K. Egiazarian, A. Beghdadi, I. Tabus, C. Larabi, F. Battisti and L. Oudre, (eds.) (2018) Proceedings of the 2018 7th European Workshop on Visual Information Processing (EUVIP). IEEE, ISSN 2471-8963 ISBN 9781538668979.

https://doi.org/10.1109/EUVIP.2018.8611663

# Real-time light-field 3D telepresence

Aron Cserkaszky, Attila Barsi, Zsolt Nagy, Gabor Puhr, Tibor Balogh, Peter A. Kara
*Holografika Ltd.*
{a.cserkaszky, a.barsi, zs.nagy, g.puhr, t.balogh, p.kara}@holografika.com

***Abstract*—Light-field technology is often looked at as the final frontier of glasses-free 3D visualization, as no additional viewing gear is required to experience its capabilities to their full extent. Among the numerous industrial and commercial use cases, light-field telepresence stands out, as such natural visualization may significantly boost the sense of presence. In this paper, we present a fully-implemented real-time light-field 3D telepresence system. We provide a comprehensive analysis of the implementation of the one-way system, highlighting how the achieved capabilities satisfy the reasonable requirements towards such system. The paper also discusses future enhancements to the 3D telepresence system, since its true potential is yet to be fulfilled.**

***Index Terms*—3D telepresence, light-field streaming, 3D display, light-field camera**

## I. Introduction

The research on light-field capture and visualization is one of the most promising and rewarding scientific efforts in the area of 3D visualization technologies at the time of this paper. Promising, as there is a long list of potential future industrial and commercial applications, and the already existing systems have managed to deliver, to live up to the expectations. Rewarding, as the successful implementations of light-field systems have proven their contributions, their real value for the world of true glasses-free 3D imaging, and new systems are expected to exceed their predecessors in the aspects of efficient usage and personal 3D experience.

Such new industrial applications include 3D design of any scale, medical diagnostics, traffic and process control, gas and oil exploration [1] and many more. Similarly to most novel technologies, military-grade applications are considered as well, such as land, sea or air vehicle control systems, or battlefield modeling. For everyday commercial usage, light-field technology offers multiple forms of large-scale and home entertainment, like light-field cinema [2], home multimedia recording (i.e., by hand-held devices, see Section II) and consumption (e.g., video streaming [3] on a TV-like light-field display such as the HoloVizio 80WLT [4]) or gaming.

Some of these applications do not pose reasonable time constraints at all, while others are delay-sensitive real-time utilizations of light-field technology. The two most common real-time applications today on conventional 2D devices are multimedia streaming — which can either be Video-on-Demand (VoD) streaming or live streaming — and two-way (or more than two-way) multi-party communication. As only genuinely time-critical applications are considered, real-time text message exchange services are excluded from the latter. Multi-party communication happens between two or more individuals, in voice-only (analogous to conventional phone calls), video-only (e.g., web camera session without sound; rarely the case today) and audiovisual forms.

The ultimate goal of such communication is the so-called "sense of presence". In the aforementioned forms of communication, it refers to the ability to experience the virtual presence of another individual. Evidently, this term can be separated into an audio and a visual dimension. With light-field technology, the visual dimension can benefit from the new frontier of glasses-free 3D, as such systems provide smooth motion parallax. Yet true sense of presence also requires life-like sizes, proportions, in order for the visualization of the individual on the other end of the communication session to feel less like an arbitrary digital interpretation and more like the actual presence of the remotely located individual. Systems that aim to achieve an extent of the sense of presence are commonly known as telepresence systems, including all previously mentioned solutions that replicate audio, video, or both, as any of those can indeed provide a given sense of presence.

In this paper, we introduce the design and implementation of a novel one-way real-time light-field 3D telepresence system. A one-way system in this context means that live images of an individual are captured and at a remote location are displayed real-time, but this individual is not provided a symmetrical visual feedback from the other end of the system. Our fully-implemented solution is a true-to-scale system, visualizing the individual in a near-life-like size. The paper details the scientific decisions, design, implementation, verification and validation of the 3D telepresence system. It is important to note that certain details regarding the implemented system are protected intellectual property, and cannot be disclosed in this paper.

The remainder of the paper is structured as follows: Section II overviews the related research in the area, including capture and display systems. The requirements that were to be satisfied are detailed in Section III. The implemented telepresence system is introduced in Section IV. Further discussion on challenges, optimization, scalability and system performance is provided in Section V. The paper is concluded in Section VI, also pointing out potential future scientific contributions.

The research in this paper was done as a part of and was funded from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 676401, ETN-FPI, and the EUREKA_15-1-2016-0003 LiveRay project supported by the NRDI Fund of Hungary.

## II. Related Research

In the general sense, the plenoptic function [5] aims to describe the totality of light rays in a scene. It is a 7D function that shows the intensity of light from every position (3 spatial dimensions) to every direction (2 angular dimensions) in every wavelength (1 wavelength dimension) in each moment (1 temporal dimension). Even though the plenoptic function omits the phase and polarization of the light rays, it is seldom used in rendering or visualization due the complexities caused by its unnecessary generality.

The more practical approach is the concept of light-field, which in the modern interpretation means a four-dimensional space that describes rays of light emitted by a convex surface [6]. The temporal dimension is left out and the continuous wavelength description is supplanted by separate light-fields for the three primary colors (red, green and blue). The definition of light-field can be used to create a "window of light" that enables the viewer to experience a 3D window to a distant or virtual place. Important properties of this light-field is the baseline, which is basically the length of the diagonal of the so-called "window of light", representing the maximum spatial relocation of the observer's perspective within the light-field. We define a light-field as having wide baseline if the baseline is more than 1 meter, and short baseline if it is in the order of centimeter. Typical wide-baseline systems compromise on the vertical baseline in favor of the horizontal baseline to reduce data and hardware requirements. Light-field systems that allow no change of perspective in the vertical direction are called horizontal-parallax-only (HPO), as they reduce the 4D light-field to a 3D subset. In this context, the 4D light-field systems are called full-parallax (FP) as they have baselines in both directions. Using HPO systems have only few disadvantages when the viewers of the light-field mostly move horizontally, as humans typically do.

Light-fields are recorded with light-field camera systems. Currently there are already short-baseline FP camera systems on the market by Lytro and Raytrix, that are based on placing a microlense array in front of a single imaging sensor [7]. Many have built linear-camera-array-based capture systems initially for the purpose of capturing so-called "bullet time effects", and lately for capturing wide-baseline HPO light-fields. Researchers have also built wide-baseline FP capture systems for experimental purposes [8]. There has also been commercial attempts to create a wide-baseline FP capture system in the form of Lytro Immerge with questionable success, as the company that produced it is in the process of shutting down at the time of this paper.

Light-field displays already exist as well, for the purpose of showing a light-field to the viewer enabling a visual experience that is much closer to reality than conventional 2D displays. Analogously to light-field cameras, there are generally short-baseline FP displays and wide-baseline HPO display systems. The former ones are nowadays used in head-mounted near-eye light-field displays [9], that might soon enter the consumer market, with the primary goal of solving the vergence-accommodation conflict. The latter ones use arrays of projectors to enable glasses-free 3D visualization for groups of viewers [10] and are already present on the market [11]. There are also wide-baseline FP displays that compromise in the temporal domain, since they use analog film as the basis of the light-field with microlense arrays providing the directional rays [12].

Complex light-field systems would need to combine light-field capture and display systems. There have already been attempts to create such systems by extending light-field camera systems with streaming and dynamic viewing capabilities on conventional 2D displays [13] [14]. Likewise, light-field display systems have been extended to receive streamed light-field content and visualize it in real time [15] [16]. Furthermore, complex systems with symmetric capture and display light-fields have been pioneered [17], that do not need to convert the light-fields but also suffer limitations from this property. These precursors have paved the way to combine integrated large-scale low-latency light-field telepresence systems.

## III. Requirements

In order for such system to serve its purpose and be used efficiently, there are several requirements to be satisfied simultaneously. One of the most critical requirements is a low and reliable system latency. Of course network latency is vital during practical use as well, but in the scope of this paper, we only discuss system delay. This requirement is analogous to what applies to videoconferencing services on 2D displays. If not fulfilled, audio/video synchronization may be compromised, unnatural pauses in visualization continuity may occur and the overall user experience may be degraded.

Another absolute must is a sufficiently high resolution. Here resolution refers to both spatial and angular resolution. Low spatial resolution may result in certain extents of blur which not only distorts the visual experience but may also make the extraction of necessary visual information (i.e., the facial expression of the individual) difficult or impossible. Insufficient angular resolution may be even worse, as such can result in the disturbance of the horizontal motion parallax. Visual phenomena in such scenario includes the crosstalk effect, and also sudden view jumps, which is a lethal nemesis of glasses-free 3D visualization. Yet going towards high-end extremes should be avoided as well, as the total system latency is also affected by the computational requirements and the bandwidth usage of the given system.

The frame rate requirement of such visualization can be taken into account similarly to 2D counterparts. At least 20-25 FPS is needed in order to preserve relatively natural human motions. High FPS values should be avoided, due to the same reasons regarding system latency as in case of resolution.

As mentioned earlier, the sense of presence also necessitates true-to-scale visualization that can give the impression of a life-size human being. This means that the screen of the system should be at least 170-180 cm tall. However, this also means that that screen does not need to be taller than

a given value, approximately 200-210 cm. There is of course a reasonable requirement for width as well, since a sufficiently tall screen with an irrationally small width of 20-30 cm would not necessarily be properly usable. Therefore, such screen should be at least roughly 70-80 cm wide, and does not need to be wider than 120 cm.

Light-field systems are typically designed for multi-user scenarios, as the multi-viewer capability is one of the greatest strengths of this visualization technology. Indeed, within the valid Field of View (FOV) of the display, there is no theoretical limitation of the number of simultaneous viewers. The limitation is only practical, determined by the size of the FOV. Trivially, the bigger the FOV, the more viewers it can accommodate. The FOV also defines the sideways mobility of a single user. In a very limited FOV (e.g., 15-20 degrees), one would need to observe the screen directly from the middle, as leaving the valid FOV would hinder the appropriate 3D view. Therefore, such telepresence system needs to comfortably accommodate multiple users and enable viewer movement — which is actually a fundamental part of the overall experience — thus the FOV of the display should be at least 100-120 degrees.

## IV. Implementation and Capabilities

According to the best knowledge of the authors, this paper presents the first fully-implemented real-time light-field 3D telepresence system. The system is comprised of the capture and display subsystem with a general purpose network in between them. Fig. 1 shows the arrangement of the subsystems and the main components.

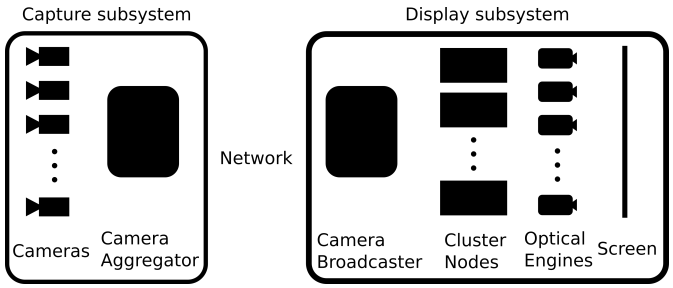


Fig. 1. The main components of the telepresence system.

The light-field capture subsystem uses 96 cameras. These cameras are arranged on an arc of 120 degrees with approximately 2.7 meters from the center of the arc. A photograph of the capture subsystem is shown on Fig. 2. The cameras capture images of $1280 \times 1024$ pixels that are output on their Gigabit Ethernet interface. The full frames of the cameras are collected by the camera aggregator and then are streamed over the IP network towards the display subsystem. The frame rate of the cameras are set to 25 FPS in the current implementation. The cameras are synchronized externally to a common clock pulse to ensure that frames are taken at the same time and the light-field created from them would not suffer temporal distortions. Another important property of the cameras is the global shutter. This helps to minimize any problematic visual artifacts created by spatial and temporal aliasing of moving objects.

Fig. 2. The cameras of the capture system arranged on an arc.

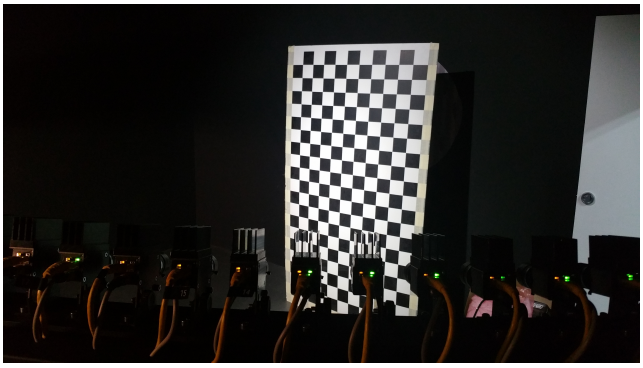

Fig. 3. Calibration of the light-field camera system.

The exact light-field structure of the capture subsystem cannot be computed from the geometric model of the system, as the real implementation of the system will always differ from the planned design. The light-field structure is particularly sensitive for the orientation and lens distortions of the cameras. We used the standard checkerboard based calibration method for the camera array as seen on Fig. 3 with algorithms based on Zhang [18]. The calibration parameters are transmitted offline to the display subsystem and are used to interpret the camera images as a light-field.

On the display subsystem side, the light-field of the camera images during a telepresence session are distributed by the camera broadcaster, then are converted by the computer clusters to the display light-field and are projected onto the holographic screen with the optical engines. The array of optical engines have a combined resolution of 100 megarays. The images of the optical engines are computed on consumer-grade Nvidia GPUs, housed in cluster node computers. Each optical engine projects a specific subset of the light-field and needs the camera images that cover this subset. The system exploits the fact that optical engines connected to a single cluster node project neighboring subsets of the light-field. The main purpose of the camera broadcaster is to select and cut

the sections of the camera images that are required during the conversion by the nodes and broadcast these sections to the respective nodes. On the nodes the image sections are converted by a look-up-table to the display light-field, which practically means that the color of the rays emitted by the optical engines is computed from adjacent capture light-field rays in the 4D light-field space. The look-up-table is computed from the combination of the calibration data of the display and capture subsystem. The calibration method of the display subsystem is proprietary, and is based on the optical engines projecting calibration patterns on the screen, while a camera at a fixed position is observing the resulting visual patterns.

Fig. 4. The light-field display system with synthetic image.

Fig. 5. A live telepresence photographed from different angles.

The holographic screen of the display subsystem is 100 cm wide and 180 cm tall, in order to visualize the entire body of a human being in near full scale. An example virtual scene is shown on Fig. 4, illustrating the operation of the light-field display. The arrangement of the optical engines behind the screen provide a full-angle 180-degree FOV. As the capture subsystem has a smaller FOV than the display subsystem, the extra viewing angles show the nearest 2D angular view of the real scene during a telepresence session. The 2D equivalent resolution of the system is $720 \times 1280$ pixels, and the angular resolution in the center is 0.9 degree, based on the amount and distribution of the emitted display light-field rays.

The current implementation of the telepresence system uses a high-bandwidth optical Local Area Network (LAN) between the capture and display subsystem, to minimize latency. The resulting total system latency — measured between the cameras recording a frame and the appearance of the light-field of this frame on the screen — is approximately 100 ms. Also, we decided to use LAN due to the fact that the subsystems are physically located near each other.

Fig. 5 shows the display subsystem in operation during a live telepresence. This was a part of the exhaustive verification and validation process of the system before public deployment. The implementation was tested with multiple different individuals at both subsystems, and simultaneous viewers at the display subsystem, focusing on the multi-user experience. Several environmental lighting configurations were used as well, since a successful implementation needs to ensure the desired visual performance in all common deployment scenarios (e.g., dark room with minimal artificial light sources, highly illuminated exhibition area, etc.).

Fig. 6. The system in operation at the Information Communication Technology center of SK Telecom in Seoul, South Korea.

The fully-implemented system is exhibited at the Information Communication Technology (ICT) Experience Center "T.um" of SK Telecom in Seoul, South Korea. It is part of the Hologram Conference Room exhibition where visitors

participate in resolving global problem scenarios with the help of modern technology. Fig. 6 captures one of the first utilizations of the deployed system on site.

## V. Discussion

One obvious limitation of the current implementation is the reliance on a high-bandwidth LAN. This could be easily addressed, with adding stream-based video compression/decompression methods, reducing the bandwidth requirement significantly. The compression and decompression would add an estimated 30-40 ms of latency on top of the inherent latency of the network. The additional computing requirements would necessitate adding processing nodes on the capture subsystem side and potentially increasing the capacity of the nodes on the display subsystem side.

Another shortcoming of the current system is the relatively limited frame rate of 25. This limitation is caused by the sole unit of camera broadcaster, that handles the images of all cameras. A partitioned and modular system with multiple camera broadcasters handling portions of the cameras would increase the frame rate potentially to the maximum allowed by the cameras.

A corollary to having multiple camera broadcasters would be the property of modularity. This could be helped by redesigning the entire system to be modular. For example, the camera arc could be sectioned in a way that enables flexible arrangement of the cameras based on specific FOV and angular resolution requirements of the served use-case. Likewise, the display subsystem could also use the same technique to make the display light-field configurable.

Going further, the system could be extended to provide a multi-way telepresence by integrating the capture and display subsystem together. The camera arc in this case could be placed just above the top of the display screen, which would induce a minimal upward shift of perspective from the ideal eye-level height of the camera arc. This could be mitigated by advanced gaze correction techniques [19] or view synthesis [20].

The limitation of horizontal-parallax-only light-field is much more difficult to address on both the capture and the display side. A full-parallax telepresence system would require 2D arrays of both cameras and optical engines with a likewise increased computing and bandwidth capacity. Currently this is beyond the practical and economical limits of the present day, and will probably also be in the near future.

## VI. Conclusion

In this paper we presented a fully-implemented real-time light-field 3D telepresence system. Even as a one-way solution, it already provides the highest level of the sense of presence that can be achieved today via the smooth horizontal motion parallax and the near-true-to-scale visualization. Its future optimization aims to further enhance the overall user experience by adjusting parameters, such as frame rate, to provide a more life-like visual sensation, and by supporting modularity, thus paving the way for broader utilization.

## References


[1] T. Balogh and P. T. Kovács, "Holovizio: The next generation of 3D oil & gas visualization," in *70th EAGE Conference and Exhibition-Workshops and Fieldtrips*, 2008.

[2] P. A. Kara, Z. Nagy, M. G. Martini, and A. Barsi, "Cinema as large as life: Large-scale light field cinema system," in *2017 International Conference on 3D Immersion (IC3D)*. IEEE, 2017.

[3] P. A. Kara, A. Cserkaszky, M. G. Martini, A. Barsi, B. Laszlo, and T. Balogh, "Evaluation of the Concept of Dynamic Adaptive Streaming of Light Field Video," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, 2018.

[4] Holografika, "Holovizio 80WLT light-field display," http://holografika.com/80wlt/ (retrieved May 2018).

[5] E. H. Adelson, J. R. Bergen *et al.*, "The plenoptic function and the elements of early vision," 1991.

[6] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[8] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 765–776.

[9] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, 2013.

[10] J.-H. Lee, J. Park, D. Nam, S. Y. Choi, D.-S. Park, and C. Y. Kim, "Optimal projector configuration design for 300-Mpixel multi-projection 3D display," *Optics express*, vol. 21, no. 22, pp. 26 820–26 835, 2013.

[11] T. Balogh, "The Holovizio system," in *Stereoscopic Displays and Virtual Reality Systems XIII*, vol. 6055. International Society for Optics and Photonics, 2006.

[12] D. Jung and R. Koch, "Efficient rendering of light field images," in *Video Processing and Computational Video*. Springer, 2011, pp. 184–211.

[13] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera." *Rendering Techniques*, vol. 2002, pp. 77–86, 2002.

[14] Y. Liu, Q. Dai, and W. Xu, "A real time interactive dynamic light field transmission system," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2173–2176.

[15] T. Balogh and P. T. Kovács, "Real-time 3D light field transmission," in *Real-Time Image and Video Processing 2010*, vol. 7724. International Society for Optics and Photonics, 2010.

[16] P. T. Kovács, A. Zare, T. Balogh, R. Bregović, and A. Gotchev, "Architectures and codecs for real-time light field streaming," *Journal of Imaging Science and Technology*, vol. 61, no. 1, pp. 10 403–1, 2017.

[17] X. Sang, F. C. Fan, C. Jiang, S. Choi, W. Dou, C. Yu, and D. Xu, "Demonstration of a large-size real-time full-color three-dimensional display," *Optics letters*, vol. 34, no. 24, pp. 3803–3805, 2009.

[18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[19] C. Kuster, T. Popa, J.-C. Bazin, C. Gotsman, and M. Gross, "Gaze correction for home video conferencing," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, 2012.

[20] D. Tian, P.-L. Lai, P. Lopez, and C. Gomila, "View synthesis techniques for 3D video," in *Applications of Digital Image Processing XXXII*, vol. 7443. International Society for Optics and Photonics, 2009.