Elemental Image and Audio Synchronization and Transmission Technique for Glasses-Free 3D TV System

by

Amit Hasan Khan
15101083
Fairoz Nower Khan
15101084
Noor E Nisa
14201044
Ashraful Jannat
18241052

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Amit Hasan Khan
15101083

Fairoz Nower Khan
15101084

Noor E Nisa
14201044

Ashraful Jannat
18241052

Approval

The thesis/project titled " Elemental Image and Audio Synchronization and Transmission Technique for Glasses-Free 3D TV System" submitted by

1. Amit Hasan Khan (15101083)
2. Fairoz Nower Khan (15101084)
3. Noor E Nisa (14201044)
4. Ashraful Jannat (18241052)

Of Spring, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 25, 2019.

Examining Committee:

Supervisor:
(Member)

____

Dr Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

____

Md. Abdul Mottalib
Head of Dept.
Department of Computer Science and Engineering
BRAC University

Abstract

Previously integral imaging based 3D technology was proposed only for still image. We propose a novel technique to synchronize elemental images and audio signal and the transmission technique for glasses-free 3D TV system based on integral imaging. The main idea behind the method is to generate 3D video based on elemental images synchronized with audio stream. The system uses the depth information and RGB data of per frame of a video through Intel RealSense 3D camera and the audio stream from microphone. The audio file is sampled according to per frame duration of the video and kept in different buffers but having same index. The frames are divided into elemental images using Elemental Image Generation algorithm and the audio signal is synchronized according to the index. Then the stream of elemental images and corresponding audio data is transmitted to data server for storage. HLS streaming protocol is used to stream the TV content. A dedicated web application was made that fetches data from the server and plays video on the user end display device. By using multi-array of lenses in front of display, the video is viewed as three-dimensional with the help of integral imaging technology. The 3D TV content has an output rate of 30 frames per second and a calculated viewing angle of 18.9°. As integral imaging is an auto stereoscopic method to represent depth perception, it frees the viewer from wearing any 3D glasses.

Keywords: Elemental image, 3D TV, Glass-Free, Integral imaging, Streaming protocol, Internet TV, Synchronization.

Acknowledgement

First of all, Thanks to Almighty for giving us the ability to complete our task successfully.

We would like to express our sincere gratitude to our supervisor Dr. Ashraful Alam for his utmost attention and valuable time. We would also like to thank him for giving us the opportunity to work on this topic and guiding us throughout the process.

Additionally, we would like to thank our family members to support us mentally in our hard times.

Nonetheless, we would like to show appreciation to the Department of Computer Science and Engineering, BRAC University and CVIS research lab for providing us with all the fundamental help.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

3D television (3DTV) delivers the interpretation of depth to the audience by exercising methods like stereoscopic display [1], multi-view display [2], or any other kind of 3D display. Nearly all contemporary 3D televisions makes use of an active shutter 3D system or a polarized 3D system, and some are auto-stereoscopic lacking any external viewing aid [3]. Multiple methods exist to generate and exhibit 3D moving pictures [4], [5].

Depth perception in human visual system (HVS) is a complex process where the information received from the two eyes is processed within the brain. Later the brain converts the raw data into meaningful visual figures. The depth perception in brain is achieved using various cues such as, monocular, binocular, stereoscopic, and monoscopic etc. Stereopsis in HVS refers to perception of depth based on binocular disparities. Left and right eyes receive two slightly dierent images of the 3D scene due to interpupillary distance (IPD) (also called interocular distance), which results in positional differences of the objects. These positional differences (usually in horizontal direction), referred to as binocular disparities, are processed in the visual cortex of the brain to yield depth perception. This phenomenon can be simulated artificially by presenting the eyes two images of a 3D scene captured at slightly different viewpoints and it is called stereoscopy [6].

3D displays can be categorized into 6 classes: stereoscopy, auto stereoscopy, multiview-autostereoscopy, integral imaging, holography, and volumetric displays. The principle of stereoscopy imitates the human visual system by presenting the viewer two images of a 3D scene captured at slightly dierent viewpoints. Such images are called stereopairs or stereograms and are usually captured with a stereo-camera. The two images are projected simultaneously. Special glasses are required to filter the two images to the appropriate eye. Active and passive glasses are two types of 3D glasses. Active glasses interact with the display device to synchronize the visualization process, whereas the passive glasses do not communicate with the display device-they separate the constant 3D stream for left and right eyes.

Multiview autostereoscopy (also called multiscopy or automultiscopy) is an extension of autostereoscopy that provides more realistic 3D experience by rendering different set of stereo images for different viewing zones. The same techniques of parallax barrier and lenticular lenses are used for stereoscopy. This technique lacks the horizontal parallax which means when the viewer moves horizontally while viewing the display, he will see the same 3D scene which is quite unnatural. Multiview autostereoscopy provides horizontal parallax by rendering more than 2 views, upto 50 views or more. As a result when the viewer roams around the display he experiences a new set of views in each viewing zone giving a natural 3D experience.

Integral imaging is an extension of lenticular multiview autostereoscopy where the one-dimensional lenticular array of lenses is replaced with two-dimensional array of microlenses. Each microlens in 3D display produces a unique perspective of the 3D scene to the viewing eye based on its distance and position in front of the display. The acquisition of 3D scene in integral imaging is carried out with an array of very small lenses in a homogeneous arrangement with the display array of microlenses. Each lens based on its position and focal length, captures a unique 2D projection of the 3D world. These 2D images are known as 'elemental images'. In integral imaging visualization, the 2D display device projects each elemental image onto the corresponding microlens in the lenticular array. Then each microlens optically relays its elemental image into the 3D space, the overlap of all elemental images creates the real 3D scene.

Holography was invented by Dennis Gabor in late 1940s and the modern form of holography was introduced by Yuri Denisyuk after the development of the laser technology in 1960s [6]. Holography is a three-dimensional visualization technique that uses photographic recording of the light field of 3D scene; these recordings are called holograms. Laser beam is used to capture the 3D light field reflected from the illuminated object. The laser beam is split into reference beam and object beam (also called illumination beam) and their interference pattern is recorded when they meet on a holographic film. For visualization the film is projected with the laser light which creates the same light field as it was originally captured, reproducing the original 3D scene.

Volumetric displays produce 3D illusion by rendering multiple slices of a 3D scene in a defined volume of space. Though the concept of volumetric 3D technology is more than a century old, it is still immature and under development - far from the public use. Various devices have been proposed in past for volumetric display; none of them is officially accepted. In swept-volume technique of volumetric display, the 3D scene is divided into slices which are one by one projected onto a rotating surface. The human eye is able to perceive the 3D scene due to the persistence of vision property.

3D televisions are getting increasingly popular as time progresses. New technologies such as Visible light communication are being integrated into 3D television as the demand for 3D TV rises. Implementation of Visible Light Communication allows information acquiring straight into TV as LED lights delivers information by flickering at high frequencies.

The continuously increasing demand of new media technologies needs the rapid improvement in 3D display technology. In present time, 3D glasses are required to convey the depth perception of a motion picture to the viewer. These glasses only work with passive 3D technology. In our system, we approached the autostereoscopic method with integral imaging technology to deduct the dependency of wearing 3D glasses. However, a glass free 3D technology was proposed for still images [7]. Our proposed model synchronizes audio with per frame of motion pictures for a 3D TV system .

1.2 Literature Review

The first step to attain glass-free 3D technology is integral imaging. In the past few years, Integral Imaging has garnered the attention of a good number of researchers and enthusiasts, evidenced by the surge in relevant journals and conference publications. First proposed by Gabriel Lippmann in 1908 [8], it is a 3D imaging method that can produce full color, full parallax 3D images without the assistance of any larger objective or viewing lens. The technique is basically composed of numerous distinct elemental images viewed through an array of micro lenses. In spite of its benefits[9][10][11], Integral Imaging has its share of constraints including narrow viewing angle, inadequate resolution and shallow image depth [12][13]. Narrow viewing angle is a major limitation which is caused by restriction in the space where each elemental image can be displayed. This drawback alone is constraining the commercialization of the technique. A number of studies have been conducted to solve it including lens switching method [14], the curved lens array technique[15], moving lenslet arrays with a low fill factor[16], micro-convex-mirror arrays[17] and multiple-axis telemetric relay system[18]. These methods however only can minimize the problem, but not to an extent so that it can resolve the drawbacks to the commercial implementation. A few efficient methods have been verified to solve the problem. Choi et al. demonstrated a multiple-viewing-zone integral image display using a dynamic barrier array that generates multiple viewing zones by directing the light rays emitted from the elemental images using a dynamic barrier array[19].

Alam, Kwon, Piao, Kim and Kim (2015) demonstrated a viewing-angle-enhanced integral imaging display system that uses a time-multiplexed, two-directional sequential projection (TTSP) scheme and a directional elemental image generation and resizing (DEIGR) algorithm[20]. The system consists of three processes: acquisition of depth and color information of a real object using sensor, generation of two sets of directional elemental images contemplating two different angular perspectives using a DEIGR algorithm, and projection of two sets of DEIs using the TTSP scheme. The experiment conducted during the research demonstrated that the viewing angle can be enhanced almost two times compared to other existing methods.

Integral imaging is made up of two stages: pickup and display. During pickup, the object is imaged using an array of lenses, and the image is produced as an array of parallax images of the object, called an elemental image array (EIA). To execute a system that is capable of imaging real-world subjects, the EIA of the subjects must be produced in real time, with the system equipped to support real-time computation of the EIA. Jeong et al. proposed a system for generating EIAs using a central processing unit (CPU) and graphics processing unit (GPU)[21]. This method utilizes the parallel computing architecture of the GPU, allowing concurrent computation of the EIA from the real-world objects in real time. A depth camera is used to obtain the depth and RGB data of real-world objects. Then GPU parallel processing is applied allowing real-time computation of the EIAs for real-world objects. The produced EIAs are exhibited on an LCD panel using a lens array. The experiment conducted during the research shows elemental images generated at a rate of more than 30 frames per second using a quad-core 3.0 GHz CPU and a 256-core GPU.

Park, Hong and Lee (2009) discusses about Integral Imaging as an alternative to other 3D techniques like stereoscopy and holography. Acquiring and reassembling light beam scattering utilizing a lens array, Integral Imaging implements an efficient approach to obtain 3D information about the subject, which can later be processed and visually recreated for 3D displays. Integral Imaging also provides an implicit benefit of uninterrupted vertical and horizontal motion parallax. Originally proposed as a method for 3D display of fixed images, the field has now broadened to involve more features of 3D data processing. Due to its' wide scope, Integral Imaging can be implemented for developing industrial applications in future [22].

Xiao, Javidi, Martinez-Coral and Stern (2013) explore the applications utilizing the Integral Imaging technique. One of the applications includes underwater 3D visualization that implements an Integral Imaging based three-dimensional reconstruction algorithm to perceive 3D image in obscure water. The paper further examines other implementation of Integral Imaging involving photon counting, 3D microscopy to discern and recognize cells, 3D tracking of objects etc. The exercise of the technique in fields among the likes of medical imaging and industrial manufacturing demonstrates its commercial viability [23].

Min, Hahn, Kim and Lee (2005) proposes and demonstrates a three dimensional display technique based on image floating and integral imaging. The integral imaging method was used because the images produced have volumetric characteristics. The proposed system consists of an electro-floating system with an integral imaging based display system. The electro-floating system can produce 3D images with a better sense of depth. The electro-floating system proposed can be used in various applications [24].

Choi, Kim, Lim and Lee (2005) presents a new computer generated holographic three dimensional display system utilizing integral imaging lens array. In a traditional holographic system, the viewing angle is inversely proportional with the resolution. The proposed system in the paper is viewing angle enhanced. The system entails captured elemental images to be converted to computer generated holography without being needed to be compressed. The transformed images can be viewed through the elemental lens array to generate a three dimensional integral image. From the followed methods one can view a full parallax three dimensional integral images with a large or enhanced viewing-angle without the aid of any supplementary device like a glass [25].

Baasantseren, Park, Kwon and Kim (2009) explores a new kind of Integral Imaging based display system. Traditional integral imaging based display systems involve an uncomplicated design with an elemental lens array and a two dimensional display panel. The system demonstrated by them entails an integral imaging display with larger viewing angle utilizing two elemental image masks. At the end of the experiment conducted, the results showed that when the new elemental images are displayed with the elemental masks, just a single three dimensional image appears in the middle of the suggested display system, since the masks successfully leads the elemental images to their respective elemental lenses. The position of the elemental images for the display panel were not restricted or stable. Moreover, the demonstration also showed that the proposed method is successful in creating three dimensional integrated images with wider viewing angle. It was discovered that the viewing angle increased along both horizontal and vertical directions [26].

Kwon et al.(2012) presents a new technique as a solution for the long processing times for elemental images to be generated. The system implemented OpenCL based GPU parallel processing. In conventional CPU based systems it huge amount of time is spent on processing only, which worsens as the number of elemental images increase. The proposed technique reduces the processing time for elemental images to be generated drastically. It is possible to implement the method on real time interactive integral imaging display system [27].

Hahn, Kim, Kim Lee (2008) presented an undistorted pickup method for integral imaging of both real and virtual objects. The proposed system consists of overlapping view volumes parallel to the ones in the integral imaging display which ultimately results in no deformation, something that often originated from the conflicting directions of elemental image from pickup to display. To implement the technique proposed, Fourier optics, lens array and telecentric lens were utilized [28].

The purpose of the review was to analyze and study different aspects of Integral Imaging and the recent techniques and applications based on it. Researches that focused on improving the image quality and enhancing viewing angles were discussed. New applications in the recent years that implement the integral imaging technique were also addressed. The findings of the conducted review also supports our justification to use the Integral Imaging technique for our 3D TV system.

1.3 Thesis Overview

This thesis is organized as following:

Chapter 2 introduces the basic principles and fundamental concepts of elemental image and integral image techniques. The applications and limitations of this display method is discussed.

Chapter 3 describes the workflow of the whole system and proposed methodology. It consists of a block diagram giving idea of the system. Five phases of the proposed methodology- acquisition, processing, synchronization, transmission and user end reconstruction of 3D TV content is stated elaborately.

In chapter 4, the experimental setup has been presented as well as the technologies that were used during the experiment. Later the results have been demonstrated.

Chapter 5 provides the summary, conclusions and future plans that can be implemented.

Chapter 2

Fundamental Concepts

2.1 Integral Imaging

Integral imaging is one of the most promising autostereoscopic 3D display techniques that provides True 3D perception to the viewers, with a full color full parallax moving 3D image with continuous viewpoints. This type of displays does not need any glasses or supplementary viewing aids. Among all the 3D technologies, only integral imaging, holography and volumetric displays meet the three requirements for true 3D experience. Holography is thought of one of the best candidates of future 3D displays but it comes with several constraints like huge data handling , coherent light source for recording and display, which is tough to implement with today's existing technology. On other hand, volumetric displays also has major constraints because it depends on mechanical movements of light rays. In those senses, integral imaging technique can be an ideal candidate for implementing a true 3D imaging display. As it can perform both the capture and display using an incoherent light source, it is considered a promising technique. Despite some tremendous benefits, integral imaging has been suffering for some drawbacks such as insufficient image resolution, limited viewing angle and a shallow depth perception that are basically restricted by the lens specifications (i.e., focal length, lens pitch etc.) of the lens array used. Because of these limitations, there were no remarkable commercial applications on this technology even after a century of its invention.

In last few decades, revolutionary development of sensors, materials and computer technologies, recently integral imaging has gained great attention of many researchers [29]. A lot of researchers have been working to overcome the limitations of this imaging technology[30][31].

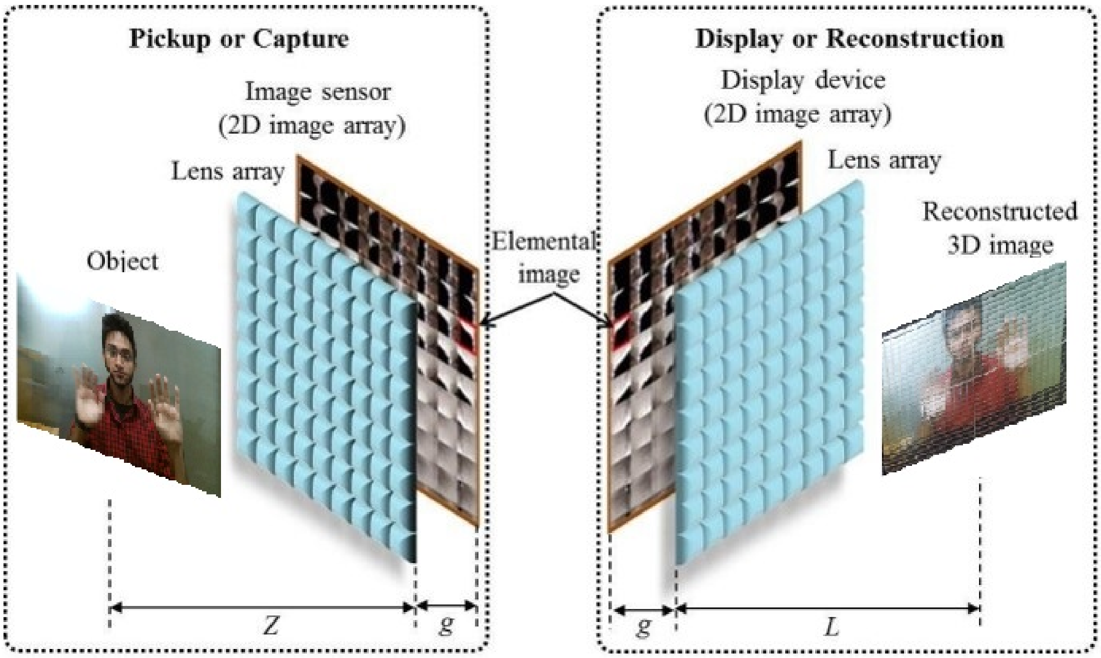

Figure 2.1: Integral Imaging.

Figure 2.1 illustrates the basic concept of integral imaging. The technique is composed of two main processes. The first is called 'pick-up' or 'capture', in which 2D image array called elemental images are captured and recorded on a 2D image sensor like photographic plate or charge coupled device (CCD) by tracing the light rays emerging from a 3D real life object. It can be done by using a micro lens array consisting of a number of convex lenses named as elemental lens. This is basically an acquisition process of the optical information of a 3D object, in which the directional and intensity information of the 3D object are spatially sampled by an array of lenses.

The second process is termed as 'display' or 'reconstruction', in which the recorded 2D image array are retraced through the same array of refractive optical elements (lens array) displaying by a high resolution display device like a LCD panel as shown in figure 2.1. In reconstruction process, a reconstructed 3D image of the same 3D object is visualized with a full color full parallax and continuous viewpoints within a certain viewing positions.

2.2 Elemental Image Generation

Elemental image generation algorithms are integral for the computerized generation of elemental images. In integral imaging based three-dimensional displays, the resolution of the elemental images(EIs) largely determines the resolution of the integrated three-dimensional image. The EIs are shown on the 2D LCD panel and the lens array is used to integrate them to form a 3D image. 3D integrated point appears at cross-section of the congregated rays that are emanated from different EIs. Each EI encompasses different directional information so that the audiences at the user end can perceive different perspective. Till date a number of elemental image generation algorithms exist including the direct pickup method of integral imaging 3D display system. In this technique, the object information are perceived through an lens array named elemental lenses and each lens generates an elemental image by following the ray emitted from the object concerned lenslet of the lens array. Finally, an array of elemental images can be precisely recorded by a CCD (charge-coupled device) camera. This method causes noise with the captured elemental image.

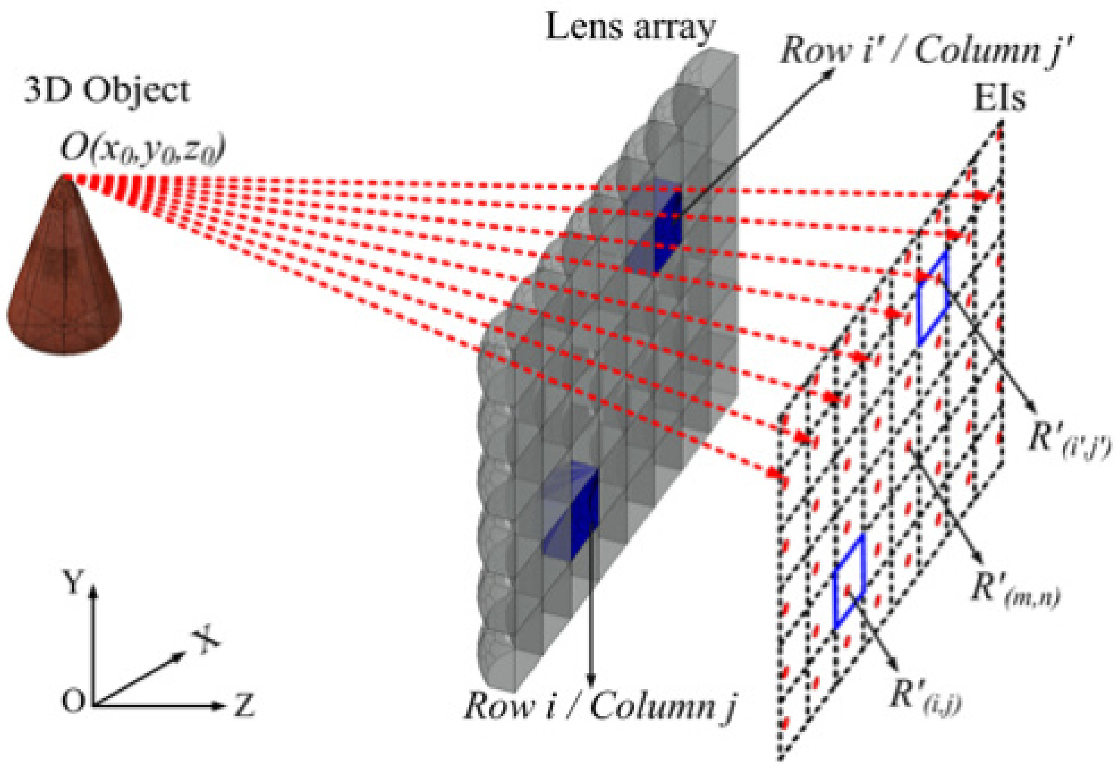

Figure 2.2: Elemental Image Formation[32].

In order to reduce the noises that were caused in direct pickup method, computer generated elemental imaging (CGEI) method was introduced. Several algorithms based on this method are present. CGEI method can only generate elemental images from synthesized 3D object points from a 2D image which are not real depths of a 3D object. An imaginary object is considered that contains all the 3D information. If it is defined that the adjacent plane which is parallel to the display panel as the x-y plane, the information relative to depth is represented by z direction. So, an imaginary object possesses information concerning a 3D location (x,y,z). This information can be used to form a 2D elemental image array assuming observation from different viewing directions through different elemental lenses in a lens array. Therefore, a point in 3D space is mapped to an array of 2D dots. It refers the process as 3D to 2D mapping. This mapping is performed for points that are used to build up a 3D object. In CGEI, computer is used to generate elemental images of imaginary objects rather than using the direct pickup process. Use of only one lens array and conventional display devices make it compact and cost effective but it was unable to map 3D objects of real world objects.

Next, elemental image generation method using real depth and color information of a real object was invented which can overcome the limitation of CGEI method. In the integral imaging pickup process, the elemental images are captured by a charge-coupled device (CCD) camera where a lens array is used so that each elemental lens forms a corresponding image of the object. The conventional pickup method has several disadvantages due to using lens array. It requires to place the lens array and a CCD camera in the exact position which is difficult most of the time. Along with that it allows unnecessary beams to pass through the lens array which hampers the quality of generated elemental images. Again, decomposition of the lenses defense for the post processing of generated elemental images. Further, it is really complex for the camera lens and lens array to generate elemental images from large objects. Therefore the new method was proposed which is efficient to generate elemental images from a real object without the need of lens array. Since it does not require any kind of lens array like conventional pickup process, it has eventually simplified the optical setup of the pickup process. A depth camera is required for this method to perceive the distance of the real object from the camera, namely depth perception.

In our proposed system, EI generation using depth and color information technique was used. The work flow of EI generation algorithm is shown in figure 2.3.

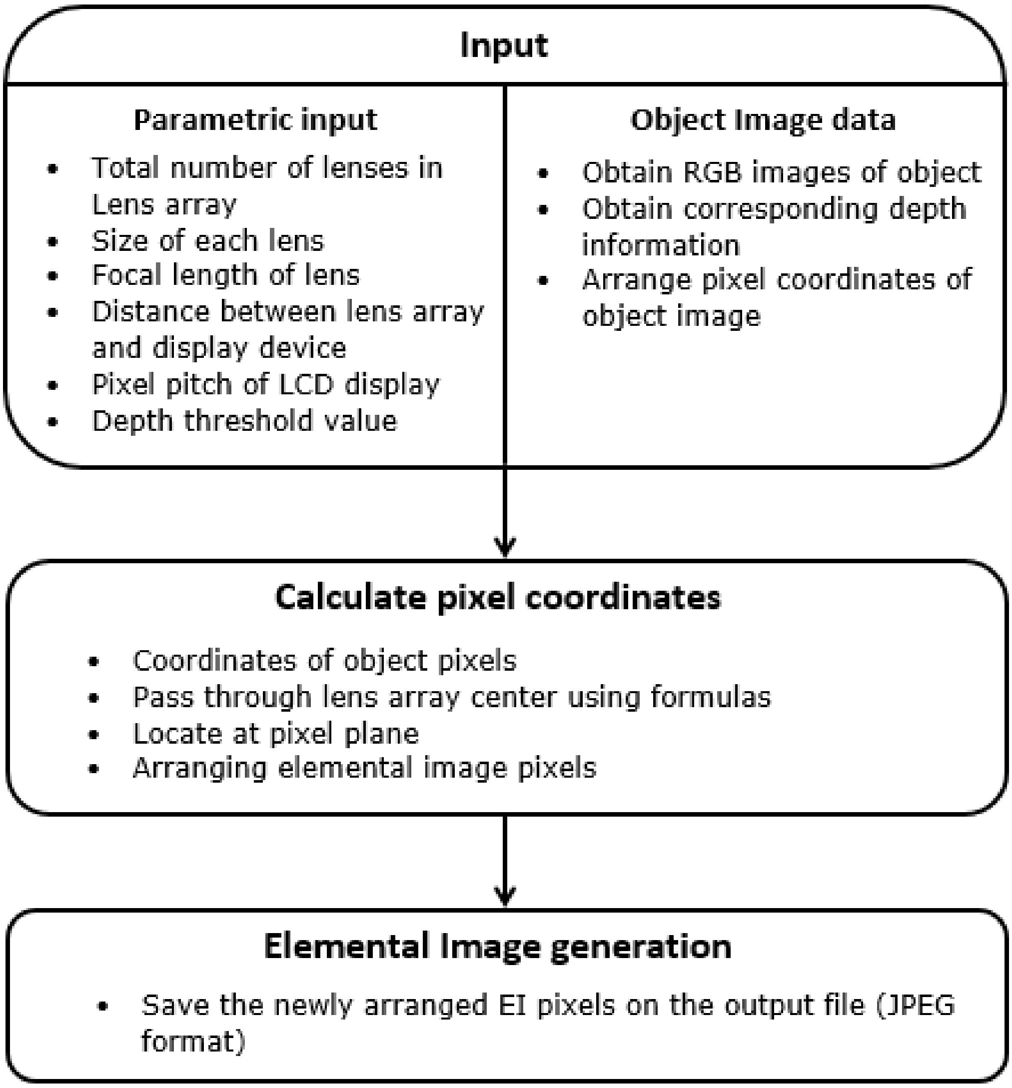

Figure 2.3: Elemental Image generation work flow.

Chapter 3

Proposed Method

A unique system is proposed here to synchronize elemental images and audio signal and the transmission technique for a glass-free 3D TV system based on integral imaging. The main idea behind the method is to generate 3D video based on elemental images synchronized with audio stream.

The process is divided into 5 phases- acquisition, processing, synchronization, transmission and 3D reconstruction at the receiving end. The acquisition part of the process focuses on capturing depth, color and audio data for the experiment. In the next phase the visual data are converted into elemental images. The elemental images are then synchronized with audio data and formed into a video in the synchronization phase. Further, the elemental image video is transferred in the server and the user receives the video using the web app. In the last phase the elemental image video is reconstructed in 3D using lens array in the receiving end. Figure 3.1 illustrates the work flow of whole system.

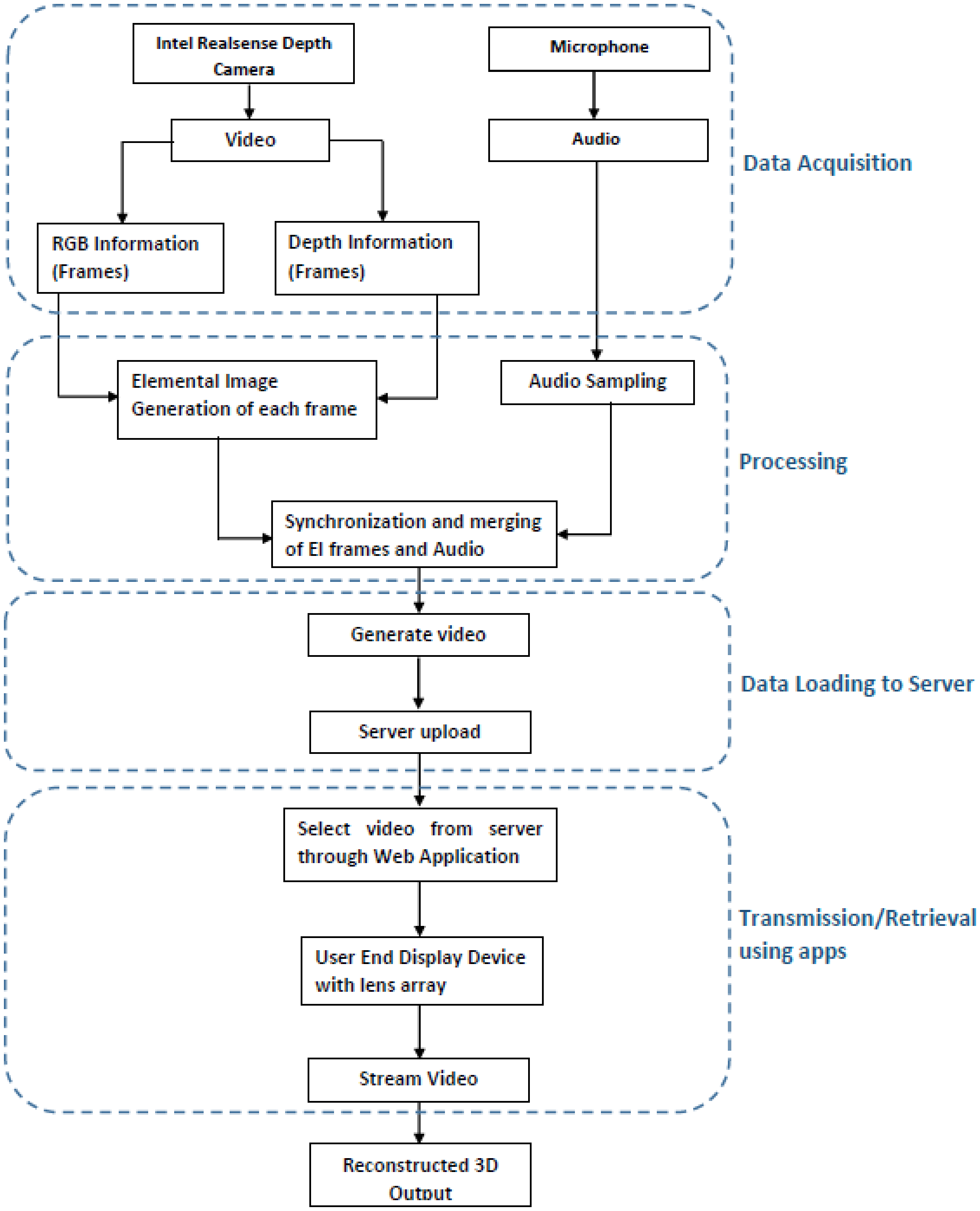

Figure 3.1: Block diagram representing the entire system.

3.1 Acquisition

The acquisition phase for the proposed system requires three types of input data- Depth image, RGB color image and audio sample. In this part all the inputs are collected using different devices to start the experiment.

The depth of a real world object can be obtained from a depth image .The gap between the of object and viewpoint is illustrated by the intensity value. It can be color coded to visually represent the close and far objects efficiently. Usually gray scale is used to represent distance of an object in depth image. Grayscale is collection of monochromic shades that range from white to black. The grayscale intensity is denoted by 8-bit integer from 0 to 255 where 0 is pure black and 255 is pure white. In a depth image if an object is closer to the camera then the object will appear brighter and objects with more distance will appear darker. Devices like stereo camera are used to acquire the depth map which can be implemented in various 3D vision algorithms.

Images with color can be represented with 24-bit RGB images. The colors in RGB images (24-bit with 8-bits for each of the red, green and blue channels) are used to show multi-channel images as red, green and blue combined in various ratio can create any color in visible spectrum. The colors are used to depict authentic colors. A RGB-D image is an amalgamation of a color image with its respective depth image. One can use Kinect sensor or the more advanced Intel RealSense cameras to acquire these information.

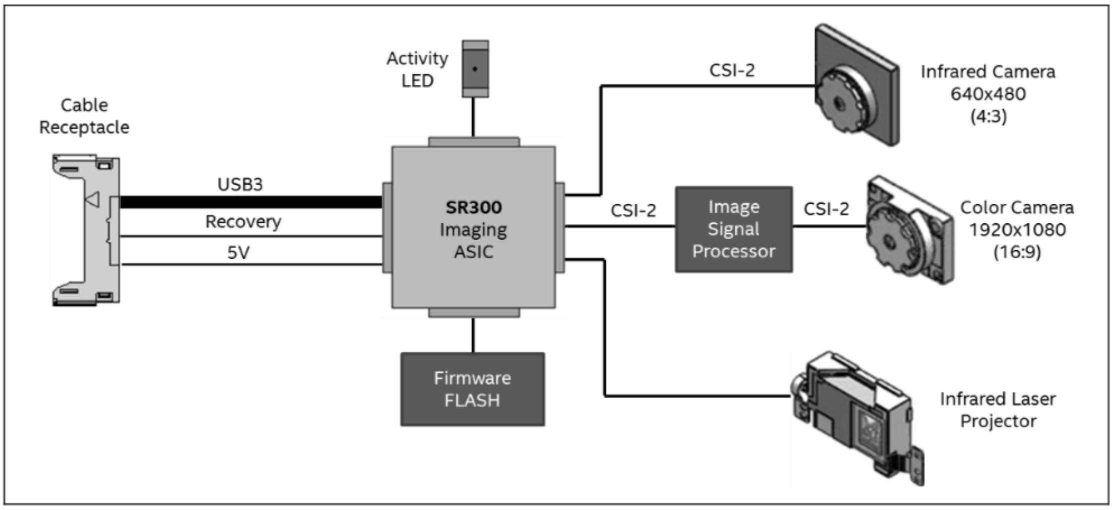

Figure 3.2: Block diagram representing intel realsense sr300[33].

3.1.1 Depth and RGB Data Capture

In the experiment, the approximate depth and RGB information of per frame of the video is gathered using Intel RealSense camera module. The depth frames are generated with the help of IR projector and IR camera placed in Intel Realsense

SR300 model. First, the IR projector projects a set of increasing spatial frequency coded IR vertical bar patterns towards the scene. These patterns are then warped by the scene and reflected back which is then captured by the IR camera. After, the pixel values captured by IR camera are handled by the imaging ASIC to create the depth frame. For the acquisition of RGB information the color camera of Intel RealSense is used. The chromatic sensor inside the color camera captures chromatic pixel values and the image signal processor processes them. With the chromatic pixel value information the color frames are generated which are then passed to imaging ASIC to form the RGB data.

3.1.2 Audio Signal Record

The audio signal is received through a separate microphone in this experiment. The sound waves generated by the scene carries energy to microphone which moves the diaphragm inside it back and forth. The coil attached to the diaphragm also moves as a result the permanent magnet inside the microphone generated a magnetic field cutting through the coil. The coil moving through the magnetic field generates electric current which is then stored as audio signal. For the sake of synchronization of audio and video the Intel RealSense camera and the microphone are turned on precisely at the same time. The depth value was calculated through Infrared projection and grey-scale depth camera.

Table 3.1: Specification of captured data.

Key Components	**Specifications**
Color Data	RGB24 1920x1080
Depth Data	640x480
Frame Rate	30
Image Format	BMP
Audio Format	WAV

Figure 3.3 shows the acquisition of grayscale image (a), color image (b) and audio signal (c) frame by frame over time T, where T represents the total recording period.

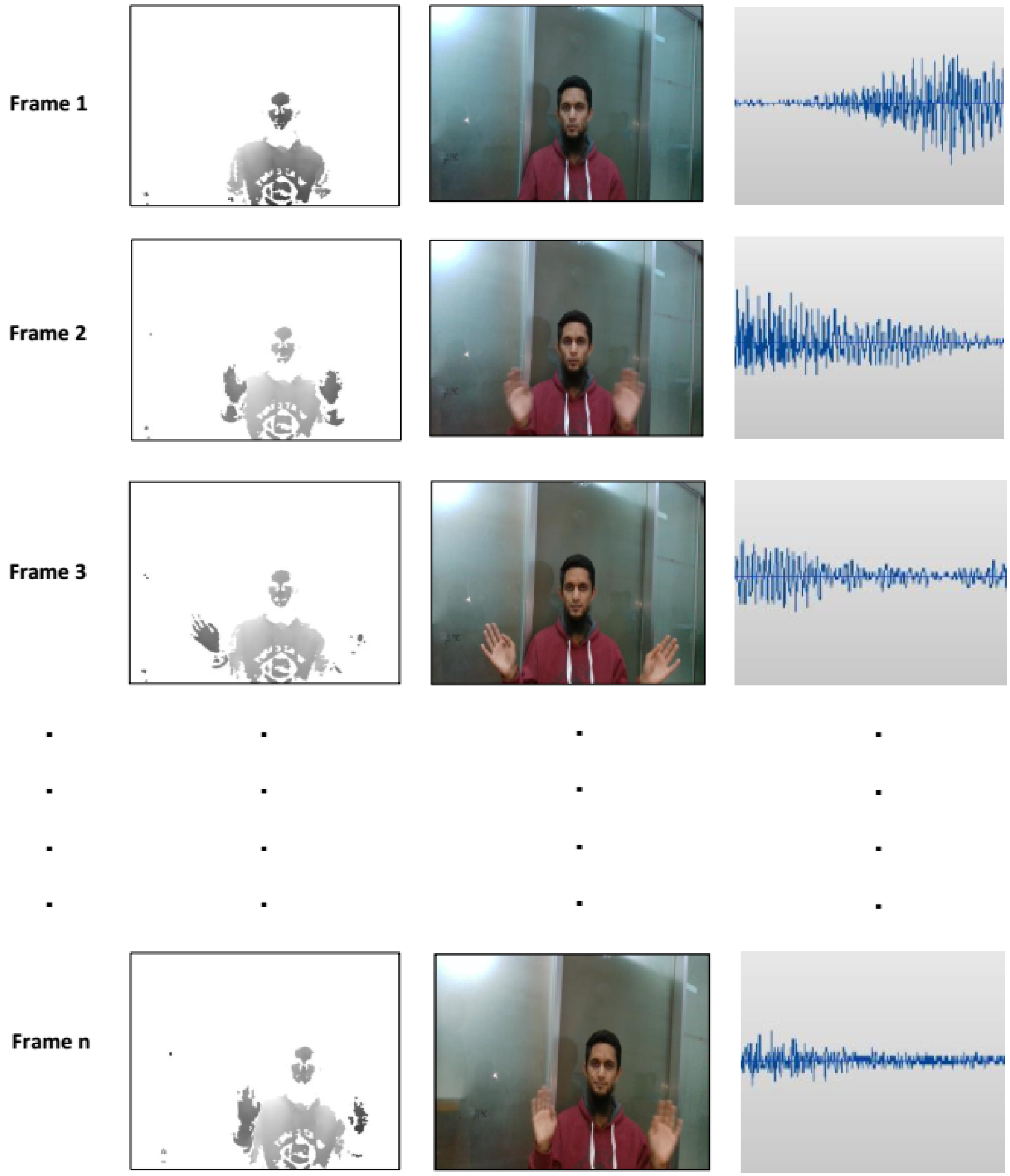

Figure 3.3: (a) Depth stream, (b) RGB stream and (c) Audio signal recording.

3.2 Processing

Initially the video captured from acquisition phase is extracted into frames. The Realsense SDK clip editor tool analyzes, visualizes and edits SDK recorded files. The tool has been used to decode video file and convert it to bitmap image frames. This created two separate streams of images- RGB color and grayscale. The audio signal is sampled using FFmpeg, a command-line-based tool for processing of video and audio files. FFmpeg includes libavcodec, an audio/video codec library used by many commercial and free software products and supports wide variety of audio formats. The audio obtained from the microphone is sampled as per the total number of image frames. As the video is captured at 30fps, each frame consumes about 1/30 namely, 33.3333ms. Therefore, the audio is sampled with the period of 33.3333ms. Color, depth and segmented audio information are then stored in data buffers with same indexing to keep track of their sequence. It can be mapped as a matrix (1) where C, D and A denotes respectively color images, depth information and audio segments starting from index 1 to n.

$$\begin{bmatrix} C_1 & D_1 & A_1 \\ C_2 & D_2 & A_2 \\ C_3 & D_3 & A_3 \\ \vdots & \vdots & \vdots \\ C_n & D_n & A_n \end{bmatrix} \tag{1}$$

The streams of RGB-D images are then processed through elemental image (E.I) generation algorithm proposed by Alam, Kwon, Piao, Kim and Kim [20]. During the initial stage the parameters of the lens and display components are recorded which are later used in the pixel mapping algorithm [34] to generate E.I. The obtained depth data was converted as stated in the algorithm [34].

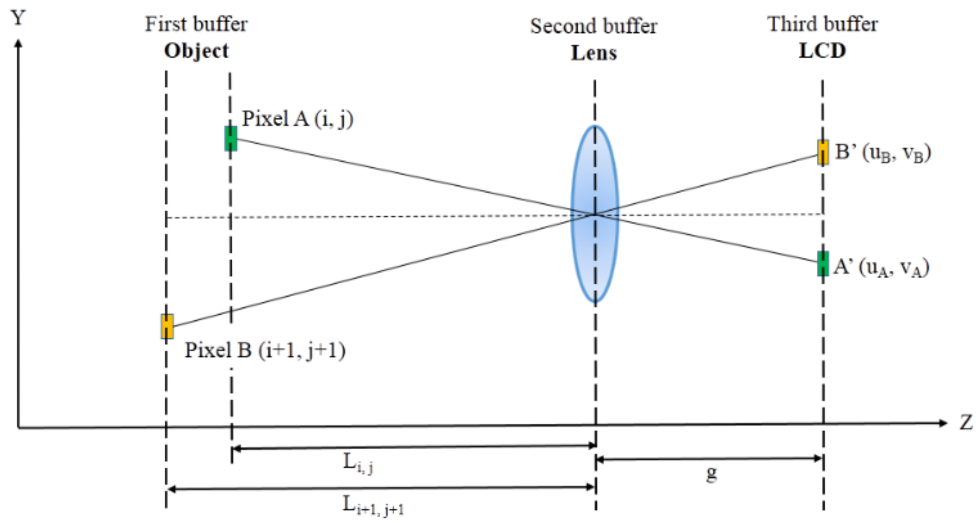

Figure 3.4: Illustration of pixels mapping from object image plane to elemental image plane for rendering elemental image pixels.

The depth data covers the distance between the surface points of the object and depth camera. Usually, the real depth data that is obtained by the camera goes beyond the depth range of the integral imaging technique. The depth range depends on the pixel size P_I and the elemental lens pitch P_L. The depth range [34] is expressed as

$$\Delta d = \frac{2d \times P_I}{P_L} \tag{2}$$

The central depth (d) [34] of the system is expressed by

$$d = \frac{f \times g}{f + g} \tag{3}$$

Where f is the focal length of the lens and g is the gap between display and lens array. The converted distance L (i, j) [34] between lens and object is expressed as

$$L_{(i,j)} = \frac{d \times (max(z) + min(z))}{2 \times Z_{(i,j)}} \tag{4}$$

Where Z (i, j) is the real depth of the pixel (i, j). The L (i, j) is later used to calculate the pixel coordinates. The pixel coordinates or (u, v) [34] is calculated by

$$u = P_L \times i_L - \frac{g \times (i \times P_I - P_L \times i_L)}{L_{(i,j)}} \tag{5}$$

$$v = P_L \times j_L - \frac{g \times (j \times P_I - P_L \times j_L)}{L_{(i,j)}} \quad (6)$$

Where i and j are the object pixel index in x and y axis. Similarly i_L and j_L are the index points of the lenses in x and y axis. The lens pitch and index are used to calculate the coordinates of the center of the lenses. Implementing these formulas, the E.I generation algorithm converted each extracted frames of the video into a sequence of elemental images (example shown in figure 3.6).

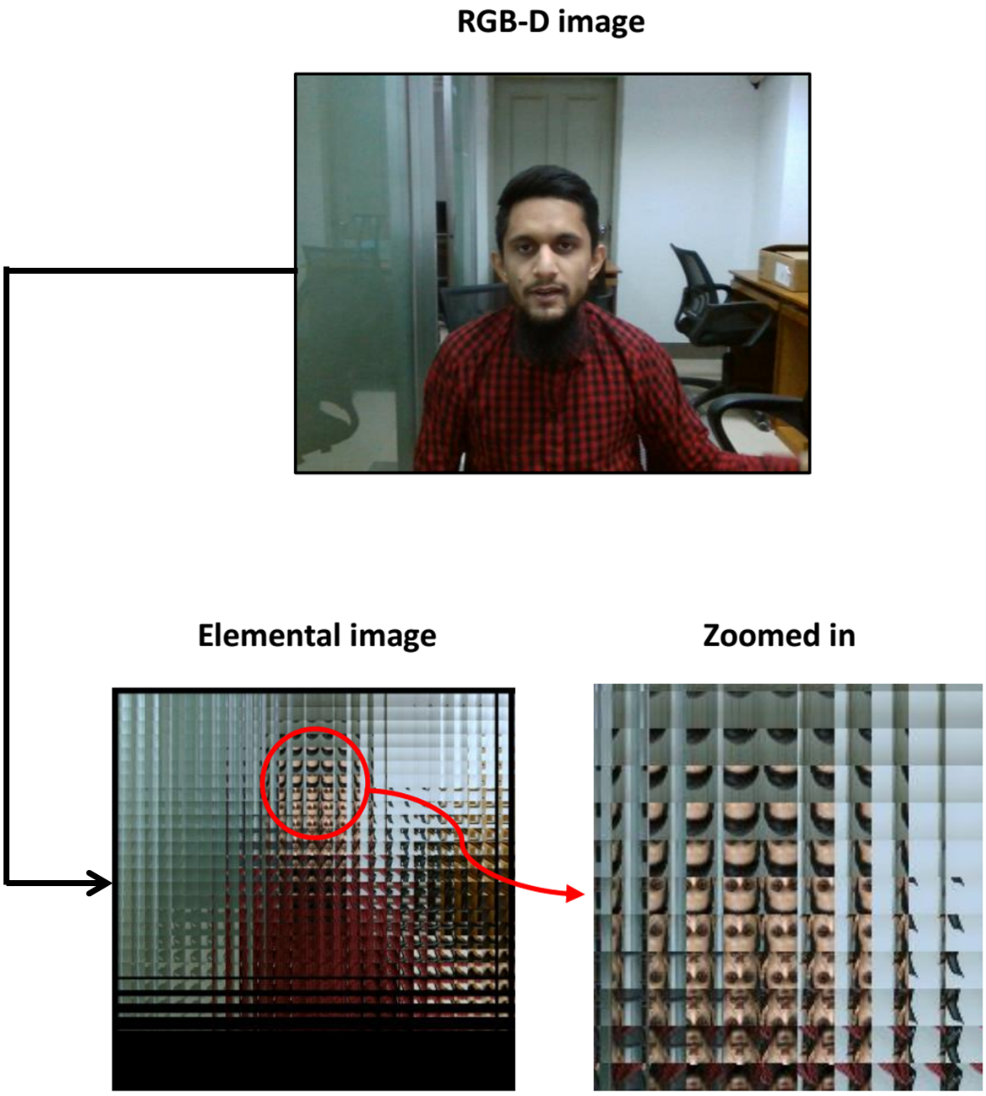

Figure 3.5: Generation of Elemental Image.

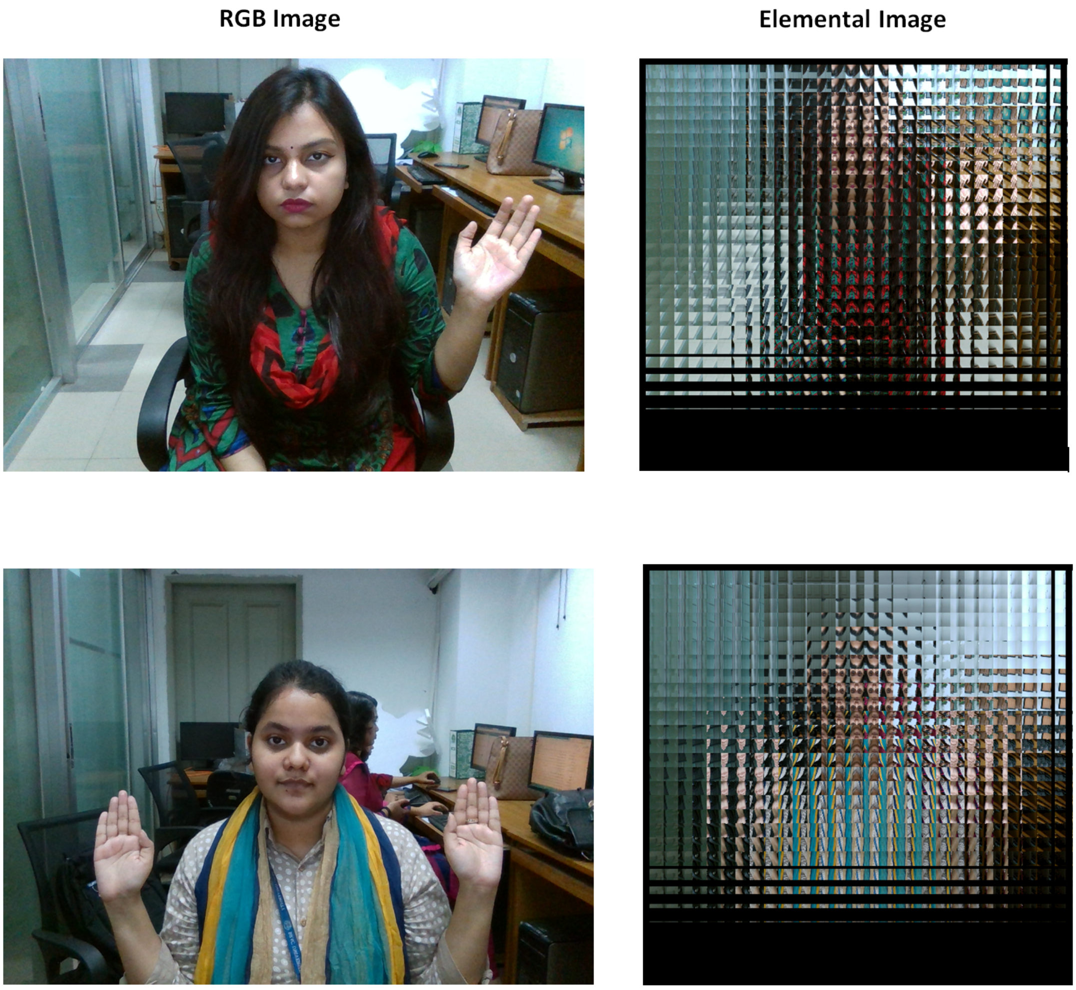

Figure 3.6: More Elemental Image samples.

3.3 Synchronization of E.I frames with Audio Samples and Video Generation

Synchronization refers to the activity of two or more things at the same time or rate. Audio-video synchronization is the relative timing of audio (sound) and video (image) parts during creation, transmission, reception and play-back. Here the synchronization means displaying image frames with the audio segment that was taken during the acquisition of that exact frame.

The sequence of elemental images previously generated and audio segments are stored in data buffers with identical indexing. It can now be mapped as

$$\begin{bmatrix} EI_1 & A_1 \\ EI_2 & A_2 \\ EI_3 & A_3 \\ \vdots & \vdots \\ EI_n & A_n \end{bmatrix} \tag{7}$$

Where EI and A defines corresponding elemental image frame and audio segment respectively. The stored E.I frames are synchronized and merged with the corresponding audio samples using FFmpeg command tool. The FFmpeg library takes the image frames and audio data in input buffer, then combines them sequentially according to index number and saves in an output buffer stream. A 30fps video file is generated after the merging of all EI frames with its respective audio samples.

However, synchronization error is one of our concern while processing audio and video. The lip sync error is represented by the number of time the audio deviates from the synchronized video. A positive value demonstrates that the audio precedes the video and a negative number demonstrates that the audio is delayed than the video. Audio and video synchronization can be a major concern in instances such as videoconferencing. During creation, AV sync error can occur if a microphone is situated at a far distance from the sound source, the audio will be desynchronized as the speed of sound is slower than the speed of light. The AV-sync error continues to rise with distance. While combining the video clips usually either the audio or video is delayed on purpose so that they are eventually synchronized.

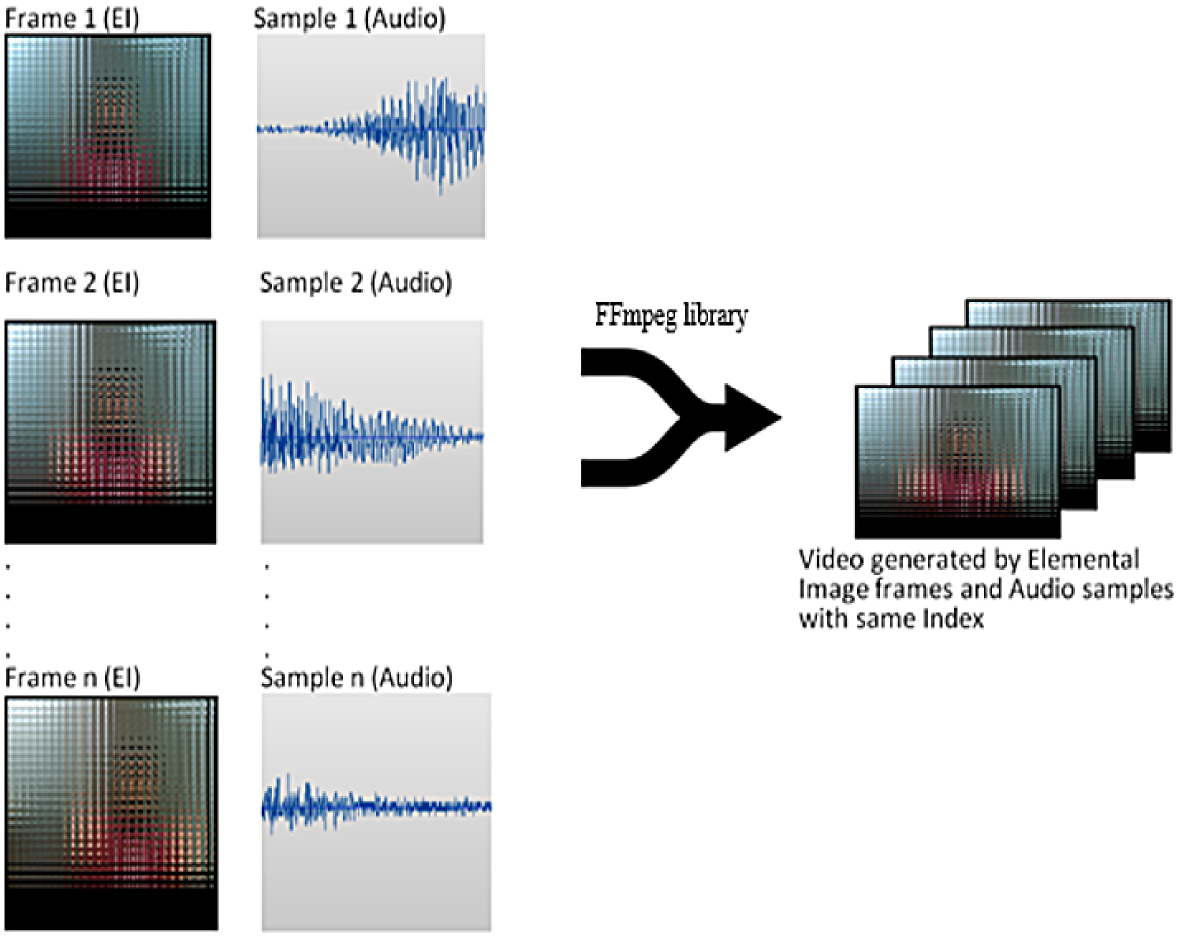

Figure 3.7: Synchronization and merging of EI frames corresponding to audio samples.

3.4 Transmission

3.4.1 Server Upload

Most digital video is designed for two things: storage and playback. For storage, the generated video file is uploaded to the 3D TV server using Transmission Control Protocol (TCP). The TCP/IP is a common protocol which provides a data transmitting mechanism to manage packet flows between devices connected to the Internet. A packet means a collection of data or the video signal in this case. The video signal is segmented into multiple IP packets so that it is ready to be sent over the IP network. Firstly, the communication begins by initiating a (TCP/IP) connection from the client to the server. This is called the handshake. When a connection has been established, both parties can send and receive data to/from each other. To upload data to a server, the client typically sends a HTTP POST request which contains the data to be uploaded. The server knows how to handle such a request and stores the data.

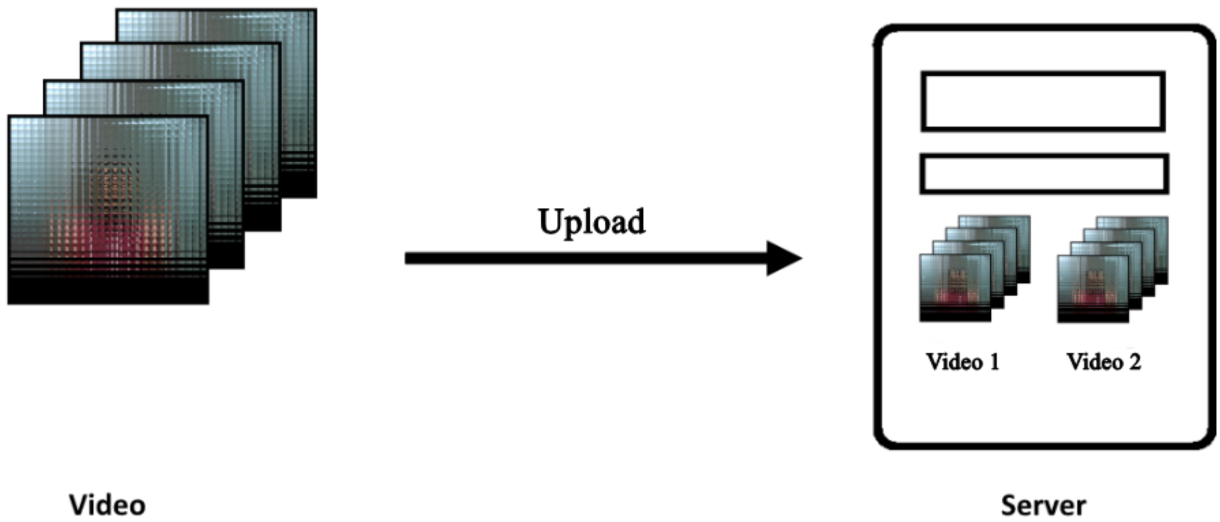

Figure 3.8: Transmission of TV content to server.

The total number of pixels per frame is given by-

$$P_N = (N_H \times N_V) \times N_{EI} \tag{8}$$

where, N_{EI} represents the number of generated elemental images.
The number of pixels in each image in the elemental image frame horizontally or vertically is expressed by-

$$N_H = N_V = \frac{P_L}{P_{XD}} \tag{9}$$

where, P_L represents the lens pitch and P_{XD} represents the pixel size of display device.
The RGB bit rate is expressed by-

$$B_{RGB} = (N_H \times N_V) \times N_{EI} \times 24 \times 30 \tag{10}$$

The audio bit rate is expressed by-

$$B_{Audio} = \frac{Audio\,Size}{Audio\,Duration} \tag{11}$$

Finally, the total transmission rate or bit rate is expressed by-

$$Total\ Bit\ Rate = B_{RGB} + B_{Audio} \tag{12}$$

Table 3.2: Calculated value for Transmission.

Parameters	Result
Total number of Pixel/ Frame, P_N	360000
Bit Rate of Video	8.64 MB/s
Bit Rate of Audio	0.32 MB/s
Total Bit Rate	8.96 MB/s

3.4.2 User End Streaming

On the user end, the TV content can be fetched from server and streamed online at any time. Video streaming consists of fragmenting the video and then sending the fragments sequentially. A video streaming protocol is a systemized delivery technique for fragmenting the video, sending it sequentially to the viewer, and recollect it for playing. On the proposed system, HTTP Live Streaming (HLS) protocol [35] has been used. This media streaming protocol is used for transmitting visual and audio media to audiences on the world wide web. This protocol disintegrates the video content into its fragments. HTTP then delivers the fragments to audiences. HLS streams are produced spontaneously which are then stored into an HTTP server. The protocol fragments the video files into short fragments with the .ts file extension (representing MPEG2 Transport Stream). The HTTP server also produces a .M3U8 playlist file. The playlist file serves as a bank that points towards additional index files for each of the existing quality options.

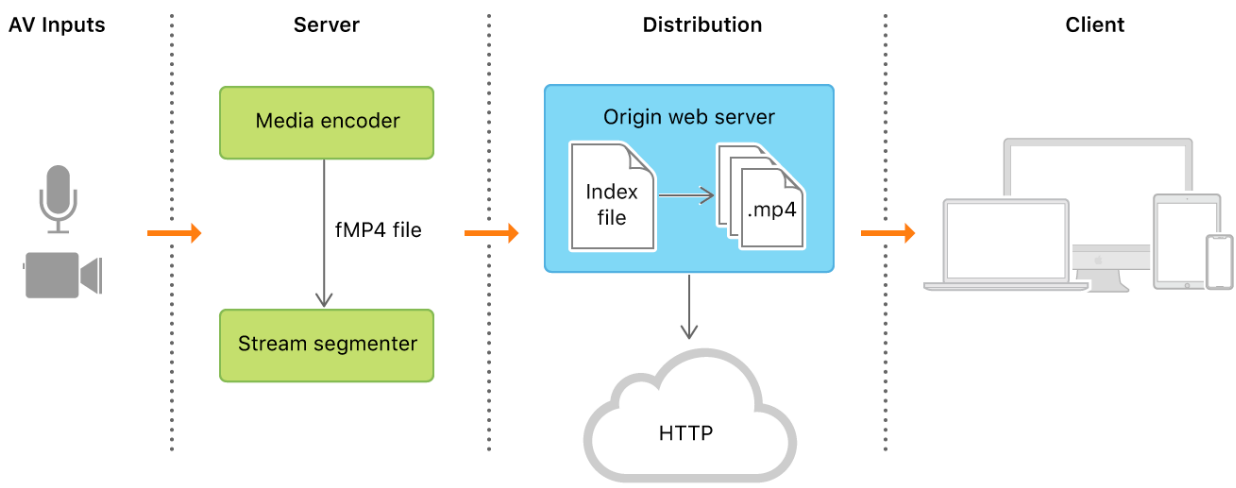

Figure 3.9: HLS Transmission.

A particular user's video player software is capable of identifying degrading or improving network conditions. If either occur, the player software reads the main index playlist, discerns the ideal quality video, and then reads the quality-specific index file to determine which video fragment match up to where the audience is viewing.

Advantages

1. HTTP live streaming is a popular streaming protocol that is compatible on desktop browsers, smart TVs, and smart phones. HTML5 video players also natively support HLS. This allows the stream to reach as many viewers as possible.

2. HLS implements a technique named Adaptive Bitrate Delivery. This technique is used to calculate the internet speed available to each viewer of a particular video. Following that, the video quality delivered is regulated accordingly. Video delivered using HLS, if formatted correctly, will dynamically deliver unparalleled video quality while reducing buffering and delay. This application refines the user experience. The video will try to resume playing to the extent feasible. The video quality will enhance given that the internet speed is fast enough to manage the load.

3. Another added benefit of HLS is that it is cheap. As mentioned before, the format is compatible on about every device via HTML5 and Media Source Extensions. There is no additional requirement of a particular device to watch the content.

4. Security is also one of the advantages. HLS is known for offering a better secure browser for the users.

3.5 Reconstruction of 3D TV Content at User End Application

3.5.1 Video Playback via Web Application

A web application was built to make the system available and accessible on worldwide web. The app was created from scratch using HTML and CSS to build the structure and design the user interface. Furthermore, Bootstrap was used to make the app responsible so that it is easily viewable on any device. Moreover, JavaScript is used to make the website fully functional. Figure 7 shows the interface of end user.

The application consists of two parts. On the left side there are thumbnails of each video and on the right side one particular video is displayed by default. User can watch any of the videos by clicking on the particular thumbnails provided for each video. All the videos are pre-uploaded to the server. Domain and hosting service was bought to make the app active on the internet. There is one more page on the website where the project documentation and features are given. The app has a user friendly interface. The selected video is fetched from the server using HTTP and HLS protocol. When the video is streamed to a viewer, it is imperative that the bits are delivered quickly so as to prevent a video from unwanted delays or buffering. The data being sent is also time-sensitive as slow data streams result in poor viewing experience.

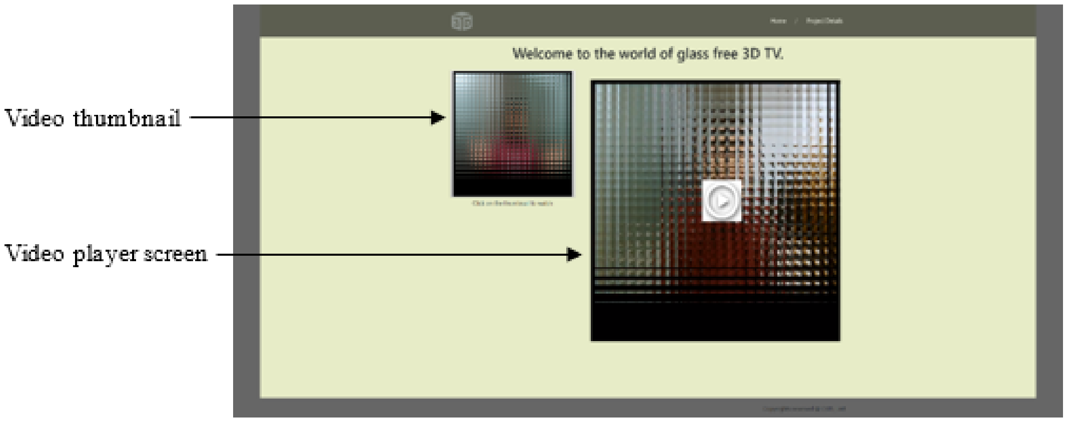

Figure 3.10: Snapshot of the user end web application.

Web applications typically employs an integration of server-side scripts (for instance PHP) to manage the storage and obtaining of the data, and client-side scripts (for instance javascript) to display information to end-users[36].

Through implementation of technologies like Java, application-specific techniques can be conducted successfully. Multiple services have operated to integrate these techniques into an amicable interface that embraces the aspects of an operating system. Simple methods are compatible on these technologies. Web developers typically utilize client-side scripting to add functionality, mostly to produce an interactive experience that eliminates the need of realoading the page. Technologies have been lately evolved to synchronize client-side scripting with server-side technologies including the likes of ASP.NET, J2EE and many more.

A web server is a prerequisite to a web application needed to handle client requests. Additionally an application server is also required to execute the asssigned jobs. A database may also be required to house the information. Application server technology includes the likes of ASP, ColdFusion, PHP and many more.

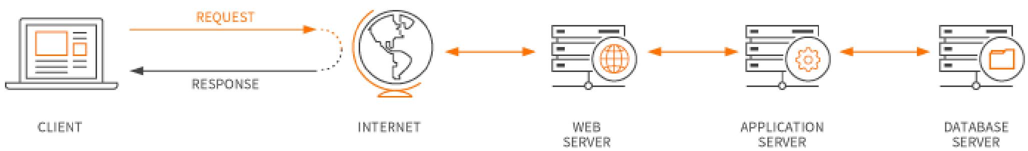

Figure 3.11: Work flow of web application [36].

Work flow of Web Application

1. The work flow begins with the User generating a request to the web server on the world wide web via a web browser or the user interface of application.
2. Upon receiving the request the web server passes it to the suitable web application server.
3. Web application server executes the requested task and produces the appropriate result.
4. Following that the web application server directs the results and the requested information to the web server.
5. In response, the web server displays the requested information on the user's display.

Benefits of using Web Based Application

1. Web applications is supported on numerous platforms.
2. No issue of compatibility or suitability.
3. No issue of space restriction.
4. Software theft issue is minimized in subscription-based web applications.
5. It is cost effective for both the business and clients.

3.5.2 Reconstruction of 3D Content using Integral Imaging

The video is reconstructed at the display monitor through Integral Imaging. As mentioned earlier, integral imaging consists of two parts- pickup and display. During displaying, the obtained elemental images on the user end monitor, light from the display device returns through the lens array and projects the generated elemental images onto the focal plane of the lens array[37].

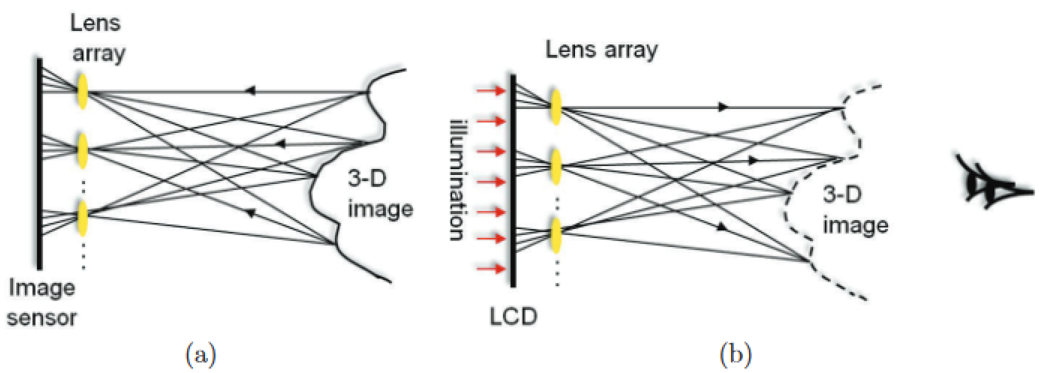

Figure 3.12: Integral imaging. (a) Pickup, (b) Display [37].

Integral imaging is basically a three dimensional imaging technique that is auto stereoscopic and manipulates light field using two dimensional array of micro lenses. With the help of integral imaging technique 3-D visuals can be provided without any need of supplementary glasses or external devices. As discussed, the transferred 3D content is a synchronized video of Elemental Image frames. Each EI frame illuminates multi directional sequential projection which is then collected by the array of lens. This causes many PLS (Point Light Sources) at various parts of the focal plane. Also, the positions of the point light sources can be manipulated by projection angle. The result of integral imaging is a complete 3-D visual with full color, full parallax in all directions and significant depth cues. Moreover, the resulting visual changes perspective depending on the distance and location of the observer. The quality of the result majorly depends on the size of each lens and resolution of the video. The resulting video gets better if the size of lens is small enough.

The lens array in this experiment consists of 30 by 30 number of lens which is in accordance to the number of elemental image segments. It is placed with a distance of 15mm from the display and in a way so that it is perfectly aligned with the video. The overlap between all the projected elemental images coincides in the 3-D space to form a real 3 dimensional image, which appears reversed since the user's perspective is opposite to that of the lens array. The depth inverted images are transformed to correct depth by rotating each image by 180 along the center. The reconstructed images differ according to perspective of the user. The resolution of each reconstructed image depends on the resolution of the camera used (Intel Realsense sr300 in this case) and the number of generated Elemental Images [38].

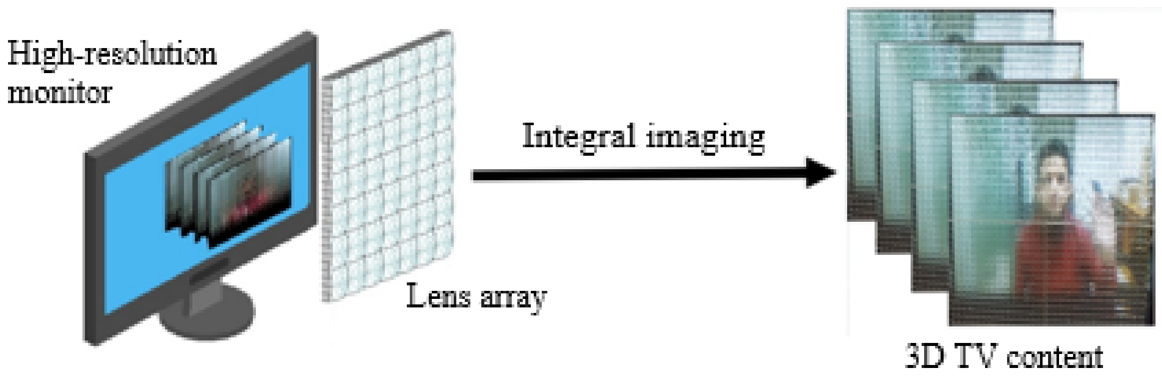

Figure 3.13: 3D reconstruction using integral imaging.

Chapter 4

Experimental Setup And Results

4.1 Experimental Setup

Our experimental setup consisted of Intel RealSense camera and microphone for the acquisition part of the experiment. Also, display monitor attached with lens array was used for user end reconstruction of the 3D TV content. The specifications of the equipment used in the experiment are mentioned below.

Table 4.1: Specification of Display Setup.

Key Components	**Specifications**	**Characteristics**
Lens array	Pitch of lens	5mm
	Focal Length	10mm
	Number of lenses	30x30

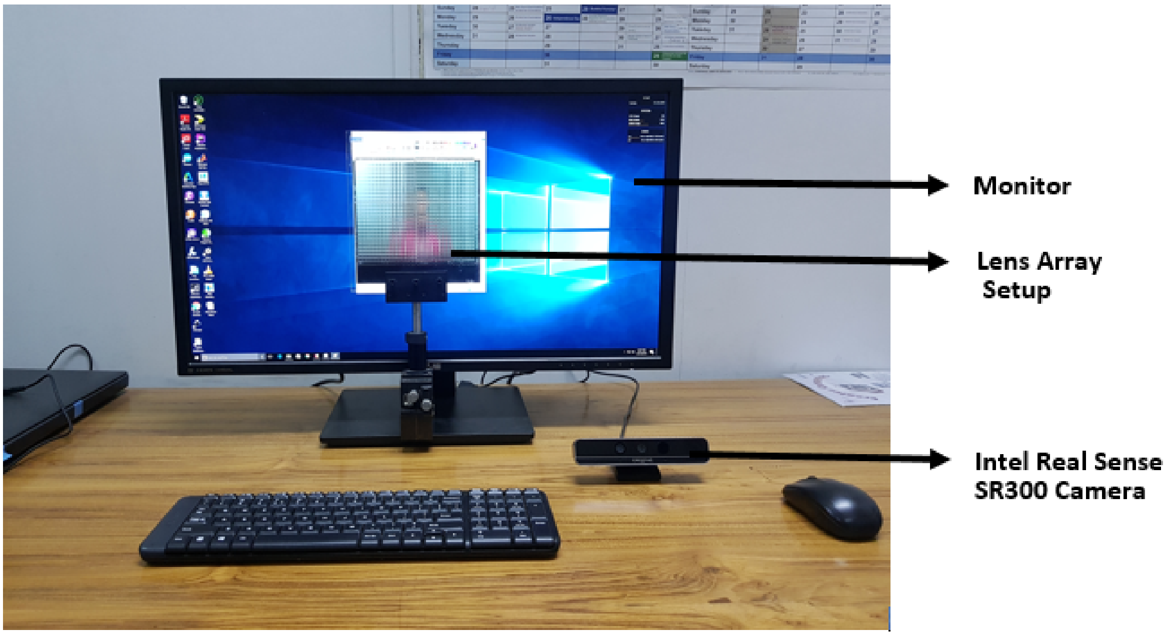

Figure 4.1: Full experimental setup.

Table 4.2: Specification of technical setup.

Key Components	**Specifications**	**Characteristics**
Intel RealSense	Model	SR300
	Color Resolution	1920x1080 pixels
	Depth Resolution	640x480 pixels
Microphone	Model	Samsung
Display Monitor	Model	ASUS LED monitor

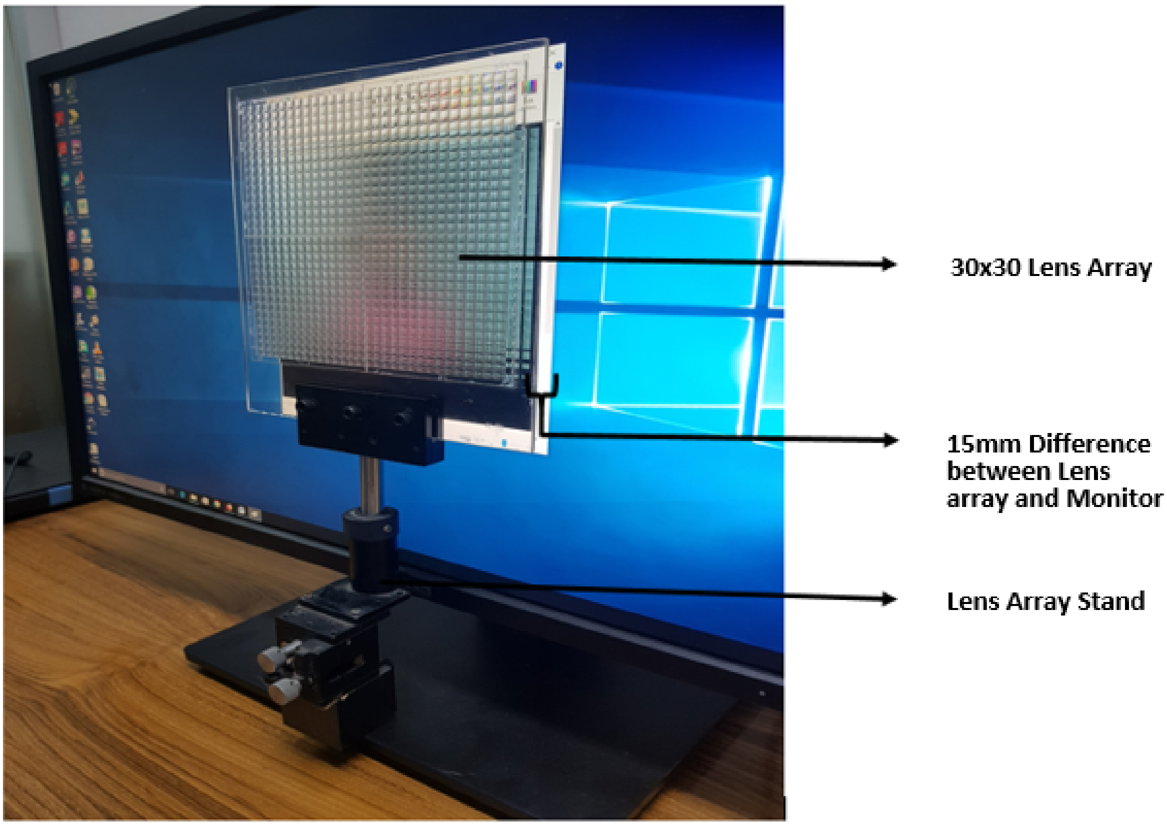

Figure 4.2: Lens array setup.

4.1.1 Technologies Used

MATLAB

MATLAB is a widely used programming language. For image processing, MATLAB function files are generally written, or script files to perform the operations. These files structure a formal record of the processing used and guarantees that the final outcomes can be tested and imitated by others should the need emerge.

The advantage of using MATLAB is that it offers numerous capabilities and functions for image processing or in general handling tasks of signal processing. Most of these functions are written in the MATLAB language as well as are publicly readable as plain text files. In this manner the usage details of these functions are accessible and open to scrutiny.

Another advantage of MATLAB is that it ensures utmost numerical precision in the result. Generally, image files store data to 8 bit precision. This corresponds to a range of integer values from 0-255. Three 8 bit numbers can be used to represent a pixel in a color image, each of them representing the blue, green and red components as an integer value within 0 to 255. Typically this is sufficient accuracy for representing general pictures. However it is very easy to generate values that lie outside the range 0-255 while one reads this image information into memory and starts to process it. For instance, to double the contrast of an image one multiplies the intensity values by 2. An image value of 200 will wind up 400 and numerical overflow will result. The numbers or occurrences may vary depending on the image processing programs at use. Some algorithms may trim the results to an integer ranging between 0-255, others may implement the mathematical operations in floating point arithmetic and then rescale the final outcome to an integer in the range 0-255. It is conceivable to have full oversight of the accuracy with which one represents data in MATLAB. An image can be read into memory and the data cast into double precision floating point values. All image processing steps can then be done in double precision floating point arithmetic, and at no intermediate stage does one need to rescale the results to integers in the range 0-255. Only at the last point when the image is to be displayed and/or written to file does it need to be rescaled. Here to eliminate extreme pixel values, one can use histogram truncation. So that the bulk of the image data at the end is properly represented.

MATLAB has a built in graphics framework. It incorporates high-level commands for various image processing related tasks, two-dimensional and three-dimensional data visualization and animation. It integrates a greatly optimized graphical output for interaction and can be plotted very easily. In Addition, MATLAB offers low-level commands that enable one to completely tweak the appearance of graphics along with building complete Graphical User Interfaces (GUI) on MATLAB applications.

Being a scientific programming language, MATLAB provides robust mathematical and numerical support for implementing advanced algorithms. Its mathematical function library includes a vast collection of computational algorithms ranging from elementary functions like sum, division, mean/max, sine, cosine, and complex arithmetic, to more sophisticated functions like Bessel functions, fast Fourier transforms, all kinds of matrix operations such as matrix multiplication, eigenvalues, inverse etc. For all these reasons, MATLAB has a mainstream use in the field of image processing and computer vision technology. Researchers very likely implement new algorithms first in MATLAB.

Intel RealSense SR300 Camera

The Intel RealSense SR300 depth camera implements a short range, coded light 3D imaging system. The wide range of 3D mode configurations and synchronization settings of the SR300 accredit the product to be an optimal solution in the category of 3D imaging applications. A handful of new features and significant improvements were added by Intel to the SR300 over the first-generation RealSense camera F200.

The SR300 camera has an improved depth range of 2 meters. It has enhanced middleware quality, provides dynamic motion capture with higher-quality depth data. Additionally, the new model has decreased power consumption while extending the robustness. With 1080p full HD video image quality at up to 30 frames per second (FPS), or 720p HD video image quality at up to 60 FPS, the SR300 model provides improved frame capture over the previous generation cameras. The SR300 model uses Fast VGA depth mode instead of native VGA depth mode that the F200 model uses. This new depth mode minimizes exposure time as well as allows dynamic motion up to 2m/s. This camera enables new platform usages by providing synchronized color, depth, and IR video stream data to the client system. Moreover, the RealSense SDK has added improved background segmentation, a new 3D Cursor mode and 3D object scanning for the SR300 camera. Therefore, the SR300 model of RealSense 3D camera has been used in the proposed system.

4.2 Experiment Results

The experiment was conducted on four separate videos with durations of 5s, 7s, 15s and 60s respectively. The most crucial part of the experiment was the synchronization of EI frames with the right audio samples. At the acquisition part shown in Figure 1, we made sure that the camera and microphone were turned on at the same time and also the duration of the audio and video files are exactly same so that the files could be equally segmented and stored with same indexing in different buffers. After the acquisition part, the Elemental Image Generation Algorithm was able to successfully process the depth and RGB values and the generated elemental images were stored with the same indexing as the depth and RGB information. Then, we could successfully merge and synchronize the audio samples with corresponding EI frames for all four videos with full accuracy. However, the only drawback was that the elemental image generation process is time consuming as for each frame it took approximately 22 seconds to generate elemental image.

The reconstruction of the 3D TV content at the user end determines the experimental result. The multi-array lens was carefully aligned with the video appearing on the user end web app and was kept at 15mm distance from the display monitor. The video is reconstructed based on integral imaging and the user is able to watch 3D content with synchronized audio. However, the result could have been improved if we could use lens array containing smaller pitched lens. Additionally, the 3D content can be watched at the viewing angle of approximately fifteen degrees from the user end display system. Nevertheless, we got the best result for all four videos when the viewing angle is at approximately zero degree. The result at the user end shows a smooth 3D video perfectly synchronized with audio. The reconstruction for two videos out of the four are shown.

First Experimental Video

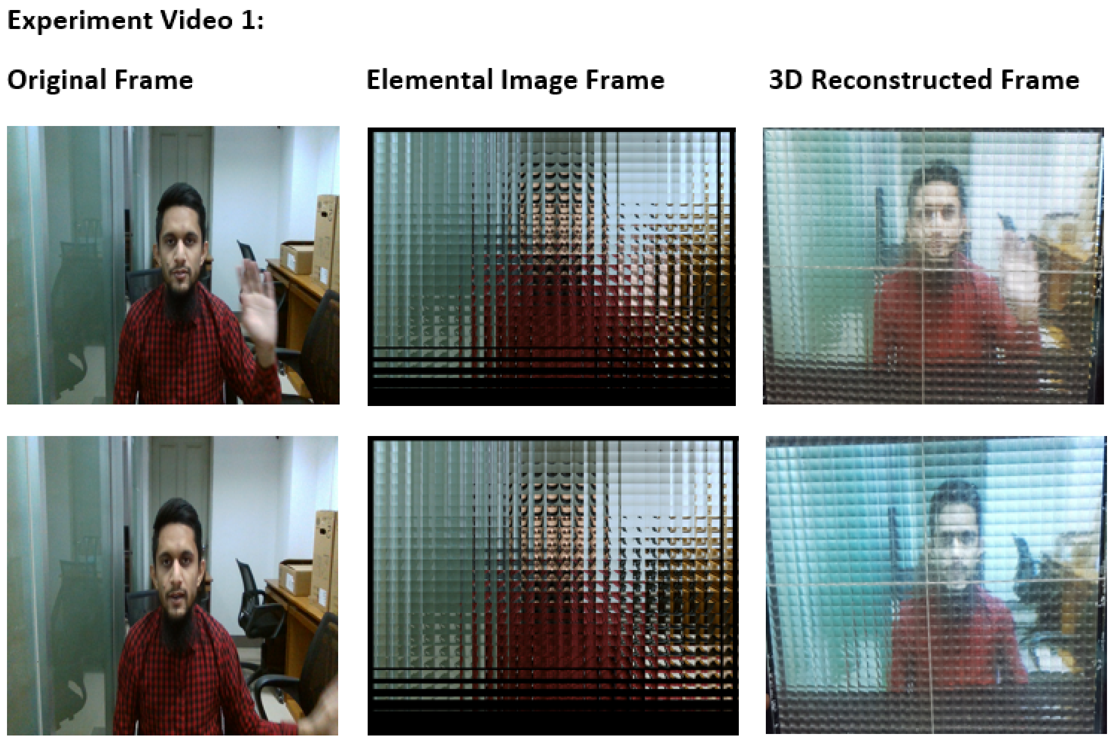

Figure 4.3: Video frames, respective elemental images generated through E.I generation algorithm, 3D reconstruction using integral imaging of 1st experimental video.

Second Experimental Video

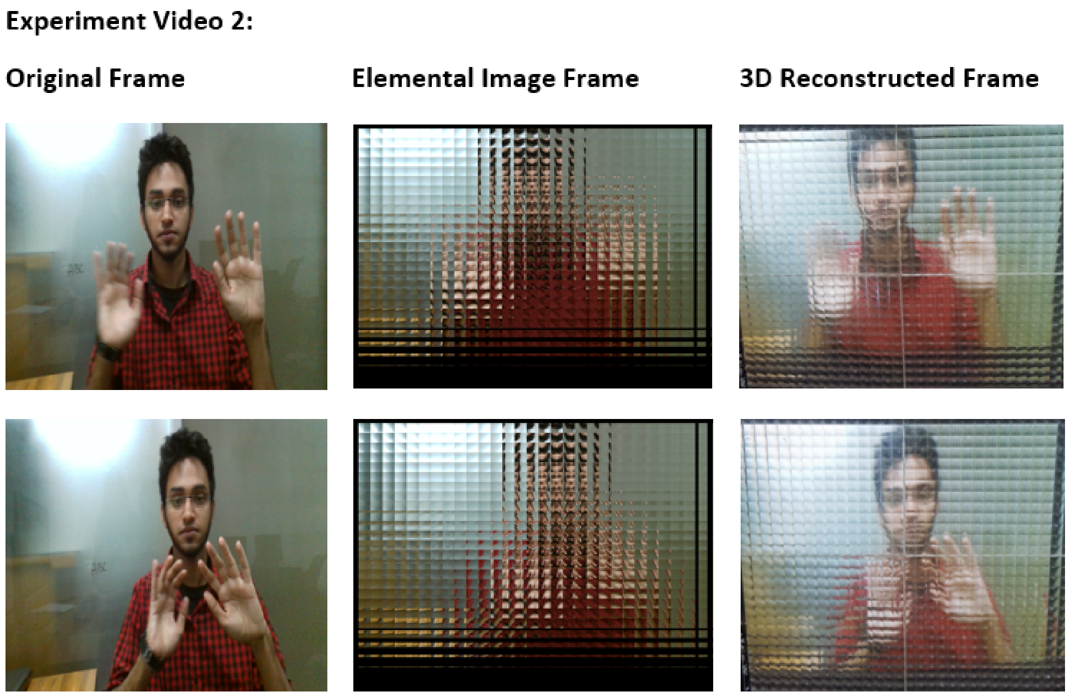

Figure 4.4: Video frames, respective elemental images generated through E.I generation algorithm, 3D reconstruction using integral imaging of 2nd experimental video.

Quality Measurement of 3D Reconstructed Image

The viewing characteristics of the integrated 3D image can be analyzed by three viewing parameters: (i) resolution, (ii) viewing angle and (iii) image depth. Pixels of the display are focused in a focal plane of the lens array. That focal plane is called a central depth plane (CDP).

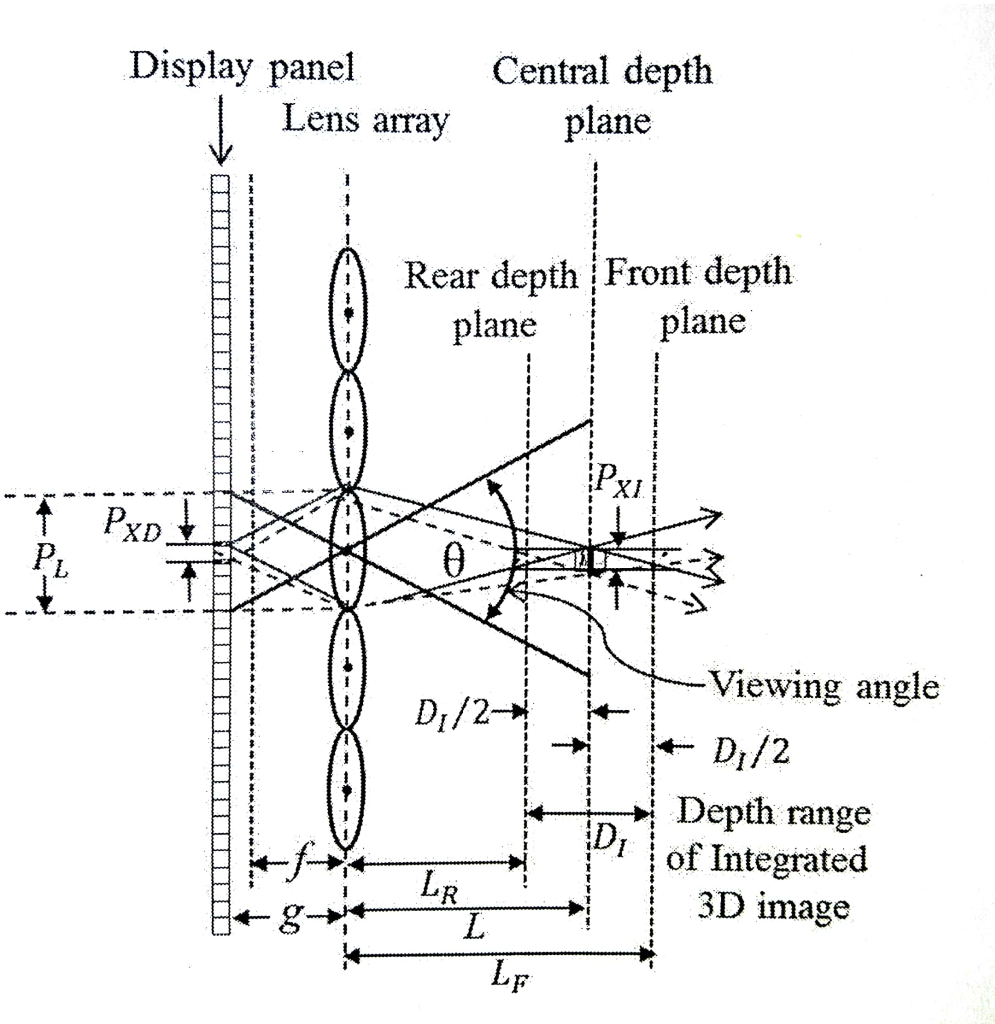

Figure 4.5: Viewing characteristics and parameters of Integral Imaging Display.

CDP is determined by the Gaussian lens equation as

$$L_{CDP} = \frac{g}{g - f} \tag{13}$$

Where f is focal length of an elemental lens and g is the gap between the lens array and display panel.

A pixel size of integrated 3D image can be determined with geometrical analysis by

$$P_{XI} = \frac{L_{CDP}}{g} \times P_{XD} \tag{14}$$

Where P_{XD} is a pixel size of 2D display device.

A depth range is equal to the difference between the rear depth plane and a front depth plane that depends on the size of the integrated 3D point. The depth range can be expressed by

$$D_I = \frac{2P_{XI}}{P_L} \times L_{CDP} \tag{15}$$

In general case a viewing angle of integral imaging display can be determined by

$$\theta = 2\arctan\frac{P_L}{2g} \tag{16}$$

Viewing Angle at -15°   Viewing Angle at 15°

Figure 4.6: Perspective in different viewing angle.

Table 4.3: Calculated value of 3D Image Quality.

Parameters	**Result**
Pixel Size of Integrated 3D Image, P_{XI}	0.5 mm
Depth Range of Integrated Image, D_I	6 mm
Viewing Angle, θ	18.9

Chapter 5

Conclusion and Future Work

5.1 Conclusion

From this study, we established a glass free 3D Internet TV system by the synchronization of Elemental image frames and audio samples converting it into a video file. The elemental image frames are generated with the depth and RGB values of the initial video. The EI video with synchronized audio is then uploaded to the 3D TV server and can be played from the user end web application. Furthermore, the user end display system consists of multi-array lens by which the video is reconstructed based on integral imaging. The user experiences 3D video with synchronized audio at the user end display.

The process involving the extraction of raw data including the RGB data, depth information and audio files were executed successfully. Utilizing the acquired data, the following steps concerned with generation of elemental images and the ultimate synchronization of audio frames with corresponding elemental images were also accomplished. Later measures taken to store the resulting synchronized videos in the central server successfully. The server allows both transmission of data into and from it. The videos were transmitted through an application to the user end device, where its reconstruction premised on the Integral Imaging technique was accomplished. However, the resolution of the videos needs to be enhanced before commercial implementation.

5.2 Future work

We have implemented the elemental image generation algorithm using MATLAB which is a CPU centered task. Since MATLAB is an interpreted language, the execution speed is a major limitation. It took approximately 22 seconds to generate each elemental image of a certain frame. In future, we will integrate the system to GPU parallel processing using NVIDIA graphics unit and their video codec SDK. This will ensure faster generation of elemental images and therefore, a more clean and efficient system.

As video and audio processing occurs in a number of steps, the lack of ability to recover means that all steps must be carefully designed to avoid drift. Additionally, even though each stage causes only minor drift, the drift can still be accumulated into an obvious one. Despite careful planning and timing calibration, errors can creep in during processing stage and measures must be taken to identify the factors that contribute to any delays caused and manage the resulting misalignment. Also, the timing of turning on the camera and microphone is extremely crucial as it determines the accuracy of synchronization of the end product. Thus, the acquisition part of the experiment has to be very precise for the best outcome.

Code efficiency is a key element to ensure a high performance and effective system. The algorithms that we have implemented in different stages like the elemental image generation or the video encoding/decoding algorithm- their efficiency and runtime execution is directly linked to code efficiency. Achieving code efficiency can be done by removing redundant segments, using optimal memory and reusable components, using best keywords, data types and variables, calling function in place of repeated code etc. We have used MATLAB to implement the algorithm and the since the execution speed is a disadvantage in future we would like to use other languages like C or C++ which would make the program faster, optimal and more efficient.

Previously, glass free 3D system could only display still images. With our algorithm, it will be able to display 3D videos. Our algorithm is developed to make the best use of latest 3D camera resources. Thus it will deliver a considerable FPS (frame per second) value and ensure pleasant watching.

Bibliography

[1] J. P. McIntire, P. R. Havig, and E. E. Geiselman, “Stereoscopic 3d displays and human performance: A comprehensive review”, *Displays*, vol. 35, no. 1, pp. 18–26, 2014.

[2] J. Matsushima, Y. Ibata, K. Masumura, T. Asai, T. Sato, and K. Shigemura, “Two-way multi-view 2-d/3-d display combining lc lens and hvxdp panel using novel pixel arrangement”, *Journal of the Society for Information Display*, vol. 24, no. 4, pp. 252–261, 2016.

[3] L. Bogaert, Y. Meuret, H. De Smet, and H. Thienpont, “Analysis of two novel concepts for multiview three-dimensional displays using one projector”, *Optical Engineering*, vol. 49, no. 12, p. 127 401, 2010.

[4] M. A. Alam, M.-L. Piao, L. T. Bang, and N. Kim, “Viewing-zone control of integral imaging display using a directional projection and elemental image resizing method”, *Applied optics*, vol. 52, no. 28, pp. 6969–6978, 2013.

[5] M. A. Alam, G. Baasantseren, M.-U. Erdenebat, N. Kim, and J.-H. Park, “Resolution enhancement of integral-imaging three-dimensional display using directional elemental image projection”, *Journal of the Society for Information Display*, vol. 20, no. 4, pp. 221–227, 2012.

[6] M. S. Farid, “Introduction to 3dtv and 3d-cinema”, in. Dec. 2015, pp. 1–26.

[7] K.-C. Kwon, J.-S. Jeong, M.-U. Erdenebat, Y.-T. Lim, K.-H. Yoo, and N. Kim, “Real-time interactive display for integral imaging microscopy”, *Applied optics*, vol. 53, no. 20, pp. 4450–4459, 2014.

[8] G. Lippmann, “La photographie integrale. comtes rendus”, *Academie des Sciences*, vol. 146, pp. 446–451, 1908.

[9] S.-H. Hong, J.-S. Jang, and B. Javidi, “Three-dimensional volumetric object reconstruction using computational integral imaging”, *Optics Express*, vol. 12, no. 3, pp. 483–491, 2004.

[10] V. J. Traver, P. Latorre-Carmona, E. Salvador-Balaguer, F. Pla, and B. Javidi, “Human gesture recognition using three-dimensional integral imaging”, *JOSA A*, vol. 31, no. 10, pp. 2312–2320, 2014.

[11] Y.-W. Song, B. Javidi, and F. Jin, “3d object scaling in integral imaging display by varying the spatial ray sampling rate”, *Optics express*, vol. 13, no. 9, pp. 3242–3251, 2005.

[12] C.-W. Chen, M. Cho, Y.-P. Huang, and B. Javidi, “Improved viewing zones for projection type integral imaging 3d display using adaptive liquid crystal prism array”, *Journal of Display Technology*, vol. 10, no. 3, pp. 198–203, 2014.

[13] Z. Kavehvash, K. Mehrany, and S. Bagheri, "Optimization of the lens-array structure for performance improvement of integral imaging", *Optics letters*, vol. 36, no. 20, pp. 3993–3995, 2011.

[14] B. Lee, S. Jung, and J.-H. Park, "Viewing-angle-enhanced integral imaging by lens switching", *Optics letters*, vol. 27, no. 10, pp. 818–820, 2002.

[15] Y. Kim, J.-H. Park, S.-W. Min, S. Jung, H. Choi, and B. Lee, "Wide-viewing-angle integral three-dimensional imaging system by curving a screen and a lens array", *Applied optics*, vol. 44, no. 4, pp. 546–552, 2005.

[16] J.-S. Jang and B. Javidi, "Improvement of viewing angle in integral imaging by use of moving lenslet arrays with low fill factor", *Applied optics*, vol. 42, no. 11, pp. 1996–2002, 2003.

[17] ——, "Three-dimensional projection integral imaging using micro-convex-mirror arrays", *Optics Express*, vol. 12, no. 6, pp. 1077–1083, 2004.

[18] R. Martınez-Cuenca, H. Navarro, G. Saavedra, B. Javidi, and M. Martinez-Corral, "Enhanced viewing-angle integral imaging by multiple-axis telecentric relay system", *Optics Express*, vol. 15, no. 24, pp. 16 255–16 260, 2007.

[19] H. Choi, S.-W. Min, S. Jung, J.-H. Park, and B. Lee, "Multiple-viewing-zone integral imaging using a dynamic barrier array for three-dimensional displays", *Optics express*, vol. 11, no. 8, pp. 927–932, 2003.

[20] M. A. Alam, K.-C. Kwon, Y.-L. Piao, Y.-S. Kim, and N. Kim, "Viewing-angle-enhanced integral imaging display system using a time-multiplexed two-directional sequential projection scheme and a deigr algorithm", *IEEE Photonics Journal*, vol. 7, no. 1, pp. 1–14, 2015.

[21] J.-S. Jeong, K.-C. Kwon, M.-U. Erdenebat, Y. Piao, N. Kim, and K.-H. Yoo, "Development of a real-time integral imaging display system based on graphics processing unit parallel processing using a depth camera", *Optical Engineering*, vol. 53, no. 1, p. 015 103, 2014.

[22] J.-H. Park, K. Hong, and B. Lee, "Recent progress in three-dimensional information processing based on integral imaging", *Applied Optics*, vol. 48, no. 34, H77–H94, 2009.

[23] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, and applications", *Applied optics*, vol. 52, no. 4, pp. 546–560, 2013.

[24] S.-W. Min, M. Hahn, J. Kim, and B. Lee, "Three-dimensional electro-floating display system using an integral imaging method", *Optics Express*, vol. 13, no. 12, pp. 4358–4369, 2005.

[25] K. Choi, J. Kim, Y. Lim, and B. Lee, "Full parallax viewing-angle enhanced computer-generated holographic 3d display system using integral lens array", *Optics Express*, vol. 13, no. 26, pp. 10 494–10 502, 2005.

[26] G. Baasantseren, J.-H. Park, K.-C. Kwon, and N. Kim, "Viewing angle enhanced integral imaging display using two elemental image masks", *Optics express*, vol. 17, no. 16, pp. 14 405–14 417, 2009.

[27] K.-C. Kwon, C. Park, M.-U. Erdenebat, J.-S. Jeong, J.-H. Choi, N. Kim, J.-H. Park, Y.-T. Lim, and K.-H. Yoo, "High speed image space parallel processing for computer-generated integral imaging system", *Optics express*, vol. 20, no. 2, pp. 732–740, 2012.

[28] J. Hahn, Y. Kim, E.-H. Kim, and B. Lee, "Undistorted pickup method of both virtual and real objects for integral imaging", *Optics express*, vol. 16, no. 18, pp. 13 969–13 978, 2008.

[29] N. Kim, M. A. Alam, L. T. Bang, A.-H. Phan, M.-L. Piao, and M.-U. Erdenebat, "Advances in the light field displays based on integral imaging and holographic techniques", *Chinese Optics Letters*, vol. 12, no. 6, p. 060 005, 2014.

[30] B. Lee, J.-H. Park, and S.-W. Min, "Three-dimensional display and information processing based on integral imaging", in *Digital Holography and Three-Dimensional Display*, Springer, 2006, pp. 333–378.

[31] B. Lee, S.-W. Min, and B. Javidi, "Theoretical analysis for three-dimensional integral imaging systems with double devices", *Applied optics*, vol. 41, no. 23, pp. 4856–4865, 2002.

[32] X. Jiao, X. Zhao, Y. Yang, Z. Fang, and X. Yuan, "Dual-camera enabled real-time three-dimensional integral imaging pick-up and display", *Optics express*, vol. 20, no. 25, pp. 27 304–27 311, 2012.

[33] Intel, *Embedded coded light 3d imaging system with full high definition color camera*, Jun. 2016. [Online]. Available: https://www.mouser.com/pdfdocs/intel_realsense_camera_sr300.pdf.

[34] G. Li, K.-C. Kwon, G.-H. Shin, J.-S. Jeong, K.-H. Yoo, and N. Kim, "Simplified integral imaging pickup method for real objects using a depth camera", *Journal of the Optical Society of Korea*, vol. 16, no. 4, pp. 381–385, 2012.

[35] M. Wilbert, *When should you use hls streaming?*, Feb. 2019. [Online]. Available: https://www.dacast.com/blog/hls-streaming-protocol/.

[36] A. Ndegwa, *What is a web application?*, cited April 2019. [Online]. Available: https://maxcdn4support.wpengine.com/visual-glossary/web-application/.

[37] B. Javidi, X. Shen, A. S. Markman, P. Latorre-Carmona, A. Martinez-Uso, J. M. Sotoca, F. Pla, M. Martinez-Corral, G. Saavedra, Y.-P. Huang, *et al.*, "Multidimensional optical sensing and imaging system (mosis): From macroscales to microscales", *Proceedings of the IEEE*, vol. 105, no. 5, pp. 850–875, 2017.

[38] *Integral imaging - optical information processing laboratory*, https://sites.google.com/site/oipcbnu/research/3dimgpro, Accessed: 2019-04-04.