Inverse Design on Meta-Optics for Augmented Reality and Depth Perception

Yao-Wei Huang*
National Yang Ming Chiao Tung University, Hsinchu, Taiwan
ywh@nycu.edu.tw

Abstract

We present a compact metasurface-based dot projector for structured light depth-sensing and facial recognition, eliminating bulky optics while achieving high-quality performance. Our approach features a smaller area, lower power consumption, more dots, and a wider field-of-view. Additionally, a high-Q metasurface with enhanced efficiency via topology optimization demonstrates precise color selectivity. Our augmented reality platform seamlessly integrates real-world scenes with virtual elements, utilizing the high-Q metasurface as an optical combiner.

Author Keywords

Metasurfaces; inverse design; photonic crystal surface-emitting laser; structured light; depth sensing; facial recognition; nonlocal; high *q*uality factor; optical combiner.

1. Introduction

In recent years, metasurfaces have enabled precise manipulation of light both spectrally and spatially at the wavelength scale. With the rapid advancements in augmented reality (AR) and virtual reality (VR) technologies, there is a growing demand for applications in optical combiners and depth sensing. However, challenges remain, including the issue of back-emission in optical combiners and the need for a more compact form factor for near-infrared dot projectors.

Structured light (SL) technology, comprising a dot projector, a camera, and an algorithm to calculate depth, is widely suitable for depth perception and facial recognition applications in smartphones and AR/VR headsets. Current commercial dot projectors, for example, still rely on diffractive optical elements (DOEs) and vertical-cavity surface-emitting laser (VCSEL) arrays. However, the beam divergence from each VCSEL spot is approximately 20 degrees, necessitating additional collimation lenses and light guides. This results in a bulky dot projector design and contributes to the large notches commonly seen on the top screens of smartphones. To address these limitations, many companies are working to advance metasurface technology by integrating it with VCSEL arrays. Notable examples include the Orion Pattern Projector by Metalenz, where thin metasurfaces effectively replace the bulky collimation lenses and DOEs. However, the overall system thickness is still limited by the spacing required between the metasurface and the VCSEL array. Recent studies, such as those by the C. Xu and P. Genevet groups, have demonstrated the seamless integration of GaAs-based metasurfaces with back-emitting VCSELs.[1] By employing a metalens, they achieved a significant reduction in divergence angle to less than 1 degree. Despite this improvement, the low output power of the VCSEL remains a challenge, making it unsuitable for structured light applications in depth perception.

On the other hand, to eliminate back-emission in optical combiners, one promising approach is the use of high-*Q* metasurfaces as optical combiners.[2] High-*Q* metasurfaces are particularly notable for their sharp spectral features and narrow angular responses, with nonlocal metasurfaces often being a prime example.[3] A standout concept in this domain is quasi-bound states in the continuum (q-BIC), which leverage subtle perturbations to break in-plane symmetry. This process enables interaction with free-space light by making the states leaky and excitable.[4] Furthermore, by rotating meta-atoms, additional geometric phases can be introduced for precise wavefront manipulation.[2] Resonant waveguide gratings (RWGs), also known as guided-mode resonance gratings, represent another significant approach to nonlocal metasurfaces. These structures utilize lateral leaky guided modes to achieve sharp reflectance or transmittance spectra.[4] However, forward design methods for RWG and q-BIC structures often face challenges in achieving high efficiency. Recent advances in inverse design strategies for metasurfaces, such as topology optimization, gradient-based inverse design, genetic-tree optimization, evolutionary optimization, deep learning, and hybrid approaches, have shown great promise. While these methods primarily enhance local metasurface characteristics, they also highlight the transformative potential of inverse design in advancing high-*Q* and nonlocal metasurfaces. These innovations open diverse applications and further emphasize the importance of optimization in shaping the next generation of metasurfaces.

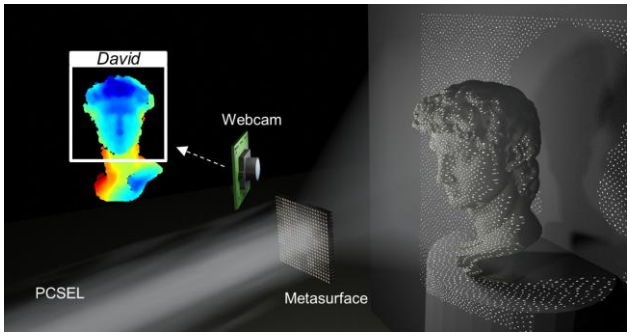

Figure 1. Illustration shows our proposed depth-sensing and facial recognition system.

2. Compact dept perception and facial recognition

Scheme: We developed a metasurface-based dot projector by integrating a metasurface with a photonic crystal surface-emitting laser (PCSEL), as shown in Fig. 1.[5] PCSELs leverage the band-edge resonance effect of a photonic crystal as the laser cavity, achieving a beam divergence of less than 1 degree, thereby eliminating the need for additional collimation lenses. The use of triangular air holes further enhances high-power operation, enabling the realization of watt-class high-power PCSELs. For this study, a commercially available PCSEL sourced from Hamamatsu was used. The laser operates at a current of 300 mA and delivers an optical power of 40 mW. The measured divergence angle is approximately 0.3 degrees, as determined by a laser beam profilometer in the far field. The laser operates in single mode with a central wavelength of 940.2 nm. Additionally, it exhibits partial polarization, with a degree of linear polarization of approximately 0.534. These characteristics highlight the potential for integrating PCSELs with metasurfaces, enabling

ISSN 0097-996X/25/5601-0410-$1.00 © 2025

compact form-factor solutions for advanced optical systems.

Metasurface Design: To accommodate the partially polarized light, the building block of our metasurface is designed as a square pillar. The metasurface and substrate are made of GaAs, a common material for VCSELs and PCSELs, enabling direct fabrication of the metasurface on top of these lasers. The metasurface design is based on computer-generated holography, with the dot pattern pre-encoded in *k*-space. The measured field of view (FOV) extends up to 158 degrees. We demonstrated three different holograms Dots in Sample 1 and Sample 2 are in array arrangement. But the dots in Sample 2 have higher density than those in Sample 1. Dots in Sample 3 are in random arrangement.

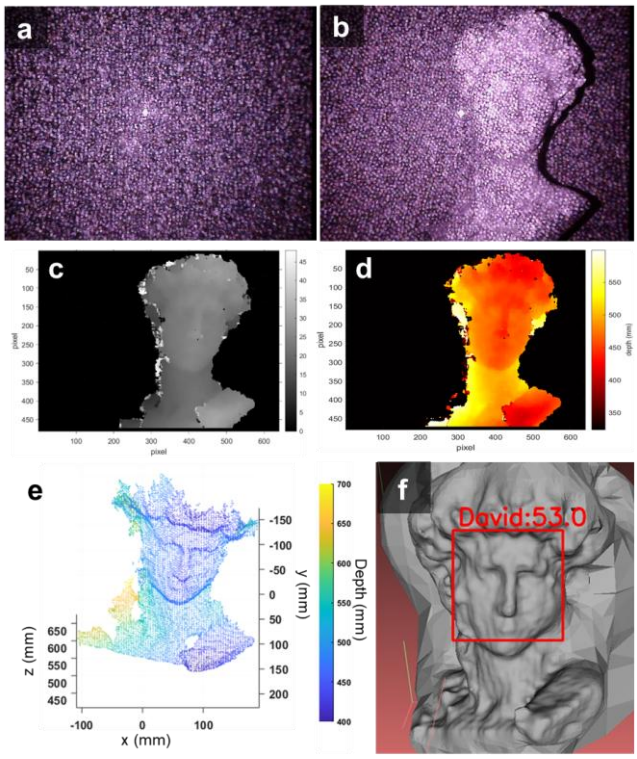

Figure 2. Process of dept sensing and facial recognition. (a) Reference image. (b) object image. (c) Disparity map. (d) depth map. (e) 3D point cloud model. (f) 3D mesh surface and facial recognition with confidence score by using our algorithm.

Dept Perception: Depth estimation is performed through coordinate transformation based on an approximation of the triangular relationship.[6] To evaluate the depth estimation, we begin with the recorded reference image (Fig. 2a) and object image (Fig. 2b). The disparity image (Fig. 2c) is generated by performing feature matching between these two images. Using the triangular relationship and the disparity image, the depth map (Fig. 2d) is calculated. Subsequently, a 3D point cloud model (Fig. 2e) is created by transforming pixel coordinates into global coordinates. Depth sensing is achieved in real-time at approximately 15 frames per second.[7]

Facial Recognition: For facial recognition, we first utilized MeshLab to reconstruct a 3D mesh surface (Fig. 2f) from the point cloud data. Next, a facial recognition classifier was trained using OpenCV. The training process involved inputting several face images across various categories, including photos of "David," into the algorithm to establish a feature extraction database. When a new image, such as a reconstructed 3D mesh surface, is input, the algorithm assigns it to the appropriate category with a confidence score. The results showed that both the real structure and the reconstructed 3D mesh surface were successfully recognized as "David." The confidence score for the 3D mesh surface of David indicates strong performance in facial recognition, demonstrating the system's accuracy and reliability.

Comparison with State-of-the-arts: We compared the results from Sample 3 with those obtained using the Face ID system. In the latter case, the dot projection from Face ID was used, but the depth estimation, 3D point cloud generation, and facial recognition were performed using our algorithm. The comparison revealed that both systems produced similar confidence scores, accurately categorizing the facial features of "David" with comparable performance. Additionally, we conducted reliability and accuracy tests for depth sensing. Across all cases, the estimated depths were close to the ground-truth values. Notably, Sample 3 demonstrated a smaller deviation compared to Sample 2 and the Face ID system. This improvement can be attributed to the random dot arrangement in Sample 3, which enhances feature matching by improving correlation calculations within block areas during disparity estimation.

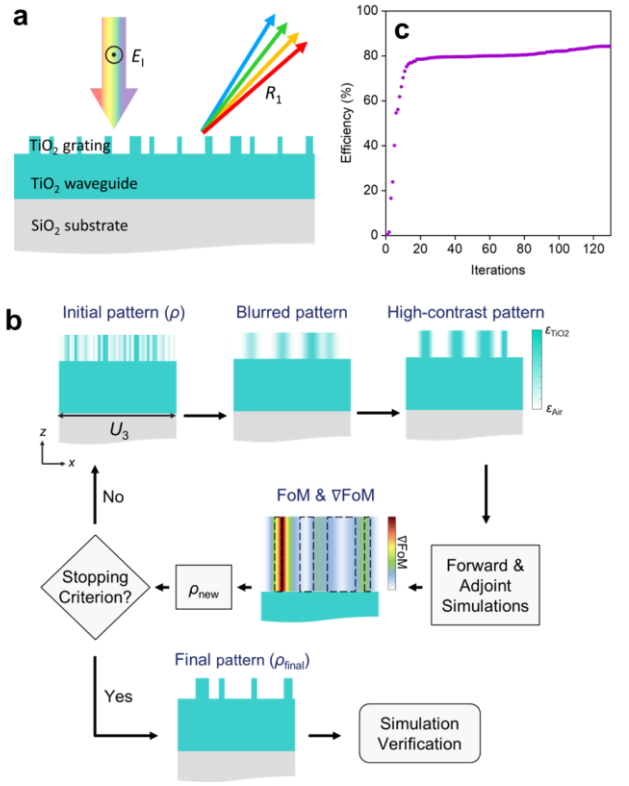

Figure 3. (a) Schematic of our high-*Q* metasurface. (b) Flowchart of the iterative optimization process. (c) Diffraction efficiency (FoM) over the optimization iteration.

3. Topology optimized high-*Q* metasurface for application in optical combiner

Inverse Design: Our high-*Q* metasurface is based on an RWG design (Fig. 3a).[8] The structure consists of a topology-optimized

titania grating on a titania waveguide, all on a glass substrate. The guided mode is determined by the thickness of the waveguide. So once the thickness is fixed, the operating wavelengths are also set. Therefore, we can focus more on the coupling efficiency, when optimizing the freeform shape of the grating.

Our inverse design process utilizes adjoint-based topology optimization (Fig. 3b).[9] The initial design begins with a random pattern representing the distribution of material between air and titania. To ensure fabrication feasibility, a blur function is applied to smooth the pattern, while a contrast function drives the material distribution toward a binary state, representing either air or titania. The figure of merit (FoM) is defined as the first-order reflective diffraction efficiency at a wavelength of 532 nm. The gradient of the FoM is computed using the electric field distributions obtained from both forward and adjoint simulations. With each iteration, the pattern evolves, and the diffraction efficiency increases until the stopping criterion is met (Fig. 3c). This process ultimately yields a binary pattern optimized for the target performance.

Simulation Verification: We validated the performance of the optimized high-*Q* metasurface by analyzing the simulated reflective diffraction results. Normal-incidence light spanning the visible wavelength range was used in the simulations. The high-*Q* metasurface exhibited reflection of four narrow spectral bands at different angles, achieving efficiencies of up to 78% and *Q*-factors as high as 1362. Analysis of the electric field distribution revealed that these resonances correspond closely to guided-mode resonances. Specifically, the modes from TE_0 to TE_3 were associated with red, yellow, green, and blue light, respectively.

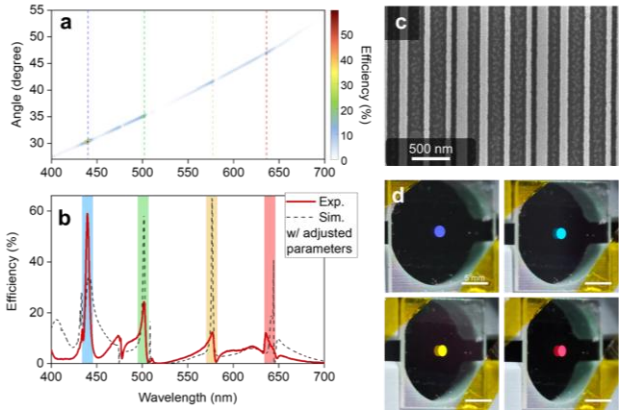

Figure 4. (a) Angular-resolved spectral measurements. (b) The overall spectrum. (c) SEM image of the sample. (d) Reflection images of the high-*Q* metasurface captured at each operational angle using a micro-focus camera

Optical Characteristics: We fabricated our high-*Q* metasurface sample using electron-beam lithography, atomic layer deposition, and the high-density plasma reactive ion etching method. Fig. 4c shows the SEM image of our fabricated grating on top of a titania waveguide. We performed angular-resolved spectral measurements of the first-order reflective diffraction efficiency and the result is shown in Fig. 4a. We observe several local peaks on the colormap. The overall spectrum is derived by extracting the envelope of all angle-dependent spectra. The experimental efficiency reaches up to 59.1% in the blue spectral region, with a *Q*-factor of 93 (Fig. 4b). However, we observe a blue shift in the spectra compared to the optimization target. To identify potential fabrication errors, we adjusted some parameters in simulation (Fig. 4b). The results show a good alignment of peak wavelengths, indicating that the waveguide in the experiment may be thinner than expected. The sample has a diameter of 1 mm, and the photographs demonstrate vivid colors at each operational narrow angle, showing colors corresponding to blue, cyan, yellow, and red light (Fig. 4d). The cyan color results from the blue shift of the green spectral region.

Figure 5. Demonstration of an AR platform utilizing a high-*Q* metasurface as the optical combiner. (a) An AR image showing the fusion of the real-world statue of David with a virtual label "D". (b) An AR image showing the fusion of the real-world fox plush toy with a virtual geolocation marker "NYCU."

Optical Combiner in AR platform: Our topology optimization of high-*Q* metasurfaces extends beyond 1D designs to encompass 2D freeform structures, providing greater flexibility for selectively sustaining and suppressing specific colors. For example, we optimized a high-*Q* metasurface that not only maintains high-*Q* spectra across RGB colors but also effectively suppresses yellow light. Furthermore, we demonstrated an AR platform where the high-*Q* metasurface serves as an optical combiner, with results shown in Fig. 5. In one practical scenario, a facial recognition system identifies a person as "David," and a virtual green "D" instantly appears, seamlessly blending into the real-world scene (Fig. 5a). Another scenario imagines an AR headset user on the National Yang Ming Chiao Tung University (NYCU) campus, looking at the real-world fox mascot plush toy, while a virtual red geolocation marker "NYCU" is displayed overlaid on the scene (Fig. 5b). In our AR demonstration, we utilized a laser back-illuminated micro-liquid crystal display (micro-LCD). Thanks to the narrow spectral bandwidth of the laser source, the virtual images are less prone to diffraction-induced dispersion. However, some blurring is still observed in the virtual image, attributed to aberrations in the lenses used in the optical setup. Addressing these optical aberrations will be an important focus for future research. These results highlight the potential of our metasurface technology for advancing AR applications.

4. Conclusion

We have developed a compact depth-sensing system that delivers high-quality depth perception and facial recognition. Compared to commercial systems, our approach offers several advantages, including a smaller footprint, lower power consumption, an increased number of near-infrared dots, and a FOV. Additionally, we demonstrated a high-*Q* nonlocal metasurface for color selectivity in RYGB, leveraging guided-mode resonance. Through topology optimization, the freeform 1D grating achieved an experimental efficiency enhancement of approximately 15.7 times, effectively addressing the limitations of traditional RWGs.[4] Simulations revealed *Q*-factors as high as 1362, while experimental results showed a *Q*-factor of 93.3. Furthermore, we

demonstrated another high-*Q* metasurface capable of operating across RGB colors. This metasurface was utilized as an optical combiner to seamlessly merge virtual images with real-world scenes. Our research also showcased several algorithms aimed at improving AR systems, including the inverse design of optical combiners, a real-time depth perception algorithm, and a facial recognition classifier. Our work lays the foundation for future applications in AR headsets, advancing both functionality and integration possibilities.

5. Acknowledgements

This work is supported by the National Science and Technology Council in Taiwan (grant nos. 113-2112-M-A49-025) and the Ministry of Education in Taiwan under the Yushan Young Scholar Program. This work was performed in part at the Nano Facility Center and the Center for Nano Science and Technology in the National Yang Ming Chiao Tung University.

6. References

1. Xie Y-Y, Ni P-N, Wang Q-H, Kan Q, Briere G, Chen P-P, et al. Metasurface-integrated vertical cavity surface-emitting lasers for programmable directional lasing emissions. *Nature Nanotechnology*. 2020; 15: 125–130.
2. Malek S C, Overvig A C, Alù A, Yu N. Multifunctional resonant wavefront-shaping meta-optics based on multilayer and multi-perturbation nonlocal metasurfaces. *Light: Science & Applications*. 2022; 11: 246.
3. Overvig A, Alù A. Diffractive nonlocal metasurfaces. *Laser & Photonics Reviews*. 2022; 16 (8): 2100633.
4. Quaranta G, Basset G, Martin O J F, Gallinet B, Color-selective and versatile light steering with up-scalable subwavelength planar optics. *ACS Photonics*. 2017; 4 (5): 1060–1066.
5. Hsu W-C, Chang C-H, Hong Y-H, Kuo H-C, Huang Y-W. Metasurface- and PCSEL-based structured light for monocular depth perception and facial recognition. *Nano Letters*. 2024; 24 (5): 1808–1815.
6. Hsu W-C, Chang C-H, Hong Y-H, Kuo H-C, Huang Y-W. Integration of metasurface and PCSEL in real-time depth-sensing micro-system for facial recognition. *CLEO conference*. 2024; ATh3O.7. doi: 10.1364/CLEO_AT.2024.ATh3O.7.
7. Huang Y-W. Compact depth perception from integration of metasurface and PCSEL. *SPIE Proceedings*. 2024; PC12991: PC1299113. doi: 10.1117/12.3017190.
8. Su H-T, Wang L-Y, Hsu C-Y, Wu Y-C, Lin C-Y, Chang S-M, et al. Topology optimization enables high-*Q* metasurface for color selectivity. *Nano Letters*. 2024; 24 (33): 10055-10061.
9. Huang Y-W, Metasurface-based depth sensing and topology optimized high-*Q* metasurfaces. *SPIE Proceedings*. 2024; 13111: 1311105. doi: 10.1117/12.3017190.