Interactive and Gesture-Capable 3D Holographic Light Field Display with Registered Interaction between User and Light Volume

Iván Alexis Sánchez Salazar Chavarría[1], and Masahiro Yamaguchi[1]

sanchez.a.aa@m.titech.ac.jp
[1]School of Engineering, Tokyo Institute of Technology, 4259-G2 Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8502, Japan
Keywords: Light-field, 3D-interaction, holography, touchable interface

ABSTRACT

To achieve the registration between the reconstructed content and the position of the user in 3D displays capable of reconstructing a real image in mid-air, we propose a method based on simultaneous use of scattered light detection and a stereo vision-based gesture sensor.

1 INTRODUCTION

In recent years, we have seen the introduction of 2D aerial displays as well as 3D displays that reproduce a light volume in mid-air. Although they still do not offer the image quality found in 2D screens, its rapid adoption as interfaces in hygienic environments, navigation applications (e.g. heads-up displays in cars), and hands-busy settings in general, demands improvements on the easiness of use and on its overall user experience. Aerial 2D displays (2D image floating in air) are well advanced into its development, with its standardization already being on progress. They are capable of detecting real time gestures and interaction, having very similar capabilities to the current touchable 2D displays. Expanding this technology to include images presented in different depths in the same frame can potentially offer more functionalities and a closer experience to interacting with objects as if they were real. Of particular interest, is the light field technique which is an active research area [7,11]

There exist not only several studies, but also commercial applications in which hand tracking and gestures are used to control 3D images reconstructed by light field displays or holographic displays. However, these interfaces do not match the location of the gesture with the reconstructed content, degrading the interaction and making it less similar to a real interaction. This problem has motivated our research, where we have used the light scattered by the reconstructed light field when the user interacts with it [1]. More recently, we have used the color of the light along with the 2-dimensional coordinates of an RGB camera to track the hand of a user in 3D [2]. To increase the variety of interactions in our system, we have also proposed the combination of scattered light with an off-the-shelf stereo vision-based gesture sensor (Leap Motion controller [3]). Since our method registers naturally the positions of the user and the content, combining it with a sensor capable of registering 3D gestures can increase the capabilities of a future system [10]. In this work, we provide details of the processing speed of the integral image, the handling of the interaction under illumination and the implementation of gestures based on the Leap Motion controller (LM)

2 BACKGROUND

There have been several proposals to create an interface where the content pops up from the screen, floating in mid-air and closer to the user. There are examples using 2D aerial interfaces [4], 3D ray reconstruction-based light field displays [5], and even 3D wave front reconstruction-based holographic displays [6]. The problem that persists in all of them is that registering the place of interaction with the place of reconstruction seems to be left unattended. This problem is very important if we are to have a high-quality interface and not just an experimental prototype. This problem has been addressed by our group before ([1-2]) by using the light that is scattered when the user interacts with the light of the reconstructed volume, resulting in a more direct interaction. In this study, we propose a method to match the reconstructed content and the place of interaction using scattered light. We also present how the use of this registration permits the implementation of gestures that can modify the light field in real time. Using this method, it is possible to grab, drop, rotate and change the size of the light field of an object.

3 PRINCIPLE

3.1 Description of the system

Our system is similar to the one described in [1-2]. It consists of an array of Holographic Optical Elements (HOEs) that have the functionality of an array of convex mirrors. The impinged light in this array is controlled with a commercial projector, whose output is collimated (see fig. 1). An alignment procedure is used to match the projector coordinates with the spatial location of the center of each HOE. The integral image [7] of the 3D object to be reconstructed is computed either by a 3D rendering software (e.g. Blender [8]) or by a GPU based implementation (e.g. OpenGL). In this study, we use an OpenGL implementation to generate moving objects in real time, while more elaborate light fields whose movement does not require real time processing are rendered with Blender. The combination of both integral images (GPU, real-time generated and images rendered beforehand) can aid in the creation of an interface with more elements (e.g. letters, indications, etc.).

3.2 3D tracking based on scattered light

When the user interacts with the reconstructed light volume using his or her finger, light will be scattered. This signal is captured by an RGB camera and processed to identify the presence of interaction. The displacement of the scattered light inside the camera indicates the movement of the user in a certain direction. When not only scattered light, but also its color detection is considered, the y-direction can be encoded in the color. More details on this implementation can be found in [2]. The color of the light is also identified and compared to a previously registered color value.

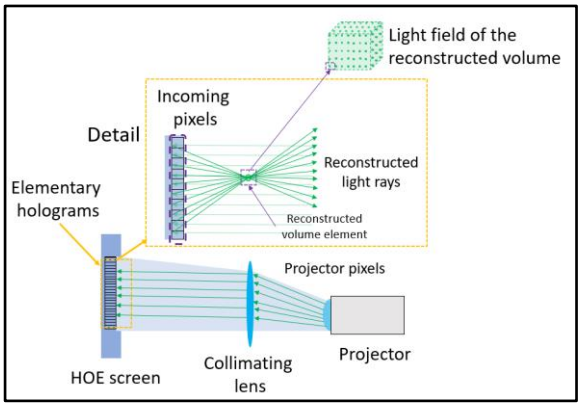

Figure 1. Reconstruction of the light field of an object in a projection-based holographic light field display

3.3 Gesture-sensor integration

Although the method described in the previous section is capable to track the position of the fingertip of a user in 3D using an RGB camera, the detection of more complicated movements and gestures that include more fingers is desirable to improve the usability of the system. In this work, , we decided to integrate into our system the off-the-shelf Leap Motion sensor (LM [3]). The LM sensor is an infrared stereo camera that uses a skeleton-based model of the hand which is combined with the depth measurement provided by the stereo vision to measure several features of the hand (e.g. fingertips positions, center of palm, direction of normal vectors, etc.)

These features are measured in a coordinate system whose origin is the center of the device itself. To use these 3D measurement capabilities and make them correspond with the position of the reconstructed objects, a registration process should be implemented. A similar procedure combining the LM sensor and a light field display has been performed in a previous study [9]. Adding to that study, here we implement gestures and real-time light field motion. Additionally, we consider how the measurement of the scattered light can aid to improve the registration between the content and the device.

3.4 Scattered light for position registration

We propose to use the scattered light as a cue to indicate to the LM controller the location of the coordinate system of the HOE screen. The color detection is also used to encode different positions within the display space. Let $\bar{p}_D^1 = (x_D^1, y_D^1, z_D^1) \in \mathbb{R}_D^3$ be a position in the display space, denoted by $\mathbb{R}_D^3$. By reconstructing a light volume in $\bar{p}_D^1$, with

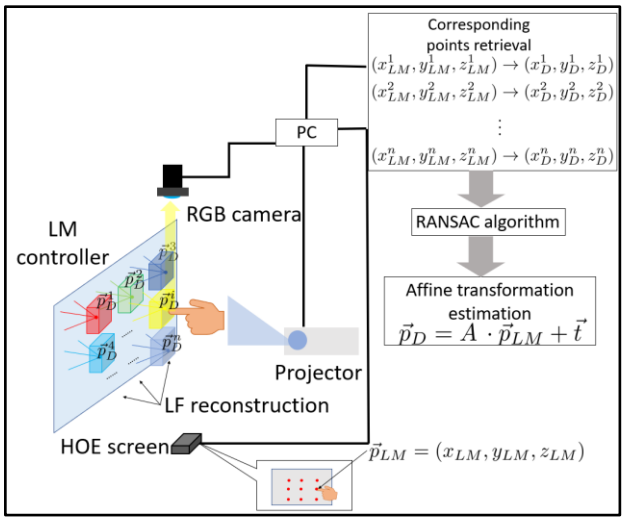

Figure 2. Position registration between the LM controller and the light field display based on scattered light

a certain color $\bar{C}_1 = (R_1, G_1, B_1)$, we can recover the position of the finger as being $\bar{p}_D^1$ whenever $\bar{C}_1$ is scattered and detected by the RGB camera (see fig. 3). Running the LM controller in parallel to this process, we can retrieve the coordinates of the finger in the LM space, obtaining a point $\bar{p}_{LM}^1 = (x_{LM}^1, y_{LM}^1, z_{LM}^1) \in \mathbb{R}_{LM}^3$, achieving a correspondence between the points $\bar{p}_D^1$ in the display space and $\bar{p}_{LM}^1$ in the LM coordinate system. Repeating this process for *n* points ($n \geq 4$) will provide the input of a RANSAC implementation that estimates an affine transformation between the display space $\mathbb{R}_D^3$ and the LM space $\mathbb{R}_{LM}^3$

3.5 Gestures implementation

The LM controller uses an implementation of a hand model to detect features of the hand extracted from the captured video of the built-in stereo camera of the device. It provides the position of all the fingers of the hand in 3D space, as well as the center of the palm, the normal vector to the palm and its Euler angles. All this information can be used to implement gestures based on either the change in distance of the different fingers (the index and the thumb, for example) or the change in orientation of the vectors (see fig. 3, c and d).

4 EXPERIMENT

The experimental setup for this implementation consisted in an array of holographic optical elements (HOEs) that has the function of an array of convex mirrors. An RGB camera is located on top of the display and the light scattered after the interaction of the user is captured and processed to identify it as belonging to one of the sets of color previously captured (see [2]). A LM controller is also located below the screen to capture the gestures of the user, as depicted in fig. 3. The position read-out by the LM controller will be saved after the camera captures the scattered color by the user and the color identification associates the scattered color to one of the registered colors and, therefore, to one of the

reconstructed positions within the display render space (see fig. 3).

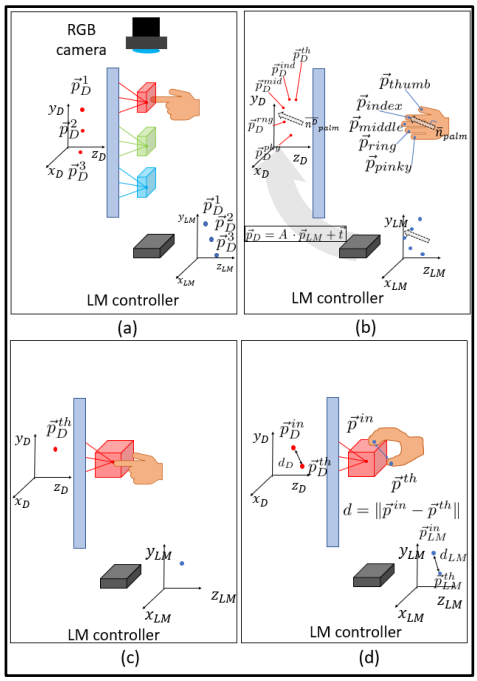

Figure 3. Process of registering the position of the LM controller and the display space through the use of scattered light (a-c). Once the calibration process is finished, the extracted features are used to implement gesture detections like grabbing (d)

4.1 Real time processing

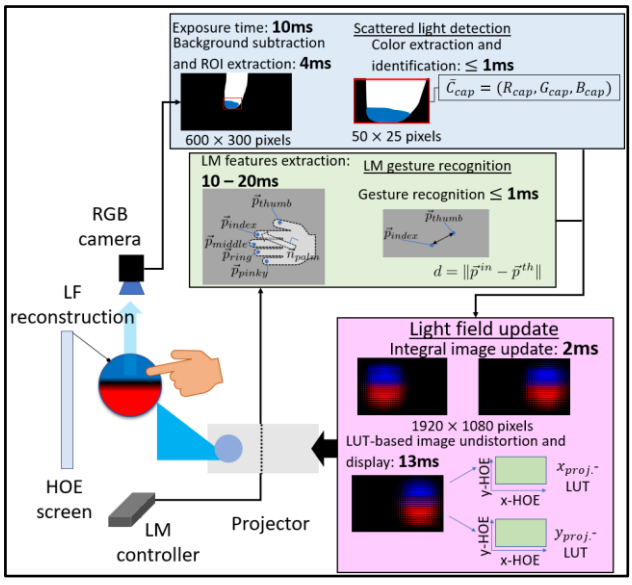

Figure 4. Detail of the processing time required for each step of the process. The total latency will depend on what routines are combined (scattered light, LM controller or both)

The light field reconstructed in this system is obtained from the projection of an integral image of the desired scene, which is obtained from decomposing the pixels of a scene captured from different views to be reconstructed in the right angular position upon projection (for details see [7]). The processing time of every step of the process from the user's interaction to the update of the light field was measured and optimized to obtain a low latency (see fig. 4). The scattered light detection takes approximately 15 milliseconds (ms), depending on the exposure time of the camera and the size of the region of interest (ROI) selected. The LM controller is also able to obtain the hand features shown in fig. 5 in 10 ms (100 fps). However, depending on the position of the hand and the gesture it performs, it can take up to 20 ms if the hand is in a position were the LM processing takes more time to detect the shape of a hand (e.g. pinching gesture). It is possible to choose between scattered light interaction and LM based interaction. When the position is registered using scattered light, this process is performed sequentially and the times are summed up. The processing of the light field using a GPU-based OpenGL implementation allows for a rapid update of 2 ms, but the distortion step detailed in [2] is still the most time-consuming routine.

4.2 Color detection under room illumination

Since this approach depends on the detection of the light scattered by the user, and this light is in general combined with the light present in the environment where the system is located (normally an indoor office), the system may not be able to identify the scattered light. To increase the usability of this approach in an illuminated environment, we propose to sample the finger of the user when not interacting with the light field and use this value

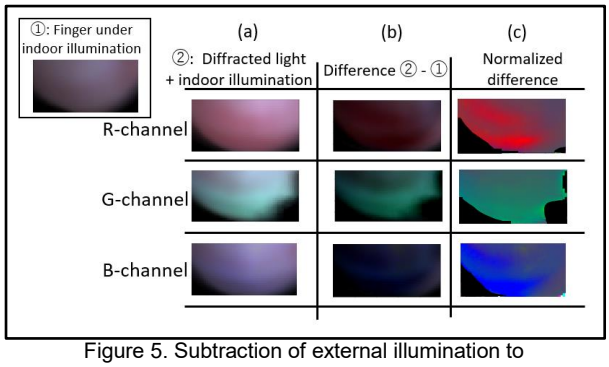

Figure 5. Subtraction of external illumination to increase the accuracy of the color detection

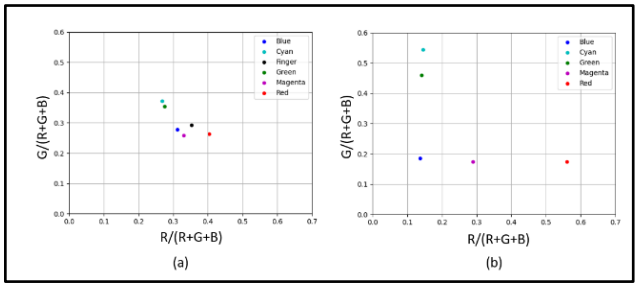

Figure 6. Change in the mean values of the detected colors after subtraction of finger illuminated by room illumination: (a) before and (b) after.

to offset the scattered light interaction. The subtraction between the scattered light and the finger under

environmental illumination is normalized and a more robust signal can be obtained even in a lit-up environment (see fig. 5).

5 RESULTS

During the color acquisition, the use of the color information with and without the proposed subtraction in section 4.2 is compared in fig. 6. Here, the separation of the main values of the acquired colors provides a more robust identification of the scattered colors.

The acquisition of the positions of the rendered colored buttons in the LM space was performed using an ATM-like interface (fig. 7), and the detected colors in the different positions were used as a cue for the LM controller to save the position of the user. These points were used to estimate the affine transformation between both coordinate systems as explained in section 3.4

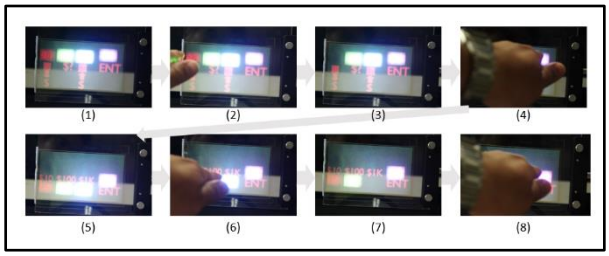

Figure 7. Retrieval of scattered light to register the LM controller position

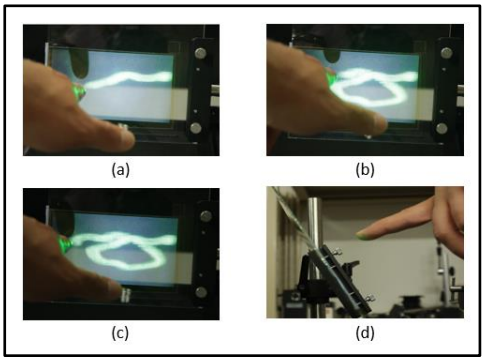

Figure 8. Generation of the light field of the trajectory traced by the user (a-c). Side view showing the scattered light on the tip of the finger (d).

This aligned system was then used to create the 3D trajectory of the finger of the user passing in front of the HOE screen (see fig. 8). The real time operation is made possible by the low latency of the GPU-based implementation.

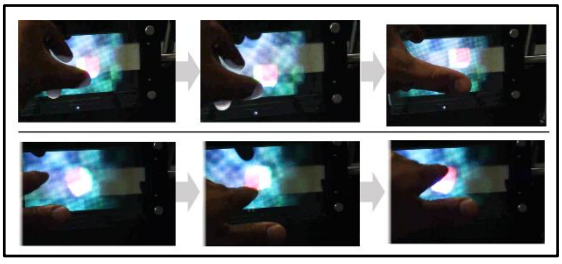

Figure 9. Gestures matching the light field registration. Grab and drop (up) and poke and swipe (down)

The alignment of the coordinate systems was used to implement gestures that match the reproduced content (fig. 9). Using the distance between the index and the thumb, a grab and drop gesture that mimics the grab and drop of the light field volume as if it were a real object was implemented. The registration procedure also allowed for the implementation of “poking and spinning” gesture in which the user was able to spin the light field when the finger position matches the location of the light field.

6 CONCLUSIONS

We have demonstrated how the scattered light that arises from the interaction between the light field and the user can be used to align the coordinate systems of a commercial tracking sensor and a holographic light field display that reconstructs a real image in air. This method creates a direct way in which the user can interact with the virtual content, and can be considered to develop more elaborated applications such as 3D pattern recognition for identification purposes or some other interactive applications.

References

[1] M. Yamaguchi and R. Higashida, Appl. Opt. Vol. 55, No. 3, A178–A183 (2016).

[2] I.A. Sánchez Salazar Chavarría, T. Nakamura and M. Yamaguchi, Opt. Express Vol. 28, No. 24, pp. 36740-36755 (2020).

[3] Ultraleap official website. https://www.ultraleap.com

[4] H. Yamamoto et al, Int. Conf. on 3D Imaging (IC3D), pp. 1-5 (2014).

[5] Looking Glass factory Inc. official website https://lookingglassfactory.com/

[6] M. Takenaka, T. Kakue, T. Shimobaba and T. Ito, IEEE Access, Vol. 9, pp. 36766-36774 (2021).

[7] Martínez-Corral, Manuel, and Bahram Javidi." *Advances in Optics and Photonics*, Vol. 10 No. 3 pp. 512-566. (2018).

[8] Blender https://www.blender.org/

[9] V. K. Adhikarla, J. Sodnik, P. Szolgay and G. Jakus, Sensors Vol. 15, No. 4, pp. 8642-8663 (2015)

[10] I. A. Sánchez Salazar Chavarría, T. Nakamura, and M. Yamaguchi, 2021 OSA Imaging and Applied Optics Congress, 3Th7E.2, (2021).

[11] Pan, X and Komatsu S., Appl. Opt. Voll. 58, No.23,6414-6418 (2019).