Gregory Wright
August 2016

The Feasibility of Using Light Field Capture for Virtual Reality on the Lunar Mission One

Contents

Abstract

There are many technologies that are being developed that want to be used on the Lunar Mission One module. The problem is that they have very limited funds so decisions have to be made for what can and can't be brought to the Moon. This study looks into light field technology and its use to produce a virtual environment (virtual reality). It starts by considering some of the basic concepts and ideas of the workings of these technologies. The study then considers what makes a technology (and then primarily these technology) viable to be on the module. This includes the many uses of these technologies, drawbacks and alternatives of those uses and also considering the potential benefits that can be gained.

Introduction

Most by now would have heard of virtual reality and it being used to view a virtual world often using a headset. This is one of the more modern, less developed technologies and uses many different ideas to immerse the user into a lifelike experience. It is very similar to first person video games except the user feels presence. Virtual reality usually uses stereoscopic vision (A separate image for each eye) to produce the effect of being in a 3D world. The development of light field capture technology has only recently leaded onto the area of VR and the potential uses of it.

Virtual Reality and 3D Imaging

3D images are often produced by overlapping two similar 2D images of the same scenery but at slightly different angles on the same screen. To produce the 3D effect, the two images produced are perpendicularly polarised. The wearer would wear glasses with each lens containing a polarising filter. Each filter is aligned with one of the images, such that only the light from one image passes through each lens. Your eyes then merge these two images together; this process is referred to as stereoscopic vision which tricks your brain into producing the 3D effect. This is one example of stereoscopy. This is because each eye is seeing the same image but at slightly different perspectives. For virtual reality headsets a similar effect is achieved normally by using two separate screens (one for each eye) with different perspectives once again tricking your brain into thinking it is viewing the real world. A virtual environment is defined as a computer simulation that immerses the user in a world that they can explore and interact with.

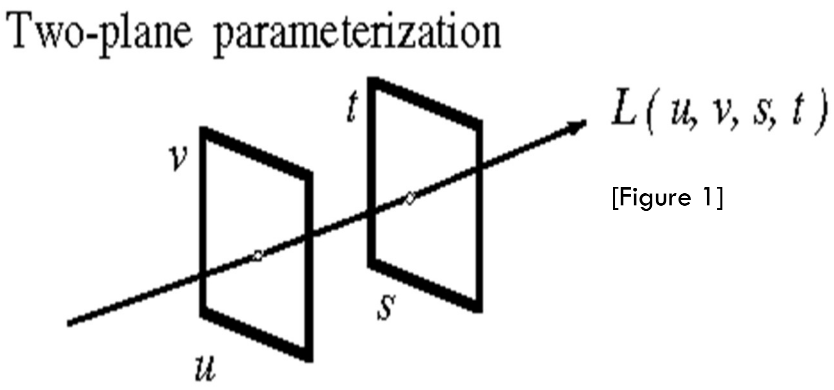
Immersion and Presence

Presence (or telepresence) is the being of within a virtual environment comprehending it as if it was reality by the user's subconscious. Immersion is where the user becomes consciously engaged with the virtual environment. For the user to feel immersed they could simply be enjoying being within the environment. It doesn't need to be a completely realistic model of an environment for presence or immersion to occur. It would be very complicated to make a user consciously believe they were within reality when in fact they were within a virtual environment. This in fact is unachievable with current technology and most likely illegal while being against human rights if it was at all possible. An acceptable level of fidelity can be achieved by the use of different visual cues as well as using other stimuli and sensors. These ideas are discussed in later sections.

Light Fields

Representations

A light field refers to every ray of light at every point in space (or the radiance at a point) travelling in a given direction. Light fields can be mathematically represented using the Plenoptic function. The 5D Plenoptic function holds five parameters. These are the coordinates of the light (i.e. x, y, z) and then the direction using two angles (i.e. ϑ, φ). Some extra parameters can also be included such as time and wavelength. This model can then be simplified by making some assumptions, this being that the light is travelling through free space and hence unaffected by external factors keeping the light travelling along a single direction and not being absorbed, reflected or refracted by other objects. This is then referred to as the 4D light field. The assumption made that leads to the loss in a dimension is that the plane that light arrives at is known. These assumptions allows for much easier computation. There are many ways to parameterise the function and each model has their individual advantages. One model is known as the 2-plane model [figure 1]. It consists of two points (one on each plane) with 2D coordinates with values in the interval [0, 1]. By convention the co-ordinates are named (s, t) and (u, v). This representation of the light field is often regarded as a light slab and a combination of light slabs can be used to represent a 360° scene.

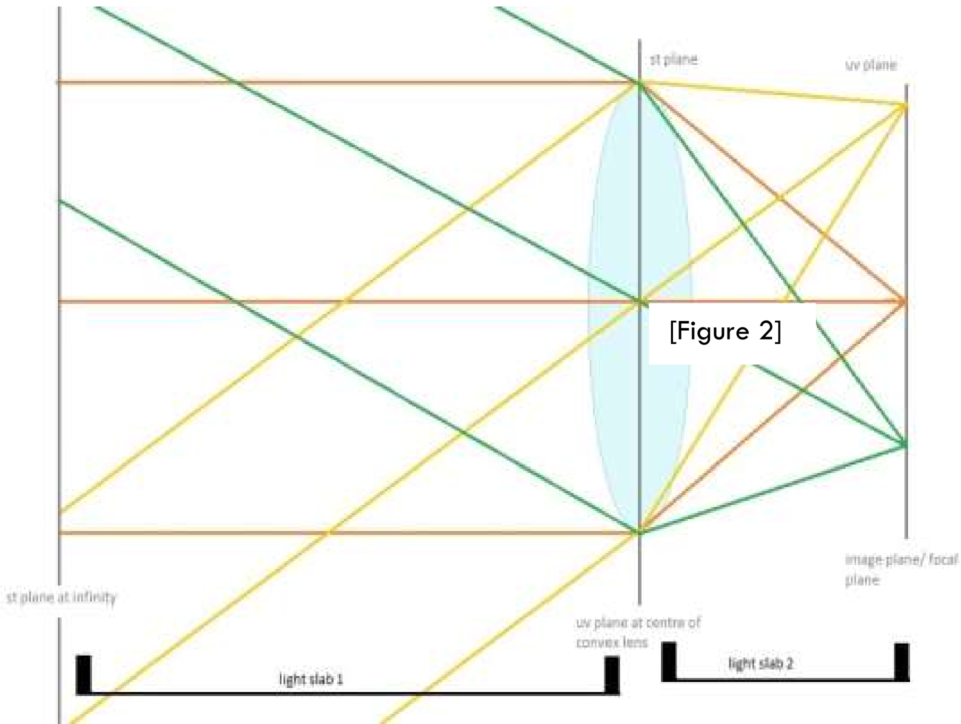


[Figure 1]

In space, objects so far away, can be assumed that the st plane is at infinity and the main lens as the uv plane. In this case all the rays of light will produce an image at the focal plane as the rays of light would be travelling parallel to one another for any given point in the light slab.

Capturing Light Fields

All film and digital cameras capture the same data on light, just different amounts of it. Cameras usually capture the light intensity and colour cameras also the colour of the light. This



[Figure 2]

image has been referred to as a two dimensional slice of the light field. For camera systems that capture the light field, they also get the direction the light is travelling. This allows for many more aspects of the scene to be looked at. Examples include depth maps as well as “refocusing” and perspective shift of the viewed 2D image.

Camera Array

It is where a magnitude of cameras (an array) each take an image of the same objects but at different perspectives at one time. Also each image could be taken at different times if the scene is considered static as there is no change within the light field. These images can then be used to produce a 3D model of a scene. There are various assortments of the array. Similarly, the same result can be achieved by using a single camera capturing multiple images. After taking each image, the cameras position is displaced before taking the next. This gives a different perspective of the scene. When taking multiple images of the same region/objects, there is difficulty finding out which parts of the images correspond with each other. This is known as the correspondence problem. For camera arrays (especially dense ones) this is an issue as there are many cameras that need to be calibrated and so complex algorithms would be needed to accurately map the light field. This is less significant for the plenoptic camera discussed in the next part. The plenoptic camera is technically a camera array because there are effectively many individual tiny sub-cameras (microlenses) fixed in place relative to one another.

Conventional Cameras

Figure 2 shows how a lens focuses light coming from point sources an infinite distance away with reference to light fields.

Before looking into how the plenoptic camera works, it is best to consider how a conventional camera works. For simplicity we’ll consider a camera with only a single convex lens. The light rays coming from an object on the focal plane refract through the convex lens and forms an image. Obviously only one side of the object would be captured. The object being captured diffusely reflects light at every point on its surface. Only some of these rays of light (at certain angles) arrive at the lens. The greater the distance of the lens, the smaller this angle becomes which tends towards zero. As well when the width of the width of the lens decreases, the smaller the angle becomes. This means that light rays entering the lens are travelling near enough parallel to one another from a great enough distance and of a small enough aperture. Parallel rays that refract through the lens form a point image on the focal plane. The parallel rays emitted from each point on the object travels a different direction to the lens and so forms a point image at different points on the focal plane. This is considered the image. The image formed is then captured by an image sensor that converts the light detected into electrical signals. This is how distant objects are captured or cameras with a very small aperture because the angle is near enough zero that all rays of light are parallel and all rays have a point image along the focal plane. This sort of setup means that close objects would become blurred. Objects closer to the lens from infinity, form an image a greater distance from the lens. This means the image sensor has to be moved closer and further away from the lens depending on the distance of the object (or vice versa) leaving other distant objects blurred. This is how a photographer changes the focus. This is one method of capturing depth information from a scene.

The change in the width of the lens can change the depth of field of an image. The smaller aperture increases the depth of field. This is because the increase in aperture size increases the circle of confusion. The acceptable circle of confusion is calculated dependant on a variety of factors. For maximum accuracy the circle of confusion should be less than the size of a pixel on the image sensor.

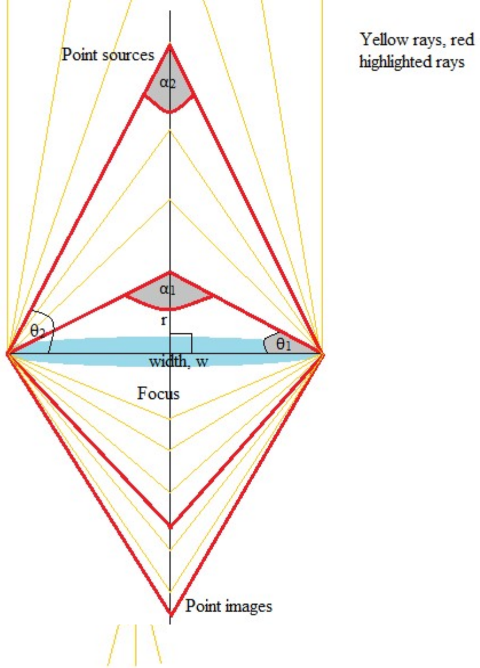
Specular reflection causes the glare and reflections seen in images on the surfaces of objects.

Figure 3 shows that as the distance of the point source tends towards infinity, angle ϑ tends towards 90° and angle α tends towards 0°. This shows that sources at infinite distance emit rays of light that are parallel.

Figure 4 shows that as the lens width reduce, the amount the light spreads out also reduce over a given distance after the point image.

Plenoptic (Light Field)

The plenoptic camera is similar to the conventional camera but instead contains a microlens array (often in the thousands) in front of the image sensor. They usually, but not necessarily, have a main lens that focuses the light onto the microlens array plane. The number of microlenses in the array often determines (and usually equivalent to) the spacial resolution. Normally the individual pixels on the sensor will average the intensity and colour of the light rays and that is then recorded. Instead each microlens refracts these rays further onto an assigned macropixel. Each macropixel contains sub pixels. Each subpixel would capture a different perspective of the lens. Each subpixel is matched with all the subpixels for a single perspective and that produces an image. The greater number of subpixels, the more perspectives can be



achieved however the spatial resolution decreases as there are the same amount of pixels available but less per image. By combining the different subpixels, different perspectives can be achieved. The plenoptic camera is limited to the fact the maximum amount of perspective shift is equal to the width of the main lens. Even the largest lenses on plenoptic cameras are no more than a few cm across.

Virtual Reality

Why should light field capture be used?

Having captured a light field, it allows for all information of a scene to be obtained. This means that any slice of the light field can be reproduced. Now an image can be obtained with any depth of field, field of view or perspective from the light field captured.

There are many differences between a 360° camera system and a light field capture system. For the 360° conventional system, to produce a 360° image from multiple images, each image has to be stitched together. There are many issues with this approach. It requires time to stitch the images together although the processing for the light fields would be significantly greater. The image quality about the stitched regions are poorer and blurred although clever image processing often hides this.

One depth cue for our mind is that objects out of focus become blurred. Virtual reality headsets currently do not have this capability. This means that all objects within the image viewed appear in focus as the screen is a fixed distance away from the retina. This reduces the immersion of the user and could worsen sickness or discomfort. The light field could be used to alter the depth of field dynamically depending upon what the user is looking at. This can be achieved by using real time eye tracking.

The light field capture correctly records the reflected specular lighting. This means correct alterations in lighting and glare when the perspective is changed. Conventional 3D modelling of the real world using a depth map produced from stereoscopic images or using a depth-sensing camera (by photogrammetry) struggle, if not fail, to alter the lighting correctly. This being a benefit as the world behaves more like one would expect. Although complete realism isn't a necessity, in this case it'll improve the immersion of the user as they normally wouldn't notice it consciously but do so at a subconscious level. This depth cue is due to the parallax caused when looking at an object from different perspectives.

Potential Involvement of Technology

How it could be used

The light field data can be reused to produce images with differing properties and features. When a new feature is developed, more data doesn't need to be collected. The original information already held on the light field can be used again.

Using light fields over 360° images

The company SpaceVR are planning to launch the first virtual reality satellite into orbit in 2017. It consists of a 360° camera system. This approach has many issues as mentioned above such as being less immersive than that could be achieved from using light field technology. When looking into space, the light is travelling such a distance that it can be assumed that all rays are travelling parallel and so produces an image along the focal plane. With such a great distance it means that there is only negligible change in perspective for the viewer (and disparity between images) so there is no need to use light field technology as everything viewed looks the same from any of the available perspectives. For objects viewed closer to the module such as viewing the Earth, the Moon or passing space debris or other vessels, light field technology is better to be used. As the light from the sun radiates these objects, specular reflection is likely to occur. This is more significant in space as there isn't an atmosphere to absorb some of the light. As the user turns with what likely would be a head mounted display, the glare doesn't change. The reflections off the surfaces don't change. This is for a conventional 360° image (or video). Light field information can be used to produce a 360° view but also can correctly display the changes to the surfaces viewed at different perspectives (from the specular reflections). This differs from the conventional camera systems.

Feasibility of Technology for Mission

Economic Feasibility

The biggest issue by far restraining the technologies being used in space is the cost to get it up there. If it costs too much it simply won't be affordable to go at all. The main consideration has to be the mass. The greater the mass of the rocket, the more thrust that's then needed to get into Earth's orbit. This would then require more fuel (and bigger engines if the increase is significant) which once again increases the mass of the rocket. This also means extra infrastructure is needed to then hold the extra fuel. The cost to send only a kilogram into Earth's orbit is currently within £10,000's. A technology would then need to be as light (low mass) as possible to reduce the economic impact for the launch. Captured light fields are usually considered as very redundant and so large amount of secondary storage (at least in the terabytes) would be needed to hold the data collected. This would then require more mass. An on-board data store would already be planned on the module but may not be considered to have such a capacity to store the light field data temporarily before it's transmitted. This then means a greater capacity would be needed which increases the mass of the module. The physical size and extra mass of the storage would be minimal but makes a significant increase in cost due to the great expense of space travel. For the mission, there would be a recommended limit of the mass of the module over which it is likely it'll becomes too expensive to launch into space. This means that there would be a need to refine the cameras to be as lightweight as possible.

Technical Feasibility

There are many methods to capture a light field and can be done so using one or multiple cameras. The plenoptic camera has its benefits of being a single camera meaning no calibration. The main issue is that the spacial resolution is reduced with a greater number of perspectives so is challenging to produce a camera with a significantly small volume but with a high enough resolution required for virtual reality. This wouldn't be as much of a challenge for a regular camera however as the camera developers only have to consider packing more pixels per unit area within the image sensor for only a single perspective. Light field camera developers have to pack more microlenses within a unit area to increase the spacial resolution but also need more pixels to have multiple perspectives (greater angular resolution).

There are going to be a number of cameras dotted around the hull of the module. These can be repurposed to capture the whole light field by becoming part of a camera array with many additional cameras. To improve the accuracy and perform easier calibration, a regular ordered array would be needed. Due to the shape of the module a spherical/cylindrical array would be more suitable than rectangular as it would mean minimal extra work is needed to position the cameras. Also these setups would consist of a smaller surface area so fewer cameras needed for the same angular resolution. To give a high angular resolution a large number of cameras are needed to reasonably immerse the user. When looking into space, there isn't a need to have many perspectives of the same scene. So only a few cameras would be needed to take images of this scene. For images taken of nearer objects, let's say the surface of the moon, much more information is required (i.e. different perspectives) and so more images need to be taken. As there would be a combination of distant space and near objects, it may be most suitable to have a squished hemi-spherical (dish like) array with a few cameras dotted on the other side of the module. The dish would be located on the side of the near object (i.e. the moon) and the few would be facing towards outer space. This would mean the dish would be on the bottom of the module when it attempts to land. This could be problematic as the module may have to rotate to take images of near objects or could have problems when there are near objects on opposite sides of the module at the same time. To get around this, the cameras may not need to be permanently fixed in place but can rotate around the module when required. This would cause more problems in itself because the system would be even more complex as well expensive due to the extra infrastructure required to physically move the cameras. Another option would be to have a consistent camera density around the module and only certain cameras are used depending upon the scene. This would mean the extra mass of cameras and infrastructure for them to be placed in a regular pattern. Another issue with this solution is that the on-board computer system would have to determine whether a light field should be captured for a scene and what images are needed. This could be achieved by predetermining points of interest and how many cameras should be used to take images within each direction before launch. Another option would be to use the disparity between images (parallax) to determine whether two objects are far away or close together. These images would be stored temporarily to not waste storage space and would be taken from a selected number of cameras such that a depth map of the complete scene can be determined. The closer objects (of most disparity) would have more images taken of. The further objects (of least disparity) would have fewer images taken of. A complex algorithm would be implemented to determine which cameras should capture images over time by using this information. This approach would add to the amount of development required but the majority would be down to programming so additional resources wouldn't be required on the module. The correspondence problem once again arises however would be a very similar issue to that for capturing the light field. One more approach would be to transmit 2D image data on a scene back to Earth. The decision is then made at mission control on what is captured and what cameras are used to capture the light field. This approach would need to be carried out within a relatively short amount of time if the scene is changing quickly.

One approach to reduce the file size would be to compress the data prior to transmission. This would be problematic as this would require large amounts of processing and to be most efficient would take a significant amount of time. Most likely a compromise could be reached between compression prior to transmission and the amount of data transmitted to allow for optimal length of transmission of the light field data without reducing the quality of the capture. Fortunately light field data captured is redundant so large amounts of compression is possible without reducing the information on the light field. Lossless compression techniques could then be used to further reduce the file size. Although lossless is considered less compressive than lossy compression, no accuracy is lost and redundancy is high so still would be a significant decrease in size.

There are many considerations on what is within the virtual world and what the user can do within it. This is dependent upon having the necessary cues available to being reproduced and manipulated. For the user to look left and right it is necessary to have information on the scene outside the field of view. In many circumstances this is the whole 360°. This is then similar for looking up and down. For the user to move left and right, different perspectives are needed to be captured. This is why multiple images are taken from different viewpoints. The data collected is used in calculating what the user would see at different viewpoints. This is done by using a technique known as interpolation. As each eye would see a different perspective of the scene, there would be parallax. This would mean that the two separate images shown to each eye would need the necessary disparity for different objects within the scene. Parallax is another cue that helps the mind to determine how far an object is away. This needs to be simulated correctly for the virtual environment to help reduce sickness and discomfort.

Either an inward or outward facing view can be captured. With the light field capture an outward facing view would be more useful for commercial use whereas an inward facing view could be better for monitoring progress of the drilling. Both of these views could be combined to produce a complete model of the light field.

Lunar Mission One is being crowd funded. They rely upon there being enough support and public interest to raise the funds they need. One way they can gain that interest is by the use of light field capture to view the Earth, Moon and stars from space. Also experiencing what it's like on the Moon's surface. If full immersion is achieved and the user of the VR headset feels presence, the experience would be much more engaging than any previous attempts. Virtual scenes that are not on the real world are easy to simulate in VR as they are based upon 3D models (synthetic virtual environments). The difficulty with VR based on the real world is that it's much harder to produce a 3D model to use. The 4D light field capture can get round this problem. The information collected can be used to model where the light would be if the camera position had been placed somewhere else, in essence virtual cameras that follow the motion of the user's retinae.

---Used for monitoring, collecting information. ---i.e. logging progress of mining. If there is an issue with the operation, may be useful to see what happening. ---Could be done with light fields.

The youth could also be influenced by the VR for the space mission and such gain an interest in science. - By them gaining a feeling of presence it could have some positive social effects on their attitude towards the world. One example would be the overview effect experienced by astronauts watching the Earth from space. They see how small it appears with the rest of the world, how small the human race is. We are only a single planet in the vast universe. All could experience this without actually having to leave the atmosphere and at relatively small cost.

Another way funds could be raised is by selling the right to use the captured scenes to businesses. One possibility could be to the universities (and later companies) developing CAVE systems (Cave Automatic Virtual Environment). They often involve putting the user into a room filled with stereoscopic wall projections. More recently the use of the light field is allowing for manipulation of real scenes and being able to replace what the user is seeing to follow the change in their perspective as they move around the room.

There are currently plans of having a robotic arm on the module. The primary use was meant to be helping with storing mine samples. This can be repurposed when not being used to do this. The robotic arm would be used on the moon's surface. By attaching a camera to the arm (which probably there would be anyway) and rotating it about its axes, a light field can then be captured. This means that only extra functionality would be required for the robotic arm which may be achieved by some additional programming. The use of the arm shouldn't then increase the amount of resources required for the mission. There may be some issues though. The arm would likely be fixed in one place and may not be able to make complete rotations or what the camera would be able to capture might not be that interesting or useful. Another concern would be if the robotic arm would shake when it is translating the camera. To get round this, the computer system would require knowing the direction the camera is facing and the position it is at for the duration of capture.

There is a variety of uses of light field capture but the level of involvement would be determined by the progress made in the technology. As a large amount of data is required to represent the light field to be captured, there may be a need to make some limitations. It may be difficult to record and store full 360° high quality information. By reducing the degrees of freedom, less data is captured reducing the file size. This would mean about only half the storage space (and cameras for an array) required if 180° was used instead. Another alternative would be to lower the angular resolution by reducing the number of cameras to be used in a camera array. Similarly this could be achieved by reducing the number of images taken from a single camera and increasing the displacements between each image. Interpolating could be used to determine what the images displayed to the user should look like when an original image wasn't taken from that perspective. This should help in smoothing any negative effects caused by this reduction. Another method could be reducing the spatial resolution of the images taken. For the user to experience presence they may not need to view an ultra HD photorealistic virtual environment. By reducing this aspect it would lead to a reduction in cost. Alternatively the pixels not used for spatial resolution could be used to improve the angular resolution with the plenoptic camera. One consideration would have to be whether to include the time dimension when capturing the light field. A still capture would be significantly less intensive than the large number of frames per second required for VR headsets. The limitations used would be dependent upon the purpose of the light field capture. If the capture was during the motion from Earth to the Moon, only a still light field would be needed as there would be little change in the field over time. Possibly a time lapse effect could be achieved by capturing a light field every few thousand kilometres. This would be acceptable as it would appear to the user that they were travelling much faster instead of jumping from one point to another in space which would otherwise increase discomfort. If the view was during the module landing, the complete light field and changes over time should be captured.

Time restraints

Another region for concern has to be the relatively short time frame that the technology requires to be developed. The launch (planned to occur in the 2020's) requires vast amounts of development and research to make the purpose of the mission achievable. There are many aspects of the technologies discussed that would require development before launch. Some of these aspects include the compression of data before transmission and the transmitter itself such that the rate of data being sent is high enough. There is also the development of light field technology. This would require many researchers and developers across many fields.

Summary

Uses of Technology

There are many uses of the light field technology which primarily consist of being a superset of the camera systems otherwise would be used on Lunar Mission One. By capturing the light field, a virtual environment should be created more accurately of the real world.

These uses include:

- Monitoring progress of the drilling systems visually.
- Providing the user an experience of the Earth from space similar to the overview effect.
- Influencing youth, encouraging them more into the world of science.
- Raising funds such as by charging users to experience space, the Earth and the Moon.
- Raising funds by selling the right to use the scenes captured to companies.

Possible Limitations

The main issue with this technology on the mission is the amount of storage space required to hold the data collected about the light fields. Because of this, the capture may be limited to some degree depending on the circumstance.

These limitations include:

- Reducing the degrees of freedom (perspective shift) that can be achieved.
- Reducing the angular resolution by either reducing the number of cameras for an array or reducing the number of images taken by a single translating camera.
- Reducing the spatial resolution.
- Reducing the frame rate by sampling less frequently.
- Capturing only a still light field and not considering the additional time dimension.

Dependencies

For the light field virtual reality system to be used for Lunar Mission One, there are some requirements that need to be filled. These requirements are dependent on the options used to capture, process and transmit the light fields from the module. The necessary dependencies are listed as such and without light field capture cannot be possible.

These dependencies include:

- A robotic arm being on the module and is available to be used and hold a device for image capture. The arm has sufficient rotation and displacement available to capture a light field. The location of the arm allows it to view regions that have some form of interest (scientific or commercial).
- The computer systems aboard the module consist of adequate processing capability to compress large amounts of data collected on the light field. This compression has to be accomplished within a trivial amount of time and can be stored within the limited secondary storage capacity. Essential.
- The scene captured for a single translating camera has to be static so there are no changes within the light field. It could otherwise lead to errors in the capture.
- The information has to be transmitted within a reasonable time frame so requires enough compression to adequately transmit the data relatively quickly.
- The module requires a transmitter which can stream data for extended periods of time and at high enough speeds.

Conclusion

The use of light field capture for virtual reality has many benefits that make it usable for Lunar Mission One. There are however many problems that would need to be overcome such that it is used effectively. There are a variety of methods that could be chosen to implement light field capture as mentioned within this report. Each method has its own unique advantages yet come at different costs. By costs referring to the difficulty in achieving technical capability and also the finance required for development and use on the module. There are likely many more concerns that haven't been raised for each approach. Overall light field capture can be used on Lunar Mission One. This information can then be used to produce virtual environments successfully. The technology is feasible in a sense of simply capturing light fields for virtual reality but may not be the case for all approaches and only to a certain extent. It is indeed achievable within the time frame. There are currently developments by many companies and researchers into the use of light fields. There are even 360° video light field camera systems currently being developed that are to be used for virtual reality. The most feasible approach would have to be the use of the robotic arm to capture a light field. It would only require a single camera system with as much data as required can then be collected. Any angular resolution could be achieved as the camera can be translated. This would be dependent upon the precision of movement of the robotic arm. The increase in mass of the module would likely be insignificant as there is already a robotic arm planned to be used. This would make it the cheapest approach to be launched. There are still many concerns to be considered such as when and where the arm could be used for this purpose. The main issue is that only static light fields can be captured. The other methods would also be feasible yet once again it is the extent which would need to be considered. The approaches would be limited to a certain quality dependent upon cost (mass). The quality is referring to the angular and special resolution but also the quality of the experience that can be achieved to immerse the user and make them feel presence within the virtual environment.

References

The knowledge and understanding on light fields and virtual environments used within this report came from many sources. These mainly included papers and online resources. The known sources are listed below.

- https://cseweb.ucsd.edu/~ravir/6160/papers/p31-levoy.pdf
- http://persci.mit.edu/pub_pdfs/plenoptic.pdf
- https://graphics.stanford.edu/papers/lfcamera/lfcamera-150dpi.pdf
- http://web.mit.edu/persci/people/adelson/pub_pdfs/elements91.pdf
- http://web.stanford.edu/~bgirod/pdfs/MagnorCSVTApril2000.pdf
- https://graphics.stanford.edu/courses/cs348b-97/lightfield/lfparam.gif [Figure 1]

Acknowledgements

In writing this report, aid was provided by Edward Zieba who helped with finding some useful resources as well insight into many of the ideas discussed.