Face Synthesis and Reconstruction – A Survey

(Comprehensive Literature Review)

Arpita S Tugave
Florida Institute of Cyber Security, University of Florida
December, 2016

***Abstract*—With recent innovations, human face modelling is not just a vision or graphics problem, but branches into deep leaning, machine learning and detailed component rendering. In this paper, we not just present available methods, but also, provide insights into the methods which could aid in realistic, faster and accurate face modeling. We make sure this is a complete study for readers reaching in this area. We also consider new methods producing better results in other applications which can have direct implication in Face rendering.**

Keywords— computer vision; computer graphics; deep learning; machine learning; unsupervised learning; shape from X, 3D morphable model

I. INTRODUCTION

Synthesis of biometric modalities such as: Face, Fingerprint, iris, handwriting, and speech have been researched extensively. SFinGe is a commercially available synthetic fingerprint generation tool developed by Cappelli et al. [1]. Similar to this, FaceGen [2] started in 1998 generates 3D face and supports other industry such as gaming, biometric security and so-on. Other modalities such as: iris, signature synthesis aren't' available at the commercial platform. Even then research in these areas is no less. In human beings, speech and emotions are directly correlated. Henceforth, using speech, emotions can be synthesized in advanced face synthesis.

There has been an increase in the demand for 3D Face synthesis in the fields of robotics, security, image recognition, virtual reality, gaming, animation studios, video conferencing, and other applications. Also, there have been other quite interesting applications such as virtual try-on and customization of eyewear; wherein, software converts a selfie vide into a highly accurate 3D model of a customer's face and then optimizes eyewear designs to match customer's unique facial anatomy.

A newer technology "Oben" is bent on exploring virtual 3D face avatar [3]. The company provides a platform wherein an individual can transport themselves into VR environment. Unlike other technologies which uses cartoon-like rendering, it uses realistic 3D avatars. Likewise, many other companies are researching 3D face generation as a part of bigger project or as a stand-alone project. Two of the million dollar applications are Disney's realistic facial hair structuring [5] and Facebook's DeepFace [4]. DeepFace is used for tagging friends in Facebook. This uses a 3D face reconstruction as its intermediate step for accurate face recognition matching human-level performance.

Fig. 1. Oben's virtual face avatar

In this paper, the first section will discuss related work. Secondly, available face databases and the ones that can be effective used for various face reconstruction and synthesis problems will be noted. We split the methods that we use into: learning, shape from X and model based reconstruction. In the end, we hope the readers will have all the available related knowledge to proceed in this field.

II. RELATED WORK

There are not many papers which survey on "Synthetic biometric" as it is vast. The very first survey [6], by S. N. Yanushkevich, is dated a decade back. The main focus of this paper is: synthetic face, fingerprint and signature. The author also came up with the book titled "Biometric Inverse Problems" extending even to synthetic DNA [7]. Here, the topic of face synthesis is described very briefly and the techniques are outdated. Nicholas and Douglas work in the area of synthetic biometric involves survey on the advantages of biometric data synthesis and criteria for it to be used in the real world applications [8]. However, the criteria aren't supported and are very generalized. Further, in their paper [9], they also provide physics based- statistical approaches for data synthesis. This concept can be extended to age based face retouching. Privacy concerns and need for synthetic biometric data has also been studied by Kazuhiko Sumi [10]. This paper uses PCA selectively on different facial parts to replace a real face with the deformed one. Thus securing person's identity anonymous to the system. Article by Mihailescu [11] describes biometric and its synthesis as inseparable. But it is lacking in detail and discusses at a very high level.

Let's look into the 3D face modeling survey papers. M. Judith Leo's paper [12] is the detailed survey paper on 3D face reconstruction. It studies active and passive modelling methods. In our paper, we extend Judith's paper with up to date passive modelling methods. Widanagamaachchi's survey paper [13] suggests drawbacks for accurate 3D face modelling system. Firstly, we need many images to capture enough details for an accurate face model. However, recent research has focused on face generation using a single image. Secondly, countless feature points or landmarks have to be initialized for higher accuracy. Many face modelling techniques have trade off wrt #feature points, speed and accuracy. Article by Georgios Stylianos [14], compares 3D Face reconstruction methods of: Example based, Stereo orthogonal, Stereo non-orthogonal, Video, and Silhouettes. We extend this with several other reconstruction techniques and suggest suitable applications for respective methods. [15] Literature review even though informative, is unfocussed and lacking in depth.

Modeling human faces used to be a computer graphics problem, with many researchers focused on this topic. To mention, Parke [41, 42, 55] is the first pioneer in this area. Various techniques model face geometries [89, 90, 95, 91, 92, 93, 94, 95, 96, 97, 81, 98, 99 ,100, 101, 102, 103, 104, 105, 106, 107, 40, 43, 44, 45, 46, 47]. University of Glasgow had very interesting research on facial expressions [118].

III. DATASET

Many Face datasets are available. We mention a few so that researchers can start experimenting with this information.

First of all, Radboud Face Database [112] consists of Caucasian males, females, children, boys, girls, Moroccan Dutch males of 67 subjects in total. Each subject has 8 emotional expression of anger, happiness, sadness, surprise, disgust, fear, contempt, and neutral. Each emotion was shown with three different gaze directions and all pictures were taken from five camera angles simultaneously.

FERET [113] is more convoluted 2D face database containing 14051 8-bit grayscale images of human – head orientation varying from left to right profile view. In DeepFace paper [4], labeled faced in Wild [114] is used. It is used to study the problem of unconstrained face recognition. The data are collected from web consisting of 13,000 images. Each image has been labeled with the name of the subject. In total there are 1680 subjects with two or more images.

In order to make a 3D Morphable model we need 3D database of MPI [115], MSU or USF [97]. These datasets are used in conjunction with 3D reconstruction algorithms. The MPI database contains 200 textured face models generated based on seven laser scans for each subject. Only 5 out of 200 are publicly available. Data acquisition is by using 3D Cyberware 3D scanner [116].

USF database consists of 218 textured 3D face data of different gender and different ethnic origin. For each sample there are over 7000 vertices defined. Images are acquired by using a Cyberware 3D laser scanner.

Fig. 2. FERET database

Fig. 3. MPI database

Fig. 4. MSU Database with varying pose and illumination

IV. METHODS BASED ON LEARNING

Nowadays, Convolutional neural networks are being refined to solve every problem in image processing. In this section we consider five image processing problems of: face segmentation, face recognition, generation of synthetic faces, 3D modelling from depth map and emotion analyzer. These problems relate to face reconstruction and synthesis directly or indirectly.

A. Face segmentaion using Learning methods

As mentioned in [13], for detailed component structuring; nose, ears, eyes, hair are segmented. These segments can then be modeled or synthesized separately to generate accurate models. Firstly, segmentation technique of Region based CNN (R-CNN) can be used. Secondly, we look into combinations of CNN and RNN to generate image descriptions. Lastly, segmentation using Conditional random fields- CRF are explored

Ross Girshick and his group at UC Berkeley came up with R-CNN [16] which could solve object detection tasks using deep learning. R-CNN uses selective search with 2000 different regions identifying the object. After extracting these regions, they are wrapped to be fed into CNNs to output feature vector for each class. These feature vectors are fed into a bounding box regressor to obtain accurate co-ordinates and then to SVM classifier to derive the classification results. Lastly, bounding boxes with maximum overlap are compressed using non-maximum suppression.

However, R-CNNs are extremely slow due to multiple layers of ConvNets, SVMs and bounding box regressor. These tasks also made R-CNN computationally expensive. Hence, Ross Girshick at Microsoft research came up with Fast R-CNN [18]. These are very fast and achieved by swiping the order of CNN and region generation. Hence the computation of CNNs are shared between the different regions. In total, image is fed to the CNN from which regions are extracted and fed into softmax classifier and bounding box regressor. Recently, Faster CNNs [19] have been proposed by using Region proposal network-RPN.

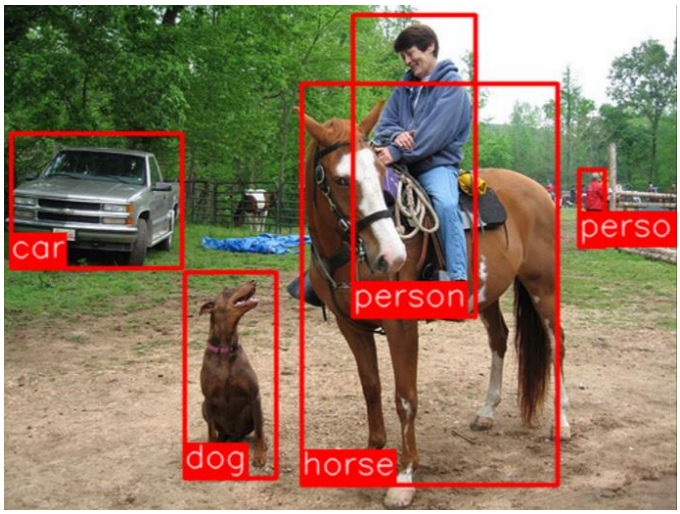


Fig. 7. R-CNN output for [17] network

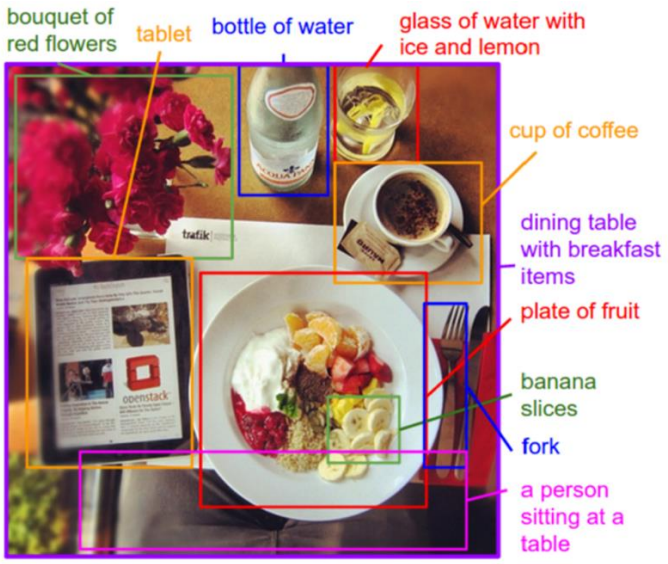


Fig. 6. Combination of CNN and RNN

RPN is used after the fully connected layers of CNN. Regions are produced in no time by using this layer.

Andrej Karpathy and Fei-Fei Li combined CNN and RNN [20] to generate natural language descriptions of image regions. Labeling of input includes weak sentences not classes. In the network architecture output of CNN is directly fed into RNN discarding softmax layer. Hence the network of RNN and CNN generates description for the test data. In order to produce semantically right descriptors, alignment model [21] by Karpathy is used. Now the model is trained with an output score for semantics. Object regions and the sentences are mapped onto dimensional space. This is done by using bi-directional RNN. At the highest level we obtain context of words in a sentence for the corresponding region. Thus we have descriptive labels or statements at the highest level for each region.

In the last two methods we saw that we can label the data using R-CNN. We can achieve this at a faster rate using Fast R-CNN and Faster R-CNN. Also, using combination of RNN and CNN can produce descriptive labels. These two techniques segment regions using bounding box with corresponding labels. Next, detailed segmentation can be achieved by using Conditional random fields- CRF [22] [23]. CRFs segment image into constituent semantic regions given labels. Hence R-CNNs or combination of CNN and RNN can be used for labeling and marking regions, followed by CRFs for accurate region segmentation.

For advanced face reconstruction these segmentation techniques can be used to reconstruct or retouch each part separately in detail. In applications of plastic surgery, detailed component structuring plays a very major role. For a nose job, nose has to be segmented. Then nose has to be independently reconstructed and transformed into a virtual viewing platform. Professionals can refine this model on the virtual platform before surgery to suit the patient's needs. Similarly, beauticians can modify the virtual model of customer's eyes and face with their design.

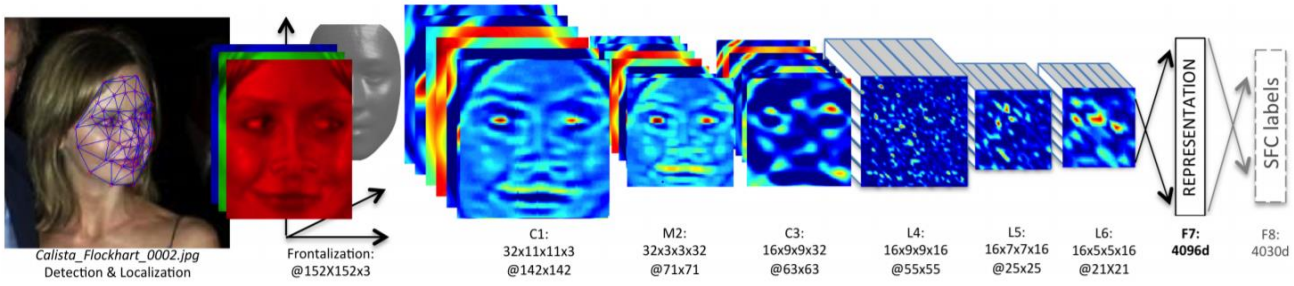


Fig. 8. Deep Face CNN architecture: In the first layer, image is decomposed into three channels for red, blue and green and resized into 152x152. In layer C1: there are 3 filters and convolution kernel of 11x11. C1 layer also has pooling layer of pooling size 2, hence images are down sampled. M2 has 32 filters with 3x3 convolution kernel. From the figure, number of filters and the size of the convolution kernel is obvious. Note that: L4 has a pooling layer.

B. Deep Face – Face Recognition

The general pipeline for face recognition involves: detect, align, represent and classify. However, scientists at Facebook modified align and represent with a 3D face model to achieve an accuracy of 97.35% on the Wild (LFW) dataset. This system [24] is as accurate as a human. In the first step alignment is used. Later on the aligned images are fed into Deep neural networks- DNN for classification. The Fully connected layer of the DNN can be modified based on number of class labels. In Fig. 2. The network has 4030 class labels.

First modification is wrt to the alignment. Image description is extracted by using LBP Histograms. From these image descriptors, Support Vector Regressor (SVR) extracts fiducial points. Using 6 fiducial points for tip of the nose, center of the eyes, and mouth location, image is cropped and aligned. In the 2D alignment, the image is scaled, rotated and translated to fit 6 similarity matrices *T*. By using this similarity matrix, the fiducial detector extracts new feature space with refined localization. The similarity matrices produce transformation which compensates in-plane rotations. For out of plane rotations 3D alignment is used.

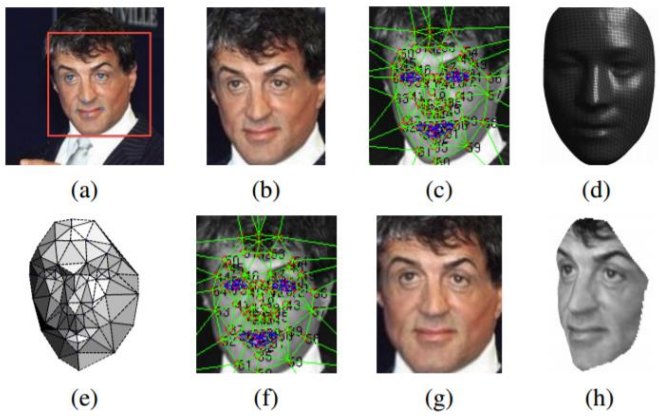
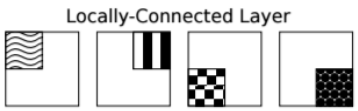


Figure 9. Alignment pipeline: (a) Face detected (b) Cropped image (c) Fiducial points (67) on the 2D-aligned crop (d) Wrapped 3D model from 2D transformations. (e) 3D camera for triangulation (f) Fiducial points (67) on the 3D aligned model for the piece-wise affine transformation. (g) Frontalized crop. (h) Side view after generating 3D model.

The average of the 3D scans from USF Human-ID database is used to generate a common 3D Morphable model. A 3D affine camera is registered for the 2D aligned crop, to wrap into 3D using 67 fiducial points. Overall, the cost function is same as least square by using Cholesky's decomposition. In the end, Frontalization relaxation is added to allow small distortions. Also, symmetricity is used to blend the components to compensate for the missing data.

The deep Neural network is trained to classify the identity of the Facebook images. The architecture is as shown in Fig above. In the figure, layers and their properties are noted. We should also understand that pooling layer that would down sample the data by half has stride of 2. Meaning the data is skipped half a time to reduce it size in half. Pooling layer is used to generalize the network for any input data, thus reducing overfitting. On the downside, there is loss in micro textures and location of the features. Layers of C1, M2 and C3 extract low level features like certain textures and edge. Pooling is used only in the first layer considering loss of information. In DeepFace these first three layers are termed as "front end adaptive pre-processing stage".

Layers of L4, L5 and L6 are locally connected [57, 58]. In locally connected each region of the feature map has different filter as shown in the figure to the left. Using this spatial temporary information of the convolution is not valid. Locally connected layers introduce a very large number of parameters. This is possible in deep face as Facebook is the biggest repository for the face images. Finally, F7 an F8 are Fully connected layers, producing class labels using softmax layer. The fully connected layers, mostly capture global information. In this network the features are normalized between 0 and one to make it immune to illumination variations. L2-normalization is used.



C. *Generating synthetic faces with deconvolutional network*

Flying chairs [61] by Alexey Dosovitskiy is a parametric CNN. The network learns 3D models of chairs by giving parameters as input. This paper uses various styles of chairs and feed them as input class. Along with this even the camera orientation and position is passed as the input. With the output being target image for the RGB network and segmented image; using the ground truth depth map. This network learns models of chairs and thus by giving camera position, orientation and style of the chair, we can get our desired chair as the output.

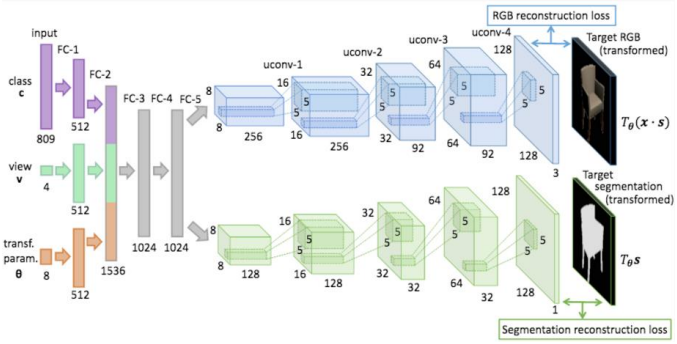


Fig. 10. Paramteric CNN Network for Flying Chair model

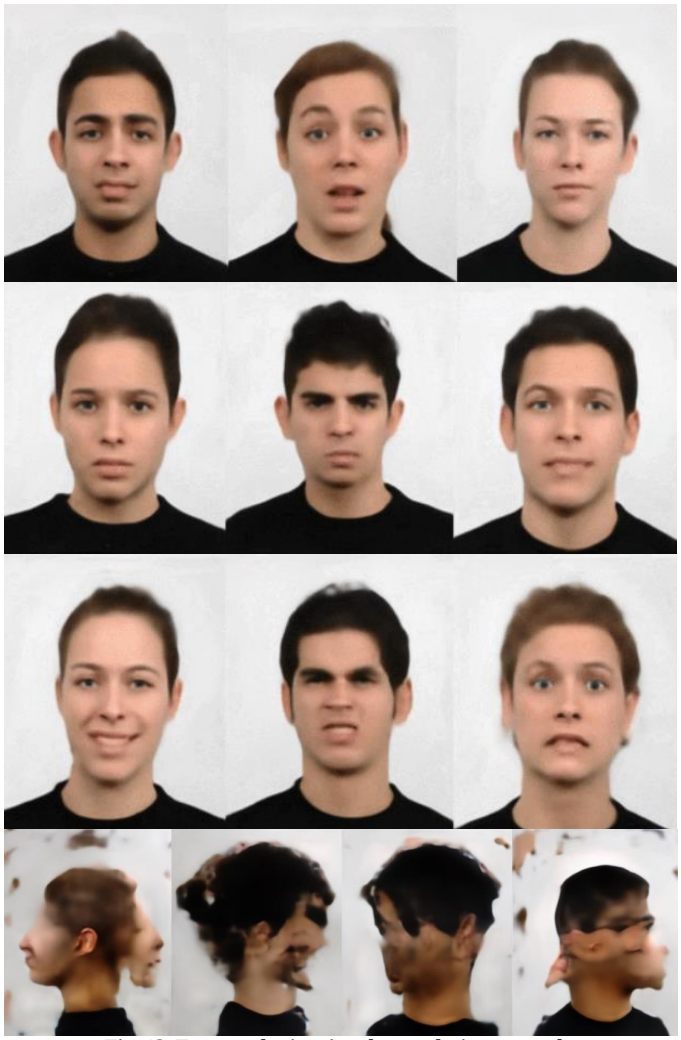


Fig. 12. Face synthesis using deconvolution networks

Fig. 11. Classifying emotions using CNN

[59] uses Radboud Faces Database to synthesis Face using Flying chairs network. Source code can be found [60]. Deconvolution is the of convolution. Here instead of feature map, images are produce from the feature map. Deconvolution layers are used in several applications like depth map generation [62], face synthesis [59], 3D modeling [61], data visualization [63]. For deconvolution we use up sampling instead of pooling layers. Batch normalization by Sergey Ioffe and Christian Szegedy [64] is used to compensate for leaky ReLU activation functions. The network of [59] is overall able to interpolate between different faces as shown the Fig. It is visually appealing to see Gif which can found in the code repository [60]. If side angle view i.e. 90° orientation is used, then illegal outputs are produced. Hence the dataset needs to be granular and large enough for the network to generate faces of all the orientations.

D. *Analyzing emotions using CNN*

To render and synthesis the human face realistically, human emotions have to be recognized and mapped on to the face. Recently, there has been a growing research on using CNNs for classifying human expressions. FER2013 dataset consists of faces with expressions. As shown by [65] disgust labelled classes create an imbalance with other labels. Hence, it can be merged with anger. Accuracy of prediction depends on the complexity of the CNNs. It is important to vary the number of CNN, fully connected and drop out layers. It is believed that if video sequence is used, CNNs can capture even non-controlled facial expression as discussed by [6].

Recent developments in this field are: work by Alizandeh and Fazel [66] not just uses raw pixel data as input, but also employs a hybrid feature strategy. In this strategy raw pixels are combined with Histogram of oriented gradients – HOG to capture more details of human expressions. Young-Hyen Byeon and Keun-Chang Kwak [67] use video data for facial emotion recognition. It uses 3D -CNN with data augmenting methods of Principal component analysis- PCA and Tensor-based Multilinear Principal Component Analysis – TMPCA for dimension reduction. Mundher Al-Shabi study hybrid model by combining SIFT and CNN [68]. Higher emotion recognition accuracy is achieved using a dataset of small size. Peter Burkert proposed usage of FeatEx Block in their DeXpression network [69]. This continuously provides feedback of the right label at each layer using ReLU activation functions. Hence correction is made at each layer which results in better accuracy.

E. Structure from depth map

Time of Flight, Structured light and Stereo technology have been used widely for Depth Map estimation. Each have these come with their own pros and cons in terms of speed of image capture, structural description and ambient light performance. Monocular cues such as: Texture and Gradient Variation, Shading, color/Haze, and defocus aid in accurate depth estimation. These are complex statistical models which are susceptible to noise. Recently, data driven approaches as in deep learning has been employed for depth estimation. These data driven approaches are less prone to noise if presented with enough data to learn coarser and finer details.

A fully automatic 2D-to-3D conversion algorithm: Deep3D [70] that takes 2D images or video frames as input and outputs 3D stereo image pairs. David Eigen from NYU proposes a single monocular image based architecture that employs two deep network stacks – Multi Scale Network [71]: one that makes a coarse global prediction based on the entire image, and another that refines this prediction locally. It is trained on real world dataset. "FlowNet: Learning Optical Flow with Convolutional Networks" [72] uses video created virtually to make the network learn motion parameters and hence forth extract optical flow. "Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches" [73] a method for extracting depth information from stereo data and their respective patches. Similar to [73] "Depth and surface normal estimation from monocular images using regression on deep features and hierarchical {CRFs}" [74] uses different scale of image patches to extract depth information. SteroConvNet [75] generates depth map and the first half of the network is shown below. Second half of the network is the mirror image of the last convolution layer, replacing convolution with deconvolution and pooling with upscaling. In Deeper Stereo ConvNet, input remains constant but architecture is modified with an extra convolution and deconvolution layer. Also, depth of the filters is increased referring to [72] in order to capture more details. Referring to [73] and [74], input stream has been increased to 6 for Patched Deeper Stereo ConvNet, by decomposing left image into 4 scaled parts. Thus, as in the referenced papers higher accuracy of the depth map is expected. Patched Deeper Stereo ConvNet predicts depth map very similar to the ground truth.

Data Driven Depth Estimation approaches would be effective provided: a) Sufficiently large descriptive labelled dataset which are rare.eg KITTI, NYU, Middlebury. b) Time and resources to train offline. C)Training time is directly proportional to the depth and the complexity of the CNN. The network model should be trained and validated for the real world data. The calibration information of the dataset is required to generate 3D Point Cloud Model.

Recently there has study related to depth estimation using unsupervised learning. In unsupervised learning ground truth is not required. Ravi Garg and his team recently came up with a very accurate unsupervised technique [79] which works similar to Auto-encoders. On similar lines, Clement Godard and his teams came up with the network [80] which produce stereo image without using ground truth depth map.

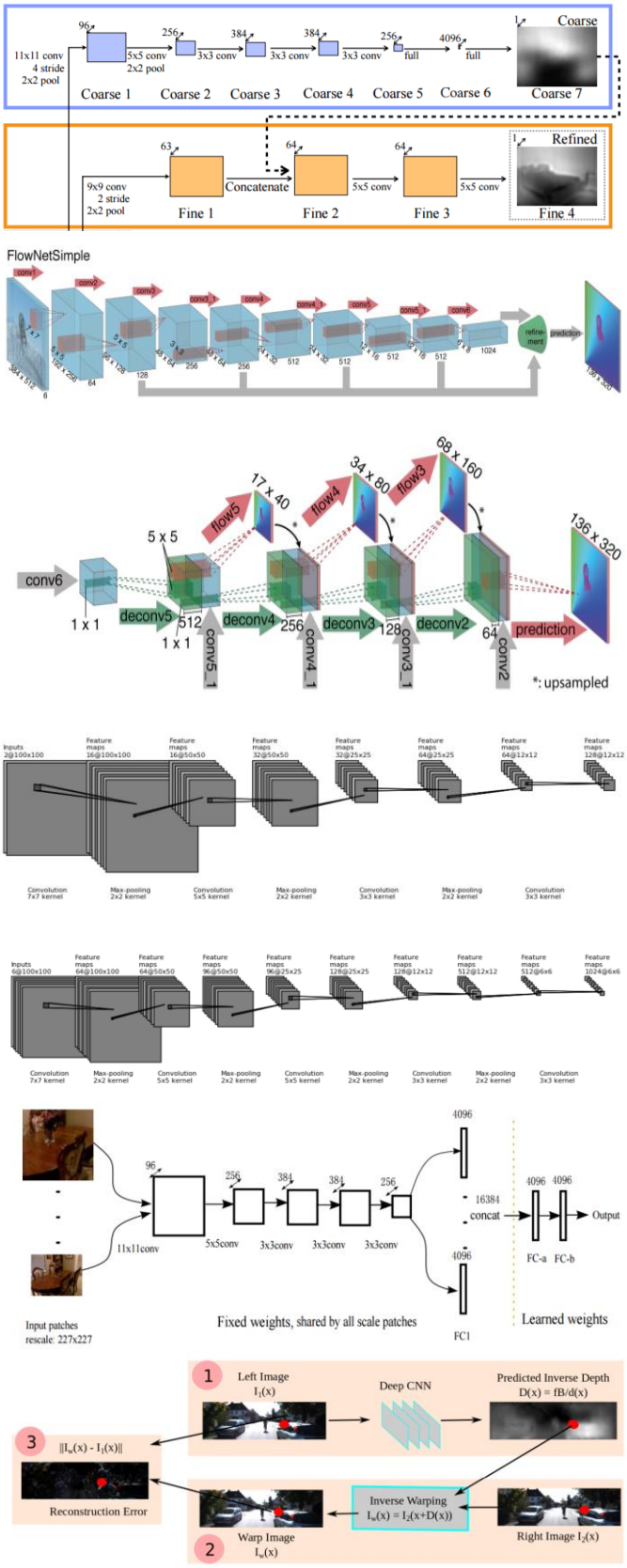


Figure 13. From top: (a) Multi-scale (b) Optical flow conv and deconv (c) SteroConvNet (d) Patched Deeper SteroConvNet (e) Multiscale net using CRF cost function (f) Unsupervised Net

V. SHAPE FROM X

Extracting shape from image cues such as: Shading, Texture, Contour, Stereo, Focus, Motion and Silhouettes is considered as Shape from X. Even if these methods are stand-alone techniques to extract shape, fusing many will produce better results. Table from [12] is shown below:

SHAPE FROM X	#INPNT IMAGES
SHADING TEXTURE CONTOUR	SINGLE
STEREO FOCUS	MANY
MOTION SILHOUETTES	VIDEO

Table. 1. Shape from X methods

A. *Shape from Shading*

Shape from shading was first stated by Horn in 1970 and later refined by Horn and Brooks in their book [25]. Considering a virtual light source, the angle between the surface normal and the light can be estimated. After this step, the problem simplifies to recovering shape from the known intensity variations.

Survey by Zhang [26], provides us with 6 shape from shading techniques and evaluates their performance. 1) Earliest technique was the Minimization approaches recovering surface gradients was studied by [27]. By introducing brightness and smoothness constraints the problem of two unknowns for the surface gradient is resolved. 2) Propagation method is also known as Horn's characteristic strip method [28]. Here, the shape information is propagated along strips, assuming no crossovers. Strips are interpolated to obtain dense model. 3) Local approaches by Pentland [29] linearize the reflectance map to solve for the shape. In here at every point the surface is assumed to vary spherically. 4) Whereas, Lee and Rosenfeld [30] calculated tilt of the surface using first derivative of the intensity. 5) and 6) Linear approaches reduce the non-linear problem into linear using linearization of the reflectance map using Pentland [31] and Tsai and Shah [32]. It is assumed that the lower order components in the reflectance map dominates. Pentland's method using Fourier transform to the linear function. Tsai's method does linear approximation on depth using iterative Jacobi scheme.

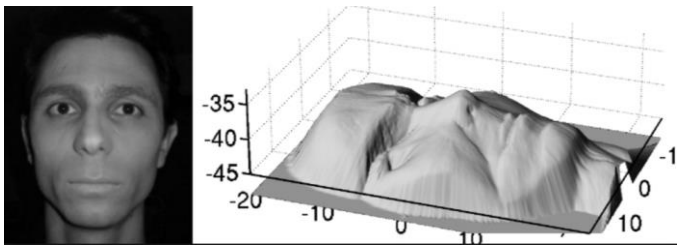


Fig. 15. Shape from Shading

Challenges produced in these techniques include: 1) Zheng and Chellappa: errors around cheeks and lips. 2) Lee and Kuo: self-shadows are not accounted for. 3) Bichsel and Pentland: The algorithm is very fast and produces best result when light source is on the side. 4) Lee and Rosenfeld: Doesn't perform well for images with non-spherical surfaces. 5) Pentland: Problems when reflectance changes non-linearly. 6) Tsai and Shah: very sensitive to noise.

Work by Prados [33] shows that shape from shading can be an understood problem if the attenuation term of the illumination is taken into account. Also, Zhu and Shi [34] show us how to resolve local ambiguities in shape from shading by recognizing mountains through global view.

B. *Shape from texture*

This technique cannot be used by itself for face reconstruction. But it stands as a supporting monocular cue for accurate face modeling. Shape from shading was first started by Gibson in his work "The Perception of the Visual World", stating texture is an important visual cue.

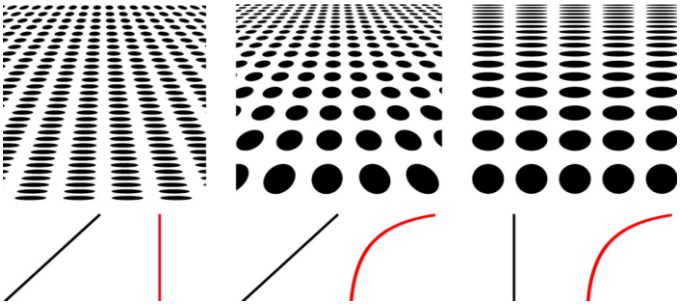


Fig. 14. Shape from Texture

Texture is assumed homogeneous by Witkin [35] and Brown [36] which is not always true if the texture rotates on the surface. Many a times assumption with the camera model induces errors. Hence many shape from texture methods for orthographic camera [37] and perspective [38], [39] have been studied. On Blender platform, we can produce 3D model using depth map and the textured image. Firstly, displacement matrix is specified for the depth map, later on texture is interpolated over the displaced image. This is an instance of shape from texture, as we produce 3D model by using texture of the ground truth depth map.

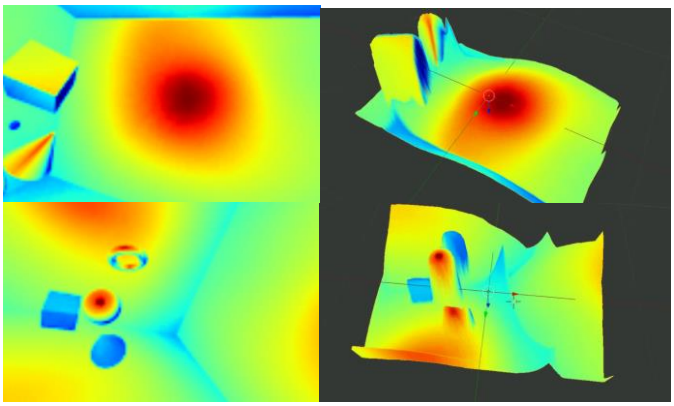


Fig. 16. 3D model using depth map on Blender

C. Shape from contour

This technique [117] was first introduced by Takeo Kanade in 1980. Similar 3D-projections can be produced by using object with different shape. Image irregularities have to be mapped into shape constraints heuristically using skew symmetries. Another novel representation of skew symmetries was introduced by Shiu-Yin [119] using rotation, midpoint estimation and Hough transform. Yet another symmetry [120] was studied by Fatig Ulupinar and Ramakant Nevatia. Symmetries that limb boundaries of Straight Homogeneous Generalized Cylinders – SHGC and Constant Cross-section Generalized Cylinders – CGCs are studied. Doug DeCarlo used suggestive contours [122] which are more accurate representation than contour alone.

Recently, Dejan Todorovic studied how the shape from contour and shape from shading are correlated [121]. Uses of Shape from shading is usually diminished due to illumination variance. However, if both were to be combined, users were able to discern 3D surface.

Fig. 17. Shape from Contour

D. Shape from stereo

Stereo image, usually consists of left and right. Fundamental matrix using camera intrinsic and extrinsic parameters are used to estimate disparity and hence the depth map. 3D modeling using depth map is direct in the specified direction. Processing this information in many orientations will help generate accurate 3D model. This technique is also called Photometric Stereo and is used by Roth et al in their recent work [89, 91] to produce highly accurate 3D face model using just one image.

Stereo- 3D Reconstruction algorithms consider orthogonal or non-orthogonal images. In first case of orthogonal we have two images of frontal and profile view. The next step would be to extract feature points corresponding to both the views. Key features usually include the ears, eyes, nose, eyebrows, mouth, and hair outlines. One of the popular very old method [123, 124, 125] uses this technique. Pandzic [126, 127] termed these as Facial definition points- FDP and in MPEG-4 format. [123] uses displacement vectors to interpolate feature vectors. [124] uses wired model and hence established one-to-one correspondence from 2D to 3D. For non-feature point region Dirichlet Free Form Deformation [128] is used. [125] uses Radial Basis function to deform the generic head model.

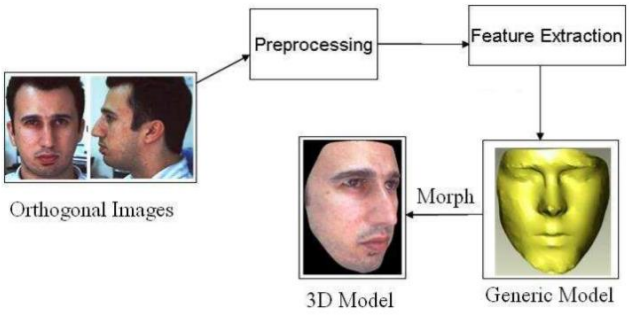


Fig. 18. Face mode from stereo using orthogonal image

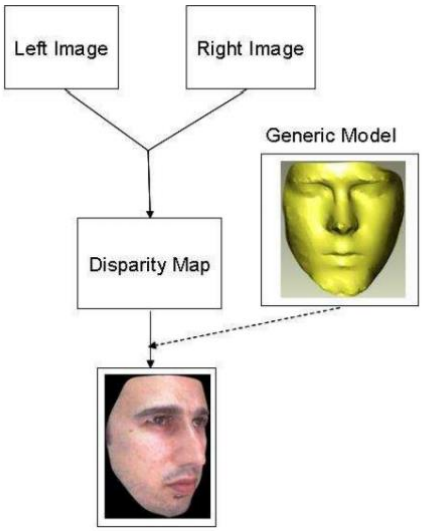


Fig. 19. Face mode from stereo using non-orthogonal image

The non-orthogonal Stereo technique require two cameras to capture left and right images. Using fundamental matrix of the camera disparity between the pixel correspondence in the two images is calculated. The depth map produced can be used in 3D reconstruction. Depth map can also be used by using Deep Neural Network as discussed in previous sections. Cameras can be placed either in horizontal or vertical epipolar directions [129, 130, 131, 48]. As studied by [14], in face reconstruction vertical epipolar configuration suits the best.

Several studies have been executed to generate accurate disparity maps. Few of them: Energy maximizing snakes-active contours [132], parallel stereo algorithm [133], simulated annealing for depth estimation [129], volumetric approach to computer disparity [130]. Comparisons of different stereo algorithms can also be found in [131, 134, 135].

Consecutive frames can also be treated as stereo input. 3D reconstruction is possible by using Bundle adjustment [136]. In stereo non-orthogonal structured light is used to improve efficiency. But, this also adds hardware constraint.

E. *Shape from focus*

Highly cited paper [137] by Sree K. Nayar gives us an idea of the importance of using Shape from focus. Taking an orthographic homogeneous image using pinhole means at any instance, camera can focus at only one point. Nayar extended his work even to Rough surface [138] by using sum-modified laplacian – SML and interpolation using Gaussian kernels. Now, there are many other techniques [139] used to evaluate shape from focus (Refer Fig below).

Focus operator	Abbr.	Focus operator	Abbr.
Gradient energy	GRA2	Gray-level variance	STA3
Gaussian derivative	**GRA1**	Gray-level local variance	**STA4**
Thresholded absolute gradient	GRA3	Normalized gray-level variance	STA5
Squared gradient	GRA4	Modified gray-level variance	**STA6**
3D gradient	GRA5	Histogram entropy	STA7
Tenengrad	GRA6	Histogram range	STA8
Tenengrad variance	**GRA7**	DCT energy ratio	**DCT1**
Energy of Laplacian	LAP1	DCT reduced energy ratio	**DCT2**
Modified Laplacian	LAP2	Modified DCT	**DCT3**
Diagonal Laplacian	LAP3	Absolute central moment	**MIS1**
Variance of Laplacian	**LAP4**	Brenner's measure	**MIS2**
Laplacian in 3D window	LAP5	Image contrast	**MIS3**
Sum of wavelet coefficients	**WAV1**	Image curvature	**MIS4**
Variance of wavelet coefficients	**WAV2**	Hemli and Scherer's mean	**MIS5**
Ratio of the wavelet coefficients	WAV3	Local binary patterns-based	**MIS6**
Ratio of curvelet coefficients	WAV4	Steerable filters-based	**MIS7**
Chebyshev moments-based	**STA1**	Spatial frequency measure	**MIS8**
Eigenvalues-based	**STA2**	Vollath's autocorrelation	**MIS9**

Fig. 20. Shape from focus methods

Similar to shape from focus, we can also use shape from defocus. 3-dimensional shape can be inferred from a set of defocused images [140]. We use Singular Value decomposition – SVD on images at various depths. This can be done in real time and holds true for any camera.

F. *Shape from motion*

Shape from motion uses video or images captured from different viewpoints. In other terms it is also referred to as Structure from motion. Various structure from motion methods are described in the popular 3D modeling system, described later in this paper. Nikos Sarris in their method [107] uses sequence of image for 3D face modeling. Yin et al. [109] could produce higher resolution model. Ypsilos et al. [106] produced 3D talking model using video and speech. Chowdhury et al. [103] combined Shape from motion and generic model. Xin et al [104] could produce real time 3D face model by tracking 500 feature points at any given instance.

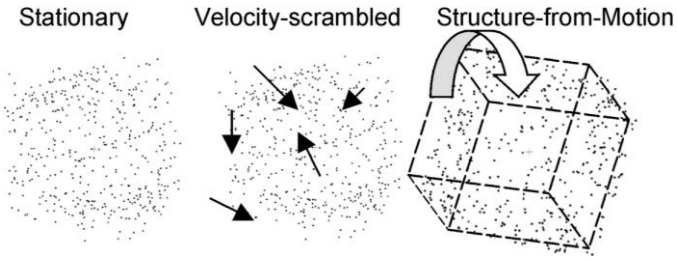


Fig. 21. Structure from motion

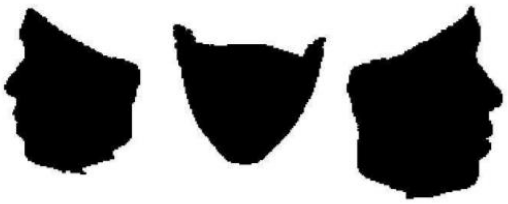


Fig. 22. Shape from Silhouette

G. *Shape from Silhouette*

Shape from silhouettes auses sequence of images or video, because silhouette at different orientation aids in 3D reconstruction. It provides as an alternative to structured lighting technique, capturing surface details. This has been used with other techniques for 3D face modeling.

H. *Combing Cues*

Combining various shape from X cues helps to produce 3D face models depending on the applications. It is assumed that the cue will make up for the cons of the other cue. Suggestive contours by Doug DeCarlo [122] use contours in conjunction with silhouette to produce realistic 2D face models. White and Forsyth combine shading and texture to produce unambiguous normals [142]. Deformable model can be generated by using stereo and shading [143].

Moghaddam et al. [144, 145] combined silhouette and Morphable model. Distance minimization is used on silhouette and Morphable model. Similar work by Wang et al. [146] combines 3D PCA based model in conjunction with silhouette. Also texture information is used in computing spherical harmonics basis texture. Ilic and Fua [147] use combination of stereo and silhouette.

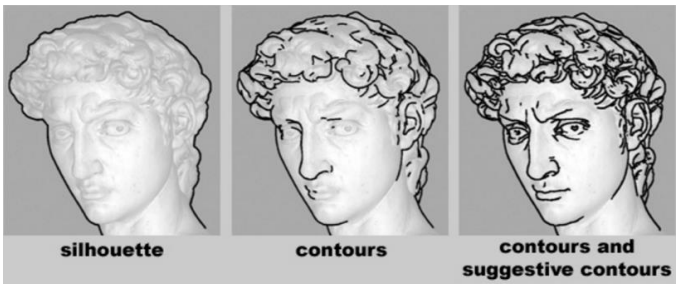


Fig. 23. Combing silhouette and contours

VI. Model Based Reconstruction

In this technique a 3D generic face model is used as the initial face model. Almost all face reconstruction methods use 3D Morphable model. Even DeepFace [4] that we discussed under learning methods uses 3D Morphable model as its intermediate step. Usually this 3D model is generated by using 3D Face database such as MPI or USF.

A. Generating Morphable Model

To generate Morphable model [40] we use 3D Face Database. This work is similar to the study [56] by DeCarlos. Correspondence is established between all the faces. However, unlike [56], no surface interpolation techniques [57, 58, 59] are required. It is assumed that valid texture values are equal to the number of vertices. Parameters to control shape and texture of the face are dynamic in nature. Probability of likelihood of the parameter variance controls the likelihood of appearance. Principal component analysis- PCA is used to reduce dimension of the feature. Arbitrary face can be generated by varying control parameters for shape and texture. To generate Morphable model, first we segment eyes, nose, mouth, ears, and surrounding region. Thus the complete face vector space is sub-divided into separate space for each segment. Each part is processed separately which results in accuracy and speed-up. Then parts are combined with blending to produce 3D Morphable model.

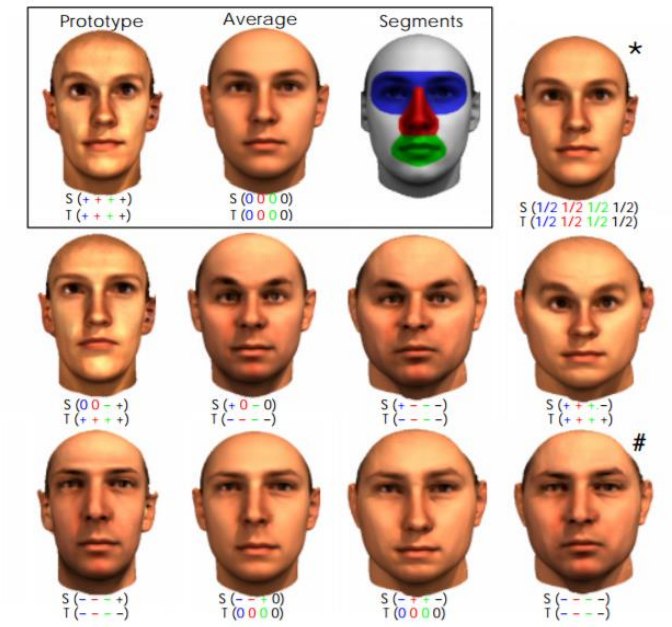


Fig. 24: Single prototype can be used to produce variety of faces using Morphable model. Segments correspond to eye, mouth, nose and surrounding area. Parameters can be varied independently for shape and texture on these four segments to produce distinctive face models. * is the standard Morphable model which is located in between average and the prototype. If the differences are subtracted from average, then we get 'anti'-face (#). The deviation of a prototype from the average is added (+) or subtracted (-) from the average.

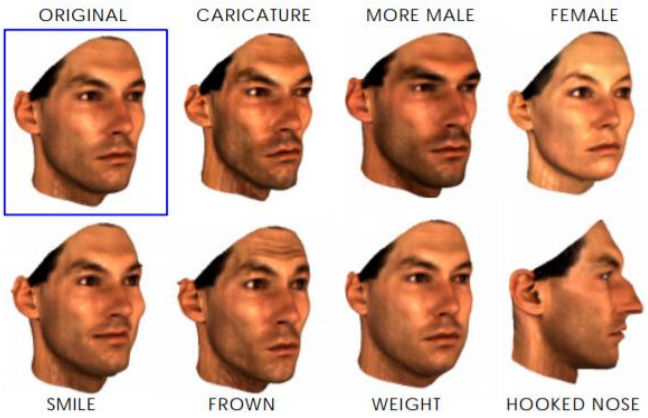


Fig. 25: Variation of facial attributes

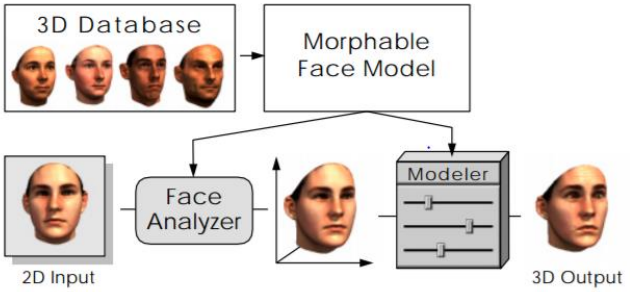


Fig. 26: 3D Morphable model as intermediate step

Facial attributes are specified at certain locations by using hand-labeled example set. At these locations, feature vector and texture parameters are defined. Varying these will change specific attribute of the face keeping the other attributes constant. Invariant attributes are difficult to isolate. In this model, facial attributes of gender, darkness of eyebrows, double chins hooked versus concave noses, fullness of faces, and darkness of eyebrows are variable. In caricatures special attribute of distinctiveness is varied. This attribute is not included in this work, however can be easily introduced. Several works on caricature modeling are: [82], [83], [84], [85], [86], and [87].

B. Others

There have been many more researchers bent on producing efficient morphable mode. Work by Xiaowen Wang [81] produced Morphable model using multiple views. Using single view introduces ambiguities in terms of actual model as only one depth map is used. Therefore, more certain model can be generated using multi angled views. High resolution face images can be rotated using Thomas Vetter and Tomaso Poggio's linear transformation [88]. As rotation is a linear transformation, it can be easily learned from basic set of 2D prototypical view. In next section, we will see popular 3D reconstruction methods.

VII. POPULAR FACE RECONSTRUCTION MODELS

This section is missing in every survey paper on synthetic biometric or face synthesis. Thus we provide researches with the direction in which face synthesis and reconstruction has been moving. First of all let us list all the work done in this area in chronoligical fashion. We skip papers discussed earlier.

Year	Research Papers
1980	Todd et al. [46]
1989	Thalmann et al. [45], Lewis et al. [47]
1992	DiPaola et al. [44]
1993	Akimoto et al. [123]
1996	Lengagne et al [133]
1997	Lee et al. [124]
1998	Pighin et al. [48], DeCarlos et al. [43]
1999	Blanz et al. [40]
2001	Sarris et al. [107], Chen et al. [130]
2003	Yin et al[1]. [102], Ansari et al. [110]
2004	Ypsilos et al. [106], Xing et al. [108]
2005	Chowdhury et al. [103], Jiang et al. [104], Zhang et al[1] [105], Park et al. [125], Xin et al. [
2009	Paysan et al. [100], Chouvatut et al. [99]
2010	Wang et al. [81]
2011	Shlizerman et al [98]
2014	Kazemi et al. [93], Behlim et al. [94]
2015	Roth et al[1] [91], Liu et al. [92]
2016	Roth et al[2] [89], Piotraschke et al. [90], Garrido et al. [95]

Table. 2. Face Reconstruction methods

Todd et al. [46] in his paper describes growth of human head. Biologically, there is a certain pattern in which human head develops. Many constraints such as: Cardiodal strain, Spiral strain, Affine shear, Reflected shear, Rotation and non-change are studied. Finally, basic pattern of human growth matches Cardiodal strain.

Thalmann et al. [45] describes face synthesis and animation at a very crude level. A very basic level, face synthesis is explored using Local transformation and shape interpolation.

Lewis et al. [47] came up with their own noise function with better efficiency and control over noise power spectrum. The new algorithm suits the purposes where in high noise is required. Face synthesis from the point of view of stochastic gradient and solid texturing is used.

DiPaola et al. [44] is a face animation model varying face parameters. In the parametric approach, facial model is generated from parameters which act as vertices. This model uses 100s of parameters to model the face and animate it. Even though warping models and twisting models are applied, this technique is very basic and doesn't produce realistic face models.

Pighin et al. [48] at Microsoft developed user-assisted technique to produce 3D-face model. User recovers the camera pose and marks the orientations. Using various camera images and images with different human expressions, 3D face is generated by blending textures. To generate transitions between various expressions, various models are morphed with texture interpolation. On the downside this is not an automatic technique and requires way too many images to model a face.

Anthropometry is the biological science of human body measurement. DeCarlos et al. [43] uses Farkas's inventory on Anthropometric landmarks on the face defining proportions. B-splines are used for surface representation with linear and non-linear constraints. Thin-plate approximates bending of surfaces. However, model produced by this method is not as accurate as the model produced by using landmarks.

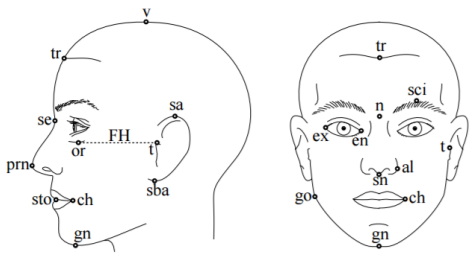


Fig. 1: Anthropometric landmarks on the face

Nikos Sarris in their method [107] produce real time 3D-human head using camera. Once users face is recognized, he Is asked to place his head in the highlighted area. By blinking, the system is signaled to capture feature points and the profile view. Non- rigid transformations are applied on the rigid 3D-model. If the rigid model is not accurate then this technique will fail to produce realistic results. Also the 3D- model synthesized is very grainy.

Yin et al[1]. [109] is similar to the previous case. However, accuracy and resolution of the 3D model is higher. While 3D modelling orientations are referenced, producing higher accuracy. Higher resolution is generated by using Hyper-resolution image enhancement and adapting 3D model for each resolution. Improved accuracy for face detection. This is the base for 3D DeepFace [4] model which uses CNN achieving best face recognition.

Ansari et al. [110] uses orthographic views of profile and frontal view with generic model to produce 3D face model. Feature points are extracted from image, which are in turn used to deform the generic 3D model. Procrustes Analysis minimizes the distance between feature points and 3D vertices. Finally, local deformation makes the model more realistic. Quality achieved is higher than those produced in the earlier papers.

Ypsilos et al. [106] could produce 3D talking model by using capture system as shown in the figure and speech. Accurate non-rigid 3D -video is produced by using optical flow and displacement map representation. This application is very appealing and can be the future of modern video conferencing.

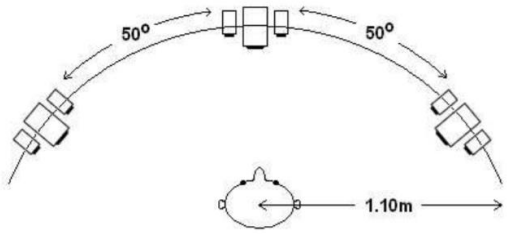


Fig. 1: Capture system

Xing et al. [108] is the first paper which uses single image to produce 3D face model. SVM is used for face detection and regressor is used to detect individual features of the class. Morphable model is used for model fitting with linear projections. Classification based refinement achieves higher accuracy than the regression system.

Chowdhury et al. [103] uses generic model slightly in a different way to make the model less dependent on the model properties. Energy function is used map 3D model using Structure from motion- SFM and generic model. Thus two of the 3D modeling techniques are combined to produce higher accuracy.

Jiang et al. [104] proposes integrated approach to model the face under variations of Pose, Illumination and Expression - PIE. Improved integration of various methods helps solve this challenge. As evaluation metric, face recognition is conducted on images and videos varying in PIE. However, this model works for only frontal face with small set of deviation in the orientation.

Zhang et al[1] [105] has similar condition as the previous case and assumes the surface to be Lambertian. Spherical harmonics is used to approximate at different lighting conditions. Hence accurate 3D model is generated under varying lighting conditions. This model can also remove cast shadows. However, the assumption that input and the target have the same cast shadows introduce error in the system.

Paysan et al. [100] uses MRI scans to model human skull. Using this 3D models of the face are structured. But due to variations between the skull and face, this system is not completely accurate.

Chouvatut et al. [99] is an improvement over previous techniques to model face using video sequences. It uses Gradient descent and Powel's Multidimensional minimizations. Using this method even the occluded regions can be accurately modeled. But, there is a limit as to how much of occlusion can be accurately depicted.

Shlizerman et al. [98] is more detailed in its study of producing 3D model using a single image. Here only a single reference model is used as its 3D Morphable model. Input image itself molds the 3D Morphable model to suit the application. Images in USF dataset with controlled and uncontrolled lighting conditions are tested for very good accuracy results. In overall, it would be better if the single 3D reference was molded based on probability conditions for the dataset.

General reconstruction techniques using depth map are computationally expensive and are non-linear wrt construction error. Kazemi et al. [93] therefore used: random forest to produce noisy correspondence fields and then fine-tuned using stochastic optimizations. Works well in low light, but is robust only slight occlusions.

Raster-Stereography is used in imaging vertebrae. Behlim et al. [94] used this technique for 3D face modeling. This technique is very similar to estimation and modeling using structured light [110, 111]. Like structured light it is easily affected by different lighting conditions.

Roth et al[1] [91] work is one of the most recent pioneering work. 3D face model is represented as watertight triangulated surface. The photometric-stereo method developed involves: 1) 3D model provides flexibility from all pose directions. 2) Combination of landmark constraints and photometric-stereo based normal is used in the face reconstruction. In this model, landmarks are not detected in automatic fashion. Further work by Roth et al. [89] is an improvement over this. In [89] personalized template is derived by fitting 3D Morphable model. Novel photo-stereometric scheme helps refine the model course to fine. Fusing [91] and [89] could produce in even better results.

Liu et al. [92] doesn't use 3D Morphable model as it is ineffective in real time applications. Here it is treated as a regressor problem and thus a cascade of regressor run offline to reduce the deviations of the input image landmarks and the landmarks from the reconstructed face. One limitation of this method is that input landmarks should be provided.

Piotraschke et al. [90] developed a very intelligent system which doesn't need labelling from outside world for better modeling. Where in quality is measured automatically in the system which helps to refine the model. Hence, this is a complete automatic system without requiring human interference.

CONCLUSION

This is the most detailed survey paper up to date. We include very basic methods like learning, shape from X, shape from shading to combining various complex methods. Popular methods discussed in this section will help researchers get on to speed with the topic.

In our future work, we implement various face reconstruction algorithms and provide comparative studies.

ACKNOWLEDGMENT

I would like thank Dr. Damon Woodard for giving me this opportunity and helping me focus on my passion for 3D face reconstruction.

REFERENCES

[1] R. Cappelli, SFinGe: Synthetic Fingerprint Generator, In Proc. Int. Workshop on Modeling and Simulation in Biometric Technology, Calgary, Canada, pp. 147–154, June 2004

[2] Face Genration software: FaceGen, http://facegen.com/index.htm

[3] Oben 3D Virtual Avatar, https://techcrunch.com/2016/11/03/oben-nabs-7-7m-series-a-as-it-looks-build-a-more-human-vr-avatar/

[4] "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", Taigman, Yaniv and Yang, Ming and Ranzato, Marc'Aurelio and Wolf, Lior, CVPR '14

[5] "Coupled 3D Reconstruction of Sparse Facial Hair and Skin", Thabo Beeler, Bernd Bickel, Gioacchino Noris, Paul Beardsley, Steve Marschner, Bob Sumner, Markus Gross, Disney Research Zuric.

[6] S. N. Yanushkevich, "Synthetic Biometrics: A Survey", The 2006 IEEE International Joint Conference on Neural Network Proceeding, 2006

[7] "Biometric Inverse Problems", Svetlana N. Yanushkevich, Adrian Stoica, Vlad P. Shmerko, Denis V. Popel, May 5, 2005 by CRC Press Reference - 416 Pages - 75 B/W Illustrations , ISBN 9780849328992 - CAT# 2899

[8] Nicholas M. Orlans, Douglas J. Buettner and Joe Marques, "A Survey of Synthetic Biometrics: Capabilities and Benefits", Proceedings of the International Conference on Artificial Intelligence, {IC-AI} '04, June 21-24, 2004, Las Vegas, Nevada, USA, Volume 1

[9] Douglas J. Buettner and Nicholas M. Orlans, "A Taxonomy for Physics Based Synthetic Biometric Models", Automatic Identification Advanced Technologies, IEEE Workshop.

[10] Kazuhiko Sumi and Takashi Matsuyama, "Privacy Protection of Biometrics Evaluation Database – A Preliminary Study on Synthetic Biometric Database"

[11] Mihailescu Marius Iulian, "Direct Problems and Inverse Problems in Biometric Systems", Journal of Knowledge Management, Economics and Information Technology, 2013

[12] M. Judith Leo and D. Manimegalai, "3D modeling of human faces- A survey", 3rd International Conference on Trendz in Information Sciences Computing (TISC2011), 2011.

[13] Widanagamaachchi, W.N. and Dharmaratne, A.T., " 3D Face Reconstruction from 2D Images - A Survey", Digital Image Computing: Techniques and Applications (DICTA), 2008}, title={3D Face Reconstruction from 2D Images, 10.1109/DICTA.2008.83

[14] STYLIANOU, GEORGIOS and LANITIS, ANDREAS, "IMAGE BASED 3D FACE RECONSTRUCTION: A SURVEY", International Journal of Image and Graphics, 2009.

[15] Deepti Chandra and Sanjeev Karmakar and Rajendra Hegadi, "Article: Techniques of Facial Synthesis: A Comprehensive Literature Review", International Journal of Computer Applications, 2013.

[16] Ross B. Girshick and Jeff Donahue and Trevor Darrell and Jitendra Malik , "Rich feature hierarchies for accurate object detection and semantic Segmentation", CoRR, 2013

[17] https://github.com/mitmul/chainer-faster-rcnn

[18] Ross B. Girshick ,"Fast {R-CNN}", CoRR, 2015

[19] Shaoqing Ren and Kaiming He and Ross B. Girshick and Jian Sun, "Faster {R-CNN:} Towards Real-Time Object Detection with Region Proposal Networks", CoRR, 2015

[20] Andrej Karpathy and Fei Fei Li, Deep Visual-Semantic Alignments for Generating Image Descriptions, CoRR, 2014

[21] Andrej Karpathy and Armand Joulin and Fei Fei Li, Deep Fragment Embeddings for Bidirectional Image Sentence Mapping, CoRR 2014

[22] Jakob Verbeek, William Triggs. Scene Segmentation with CRFs Learned from Partially Labeled Images. John C. Platt and Daphne Koller and Yoram Singer and Sam Roweis. *NIPS 2007 - Advances in Neural Information Processing Systems*, Dec 2007, Vancouver, Canada. MIT Press, 20, pp.1553-1560, 2008

[23] Nils Plath, Marc Toussaint, and Shinichi Nakajima. 2009. Multi-class image segmentation using conditional random fields and global classification. In *Proceedings of the 26th Annual International Conference on Machine Learning* (ICML '09). ACM, New York, NY, USA, 817-824. DOI: http://dx.doi.org/10.1145/1553374.1553479

[24] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH,2014, pp. 1701-1708.

[25] K. P. Horn, and M. J. Brooks, Shape from Shading, MIT Press, Cambridge, Massachusetts, 1989.

[26] S R. Zhang, P. Tsai, 1. Cryer, and M. Shah, "Shape from Shading: A Survey", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), pp. 690-706, 1999.

[27] K. Ikeuchi and B.K.P. Horn. Numerical shape from shading and occluding boundaries. Artificial Intel ligence, 17(1-3):141{184, 1981.

[28] B. K. P. Horn. Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View. PhD thesis, MIT, 1970.

[29] A. P. Pentland. Local shading analysis. IEEE Transactions on Pattern Analysis and Machine Intel ligence, 6:170{187, 1984.

[30] C.H. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. Articial Intel ligence, 26:125{143, 1985.

[31] A. Pentland. Shape information from shading: a theory about human perception. In Proceedings of International Conference on Computer Vision, pages 404{413, 1988.

[32] P.S. Tsai and M. Shah. Shape from shading using linear approximation. Image and Vision Computing Journal, 12(8):487{498, 1994.

[33] Emmanuel Prados, Olivier Faugeras. Shape from Shading: a well-posed problem?. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'05*, Jun 2005, San Diego, United States. IEEE, 2, pp.870-877, 2005

[34] Qihui Zhu and Jianbo Shi , "Shape from Shading: Recognizing the Mountains through a Global View", 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006

[35] A.P. Witkin. Recovering surface shape and orientation from texture. AI, 17(1-3):17– 45, August 1981.

[36] L.G. Brown and H. Shvaytser. Surface orientation from projective foreshortening of isotropic texture autocorrelation. PAMI, 12(6):584–588,June 1990.

[37] D.A. Forsyth. Shape from texture without boundaries. In ECCV02, page III: 225 ff., 2002.

[38] M. Clerc and S. Mallat. The texture gradient equation for recovering shape from texture. PAMI, 24(4):536–549, April 2002.

[39] J.Y. Jau and R.T. Chin. Shape from texture using the wigner distribution. CVGIP, 52(2):248–263, November 1990

[40] Blanz, Volker and Vetter, Thoma, "A Morphable Model for the Synthesis of 3D Face", Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Technique, SIGGRAPH '99

[41] F.I. Parke. Computer generated animation of faces. In ACM National Conference. ACM, November 1972.

[42] F.I. Parke. A Parametric Model of Human Faces. PhD thesis, University of Utah, Salt Lake City, 1974.

[43] D. DeCarlos, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In Computer Graphics Proceedings SIGGRAPH'98, pages 67–74, 1998

[44] S. DiPaola. Extending the range of facial types. Journal of Visualization and Computer Animation, 2(4):129–131, 1991.

[45] N. Magneneat-Thalmann, H. Minh, M. Angelis, and D. Thalmann. Design, transformation and animation of human faces. Visual Computer, 5:32–39, 1989.

[46] J. T. Todd, S. M. Leonard, R. E. Shaw, and J. B. Pittenger. The perception of human growth. Scientific American, 1242:106–114, 1980

[47] J. P. Lewis. Algorithms for solid noise synthesis. In SIGGRAPH '89 Conference Proceedings, pages 263–270. ACM, 1989.

[48] F. Pighin, J. Hecker, D. Lischinski, Szeliski R, and D. Salesin. Synthesizing realistic facial expressions from photographs. In Computer Graphics Proceedings SIGGRAPH'98, pages 75–84, 1998.

[49] B. Guenter, C. Grimm, D. Wolf, H. Malvar, and F. Pighin. Making faces. In Computer Graphics Proceedings SIGGRAPH'98, pages 55–66, 1998.

[50] Y.C. Lee, D. Terzopoulos, and Keith Waters. Constructing physics-based facial models of individuals. Visual Computer, Proceedings of Graphics Interface '93:1–8, 1993.

[51]] D. Terzopoulos and Keith Waters. Physically-based facial modeling, analysis, and animation. Visualization and Computer Animation, 1:73–80, 1990.

[52] N. Magneneat-Thalmann, H. Minh, M. Angelis, and D. Thalmann. Design, transformation and animation of human faces. Visual Computer, 5:32–39, 1989.

[53] Keith Waters. A muscle model for animating three-dimensional facial expression. Computer Graphics, 22(4):17–24, 1987.

[54] S. Platt and N. Badler. Animating facial expression. Computer Graphics, 15(3):245–252, 1981.

[55]] F. I. Parke and K. Waters. Computer Facial Animation. AKPeters, Wellesley, Massachusetts, 1996

[56] https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

[57] K. Gregor and Y. LeCun. Emergence of complex-like cells in a temporal product network with local receptive fields. arXiv:1006.0448, 2010

[58] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In CVPR, 2012.

[59] https://zo7.github.io/blog/2016/09/25/generating-faces.html

[60] https://github.com/zo7/deconvfaces

[61] Alexey Dosovitskiy and Jost Tobias Springenberg and Thomas Brox, "Learning to Generate Chairs with Convolutional Neural Networks", CoRR, 2014

[62] https://github.com/ArpitaSTugave/Depth-Estimation-using-CNN

[63] Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks" , CoRR, 2013

[64] Sergey Ioffe and Christian Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", CoRR, 2015

[65] https://github.com/JostineHo/mememoji

[66] "Convolutional Neural Networks for Facial Expression Recognition", Shima Alizadeh and Azar Fazel, Stanford University , 2016

[67] "Facial Expression Recognition Using 3D Convolutional Neural Network", Young-Hyen Byeon and Keun-Chang Kwak, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 12, 2014

[68] Mundher Al Shabi and Wooi Ping Cheah and Tee Connie, Facial Expression Recognition Using a Hybrid {CNN-SIFT} Aggregator, CoRR, 2016

[69] Peter Burkert and Felix Trier and Muhammad Zeshan Afzal and Andreas Dengel and Marcus Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression RecognitioN", CoRR 2015

[70] "Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks" Junyuan Xie, Ross Girshick, Ali Farhadi,University of Washington.

[71] "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network" David Eigen, Christian Puhrsch, Rob Fergus Dept. of Computer Science, Courant Institute, New York University.

[72] "FlowNet: Learning Optical Flow with Convolutional Networks", A. Dosovitskiy and P. Fischer, ICCV , 2015.

[73] "Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs" by Bo Li1, Chunhua Shen , Yuchao Dai , Anton van den Hengel, Mingyi He, IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15).

[74] "Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches" by Jure Zbontar ,University of Ljubljana Vecna ,Yann LeCun, Journal of Machine Learning Research 17 (2016).

[75] https://github.com/LouisFoucard/StereoConvNet

[76] Depth Estimation using Monocular and Stereo Cues, Ashutosh Saxena, Jamie Schulte, Andrew Y.ng, IJCAI'07 Proceedings of the 20th international joint conference on Artifical intelligence Pages 2197-2203

[77] High-Accuracy Stereo Depth Maps Using Structured Light, Daniel Scharstein ,Middlebury College, In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)

[78] https://indico.io/blog/exploring-computer-vision-convolutional-neural-nets/

[79] Ravi Garg and Vijay Kumar B. G and Ian D. Reid, "Unsupervised {CNN} for Single View Depth Estimation: Geometry to the Rescue", CoRR, 2016

[80] Cl'ement Godard and Oisin Mac Aodha and Gabriel J. Brostow, "Unsupervised Monocular Depth Estimation with Left-Right Consistency", CoRR, 2016

[81] Xiaowen Wang, Wei Liang, Liuxin Zhang, "Morphable Face Reconstruction with Multiple Views", Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 250-253, Aug. 2010.

[82] Ergun Akleman, "Automatic Creation of Expressive Caricatures: A Grand Challenge For Computer Graphics"

[83] ROBERT L. GOLDSTONE, MARK STEYVERS, BRIAN J. ROGOSKY, "Conceptual interrelatedness and caricatures", Memory & Cognition 2003, 31 (2), 169–180

[84] Pei-Ying Chiang, Wen-Hung Liao, Tsai-Yen Li, "AUTOMATIC CARICATURE GENERATION BY ANALYZING FACIAL FEATURES"

[85] Yang, W., Toyoura, M., Xu, "Example-based Caricature Generation with Exaggeration", J. et al. Vis Comput (2016) 32: 383. doi:10.1007/s00371-015-1177-9

[86] Lyndsey Ann Clarke , "The Automatic Generation of 3D Caricatures from a single Facial Photograph", A thesis submitted to the University of Wales in fulfilment of the requirements for the Degree of Doctor of Philosophy

[87] Chris Johnson , "Faces and Caricatures: 3D Caricature Generator"

[88] T. Vetter and T. Poggio, " Linear object classes and image synthesis from a single example image", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.733-742,1997.

[89] "Adaptive 3D face reconstruction from unconstrained photo collections", Roth, Joseph and Tong, Yiying and Liu, Xiaoming", 2016, CVPR

[90] "Automated 3D Face Reconstruction From Multiple Images Using Quality Measures", Marcel Piotraschke, Volker Blanz; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3418-3427

[91] Joseph Roth and Yiying Tong and Xiaoming Liu, "Unconstrained 3D Face Reconstruction", Proc. IEEE Computer Vision and Pattern Recognition, 2015

[92] Feng Liu and Dan Zeng and Jing Li and Qijun Zhao, "Cascaded Regressor based 3D Face Reconstruction from a Single Arbitrary View Image", CoRR, 2015

[93] Vahid Kazemi, Cem Keskin, Jonathan Taylor, Pushmeet Kohli, Shahram Izadi, “Real-time Face Reconstruction from a Single Depth Image”, 3DV, 2014

[94] S. I. Behlim and T. Q. Syed, 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), “3D facial reconstruction from a single 2D rasterstereography image”, 2014

[95] Pablo Garrido and Michael Zollhoefer and Dan Casas and Levi Valgaerts and Kiran Varanasi and Patrick Perez and Christian Theobalt, “Reconstruction of Personalized 3D Face Rigs from Monocular Video”, {ACM} Trans. Graph. (Presented at SIGGRAPH 2016)

[96] Pramila D. Kamble, Bharti W. Gawali - Face Detection with Photo-Sketch using 3D Face Expressions Synthesis and Recognition - published at: "International Journal of Scientific and Research Publications (IJSRP), Volume 2, Issue 9, September 2012 Edition".

[97] USF, “Human id database,” http://www.cse.usf.edu/ sarkar/index files/DataAndCode.htm, Last Accessed June 2008

[98] “3D face reconstruction from a single image using a single reference face shape”, Kemelmacher-Shlizerman I[1], Basri R, IEEE Trans Pattern Anal Mach Intell. 2011 Feb;33(2):394-405. doi: 10.1109/TPAMI.2010.63.

[99] Varin Chouvatut and Suthep Madarasmi and Mihran Tuceryan, “Face Reconstruction and Camera Pose Using Multi-dimensional Descent”, International Journal of Computer, Electrical, Automation, Control and Information Engineering}, volume 3, 2009

[100] Paysan, Pascal and Lüthi, Marcel and Albrecht, Thomas and Lerch, Anita and Amberg, Brian and Santini, Francesco and Vetter, Thomas, Springer,Lecture Notes in Computer Science, Face Reconstruction from Skull Shapes and Physical Attributes”,volume = 5748, year = 2009

[101] Y. Zhang and Q. Ji and Z. Zhu and B. Yi, Journal, IEEE Transactions on Circuits and Systems for Video Technology, “Dynamic Facial Expression Analysis and Synthesis With MPEG-4 Facial Animation Parameters”, 2008

[102] L.Yin and M. Yourst, "3D face recognition based on high-resolution 3D face modeling from frontal and profile views" In ACM Workshop on Biometric Methods and Applications, pp. 1-8,2003.

[103] Amit K. Roy Chowdhury and Rama Chellappa, “Face reconstruction from monocular video using uncertainty analysis and a generic model”, Pattern Recognition, 2005

[104] Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, " Efficient 3D reconstruction for face recognition", Pattern Recognition, pp. 787-798, 2005.

[105] L. Zhang and S. Wang and D. Samaras, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), “Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model”, 2005

[106] I. A. Ypsilos and A. Hilton and A. Turkmani and P. J. B. Jackson, 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on, “Speech-driven face synthesis from 3D video”, 2004

[107] Building Three Dimensional Head Models Nikos Sarris, Nikos Grammalidis, and Michael G. Strintzis

[108] Y. Xing, C. Guo, and Z. Tan, "Automatic 3D facial model reconstruction from single front view image", Proc. of 4th Int'l Conf. On Virtual Reality and Its Applications in Industry, SPIE, pp. 149-152, 2004.

[109] Ansari and M. Mottaleb, "3D face modeling using two orthogonal views and a generic face model", Proc. of Int'! Conf. on Multimedia and Expo (ICME), pp. 289-292, 2003.

[110] Li Zhang, Brian Curless, and Steven M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. In *Proceedings of the 1st International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, Padova, Italy, June 19-21, 2002, pp. 24-36.

[111] Sean Ryan Fanello∗, Christoph Rhemann∗, Vladimir Tankovich ,Adarsh Kowdle, Sergio Orts Escolano, David Kim, Shahram Izadi , “HyperDepth: Learning Depth from Structured Light Without Matching”, CvPR, 2016

[112] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. Cognition & Emotion, 24(8), 1377—1388. DOI: 10.1080/02699930903485076

[113] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. 2000. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*22, 10 (October 2000), 1090-1104. DOI=http://dx.doi.org/10.1109/34.879790

[114] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In ECCV Workshop on Faces in Real-life Images, 2008

[115] MPI, “Max planck institute face database,” http://faces.kyb.tuebingen.mpg.de, Last Accessed June 2008.

[116] Cyberware, “Rapid 3d scanners,” http://www.cyberware.com/, Last Accessed June 2008.

[117] Kanade, T. "Recovery of the three-dimensional shape of an object from a single view" Artificial Intelligence Vol. 17(1981) pp409-460.

[118] http://www.di4d.com/work/research/

[119] “Shape from Contour Using Symmetries”, Shiu-Yin Kelvin Yuen, 1989, Alvey Vision Conference 1989: 1-4

[120] “Shape from Contour: Straight Homogeneous Generalized Cylinders and Constant Cross Section Generalized Cylinders”, Fatig Ulupinar and Ramakant Nevatia, 1995

[121] “How shape from contours affects shape from shading”, Dejan Todorović, Vision Research 103 · October 2014, DOI: 10.1016/j.visres.2014.07.014

[122] “Suggestive Contours for Conveying Shape”, Doug DeCarlo, Adam Finkelstein, Szymon Rusinkiewicz, Anthony Santella, In SIGGRAPH 2003

[123]. T. Akimoto, Y. Suenaga, and R.S. Wallace, “Automatic creation of 3d facial models,” IEEE Computer Graphics and Applications, 1993.

[124] W. Lee, P. Kalra, and N.M. Thalmann, “Model based face reconstruction for animation,” in Int. Conf. on Multimedia Modeling, 1997. 49.

[125]. I.K. Park, H. Zhang, and V. Vezhnevets, “Image-based 3d face modeling system,” EURASIP Journal on Applied Signal Processing, , no. 13, pp. 2072–2090, 2005.

[126] MPEG, “Mpeg4 description,” http://www.chiariglione.org/MPEG/standards/mpeg- 4/mpeg-4.htm, Last Accessed June 2008.

[127] I. S. Pandzic and R. Forchheimer, “Mpeg4 facial animation: The standard, implementation and applications,” John Wiley & Sons, 2002.

[128] G. Farin, “Surfaces over dirichlet tessellations,” Computer Aided Geometric Design, vol. 7, no. 1–4, pp. 281–292, 1990.

[129] D. Onofrio, S. Tubaro, A. Rama, and F. Tarres, “3d face reconstruction with a four camera acquisition system,” in Int. Workshop on Very Low Bit-Rate Video-Coding, 2005.

[130] Q Chen and G Medioni, “Building 3-d human face model from two photographs,” Journal of VSLI Signal Processing, pp. 127–140, 2001.

[131] P. Leclercq, J. Liu, A. Woodward, and P. Delmas, “Which stereo matching algorithm for accurate 3d face creation?,” in Int. Workshop on Combinatorial Image Analysis, Lecture Notes in Computer Science, 2004.

[132] S. Huq, B. Abidi, A. Goshtasby, and M. Abidi, “Stereo matching with energy minimizing snake grid for 3d face modeling,” in Defense and Security Symposium, 2004

[133] R. Legnagne, J. Tarel, and O. Monga, “From 2d images to 3d face geometry,” in Int. Conf. on Automatic Face and Gesture Recognition, 1996.

[134] M. Chan, C. Chen, and G. Barton, “A strategy for 3d face analysis and synthesis,” in Proc. of Image and Vision Computing, 2003.

[135] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” International Journal of Computer Vision, vol. 47, no. 1-3, pp. 7–42, 2002.

[136] P Fua, “Regularized bundle-adjustment to model heads from image sequences without calibration data,” Int, Journal of Computer Vision, vol. 38, no. 2, pp. 153–171, 2000.

[137] Shape from Focus Shree K. Nayar CMU-RI-TR-89-27

[138] “Shape from Focus: An Effective Approach for Rough Surfaces”, Shree K. Nayar, Yasuo Nakagawa

[139] “Analysis of focus measure operators for shape-from-focus”, Said Pertuz, Domenec Puig, Miguel Angel Garcia ,http://dx.doi.org/10.1016/j.patcog.2012.11.011, Volume 46, Issue 5, May 2013, Pages 1415–1432

[140] Paolo Favaro and Stefano Soatto. 2002. Learning Shape from Defocus. In *Proceedings of the 7th European Conference on Computer Vision-Part II* (ECCV '02), Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen (Eds.). Springer-Verlag, London, UK, UK, 735-745.

[141] . L. Xin, Q. Wang, J. Tao, X. Tang, T. Tan, and H. Shum, “Automatic 3d face modeling from video,” in IEEE Int. Conf. on Computer Vision, 2005.

[142] R. White, D.A. Forsyth "Combining Cues: Shape from Shading and Texture", IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[143] P.Fua and Y. Leclerc, "Object-centered surface reconstruction: Combining multi-image stereo and shading",International Journal of Computer Vision, 16:35-56,1995.

[144] B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju, “Model-based 3d face capture with shape-from-silhouettes,” in Analysis and Modeling of Faces and Gestures, October 2003.

[145] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, “Silhouette-based 3d face shape recovery,” in Graphics Interface, June 2003.

[146] S. Wang, L. Zhang, and D. Samaras, “Face reconstruction across different poses and arbitrary illumination conditions,” in Proc. of Audio- and Video-based Biometric Person Authentication, 2005, pp. 91–101.

[147] S.llic and P. Fua, " Implicit meshes for modeling and reconstruction", Conference on Computer Vision and Pattern Recognition, Madison, WI, 2003.