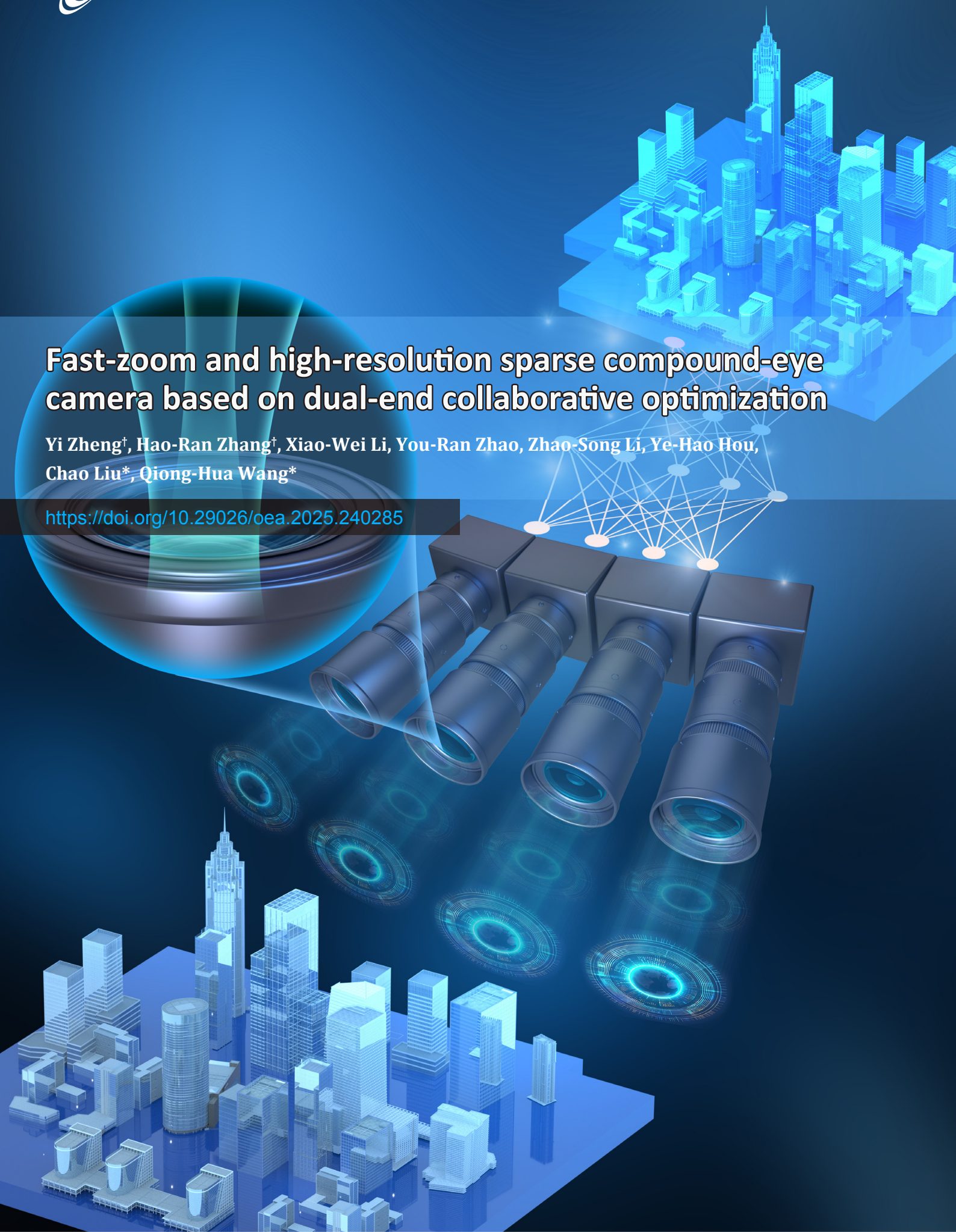# Fast-zoom and high-resolution sparse compound-eye camera based on dual-end collaborative optimization

Yi Zheng[†], Hao-Ran Zhang[†], Xiao-Wei Li, You-Ran Zhao, Zhao-Song Li, Ye-Hao Hou, Chao Liu*, Qiong-Hua Wang*

Fast-zoom and high-resolution sparse compound-eye camera based on dual-end collaborative optimization

Yi Zheng, Hao-Ran Zhang, Xiao-Wei Li, You-Ran Zhao, Zhao-Song Li, Ye-Hao Hou, Chao Liu and Qiong-Hua Wang

---

Related articles

Integral imaging-based tabletop light field 3D display with large viewing angle

Yan Xing, Xing-Yu Lin, Lin-Bo Zhang, Yun-Peng Xia, Han-Le Zhang, Hong-Yu Cui, Shuang Li, Tong-Yu Wang, Hui Ren, Di Wang, Huan Deng, Qiong-Hua Wang

Opto-Electronic Advances   2023 **6**, 220178   doi: 10.29026/oea.2023.220178

Advanced biological imaging techniques based on metasurfaces

Yongjae Jo, Hyemi Park, Hyeyoung Yoon, Inki Kim

Opto-Electronic Advances   2024 **7**, 240122   doi: 10.29026/oea.2024.240122

Large-field objective lens for multi-wavelength microscopy at mesoscale and submicron resolution

Xin Xu, Qin Luo, Jixiang Wang, Yahui Song, Hong Ye, Xin Zhang, Yi He, Minxuan Sun, Ruobing Zhang, Guohua Shi

Opto-Electronic Advances   2024 **7**, 230212   doi: 10.29026/oea.2024.230212

More related article in Opto-Electronic Journals Group website

Fast-zoom and high-resolution sparse compound-eye camera based on dual-end collaborative optimization

Yi Zheng[1†], Hao-Ran Zhang[1†], Xiao-Wei Li[1], You-Ran Zhao[1], Zhao-Song Li[1], Ye-Hao Hou[1], Chao Liu[1,2]* and Qiong-Hua Wang[1,2]*

Due to the limitations of spatial bandwidth product and data transmission bandwidth, the field of view, resolution, and imaging speed constrain each other in an optical imaging system. Here, a fast-zoom and high-resolution sparse compound-eye camera (CEC) based on dual-end collaborative optimization is proposed, which provides a cost-effective way to break through the trade-off among the field of view, resolution, and imaging speed. In the optical end, a sparse CEC based on liquid lenses is designed, which can realize large-field-of-view imaging in real time, and fast zooming within 5 ms. In the computational end, a disturbed degradation model driven super-resolution network (DDMDSR-Net) is proposed to deal with complex image degradation issues in actual imaging situations, achieving high-robustness and high-fidelity resolution enhancement. Based on the proposed dual-end collaborative optimization framework, the angular resolution of the CEC can be enhanced from 71.6" to 26.0", which provides a solution to realize high-resolution imaging for array camera dispensing with high optical hardware complexity and data transmission bandwidth. Experiments verify the advantages of the CEC based on dual-end collaborative optimization in high-fidelity reconstruction of real scene images, kilometer-level long-distance detection, and dynamic imaging and precise recognition of targets of interest.

Keywords: compound-eye camera; zoom; high resolution; collaborative optimization

Introduction

Modern optical imaging systems are expected to have characteristics of large field of view, high resolution, and fast imaging speed[1–6]. However, due to the limitations of spatial bandwidth product and data transmission bandwidth, field of view, resolution, and imaging speed constrain each other in an optical imaging system[7–9]. Insect compound eye is a multi-aperture visual perception organ composed of numerous ommatidia, providing solutions to break the spatial bandwidth product limitation of traditional single-aperture imaging systems[10–15]. Inspired by the insect compound eye, many artificial compound-eye array cameras have been developed and possible to achieve both large field of view and high resolution[16–18]. However, the dense sub-camera array not only generates massive amounts of data, but also leads to a dramatic increase in system size, weight, and power consumption, which severely limits the imaging speed and

portability of the system.

To solve these issues, some representative works have been proposed in recent years. Some researchers proposed compound-eye array cameras with multi-scale optical structures, where the array cameras share a large-scale objective lens at the front end[19−22]. This reduces the number of lenses and improves the compactness of the system to a certain extent, while ensuring a large aperture and achieving well aberration correction. Some researchers proposed array cameras consisting of two types of sub-cameras with different resolutions, which can operate with active selective local high-resolution imaging within a large field of view, reducing data transmission bandwidth demand by adaptively allocating image resources[23,24]. In addition, achieving non-uniform resolution imaging modes through irregular arrangement of different types of lenses also provides a solution to achieve large field-of-view imaging and improve imaging speed[25,26]. However, the demands for a large number of image sensors with high pixel density, or a mass of lens units are always inevitable. Therefore, it is still challenging to achieve large field of view, high resolution, and fast imaging speed with low optical hardware complexity and data transmission bandwidth.

In fact, zoom imaging has been widely used to deal with the conflict between field of view and resolution. Zoom imaging systems can obtain high resolution images of key targets using limited photosensitive pixels by adjusting the imaging focal length[27,28]. Introducing zoom function in compound-eye array cameras will be of great significance in reducing the number of sub-cameras and alleviating the demand of data transmission bandwidth while maintaining the large field of view and high-resolution imaging capability. However, the zoom mode through mechanically moving lens group lacks fast adjustment speed, usually costing several seconds, and requires high complexity causing a relatively large volume and weight. Liquid lenses emerging in recent years have attracted much attention due to the fast focusing characteristics and high compactness[29−32], but the design method of zoom imaging systems based on liquid lenses still needs further exploration. Moreover, zoom imaging systems still suffer degradation of imaging quality caused by many factors such as insufficient aberration correction, manufacturing tolerance and environmental interference. Some deep learning based super-resolution algorithms that do not rely on optical hardware have achieved impressive results, but they may be inevitably constrained by generalization[33−35]. And although point spread function engineering has good interpretability[36−38], it often necessitates a complex calibration and iterative calculation process, with executing difficulty increasing sharply with the number of sub-cameras in compound-eye array cameras. In summary, fast zoom and high-fidelity resolution enhancement are crucial for compound-eye array cameras, because of the potential to deal with the conflict among the field of view, resolution, and imaging speed.

Here, a fast-zoom and high-resolution sparse compound-eye camera (CEC) based on dual-end collaborative optimization is proposed, which provides a cost-effective way to break through the trade-off among the field of view, resolution, and imaging speed, as shown in Fig. 1. Through fast zooming of the optical end and information demodulation of the computational end, the designed sparse CEC can achieve both real-time large-field-of-view imaging and high-resolution imaging. Specifically, in the optical end, a sparse CEC based on liquid lenses is designed, which can achieve fast adjustment of imaging magnification and aberrations. Each sub-camera of the CEC has a field of view more than 15.8° × 11.8° and 2× optical zoom imaging capability, with a zoom response time of only 5 ms. In the computational end, a disturbed degradation model driven super-resolution network (DDMDSR-Net) is proposed, which achieves high-robustness and high-fidelity resolution enhancement. By introducing an imaging degradation model considering environment interference, manufacturing tolerance and acquisition noise, the trained DDMDSR-Net is well suited for the resolution enhancement of the proposed CEC without complex calibration processes. Moreover, a channel attention mechanism is used during the training and reconstruction process of the DDMDSR-Net, ensuring the strong capability of feature information extraction. Through the collaborative zoom and optimization of both the optical and computational ends, the angular resolution of the CEC can be enhanced from 71.6″ to 26.0″, dispensing with high optical hardware complexity and data transmission bandwidth. Experiments verify the advantages of the CEC in high-fidelity reconstruction of real scene images, kilometer-level long-distance detection, and dynamic imaging and precise recognition of targets of interest. The proposed CEC has important application value in fields such as geodesy, search and track, and urban traffic monitoring.
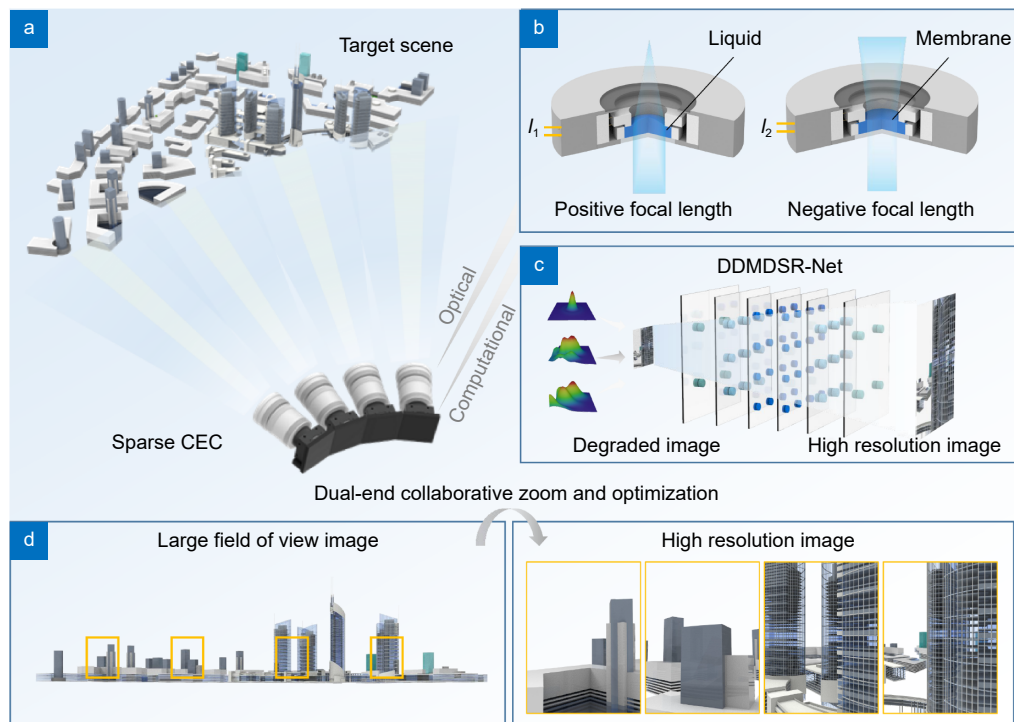
Fig. 1 | Concept of fast-zoom and high-resolution sparse CEC based on dual-end collaborative optimization. (**a**) Schematic diagram of imaging scene and the proposed sparse CEC based on liquid lenses. (**b**) Structure and zoom principle of the liquid lens. (**c**) Schematic diagram of the proposed DDMDSR-Net. (**d**) Schematic diagram of the imaging characteristics of the proposed sparse CEC.

Results and discussion

Principle of the sparse CEC based on dual-end collaborative optimization

The core of the sparse CEC based on dual-end collaborative optimization is to achieve resolution enhancement through the collaboration of fast zooming in the optical end and information demodulation in the computational end. This strategy allows the proposed CEC to operate using sparse sub-camera array and image sensors with limited pixel density, ensuring low optical hardware complexity and data transmission bandwidth. Meanwhile, the large-field-of-view and high-resolution imaging can also be achieved. The workflow diagram of the sparse CEC based on dual-end collaborative optimization is shown in Fig. 2(a). Firstly, the sub-cameras of the sparse CEC are used to capture images of different regions, which can be stitched to form large field of view images. Then, by controlling the liquid lenses in the sub-cameras, the imaging magnification and aberrations of the sub-cameras can be fast adjusted to obtain zoom images with wider information frequency domain. Through the reconstruction by the proposed DDMDSR-Net, the resolution of the images can be further improved.

At the optical end, the liquid lens based sparse CEC is designed, composed of multiple identical sub-cameras arranged on a curved surface. By reasonably setting the angle between the optical axes of the sub-cameras to be slightly smaller than the maximum field of view of the sub-camera, there is an overlap between the fields of view of the sub-cameras, allowing for stitching of large field of view images. By pre-calibrating the homography between the sub-images captured by the sub-cameras and the large field of view images, and combined with adjustments of image brightness and white balance coefficient, stitched large field of view images can be generated. Each sub-camera contains several liquid lenses and solid lenses, and the focal length as well as imaging aberrations can be adjusted by controlling the liquid interfaces of the liquid lenses, which is different from traditional zoom methods that rely on the movement of lens groups. Specifically, the liquid lens are driven by a voice coil motor, which squeezes the cavity through Ampere's force and causes deformation of the liquid interface, thus realizing zoom function. Therefore, the zoom response time of the designed sparse CEC based on liquid lenses can achieve millisecond level. When executing specific design, we use four sub-cameras to form the CEC as an
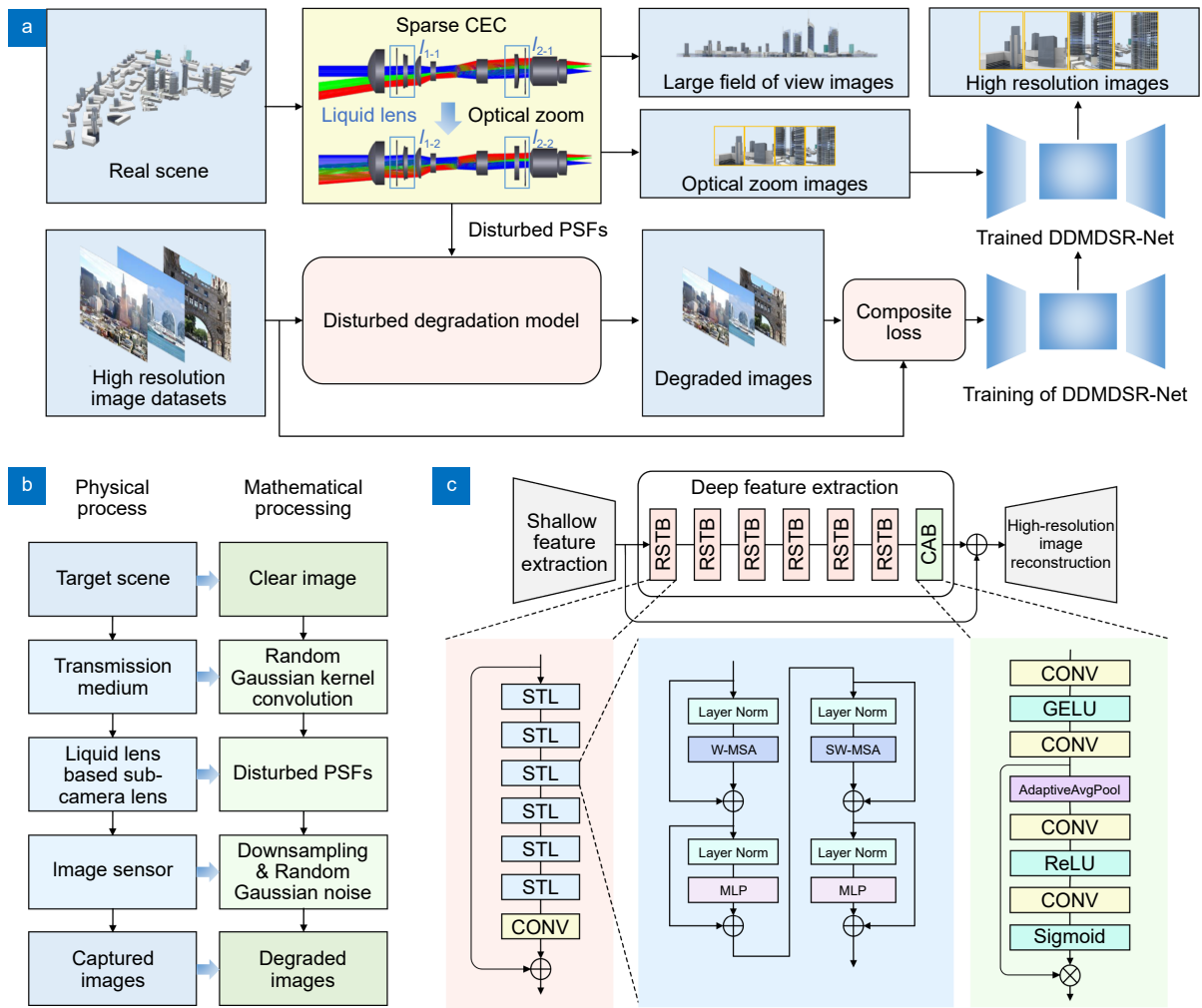
Fig. 2 | Fast-zoom and high-resolution imaging mechanism of the proposed sparse CEC. (**a**) Dual-end collaborative optimization framework. (**b**) Architecture of disturbed degradation model and generation of image training pairs. (**c**) DDMDSR-Net using a multi-channel attention mechanism.

example (Supplementary information Section 1). Each sub-camera contains two liquid lenses, four single solid spherical lenses and two spherical doublet lenses, operating with a field of view more than 15.8° × 11.8° and 2× optical zoom imaging capability. The two liquid lenses are located in the zoom group and compensation group, respectively, with sufficient spacing between them, so as to achieve sufficient modulation of the optical power of the sub-camera.

At the computational end, the DDMDSR-Net is designed to deal with complex image degradation issues in actual imaging situations and achieve further resolution enhancement. A disturbed degradation model is first built for the training datasets generator, considering the composite degradation factors of transmission medium, lens aberration, and image acquisition. The physical process and corresponding mathematical description of the disturbed degradation model are shown in Fig. 2(b), and the imaging degradation process can be represented as

$$I(x, y) = [\mathrm{PSF}(x, y) \otimes T(x, y) \otimes O(x, y)] \downarrow + N(x, y), \quad (1)$$

where $I(x, y)$ and $O(x, y)$ represent degraded image and original image, respectively, $T(x, y)$ is the degradation function of the transmission medium, $\mathrm{PSF}(x, y)$ is the degradation function of the sub-camera lens also called point spread function (PSF), $\downarrow$ represents down-sampling process of the image sensor, and $N(x, y)$ is the noise. The degradation functions and noise are set to vary according to the random disturbed factors, rather than being set as fixed functions, so as to deal with various complex degradation scenarios. Specially, Monte Carlo method is used to generate the degradation functions of the sub-camera lens, which can be expressed as

$$\mathrm{PSF}(x, y) = \mathrm{PSFMonte}(\mathrm{idealPSF}(x, y), d_1, d_2, ..., d_n), \quad (2)$$

where idealPSF(x, y) is the ideal PSF without disturbance degradation, and $d_1, d_2, ... , d_n$ represent values of a set of predefined disturbance factors related to manufacturing tolerances of the lenses and wavefront errors of the liquid lenses. The details of image degradation processing and specific settings can refer to Supplementary information Section 2. Combining high resolution image datasets with the proposed disturbed degradation model, training pairs can be easily generated for supervised learning of the DDMDSR-Net.

The DDMDSR-Net consists of three modules: shallow feature extraction module, deep feature extraction module, and image reconstruction module, as shown in Fig. 2(c). Among them, the shallow feature extraction module consists of 3×3 convolution layers, which are used to extract shallow information from the image and expand the dimension of the feature space. The deep feature extraction module consists of six residual Swin Transformer blocks (RSTBs) and one channel attention block (CAB), which are used to further extract features in high dimensional space[39]. The details of the network can refer to Methods and Supplementary information Section 3. Through the reasonable generalization brought by the proposed disturbance degradation model and the strong feature extraction ability of the network, high-robustness and high-fidelity resolution enhancement can be achieved. Such strategy is well suited for the resolution enhancement of the CEC with multiple identical sub-cameras, dispensing with complex calibration processes.

Fabrication and basic imaging performance test

According to the above design, a prototype of the sparse CEC is fabricated, as shown in Fig. 3(a). Each solid lens in the sub-cameras is coated with a visible light band anti-reflective film, and each sub-camera contains two liquid lenses with the type of EL-12-30-TC provided by Optotune Switzerland AG. The low bandwidth image sensor with the type of VEN-161-61U3C provided by Daheng Imaging Inc. has a resolution of 1080×1440, and supports a maximum frame rate of 61.2 fps for image acquisition. Four sub-cameras are fixed on an installation bracket, the mechanical length of each sub-camera is within 10 cm and the weight is approximately 250 g. And each sub-camera can be independently or synchronously driven to achieve zoom function through a customized current driver. In the experiment, we mainly control the sparse CEC to switch between two critical focal length states with a fastest response time of approximately 5 ms, and the relationship between driving currents and focal length is shown in Fig. 3(b).

To test the basic imaging performance, the sparse CEC is used to capture images of a resolution target. The images captured by one of the sub-cameras before and after dual-end collaborative zoom and optimization are shown in Fig. 3(c, d), being taken as an example. It can be tested that the sub-camera can capture large field of view images as expected, and the total horizontal field of view of the sparse CEC with a single exposure can reach more than 40°. The line pairs of the fourth group elements can be distinguished, which means the initial angular resolution reaches 71.6". And after dual-end collaborative zoom and optimization, the line pairs of the eleventh group elements can be clearly distinguished, which means the angular resolution reaches 26.0", and significant improvement in the contrast of the line pairs of the fourth group elements can also be observed. The imaging performance test results indicate the proposed method can not only enhance low-frequency information, but also successfully overcomes the interference of blur and noise, extracting and amplifying the required high-frequency information (Supplementary information Section 4).

Imaging performance for real scenes and applications

Imaging and high-fidelity reconstruction for real scenes

In order to test the imaging performance of the sparse CEC based on dual-end collaborative optimization framework in more general real-world scenarios, it is used to capture images of a parking lot scene. After pixel mapping and white balance and image brightness correction, we successfully complete the stitching of four sub field of view images, forming a large field of view image of the scene, as shown in Fig. 4(a). We enlarge and compare some key areas, especially those containing abundant semantic and texture information, to verify the effectiveness of high-resolution reconstruction, as shown in Fig. 4(b, c).

It can be seen that after dual-end collaborative zoom and optimization, the texts on the landmark, license plates, and road sign becomes clearer, and the texture
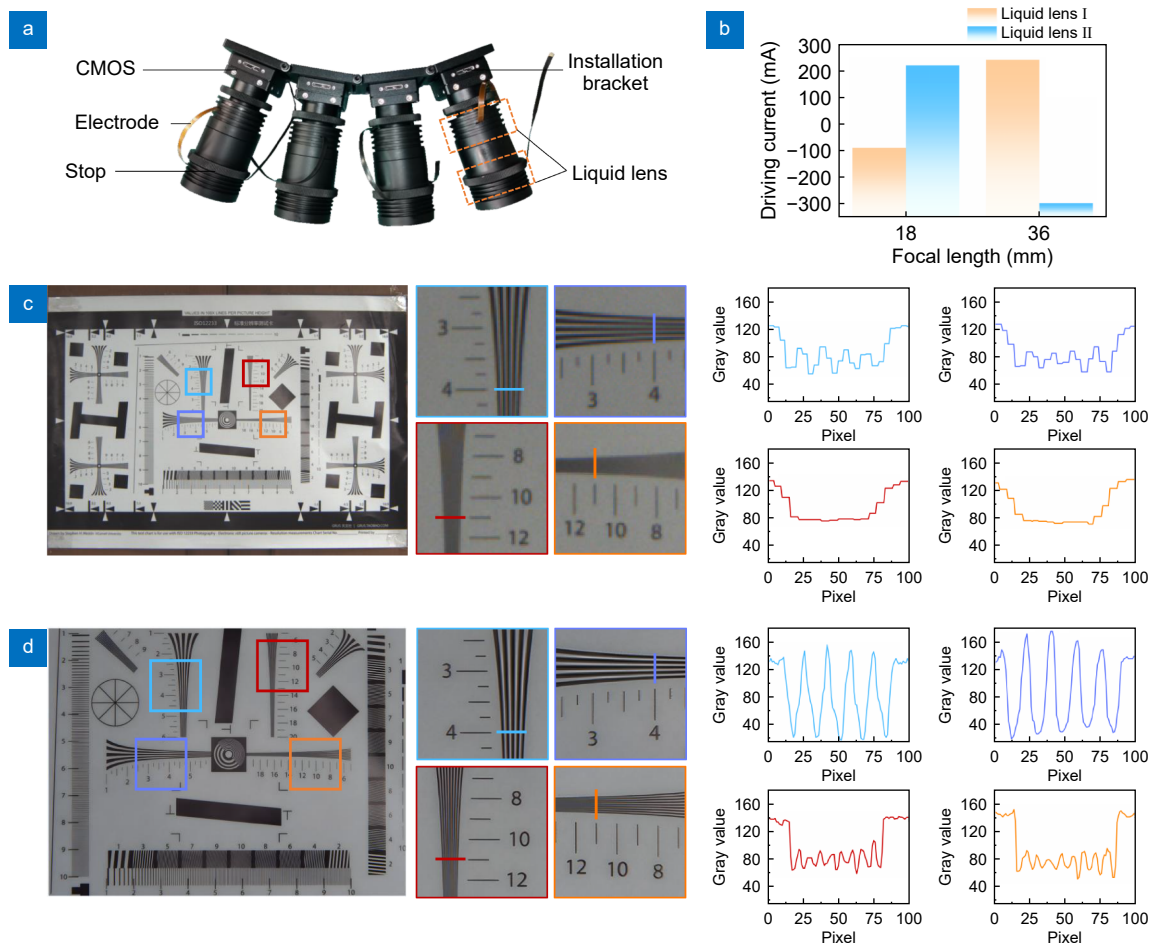
Fig. 3 | Fabrication and imaging performance test of the proposed sparse CEC. (**a**) Prototype of the proposed sparse CEC. (**b**) Relationship between the driving currents of the liquid lenses and the focal length of the sub-camera. (**c**) Large field of view image of the resolution target captured by a sub-camera in the CEC. (**d**) High resolution image of the resolution target through the dual-end collaborative zoom and optimization.

information of the speed bump is also successfully enhanced. Although humans can mentally supplement information based on life experience, it is difficult for machines to do so. The proposed strategy successfully enhances the effective information of key target images in real scenes by combining optical and computational methods, demonstrating significant practical application value in machine vision.

Imaging performance for targets at different distances
When capturing targets at different distances, the environmental interference factors in the imaging may change, which brings challenges to the robustness and generalization of the method. To further test the imaging performance when dealing with different imaging conditions, the sparse CEC is used to capture an urban scene, which contains targets at different distances.

Similar to the previous experimental steps, we stitch the sub field of view images captured by four sub-cameras to obtain a large field of view image of the scene, as shown in Fig. 5(a). Then we compare the locally magnified images of targets at different distances before and after dual-end collaborative zoom and optimization, including buildings located approximately 0.15 km and 0.2 km away, as well as signs located approximately 1.7 km away from the sparse CEC. It can be seen that the proposed strategy is still effective, bringing obvious improvement of image resolution and significant suppression of noise and blur, as shown in Fig. 5(b). This is the result of the combination of fast optical zoom which effectively extends the captured information frequency domain, and disturbed degradation model which helps demodulate and enhance effective information.

Moreover, we also use a digital single lens reflex (DSLR) camera (Canon EOS 6D Mark II, with 26.2 million pixels) with a lens (RF24-105mm F4 L IS USM) to

Fig. 4 | Imaging and high-fidelity reconstruction results for real scenes. (**a**) Large field of view image captured by the proposed sparse CEC. (**b**) Local magnification of the large field of view image. (**c**) Reconstructed high resolution images through the dual-end collaborative zoom and optimization.

Fig. 5 | Imaging and high resolution reconstruction results of targets at different distances. (**a**) Large field of view image captured by the proposed sparse CEC. (**b**) Comparison between the raw images, reconstructed images through the dual-end collaborative zoom and optimization, and the images captured by a DSLR camera.

capture images of the targets under telephoto shooting status, for comparing the fidelity of high-resolution reconstruction of the images. Although the exposure time of the DSLR camera is set to be the same as the proposed CEC, due to differences in ambient lighting conditions and spectral responsiveness of image sensors, there are some differences in the brightness and visual contrast of the images. Nevertheless, the feature information of the targets remains basically consistent, and the reconstructed images show better denoising and anti-ghosting effects. This means that the proposed sparse CEC operating with dual-end collaborative optimization framework achieves comparable imaging quality with a data transmission bandwidth requirement around four times lower than the DSLR camera.

Dynamic imaging and precise recognition of targets of interest

To demonstrate the fast imaging speed characteristics, we construct a dynamic imaging pipeline for the camera and select a highway as the imaging test scenario, as shown in Fig. 6(a). Although each sub-camera can be in-

dependently controlled and selectively trigger zoom imaging, to test the critical situation of dynamic imaging, the zoom imaging time duty cycle is set to 50%. That is to say, the imaging focal length state of each sub-camera is controlled by a square wave signal with a duty cycle of 50%, and dual focus images are alternately captured, as shown in Fig. 6(b).

Initialization processing is performed at the beginning of the operation, during which the white balance coefficient, gain coefficient, and internal and external parameters of the sub-cameras are pre-calibrated. During the first duration T, the focal lengths of the sub-cameras are synchronously adjusted to 18 mm, and images captured by the four sub-cameras can be directly stitched through pixel mapping to form a large field of view image. Then, during the second duration T, the focal lengths of the sub-cameras are synchronously adjusted to 36 mm, and images captured by the four sub-cameras can be stored, then reconstructed through the DDMD-SR-Net in the background or executed other processing. Due to lower bandwidth requirements, the storage can be completed in real-time on a single computer. The
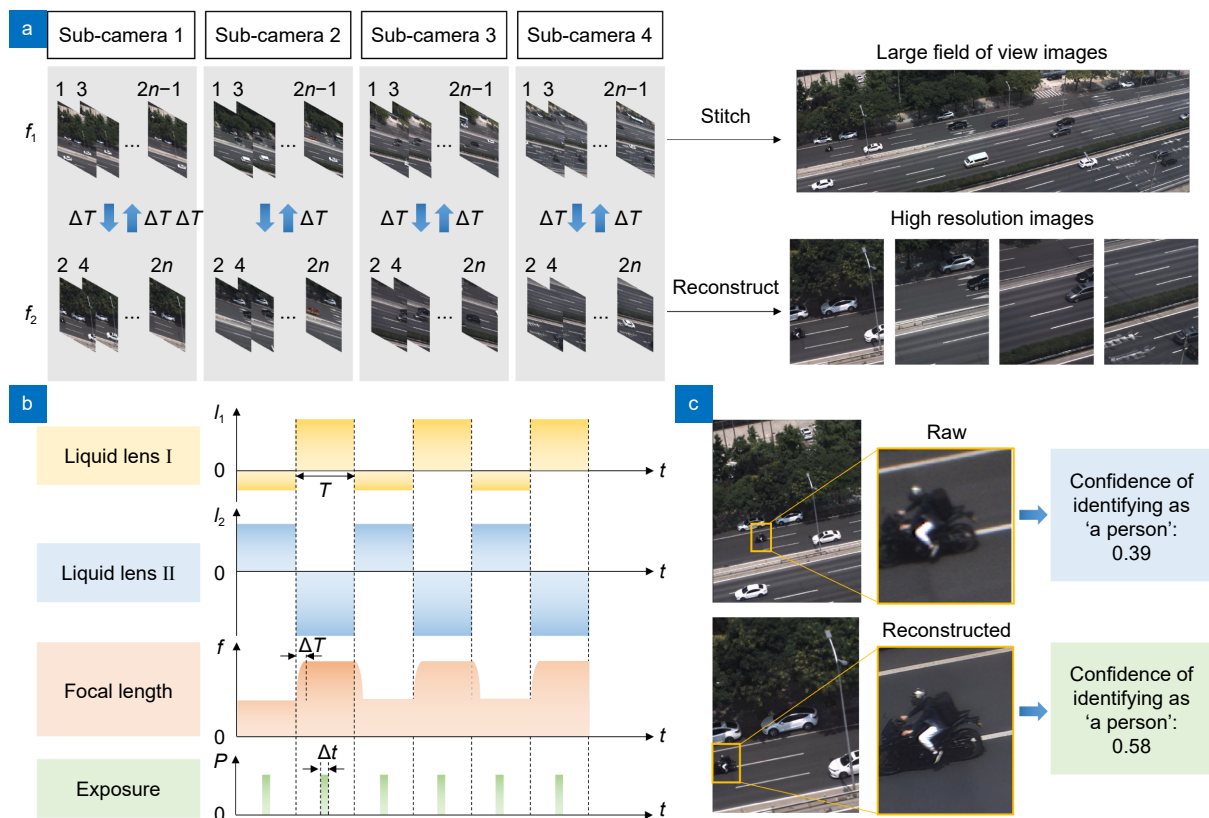

Fig. 6 | Dynamic imaging pipeline and precise recognition of targets of interest. (**a**) Dynamic imaging pipeline of the proposed sparse CEC. (**b**) Timing control of zoom and image acquisition. (**c**) Recognition results of targets of interest before and after dual-end collaborative zoom and optimization.

acquisition speed of dynamic dual focus images is mainly determined by the zoom response time of the liquid lenses and the exposure time of the image sensors. During the operation, it is found that although the focal length of the liquid lens can be basically achieved within 5 ms after the zoom signal is triggered, it is best to delay the start time of exposure appropriately due to the continuous slight oscillation of the liquid surface, which can help to avoid image integration blur. The supportable acquisition period of dual focus images is limited to:

$$2T \geqslant 2(\Delta T + \Delta t + \varepsilon) , \qquad (3)$$

where ΔT is the zoom response time of the liquid lenses considering the duration of slight oscillation, Δt is the exposure time, and ε is the time delay caused by synchronization error of the sub-cameras or other interference factors. In the demonstration experiment (Supplementary Media 1), ΔT is set to 20 ms, Δt is set to 5 ms, and ε is set to 13 ms, which means the acquisition frame rate of each focal length image can reach 13 fps and the time difference between obtaining images of the same target with two focal lengths does not exceed 38 ms, which is sufficient to deal with most dynamic scenes. It should be noted that by optimizing the synchronization triggering mechanism, ε can theoretically be further compressed and faster dynamic dual focus imaging speed can be achieved. The maximum acquisition frame rate of each focal length image can reach around 20 fps, which is almost impossible to achieve by using traditional mechanical zoom mode.

In addition, we also demonstrate the potential applications of the sparse CEC based on dual-end collaborative optimization framework in target search and precise recognition. When used for urban traffic monitoring, the sparse CEC can achieve real-time acquisition of large field of view images, which means that it has efficient search capabilities. Meanwhile, for potential targets of interest, precise recognition can be achieved through dual-end collaborative zoom and optimization. For example, when using YOLOV5 algorithm[40] for detection for the motorcycle rider in the scene, as shown in Fig. 6(c), the confidence of identifying the target as 'a person' before dual-end collaborative zoom and optimization is only 0.39, indicating a risk of misjudgment. After dual-end collaborative zoom and optimization, the confidence of identifying the target as 'a person' reaches 0.58, basically achieving accurate recognition. The reasons for not achieving higher confidence may include the pres-

ence of motion blur of dynamic targets. It is possible to deal with this issue by further reducing the image exposure time and optimizing the imaging frame rate, or using additional algorithms for motion blur suppression. In addition, when the imaging frame rate is high enough, the display frame rate of the display device also needs special attention to avoid possible information loss[41] in interactive usage scenarios with user participation.

In summary, a fast-zoom and high-resolution sparse CEC based on dual-end collaborative optimization framework is developed in this paper, which combines the advantages of liquid lens technology and computational imaging technology. By designing the sub-cameras based on liquid lenses, effective extension of captured information frequency domain for the targets of interest can be achieved through fast optical zoom. And through the designed DDMDSR-Net, effective information demodulation and resolution enhancement can be achieved with high robustness and fidelity. The designed network also shows significant advantages compared to other methods, including more effective reconstruction method based on the proposed disturbed degradation model, and stronger data learning ability (Supplementary information Section 5).

We not only present the operation process and effectiveness of the dual-end collaborative optimization framework through resolution testing experiments, but also demonstrate the ability to deal with complex image degradation issues through actual imaging experiments in various real scenarios. The successful implementation of the results is attributed to the combination of the physical driving optical zoom method based on liquid lenses and the computational demodulation method based on the proposed DDMDSR-Net. The liquid lens based optical zoom method can fast expand the frequency domain of obtainable information when capturing images of the target area of interest, and the DDMDSR-Net can extract and enhance effective information with high-robustness and high-fidelity within the frequency domain of obtainable information in actual imaging scenarios with multiple interferences. The designed compound eye camera has strong dynamic imaging capability, and can effectively complete the tasks of efficient search and accurate recognition of dynamic targets. Moreover, we believe that our proposed dual-end collaborative optimization framework can help achieve sparsity in compound-eye array cameras and significantly reduce the cost of use. For example, compared to the

classic Aqueti Mantis series compound-eye array camera consisting of 18 sub cameras with 107 megapixels totally[42], our designed CEC consisting of 4 sub cameras can achieve comparable field of view and resolution during real-world imaging, and can reduce data transmission bandwidth and weight by about an order of magnitude, which indicates the potential to be used on mobile platforms.

The proposed CEC has the potential for further improvement through some ways, such as using liquid lenses with larger zoom range[43], increasing the number of liquid lenses used in each sub-camera, optimizing the system design, or using some advanced photoelectric devices[44] to further enhance the zoom capability of the CEC, achieving higher resolution. However, the trade-off between zoom imaging capability and hardware cost needs to be considered. Besides, some advanced optical and digital methods can also be deployed to improve imaging performance, including further suppressing image artifacts caused by color dispersion[45,46] and reducing image stitching gaps[47]. With further hardware optimization and on-chip acceleration algorithm design, the proposed sparse CEC can provide solutions for many practical application tasks.

Conclusions

In conclusion, a fast-zoom and high-resolution sparse CEC based on dual-end collaborative optimization framework is proposed, which provides a cost-effective way to break through the trade-off among the field of view, resolution, and imaging speed. The proposed sparse CEC enables large-field-of-view imaging in real time, can achieve fast zoom imaging with a response time of 5 ms, and can successfully enhance the angular resolution from 71.6″ to 26.0″ through dual-end collaborative optimization. This work represents a significant advancement combining new optical devices and design methods with advanced computational imaging methods, which is effective to improve imaging performance while maintaining a low demand of optical hardware complexity and data transmission bandwidth. The proposed CEC has important application value in fields such as geodesy, search and track, urban traffic monitoring.

Methods

Structure of the DDMDSR-Net
The proposed DDMDSR-Net uses SWinIR[48] as the fun-

damental model, and a channel attention mechanism is introduced to the model to deeply fuse the extracted feature maps. In the deep feature extraction module of the DDMDSR-Net, the RSTB module is a residual module with Swin Transformer layers (STLs) and convolution layer, where STL is the core of RSTB. By using the shift window mechanism of STL, the input tensor is divided into multiple non-overlapping windows, and the local attention of each window is calculated separately, which has less computational resource consumption and faster speed than traditional Vision Transformer (ViT). And CAB can perform channel based wise weighted allocation on the obtained feature maps, activate and amplify useful features, and suppress useless features.

Training details
During the process of network training, a composite loss function is set as

$$L = \alpha L_{L1} + \beta L_{GAN} + \eta L_{PE} , \qquad (4)$$

where L_{L1}, L_{GAN} and L_{PE} are the naive L1 pixel loss, GAN loss and perceptual loss[49], respectively, α, β and η are the corresponding weights, respectively, which are specifically set as 1, 0.1 and 1 in the experiment. For the GAN loss, we use the PatchGAN model[50] as the discriminator network, which is trained jointly with the generator network to facilitate the recovery of high-frequency features. For the perceptual loss, we use the VGG19 network to extract image features, which encourages the network to generate images with more realistic details and better visual effects.

The Flickr2K and part of the OST datasets are used as the training set containing a total of 6500 images. We finetune the pretrained SwinIR model in realword super-resolution task to accelerate the training process. In the training phase, the images are cropped into a set of small patches with a size of 256×256×3. The learning rate and batch size are set to 1×10^{-6} and 16 respectively. We run a total of 300 K iterations with 2 NVIDIA 3090 GPUs, and get the final training weights for testing.

References

1. Zhang YL, Song XF, Xie JC et al. Large depth-of-field ultra-compact microscope by progressive optimization and deep learning. *Nat Commun* **14**, 4118 (2023).
2. Xu X, Luo Q, Wang JX et al. Large-field objective lens for multi-wavelength microscopy at mesoscale and submicron resolution. *Opto-Electron Adv* **7**, 230212 (2024).
3. Zhang YL, Wang MR, Zhu QY et al. Long-term mesoscale imaging of 3D intercellular dynamics across a mammalian organ.

Cell **187**, 6104–6122.e25 (2024).

4. Zheng Y, Wang X, Jiang Z et al. Adaptive multiscale microscope with fast zooming, extended working distance, and large field of view. *Light Adv Manuf* **5**, 62–74 (2024).

5. Xing Y, Lin XY, Zhang LB et al. Integral imaging-based tabletop light field 3D display with large viewing angle. *Opto-Electron Adv* **6**, 220178 (2023).

6. Lee YH, Zhan T, Wu ST. Enhancing the resolution of a near-eye display with a Pancharatnam-Berry phase deflector. *Opt Lett* **42**, 4732–4735 (2017).

7. Xue YJ, Yang QW, Hu GR et al. Deep-learning-augmented computational miniature mesoscope. *Optica* **9**, 1009–1021 (2022).

8. Chen R, Tang X, Zhao YX et al. Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. *Nat Commun* **14**, 2854 (2023).

9. Jo Y, Park H, Yoon H et al. Advanced biological imaging techniques based on metasurfaces. *Opto-Electron Adv* **7**, 240122 (2024).

10. Zhang L, Zhan HY, Liu XY et al. A planar compound eye based microsystem for high precision 3D perception. *PhotoniX* **5**, 21 (2024).

11. Bae B, Lee D, Park M et al. Stereoscopic artificial compound eyes for spatiotemporal perception in three-dimensional space. *Sci Robot* **9**, eadl3606 (2024).

12. Song YM, Xie YZ, Malyarchuk V et al. Digital cameras with designs inspired by the arthropod eye. *Nature* **497**, 95–99 (2013).

13. Jiang H, Tsoi CC, Yu WX et al. Optical fibre based artificial compound eyes for direct static imaging and ultrafast motion detection. *Light Sci Appl* **13**, 256 (2024).

14. Zhou Y, Sun ZB, Ding YC et al. An ultrawide field-of-view pinhole compound eye using hemispherical nanowire array for robot vision. *Sci Robot* **9**, eadi8666 (2024).

15. Wang L, Gong W, Cao XW et al. Holographic laser fabrication of 3D artificial compound μ-eyes. *Light Adv Manuf* **4**, 26 (2023).

16. Zhou KC, Harfouche M, Cooke CL et al. Parallelized computational 3D video microscopy of freely moving organisms at multiple gigapixels per second. *Nat Photonics* **17**, 442–450 (2023).

17. Afshari H, Popovic V, Tasci T et al. A spherical multi-camera system with real-time omnidirectional video acquisition capability. *IEEE Trans Consum Electron* **58**, 1110–1118 (2012).

18. Cheng Y, Cao J, Zhang YK et al. Review of state-of-the-art artificial compound eye imaging systems. *Bioinspir Biomim* **14**, 031002 (2019).

19. Brady DJ, Gehm ME, Stack RA et al. Multiscale gigapixel photography. *Nature* **486**, 386–389 (2012).

20. Llull P, Bange L, Phillips Z et al. Characterization of the AWARE 40 wide-field-of-view visible imager. *Optica* **2**, 1086–1089 (2015).

21. Fan JT, Suo JL, Wu JM et al. Video-rate imaging of biological dynamics at centimetre scale and micrometre resolution. *Nat Photonics* **13**, 809–816 (2019).

22. Xie H, Han XF, Xiao GH et al. Multifocal fluorescence video-rate imaging of centimetre-wide arbitrarily shaped brain surfaces at micrometric resolution. *Nat Biomed Eng* **8**, 740–753 (2024).

23. Yuan XY, Ji MQ, Wu JM et al. A modular hierarchical array camera. *Light Sci Appl* **10**, 37 (2021).

24. Wu Q, Zhang SS, Liao WX et al. Heterogeneous compound eye camera for dual-scale imaging in a large field of view. *Opt Express* **30**, 45143–45155 (2022).

25. Thiele S, Arzenbacher K, Gissibl T et al. 3D-printed eagle eye: compound microlens system for foveated imaging. *Sci Adv* **3**, e1602655 (2017).

26. Toulouse A, Drozella J, Motzfeld P et al. Ultra-compact 3D-printed wide-angle cameras realized by multi-aperture freeform optical design. *Opt Express* **30**, 707–720 (2022).

27. Chia YH, Liao WH, Vyas S et al. In vivo intelligent fluorescence endo-microscopy by varifocal meta-device and deep learning. *Adv Sci* **11**, 2307837 (2024).

28. Liu C, Jiang Z, Wang X et al. Continuous optical zoom microscope with extended depth of field and 3D reconstruction. *PhotoniX* **3**, 20 (2022).

29. Shi RH, Chen XY, Deng JH et al. Random-access wide-field mesoscopy for centimetre-scale imaging of biodynamics with subcellular resolution. *Nat Photon* **18**, 721–730 (2024).

30. Mazzella L, Mangeat T, Giroussens G et al. Extended-depth of field random illumination microscopy, EDF-RIM, provides super-resolved projective imaging. *Light Sci Appl* **13**, 285 (2024).

31. Liu C, Zheng Y, Yuan RY et al. Tunable liquid lenses: emerging technologies and future perspectives. *Laser Photonics Rev* **17**, 2300274 (2023).

32. Guo CL, Blair GJ, Sehgal M et al. Miniscope-LFOV: a large-field-of-view, single-cell-resolution, miniature microscope for wired and wire-free imaging of neural dynamics in freely behaving animals. *Sci Adv* **9**, eadg3918 (2023).

33. Rivenson Y, Göröcs Z, Günaydin H et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).

34. Liu SB, Xie BK, Yuan RY et al. Deep learning enables parallel camera with enhanced-resolution and computational zoom imaging. *PhotoniX* **4**, 17 (2023).

35. Zhang K, Liang JY, Van Gool L et al. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of IEEE/CVF International Conference on Computer Vision* 4771–4780 (IEEE, 2021). http://doi.org/10.1109/ICCV48922.2021.00475.

36. Weiss LE, Shalev Ezra Y, Goldberg S et al. Three-dimensional localization microscopy in live flowing cells. *Nat Nanotechnol* **15**, 500–506 (2020).

37. Liu S, Chen J, Hellgoth J et al. Universal inverse modeling of point spread functions for SMLM localization and microscope characterization. *Nat Methods* **21**, 1082–1093 (2024).

38. Belthangady C, Royer LA. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat Methods* **16**, 1215–1225 (2019).

39. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition* 7132–7141 (IEEE, 2018). http://doi.org/10.1109/CVPR.2018.00745.

40. Ultralytics. YOLOv5 (Version v3.0) [Computer software]. Zenodo (2020). https://doi.org/10.5281/zenodo.3908559.

41. Peng FL, Chen HW, Gou FW et al. Analytical equation for the motion picture response time of display devices. *J Appl Phys* **121**, 023108 (2017).

42. Brady DJ, Pang W, Li H et al. Parallel cameras. *Optica* **5**, 127–137 (2018).

43. Li ZS, Liu C, Li XW et al. Real-time holographic camera for obtaining real 3D scene hologram. *Light Sci Appl* **14**, 74 (2025).

44. Wang D, Li YL, Zheng XR et al. Decimeter-depth and polarization addressable color 3D meta-holography. *Nat Commun* **15**, 8242 (2024).

45. Zhan T, Zou JY, Xiong JH et al. Practical chromatic aberration correction in virtual reality displays enabled by cost-effective ultra-broadband liquid crystal polymer lenses. *Adv Opt Mater* **8**, 1901360 (2020).

46. Luo ZY, Li YNQ, Semmen J et al. Achromatic diffractive liquid-crystal optics for virtual reality displays. *Light Sci Appl* **12**, 230 (2023).

47. Wang S, Liu XX, Li YY et al. A deep learning-based stripe self-correction method for stitched microscopic images. *Nat Commun* **14**, 5393 (2023).

48. Liang JY, Cao JZ, Sun GL et al. SwinIR: image restoration using Swin Transformer. In *Proceedings of IEEE/CVF International Conference on Computer Vision Workshops* 1833–1844 (IEEE, 2021). http://doi.org/10.1109/ICCVW54120.2021.00210.

49. Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the 14th European Conference on Computer Vision* 694–711 (Springer, 2016). http://doi.org/10.1007/978-3-319-46475-6_43.

50. Isola P, Zhu JY, Zhou TH et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 5967–5976 (IEEE, 2017).http://doi.org/10.1109/CVPR.2017.632.

Author contributions

Y. Zheng and H. R. Zhang proposed the idea and performed the experiments. X. W. Li, Y. R. Zhao, Z. S. Li and Y. H. Hou analyzed the data. C. Liu and Q. H. Wang conceived the project. All authors discussed the results and commented on the paper. Y. Zheng and H. R. Zhang contributed equally to this work.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at
https://doi.org/10.29026/oea.2025.240285

Scan for Article PDF