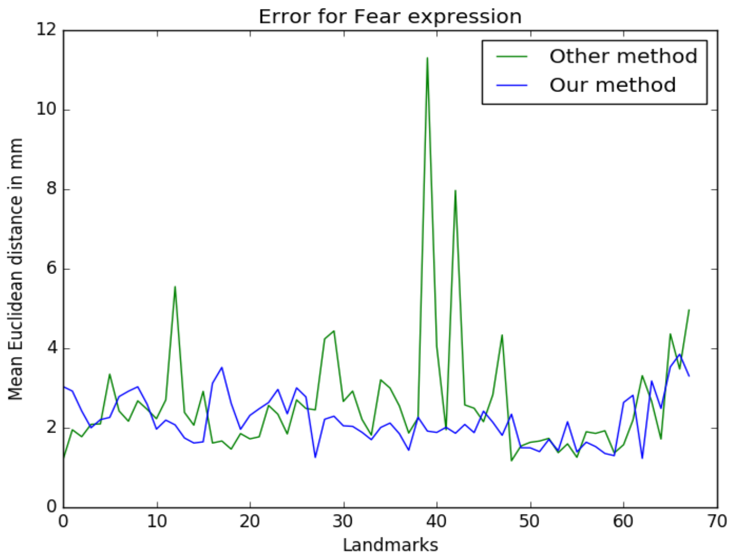
LANDMARK	Method1		Our Method	
	Mean	Std	Mean	Std
Mouth	2.73	0.68	**2.44**	0.61
Right Eyebrow	**2.19**	0.82	3.11	0.93
Right Eye	3.84	1.93	**2.64**	0.66
Left Eyebrow	**2.41**	0.74	3.18	1.08
Left eye	3.87	2.33	**2.06**	0.93
Nose	2.8	1.25	**2.1**	0.74
Jaw	**2.66**	1.98	3.02	1.35

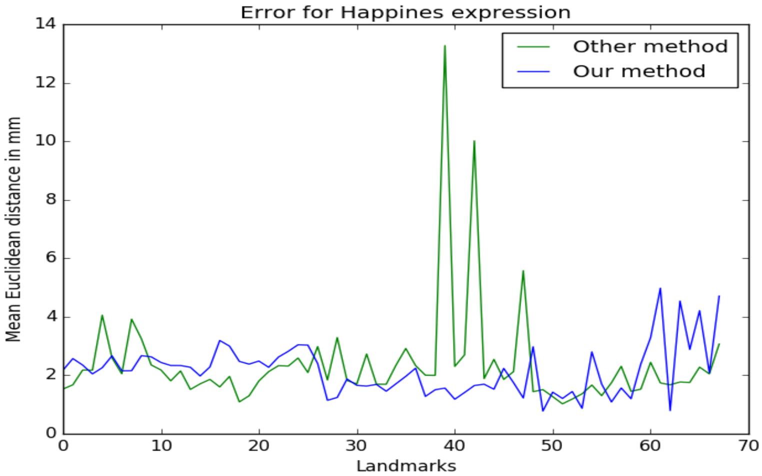
Table 1: Mean error for region of face

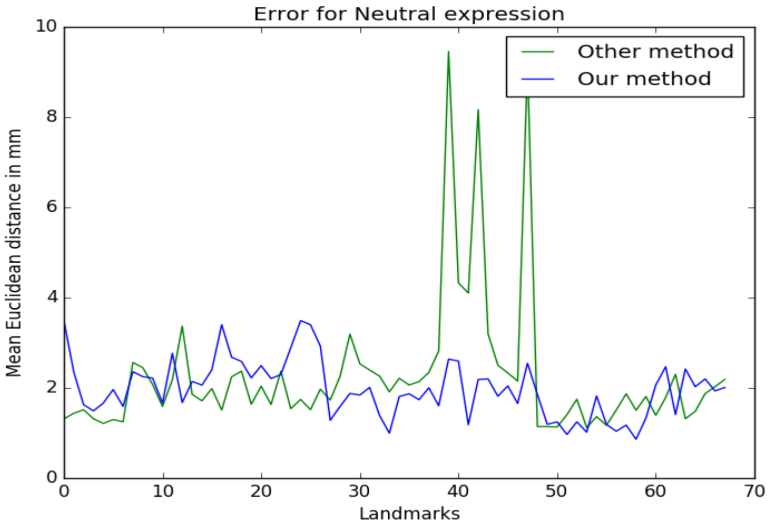
Figures 10-16 illustrate the errors in the 6 expressions:

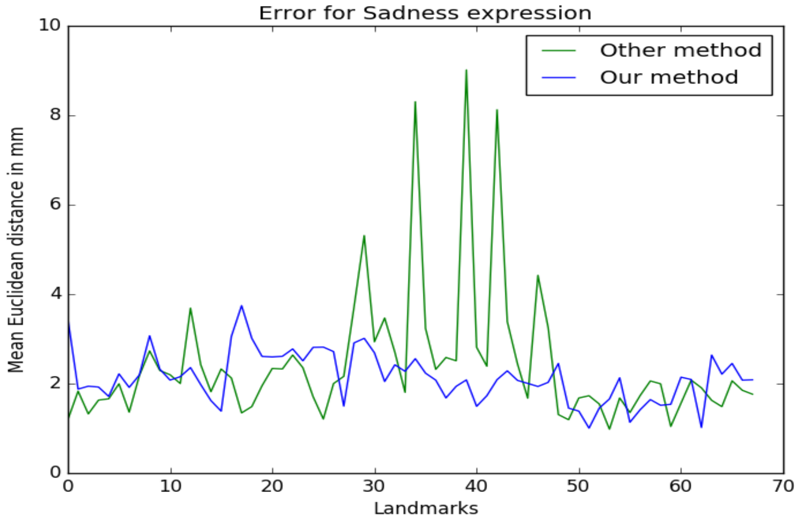

Error for Anger expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks

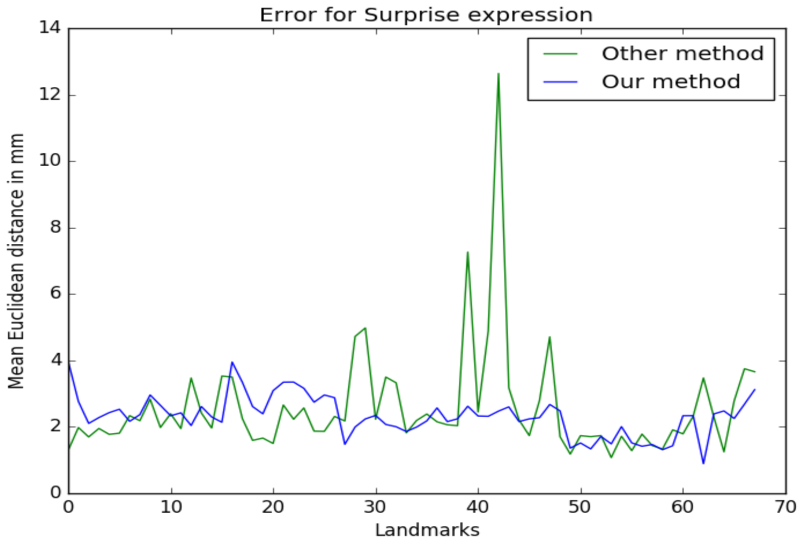

Error for Disgust expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks


Error for Fear expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks


Error for Happines expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks
0
2
4
6
8
10
12
14
0
10
20
30
40
50
60
70


Error for Neutral expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks
0
2
4
6
8
10
0
10
20
30
40
50
60
70


Error for Sadness expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks
0
2
4
6
8
10
0
10
20
30
40
50
60
70


Error for Surprise expression
Other method
Our method
Mean Euclidean distance in mm
Landmarks
0
2
4
6
8
10
12
14
0
10
20
30
40
50
60
70

VII. CONCLUSIONS AND FUTURE WORK

The aforementioned method of the above study shows an improvement in the quality of the automatic procedure of 3D localization based on the texture of the shapes and can be utilized in combination with existed algorithms. By learning a 3D identity and expression agnostic model I leveraged the 3D statistics that is not considered. This linear model captured the variance of every landmark and removed the noise in the landmarks. As a result, this suggested procedure improves prediction of the landmarks especially concerning the occasion of inner area of face like mouth, eyes as well as nose. On the other hand, joint method of "Zhang et al. (2016)" offers better results in outer region of face like jaws, eyebrows. This outcome may be improved if there were more data.

Regarding the future work, new approaches based on Deep Learning models might produce better results.

References

Abrevaya, V. F., Wuhrer, S., & Boyer, E. (2018, September). Spatiotemporal modeling for efficient registration of dynamic 3d faces. In 2018 International Conference on 3D Vision (3DV) (pp. 371-380). IEEE.

Achenbach, J., Brylka, R., Gietzen, T., Hebel, K. Z., Schömer, E., Schulze, R., ... & Schwanecke, U. (2018, September). A multilinear model for bidirectional craniofacial reconstruction. In Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine (pp. 67-76). Eurographics Association.

Ackermann, J., & Goesele, M. (2015). A survey of photometric stereo techniques. Foundations and Trends® in Computer Graphics and Vision, 9(3-4), 149-254.

Alabort-i-Medina, J., Antonakos, E., Booth, J., Snape, P., & Zafeiriou, S. (2014, November). Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 679-682). ACM.

Aldrian, O., & Smith, W. A. (2012). Inverse rendering of faces with a 3D morphable model. IEEE transactions on pattern analysis and machine intelligence, 35(5), 1080-1093.

Allen, B., Curless, B., Curless, B., & Popović, Z. (2003, July). The space of human body shapes: reconstruction and parameterization from range scans. In ACM transactions on graphics (TOG) (Vol. 22, No. 3, pp. 587-594). ACM.

Alotaibi, S., & Smith, W. A. (2017). A Biophysical 3D Morphable Model of Face Appearance. In Proceedings of the IEEE International Conference on Computer Vision (pp. 824-832).

Amberg, B., Knothe, R., & Vetter, T. (2008, September). Expression invariant 3D face recognition with a morphable model. In 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 1-6). IEEE.

Atick, J. J., Griffin, P. A., & Redlich, A. N. (1996). Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. Neural computation, 8(6), 1321-1340.

Beeler, B., Good, B., Rashkeev, S., Deo, C., Baskes, M., & Okuniewski, M. (2010). First principles calculations for defects in U. Journal of Physics: Condensed Matter, 22(50), 505703.

Beeler, T., Bickel, B., Noris, G., Beardsley, P., Marschner, S., Sumner, R. W., & Gross, M. (2012). Coupled 3D reconstruction of sparse facial hair and skin. ACM Transactions on Graphics (ToG), 31(4), 117.

Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., ... & Gross, M. (2011, August). High-quality passive facial performance capture using anchor frames. In ACM Transactions on Graphics (TOG) (Vol. 30, No. 4, p. 75). ACM.

Beeler, T., & Bradley, D. (2014). Rigid stabilization of facial expressions. ACM Transactions on Graphics (TOG), 33(4), 44.

Bérard, P., Bradley, D., Nitti, M., Beeler, T., & Gross, M. (2014). High-quality capture of eyes. ACM Transactions on Graphics (TOG), 33(6), 223.

Bermano, A., Beeler, T., Bradley, D., Bickel, B., Gross, M., & Kozlov, Y. (2017). U.S. Patent No. 9,691,165. Washington, DC: U.S. Patent and Trademark Office.

Berretti, S., Amor, B. B., Daoudi, M., & Del Bimbo, A. (2011). 3D facial expression recognition using SIFT descriptors of automatically detected key points. The Visual Computer, 27(11), 1021.

Blanz, V., & Vetter, T. (1999, July). A morphable model for the synthesis of 3D faces. In Siggraph (Vol. 99, No. 1999, pp. 187-194).

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. IEEE Transactions on pattern analysis and machine intelligence, 25(9), 1063-1074.

Bolkart, T., & Wuhrer, S. (2015). A groupwise multilinear correspondence optimization for 3d faces. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3604-3612).

Bolkart, T., & Wuhrer, S. (2015). 3D faces in motion: Fully automatic registration and statistical analysis. Computer Vision and Image Understanding, 131, 100-115.

Booth, J., Antonakos, E., Ploumpis, S., Trigeorgis, G., Panagakis, Y., & Zafeiriou, S. (2017, July). 3D face morphable models" In-The-Wild". In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5464-5473). IEEE.

Booth, J., Roussos, A., Ververas, E., Antonakos, E., Ploumpis, S., Panagakis, Y., & Zafeiriou, S. (2018). 3D Reconstruction of “In-the-Wild” Faces in Images and Videos. IEEE transactions on pattern analysis and machine intelligence, 40(11), 2638-2652.

Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A., & Dunaway, D. (2016). A 3d morphable model learnt from 10,000 faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5543-5552).

Booth, J., & Zafeiriou, S. (2014, October). Optimal uv spaces for facial morphable model construction. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 4672-4676). IEEE.

Bouaziz, S., Wang, Y., & Pauly, M. (2013). Online modeling for real time facial animation. ACM Transactions on Graphics (ToG), 32(4), 40.

Bradley, D., Heidrich, W., Popa, T., & Sheffer, A. (2010, July). High resolution passive facial performance capture. In ACM transactions on graphics (TOG) (Vol. 29, No. 4, p. 41). ACM.

Brunton, A., Salazar, A., Bolkart, T., & Wuhrer, S. (2014). Review of statistical shape spaces for 3D data with comparative analysis for human faces. Computer Vision and Image Understanding, 128, 1-17.

Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In Proceedings of the IEEE International Conference on Computer Vision (pp. 1021-1030).

Cao, C., Weng, Y., Zhou, S., Tong, Y., & Zhou, K. (2013). Face warehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20(3), 413-425.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998, June). Active appearance models. In European conference on computer vision (pp. 484-498). Springer, Berlin, Heidelberg.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. Computer vision and image understanding, 61(1), 38-59.

Cosker, D., Krumhuber, E., & Hilton, A. (2011, November). A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In 2011 International Conference on Computer Vision (pp. 2296-2303). IEEE.

Craw, I., & Cameron, P. (1991). Parameterizing images for recognition and reconstruction. In BMVC91 (pp. 367-370). Springer, London.

Creusot, C., Pears, N., & Austin, J. (2013). A machine-learning approach to Key Point detection and landmarking on 3D meshes. International journal of computer vision, 102(1-3), 146-179.

Cudeiro, D., Bolkart, T., Laidlaw, C., Ranjan, A., & Black, M. J. (2019). Capture, Learning, and Synthesis of 3D Speaking Styles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 10101-10111).

Cunningham, S. (2000). GVE'99: Report of the 1999 eurographics/SIGGRAPH workshop on graphics and visualization education. ACM SIGGRAPH Computer Graphics, 33(4), 96-102.

Dai, H., Pears, N., & Smith, W. (2018, May). A Data-augmented 3D Morphable Model of the Ear. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 404-408). IEEE.

Dai, H., Pears, N., Smith, W. A., & Duncan, C. (2017). A 3d morphable model of craniofacial shape and texture variation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3085-3093).

Debevec, P., Hawkins, T., Tchou, C., Duiker, H. P., Sarokin, W., & Sagar, M. (2000, July). Acquiring the reflectance field of a human face. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 145-156). ACM Press/Addison-Wesley Publishing Co..

Dessein, A., Smith, W. A., Wilson, R. C., & Hancock, E. R. (2015). Example-based modeling of facial texture from deficient data. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3898-3906).

Dibeklioglu, H., Salah, A. A., & Akarun, L. (2008, September). 3D facial landmarking under expression, pose, and occlusion variations. In 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems (pp. 1-6). IEEE.

Dorai, C., & Jain, A. K. (1997). COSMOS-A representation scheme for 3D free-form objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(10), 1115-1130.

Dovgard, R., & Basri, R. (2004, May). Statistical symmetric shape from shading for 3D structure recovery of faces. In European Conference on Computer Vision (pp. 99-113). Springer, Berlin, Heidelberg.

Dryden, I. L., & Mardia, K. V. (1998). Statistical shape analysis: Wiley series in probability and statistics.

Echevarria, J. I., Bradley, D., Gutierrez, D., & Beeler, T. (2014). Capturing and stylizing hair for 3D fabrication. ACM Transactions on Graphics (ToG), 33(4), 125.

Egger, B. (2017). Semantic morphable models (Doctoral dissertation, University of Basel).

Egger, B., Kaufmann, D., Schönborn, S., Roth, V., & Vetter, T. (2016, February). Copula eigenfaces. In 11th International Conference on Computer Graphics Theory and Applications (GRAPP) (Vol. 1, pp. 50-58).

Egger, B., Kaufmann, D., Schönborn, S., Roth, V., & Vetter, T. (2016, February). Copula Eigenfaces with Attributes: Semiparametric Principal Component Analysis for a Combined Color, Shape and Attribute Model. In International Joint Conference on Computer Vision, Imaging and Computer Graphics (pp. 95-112). Springer, Cham.

Edwards, G. J., Taylor, C. J., & Cootes, T. F. (1998, April). Interpreting face images using active appearance models. In Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 300-305). IEEE.

Fang, T., Zhao, X., Ocegueda, O., Shah, S. K., & Kakadiaris, I. A. (2012). 3D/4D facial expression analysis: An advanced annotated face model approach. Image and vision Computing, 30(10), 738-749.

Feng, Z. H., Hu, G., Kittler, J., Christmas, W., & Wu, X. J. (2015). Cascaded collaborative regression for robust facial landmark detection trained using a mixture of synthetic and real images with dynamic weighting. IEEE Transactions on Image Processing, 24(11), 3425-3440.

Ferrari, C., Lisanti, G., Berretti, S., & Del Bimbo, A. (2015, October). Dictionary learning based 3D morphable model construction for face recognition with varying expression and pose. In 2015 International Conference on 3D Vision (pp. 509-517). IEEE.

Furukawa, Y., & Ponce, J. (2009, June). Dense 3d motion capture for human faces. In 2009 IEEE conference on computer vision and pattern recognition (pp. 1674-1681). IEEE.

Garrido, P., Zollhöfer, M., Wu, C., Bradley, D., Pérez, P., Beeler, T., & Theobalt, C. (2016). Corrective 3D reconstruction of lips from monocular video. ACM Trans. Graph., 35(6), 219-1.

Geng, J. (2011). Structured-light 3D surface imaging: a tutorial. Advances in Optics and Photonics, 3(2), 128-160.

Georghiades, A. S. (2003, October). Incorporating the Torrance and Sparrow Model of Reflectance in Uncalibrated Photometric Stereo. In iccv (Vol. 3, p. 816).

Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., & Vetter, T. (2018, May). Morphable face models-an open framework. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 75-82). IEEE.

Ghosh, A., Fyffe, G., Tunwattanapong, B., Busch, J., Yu, X., & Debevec, P. (2011, December). Multiview face capture using polarized spherical gradient illumination. In ACM Transactions on Graphics (TOG) (Vol. 30, No. 6, p. 129). ACM.

Gotardo, P., Riviere, J., Bradley, D., Ghosh, A., & Beeler, T. (2018). Practical dynamic facial appearance modeling and acquisition.

Gotardo, P. F., Simon, T., Sheikh, Y., & Matthews, I. (2015). Photogeometric scene flow for high-detail dynamic 3d reconstruction. In Proceedings of the IEEE International Conference on Computer Vision (pp. 846-854).

Guo, J., Mei, X., & Tang, K. (2013). Automatic landmark annotation and dense correspondence registration for 3D human facial images. BMC bioinformatics, 14(1), 232.

Hallinan, P. W., Gordon, G., Yuille, A. L., Giblin, P., & Mumford, D. (1999). Two- and three-dimensional patterns of the face. AK Peters/CRC Press.

Han, F., & Liu, H. (2012). Semiparametric principal component analysis. In Advances in Neural Information Processing Systems (pp. 171-179).

Hernández, C., Vogiatzis, G., Brostow, G. J., Stenger, B., & Cipolla, R. (2007, October). Non-rigid photometric stereo with colored lights. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE.

Hewer, A., Wuhrer, S., Steiner, I., & Richmond, K. (2018). A multilinear tongue model derived from speech related MRI data of the human vocal tract. Computer Speech & Language, 51, 68-92.

Huber, P., Hu, G., Tena, R., Mortazavian, P., Koppen, P., Christmas, W. J., ... & Kittler, J. (2016, February). A multiresolution 3d morphable face model and fitting framework. In Proceedings of the 11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.

Hutton, T. J., Buxton, B. R., & Hammond, P. (2001, December). Dense surface point distribution models of the human face. In Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001) (pp. 153-160). IEEE.

Jahanbin, S., Choi, H., & Bovik, A. C. (2011). Passive multimodal 2-D+ 3-D face recognition using Gabor features and landmark distances. IEEE Transactions on Information Forensics and Security, 6(4), 1287-1304.

De Jong, M. A., Wollstein, A., Ruff, C., Dunaway, D., Hysi, P., Spector, T., ... & Wolvius, E. B. (2015). An automatic 3d facial landmarking algorithm using 2d gabor wavelets. IEEE Transactions on Image Processing, 25(2), 580-588.

Jones, M. J., & Poggio, T. (1998). Multidimensional morphable models: A framework for representing and matching object classes. International Journal of Computer Vision, 29(2), 107-131.

Jones, M. J., & Poggio, T. (1998, January). Multidimensional morphable models. In Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271) (pp. 683-688). IEEE.

Kakadiaris, I. A., Passalis, G., Theoharis, T., Toderici, G., Konstantinidis, I., & Murtuza, N. (2005, June). Multimodal face recognition: Combination of geometry with physiological information. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 2, pp. 1022-1029). IEEE.

Kato, H., Ushiku, Y., & Harada, T. (2018). Neural 3d mesh renderer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3907-3916).

Khamis, S., Taylor, J., Shotton, J., Keskin, C., Izadi, S., & Fitzgibbon, A. (2015). Learning an efficient model of hand shape variation from depth images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2540-2548).

Kim, H., Carrido, P., Tewari, A., Xu, W., Thies, J., Niessner, M., ... & Theobalt, C. (2018). Deep video portraits. ACM Transactions on Graphics (TOG), 37(4), 163.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., ... & Shade, J. (2000, July). The digital Michelangelo project: 3D scanning of large statues. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 131-144). ACM Press/Addison-Wesley Publishing Co..

Lewis, J. P., Mo, Z., Anjyo, K., & Rhee, T. (2014). Probable and improbable faces. In Mathematical Progress in Expressive Image Synthesis I (pp. 21-30). Springer, Tokyo.

Li, H., Yu, J., Ye, Y., & Bregler, C. (2013). Realtime facial animation with on-the-fly correctives. ACM Trans. Graph., 32(4), 42-1.

Li, T. M., Aittala, M., Durand, F., & Lehtinen, J. (2018, December). Differentiable monte Carlo ray tracing through edge sampling. In SIGGRAPH Asia 2018 Technical Papers (p. 222). ACM.

Li, T., Bolkart, T., Black, M. J., Li, H., & Romero, J. (2017). Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics (TOG), 36(6), 194.

Liang, S., Wu, J., Weinberg, S. M., & Shapiro, L. G. (2013, July). Improved detection of landmarks on 3D human face data. In 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 6482-6485). IEEE.

Lin, T. H., Shih, W. P., Chen, W. C., & Ho, W. Y. (2006, March). 3D face authentication by mutual coupled 3D and 2D feature extraction. In Proceedings of the 44th annual Southeast regional conference (pp. 423-427). ACM.

Liu, S., Li, T., Chen, W., & Li, H. (2019). Soft Rasterizer: A Differentiable Renderer for Image-based 3D Reasoning. arXiv preprint arXiv:1904.01786.

Lu, X., Jain, A. K., & Colbry, D. (2005). Matching 2.5 D face scans to 3D models. IEEE transactions on pattern analysis and machine intelligence, 28(1), 31-43.A.S. Mian , M. Bennamoun , R. Owens , An efficient multimodal 2D–3D hybrid approach to automatic face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (11) (2007) 1927–1943 .

Ma, W. C., Hawkins, T., Peers, P., Chabert, C. F., Weiss, M., & Debevec, P. (2007, June). Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In Proceedings of the 18th Eurographics conference on Rendering Techniques (pp. 183-194). Eurographics Association.

Marschner, S. R., Westin, S. H., Lafortune, E. P., Torrance, K. E., & Greenberg, D. P. (1999). Image-based BRDF measurement including human skin. In Rendering Techniques' 99 (pp. 131-144). Springer, Vienna.

Milborrow, S., Bishop, T., & Nicolls, F. (2013). Multiview active shape models with sift descriptors for the 300-w face landmark challenge. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 378-385).

Moschoglou, S., Ververas, E., Panagakis, Y., Nicolaou, M. A., & Zafeiriou, S. (2018). Multi-attribute robust component analysis for facial UV maps. IEEE Journal of Selected Topics in Signal Processing, 12(6), 1324-1337.

Mpiperis, I., Malassiotis, S., & Strintzis, M. G. (2008). Bilinear models for 3-D face and facial expression recognition. IEEE Transactions on Information Forensics and Security, 3(3), 498-511.

Nehab, D., Rusinkiewicz, S., Davis, J., & Ramamoorthi, R. (2005). Efficiently combining positions and normals for precise 3D geometry. ACM transactions on graphics (TOG), 24(3), 536-543.

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., ... & Fitzgibbon, A. W. (2011, October). Kinect fusion: Real-time dense surface mapping and tracking. In ISMAR (Vol. 11, No. 2011, pp. 127-136).

Papaioannou, A., Antonakos, E., & Zafeiriou, S. (2017). Complex representations for learning statistical shape priors. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1180-1184). IEEE.

Park, J. J., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S. (2019). Deepsdf: Learning continuous signed distance functions for shape representation. arXiv preprint arXiv:1901.05103.

Passalis, G., Perakis, P., Theoharis, T., & Kakadiaris, I. A. (2011). Using facial symmetry to handle pose variations in real-world 3D face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(10), 1938-1951.

Patel, A., & Smith, W. A. (2011, November). Simplification of 3D morphable models. In 2011 International Conference on Computer Vision (pp. 271-278). IEEE.

Patel, A., & Smith, W. A. (2012). Driving 3D morphable models using shading cues. Pattern Recognition, 45(5), 1993-2004.

Paysan, P. (2010). Statistical modeling of facial aging based on 3D scans (Doctoral dissertation, University of Basel).

Perakis, P., Theoharis, T., & Kakadiaris, I. A. (2014). Feature fusion for facial landmark detection. Pattern Recognition, 47(9), 2783-2793.

Pierrard, J. S. (2008). Skin segmentation for robust face image analysis (Doctoral dissertation, University of Basel).

Qu, C., Monari, E., Schuchert, T., & Beyerer, J. (2014, August). Fast, robust and automatic 3D face model reconstruction from videos. In 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 113-118). IEEE.

Saito, S., Hu, L., Ma, C., Ibayashi, H., Luo, L., & Li, H. (2018, December). 3D hair synthesis using volumetric variational autoencoders. In SIGGRAPH Asia 2018 Technical Papers (p. 208). ACM.

Salazar, A., Wuhrer, S., Shu, C., & Prieto, F. (2014). Fully automatic expression-invariant face correspondence. Machine Vision and Applications, 25(4), 859-879.

Schneider, A., Egger, B., & Vetter, T. (2018, May). A Parametric Freckle Model for Faces. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 431-435). IEEE.

Schneider, T., Dymczyk, M., Fehr, M., Egger, K., Lynen, S., Gilitschenski, I., & Siegwart, R. (2018). maplab: An open framework for research in visual-inertial mapping and localization. IEEE Robotics and Automation Letters, 3(3), 1418-1425.

Schumacher, M., & Blanz, V. (2015, May). Exploration of the correlations of attributes and features in faces. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (Vol. 1, pp. 1-8). IEEE.

Shelton, C. R. (2000). Morphable surface models. International Journal of Computer Vision, 38(1), 75-91.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Josa a, 4(3), 519-524.

Smith, W. A. (2016). The perspective face shape ambiguity. In Perspectives in Shape Analysis (pp. 299-319). Springer, Cham.

Song, M., Tao, D., Sun, S., Chen, C., & Maybank, S. J. (2014). Robust 3D face landmark localization based on local coordinate coding. IEEE transactions on image processing, 23(12), 5108-5122.

Tewari, A., Zollhöfer, M., Garrido, P., Bernard, F., Kim, H., Pérez, P., & Theobalt, C. (2018). Self-supervised multi-level face model learning for monocular reconstruction at over 250 hz. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2549-2559).

Tewari, A., Bernard, F., Garrido, P., Bharaj, G., Elgharib, M., Seidel, H. P., ... & Theobalt, C. (2019). FML: face model learning from videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 10812-10822).

Tran, L., Liu, F., & Liu, X. (2019). Towards high-fidelity nonlinear 3D face morphable model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1126-1135).

Tran, L., & Liu, X. (2018). Nonlinear 3D face morphable model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7346-7355).

Tran, L., & Liu, X. (2019). On learning 3d face morphable model from in-the-wild images. IEEE transactions on pattern analysis and machine intelligence.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), 71-86.

Vetter, T., & Poggio, T. (1997). Linear object classes and image synthesis from a single example image. IEEE Transactions on pattern analysis and machine intelligence, 19(7), 733-742.

Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., & Samaras, D. (2008). Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), 1968-1984.

Weise, T., Li, H., Van Gool, L., & Pauly, M. (2009, August). Face/off: Live facial puppetry. In Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer animation (pp. 7-16). ACM.

Wiskott, L., Fellous, J. M., Krüger, N., & Von Der Malsburg, C. (1997, September). Face recognition by elastic bunch graph matching. In International Conference on Computer Analysis of Images and Patterns (pp. 456-463). Springer, Berlin, Heidelberg.

Xu, Z., Wu, H. T., Wang, L., Zheng, C., Tong, X., & Qi, Y. (2014). Dynamic hair capture using spacetime optimization. To appear in ACM TOG, 33, 6.

Zafeiriou, S., Atkinson, G. A., Hansen, M. F., Smith, W. A., Argyriou, V., Petrou, M., ... & Smith, L. N. (2012). Face recognition and verification using photometric stereo: The photoface database and a comprehensive evaluation. IEEE transactions on information forensics and security, 8(1), 121-135.

Zhang, C., Smith, W. A., Dessein, A., Pears, N., & Dai, H. (2016). Functional faces: Groupwise dense correspondence using functional maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5033-5041).

Zhang, L., Snavely, N., Curless, B., & Seitz, S. M. (2008). Spacetime faces: High-resolution capture for~ modeling and animation. In Data-Driven 3D Facial Animation (pp. 248-276). Springer, London.

Zhou, Y., Deng, J., Kotsia, I., & Zafeiriou, S. (2019). Dense 3D Face Decoding over 2500FPS: Joint Texture & Shape Convolutional Mesh Decoders. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1097-1106).

Zivanov, J., Paysan, P., & Vetter, T. (2009). Facial normal map capture using four lights–an effective and inexpensive method of capturing the fine scale detail of human faces using four-point lights. In In Proc. GRAPP.

Zoss, G., Beeler, T., Gross, M., & Bradley, D. (2019). Accurate markerless jaw tracking for facial performance capture. ACM Transactions on Graphics (TOG), 38(4), 50.

Zoss, G., Bradley, D., Bérard, P., & Beeler, T. (2018). An empirical rig for jaw animation. ACM Transactions on Graphics (TOG), 37(4), 59.

Zuffi, S., Kanazawa, A., & Black, M. J. (2018). Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3955-3963).