EMITTER International Journal of Engineering Technology
Vol. 10, No. 1, June 2022, pp. 1~13
DOI : 10.24003/emitter.v10i1.604

Integrated Multi-view 3D Image Capture and Motion Parallax 3D Display System

Madan Lal[1], Shadi Khan Baloch[2], Shoaib R. Soomro[1]

[1]Electronic Engineering Department, Mehran University of Engineering and Technology, Jamshoro, Pakistan.
[2]Mechatronics Engineering Department, Mehran University of Engineering and Technology, Jamshoro, Pakistan.
Correspondence Author: shoaib.soomro@faculty.muet.edu.pk

Received January 3, 2022; Revised February 5, 2022; Accepted March 8, 2022

Abstract

We propose an integrated 3D image capture and display system using a transversely moving camera, regular 2D display screen and user tracking that can facilitate the multi-view capture of a real scene or object and display the captured perspective views in 3D. The motion parallax 3D technique is used to capture the depth information of the object and display the corresponding views to the user using head tracking. The system is composed of two parts, the first part consists of a horizontally moving camera interfaced with a customized camera control and capture application. The second part consist of a regular LCD screen combined with web camera and user tracking application. The 3D multi-view images captured through the imaging setup are relayed to the display based on the user location and corresponding view is dynamically displayed on the screen based on the viewing angle of the user with respect to the screen. The developed prototype system provides the multi-view capture of 60 views with the step size of 1 cm and greater than 40˚ field-of-view overlap. The display system relays 60 views providing the viewing angle coverage of ±35˚ where the angular difference between two views is 1.2˚.

Keywords: Multi-view Imaging, 3D Imaging, Motion Parallax Display, Image Rendering.

1. INTRODUCTION

The role of technology in the field of art, archeology, medicine, education, and entertainment is expanding over the past few decades. The 3D image capture and display technologies are widely used in such applications to capture, record, and visualize the real and virtual objects. A regular display screen used in everyday life is capable of displaying only two-dimensions of the space. On the other hand, a 3D display screen has the tendency to show the third dimension of the object or scene through capture and display of the depth information. A 3D display system provides enhanced and natural viewing experience as compared to the 2D display system. The depth perception in 3D display and imaging system is provided by triggering one of the depth cues of human vision. A major source to provide 3D perception is by using stereoscopy as depth cue [1, 2]. Stereoscopy is the process of capturing two views through the pair of cameras placed slightly apart from each other. The pair of cameras act as human eyes and record the scene from two slightly different positions. The recorded content is then displayed to a specialized stereoscopic screen, which can separate the stereoscopic views intended for each eye. The stereoscopy has been extensively used by display industry in recent years [2]. However, most of the stereoscopic displays require either high frequency LCD panels or polarizer sheets along with viewing glasses for the users. Another approach to provide 3D perception is through motion parallax depth cue. The motion parallax is a monocular depth cue which provides depth perception from the relative velocity of the object moving across retina of the user [1, 3]. The term motion here refers to the change in the position of the user with respect to object, while parallax refers to the change in the position of object image formed on the user's retina.

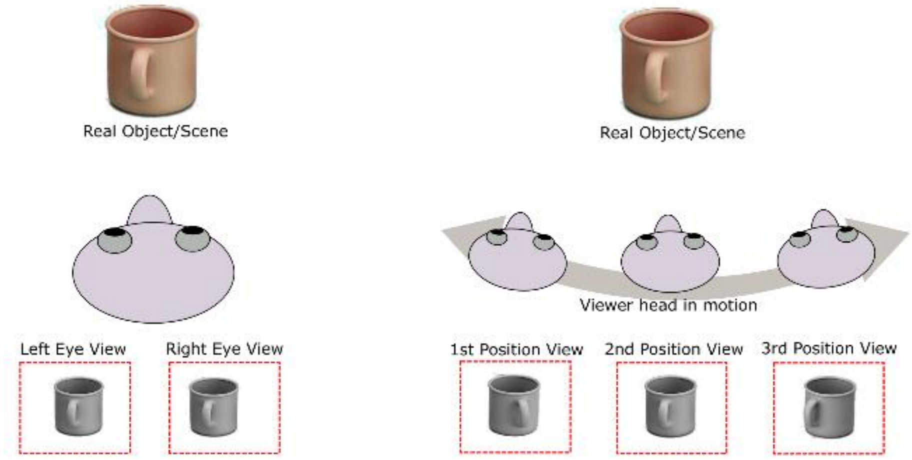

Figure 1: Illustration of the 3D depth perception through stereo parallax (left) and motion parallax (right). The arrow in right figure shows the direction of head motion.

In this paper, we present a motion parallax 3D display and multi-view imaging system in a simple and easily implementable format using a regular LCD screen, moving camera assembly and user tracking. Compared to the stereoscopic system, which captures only two views, the proposed system captures and displays the 60 views of the scene/object and covers the wide viewing range. A single moving camera precisely captures the object/scene from different viewing positions, which are processed and then visualized dynamically on the regular screen based on the viewing angle of the viewer. The multi-view capture is facilitated by using a conveyor-belt based mechanical assembly to move the camera in a controlled fashion. On the display side, the user tracking is performed by detecting the viewer's face through web camera and calculating the viewing angle with respect to the screen.

2. RELATED WORKS

There are wide variety of 3D image capturing techniques available in the literature. The simplest 3D imaging technique in the state-of-the-art consumer devices and smartphones is based on stereoscopy where two camera positioned close to each other are used to record the object from two different viewpoints [4]. The advanced imaging schemes involve variety of capture techniques including camera arrays, integral imaging, active depth cameras, micro lens array cameras and single moving camera-based systems. The camera array systems use large number of cameras organized in the form of 2D grid to capture the object/scene from different perspectives simultaneously [5, 6].

The integral imaging multi-view systems use lenslet arrays in front of the camera to capture and record the depth information of the scene. In [7], a double snapshot of the object is captured through field-like lens to image the reference lens on the micro lens array. In [8], the multi-view integral imaging scheme is proposed to perform gesture recognition through occluded spaces. In [9], a real-time multi-view capture and display system is proposed using conventional integral imaging pickup system technique combined with the computational algorithms. On the other hand, active depth cameras are either based on structured light where projected infrared light captured [10] or time-of-flight depth cameras where depth map of the object is generated by tracking the time taken by light rays to reflect through object [11].

The multi-view capture techniques proposed in this paper falls under the category of moving camera-based 3D imaging systems. Such systems use single camera and move it to the predefined positions to capture the object from different perspective. One of such technique is called axial 3D imaging in which the camera is moved along its optical axis and Multiview images are captured [12]. An enhanced version of axial 3D imaging is off-axial imaging in which the camera is moved in the diagonal direction with respect to its optical axis to address both transverse and longitudinal directions [13, 14].

The research work on motion parallax 3D displays is substantially increased due to its dreadful demand for augmented reality applications and systems. The motion parallax multi-view 3D displays have been proposed in literature using light field displays, head-mounted displays, and holographic display systems. The light field based motion parallax displays use the concept of rays projections in 3D space to address the parallax depth cue [15–17]. The full parallax multi-view 3D displays are realized by using holography in which the light wave pattern is produced using spatial light modulators [18, 19].

3. ORIGINALITY

The technology to capture, process and visualize 3D content is developing rapidly. In most of the previous research, the 3D capture and display has mostly been focused on the binocular depth cues, which require high speed liquid crystal panels and the polarizer or active shutter glasses for the viewer. On the other hand, the advanced 3D capture and display systems either require higher number of cameras for multi-view capture or complex optical components for the display, which make such systems expensive and out of reach from the common people.

Contributing from this study, we implemented the system by using off-the-shelf hardware components to achieve both multi-view image capture and motion parallax 3D display features. The approach presented in this paper is relatively easy and widely implementable in middle and low-income countries. We also provide the design details and real-time demonstration of the developed prototype, which make the proposed system realizable for various applications including online education, learning, medical visualization, art and archeology.

4. SYSTEM DESIGN

The proposed system consists of two main elements, 1) Multiview 3D capture assembly and 2) Motion parallax 3D display assembly. Figure 2 shows the block diagram of the system. The imaging system consist of a 2D imaging camera mounted on a horizontally moving motorized mechanical assembly. The moving assembly allows the motion of camera along horizontal direction. The motorized assembly is controlled through a microcontroller. The camera capture operation and its movement are commanded through a python program running on the computer. The step size of camera movement and number of required multi-view images are defined through the developed program application.

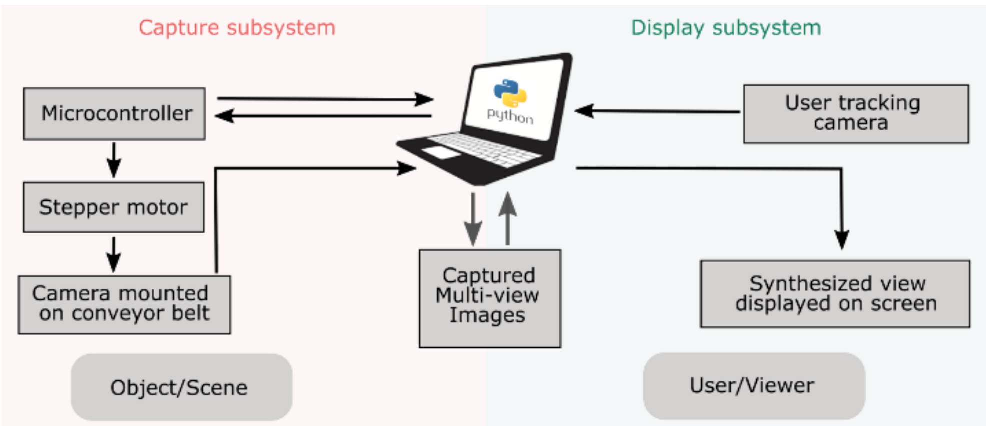

Figure 2: Block diagram of the proposed system showing capture and display subsystems.

The display assembly consist of an LCD display and web-camera connected to a CPU. The web-camera is used to track the user within the viewing range of the display and is fixed positioned at the center of LCD screen. The display subsystem is also controlled by the python program application, which detects the user position by detecting the face, calculates the distance and viewing angle of the user with respect to the screen and relays the desired view from the set of captured multi-view images.

4.1 Optical Geometry of the System

Some of the important parameters of moving camera or camera array based multi-view imaging systems are field-of-view (FoV) overlap between different views and optimal distance of the captured object/scene. These parameters mainly depend on the optical properties of the camera and geometric configuration of the system as illustrated in figure 3.

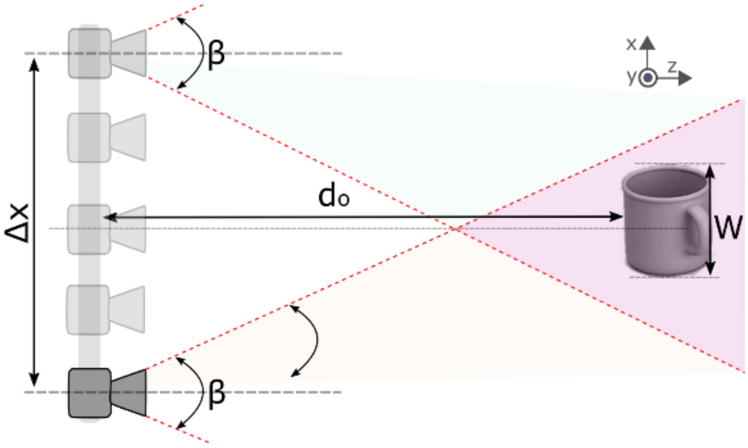

Figure 3: Geometric setup of the proposed imaging scheme.

Consider the β as horizontal of the camera in degree. The camera is translated along the horizontal axis within the range of 0 to Δx cm and the object of width W cm is placed at the distance of do from the capture setup. The proposed system moves the capture camera in straight direction i.e. x-axis only. Compared to the curved camera path, the straight-line conveyor belt movement provides the linear offset in the field-of-view of adjacent perspective views and avoids the need of projective goemetric correction of each captured view as required in curved camera assembly. A typical multi-view imaging system requires the presence of captured object in all captured views. Therefore, the object is required to be at the minimum distance from the camera, so that it is fully visible from all camera positions. For the proposed system, the minimum object distance depends on the horizontal field-of-view camera and width of object, which can be expressed as:

$$d_o = \frac{\Delta x - W}{2\tan\left(\frac{\beta}{2}\right)} \tag{1}$$

Since the proposed system has camera moving in single/horizontal direction only, the vertical field-of-view of different multi-views is fully overlapped. However, the amount of horizontal field-of-view overlap α depends on the object distance, maximum translation range Δx and native horizontal field-of-view β of the camera, which can be expressed as:

$$\alpha = \beta - 2\tan^{-1}\left(\frac{\Delta x}{2d_o}\right) \tag{2}$$

The motion parallax cue is mainly based on the direction from which the object/scene is being viewed. Therefore, it is important to determine the viewing angle of the user with respect to screen. Since the proposed system accounts for the horizontal motion parallax only, the viewing angle in this case is defined as the horizontal component of the angle between the user's face and center of the screen. Assuming Wp as pixel width, γ as horizontal field-of-view of the user tracking camera, the viewing angle ψ can be expressed as:

$$\psi = \left(\frac{2X_P - W_P}{2W_P}\right)\gamma \tag{3}$$

Where Xp is the horizontal pixel value of the centroid of the detected face.

4.2 Application Software

The application software for multi-perspective image capture and motion parallax display was implemented using python programming. Figure 4(a) shows the flowchart of the capture program. The capture program starts with the initialization of the camera and serial port, and the flag signal is sent to the microcontroller to move the camera one-step (i.e. 1 cm). Once the camera is moved, the perspective view of the object is captured, and the overlapping image region is cropped using equation 2. The captured view/image is then mapped with the viewing angle using equation-3 and the resultant perspective-view and its corresponding viewing angles are stored. The process is repeated until the camera is fully translated to 60 cm and all the perspective views are recorded.

Figure 4(b) shows the flowchart of display process. At first, the user tracking camera is initialized, new image frame is captured and converted to grayscale. The grayscale image is used as input for face detection. The Haar cascade classifier technique is used to detect the faces in the captured frame[20, 21]. The centroid of the face is determined by using bounding box coordinates of the detected face. The width and height of the bounding box is used to calculate the centroid of the face. The centroid is the geometric center of the bounding box of the face and may slightly vary from the actual centroid and center of the face due to instablity of Haar-cascade classifier. Since the view rendering process of the proposed system has minimum angular separation of 1.2° between adjacent views, the slight variation in the centroid of the detected face does not affect the overall display performance of the system.

The viewing angle is then estimated by using the centroid coordinates of the face, intrinsic properties of the camera and equation-3. The associated perspective image is then fetched and displayed to the screen. To provide the perception of motion parallax, the process is repeated in real-time until interrupted by the viewer/user.

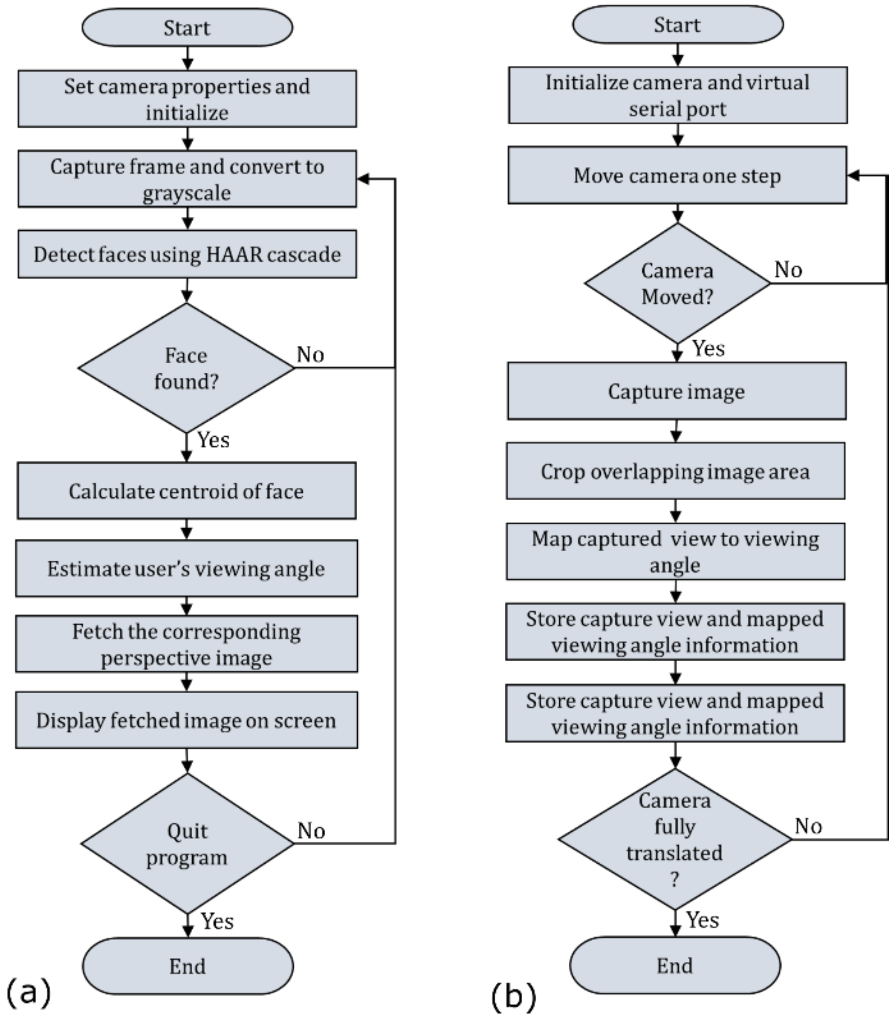

Figure 4: (a) shows the flowchart of the motion parallax display software application and (b) shows the flowchart of the multi-view capture software application.

5. EXPERIMENT AND ANALYSIS

5.1 Experimental Prototype

An experimental prototype setup was developed to practically demonstrate the process of multi-view 3D imaging and motion parallax 3D display as shown in figure 5. The 3D imaging setup, shown in figure 5(a), consisted of an off-the-shelf UBISOFT web-camera, a stepper motor NEMA17 along with A4988 motor driver module, Arduino Uno microcontroller, an in-house developed conveyor belt assembly fixed on a wooden base and a computer unit running the application program. The communication between python program and microcontroller was handled though a virtual serial port over USB interface, while the camera was connected to computer using USB interface. The camera has resolution of 640×480 pixels with 80˚×60˚ field-of-view. The conveyor belt assembly was constructed using aluminum railing, sliding bearings, timing belt and a pair of 16-teeth pully. The conveyor belt length was set to 80 cm which provided the maximum camera translation (Δx) of 60 cm. The minimum step length of the camera was set to 1 cm. In that way, the maximum 60 images each corresponding slightly different perspective view of the object were recorded in single capture event. On the other hand, the motion parallax display setup, shown in figure 5(b), consisted a 27' full-HD LCD screen and an A4tech web-camera, both connected to the computer terminal running a separate python program script for motion parallax display. The web-camera was placed at the top center of the screen and provided the viewing angle detection up to ψ=±35°. The user/viewer was positioned within the distance of 80 to 200 cm from the screen. The minimum and maximum user distance is limited by the field-of-view of the tracking camera and pixel resolution of the face image respectively. When the user distance is less than 80 cm, the tracking camera detects the face within the small area only. Similarly, for the distance greater than 200 cm, face resolution in the image is degraded and the face is not detected by the tracking camera. For the test experiments we used the optimum and comfortable viewing distance which was found to be in the rage of 100 cm.

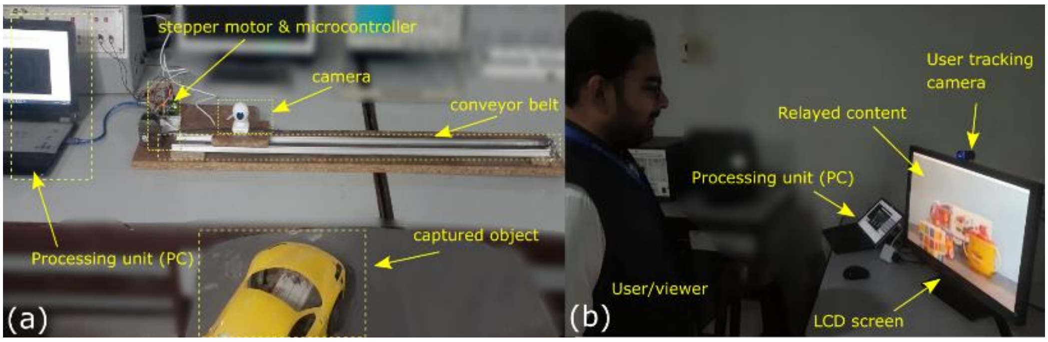

Figure 5: Experimental prototype setup, (a) shows the multi-view capture subsystems and, (b) shows the motion parallax display subsystem.

5.2 Results and Discussion

The multi-view image capture was demonstrated by using the developed prototype shown in figure 5(a). First, the object of interest was placed at a fixed distance from the setup. The optimal distance was determined using equation-1 and considering the physical dimensions of the object to make sure that the captured object is fully visible in all captured views. The capture process was trigged through the program application, the program application first sends the camera movement signal to the microcontroller. The microcontroller sends command to the motor. The motor moves at the rate of 1.8 degree/step, which corresponds to 3 mm translation of conveyor belt. Once the camera is moved to 1 cm, the microcontroller sends the flag signal to python program. After the flag signal is received, the camera captures the snapshot of scene and stores it in the hard drive.

Figure 6: The set of multi-perspective views of a real object i.e. toy car recorded using the developed prototype. 60 views are recorded with camera separation of 1 cm.

Figure 6 shows the complete set of multi-view images of the real object captured through the developed system. The captured object is a toy car of size 32 cm by 15 cm placed at 50 cm from the setup. The figure shows the smooth transition in perspective views of the object as the camera is translated.

Figure 7 show three camera positions and corresponding perspective views captured from those positions. Figure 7(a) shows the image result of the objects when captured from the initial camera position while figure 7(b) and figure 7(c) shows the recorded perspective view from mid position and extreme right position respectively. The figure clearly shows the horizontally shifting in the perspective views of the objects when moving from left to right direction, which are further used at display end to provide sense of motion parallax 3D.

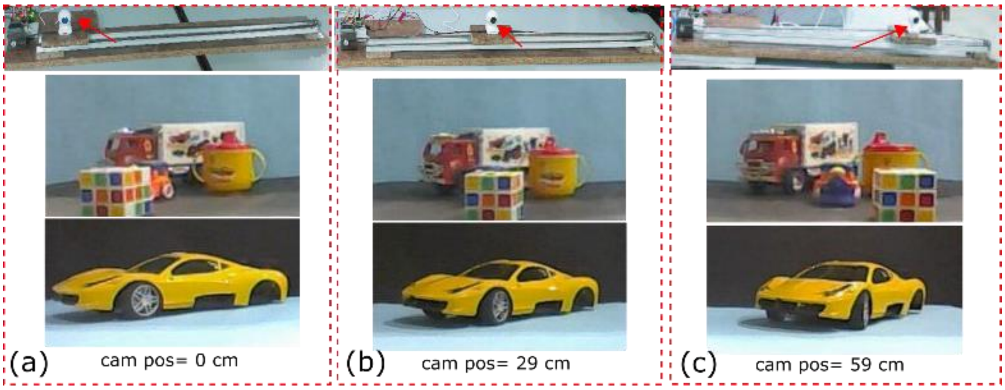

Figure 7: The results of captured multi-view images from different camera positions, (a) extreme left view from 1cm, (b) center view from 30 cm and (c) extreme right view from 60cm.

The multi-view capture capability of the system was analyzed by equation-2 and specifications of the experimental prototype. Figure 8 shows the amount of angular overlap between all captured perspective views when the distance of object is varied from 10 to 100 cm. Using the camera having 80° horizontal field-of-view, a maximum of 65° overlap can be achieved when the camera translation is limited to 20 cm and the object is placed at the distance of 100 cm from the capture setup. Figure 8 shows noticeable decrease in the field-of-view overlap when the camera translation range is increased, or the distance of object is decreased. The negative value of field-of-view overlap in figure 8 shows the amount of angler separation of first and last perspective views. In our experimental prototype, the maximum camera translation was set to 60 cm which provided more than 40° FoV overlap at the object distance of greater than 85 cm.

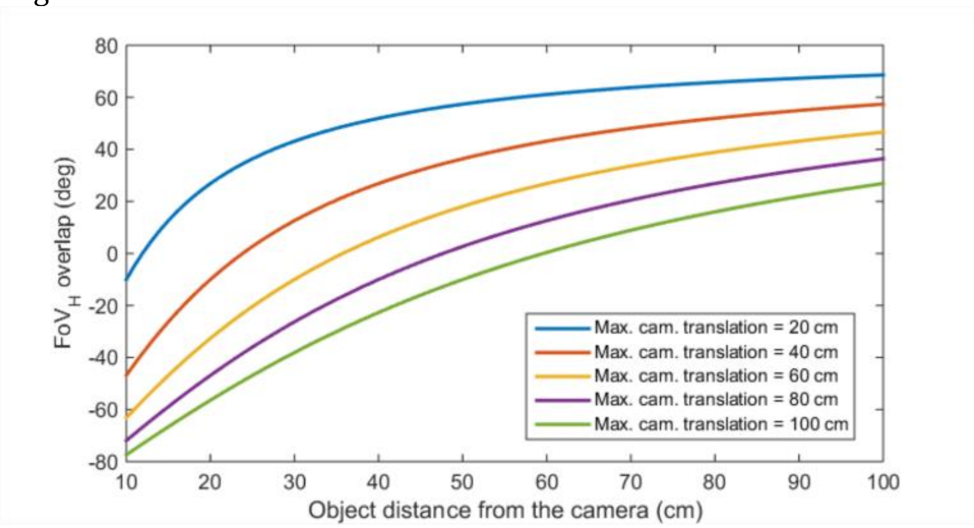

Figure 8: Horizontal field-of-view overlap as the function of distance between object and imaging subsystem when the maximum camera translation range is varied from 20 cm to 100 cm.

The real-time motion parallax display results are shown in figure 9 which show the user's face detection and the desired perspective view corresponding to the viewing angle as seen by the user.

The user is positioned at the distance of 100 cm from the screen. The number of perspective views are evenly mapped to the range of viewing angles. The demonstrated system covers the viewing angle of -35° to +35°. The angular resolution of the horizontal parallax is 0.85 views/degree. Figure 9(a) shows the user standing at the right side of the screen at the viewing angle of -25° and the corresponding perspective view of the captured object. Similarly, figure 9(b) and figure 9(c) show the resultant perspective views at the viewing angle of 0°and +20° respectively. Due to the real-time view rendering and tracking of the user's position, the system provides a perception of motion parallax depth cue.

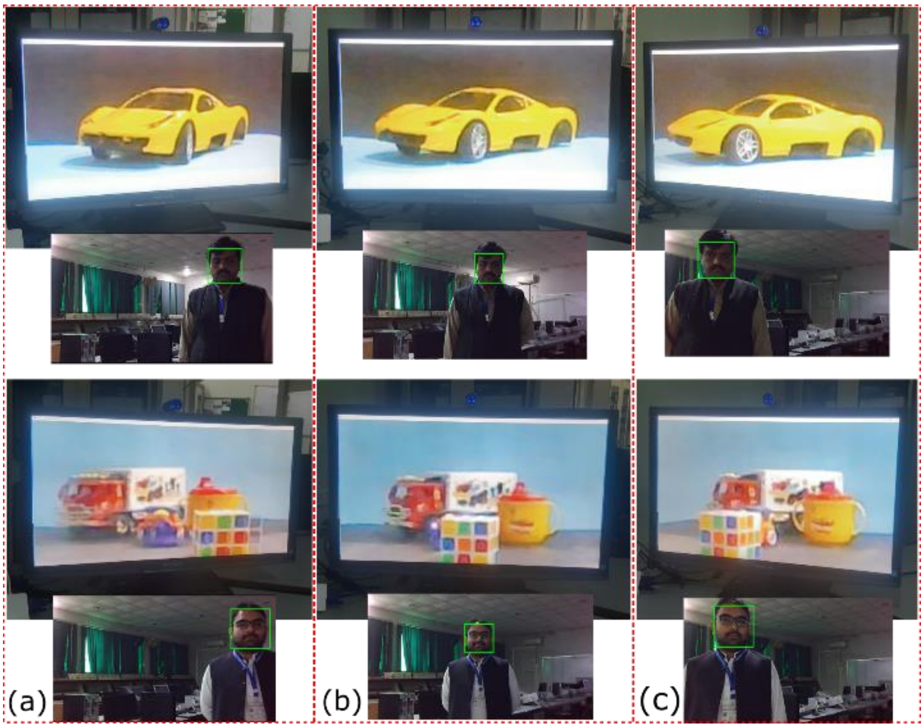

Figure 9: The motion parallax display results using developed prototype where two users are viewing the different objects on the screen when moving through different viewing angles.

6. CONCLUSION

A combined multi-view 3D imaging and motion parallax 3D display system was proposed and demonstrated using a 2D capture camera, moving mechanical assembly and regular LCD screen. The perspective views of the real objects were recorded by moving a camera in transverse direction. The motion parallax depth cue was exploited to provide the sense of depth using 2D screens. The optical geometry of the system was studied to understand the field-of-view overlap between captured views and calculate viewing angle of the user while displaying the captured content. Python programming-based software application was developed for the imaging and display procedures. The developed experimental prototype was tested by capturing the multi-view images of the real-objects and displaying them on the screen. The test prototype was able to capture the 60 perspective views along the camera translation range of 60 cm, which provided the separation of 1 cm between each perspective views. A 65° field-of-view overlap between the all perspective images was achieved at the object distance of 80 cm. The user tracking was performed within the viewing angle range of ±35°. The image capture quality can be further improved by replacing off-the-shelf web-camera with high resolution CMOS camera. The current implementation focuses on the horizontal parallax only which can be further extended to vertical parallax capture as well by the camera in 2D grid pattern.

Acknowledgements

The authors are thankful to electronic engineering department of Mehran University of Engineering and Technology for providing the resources to conduct this research work.

REFERENCES

[1] B. Rogers and M. Graham, **Similarities between motion parallax and stereopsis in human depth perception**, *Vision Res.*, vol. 22, no. 2, pp. 261–270, 1982.

[2] J. Geng, **Three-dimensional display technologies**, *Adv. Opt. Photonics*, vol. 5, no. 4, pp. 456–535, 2013.

[3] B. Rogers and M. Graham, **Motion parallax as an independent cue for depth perception**, *Perception*, vol. 8, no. 2, pp. 125–134, 1979.

[4] R. Kumar, **Smartphone 3D Camera Market by Technology,** 2018. [Online]. Available: https://www.alliedmarketresearch.com/smartphone-3d-camera-market. [Accessed: 02-Feb-2021].

[5] W. Matusik and H. Pfister, **3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes**, *ACM SIGGRAPH 2004 Pap. - SIGGRAPH '04*, vol. 23, no. 3, p. 814, 2004.

[6] B. Wilburn *et al.*, **High performance imaging using large camera arrays**, *ACM Trans. Graph.*, vol. 24, no. 3, p. 765, 2005.

[7] H. Navarro, J. C. Barreiro, G. Saavedra, and B. Javidi, **High-resolution far-field integral-imaging camera by double snapshot**, *Opt. Express*, vol. 20, no. 2, pp. 890–895, 2012.

[8] V. J. Traver, P. Latorre-carmona, E. Salvador-balaguer, F. Pla, and B. Javidi, **Three-Dimensional Integral Imaging for Gesture Recognition Under Occlusions**, vol. 24, no. 2, pp. 171–175, 2017.

[9] J. Kim, J.-H. Jung, C. Jang, and B. Lee, **Real-time capturing and 3D visualization method based on integral imaging**, *Opt. Express*, vol. 21,

no. 16, pp. 18742–18753, 2013.

[10] Z. Cai, X. Liu, G. Pedrini, W. Osten, and X. Peng, **Structured-light-field 3D imaging without phase unwrapping**, vol. 129, no. January, 2020.

[11] V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun, **Real time motion capture using a single time-of-flight camera**, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 755–762.

[12] R. Schulein, M. Daneshpanah, and B. Javidi, **3D imaging with axially distributed sensing**, *Opt. Lett.*, vol. 34, no. 13, pp. 2012–2014, 2012.

[13] Y. Piao, H. Qu, M. Zhang, and M. Cho, **Three-dimensional integral imaging display system via off-axially distributed image sensing**, *Opt. Lasers Eng.*, vol. 85, pp. 18–23, 2016.

[14] D. Dunn *et al.*, **Wide Field of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors**, *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 4, pp. 1275–1284, 2017.

[15] M. Date, S. Shimizu, D. Mikami, and Y. Kusachi, **Smooth motion parallax and high resolution display based on visually equivalent light field 3D**, vol. 1140205, no. April 2020, p. 4, 2020.

[16] H. Watanabe, N. Okaichi, T. Omura, M. Kano, H. Sasaki, and M. Kawakita, **Aktina Vision: Full-parallax three-dimensional display with 100 million light rays**, *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.

[17] P. Wang *et al.*, **Demonstration of a low-crosstalk super multi-view light field display with natural depth cues and smooth motion parallax**, *Opt. Express*, vol. 27, no. 23, p. 34442, 2019.

[18] A. Maimone, A. Georgiou, and J. S. Kollin, **Holographic near-eye displays for virtual and augmented reality**, *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–16, 2017.

[19] K.-H. L. Son, Jung-Young, Hyoung Lee, Beom-Ryeol Lee, **Holographic and Light-Field Imaging as Future 3-D Displays**, *Proc. IEEE*, vol. 105, no. 5, pp. 789–804, 2017.

[20] M. H. Abdurrahman, H. Amran Darwito, and A. Saleh, **Face Recognition System for Prevention of Car Theft with Haar Cascade and Local Binary Pattern Histogram using Raspberry Pi**, *Emit. Int. J. Eng. Technol.*, vol. 8, no. 2, pp. 407–425, 2020.

[21] C. Li, Z. Qi, N. Jia, and J. Wu, **Human face detection algorithm via Haar cascade classifier combined with three additional classifiers**, *ICEMI 2017 - Proc. IEEE 13th Int. Conf. Electron. Meas. Instruments*, vol. 2018-Janua, pp. 483–487, 2017.