The potential synergies of visual scene reconstruction and medical image reconstruction

Aron Cserkaszky[a], Peter A. Kara[b],
Attila Barsi[a], and Maria G. Martini[b]

[a]Holografika, Budapest, Hungary
[b]WMN Research Group, Kingston University, Kingston upon Thames, UK

ABSTRACT

Image reconstruction in nuclear medicine produces valuable volumetric data of vital markers in living bodies. Visual scene reconstruction methods, that aim to recreate a scene from camera images, are also continuously improved by the recent advancements of light-fields and camera systems. The parallels of the two fields are increasingly noticeable as we now have the computing power and methods to take into account transparent materials and ray trace the scattering and other effects of lights for visual scene reconstruction. In this paper, we aim to highlight and analyze the similarities and potential synergies of the two methods.

Keywords: light-field, nuclear medicine, 3D visualization

1. INTRODUCTION

At the time of this paper, methods of visual scene reconstruction and medical image reconstruction develop rapidly. With the significant recent advances in computing power, computationally expensive techniques are now commonly available and highly benefit both fields. Although due to certain differences, they progress a lot independently, yet there is quite some common ground on which synergies could reach their full potentials.

In this paper, we provide an overview of the two fields, and analyze their similarities and synergies. The paper also discusses the practical requirements of future glasses-free 3D visualization of medical data, particularly on light-field displays.

The remainder of this paper is structured as follows: Section 2 overviews the problems of reconstruction. The processes of visual scene reconstruction and medical image reconstruction, together with their relevant properties, are detailed in Section 3 and 4, respectively. The similarities and synergies of these fields are analyzed in Section 5. Section 6 provides a discussion regarding the visualization of medical information. The paper is concluded in Section 7.

2. RECONSTRUCTION PROBLEMS

Reconstruction problems are by their nature inverse problems, where a signal is measured at discreet points after being affected by a system, and the ultimate goal is to recover the original high-resolution signal without the influence of the system. Naturally, depending on the complexity and the influence of the system on the signal, this can be a hard challenge. The number of available measurement points compared to the desired reconstruction quality also modifies the difficulty of the problem. The system — that the signal is in — contains all physical effects related to the propagation of the signal to the measuring device, the process of measurement and various other relevant noise factors.

Further author information: (Send correspondence to Aron Cserkaszky)
Aron Cserkaszky: E-mail: a.cserkaszky@holografika.com
Peter A. Kara: E-mail: p.kara@kingston.ac.uk
Attila Barsi: E-mail: a.barsi@holografika.com
Maria G. Martini: E-mail: m.martini@kingston.ac.uk

A relatively simple example of a reconstruction problem is the deblurring of 2D images. Depending on the magnitude of the blur, often a deconvolution of the blurring kernel results in a good solution. Harder or sparser problems usually require an assumption or prior information about the original signal that regularizes the reconstructed solution. A sparse problem in this context means that we have fewer measurements (known values) than values to be reconstructed (unknowns), making the reconstruction problem undetermined and insolvable without additional assumptions and regularizations. In a severely blurred image, this might mean assuming the feature structures or the present frequencies in the original image.

Truly difficult reconstruction problems require iterative solving methods, where the system is modeled with a system operator, and an iterating algorithm searches for a hypothesized signal, that after applying the system operator produces simulated measurements close enough to the actual measurements. Iterative algorithms have of course their own issues with convergence and stopping criteria. Furthermore, any mismatch between the real behavior of the system and the system operator decreases the stability of the iteration. In most cases, we can assume a linearity of the inverse problem, therefore the system operator is a system matrix, that contains the detection weights for possible signal elements to each measurement. Updating the hypothesized signal according the difference of actual and simulated measurements needs to be domain specific to be tolerant to the most relevant noises of the system.

3. MEDICAL IMAGE RECONSTRUCTION

Medical image reconstruction aims to reconstruct the distribution of a physical quantity in a volume, that will be used for establishing a clinical diagnosis, therapy monitoring or other medical purposes. In this section, we will consider positron emission tomography (PET), single photon emission tomography (SPECT) and computed tomography (CT), and exclude the modalities of magnetic image resonance (MRI) and ultrasound (US), as those do not require a line-integral-based reconstruction.

The basic idea of CT is that we have a collimated source of X-rays, we direct these rays to pass through a volume with unknown materials inside, and we measure on the other side the amount of X-ray photons that were able to pass through the volume. So along this line of measurement, we get an integrated pass through probability, but have no localized information of the materials along the segments of the line. The X-rays scatter or get absorbed by the electron clouds of atoms, therefore the probability of detection is related to the composition and density of the materials. With sufficient amount of lines of measurement the volume of materials can be reconstructed, yielding a useful anatomical map of a patient with sub-millimeter resolution for example.

SPECT is a technique when a radioactive isotope bound to some molecules of interest is distributed in a volume, and the that gamma rays emitted by the isotope are measured with a collimated detector, that ensures that the gamma ray came from a certain direction. The number of detection on each line of measurement is proportional to the amount of the isotope along the line. For precise reconstruction, the gamma scattering effects have to be taken into account both in the detector collimator and the volume — preferably with advance knowledge of the distribution of materials from a CT scan. The reconstructed volumetric image shows the distribution of the marked molecule and can supply valuable information for the diagnoses of a range of diseases.

PET is similar to SPECT, but in this case the radioactive isotope is not emitting a single gamma ray, but a positron, the anti-particle of electron. Once the positron meets an electron, the matter anti-matter annihilation occurs, that creates two high-energy gamma rays that fly out in opposite directions from the annihilation point. This creates a unique opportunity, since if the detectors at the opposite sides of the volume detect gamma rays at the same time, there is a good chance that they came from the same annihilation event on a line between the two detection points. Therefore, collimation of the rays is inherently accomplished by this measurement technique, allowing higher detection rates.

These three medical imaging techniques share the common feature of that they primarily measure a quantity along lines and air has negligible effect on the rays. Due to the similarity of the inverse problem, their reconstruction techniques are likewise similar. Initial reconstruction methods focused on 2D slices of the volume of interest, and were using the sinogram to visualize the measured data for individual slices, as shown of Fig. 1*. A sinogram simply is a parametrization of the measured lines in the 2D slice, where each line gets assigned two coordinates:

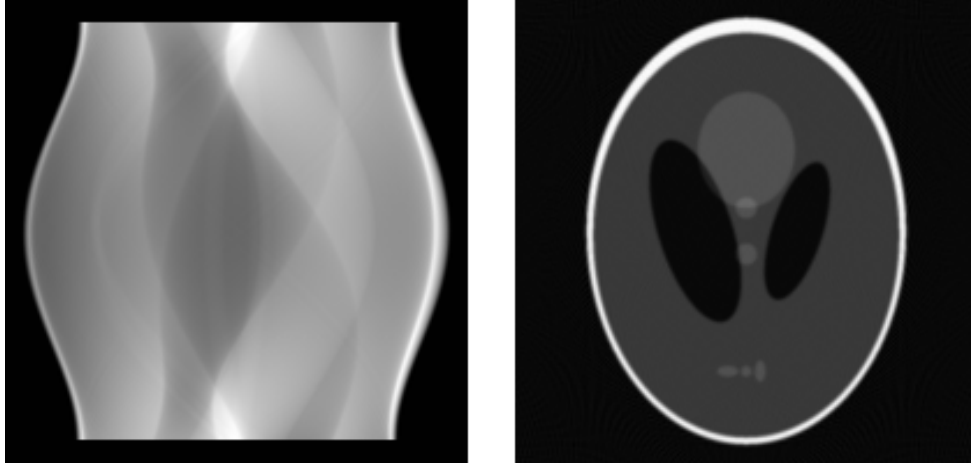*https://tomroelandts.com/articles/tomography-part-3-reconstruction

Figure 1. Example of a 2D sinogram (left) of the famous Shepp-Logan phantom (right).

the distance from the center of the slice, and the intercept angle with the x-axis. Early reconstruction methods were based on projecting the measured counts along the measurement lines into the slices of the volume with an initial filtering of the data — a technique known as Filtered Back Projection[1].[2] Then the reconstructed slices were assembled into the volumetric result.

Modern high-quality and inherently 3D techniques use iterative reconstructions like the maximum likelihood estimation maximization (ML-EM).[3] In this case, the all relevant physical effects can be modeled, which became only possible with the computing power of modern computers. To date, Monte-Carlo simulations have been the most precise ways to simulate the full system operator of medical imaging systems.[4] In a fully 3D case, the measurement lines have 4 dimensions, so there is no easy visualization of the data, as a 4D sinogram[5] would not be meaningful for humans.

4. VISUAL SCENE RECONSTRUCTION

Visual scene reconstruction is the process of building a 3D understanding of a scene from visual information — typically ordinary camera images. A simple scene consisting of matte everyday objects can be easily imagined in 3D by a human observer, just by looking at one or maybe a few photographs of the scene. This happens because the observer uses a lot of assumptions of the objects and the scene, and there are no shining, reflecting or other visual illusions inducing tricky effects in the scene. However, the goal is to algorithmically solve this reconstruction problem, even for scenes including difficult effects.

Additionally to the visible light information captured by normal cameras, other modalities of light might also be used to augment and improve the quality of the scene reconstruction. Infrared (IR) light is used often when extra information is needed from the scene without disturbing the visual color information. This can mean just another camera image from a different wavelength interval or IR, with additional useful image features to be processed. It can also be a structured IR light like with the relatively popular Kinect v1 sensors,[6] that emit an IR pattern and a low-resolution depth of the scene is computed by the warping of the pattern. Another IR modality often used is the time-of-flight (ToF) camera concept, where an IR pulse lights up the scene, and by the response time delay of various parts of the scene the depth can be estimated by using the speed of light in air. The Kinect v2 and many other color-plus-depth cameras employ this concept, as well as the LIDAR (light-radar) systems.

However, the normal color camera based scene reconstruction methods[78] can only rely on the consistency that if intersecting camera rays show similar color value, then there is a probability that an object of that color is at the intersection of the rays. There are many corollaries of this statement. First of all, the camera system has to be designed in a way that rays of individual cameras intersect often and preferably everywhere, in the volume encompassing the scene that is to be reconstructed. The second corollary is that the better this assumption holds — that an objects color is the same regardless of viewing direction — the higher the quality of the reconstructed
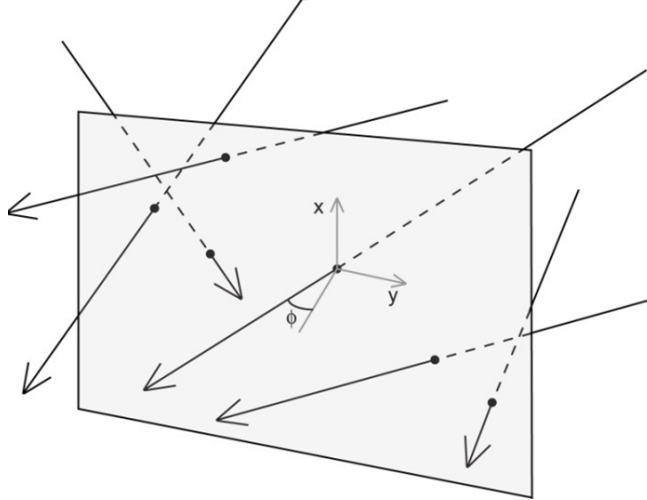
Figure 2. Illustration of the light-field describing each ray with spatial coordinates of the intersection, and angular coordinates of the direction.

result can be. In the context of the inverse reconstruction problem, the known measurements are the color values measured by each pixel of the camera system, and the unknowns are the lack or the color of the objects of each point in the scene with a certain resolution.

The light-field[9] concept is a very useful tool to organize the rays of scene capture systems. The light-field is a four-dimensional ray space that describes the color of rays that a convex surface emits outwards, Fig. 2 shows and example parametrization. Air can be regarded as a non participating media with regard to the visible spectrum, so the position of the cameras is not relevant, and the reconstruction algorithms can work efficiently with just the rays in the 4D light-field. One use case for light-fields is that in this general ray representation the dense light-field can be reconstructed[10] from sparse samples, and the dense reconstructed light-field can then be used for high-quality full scene reconstruction.[11]

In the future, we can not rule out complicated or peculiar scene reconstruction problems, where the exact wavelength, the polarization, the phase or the timing of the detected visual photons might be essential for a high-quality reconstruction result. These properties are not described by the 4D light-field, so additional dimensions will have to be taken into account.

Iterative techniques have also been used for scene reconstruction,[12] however, to date, no published algorithm exist that could do an iterative reconstruction of a natural 3D scene modeling all non-lambertian effects in a scene with high-quality result. In this regard, scene reconstruction is yet to reach the level of medical reconstruction techniques.

5. SIMILARITIES AND SYNERGIES

The fields of medical image reconstruction and scene reconstruction have many features in common. First and foremost, both reconstruct the 3D distribution of properties in a volume, by detecting and measuring the energy of photons exiting the volume. In some cases, the initial photons are emitted by controlled of semi-controlled sources, like in CT or when a scene is illuminated by lights. In other cases, the distribution of the photon sources is the main goal of the reconstruction. The correct estimation of the paths of photons in the volume of interest is essential in gaining information out of the detected events. In the case of scene reconstruction and SPECT, this is primarily accomplished by elements directing the photons into the detector: optical elements or collimators. In case of CT, the direction is determined by controlling the source, and for PET, the coincidence measurement supplies this information. It is important to note the difference of photon energies though, which is directly related to the fact that while visual photons with around 1 eV energy can be redirected by conventional optical lenses, the gamma and X-rays of medical imaging are in the range of 100–1000 keV, and can only be collimated by the absorbing materials.

The second main similarity is that both techniques operate on the inherently 4D data structure of lines described in a 4D space. For visual scenes, this is called light-field, and for medical images, it is the 4D sinogram. This primary data structure is almost interchangeable between the two methods, though the different effects governing the photon transport — like scattering — would disallow the direct interchange of the reconstruction algorithms, if those effects are not negligible in the dataset.

Ray-tracing is a technique for generating realistic camera images from virtual scenes, often used in movies. In forward ray-tracing, the simulation of photons starts at the light sources and relevant reflection and non-lambertian effects are modeled. The ray-tracing technique is analogous to initial methods of simulating the detector result for medical image reconstruction. These ray-tracer-based methods have been surpassed by fully physical simulators for both applications. Employing essentially the same Monte-Carlo techniques, when the photon transport is simulated with all relevant physical effects without bias or restrictions, for both problems yield the most realistic results. Naturally, this is also the most computationally expensive, and somewhat wasteful, since as in reality, only a fraction of all photons reach the camera or the detector. For physical simulators, we can also observer the similarity between the complex energy-specific Compton scattering[13] models for gamma rays, and the bidirectional reflectance distribution function[14] that models the visual reflectance of materials. We hope that soon, iterative reconstruction techniques will also be standard for scene reconstruction, to achieve the highest quality results from the available visual information.

6. DISCUSSION ON VISUALIZATION

In this section, we discuss the practical properties and attributes of visualization for the relevant medical use cases. The aim of introducing the glasses-free 3D of light-field systems is to further increase the current levels of diagnostics accuracy[15] and surgical precision.[16] At the time of this paper, 3D medical content is already frequently used in practice and their contributions to the work of medical experts is evident, yet the majority of such content is still visualized on conventional 2D screens, creating an inherent set of limitations. There are of course 3D viewing solutions in use as well, which necessitate 3D viewing gears, i.e., 3D glasses and head gears. When light-field technology is utilized for this purpose, no such equipment is needed, hence the name "glasses-free 3D". It needs to be noted that light-field displays are not only potentially beneficial for general 3D medical image visualization, but also for 4D sinograms. In the scope of our analysis, we mainly consider horizontal-parallax-only (HPO) visualization, and do not directly address full-parallax (FP) displays.

There are certain requirements that are vital to be met in order for such displays to fully benefit the medical sector. The first property is the field of view (FOV), which determines the angle of the area in which light-field visualization can be validly observed. While in case of stereoscopic 3D, the total number of simultaneous viewers is determined by the number of glasses that connect to the system, light-field displays can accommodate any given number of viewers that can physically fit inside the valid FOV. Prior research indicates that the human visual system (HVS) of the viewers can benefit up to 135 degrees of FOV, but commonly fails to differentiate above that.[17] On that other hand, having an insufficiently small FOV not only limits the number of simultaneous viewers, but the extent of supported user mobility as well. Therefore, such system should have a FOV of at least 45 degrees, at most 135 degrees, and ideally should be in the interval between 90 and 120 degrees.

Possibly the most important property of such system is the resolution. We can differentiate between display and content resolution, and between spatial and angular resolution. For system requirements, the values of display resolution are more relevant; however, at the end of day, content resolution has an equal weight in the eyes of the end user, who are medical experts in this specific scenario.

Generally speaking, the visualization quality of light-field displays has great importance, but in case of medical imaging, it can actually be a "matter of life and death". This directly implies that resolution values cannot be compromised, neither spatial nor angular resolution. Suboptimally low spatial resolution manifests in blur that is not uniform across the volume of visualization. While in multimedia use cases, blurred content is often tolerable[18] and can also be visually appealing particularly for computer-generated models and scenes, it may result in false positive or false negative diagnosis in medical imaging. However, at the time of this publication, according to the best knowledge of the authors, no research reports that light-field content above 1920×1080 spatial resolution could be efficiently distinguished by viewers, thus capturing and displaying in a resolution higher than that might not be necessary.

Angular resolution may even be seen a more critical parameter, as the insufficiency of which disturbs the smoothness of the horizontal motion parallax. Prior works show that undisturbed parallax effect may not be necessarily guaranteed below an angular resolution of 2 source views per degree (or 0.5 degree),[19] which can serve as the lower bound for visualization excellence. Surely, the half of this given angular resolution may prove to be sufficient for general purposes, but as it has already been stated, the use case dictates high-quality visualization.

In case the captured medical content does not have the required angular resolution, estimative techniques such as interpolation can be used to increase the density. Although it has already proven to enhance the perceived quality despite the inaccuracies of synthesis,[20] and the content may become more appealing visually through the increased content contrast,[21] again, it may result in errors in medical diagnostics. Future research on interpolation involving medical experts shall prove to be beneficial in the design and standardization of light-field visualization for medical use cases.

The interdependence between spatial and angular resolution may become relevant if artifacts and other distortions appear.[22] These visual phenomena include (but are not limited to) the crosstalk effect, discrete image borders, ghosting and sudden view jumps. However, due to the high level of requirements, the connections between the parameters may be discarded and thus they can be approached independently.

The depth budget of the system defines how much the content may spatially deviate from the plane of the screen of the display. This not only affects the perception of depth, but also has an effect on the area of the valid FOV, as viewing the display from a position that is closer to the screen than the measure of the utilized depth budget results in the perception of invalid FOV. This means that a portion of visual information becomes perceptually inaccessible to the viewer, which also degrades the overall visual experience. As for the valid FOV, although common medical use cases imply viewers being relatively close to the display, but if larger distances are targeted as well, the additional requirement on angular resolution must be taken into consideration.

Lastly, it is important to state that the vast majority of such medical data is still image data. Therefore, the more lenient parameters of light-field video visualization[23] constrains do not apply. In this discussion, locally stored data was mainly considered, or at least the lack of real-time use cases. If transmission is key — e.g., in case of remote surgical planning — then certain levels of quality and data volume trade-offs must be taken into consideration.

7. CONCLUSION

In this paper, we have analyzed the potential synergies of visual scene reconstruction and medical image reconstruction. The paper highlights the practical use cases of tomography systems, such as CT, PET and SPECT. We have also provided a discussion on the visualization of medical information, with strong considerations on the requirements towards visual quality.

ACKNOWLEDGMENTS

REFERENCES

[1] Brooks, R. A. and Di Chiro, G., "Theory of image reconstruction in computed tomography," *Radiology* **117**(3), 561–572 (1975).

[2] Pan, X., Sidky, E. Y., and Vannier, M., "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?," *Inverse problems* **25**(12), 123009 (2009).

[3] Wilson, D. W. and Tsui, B. M., "Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images," *IEEE transactions on nuclear science* **40**(4), 1198–1203 (1993).

[4] Wirth, A., Cserkaszky, A., Kári, B., Legrady, D., Feher, S., Czifrus, S., and Domonkos, B., "Implementation of 3D Monte Carlo PET reconstruction algorithm on GPU," in [*Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*], 4106–4109, IEEE (2009).

[5] Gac, N., Mancini, S., Desvignes, M., and Houzet, D., "High speed 3D tomography on CPU, GPU, and FPGA," *EURASIP Journal on Embedded systems* **2008**, 5 (2008).

[6] Zhang, Z., "Microsoft kinect sensor and its effect," *IEEE multimedia* **19**(2), 4–10 (2012).

[7] Seitz, S. M. and Dyer, C. R., "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision* **35**(2), 151–173 (1999).

[8] Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., Hoppe, H., Kirk, A., and Sullivan, S., "High-quality streamable free-viewpoint video," *ACM Transactions on Graphics (TOG)* **34**(4), 69 (2015).

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

[10] Vagharshakyan, S., Bregovic, R., and Gotchev, A., "Light field reconstruction using shearlet transform," *IEEE transactions on pattern analysis and machine intelligence* **40**(1), 133–147 (2018).

[11] Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., and Gross, M. H., "Scene reconstruction from high spatio-angular resolution light fields.," *ACM Trans. Graph.* **32**(4), 73–1 (2013).

[12] Dyer, C. R., "Volumetric scene reconstruction from multiple views," in [*Foundations of image understanding*], 469–489, Springer (2001).

[13] Cromer, D. T. and Mann, J. B., "Compton scattering factors for spherically symmetric free atoms," *The Journal of Chemical Physics* **47**(6), 1892–1893 (1967).

[14] Schlick, C., "An inexpensive BRDF model for physically-based rendering," in [*Computer graphics forum*], **13**(3), 233–246, Wiley Online Library (1994).

[15] Magalhães, D. S., Serra, R. L., Vannucci, A. L., Moreno, A. B., and Li, L. M., "Glasses-free 3D viewing systems for medical imaging," *Optics & Laser Technology* **44**(3), 650–655 (2012).

[16] Narita, Y., Tsukagoshi, S., Suzuki, M., Miyakita, Y., Ohno, M., Arita, H., Saito, Y., Kokojima, Y., Watanabe, N., Moriyama, N., and Shibui, S., "Usefulness of a glass-free medical three-dimensional autostereoscopic display in neurosurgery," *International journal of computer assisted radiology and surgery* **9**(5), 905–911 (2014).

[17] Kara, P. A., Kovács, P. T., Martini, M. G., Barsi, A., Lackner, K., and Balogh, T., "From a Different Point of View: How the Field of View of Light Field Displays affects the Willingness to Pay and to Use," in [*Eighth International Workshop on Quality of Multimedia Experience (QoMEX)*], (2016).

[18] Kara, P. A., Kovács, P. T., Martini, M. G., Barsi, A., Lackner, K., and Balogh, T., "Viva la Resolution: The Perceivable Differences between Image Resolutions for Light Field Displays," in [*5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*], (2016).

[19] Kara, P. A., Martini, M. G., Kovács, P. T., Imre, S., Barsi, A., Lackner, K., and Balogh, T., "Perceived quality of angular resolution for light field displays and the validity of subjective assessment," in [*International Conference on 3D Imaging (IC3D)*], (2016).

[20] Cserkaszky, A., Barsi, A., Kara, P. A., and Martini, M. G., "To interpolate or not to interpolate: Subjective assessment of interpolation performance on a light field display," in [*IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*], 55–60, IEEE (2017).

[21] Kara, P. A., Kovács, P. T., Vagharshakyan, S., Martini, M. G., Barsi, A., Balogh, T., Chuchvara, A., and Chehaibi, A., "The Effect of Light Field Reconstruction and Angular Resolution Reduction on the Quality of Experience," in [*12th International Conference on Signal Image Technology & Internet Based Systems (SITIS) 3rd International Workshop on Quality of Multimedia Services (QUAMUS)*], (2016).

[22] Kara, P. A., Cserkaszky, A., Barsi, A., Papp, T., Martini, M. G., and Bokor, L., "The Interdependence of Spatial and Angular Resolution in the Quality of Experience of Light Field Visualization," in [*International Conference on 3D Immersion (IC3D)*], IEEE (2017).

[23] Kara, P. A., Cserkaszky, A., Martini, M. G., Barsi, A., Bokor, L., and Balogh, T., "Evaluation of the Concept of Dynamic Adaptive Streaming of Light Field Video," *IEEE Transactions on Broadcasting* **64**(2), 407–421 (2018).