

Copyright 2021 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this publication for a fee or for commercial purposes, and modification of the contents of the publication are prohibited.

This is the accepted version of the paper available at <https://doi.org/10.1117/12.2589129>

# On the simulation of hand-held cameras in light field rendering

Mary Guindy<sup>ab</sup>, Attila Barsi<sup>b</sup>, Peter A. Kara<sup>c</sup>, Tibor Balogh<sup>a</sup>, and Aniko Simon<sup>d</sup>

<sup>a</sup>Holografika, Budapest, Hungary

<sup>b</sup>Pazmany Peter Catholic University, Budapest, Hungary

<sup>c</sup>Budapest University of Technology and Economics, Budapest, Hungary

<sup>d</sup>Sigma Technology, Budapest, Hungary

## ABSTRACT

Camera motions are essential to numerous fields of arts and science, such as cinematography, video games, 3D simulations and many more. Some of these applications rely on physical cameras, while others use virtual cameras. Despite the efficiency of virtual cameras, their motion misses some of the subtle details of real camera movements. Since hand-held camera motion accounts for the noise factor produced by hand tremors and muscle fatigue, the resulting output is more realistic than what virtual cameras produce in general. Much research has been done to produce simulations for realistic hand-held cameras; however, implementing these techniques for light field rendering has not been investigated yet. Among the different solutions to produce realistic hand-held camera motions, databases could be set up by collecting data from real hand-held cameras, although this method requires extensive data to be recorded from different hand-held cameras in order to produce reliable results. In addition to generating databases for hand-held camera simulation, jitter could be used. Since hand tremors and muscle fatigue result in adding slight details to the camera motion paths, jitter models could be used to simulate deviations. A camera motion path is defined by a set of curve functions, which could be taken into account when adding noise models to produce the hand-held camera effect. In addition to the defined camera motion path, camera orientation, location, speed and acceleration can be also considered when adding jitter. For example, if the camera accelerates, jitter shall increase accordingly. Even though the work on hand-held camera simulation techniques is an ongoing research and valid solutions are already available, applying those techniques to light field visualization have not been performed yet. Contrary to rendering light fields from physical camera arrays, virtual scenes can be rendered for each ray of the display's light field. Rendering can be performed with both ray tracing and rasterization techniques. Both of these techniques involve camera (region of interest) placement and, therefore, they allow us to perform camera motion simulations. Rendering in such a manner also eliminates all sampling and conversion artifacts, thus making it more suitable for the evaluation of visual comfort. In this paper, we introduce the scientific considerations regarding our on-going long-term work on the simulation of hand-held camera motions for light field displays by means of jitter. The technical discussion covers every aspect of the procedure, from research goal and measurement utilities to visualization and quality assessment.

**Keywords:** Light field, hand-held cameras, camera motion, camera motion path, 3D rendering.

## 1. INTRODUCTION

Light field technology has been continuously emerging in the past decades. At the time of this paper, the implementation of real light field displays is already a reality, as such glasses-free 3D displays are present in the industry and novel research is pointing towards commercial use cases. Although they are yet to appear on the consumer market, the practical potential of this so-called “natural 3D” is nearly limitless.

---

Further author information:

Mary Guindy: E-mail: m.guindy@holografika.com / guindy.mary.mohsen.messak@ppke.hu

Attila Barsi: E-mail: a.barsi@holografika.com

Peter A. Kara: E-mail: kara@hit.bme.hu

Tibor Balogh: E-mail: t.balogh@holografika.com

Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

Nearly limitless is of course not completely limitless, as there are indeed certain challenges and limitations,<sup>1</sup> yet the great efforts of the scientific community are contributing immensely to overcome them. Through innovative technologies such as the HoloVizio displays<sup>2</sup> and the better understanding of the Key Performance Indicators (KPIs) of light field visualization,<sup>3</sup> notable progress is being made to achieve completely new heights of realism in visual experience. Summa summarum, light field has come a very long way since the early days of integral imaging.<sup>4</sup>

However, in so many cases, realism is rather difficult to achieve solely through technology, as numerous aspects originate from the physiology and behavior of human beings. They need to be considered and adapted accurately, authentically, in order to properly mimic real life. Let us not delay the introduction of the research question of this paper any longer and elaborate in medias res: such instance of relevant human physiology and behavior is the usage of hand-held cameras in the context of light field technology.

At this point, two relevant facts must be highlighted: (i) the handling of cameras for light field capture has never been addressed before, and (ii) investigating this phenomenon is far from being straightforward. Therefore, in this paper, as our primary scientific contribution, we provide a comprehensive guideline, a set of exhaustive technical considerations regarding such research.

The remainder of this paper is structured as follows: Section 2 introduces the considerations regarding the simulation of hand-held cameras, separately for each individual topic and aspect. The associated subsections are ordered by research phase maturity. The paper is concluded in Section 3.

## 2. CONSIDERATIONS REGARDING SIMULATION

### 2.1 Research goal

The research goal in this specific area is to study recorded contents of hand-held cameras on light field displays. Such scientific efforts aim to prepare hand-held light field cameras to create content analogous to those of conventional 2D cameras. Furthermore, recording camera contents are necessary for the simulation of hand-held cameras in virtual spaces. The ultimate aim of such simulations is to replicate real-life camera motion. In order to achieve this, devices are required to measure the precise data regarding the spatial parameters of the camera (i.e., current position and motion).

### 2.2 Tracking via markers

One possible way to measure spatial parameters is to attach markers to the camera rig. Markers can be optical,<sup>5</sup> magnetic<sup>6</sup> or radio. Optical markers are more demanding in numbers than magnetic solutions, yet magnetic markers can be disrupted by other magnetic devices in the vicinity of the measurement. Interference is also an issue when measuring the position and orientation of radio markers. The most reliable and the most accurate solution is the utilization of magnetic markers, if and only if no other magnetic devices interfere. Regarding optical markers, measurement accuracy fundamentally depends on the resolution and recording speed of the camera, which can be somewhat compensated by the use of multiple cameras, and of course by attaching more markers. The latter is limited by the fact that in case of certain types of optical markers (i.e., larger physical markers, such as “constellations”, and not light-weight stickers), the additional optical markers may affect the weight and the weight distribution of the investigated entity (i.e., physical object), and may also change the requirements towards software complexity. It also needs to be noted that using multiple cameras requires camera calibration. Calibration is also a de facto necessity for magnetic and radio markers. Optical markers typically vary in color, in order to support the direction assessment (along the  $x$ ,  $y$  and  $z$  axes) of the entity. Radio markers are similar to magnetic markers, however, they are significantly more expensive. In case one wishes to establish a highly cost-efficient solution, one could simply attach a smart phone to the rig and use its sensors.<sup>7</sup> Yet keep in mind that the previously mentioned concern about weight and weight distribution still applies, which may result in certain issues regarding the validity of the outcome of the measurement.

### 2.3 Connecting the rig to the display

It is possible to connect the rig to a light field display, and thus provide real-time data to the rendering subsystem during the measurement. This is only a valid option for rendering virtual scenes, since according to the operational properties of the state-of-the-art light field systems, arbitrary real-time light field visualization from a rig that allows variations between the extrinsic camera parameters (i.e., the precise spatial allocation of the cameras with respect to each other may be altered) is not possible. At the time of this paper, such technique can only be applied in the case of telepresence systems,<sup>8</sup> where the camera rig is immobile, fixed, and therefore, the extrinsic camera parameters are constant.

### 2.4 Camera rig types

Depending on the type of camera we wish to simulate, we can distinguish wide-baseline<sup>9</sup> and narrow-baseline<sup>10</sup> camera systems. Most of the currently available real light field displays necessitate wide-baseline inputs, but the appropriate generation of such input is complicated and quite expensive. This can be substituted by a rig that only relies on a single camera, but takes into account the spatial parameters (i.e., the implemented rig is larger in size) and uses markers accordingly, and in fact, we may even use a rig without any camera whatsoever and track its movement. This solution is already sufficient to register the investigated movements (e.g., hand tremors) that apply to hand-held cameras. In experimental practice, following a moving object with the rig is advised, as it is a common recording scenario in everyday use. Therefore, it is important to highlight that using a rig without a camera removes the opportunity to provide visual feedback to the operator of the rig. For moving objects, placing two cameras instead of a single camera on the rig is highly recommended, since they form a stereo camera pair, capable of measuring depth information based on the disparity between the two images. This acquired information may be enable a more cost-efficient solution compared to conventional tracking via markers, as similarly accurate distance measures can be obtained without the need to extend the tracking system to cover the target. Note that using two cameras requires calibration.

### 2.5 Camera rig weight and size (design)

Regarding the rig itself, the most important considerations are of weight and size. First of all, equipping a rig with a full array of cameras may make it virtually immobile in the investigated use case, as the resulting weight may exceed the limitations of the human body (i.e., it could not be lifted, and even if it is actually lifted, one could not perform the recording task properly). Even if weight is not an issue, a greater size may introduce oscillation at the left and right ends of the rig. Oscillation can easily be detected by placing markers at several positions of the rig, as the alterations in relative positions accurately reveal this phenomenon. Another issue posed by the collective weight of the cameras is the stress on the rig. Frequent rig assessment may be necessary, as structural failure can lead to immense damage (i.e., the cameras fall off the rig and break).

### 2.6 Usability, calibration and blur

Once the rig is complete, its setup should of course consider the comfort of the user (in order to achieve realistic usage), while keeping in mind the aforementioned requirements. Fatigue can be measured subjectively, but objectively as well by measuring the orientation accuracy over time when following a moving object. The stability of the camera itself can be measured by using a calibration image (e.g., chessboard) as the recorded object. A static chessboard is also suitable to measure camera position accuracy within the rig (i.e., detect whether any camera is dislocated). It is important to avoid motion blur during the movement of both the camera and the object. This can be achieved by using global shutter. Furthermore, synchronization is evidently of key importance, thus the entire system requires a centralized clock, and recording is performed via a centralized trigger.

### 2.7 Correct time stamps

If we use optical markers, a more-or-less trivial requirement is to have faster supporting cameras (pointing towards the markers) than the camera(s) on the rig. At the time of recording with the rig, a precise optical marker measurement is already required. Such cameras should support at least 120 FPS, but 240 FPS is preferable. It is also imperative to have highly precise time synchronization, in order to have sufficiently accurate time stamps. In case of magnetic and radio triggers, time stamp issues are less relevant, as measurements are more frequent.

## 2.8 Rig position in the virtual world

If we measure with both the rig and the markers, then the rig should be registered in the real world. This means that a suitable absolute (i.e., not relative) starting position should be assigned to the rig, and subsequent positions shall be relative to this. By doing so, it becomes possible to portray both the rig and the object in the virtual world, opening new venues of experimental analysis.

## 2.9 Camera rig weight and size (assessment)

Cameras come in different shapes and sizes, or rather weights and sizes, so measurements should be performed by using a variety of cameras. This may contribute towards the optimization of hand-held light field cameras, with regard to weight and size, while avoiding critical issues, such as oscillation. Industrial cameras are typically 4-centimeter cubes, but using sport and mobile cameras are expected to be more preferable, as they tend to be smaller and lighter. Eventually, as quality is of the essence, cinematic use cases shall demand more professional cameras. Keep in mind that the previously mentioned considerations regarding synchronization and global shutter still apply.

## 2.10 Representations on the display

In order to visualize the captured results on a light field display, virtual worlds need to be rendered<sup>11</sup> until the appropriate light field camera becomes available, which will use the weight distribution defined during the measurements. In the virtual world, the camera rig and the object are commonly represented by plain boxes. More complex representations are used for Quality of Experience (QoE) studies, such as perceptually tolerable camera movements.

## 2.11 Conversion

Measuring rig position and orientation and capturing actual content are vastly different. In case of the first experimental method, camera positions during rendering move together, for which the Region of Interest (ROI) matrix is used, and the defined virtual space is moved accordingly. In case of the latter, we may actually use all the marker data, so the position and orientation of individual cameras may be considered during conversion. For this, one-by-one preprocessing of the individual cameras is required for appropriate conversion, as issues of rig rigidity may potentially alter calibration. If this is not the case (cameras are guaranteed to avoid any level of dislocation), then visualization is rather analogous to the aforementioned telepresence system.

## 2.12 Final QoE assessments

Once the entire solution is available, final QoE assessments are highly recommended. These are to be carried out with the inclusion of naive (i.e., not expert) individuals as test participants. The different types of camera movements should be evaluated separately. Such assessments should measure perceptual thresholds of acceptance, tolerance regarding movement, rotation and angular velocity and acceleration. Interaction on its own also contains rather interesting research questions. In case of a large-scale light field display, such as the HoloVizio C80 light field cinema system,<sup>12</sup> certain motion parameters may invite user movements (e.g., moving around within the valid field of view, changing the angle and distance of observation), while others may evoke a more static, theatrical behavior (i.e., the users view the content from a given position). Blur is also to be addressed, especially since certain parts of light field visualization may be located outside the sharp domain of the screen.

## 2.13 Jitter

The measurements evidently lead to the collection of data that contains the jitter. Jitter itself can be determined by the performing a comparative analysis between the theoretical camera movement (i.e., ideal movement, since it is not affected by any jitter or tremor) and the recorded camera movement. Via a series of QoE experiments on subjective threshold, tolerance levels can be determined. These levels are then applied to the raw data, the result of which is used to create functions (e.g., on distribution) of the 6 DOF motion. The data obtained through this process is the desired outcome that is utilized in the virtual world. Existing research<sup>13</sup> and commercial hand-held camera motion databases<sup>14</sup> containing raw or hand-tuned camera motion data can be examined for additional input, to further improve the mathematical representation of the jitter function. Furthermore, jitter can also be generated artificially, automatically for experimental purposes.<sup>15</sup>

## 2.14 Display size

While the HoloVizio C80 is a large display, and in fact, the largest light field display at the time of this paper, initial research shall definitely target smaller displays, especially when using hand-held cameras. Based on the state-of-the-art light field technology, only static rigs are appropriate for large displays, since camera oscillation cannot be completely avoided, and therefore, the recorded visual information becomes practically unusable. For smaller light field displays, and for content relatively close to the camera, the rig size should not exceed 500 millimeters.

## 3. CONCLUSION

In this paper, we have systematically presented the exhaustive list of scientific considerations regarding the simulation of hand-held cameras in light field rendering. There are multiple valid experimental methodologies to choose from, such as the different types of markers, and all of them can be attributed their specific advantages and disadvantages. Following the guideline, the framework for such research enables resource-efficient and reliable solutions, leading to valuable scientific contributions and pragmatic optimization.

## Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813170. Also received funding by 2018-2.1.3-EUREKA-2018-00007 and KFI 16-1-2017-0015, NRDI Fund, Hungary.

## REFERENCES

- [1] Cserkaszkzy, A., Kara, P. A., Tamboli, R. R., Barsi, A., Martini, M. G., and Balogh, T., “Light-field capture and display systems: limitations, challenges, and potentials,” in [*Novel Optical Systems Design and Optimization XXI*], International Society for Optics and Photonics (2018).
- [2] Balogh, T., “The holovizio system,” in [*Stereoscopic Displays and Virtual Reality Systems XIII*], International Society for Optics and Photonics (2006).
- [3] Kara, P. A., Tamboli, R. R., Doronin, O., Cserkaszkzy, A., Barsi, A., Nagy, Z., Martini, M. G., and Simon, A., “The key performance indicators of projection-based light field visualization,” *Journal of Information Display* (2019).
- [4] Lippmann, G., “Epreuves réversibles photographiques intégrales,” *Comptes-Rendus Academie des Sciences* **146**, 446–451 (1908).
- [5] Maida, M., Didier, J.-Y., Ababsa, F., and Mallem, M., “A performance study for camera pose estimation using visual marker based tracking,” *Machine Vision and Applications* **21**(3), 365–376 (2010).
- [6] Huang, J., Mori, T., Takashima, K., Hashi, S., and Kitamura, Y., “IM6D: magnetic tracking system with 6-DOF passive markers for dexterous 3D interaction and motion,” *ACM Transactions on Graphics (TOG)* **34**(6), 1–10 (2015).
- [7] Wagner, D., Langlotz, T., and Schmalstieg, D., “Robust and unobtrusive marker tracking on mobile phones,” in [*2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*], 121–124, IEEE (2008).
- [8] Cserkaszkzy, A., Barsi, A., Nagy, Z., Puhr, G., Balogh, T., and Kara, P. A., “Real-time light-field 3D telepresence,” in [*7th European Workshop on Visual Information Processing (EUVIP)*], 1–5, IEEE (2018).
- [9] Ziegler, M., op het Veld, R., Keinert, J., and Zilly, F., “Acquisition system for dense lightfield of large scenes,” in [*2017 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4, IEEE (2017).
- [10] Raytrix.  
<https://raytrix.de/>.
- [11] Balázs, Á., Barsi, A., Kovács, P. T., and Balogh, T., “Towards mixed reality applications on light-field displays,” in [*2014 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*], 1–4, IEEE (2014).
- [12] Holografika, “HoloVizio C80 glasses-free 3D cinema system.”  
<https://holografika.com/c80-glasses-free-3d-cinema/>.

- [13] Kurz, D., Meier, P. G., Plopski, A., and Klinker, G., “Absolute spatial context-aware visual feature descriptors for outdoor handheld camera localization overcoming visual repetitiveness in urban environments,” in [*2014 International conference on computer vision theory and applications (VISAPP)*], **2**, 56-67, IEEE (2014).
- [14] Unreal Engine, “Cinemotion 2 Real Handheld Camera Motion Kit.”  
<https://www.unrealengine.com/marketplace/en-US/product/cinemotion-real-handheld-camera-motion-kit>.
- [15] Kurz, C., Ritschel, T., Eisemann, E., Thormählen, T., and Seidel, H.-P., “Generating realistic camera shake for virtual scenes,” *JVRB-Journal of Virtual Reality and Broadcasting* **10**(7) (2014).