Computer Graphics Comp 91.5460 **Rohan Girase**
Spring 2018 **0166353**
Rohan_girase@student.uml.edu

Literature Review 3

The primary article chosen by me for the literature review is **Hands in Space Gesture Interaction with Augmented-Reality Interfaces** , by the authors-"*Mark Billinghurst, Tham Piumsomboon, and Huidong Bai".* The above article was published in the journal " Computer Graphics and Applications, IEEE (Volume:34 ,Issue: 1)".

I particularly choose this primary article, as for me this is not just a random Computer Graphics related article , I am to this day fascinated by the concepts of gesture interactions and 3D designs, may be Hollywood is to blame here , watching movies like Starwars, Minority Report, Ironman, the Avengers and many more while growing up had a subconscious effect on my love for this particular type of technology where the protagonist just by the movement of the fingers on his hand is able to control a complex and highly intelligent machine like a Computer, It is as closest as possible to the feeling a Genie must be having, when he snaps his finger and voila there it is whatever he needs. .

Now enough of my background story ,the research paper under consideration is mainly about the overlaying of 3D graphics on the real world through augmented reality technology. The paper focuses on implementing the technology on both the desktop machines used by customers in homes and by big corporate in their swanky offices and also on handheld mobile devices held by the teen to the sexagenarian (person in his sixties).

The article further says that with the help of depth sensing and gesture-tracking technologies such as Microsoft kinect, Nintendo Wii, or the leap Motion controller we can track hands in space and provide free hand gesture input and thus we can enable the user to reach and manipulate virtual contents.

The second paper which is referenced in the primary paper selected by me for the literature review is **–"Markerless Fingertip-based 3D interaction for handheld Augmented reality in a small workspace"**, by the authors-"Huidong Bai, Tham Piumsonboon and Huidong Bai". The above article was published in the journal-" Wearable Computers, 2007 11th IEEE International Symposium".

This main reason behind selection of this paper is its resemblance with my primary papers subject, both these papers point towards development of interaction and rendering in the 3D space. This paper explains us briefly about the technologies and methodologies behind the area of interests such as-

1. Fingertip Detection 2. Fingertip Depth Acquisition 3. Marker Tracking 4. Co-ordinate transformation 5. Data Communication 6. Graphic Rendering 7. Gesture Interaction

Now with the desktop Depth-sensing and gesture-tracking technologies such as Microsoft Kinect , Nintendo Wii or the Leap Motion Controller we are asked to track hands in space and provide freehand gesture input. The way these technologies work is that they have developed desktop-based systems that exploit this technology to reconstruct a person's hands in real time and let him or her reach out and manipulate virtual content. For example, in the Phobia air spider phobia treatment application, users see virtual spiders crawling in the real world and can pick them up with their bare hands. A Kinect connected camera mounted above a tabletop tracks the user's hands. It captures point cloud data within a fixed distance so that the system can detect hands or other foreground objects over the table and segment them from the background. Once the system has found the hands, the virtual spiders respond realistically, crawling over the user's arms or other objects in the scene. The results appear on a desktop monitor, letting users see themselves and the virtual content overlaid on the real world. Depth sensors also enable an AR application to understand the real world surrounding the user and support gesture-based input using real objects.

The AR Micro machines game employs a Kinect for surface reconstruction on a tabletop, letting players drive virtual cars in the real world. A realistic physics simulation and interaction between the real and virtual objects let players place real objects for the cars to drive on or use gestures to directly manipulate the virtual content. AR Micro machines uses real-world point cloud data to drape virtual meshes over the real objects. This scene mesh updates in real time as the user moves objects. A physics engine checks for collisions between the virtual cars and mesh. Physics proxy particles represent the users' hands so that they can physically interact with the virtual objects, through collision detection. The system finds the hands by searching for skin colored regions in the point cloud data.

Whereas when it comes Mobile Devices the authors and its team has developed a visual marker less fingertip detection engine that can be used to manipulate virtual objects.3 Users can select a virtual object by pointing at it and can grab, drag, and drop it by making a pinching gesture with their thumb and index finger. This system uses skin detection and contour tracking to find the 2D position of users' fingertips. The algorithm runs in real time using the camera on an Android mobile phone with a 1-GHz processor. Implementing 3D hand tracking and gesture interaction on a mobile device is more challenging owing to hardware

Early research in gesture-based interaction with augmented reality (AR) interfaces used physical objects as input devices. For example, in Hirokazu Kato and his colleagues' VOMAR (Virtual Object Manipulation in Augmented Reality) application, users manipulated a handheld paddle to arrange virtual furniture in a scene. The paddle had an AR tracking symbol on it. As long as the paddle was in view of a camera, computer vision software tracked its position and calculated the user's hand position. The system responded to several simple gestures such as hitting, tilting, and shaking, each of which executed a different command. For example, users could move a virtual object by pushing it with the paddle or could delete it by hitting it with the paddle. More recently, Volkert Buchmann and his colleagues mounted AR tracking targets on a glove and used hand gestures to directly manipulate virtual objects. Users employed pinching gestures to pick up objects and then positioned them simply by moving the hand. However, users had to wear specially modified gloves, and the fingertip input worked only when the tracking symbols were in view of the system's camera.

Now from the secondary paper we can understand about the various prototypical implementation are mentioned as follows:-

1. Fingertip detection from RGB frame- Segment the hand region from the RGB frame based on a generalized statistical skin color model, and then use the threshold of the marker depth in the depth image to remove the noise, such as the table in the background with a similar color to human skin. After that, a distance transformation is applied to find a single connected component in the altered hand image, and curvature-based and ellipse-fitting algorithms are adopted to identify the fingertips of the hand region. A hand contour point is considered to be a fingertip with 2D coordinate once it has a higher curvature value than a defined threshold.
2. Fingertip depth acquisition from depth frame- The fingertip's depth that is defined is not the point value in the corresponding coordinate of the depth frame mapping from the RGB frame, but the average depth value of a small nearby region, which circularly fits the fingertip curvature while excluding the non-skin part. This can avoid the invalid depth value of a single fingertip point, which could be frequently located in the shadowing of the infrared area in the depth frame.
3. Market tracking- The AR tracking is implemented on top of a natural feature tracking library they have developed. To build this, first port the BRISK algorithm to Android with extra performance optimizations. Specifically, the detector checks each RGB frame to detect distinctive key points, and each key point is characterized by the descriptor based on the distribution of neighbor points. Then feature matching with a pre-defined target template is conducted and outliers are removed. Combine the detection and description parts with the optical-flow tracking and created a full-featured robust natural feature-tracking system. The optical-flow-based tracker will automatically start to track the matched points, and successfully tracked points are used to calculate the homograph matrix.
4. Co-ordinate transformation- The system projects the fingertip 2D position and the depth into the marker's WCS by using the homograph matrix obtained from the server tracking section, and then sends this data to the mobile client, in which this 3D coordinate is projected again into the mobile camera coordinate system based on the homograph matrix calculated from the mobile tracking procedure.
5. Data communication-Deploy a small socket library both on the desktop server and the mobile client, and use it for real time User Datagram Protocol (UDP) based

Rohan Girase
0166353
Rohan_girase@student.uml.edu

communication. The data communication is stable enough with slight delay in our private testing network.

6. Gesture Interaction- Three atomic operations (translation, rotation, scaling) can be conducted to manipulate virtual objects by using finger gestures. To do this, we map the translation value onto the position change of a single fingertip movement , while the rotation and scale values are controlled by the orientation and distance change between two fingertips in the midair respectively. To initialize this, three operational modes can be selected directly on the touch screen by pressing one of three buttons and a corresponding mode will be activated. Use a countdown timer to start or end an operation: keeping the finger tip relatively still in a tiny pixel region around the current position longer than a certain amount of time.

7. Visual scene rendering- We use the Open GLES 2.0 library to construct and render virtual objects for AR applications on the mobile phone.

As we can see from the above passages that both the primary and secondary papers are dealing with concept of tracking movement of hand gestures in space and provide free hand gesture input in order to enable the user to reach, control and manipulate virtual contents of the machine, device on which he is operating. Though all this appears to be path breaking and fascinating still there are many areas which are to be improved and researched upon.
So that we can come closer to dream of achieving what Hollywood has already achieved in movies and make this technology available to the common man at a reasonable cost.