Advancing depth perception in spatial computing with binocular metalenses

Junkyeong Park, Gyeongtae Kim and Junsuk Rho

Related articles

More related article in Opto-Electronic Journals Group website

OEA | Opto-Electronic Advances

http://www.oejournal.org/oea

OE_Journal   @OptoElectronAdv

Advancing depth perception in spatial computing with binocular metalenses

Junkyeong Park[1], Gyeongtae Kim[1] and Junsuk Rho[1,2,3,4]*

Spatial computing and augmented reality are advancing rapidly, with the goal of seamlessly blending virtual and physical worlds. However, traditional depth-sensing systems are bulky and energy-intensive, limiting their use in wearable devices. To overcome this, recent research by X. Liu et al. presents a compact binocular metalens-based depth perception system that integrates efficient edge detection through an advanced neural network. This system enables accurate, real-time depth mapping even in complex environments, enhancing potential applications in augmented reality, robotics, and autonomous systems.

Keywords: metasurface; metalens; deep learning; depth perception; edge detection

The fields of spatial computing[1] and augmented reality (AR) are on the cusp of a transformative evolution, as researchers and engineers develop technologies to seamlessly blend the virtual and physical worlds. To make this vision a reality, devices must accurately perceive depth, allowing virtual objects to integrate naturally within real-world environments. Methods for achieving depth perception include stereo cameras[2], structured light[3], and time-of-flight systems[4]. Among these, stereo cameras are widely used because they do not require an active light source, resulting in lower power consumption. However, traditional lenses tend to be thick and heavy, which is a significant drawback.

Metasurfaces have emerged as a promising alternative to conventional optical components[5–8]. These ultra-thin, lightweight surfaces can manipulate light at a subwavelength scale, enabling advanced optical functionalities

that were previously unattainable with traditional lenses. In depth-sensing applications, metasurfaces have been effectively applied to point spread function (PSF) engineering[9] and structured light[10, 11], showing great potential for developing more compact and efficient depth perception systems. As the demand for lightweight and compact depth cameras grows, research on metasurface-based depth perception has accelerated. In a recent work published in Opto-Electronic Science[12], X. Liu et al. has recently introduced a groundbreaking binocular metalens-based depth perception system. This compact and lightweight solution promises to enhance next-generation wearable devices, bringing us closer to a more immersive and practical spatial computing experience.

The physical operation of the device relies on a binocular metalens, a 532 nm filter, and a CMOS sensor, to function like human binocular vision (Fig. 1). This novel

[1]Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea;
[2]Department of Chemical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea;
[3]Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea; [4]POSCO-POSTECH-RIST Convergence Research Center for Flat Optics and Metaphotonics, Pohang 37673, Republic of Korea.
*Correspondence: Rho J, E-mail: jsrho@postech.ac.kr

metalens employs an array of nanopillars, each crafted from gallium nitride (GaN) on a sapphire substrate. Unlike conventional lenses, which often sacrifice form for function, these metalenses are exceptionally thin, lightweight, and efficient. Each metalens has a diameter of 2.6 mm and a volume of 4.25×10^{-6} cm³, with a weight of 2.61×10^{-5} g, making it lighter than one percent of a typical hair's weight. This enables a device to capture depth data with greater accuracy and clarity, all within a design optimized for portability and comfort.
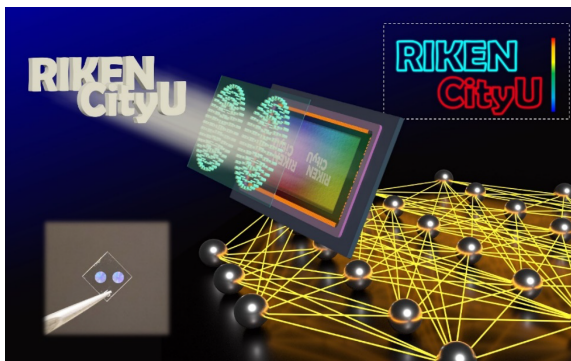

Fig. 1 | Schematic of the edge-enhanced spatial computing with binocular metalens.

The researchers developed H-Net, a stereo-matching neural network that follows an end-to-end learning framework from stereo input images to disparity map prediction. They could address the limitations of traditional stereo-matching algorithms, which frequently struggle in ambiguous environments such as flat, textureless surfaces or reflective areas where it is challenging to capture accurate depth data. To overcome these difficulties, H-Net's architecture incorporates a novel "H-Module," which includes cross-pixel interaction and cross-view interaction mechanisms to aggregate contextual information and combine diverse perspectives effectively. These interactions allow the system to dynamical-

ly prioritize contextual information from stereo views, enhancing edge detection within the images.

As a result, the H-Module filters out unreliable depth data from non-textured regions, resulting in a more precise and stable depth map. This edge-based approach to depth perception, enhanced by cross-pixel and cross-view interaction, minimizes errors in depth prediction. Traditional depth-sensing algorithms often struggle in low-texture regions, as these areas lack sufficient feature points for depth mapping. With the H-Module in H-Net, contextual information is preserved and processed through the 4D cost volume created from left and right image features, and depth estimation is refined through a 3D CNN, followed by a disparity regression module before final disparity map prediction. This approach allows the system to consistently produce accurate depth perception even in complex environments, preserving edges that indicate depth shifts and achieving a more realistic 3D model of the scene.

The research team tested the metalens system's capabilities across various challenging scenarios. For example, with transparent backgrounds, plastic sheets printed with "RIKEN" and "CITYU" placed at 16.0 cm and 12.8 cm were accurately distinguished. Similarly, in a scene with a 3D building sketch at 17.3 cm, along with toy cars positioned at 12.9 cm and 15.7 cm, the system precisely captured the depth differences (Fig. 2). Further tests included architectural sketches at 13.5 cm and 16.5 cm, as well as a toy car body ranging from 12.5 cm to 15.5 cm in depth. The system's ability to identify depth variations, coupled with a processing time of less than 0.15 seconds, highlights its potential for real-time applications in areas like AR, robotics, autonomous navigation, and other time-sensitive uses.

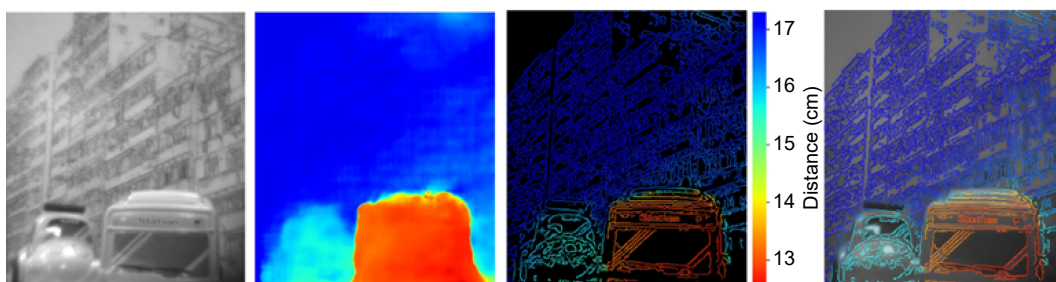The fusion of spatial computing and meta-optics represents a new era, enabling devices like AR glasses and

Fig. 2 | Edge-enhanced depth perception of various objects. The first image displays the original left view, while the second image shows the corresponding depth map. The third image presents the edge-enhanced depth map, using a shared color scale shown to the right of the third image. The fourth image is the integration of the raw image with the edge-enhanced depth map to provide a combined view.

drones to seamlessly integrate digital data with the physical world. This compact, efficient system lays the foundation for immersive and practical devices, driving spatial computing's potential in daily life. Innovations like the binocular metalens system are paving the way for real-time, high-quality 3D modeling across fields as diverse as robotics, healthcare, and virtual reality, marking an essential step toward a more integrated technological future.

References

1. Yenduri G, Ramalinga M, Maddikunta KR et al. Spatial computing: concept, applications, challenges and future directions. arXiv preprint arXiv: 2402.07912, 2024.
   https://doi.org/10.48550/arXiv.2402.07912

2. Keselman L, Woodfill JI, Grunnet-Jepsen A et al. Intel(R) realsense(TM) stereoscopic depth cameras. In *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops* 1–10 (IEEE, 2017);
   http://doi.org/10.1109/CVPRW.2017.167.

3. Geng J. Structured-light 3D surface imaging: a tutorial. *Adv Opt Photonics* **3**, 128–160 (2011).

4. Martins RJ, Marinov E, Youssef MAB et al. Metasurface-enhanced light detection and ranging technology. *Nat Commun* **13**, 5724 (2022).

5. Yang Y, Seong J, Choi M et al. Integrated metasurfaces for re-envisioning a near-future disruptive optical platform. *Light Sci Appl* **12**, 152 (2023).

6. Yu N, Genevet P, Kats MA et al. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334**, 333–337 (2011).

7. Moon SW, Lee C, Yang Y et al. Tutorial on metalenses for advanced flat optics: design, fabrication, and critical considerations. *J Appl Phys* **131**, 091101 (2022).

8. Jeon D, Shin K, Moon SW et al. Recent advancements of metalenses for functional imaging. *Nano Converg* **10**, 24 (2023).

9. Shen ZC, Zhao F, Jin CQ et al. Monocular metasurface camera for passive single-shot 4D imaging. *Nat Commun* **14**, 1035 (2023).

10. Choi E, Kim G, Yun J et al. 360° structured light with learned metasurfaces. *Nat Photon* **18**, 848–855 (2024).

11. Kim G, Kim Y, Yun J et al. Metasurface-driven full-space structured light for three-dimensional imaging. *Nat Commun* **13**, 5920 (2022).

12. Liu XY, Zhang JC, Leng BR et al. Edge enhanced depth perception with binocular meta-lens. *Opto-Electron Sci* **3**, 230033 (2024).

Acknowledgements

Scan for Article PDF