VOL 3 (2019) NO 2

e-ISSN : 2549-9904
ISSN : 2549-9610

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

A Novel to Identify Multiple faces by tracking 2D Face Images over 3D Plane

Dayanand G Savakar[#], Ravi Hosur[*]

[#] *Department of Computer Science, Dr. P. G. Halakatti Post Graduate Centre, Rani Channamma University, India*

[*] *Department of Computer Science & Engineering, BLDEA's V. P. Dr. P. G. Halakatti College of Engineering & Technology, India*
E-mail: hosuravi@gmail.com

Abstract*— Nowadays, face recognition is very much effective way of counter security threats in various aspects of human life. Though, other means of defending security attacks exists but they have their own drawbacks and overheads. Human face can be recognized from 2D face images or from 3D geometry of human faces. Although very popular, 2D face recognition algorithms are constrained by various factors like change in illumination, varying facial expressions, make-up on the face and orientation of the head. On the other hand face recognition based on 3D geometry of the faces has been shown to have more correctness than 2D face recognition. The only technological drawback however is that the 3D cameras are much less common than 2D based cameras. Therefore the work propose a novel facial expression recognition in real-time by tracking 2D face over a 3D plane. We use multiple 2D planes projected on each-other such that a 2D facial feature point is projected over all the planes, selecting the closest points over the plane and finally creating a contour enclosing the projected feature point over multiple planes, thereby creating a 3D tracking plane with 2D feature point. We use Cambridge landmark markers [4, 5] for facial tracking with multiple Homomorphic projection [5] for creating the 3D feature points. This technique has been proven better in terms of accuracy improvement.

***Keywords*— Landmark, projection, geometry, 2D face, plane, tracking.**

I. INTRODUCTION

The media driven world and video centric information system under growth, tracking video statistics becomes an integral part of the growth. People expect to know the number of times a particular brand ambassador appears in video, the sentiment in the video and related statistics. Current states of art in the face detection in the videos are 2D. The problem with such techniques is that in real videos which are best described as 2.5D, are often not processed qualitatively with existing technique. Though 2D face recognition systems have been successfully implemented commercially, it suffers with two main drawbacks: variation in illumination and pose. As the human face is not a solid and rigid object and 2D algorithms represent a face by intensity variation, variations in light illumination may result in variable reflection of light from skin of the face which can result in differences in images. Most of the 2D recognition algorithms perform well in controlled illumination condition and performance drops with non-controlled illumination.

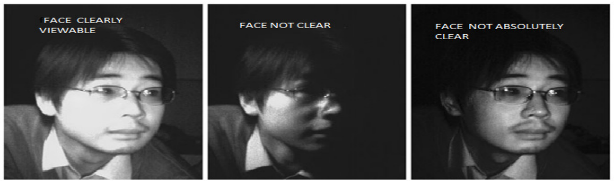


Fig1. Variation in images due to change in illumination

Another drawback of 2D face recognition algorithms is pose variation which results from movement of the head. This results full frontal view of the face which may not be available to camera. The different positions and orientations of the head with respect to some coordinate system gives rise to different poses like front profile, left profile, right profile etc. The changes in the pose may block the view of the face ultimately resulting in degradation of performance of the recognition algorithm. In an uncontrolled environment, variation due to changes in pose resulting from up, down, left and right head movements can hamper recognition task to a great extent.

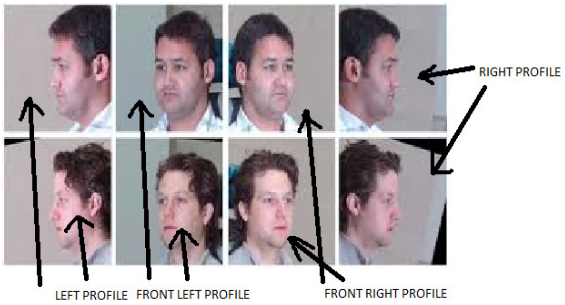


Fig 2. Variation in images due to change in pose

To overcome above limitation and to process the videos better for identifying face and the expression associated with it, we propose a novel projection based emotion detection model that projects 2D and 2.5D scenes onto a 3D plane and recognizes facial boundary on the plane. The mapped facial plane is projected back to the 2D video. The system provides better robustness to the detection.

II. LITERATURE REVIEW

A face recognition system is a computer vision oriented technology that identifies and recognizes human faces from digital images or frames of video. In recent times, face recognition systems have drawn tremendous attention over other biometrics as they do not require user cooperation and contact to the system. Various application areas where face recognition systems are being widely used like in social media, identity verification process, security systems, and fraud detection and so on.

As 2D face recognition algorithms suffer from variation in lighting, position and orientation of the head and facial expressions, 3D face recognition has become the focus of the researchers which can attain more correctness in comparison to 2D face recognition algorithms. But capturing a 3D image is a technological drawback as 3D cameras are seldom and costly. Some of the recent research works on 3D face recognition by converting 2D images have been discussed below.

In this paper [1], a novel method for detecting and tracking facial landmark features on 3D static and 3D dynamic range data. The efficiency of the detected landmarks is validated through applications for geometric based facial expression classification for both posed and spontaneous expressions, and head pose estimation.

The authors in [2] elaborated a method of reconstructing a 3D face model from lower quality 2D images acquired under various facial expression, head orientation and illumination conditions. They have accomplished the task using a 3D morphable model forming personalized templates.

The paper in [3] presents a novel method for 3D face reconstruction captured from front camera of a smartphone suffering from motion blur, non-frontal profiles and low resolution. The 3D face reconstruction has been performed by using Scale-Invariant Feature transformation first and then feature matching is used to generate consistent tracks which produce point clouds on subsequent processing by the use of Point/Cluster based Multi-view stereo(PMVS/ CMVS).

The research work in [4] focus on large pose based face recognition method by fusing CNN regression and 3D morphable model. The method fits a dense 3D morphable model to 2D face images having arbitrary pose. Cascaded CNN-based regressors have been used to determine the fitting parameters, 3D shape parameters and the parameters of projection matrix. The learning of CNN is achieved by designing pose-invariant appearance features using dense 3D shapes. The face alignment have its own importance, hence it is considered fine throughout decades [6] with additional approach say Active Shape Model [7] and Active Appearance Model (AAM) [8,9] . face alignment recognized as a top vision issue due to its popularity. So that it gains plenty of attentions to achieve better results. The present method are partitioned into three clusters say AAM-based methods [8, 9,10], Regression-based methods [11, 12, 13]

III. METHODOLOGY

Face is captured using effective multi face SDK. Face is detected by tracking the geometric points of human face. Total 70 points are matched to detect the face. We have two algorithms i.e. appearance based algorithm and model based algorithm. We mainly deal with model based algorithm where a model of the face is created and tracked and a projection model is built with the facial model. Face identification is achieved by using Viola-Jones method. This method has special feature of combining several weak classifiers and making strong classifier.

For efficiently detecting the faces in more natural videos, the scenes must have a 3D perception. Such a perception can be mathematically represented by projecting the points of 2/2.5D system over a 3D plane. Various projection techniques are popular that includes isometric projection, homomorphic projection.

A Homomorphic projection is often used in 3D video projection system. We map the facial feature points over the homomorphic plane in following ways. The geometric points of the person are matched with geometric model and all the features are matched properly. Human face contain 70 geometric points covering all the features of human face like eyes, nose, lips, mouth, chin, ears, eye brows etc. The video is captured and stored in android phone with option called choose video.

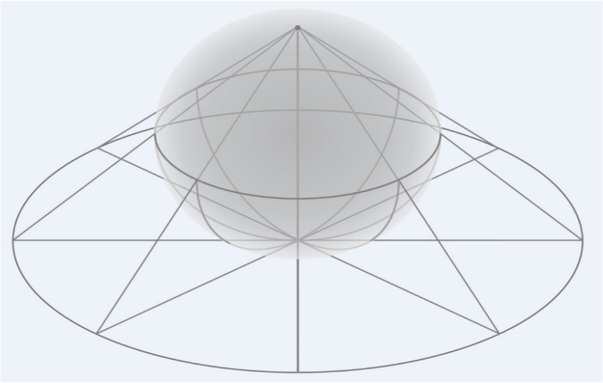


Fig 3. Typical projection

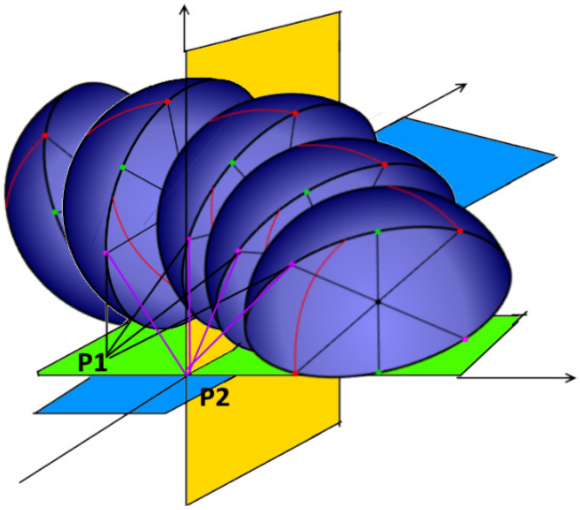


Fig 4. Typical Homomorphic projection of points

In the figure 4, we can see a typical Homomorphic projection of circular contour on a spherical plane. The proposed work uses 68 feature points, 68 landmark points as facial feature points, then projects them on a virtual spherical Homomorphic plane. This plane consists of several individual planes. By projecting on to the plane a feature point is made to be the candidate of one plane while the user moves his head the feature point's changes from one plane to another but then the parent plane. Therefore if we track the movement of a particular point across one plane and when it is in the junction of other neighboring plane by measuring 3D Mahalanobis distance.

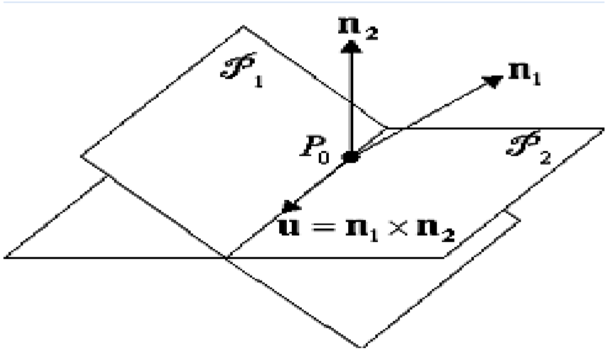


Fig 5. Representing planes to track a point

The mobility of the points across different planes is represented in figure 5. In the case of 2D features tracking a candidate landmark point is tracked only across X and Y planes. Therefore the rear movement of the person is effected due to the loss of planar information. However such a loss is negated by the proposed system by tracking the points across different planes as shown in figure 5.

The figure 5 also shows one point been tracked through multiple planes then their movements. Once the feature vectors are tracked they are being matched for patterns. We use convolution Neural Network to match the number of 68 feature points normalized across multiple 3D planes to crossing face. As points like the lips landmark points cannot be too far away from each other the multiple points tracking offers much better detection and classification result than other techniques. The facial point tracking through multiple planes is as shown in figure 6. Here, we can observe how a set of feature points are tracked from one plane to another. The difference between tracking of a point through only X and Y coordinate versus the same in the 3D planar system is visible.

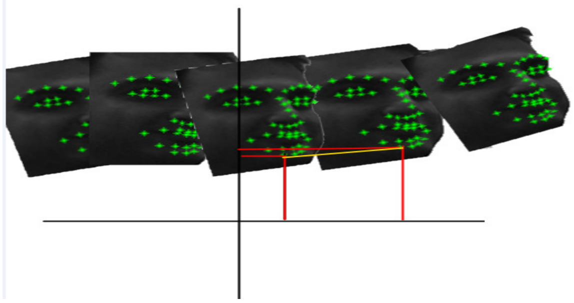


Fig 6. Tracking feature points of a face with change in plane

As number of possible facial feature points are considered to be part of multiple planes in the proposed system and are been processed in parallel for maximum likelihood of the point to be belonging to the part of plane, the proposed system is extremely parallel-able due to non-dependency on the texture or the color data, our system eliminates the skin, lightening, gender, age group of the user and is capable of detecting the face with complete geometry model of the face. Multi-plane and multiple classifiers provide immense amount of granularity to the proposed system. The methodology can be represented by the following figure 7.

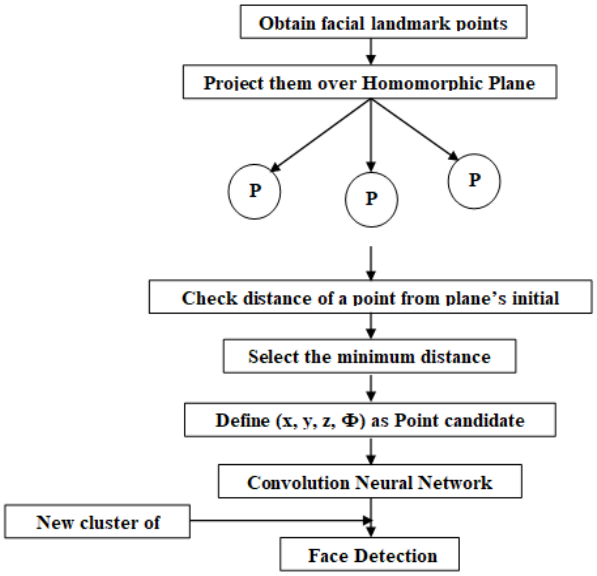

Obtain facial landmark points
Project them over Homomorphic Plane
P
P
P
Check distance of a point from plane's initial
Select the minimum distance
Define (x, y, z, Φ) as Point candidate
Convolution Neural Network
New cluster of
Face Detection

Fig 7. Proposed Methodology

IV. Results and Discussion

4. 1. 2D facial Tracking

4.1.1. *Scale Invariant*

In this experiment, we vary the resolution of the sensor to the specifications like 240×180,320×240,640×480 and 900×600. We obtain the following tracking time, tracking period, accuracy and frame rate.

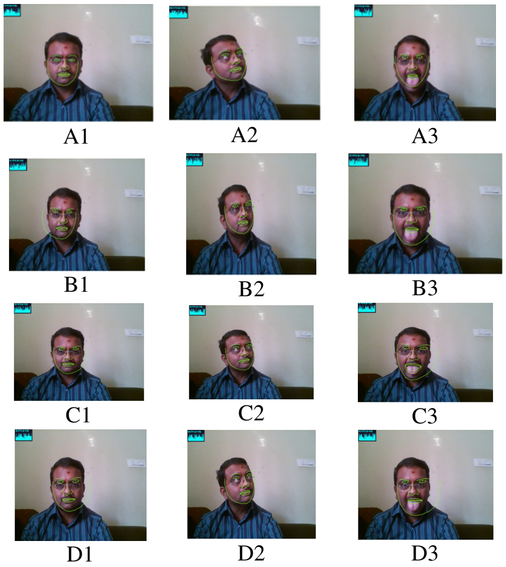


Fig 8. Scale Invariant face tracking (A_i to D_i)

TABLE I
RESULTS OF TRACKING A FACE WITH SCALE INVARIANT

PARAMETER	TRACKING TIME (IN SEC)	TRACKING PERIOD (%)	ACCURACY (%)	FRAME RATE (PER SEC)
RESOLUTION				
240×180	**03**	**37**	**56**	**92**
320×240	**07**	**83**	**77**	**83**
640×480	**13**	**71**	**84**	**80**
900×600	**14**	**67**	**91**	**78**

Tracking time is defined as the time from the beginning to the period for the landmark points to be converged. Tracking period is defined as total time fraction for which the landmark points remain converged for mild movement of head. Accuracy is referred to as total period of time for which a dominant face was detected over the time for which single dominant face was presented by author. Frame rate is defined as the number of frames being processed by the system per second.

From the above experiment, we conclude that high resolution takes longer period of time for tracking and hence frequently loses the tracking points. Very low frame rate results in bad tracking because of the increase in the confusion matrix. Hence we can say that 640×480 is the best resolution for the tracking. However this experiment also exposes the inherent drawback of the 2D based system because numbers of feature point is low.

4.1.2. *Distance Invariant*

In this experiment, we change the distance of the user from the camera and study the same tracking performance.

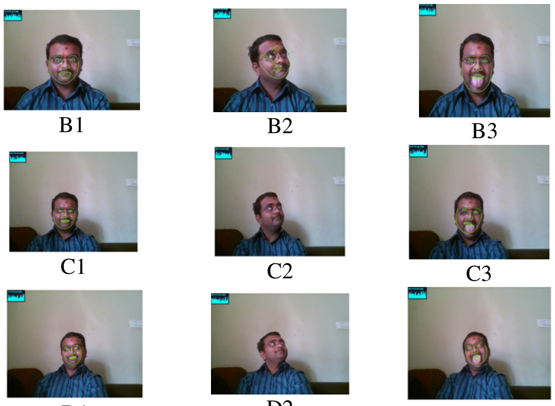


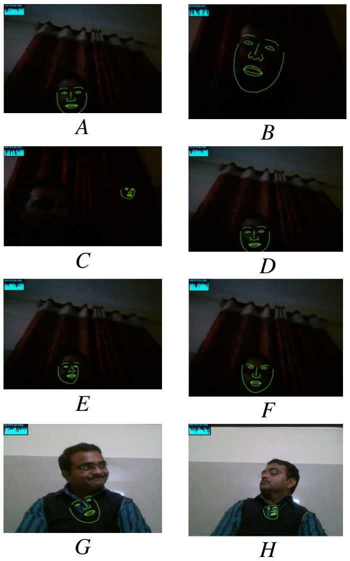
Fig 9. Distance Invariant face tracking

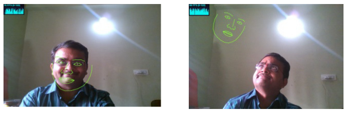
TABLE II
RESULTS OF TRACKING A FACE WITH DISTANCE INVARIANT

Parameters	Tracking Time (In seconds)	Tracking Period (%)	Accuracy (%)	Frame rate (Per second)
Resolution				
240×180	**03**	**37**	**56**	**92**
320×240	**07**	**83**	**77**	**83**
640×480	**13**	**71**	**84**	**80**
900×600	**14**	**67**	**91**	**78**

From above experiment we found that too nearer to the screen or too far away from the screen makes the tracking unstable. However the intermediate distances perform better intern of tracking accuracy and consistency. In the context of 2D facial tracking because number of tracking points are limited only a single plane slightest of the head movement requires the model to be reconfigured. This is one of the reasons why 3D facial tracking is presented in recent times because consistency for the tracking period is significantly high. The other drawback of 2D based facial tracking is, while the distance from camera is varied, the accuracy changes and retracking needs to be reinitialized.

4.1.3 Light Invariance





I *J*

Fig. 10 Light Invariant (in Lumens) face tracking (A to J)

The above experiment was conducted by varying the room light controlled by dimmer keeping the user steady and keeping resolution at 640×480.

TABLE III
RESULTS OF TRACKING A FACE WITH LIGHT INVARIANT

Parameters	**Tracking Time (In seconds)**	**Tracking Period (%)**	**Accuracy (%)**
Type of Light Intensity (in lumens)			
Low Light	1.3	98	82
Medium Light	1.427	78	89
Adequate Light	1.597	91	96

From this experiment, it is clear that Low Light intensity results in drop in accuracy as well as moderate drop in tracking time. This is because the Low Light intensity generates several shadows over the visible plane, which results in fluctuating feature points because of which the accuracy is reduced. We overcome this problem of 2D tracking by using homomorphic projection over a higher dimensional 3D plane and thereafter voting the parameters to be the part of 3D plane because it generates a better overall plane by correlating feature points from multiple planes.

4.1.4 Plane Invariant

In this method we propose different angle of view to the system where the role of system is to continuously track the emotion. We take the plane of view versus accuracy.

TABLE IV
RESULTS OF TRACKING A FACE WITH PLANE INVARIANT

Angle View	**Accuracy**
0^0	96
30^0	92
$-30^0(330^0)$	86

From the above experiment, we can see that keeping the face constant and presenting it with the control on view angle where the sensor is lower to the center is in the same plane that of the center and at the lower angle. It is clear that central axis gives the better performance in terms of tracking. Due to limited view angle in the 2D mode often it is found to be more accurate when the face is presented along with the same view angle. To overcome the drawbacks appearing from the experiment over 2D based system, we adopt a 3D system whose results have been tabulated.

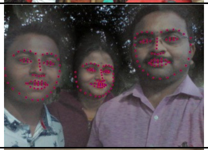
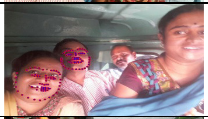
4.2. *3D Facial tracking*

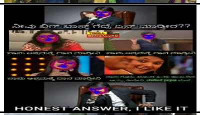
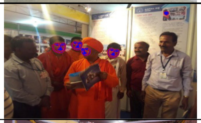
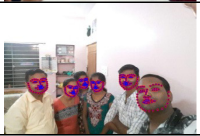
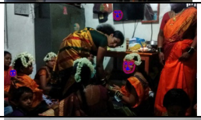
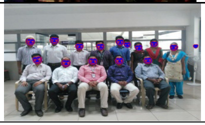
The problems that are identified in the above section are overcome by the use of proposed multiple plane, homomorphic projection and tracking 2D facial feature points. We validated the proposed technique with respect to x, y, z database which are essentially video expression databases with proposed system. We did following experimentation to prove the efficiency of our proposed system with same state of art.

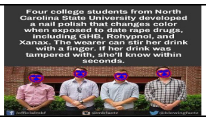
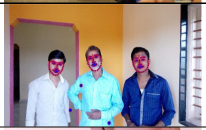
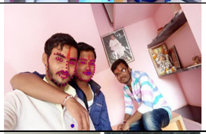
4.2. 1. Tracking of multiple faces in complex background

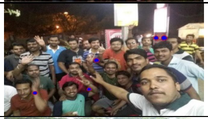
In this experiment we have used videos with multiple faces presenting different expressions across different angles and planes. We evaluated the performance with following matrices.

TABLE V
MULTIPLE FACES TRACKED FROM AN IMAGE

Sl. No	**Image**	**Actual No. of Faces**	**Detected Number of faces**
1.		**2**	**2**
2.		**2**	**2**
3.		**23**	**20**
4.		**3**	**3**
5.		**4**	**2**
6.		**4**	**2**
7.		**30**	**24**
8.		**30**	**16**
9.		**4**	**3**
10.		**4**	**2**
11.		**2**	**2**

12.		**6**	**3**
13.		**6**	**5**
14.		**5**	**5**
15.	Height Of Taking Selfie.. :V	**3**	**2**
16.		**3**	**2**
17.		**6**	**4**
18.		**28**	**28**
19.		**7**	**4**
20.		**7**	**1**
21.		**3**	**1**
22.		**6**	**6**
23.		**21**	**21**
24.		**8**	**6**
25.		**4**	**3**
26.		**14**	**14**

27.	Four college students from North Carolina State University developed a nail polish that changes color when exposed to date rape drugs, including GHB, Rohypnol, and Xanax. The wearer can stir her drink with a finger. If her drink was tampered with, she'll know within seconds.	**4**	**4**
28.		**10**	**8**
29.		**9**	**8**
30.		**12**	**10**
31.		**18**	**16**
32.		**7**	**1**
33.		**5**	**5**
34.		**3**	**3**
35.		**3**	**2**
36.		**2**	**2**
37.		**10**	**10**
38.		**5**	**5**
39.		**10**	**5**
40.		**3**	**2**
41.		**3**	**1**

42.		**9**	**3**
43.		**8**	**8**
44.		**6**	**3**
45.		**5**	**5**
46.		**6**	**4**
47.		**6**	**4**
48.		**4**	**4**
49.		**20**	**5**
50.		**5**	**4**

We can see from the above table that the proposed system is more robust in terms of accurately detecting the faces with the existing methods Cambridge landmark markers [4, 5] for facial tracking with multiple Homomorphic projection resulting in 74.265% recognition of faces in large crowds, images including disturbances, etc.

V. Conclusion

3D based facial tracking, facial feature tracking, facial recognition and other allied areas of study are being gaining significant popularity. Due to the ease of availability of 3D cameras like Intel RealSense, more and more research and communicational work is been observed in this area. However, as many low cost 2D cameras are being deployed as either as security cameras or as a standard practice of photography and videography such 3D based techniques cannot leverage this immensely popular 2D videos. As already discussed in results section 2D tracking has its own natural limitation over 3D tracking which includes loss of phase, plane, data which results in both poor tracking as well as recognition. Our proposed method is an inexpensive enhancement of the 2D and 3D technology called Homomorphic projection which is used in several 3D projectors. Our technique is not only robust but can support an overall tracking angle of up to 240^0 across the 3-principle planes.

References

[1] Laszlo Jeni, Jeff Cohn, and Takeo Kanade, "Dense 3D face alignment from 2D videos in real-time", IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2015

[2] Joseph Rath et al. "Adaptive 3D Face Reconstruction from Unconstrained Photo Collections", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4197-4206

[3] Raghabendra et al. "3D Face Reconstruction and Multimodal Person Identification from Video Captured Using Smartphone Camera", IEEE International Conference on Technologies for Homeland Security (HST), 2013

[4] Amin et al. "Large-pose Face Alignment via CNN-based Dense 3D Model Fitting", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4188-4196

[5] Syed Zulqarnain, Gilani Faisal, Shafait Ajmal Mian, "Shape-based Automatic Detection of a Large Number of 3D Facial Landmarks", CVPR 2015, PP. 4639-4648

[6] Wang, N., Gao, X., Tao, D., & Li, X. (2014). Facial feature point detection: A comprehensive survey. arXiv preprint arXiv:1410.1037.

[7] Cootes, T., Taylor, C., & Lanitis, A. (1994) Active shape models: Evaluation of a multi-resolution method for improving image search. In BMVC vol. 1, (pp. 327–336).

[8] Matthews, I., & Baker, S. (2004). Active appearance models revisited. International Journal of Computer Vision, 60(2), 135–164.

[9] Liu, X. (2009). Discriminative face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), 1941–1954

[10] Liu, X. (2010). Video-based face model fitting using adaptive active appearance model. Journal of Image Vision Computing, 28(7), 1162–1172.

[11] Valstar, M., Martinez, B., Binefa, X., & Pantic, M. (2010) Facial point detection using boosted regression and graph models. In CVPR pp. 2729–2736.

[12] Cao, C., Weng, Y., Zhou, S., Tong, Y., & Zhou, K. (2014). Facewarehouse: A3Dfacial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20(3), 413–425.

[13] Zhang, J., Zhou, S.K., Comaniciu, D., & McMillan, L. (2008). Conditional density learning via regression with application to deformable shape segmentation. In CVPR (pp. 1–8).