3D Face Tracking Using Stereo Camera

Faleh Alqahtani
Queensland University of Technology
Brisbane, Australia
Falehmohammeda.alqahtani@hdr.qut.edu.au

Prof Vinod Chandran
Queensland University of Technology
Brisbane, Australia
v.chandran@qut.edu.au

Dr. Jasmine Banks
Queensland University of Technology
Brisbane, Australia
jasmine.banks@qut.edu.au

***Abstract*—The content of this article explores the use of 3D face tracking systems by implementing of active stereo vision cameras to ascertain the position of a person's face. The various experiments conducted in-depth have produced both promising and satisfying results for images that have enabled examiners to determine the disparity between images. The paper also explores some of the various challenges researchers are facing with the implementation of algorithms to construct cloud-points in from stereo-based images. The reviewed recommendations suggest on better software components that would avail final 3D computational images or reconstructions that can easily be matched to the original. The tracking system modules address the challenges of the practical application of face tracking including pose illumination and occlusion. The content of the paper evaluates the putting into practice of multi-view or multiple stereo cameras enhance the field of view to improve the performance of a 3D tracking system. Face tracking using functional multi-view stereo camera systems can significantly solve the correspondence problem and the issue of comparing scenes or image points given that extra views reduced ambiguity in matching.**

Keywords-3D, face tracking, stereo vision, and stereo camera

I. Introduction

The subject of Face tracking has been an interesting topic of study over the years for people interested in researching computer visualization particularly in analyzing the quality of images matched to 3D simulations or scans. The trending themes in the picture analysis domain gradually move towards the aspect of examining the movement of an object in a selected image plane. The sequential process of extracting information from the captured and analyzed motion images can be difficult depending on the lighting and angles at which the original image is scanned. Although, research [3] shows that the human face is often affected by multiple changes, which are dependent on the sources that cause the change in motion of the face.

Numerous approaches employ machine vision in analyzing the images obtained from stereo cameras. Some of these techniques include the state of art tracking methods like Mean Shift/Cam Shift, TLD method, the KLT method, the Particle Filter and PHD filters for multiple target tracking methods. The fact is that the appearance of the human face can fundamentally change depending on the physiology, communication, speech and emotion state, which tends to make it hard to recreate quality 3D photo models of the human face.

II. Face Tracking Method

A. *Background*

The information highlighting the progress of research in the field of face tracking applies or employs different methods each of which offer diverse features in processing an image [1]. The findings presented in the paper [2] state and highlighted the various tracking methods implemented in the picture analysis environment human- machine interactions. The face tracking systems vary depending on the capabilities of the hardware and software capabilities used to analyze a particular image module. Here are some of the identified methods that are commonly utilized in examining pictures and object movement by relating the facial features of the person in the images.

B. *The Particle Filters Methods*

Particle filter methods use an inference method that implements sequential sampling by the importance of the particle in a particular simulation. The objective of particle filters in robustly tracking 3D facial appearance depends on determining a preliminary number of set particles that measure the images analyzed in a sequential fashion.

The steps implemented in robotics experiments require these machines to monitor emotion, which is considered a big challenge when it comes to tracking the human face. According to Cho, Lee and Suh on the uses of particle filters [10], the authors propose the utilization of active appearance models that change the number of particles depending on the situation and degree of importance of the features in the image sequence. Applying the use of particle filters in analyzing the gestures and facial expressions of the human face is expected to enable the continuous capture of the face unless the face is purposely hidden. The face expresses information by utilizing the movement of facial muscles, which entails tracking the location and shape of the facial features including the eyebrows, nose, and mouth movement [10].

In computer vision, the particle filter model is proposed in testing 3D data for facial action because they can assist in the detection of the performance of boundaries in a 3D image. Researchers have investigated the use of stereo cameras in tracking and analyzing moving images. Some of the results demonstrate that there is a significant advantage of using 3D face tracking is that it can perform linear separation on higher dimensions. The idea of particle filtering coupled with 3D face models can aid in face achieving a higher estimation of the boundary features for a given image being analyzed.

C. The Kanade–Lucas–Tomasi (KLT) Method

This method is referred to as the Kanade–Lucas–Tomasi (KLT), which is a feature tracker that uses spatial intensity information in extracting the best match for an image registration. The traditional image registration methods tend to be costly, therefore, making the KLT method an appropriate approach that can use in face tracking. Although, vision-based applications that use the Kanade–Lucas–Tomasi (KLT) algorithm provide the information in the two-dimensional symmetry, the features of the package can achieve accurate and robust corresponding features of the 3D simulations of original [8].

The Kanade–Lucas–Tomasi (KLT) algorithm uses features or points that mainly track the trajectory of the face in a given video sequence. The image analysis can be head motions where the system tracks the features of the face boundary for each frame rate from one image to the next. The KLT tracking algorithm mainly detects the face in the initial frame, selects the trackable points on the face contours by determining the equal mutual distance. The selection of features in the KLT model can adopt any criterion in identifying points between two images [8].

The face detection systems that apply the KLT face tracking algorithm in image processing implements the computational concept that integrates an approach suitable for video monitoring. The efficiency of the model is that it achieves the aim of tracking an object using a number frames captured from an image sequence taken of the object. The primary technique uses trajectory estimation of the target object over multiple frames then evaluating the variation for every pixel in the in a small neighborhood.

In a paper talking about tracking of multiple objects, the research presents information highlighting the importance of object detection in the development of predefined features that can be tracked over a number of frames for consistency. Authors' [8] argue that the challenge to implementing this method in object tracking be that sometimes the visual results are affected by abrupt camera motion or first illumination variation [21]. This means that sometimes the features that are tracked by the KTL algorithm over long sequence can be lost after multiple frames.

In the experiment, the object tracking scheme is represented with more consistent features that ensure that the trajectory of weighted functions is maintained to eliminate any tracking errors in the frames. For instance, in the image shown above, the camera can track both objects in the sequence using the stereo camera [21]. Therefore, the implementation of weighted functions is proposed to increase the performance of the algorithm in identifying distinct points after each frame in the sequence. The algorithm would be beneficial in the 3D face tracking since it is computationally intuitive in tracking the morphology transforms on the face [8].

The implementation of KLT detection features is known to have the capability to approximate the global motion of the Stereo camera area movement. As a result, the motion of the features could better represent the motion of the object as exemplified in the picture 1a and b above showing the vector representation of the object, which is a boat in this case.

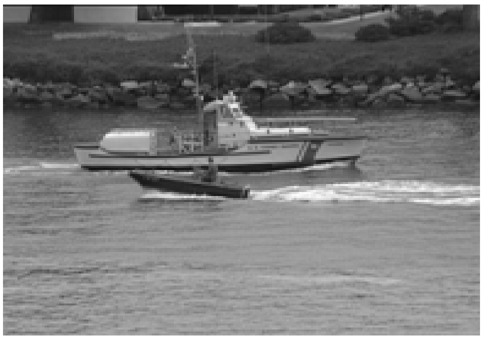


Figure 1

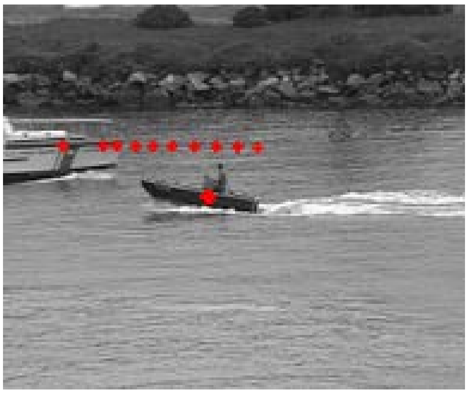


Figure 2

D. The PHD Filters for multiple Target Tracking

The accurate evaluation of frames for a given multiple objects being tracked in real world scenarios can be accomplished through using e Probability Hypothesis Density (PHD) filters and graph matching in the data associations. The PHD filtering process can be useful in reducing the noise and clutter generated in the 3D data input. The concept of PHD filters uses an algorithm function that peaks the information of the particles when it identifies the estimated target of a visualized object [11].

Although, the parameter of performing the filtering techniques on 3D image data cannot be applied to visual tracking for multiple targets. The proposed approaches to solving the issue are employing independent algorithms that track each target separately. Another drawback of using the PHD filtering method on face detection applications is the slow reaction of the system in sensing new real targets [12]. Nevertheless, the coherence of consecutive detections is improved using Particle PHD Filters, which develop the robustness of the face tracker [11].

In Figure 1 and Figure 2, the experiment aimed to demonstrate the difference between using Particle PHD Filtering (PHD-MT) and without using it in multi-agent tracking [11]. As shown in the picture frames on the right side, the application of PHD-MT successfully recovers the face of the multiple target faces in the picture frames without generating false tracks.

E. The Mean Shift/Camshift Method

The prediction implemented by the Camshift model is aimed at improving the performance of surveillance systems. Stereo Cameras that offer 3D tracking and scanning capabilities for moving objects can be quite effectively by processing the structure of the object position. The Mean Shift/Camshift Method enables the surveillance systems to be more intelligent for human life by efficiently tracking targets and objects in occlusion [18]. This model is especially useful in the design of monitoring systems.

The mean shift algorithm can also be used for smoothing and segmentation. The main approach followed in testing this method detects the face of a moving target in the image sequence [20]. The initial step involves the detection and segmentation of the moving target to enable better distinction of the target object in terms of analyzing and providing disparity between the from the background and other objects, even when they are similar.

The Mean Shift method if used together with a filtering process, for example, when used with the Kalman filter, it increases the probability of the method successfully providing the estimated positions even when the target is in occlusion. The Kalman filter process is known to offer an acceptable solution for position approximation of the object being tracked even when the locus of the target is occluded shortly [20].

In an experiment conducted by Shao results, Kuo and Juhng [12], the tracking of two target objects, that is, a diskette and a moving person proves that the CamShift tracking method is able to provide continuous positioning of a moving target.

The experiment shows that the CamShift tracking method improves performance, is an excellent one and robust. The ability to track moving people even after they pass each other and are shadowed or occluded by each other is sustained by the use of the Mean Shift method. This selected Kalman filter has the added advantage of not only tracking the human gestures [2] but also can highly predict the trajectory of the moving individuals. The Camshift method is an area of research given that it would offer promising results by solving three major issue associated with object tracking. These problems are inclusive of object tracking in occlusion, moving object detection and moving object tracking.

F. The Chehra Algorithms and TLD method

Chehra, meaning face in Hindi, is a fully automated real-time eyes and face landmark tracking and detection software. The Chehra Algorithm is another face detection software package that can be used in 3D face tracking. Researchers who have selected this method in performing the image analysis noted that the Chehra software provides 3D head pose estimation and 49 facial landmark points when it detects a face. The algorithm is reported to be able to achieve high-quality predictions and real-time performance.

The Chehra algorithm has a steady tracking phase after it initially detects the face, and it does this by tracking already identified landmarks. In a report presented by Marko, Juho, and Essa [12] the software proved to have beneficial outcomes particularly in cases that involved operating in an uncontrolled handling natural setting. The setback is that the camera loses landmarks. Thus, the detection phase of Chehra system becomes slow in computing and executing the algorithm [12].

The TLD method, which stands for Tracking-Learning-Detection, implements a process, which decomposes the tasks to improve the runtime of the algorithm or system [14]. In a paper written by Zdenek, Krystian, and Jiri [13], the decomposition of tasks by using the TLD method makes it easier to track errors given that the multiple frames are classified and stored in the system database and is in turn memorized for future reference.

III. Challenges in Face Tracking Systems

The challenges identified in this field of study include the ability of the PC camera device to track the object through facial deformations. In order to achieve effective human-computer interaction, previous research has shown that method implementations that adopt 3D face tracking image processing tend to overcome the problem if object tracking in occlusion [18].

A. Tracking through Facial Deformations

The ability of the algorithm used to monitor through facial deformations is a problem especially in the third domain of social perception. By chance, significant progress has been made in this field of research with many solutions being proposed involving smoothening face deformations gradients [16]. This was aided by the identification of deformation parameters that relate to the exterior texture of the face. Tracking and computing deformations depend on the depth perception of the object being tracked.

A matrix representation is proposed to compute the deformation space. To collect the motion data for a deformation model, the detailed features of the dimensional vector for the face being tracked. The formed gradient is independent for an individual face and needs to be calculated only once to approximate or predict the facial movement of a given subject. In an experiment conducted by Dominik and Leif, a facial deformation system model was used to track and reconstruct stereo images by tracking 3D face gestures. Optimization of the deformation procedures for a clear face in 3D tracking aids in the better analysis of the facial expressions.

B. Occlusion and Clutter

The use of 3D face tracking in the real world can be subjected to noisy information, for instance, occlusion and clutter. The one proposed method for solving this issue is the implementation of color-based tracking. As presented in the experiment-tracking people moving past each other, employing a kernel-based tracking method where the visual model of the tracked object is represented in a color histogram [11]. The outer and inner outlines of the image are detected and tracked.

The issue of clutter can be reduced by using filters that explore components that improve the distinctiveness of the boundaries on the face being tracked. One good example that has been highlighted before is the Kalman filtering model and active optical flow measurements, which monitor and predict the trajectory of video objects. A survey of 3D scanning techniques [14] reveals that occlusion can be eliminated by using projectors that lighten up the object from multiple directions to enable the capture of a clearer or high-quality 3D image.

Object occlusion is a key challenge entailed in the triangulation with this specific method, i.e., the use of optical flow measurements. If a segment of the object blocks the camera's vision of the light stripe, then the resulting 3D image will have a hole in the location that has the occluded geometry [21]. This issue can be seen in 3D descriptions of images retrieved from triangulation-based 3D face scanners. Mostly this type of clutter can be seen in concavities such as the eye sockets and the image sections around the nose. In addition, adding more dynamic light arrays solves the problem of occlusion since it is possible to track an object through using light arrays or patterns or computing the level of human interaction required to develop the final 3D appearance generated from a given scenario.

In an experiment conducted by researchers to track objects even after a temporary state of occlusion, a setup of a linear table to move the object through the light screen while two cameras move at different angles. The static light screen created supports in the capture of the laser line images, which presented a basic and constructive model of a 3D scanner.

However, not all objects that self-occlude can be tracked with the help of multi-camera device systems integrated with a static light screen. This issue can be handled by weighted functions or masks that used to measure the pose of the face. The approach of tracking the face by using stereo information in a surveillance scenario would be difficult given that one has to find a matching algorithm to aid in filling the occlusions in a given image. As proposed [16], the use of packet filters would improve the performance of face trackers given their ability to recover from the loss of track. Nevertheless, in practical terms, occlusion and clutter is still a serious obstacle to the development and use highly robust face tracking systems.

C. Robustness to Pose and Illumination Variations

The ability for algorithms to detect and register the images is contingent on many training illuminations, which will enable the system to handle possible illumination variations. Techniques that test the 3D pose variation tend to typically require a dense sample for both the illumination and pose. The 3D face tracking algorithms that implement techniques that deliver a high degree of robustness are predominantly suitable for sensing or perceiving circular shapes, for example, a laser pointer.

The use of 3D face tracking in the real world can be subjected to noisy information, for instance, clutter and is less robust against illumination and appearance changes. The issue of illumination variations can be dealt with by expressing the stereo images being tested as an appropriate linear grouping set of training images.

In an empirical research study by [14], the results showed that the use of temporal domain processing for images is the best alternative for triangulation. This is because it likely marks higher accuracy levels and augmented robustness as associated to scenarios that entail image processing in the spatial domain. The findings from the test revealed that the researchers detected higher spatial frequencies, which tend to be supportive for spatial domain processing. Then again, it is worth noting that having variations in illumination for the test images can be very helpful in the time domain processing.

Various commercially accessible 3D trackers or sensors that are suitable for face capture often implements texture-assisted or passive stereo imaging that uses controlled illumination with a movable light stripe [17]. Although, these methods or algorithmic systems only require a motionless subject which, therefore, means that it is challenging to track motion objects.

Robustness to pose and illumination variation many times leads to the loss of tracking capabilities for the facial tracking algorithms. Thus, the well-known technique adopted by [16] recommend the use of parameterized functions, which describe the motion of the image points. The 3D tracking model requires the consideration of training images to build the model framework, although, it greatly depends on the quality and variety of the data used. Given that 3D tracking, models are more robust to post variation compared to illumination, in [16] the authors proposed an analytically derived model that implements a structure for describing the appearance of the face regarding the shape, pose, surface reflectance, and the incident lighting.

D. Facial Resolution

Facial Resolution is a major challenge to video-based face tracking systems given that it majorly depends on the illumination and motion models. To test and research the impact of facial resolution on detection methods or algorithms, one has to track through illumination and scale changes. Low facial resolution affects the performance of any face detection or tracking algorithm [15]. It is one of the major weaknesses of video-based face monitoring scenarios [16]. The solution to this problem recommends the implementation of super-resolution methodologies which can help to overcome the problematic issue of low resolution to some extent.

Although, using super-resolution on face detection systems can present an additional challenging problem by itself since the detailed facial features being tracked ought to be modeled precisely. In recent times, findings discovered in research conducted by [16] suggested a technique for face super-resolution that encouraged the use of Active Appearance Models (AAMs). The use of super-resolution requires the registering of several images; then the interpolation process follows it. These two phases of super-resolution are usually treated independently, that is, the action of registering is achieved through a tracking technique that follows the super-resolution method process. In a previous research paper [16], the scientist recommended that feeding back the super-resolved texture in the nth frame for tracing the (n + 1) th frame. The technique ultimately recovers and improves the tracking method, which, in turn, develops the super-resolution yield/output. By considering the issues associated with the convergence and stability in 3D face tracking, the topic of facial resolution can be an exciting field of study for undertaking future work and research investigation.

IV. Real-Time application of target models

The techniques utilized by the tracking models are vital components in real-time application or integration in the actual world environment. Some interesting real-time based application areas include biometrics, video communication, face modeling, video surveillance and multimedia systems.

A. *Applications of Face Tracking*

The techniques or systems of face tracking can be applied to biometrics technology environs. These act as security agents by offering service in terms of offering surveillance solutions to increase its significance to security applications. Face tracking or facial tracking technology is used in both the public and the private sectors of the economy [16]. Face tracking is highly significant in virtual reality, human-computer interaction, multimedia, database recovery, information security, and computer entertainment. The applications of face tracking are inclusive of the following:

a) Film Editing

The movie industry implements the use of computer vision and graphics in video production. There are plenty of programs that provide algorithms that be used for 3D tracking systems, for example, Mocha software. The applications and add-ons within the software package gives the movie developers the ability elimination or add seamless graphical video or images to a movie section.

b) Access control

Numerous control applications such as office computer login or access use face-tracking systems in allowing permissions. Increased accuracy is achieved because of the relatively small size of the group that needs to be recognized and the capture of face pictures under natural conditions. Here, face tracking allows the monitoring of individuals in front of computer terminals and conceals their work in case they leave the terminal without logging out denying access to any other user who is unidentified.

c) Security

Security concerns are growing rapidly thus increasing the significance of face tracking. Face tracking is useful in boosting security at airports, where face-tracking technology is implemented to assist in addressing global crimes. The technologies used in security include; Viisage face tracking technology that alerts officials at security checkpoints. In addition, computer security is used in face tracking technology to prevent intrusion and unauthorized access of files and sensitive documents.

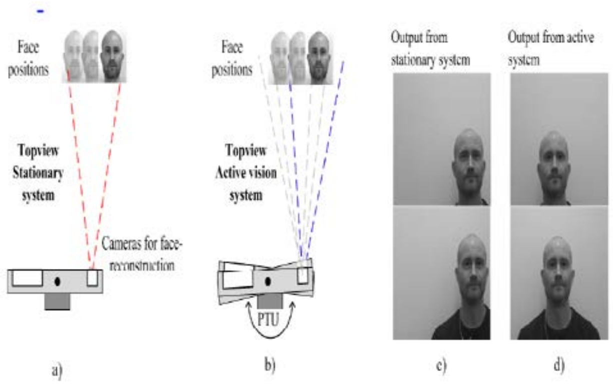

Face positions
Topview Stationary system
Cameras for face-reconstruction
a)
Face positions
Topview Active vision system
PTU
b)
Output from stationary system
c)
Output from active system
d)

Figure 3

d) Surveillance

Similar to security application of face tracking, surveillance is equally achieved. However, surveillance is increasingly challenging because of factors such as; face orientation and free lighting conditions. For example, in monitoring long-distance drivers who are prone to drowsiness.

B. *Use of Stereo Cameras to Track Faces*

The use of stereo cameras in face tracking mainly implements active vision for enhancing 3D tracking and reconstruction of the face. In a test conducted by [32] the fitting procedure uses a set of parameters already stored in the database containing information of each a person. The current information is then compared with 3D facial information stored in the database. An active vision system uses a set of stereo cameras that are used to collect multi-view images, which can be reconstructed using a 3D modelling system given the assumption that no motion occurs during the capture.

The theory presented in [32] mainly reasons that the improvement of facial 3D reconstruction can be achieved by using the stereo multi-view cameras that collect distinctive information, i.e., one camera tracks the face while the other identifies the features of the image. In the investigation two systems are deployed, that is, a top view stationery system and a top view active vision system as shown in Figure 3.

The results proves that the use of multi view stereo cameras significantly increased since the tracked face always stays in the center of the captured high resolution images. The table below shows the face recognitions made based on 10 attempts made on the subject portrayed as shown in Table 1.

Table 1

	Left	**Center**	**Right**
No PTU	***69.1%***	***85.7%***	***61.6%***
With PTU	***88%***	86%	91%

The principle of the test setup implements the pan tilt unit to enable the active vision system. The subject was recorded and pictures taken for three different positions; left, center and the right side, all of which As can be seen, the results denote that better face tracking, recognition and 3D modelling is accomplished with the pan-tilt-unit (PTU) which enables both narrow and wide field of view (FOV) capture for the stereo cameras. The cameras mounted on PTU help in attaining better quality for 3D face reconstructions. The unit allows for flexible collection of imagery data used for modelling the facial structure of the subject. The tests prove to be efficient when collectively integrated with an AAM model that defines the parameters for comparison. As long as the PTU motion can be regulated then an ideal capture can be obtained for the desired orientation. Face tracking and recognition by means of matching the data for a given face to database of faces.

C. Stereo Face Tracking System Performance

The face tracking performance is an issue that can be tactically solved by theoretical means of image positioning or calibration to diminish where the 3D model is generate. As described in [32], there is only one common method for achieving active stereo vision in face tracking and that involves the use of two pan-tilt-zoom (PTZ) camera sets in arrangement with a stationery camera for locating and tracking faces.

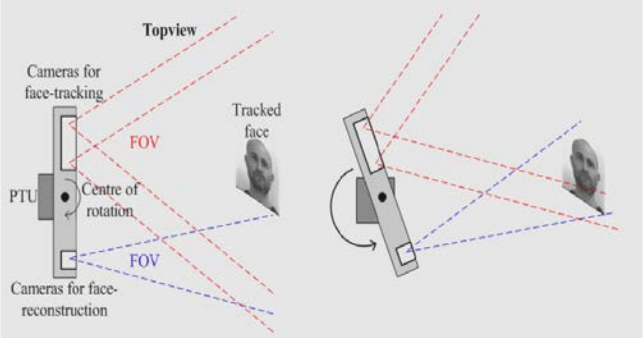


Figure 4

Although, the use of two stereo cameras mounted on a pan-tilt unit with fixed focal length offers more promising results for 3D Face tracking and recognition. In the method investigated, a nice review of the process is shown in the figure where one stereo camera is set to track the face in 3D and while the other stereo camera system performs a high quality reconstruction as a preliminary approach to 3D shape based face recognition [32]. Unlike the difficult process of using 2D image sets, the use of tilting cameras offers high quality 3D reconstruction, which enables consistent face recognition. Ultimately, the face tracking systems that are not based on 3D information are affected by number variants including illumination and pose. Since the shape of faces are not affected by the alteration pose or lighting, the use of a 3D face tracking and recognition system model has the prospective to improve and increase the performance of the Active vision stereo systems even under these conditions.

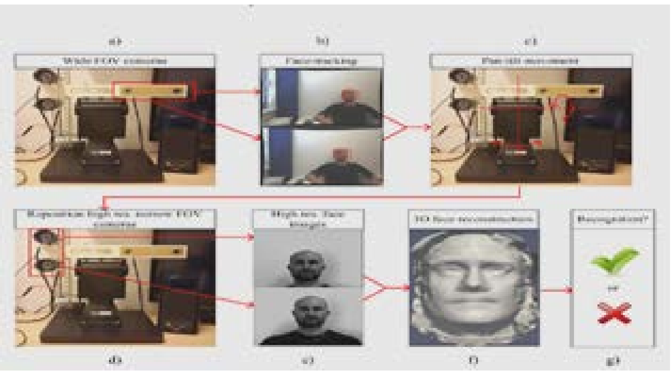


Figure 5

V. Challenges in Stereo Vision Cameras

Stereo vision cameras are used to stimulate triangulation methods in image processing. The utilization of images captured on stereo cameras coupled with 3D facial tracking methods tends to use the facial illumination changes and facial color variations, unlike 2D face techniques, which sometimes is difficult to triangulate especially at close ranges. The added advantage is that 3D facial tracking can exploit the relative insensitivity of facial shape images, thus, making is simpler to triangulate the 3D images of objects acquired from the stereo camera at short range.

Some of the challenges that face the ability for tracking systems to implement triangulation techniques like methods meant for the stereo vision based images include structured motion. Triangulation using structured light are subject to challenges like illumination or correspondence problem. There is also the issue of 3D reconstruction issues given that it is very hard to develop strategies or schemes that implement robust system algorithms for facial feature detection [21]. This will help in controlling certain challenges like the face orientation, facial expression, and variable illumination. For instance, when extracting image feature like the eyes, it usually occurs that they are open and covered in glass, or they are simply just closed; that is why these imagery structures are commonly referred as rigid features because they can only present rigid reactions or gestures. Despite the extensive research in 3D stereo vision, there are numerous concerns that still exist. Provided there are suitable similarity measures and a set of well-selected constraints, numerous methods may be able to produce excellent correlations and disparity maps. The separation of each of the cameras in the final reconstruction is an important factor in a standard stereo rig.

Accurate reconstruction is allowed by a wider separation of the cameras. However, increased difficulty in matching across stereo pairs is difficult because of increase of perspective distortion between the two images. The installation of Multi-view camera stereo systems is expensive considering it mandates more hardware and computational input to achieve accurate matching or correspondence.

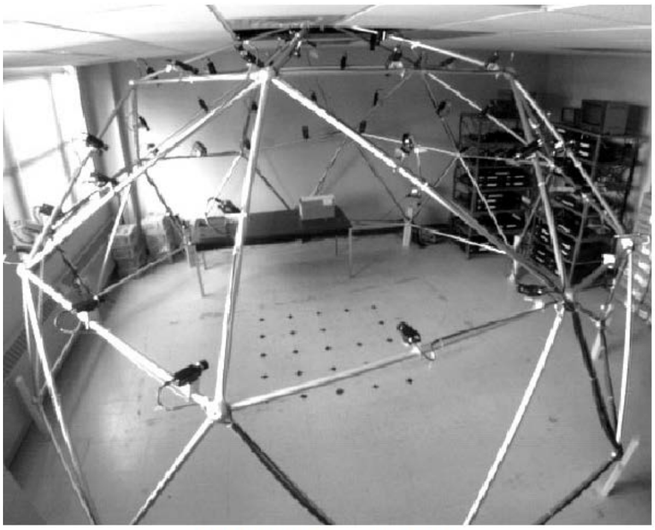
Another concern noted by scholars is occlusion associated with the position of the stereo cameras. For successful reconstruction to occur, it is significant to detect occlusion because matching occluded pixels normally results in errors. To compensate for this particular error, two or more cameras are used ensuring no point on the object is occluded. A process of constrained optimization is because of the dependence of each stage on previous stages and the nature of reconstruction. The occurrence of this error is common in every stage and hence the errors in correlating features between images results in errors at the next stage of reconstruction.

Another challenge in 3D stereo vision cameras is the issue of depth resolution affected largely by the accuracy of the correlation algorithm and the resolution of input images. Often, structured light based 3DE capture and laser scanner devices are accurate to less than a millimeter.

Other facial features that prove to be difficult to extract include the mouth. Compared to the nose, the mouth tends to be rigid. Thus, problems can arise when it comes to mining facial characteristics related to the mouth due to its rigidity. This is possibly caused by the faint image of the groove instigated by the open mouth features. Some exemplify challenges are further discussed below.

A. Correspondence Problems

The feature points in a correspondence match is what is referred to as the correspondence problem [17]. Scholars describe correspondence issues as the search in two or more 2D images for pairs of points that are projections of similar points in the scene. The task of 3D construction and depth extraction necessitates the establishment of precise correspondence. Failure to establish accurate correspondence makes reconstruction quite difficult. Bearing the assumption of accuracy of camera calibration information, communication errors are manifested in the 3D reconstructions and distortions.



CMU multi-camera stereo
51 video cameras mounted on a 5-meter diameter geodesic dome

Figure 6

Studies by Lina and Barron suggest the use of forward and backward image reconstruction as a means of quantitatively evaluating image-matching algorithms [27]. It is argued that the solutions to the correspondence problem can be tackled universally applicable techniques that apply multi-view geometry (epipolar geometry) for example as shown in the figure above [23]. Reports show that correspondence problem in stereo vision can be eliminated by the increased depth definition that multiple stereo cameras can offer.

Correct camera calibration is critical for stereo device settings to ensure that pixel matching in one image corresponds to that of the other image. The tracking problem brought about the geometric constraints of a 2D derived image makes it difficult to uniquely map and identify similar points in a given scene. Ambiguous correspondence can be diminished by deploying area-based stereo vision methods or feature-based stereo vision to compute disparity information.

As demonstrated by [32], both the area and feature-based approaches are modelled in the investigational set up by the use of the multi view stereo cameras. The paper proved that comparing existing entries of already generated 3D face models to that reconstructed that from high resolution stereo cameras showed a high level matching provided that field of view can be extended. This increases the workspace for tracking thus improving the reconstruction of 3D information collected. Multi-view cameras as exemplified in the figure 6 above directed towards a focal point or center of interest tends to limit the ambiguity in face tracking of points given that it provides a wide angle for information collection form left to right of the object.

B. 3D Reconstruction Issues

The reconstruction of faces from the collection of images under uncontrolled illumination and bearing different facial expressions and pose is a long-standing concern in computer vision. Most often, the state of the art methods is based on photometric stereo such as the pioneering work by Kemelmacher-Schlizerman and Seitz. Studies by Roth et al. has improved on the state of the art methods based on photometric stereo who unravel for a complete 3D mesh instead of a leveraged state of the art 2D facial landmark estimation and 2.5D height field [25].

Studies based on video data have yielded impressive results regarding the problem of reconstruction, tracking, and animating faces. Blended shape models used in the general state of the art techniques are also referred to as dynamic expression models representing faces and facial expressions.

According to studies in [22], earlier work relied on RGBD whereas it is now possible to address the reconstruction and tracking based on RGB video data in real time at the remarkable quality. Studies in [24] suggests regression models mapping 2D video frames to three dimension facial landmarks and then to register the DEM to the three dimension landmarks. The most recent addition to reconstruction issues in studies by is the synthesis of an increasingly detailed geometry inclusive of wrinkles.

VI. Conclusion and Future Work

Lately, methods that are more consistent have been deployed in test environments to contribute towards overcoming certain restrictions that come up when utilizing stereo vision cameras for 3D face tracking. Some the techniques being introduced are inclusive of the well-known structural matching approaches such as Active Appearance Models (AAP) and Active Shape Model (ASM). These methods are more robust compared to previous systems that manage the variations in the feature shape and image intensity.

Other even more interesting areas for future research include situations or circumstances that might require feature restoration due to noise, for example, an event might necessitate one to analyze and try to recover features related to the eye, nose, and mouth. Face tracking is an imperative problem for numerous applications, like biometrics, video communications, and video surveillance. Various approaches and techniques have been offered that realistically work well under reasonable variations of scale, lighting, and pose. The productivity level of these systems vary from tracked facial features to 3D pose estimation for the head location in a given or provided image frame. The leading challenge that future researchers should address ought to look into the robustness to changing environmental settings, occlusions, clutter facial resolution, and expressions.

References

[1] C.-J. Liao, S.-F. Su, and M.-C. Chen, “Vision-Based Hand Gesture Recognition System for a Dynamic and Complicated Environment,” 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[2] X. Zabulis, H. Baltzakis, and A. Argyros, “Vision-Based Hand Gesture Recognition for Human-Computer Interaction,” *Human Factors and Ergonomics The Universal Access Handbook*, pp. 1–30, Nov. 2009.

[3] M. A. Laskar, A. J. Das, A. K. Talukdar, and K. K. Sarma, “Stereo Vision-based Hand Gesture Recognition under 3D Environment,” Procedia Computer Science, vol. 58, pp. 194–201, 2015.

[4] P. Jiménez, L. M. Bergasa, J. Nuevo, and P. F. Alcantarilla, “Face pose estimation with automatic 3D model creation in challenging scenarios,” Image and Vision Computing, vol. 30, no. 9, pp. 589–602, 2012.

[5] M. Lauer, M. Schönbein, S. Lange, and S. Welker, “3D-objecttracking with a mixed omnidirectional stereo camera system,” Mechatronics, vol. 21, no. 2, pp. 390–398, 2011.

[6] N. B. W. Macfarlane, J. C. Howland, F. H. Jensen, and P. L. Tyack, “A 3D stereo camera system for precisely positioning animals in space and time,” Behavioral Ecology and Sociobiology, vol. 69, no. 4, pp. 685–693, 2015.

[7] J.-H. Lee, J.-H. Ko, and E.-S. Kim, “A real-time face detection and tracking for surveillance system using pan/tilt controlled stereo camera,” Real-Time Imaging VIII, 2004.

[8] M. Pfingsthorn, R. Rathnam, T. Luczynski, and A. Birk, “Full 3D navigation correction using low-frequency visual tracking with a stereo camera,” OCEANS 2016 - Shanghai, 2016.

[9] F. Abdat, C. Maaoui, and A. Pruski, “Real Time Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm,” *Human-Robot Interaction*, Jan. 2010.

[10] R. Okada, Y. Shirai, and J. Miura, “Tracking a person with 3-D motion by integrating optical flow and depth,” *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580).*

[11] D. Cho, S. Lee, and I. H. Suh, “Facial Feature Tracking Using Efficient Particle Filter and Active Appearance Model,” *International Journal of Advanced Robotic Systems*, p. 1, 2014.

[12] E. Maggio, E. Piccardo, C. Regazzoni, and A. Cavallaro, “Particle PHD Filtering for Multi-Target Visual Tracking,” *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 2007.

[13] M. Linna, J. Kannala, and E. Rahtu, “Online Face Recognition System Based on Local Binary Patterns and Facial Landmark Tracking,” *Advanced Concepts for Intelligent Vision Systems Lecture Notes in Computer Science*, pp. 403–414, 2015.

[14] Z. Kalal, K. Mikolajczyk, and J. Matas, “Tracking-Learning-Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[15] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, “Towards a practical face recognition system: Robust registration and illumination by sparse representation,” *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[16] T. Marciniak, A. Chmielewska, R. Weychan, M. Parzych, and A. Dabrowski, “Influence of low resolution of images on reliability of face detection and recognition,” *Multimed Tools Appl Multimedia Tools and Applications*, vol. 74, no. 12, pp. 4329–4349, Mar. 2013.

[17] A. K. Roy-Chowdhury and Y. Xu, “Face Tracking,” *Encyclopedia of Biometrics*, pp. 1–7, 2014.

[18] A. Hayasaka, T. Shibahara, K. Ito, T. Aoki, H. Nakajima, and K. Kobayashi, “A 3D Face Recognition System Using Passive Stereo Vision and Its Performance Evaluation,” *2006 International Symposium on Intelligent Signal Processing and Communications*, 2006.

[19] X. Yu, Z. Lin, J. Brandt, and D. N. Metaxas, “Consensus of Regression for Occlusion-Robust Facial Feature Localization,” *Computer Vision – ECCV 2014 Lecture Notes in Computer Science*, pp. 105–118, 2014.

[20] C. B. Liu, C. C. Chen, and X. Li, “Object Tracking System in Dynamic Scene Based on Improved Camshift Algorithm and Kalman Filter,” AMM Applied Mechanics and Materials, vol. 602-605, pp. 2061–2064, 2014.

[21] L.-W. Zheng, Y.-H. Chang, and Z.-Z. Li, “A study of 3D feature tracking and localization using a stereo vision system,” *2010 International Computer Symposium (ICS2010)*, 2010.

[22] S. Bouaziz, Y. Wang and M. Pauly, "Online modeling for realtime facial animation", *ACM Transactions on Graphics*, vol. 32, no. 4, p. 1, 2013.

[23] J. Read, "A Bayesian Approach to the Stereo Correspondence Problem", *Neural Computation*, vol. 14, no. 6, pp. 1371-1392, 2002.

[24] C. Cao, Y. Weng, S. Lin, and K. Zhou, “3D shape regression for real-time facial animation,” *ACM Transactions on Graphics*, vol. 32, no. 4, p. 1, Jan. 2013.

[25] J. Roth, Y. Tong, and X. Liu, “Unconstrained 3D face reconstruction,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[26] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, “PMF: A stereo correspondence algorithm using a disparity gradient limit,” *Perception*, vol. 14, no. 4, pp. 449–470, 1985.

[27] T. Lin and J. L. Barron, “Image Reconstruction Error for Optical Flow,” *Research in Computer and Robot Vision*, pp. 269–290, 1995.

[28] L. Hong and G. Chen, “Segment-based stereo matching using graph cuts,” *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*

[29] V. Kolmogorov and R. Zabih, “Multi-camera Scene Reconstruction via Graph Cuts,” *Computer Vision — ECCV 2002 Lecture Notes in Computer Science*, pp. 82–96, 2002.

[30] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*

[31] J. C. A. Read, “A Bayesian Approach to the Stereo Correspondence Problem,” Neural Computation, vol. 14, no. 6, pp. 1371–1392, 2002.

[32] “Enhanced 3D Face Processing using an Active Vision System,” Proceedings of the 9th International Conference on Computer Vision Theory and Applications, 2014.