Computational Imaging and Artificial Intelligence: The Next Revolution of Mobile Vision

This article reviews the history of mobile vision and surveys the use of artificial intelligence for computational imaging.

By Jinli Suo, Weihang Zhang, Jin Gong, Xin Yuan, Senior Member IEEE, David J. Brady, Fellow IEEE, and Qionghai Dai, Senior Member IEEE

ABSTRACT | Signal capture is at the forefront of perceiving and understanding the environment; thus, imaging plays a pivotal role in mobile vision. Recent unprecedented progress in artificial intelligence (AI) has shown great potential in the development of advanced mobile platforms with new imaging devices. Traditional imaging systems based on the “capturing images first and processing afterward” mechanism cannot meet this explosive demand. On the other hand, computational imaging (CI) systems are designed to capture high-dimensional data in an encoded manner to provide more information for mobile vision systems. Thanks to AI, CI can now be used in real-life systems by integrating deep learning algorithms into the mobile vision platform to achieve a closed loop of intelligent acquisition, processing, and decision-making, thus leading to the next revolution of mobile vision. Starting from the history of mobile vision using digital cameras, this work first introduces the advancement of CI in diverse applications and then conducts a comprehensive review of current research topics combining CI and AI. Although new-generation mobile platforms, represented by smart mobile phones, have deeply integrated CI and AI for better image acquisition and processing, most mobile vision platforms, such as self-driving cars and drones only loosely connect CI and AI, and are calling for a closer integration. Motivated by this fact, at the end of this work, we propose some potential technologies and disciplines that aid the deep integration of CI and AI and shed light on new directions in the future generation of mobile vision platforms.

KEYWORDS | Artificial intelligence (AI); autonomous driving; brain science; cameras; cloud computing; computational imaging (CI); deep learning; edge cloud; edge computing; machine learning; machine vision; mobile vision; neural networks; optics; self-driving.

I. INTRODUCTION

Significant changes have occurred in mobile vision over the past few decades. Inspired by the recent advances in artificial intelligence (AI) and the emerging field of computational imaging (CI), we are heading to a new era of mobile vision, where we anticipate deep integration of CI and AI. New mobile vision devices equipped with intelligent imaging systems will be developed and deployed widely in our daily lives.

The history of mobile vision using digital cameras [1] can be traced back to the 1990s when digital event data recorders (EDRs) were first installed on vehicles [2].

Interestingly, almost simultaneously, cell phones with built-in cameras were invented [3]. Following this, the laptop began to embed cameras inside [4]. The emergence of mobile vision and the explosion of digital data at that time set the stage for the beginning of the “big data” era.

Manuscript received 18 September 2021; revised 28 March 2023 and 6 October 2023; accepted 28 November 2023. Date of publication 12 December 2023. Date of current version 22 December 2023. This work was supported by the National Natural Science Foundation of China under Grant 61931012 and Grant 62088102. (Corresponding authors: Jinli Suo; Xin Yuan; Qionghai Dai.)
(Jinli Suo and Weihang Zhang contributed equally to this work.)

Jinli Suo is with the Department of Automation and the Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China, and also with the Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China (e-mail: jlsuo@tsinghua.edu.cn).

Weihang Zhang, Jin Gong, and Qionghai Dai are with the Department of Automation and the Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China (e-mail: zwh19@mails.tsinghua.edu.cn; gong20@mails.tsinghua.edu.cn; qh dai@tsinghua.edu.cn).

Xin Yuan is with the Research Center for Industries of the Future and School of Engineering, Westlake University, Hangzhou, Zhejiang 310030, China (e-mail: xyuan@westlake.edu.cn).

David J. Brady is with the Wyant College of Optical Sciences, The University of Arizona, Tucson, AZ 85721 USA (e-mail: djbrady@arizona.edu).

Digital Object Identifier 10.1109/JPROC.2023.3338272
EDR, computers, and cell phones are the three pillars of mobile vision that play a role in the battlefront to capture digital data.

We term this the **first revolution**, i.e., the inception of modern mobile vision. In this generation, the mode of mobile vision was to **capture and save on the device**. The main goal was to capture the data.

Equipped with the Internet and mobile communication, mobile vision is developing rapidly and revolutionizing communications and commerce. Examples of representative applications include web conferencing, online shopping, and social networks. The completion of the chain of capture, encoding, transmission, and decoding is what we refer to as the **second revolution** of mobile vision. This has been further advanced by the wide deployment of the fourth generation (4G) of broadband cellular network technology since 2009 [5]. The mode of mobile vision has then been changed to **capture, transmit, and share images**. Meanwhile, new mobile devices, such as tablets and wearable devices, have also been invented. Based on this, extensive social media have been developed and advanced. Sharing has become the primary role in this generation of mobile vision.

Although the second mobile vision revolution has brought about enormous changes to our lifestyle, it has also brought about unexpected challenges. For instance, the bandwidth is usually limited, while the end users need to capture more data. On the other hand, although a significant amount of data is captured, the growth of information is limited. For instance, most videos and images shared on social media are barely watched in detail, especially at the pixel level. For example, even though the mobile phone camera resolution is already very high, the area we are interested in will not be clearer or contain more details than the broad background where we pay no attention, which, on the other hand, frequently occupies the storage space of the mobile phone. In the same photograph, the size may be about several megabits per second or even more than 10 Mb/s. This poses two questions.

1. How to capture more information?
2. How to extract useful information from the data?

These two questions can be addressed by CI and AI, respectively, which are the two main topics in this study. Moreover, we take one step further to answer both questions in one shot.

1. Can we build new devices to capture more useful information and directly use the information to perform the desired task and maximize performance?

The answer to this question leads to the **next (third) revolution** of mobile vision, which will integrate CI and AI deeply and make the imaging systems a combination of intelligent acquisition, processing, and decision-making in high-speed modes. Mobile vision will now be in the mode of **intelligent vision**.

In this third revolution, using AI, we can build imaging systems that are optimized to functional metrics such as detection accuracy and recognition rate, not just the amount of information they capture. One function of the third revolution will still be to communicate and share images, but rather than sharing the actual physical image, the data are mined and transformed by AI to be optimized for communication. Similarly, the imaging system can understand the intent of its use in different situations, for example, to measure objects in a scene, to help quantify objects, to identify an object, and to help with safety. The system tries to intelligently use resources for task-specific rewards. Rather than maximizing data or information, it maximizes performance. These new imaging devices are referred to as **CI systems** in this study.

Equipped with this new revolution and the rapid development of machine vision and AI, we anticipate that intelligent mobile vision platforms, such as self-driving vehicles, drones, and autonomous robots, will gradually become a reality. With advances in AI, the boundary between mobile vision and intelligent imaging is becoming blurry. When applied to mobile platforms, intelligent imaging systems need to consider the load, available resources, and granted response time. These mobile vision platforms aim to replace or even improve the ability of humans to perform specific tasks, such as navigation, transportation, and tracking. In summary, the operations of mobile intelligent platforms mainly include five stages: **acquisition, communication, storage, processing, and decision-making**. In other words, these systems capture real-time visual information from the external environment, fuse more information through communication and sharing, digest and process the information, make timely decisions, and successively take corresponding actions. Meanwhile, emerging topics and techniques will support and promote the development of new mobile vision platforms. For instance, big data technology can help process and analyze a large amount of collected data within an allocated time period. Edge (cloud) computing technology can be used to access powerful and reliable distributed computing resources for data storage and processing. In addition, the fifth generation (5G) of mobile communication technology applied to smartphones would support high-speed and high-bandwidth data transmission.

Table 1 summarizes the representative events of the three revolutions in mobile vision described above. Notably, although intelligent vision has recently been advanced significantly by AI in the third revolution, automation and intelligent systems have been long-time desires of human beings. For example, the idea of using AI and cameras to control mobile robots dates back to the 1970s [6], and the use of video cameras and computer software to develop self-driving cars was initiated in 1986 [7].

### A. Current Status

Most existing mobile vision systems are a combination of a camera and a digital image processor. The former maps the high-dimensional continuous visual information into
Table 1 Three Revolutions of Mobile Vision Systems

| Figures | Events |
|---------|--------|
| **First revolution:** capture, save on device | 1990s  
  - Popularization of digital event data recorders 1999-2000  
  - Invention of cell phones with built-in camera (Kyocera/Samsung SCH-V200/Sharp J-SH04) 2006  
  - Invention of laptop with built-in webcam (Apple MacBook Pro) |
| **Second revolution:** capture, transmit and share images | 2000s  
  - Booming of video conferences and instant video messaging (Skype/Chat)  
  - Popularization of social networks sharing images and videos (Friendster/MySpace/Twitter/Instagram, etc.) |
| **Third revolution:** Intelligent vision | Ongoing process, with preliminary research studies:  
  - Connected and autonomous vehicles (CAVs) using computational imaging systems  
  - Robotic vision using computational cameras |

a 2-D digital image, during which only a small portion of the information is recorded such that the image systems are of low throughput. The latter extracts semantic information from the recorded measurements, which are intrinsically redundant and limited for decision-making and following actions. Both image acquisition (cameras) and the subsequent processing algorithms face significant challenges in real applications, including insufficient image quality, unstable algorithms, long running time, and limited resources such as memory, processing speed, and transmission bandwidth. Fortunately, we have witnessed the rapid development in the camera industry and AI. The progress in other related fields, such as cloud computing and 5G, is potentially providing solutions for these resource limitations.

**B. Computational Imaging**

In contrast to the above “capturing images first and processing afterward” regime, the emerging computation imaging technique [8], [9] provides a new architecture integrating these two steps (capturing and processing) to improve the perception capability of imaging systems (Fig. 1). Specifically, CI combines optoelectronic imaging with algorithmic processing and designs new imaging mechanisms to optimize the performance of computational analysis. Complex algorithms created for image estimation must regulate the parameters settings, such as focus, exposure, and illumination, in order to integrate computing into the imaging setup [10]. Benefiting from the joint design, CI can better extract the scene information by optimized encoding and corresponding decoding. Such architecture might help mobile vision in scenarios challenging conventional cameras, such as low-light imaging, non-line-of-sight (NLOS) object detection, and high-frame-rate acquisition.

**C. Incubation of Joint CI and AI**

In CI systems, sensor measurements are usually recorded in an *encoded* manner; therefore, some complex decoding algorithms are required to retrieve the desired signal. AI has been widely used in the past decade due to its high performance and efficiency, which can be a good option for fast high-quality decoding. In addition, AI has achieved significant success in various computer vision tasks and might even inspire researchers to design new CI systems for reliable perception with compact structures.

Recently, various AI methods represented by deep neural networks that use large training datasets to solve inverse problems in imaging [11], [12], [13] have been proposed to conduct high-level vision tasks, such as detection, recognition, and tracking. On the one hand, this has advanced the field of CI to maximize the performance for specific tasks, rather than just capturing more information, as mentioned earlier. On the other hand, this scheme is different from the human vision system. As the fundamental objective of the human vision system is to perceive and interact with the environment, the vision cortex extracts semantic information from encoded visual information, and the intermediate image might be skipped/unnecessary, leading to a higher processing efficiency. Hence, it is of great significance to develop new visual information encoding and decoding informative semantics from the raw data, which is also the main goal of CI. This also shares the same spirit as smart cameras [10], i.e., to intelligently use resources of task-specific rewards to optimize the desired performance.

In recent years, smartphones, as the most popular mobile imaging platform, have been fueling renaissance in the applications of optimized CI techniques and AI algorithms. Outstanding CI methods have been applied
to video cameras equipped on mobile phones, including high-dynamic-range (HDR) imaging/HDR+ [14], depth estimation [15], and feature recognition [16]. On the other hand, the customized silicon chips [17] and platform-aware neural networks [18], [19], [20] have been developed, paving the way for efficient implementation of AI algorithms. Because both AI and CI are developing rapidly and diverse mobile platforms urgently require advanced intelligent vision systems, the convergence of CI and AI is expected to play a significant role in the next revolution of mobile vision platforms.

In this comprehensive survey, we will review the early advances and ongoing topics integrating CI and AI, and outlook at the trends and future work in mobile vision, hoping to provide highlights for the research of the next-generation mobile vision systems. Specifically, after reviewing existing representative works, we propose the prospect of next-generation mobile vision based on the fact that the combination of CI and AI has made great progress but still cannot be fully used on mobile platforms such as self-driving cars and unmanned aerial vehicles (UAVs).

CI systems have started being used in many fields of our daily life (Section II), and AI algorithms also play key roles in each stage of the CI process (Section III). It is promising to leverage the developments of these two fields to advance the next-generation mobile vision systems. Based on this, in Section IV, we introduce several key technologies for image perception, information exchange, and decision-making, as well as some related emerging research disciplines that can be used for the development of CI and AI to revolutionize mobile vision systems in the future.

II. EMERGING AND FAST DEVELOPING CI

Conventional digital imaging maps high-dimensional continuous information to 2-D discrete measurements via projection, analog-to-digital conversion, and quantization. After decades of development, digital imaging has achieved great success and promoted a series of related research fields, such as computer vision and digital image processing [21]. The processing line is shown at the top of Fig. 1, and related image-processing technologies, such as pattern recognition, have significantly advanced the development of machine vision. However, due to the inherent limitation of this “capturing images first and processing afterward” scheme, machine vision has encountered bottlenecks in both imaging and successive processing, especially in real applications. The dilemma is mainly due to the limited information capture capability, since during the (image) capture, a large amount of information is filtered out (e.g., spectrum, depth, and dynamic range), and the subsequent processing is thus highly ill-posed.

CI, by contrast, considers the imaging system and the consequent processing in an integrated manner. This new scheme blurs the boundary between optical acquisition and computational processing and moves the calculation forward to the imaging procedure, i.e., introduces computing to design new task-specific systems. In particular, as shown at the bottom of Fig. 1, in some cases, CI changes the illumination and optical elements (including apertures, light paths, and sensors) to encode the visual information and then decodes the signal from the coded measurements computationally. In general, the encoding elements in a CI setup aim to generate coded measurement(s) of the high dimensional but redundant data in a compact manner, such as a single snapshot or a few measurements. Take the snapshot compressive imaging (SCI) [22] as an example. Denoting the target visual data as $X \in \mathbb{R}^{H \times W \times N}$, a single-shot coded measurement $Y \in \mathbb{R}^{H \times W}$ in CI methods can be generally described as a linear equation system

$$Y = \sum_{k=1}^{N} M_k \odot X_k + E$$

where $M_k \in \mathbb{R}^{H \times W}$ depicts the modulation pattern encoding kth slice of the target data $X_k$, $\odot$ denotes the Hadamard (elementwise) product, and $E \in \mathbb{R}^{H \times W}$ is the measurement noise. Equation (1) can also be simplified in vector form as

$$y = Hx + e$$

in which the sensing matrix $H \in \mathbb{R}^{HW \times HWN}$ is

$$H = [\text{diag}(\text{vec}(M_1)), \ldots, \text{diag}(\text{vec}(M_N))]$$

$x \in \mathbb{R}^{HWN}$ is the concatenation of the vector form of each slice of the target data, and $y, e \in \mathbb{R}^{HW}$ are the vectorized representation of $Y$ and $E$, respectively. The encoding process can encompass more visual information than conventional imaging systems or task-oriented information for better post-analysis. In some cases, CI generalizes multiplexed imaging to high dimensions of the plenoptic function [23]. On the other end, the decoding process generally aims to recover the latent visual data from its coded measurement(s) and is mathematically an ill-posed
inversion problem that can be solved by optimizing the following problem:

\[
\hat{x} = \arg\min_x \frac{1}{2} \|y - Hx\|^2 + \lambda R(x).
\] (4)

Here, the first term is defined to fit the forward encoding process of the CI imaging model described in (2), \(R(x)\) denotes the regularization term to constrain the solution space, and \(\lambda\) is a balancing parameter. For diverse CI systems, the encoding \(Hx\) and regularizer \(R(x)\) take different forms but conform to the same or similar framework.

The innovation of CI (by integrating sensing and processing) has strongly promoted the development of imaging research and produced a number of representative research areas, which have significantly enhanced the imaging capabilities of vision systems, especially mobile vision. In the following, we show some representative examples: depth, high speed, hyperspectral, wide field-of-view (FOV), HDR, NLOS, radar, super-resolution (SR), single pixel, through fog, and low-light imaging. In addition, the CI systems reviewed in this study are more general than conventional cameras used for photography at the visible wavelength. All these progresses demonstrate the advantages of CI and the large potential of using CI in mobile vision systems. As mentioned previously, CI systems are not limited to visible wavelengths for photography purposes. In this study, we consider diverse systems aimed at capturing images for mobile vision in the future. We summarize the recent advances in representative CI in Table 2 and describe some research topics in the following; however, this list is by no means exhaustive.

A. Depth Imaging

A depth image refers to an image with intensity describing the distance from the target scene to the camera, which directly reflects the geometry of the (usually visible) surface of the scene. As an extensively studied topic, there exist various depth imaging methods, including light detection and ranging (LiDAR), stereo vision imaging, and structured illumination methods [24]. In the following, we list some examples.

1) Stereo vision or binocular systems might be the first passive and well-studied depth imaging technique. It has a solid theory based on geometry [25] and is widely used in current CI systems [26].

2) Structured illumination, as an active method, has been widely used in our daily life for high-precision and high-speed depth imaging [27]. Researchers have developed different structured illumination methods for CI systems [28], [29], [30]. Among them, the common principle is to project a certain pattern \(M^0\) onto the target scene and retrieve the scene depth from the depth-dependent pattern distortion, with the distortions calibrated by capturing the intensity distribution of the same mask projected at different depths \(z\).

\[
M^*(x, y) = M^0 \left( \frac{f_p}{z} x + \frac{f_p}{z} d, \frac{f_p}{z} y \right) \] (5)

where \(f_p\) is the projector’s focal length and \(d\) is the lateral offset between the camera and the projector [28].

3) Light-field imaging is another way to perform depth imaging. In this case, depth is calculated from the 4-D light field, in addition to confining the scene information to 3-D spatial coordinates [61]. Ng et al. [31] proposed a plenoptic camera that captures 4-D light field with a single exposure, which is of compact structure, portable, and can also help refocusing. For a 4-D light field [61], each point in a 3-D scene can be projected into a series of 2-D spatial–angular slices, i.e., epipolar plane images (EPIs), in which its depth is linearly related to the limit slope [62], [63] and can be inferred directly. The community has witnessed different variants sharing the same spirit with increasing accuracy [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74]. To cope with the time-consuming calculation via exhaustive estimation of each EPI, deep learning methods are proposed to achieve faster speed and higher accuracy [75], [76], [77]. With CI, a light-field camera has also been employed to overcome the spatial–angular resolution tradeoff [32] and extended to capture dynamic and transient scenes [33], [34], [35].

4) Depth from focus/defocus estimates the 3-D surface of a scene from a set of two or more images of that scene. It can be implemented by changing the camera parameters [36], [37], [38], [39], [40] or using coded apertures [41], [42], [43].

5) Time-of-flight (ToF) cameras obtain the distance of the target object by detecting the round-trip time of the light pulse. Recently, Kazmi et al. [44] have verified through experiments that ToF can provide accurate depth data with a high frame rate under suitable conditions. In conventional pulse modulation in ToF, a photosensitive unit (such as a photodiode) in each pixel is used at the receiving end to control two capacitors to store the reflected charges, and the depth of the reflection is determined by the ratio of the charges in the capacitors [80]. However, this pulse-based method requires high-quality hardware and cannot eliminate the influence of ambient light. To address this challenge, continuous wave modulation is commonly used, whereas sinusoidal modulation is a simple and general approach. Assuming that the modulation frequency of the emitted sinusoidal signal is \(f\) and the four received signals sampled equidistantly in a sinusoidal period are \(r_i, i = 0, 1, 2, 3\), then the phase offset can be deduced as

\[
\Delta \phi = \phi_0 + \arctan2(r_2 - r_0, r_1 - r_3) \] (6)
Table 2: Representative Works of CI Research. We Emphasize That This List Is by No Means Exhaustive.

| CI Topic | Methods | Applications | References |
|----------|---------|--------------|------------|
| Depth    | • Binocular system obtains depth information through the parallax of dual cameras.  
          • Structured light projects given coding patterns to improve the feature-matching effect.  
          • Light-field imaging calculates the depth from the four-dimensional light field.  
          • Depth from focus/defocus estimates the depth from relative blurring.  
          • Time of flight and LiDAR continuously transmit optical pulses and detect the flight time. | 1. Self-driving  
2. Entertainment  
3. Security system  
4. Remote sensing | [24]-[81] |
| Motion   | • Coded exposure is widely applied to capture fast motions by recovering a sequence of frames from their encoded combination.  
          • Snapshot imaging uses a spatial light modulator to modulate the high-speed scenes and develops optimization algorithms, end-to-end neural networks, or plug-and-play frameworks. | 1. Surveillance  
2. Sports  
3. Industrial inspection  
4. Visual navigation | [26], [82]-[95] |
| Spectrum | • Hyperspectral imaging combines imaging technology and spectral technology to detect the 2D geometric space and one-dimensional spectral information of the target and obtain the continuous and narrow band image data of hyperspectral resolution. In the spectral dimension, the image is segmented among not only R, G, and B, but also many channels in the spectral dimension, which can feed back continuous spectral information. | 1. Agriculture  
2. Food detection  
3. Mineral testing  
4. Medical imaging | [96]-[109] |
| Field-of-view | • Large field-of-view imaging methods design a multi-scale optical imaging system, in which multiple small cameras are placed in different fields of view to segment the image plane of the whole field and the large field-of-view image is stitched through subsequent processing. | 1. Security  
2. Agriculture  
3. Navigation  
4. Environment monitoring | [93], [110]-[113] |
| Dynamic range | • The spatial light modulator based method performs multiple-length exposures of images in the time domain.  
          • Encoding-based methods encode and modulate the exposure intensity of different pixels in the spatial domain. Different control methods have been used to improve the dynamic range through feedback control. | 1. Digital photography  
2. Medical imaging  
3. Remote sensing  
4. Exposure control | [114]-[124] |
| NLOS     | • Non-Line-Of-Sight (NLOS) imaging can recover images of the scene from the indirect light.  
          • Active NLOS uses a laser as source and single-photon avalanche diode (SPAD) or streak camera as the receiver.  
          • Passive NLOS uses a normal camera to capture the scattered light and uses algorithms to reconstruct the scene. | 1. Self-driving  
2. Military reconnaissance | [125]-[171] |
| Radar    | • Synthetic Aperture Radar (SAR) uses the reflection of the ground target on the radar beam as the image information formed by the backscattering of the ground target. SAR is an active side-view radar system and its imaging geometry is based on oblique projection.  
          • Through-the-wall radar can realize motion detection without cameras.  
          • WiFi has been used for indoor localization. | 1. Drones  
2. Military inspection  
3. Elder care | [172]-[190] |
| Super resolution | • Super-resolution imaging aims to reconstruct a high-resolution image through one or a few low-resolution images of the same scene.  
          • The sub-pixel difference between multiple low-resolution images can be utilized to retrieve high-resolution details.  
          • Various optimization methods have been proposed to infer lost details by solving an ill-posed linear system.  
          • Deep learning architectures have been developed for super-resolution from a single snapshot in order to raise the efficiency and adapt for dynamic scenes. | 1. Remote sensing  
2. Surveillance  
3. Astronomical observation  
4. Microscopy | [191]-[215] |
where \( \phi_0 \) is the phase of the emitted signal at the first sampling time, and then, the estimated depth is

\[
d = \phi_0 + \frac{c \times \Delta \phi}{4\pi f}
\]

(7)

where \( c \) denotes the light speed. Similarly, the continuous wave modulation obtains wider application by eliminating the error brought by the hardware or ambient light [81]. Doppler ToF imaging [45] and CI with multicamera ToF systems [46] have also been built recently. The end-to-end optimization of imaging pipelines [47], [48] for ToF further overcomes interference using deep learning. Based on the ToF principle, but with sensor arrays, LiDAR has been widely used in autonomous driving and remote sensing; it can be categorized into two classes according to the measurement modes: scanning-based and nonscanning-based. The former obtains the real-space image of the target by point-by-point (or line-by-line) scanning with a pulsed laser, while the latter covers the entire field of the target scene and obtains the 3-D map in a single exposure using a pulsed flash laser imaging system [49]. Kirmani et al. [50] proposed to use the first photon received by the detector for high-quality 3-D image reconstruction, applying computational algorithms to LiDAR imaging in the low-flux environment.

In addition to the above-categorized depth imaging approaches, single-photon sensors have recently been used in CI for depth estimation [51], [52], [53]. In computer vision, single-image 3-D is another research topic [54], [55], [56], [57]. By adapting local monocular cameras to the global camera, deep learning helps solve the joint depth estimation of multiscale camera arrays and panoramic object rendering in order to provide higher 3-D image resolution and better quality virtual reality (VR) experience [59]. Another idea is depth estimation and 3-D imaging based on polarization. As the light reflected off the object has a polarization state corresponding to the shape, Ba et al. [58] input the polarization image into a neural network to estimate the surface shape of the object and establish a polarization image dataset. Regarding the applications, depth imaging has been widely used in our daily lives, such as in self-driving vehicles and entertainment. Recently, face-identification systems have started using depth information for security considerations. Depth imaging systems with varying resolutions ranging from meters to micrometers have been widely used in remote sensing and industrial inspection.

B. High-Speed Imaging

For an imaging system, the frame rate is a tradeoff between the illumination level and sensor sensitivity. An
insufficient frame rate results in missing events or motion blur. Therefore, high-speed imaging is crucial for recording highly dynamic scenes. In CI, temporal resolution can be improved using temporal compressive coding.

Hitomi et al. [82] employed a digital micromirror device (DMD) to modulate a high-speed scene and used a dictionary learning method to reconstruct a video from a single compressed image with high spatial resolution. A liquid crystal on silicon (LCOS) device was used for modulation in [83] to achieve temporal compressive imaging. Llull et al. [84] used mechanical translation of a coded aperture for the encoded acquisition of a sequence of time stamps, dubbed coded aperture compressive temporal imaging (CACTI), in which a static mask is placed at the image plane for pixelwise modulation, and the modulated image is then relayed to the sensor. Frame-dependent modulation is achieved by moving the mask mechanically within the exposure elapse. According to (1), the coded snapshot \( Y \) aggregates the modulated images in the form of

\[
Y(x, y) = \sum_{k=1}^{N} M(x - s(t_k), y) X_k(x, y) + E(x, y) \quad (8)
\]

where \( s(t_k) \) is the mechanical displacement at the time slot corresponding to the \( k \)th encoded frame, and \( N \) images can be retrieved from the single measurement by solving (4). Deng et al. [85] used the limited spatiotemporal distribution of the above-coded measurements and further increased the frame rate by a novel compressive video acquisition technique, termed sinusoidal one and successfully multiplexed a group of coded measurements within a snapshot. Another representative work is by Liu et al. [86], who implemented coded exposure on a complementary metal–oxide–semiconductor (CMOS) image sensor with an improved control unit instead of a spatial light modulator (SLM) and conducted reconstruction by utilizing the sparse representation of the video patches using an overcomplete dictionary. All these methods increase the imaging speed while maintaining a high spatial resolution. Solely aiming to capture high-speed dynamics, the streak camera is a key technology for obtaining femtosecond photographs [90]. Combining streak cameras with compressive sensing (CS), compressed ultrafast photography (CUP) [87] was developed to achieve billions of frames per second [88] and has been extended to other dimensions [89]. However, there might be a long way to implement CUP on mobile platforms.

High-speed imaging has wide applications in sports, surveillance, industrial inspection, and visual navigation.

C. Hyperspectral Imaging

Hyperspectral imaging technology aims to capture the spatiotemporal data cube of a target scene by combining imaging with spectroscopy. Conventional hyperspectral imaging methods include spectrum splitting by grating, acousto-optic tunable filter (AOTF) or prism. Subsequently, separate recordings are made through scanning, which limits the imaging speed and makes them inapplicable to capture dynamic scenes. Since the hyperspectral data are intrinsically redundant, compressive sensing theory [96], [97] provides a solution for compressive spectral sampling. The coded aperture snapshot spectral imaging (CASSI) system proposed by Wagadarikar et al. [105] uses a 2-D detector to compressively sample 3-D (spatiotemporal) information by encoding, dispersing, and integrating the spectral information within a single exposure. Unlike CACTI, wavelength-dependent modulation is implemented by shearing different channels of the scene’s image by a dispersive element. Therefore, the forward model for the pixel at spatial location \((x, y)\) can be specified from (1) as

\[
Y(x, y) = \sum_{k=1}^{N} M(x - \alpha(\lambda_k - \lambda_c), y) \times X_k(x - \alpha(\lambda_k - \lambda_c), y) + E(x, y) \quad (9)
\]

in which \( N \) is the number of retrieved spectral channels, \( \lambda_k \) is the central wavelength of the \( k \)th channel, and \( \alpha \) and \( \lambda_c \) are the linear dispersion and center wavelength of the dispersive element, respectively. Subsequently, other types of methods achieve hyperspectral imaging by designing and improving masks. For a compact design, Zhao et al. [98] proposed a method for acquiring coded images using a random-colored mask by a consumer-level printer in front of the camera sensor and computationally reconstructed the hyperspectral data cube using a deep-learning-based algorithm. Zhang et al. [99] also proposed to manufacture a thin film mask with low cost by conventional film photography and achieve tens of megapixel spatial resolution. In addition, the hybrid camera systems [100], [101], which integrate high-resolution RGB video and low-resolution (LR) multispectral video, have also been developed. For a more comprehensive review of hyperspectral CI systems, refer to the survey by Cao et al. [102]. Another related work is to recover the hyperspectral images from an RGB image by learning-based methods [103], which is inspired by the deep generative models [104].

Equipped with rich spectral information, hyperspectral imaging has wide applications in agriculture, food detection, mineral detection, and medical imaging [109].

D. Wide-FOV High-Resolution Imaging

Wide FOV CI with both high spatial and temporal resolutions is an indispensable tool for vision systems operating in large-scale scenes. On the one hand, high resolution helps observe details, distant objects, and small changes, while a wide FOV helps observe the global structure and connections in-between objects. However, conventional imaging equipment has long been constrained by...
the tradeoff between FOV and resolution due to limited spatial bandwidth product [110]. With the development of high-resolution camera sensors, geometric aberrations fundamentally limit the resolution of the camera in a wide FOV. In order to solve this problem, Cossairt et al. [111] proposed a method to correct aberrations by calculations. Specifically, they use a simple optical system containing a ball lens cascaded by hundreds of cameras to achieve a compact imaging architecture and realize gigapixel imaging. Brady et al. [112] proposed a multiscale gigapixel imaging system by using a shared objective lens. Instead of using a specific system, Yuan et al. [113] integrated multiscale scene information to generate gigapixel wide-field video. A joint FOV and temporal compressive imaging system are built in [93] to capture high-speed multiple FOV scenes. In the spatiotemporal compressive scheme, the random modulation is introduced by a DMD, and the single measurement simultaneously records two modulated scenes with a slight lateral displacement from each other. Therefore, the extended FOV combined with two dynamic scenes is achieved by solving the following problem with the same form as (1):

$$Y = \sum_{k=1}^{N} M_k \otimes X_{1,k} + \tilde{M}_k \otimes \tilde{X}_{2,k} + E$$ (10)

where \(\tilde{M}\) and \(\tilde{X}\) denote the shifted mask and scene, respectively. Large-FOV imaging systems have wide applications in environmental pollution monitoring, security cameras, and navigation, especially for multiple object recognition and tracking.

E. HDR Imaging

There often exists a drastic range of contrast in a scene that goes beyond a conventional camera sensor. One widely used method is to synthesize the final HDR image according to the best details from a series of low-dynamic-range (LDR) images with different exposure elapses, which improves the imaging quality. Nayar and Mitsunaga [114] proposed the use of an optical mask with spatially varying transmittance to control the exposure time of adjacent pixels before the sensor and reconstruct the HDR image using an effective algorithm. With the saturated and low-intensity values discarded, the remaining pixels \(I(x, y)\), i.e., the on-grid values, are normalized by the respective transmittance on the mask and then used to calculate the off-grid values \(I_0(x, y)\) according to

$$I(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3} f(1.5 - m, 1.5 - n) \times I_0(x + 1.5 - m, y + 1.5 - n)$$ (11)

where \(f\) can be the cubic convolution kernel, and the pseudoinverse can be employed to solve the vector form of the problem. Finally, the missing on-grid values were derived from the off-grid values. With the development of SLM, the authors further proposed an adaptive method to enhance the dynamic range of the camera [119]. Most recently, deep learning has been used in the single-shot HDR system design [115], [116]. With these approaches, CI techniques are able to achieve decent recording within both ends of a large radiance range [117]. Researchers have used the proportional–integral–derivative (PID) method to increase the dynamic range through feedback control [118].

HDR imaging aims to address the practical imaging challenges when very bright and dark objects coexist in the scene. Therefore, it is widely used in digital photography by exposure control and has also been used in medical imaging and remote sensing.

F. NLOS Imaging

NLOS imaging can recover details of a hidden scene from the indirect light that has scattered multiple times [125]. During the very first research period, a streak camera was used [126] to capture the optical signal reflected by the underlying scene illuminated by the pulsed laser but limited to 1-D capture in one shot. Later, single-photon avalanche diode (SPAD) has gained popularity in NLOS imaging to increase temporal and spatial resolution [127], [128], [129]. Its benefit is that it does not require mechanical scanning, and it can be used with single-pixel imaging to produce 3-D capture [130], [131]. After capturing the transient data, the NLOS scene can be reconstructed in the form of volume [126], [127], [128], [129] or surface [134]. In the branch of volume-based reconstruction, Velten et al. [126] used higher-order light transport to model the reconstruction of an NLOS scene as a least-squares problem based on transient measurements and solve it approximately by filtered backprojection (FBP). This paves the way for solving this problem by deconvolution. For example, O’Toole et al. [127] represented the higher-order transport model as light cone transform (LCT) in the case of confocal and reconstructed the scene by solving the 3-D signal deblurring problem. In the forward model of confocal NLOS imaging, the 2-D transient measurement can be expressed as

$$\tau(x’, y’, t) = \iiint_{\Omega} \frac{1}{(x’ - x)^2 + (y’ - y)^2 + z^2} \rho(x, y, z) \times \delta \left(2 \sqrt{(x’ - x)^2 + (y’ - y)^2 + z^2 - t^2} \right) \times dx \, dy \, dz$$ (12)

where \(\rho(x, y, z)\) is the albedo distribution of the target 3-D scene, \(\delta\) is the Dirac delta function, and \(c\) is the light speed. Let \(z = \sqrt{\alpha} \) and \(v = (tc/2)^2\), (12) can be simplified into the
3-D convolution form as $R_{\ell}(\tau) = h \ast R_{\rho}(\rho)$, where

$$R_{\ell}(\tau)(x', y', v) = \nu^{3/2} \tau(x', y', z')$$

$$R_{\rho}(\rho)(x, y, u) = \frac{1}{2\nu u} \rho(x, y, v/u))$$

$$h(x' - x, y' - y, v - u) = \delta((x' - x)^2 + (y' - y)^2 + u - v).$$

(13)

The discretized form of the above convolution can be solved by optimization methods similar to (4), where the sensing matrix $H$ is the 3-D convolution operation [127], [138]. Based on the idea of backpropagation, La Manna et al. [132] proved that the iterative algorithm can optimize the reconstruction results and is more sensitive to the optical transport model. In addition, Lindell et al. [128] used the 3-D propagation of waves and recovered the geometry of hidden objects by resampling along the time–frequency dimension in the Fourier domain. Heide et al. [133] added the estimation of the surface normal vector to the reconstruction result by an optimization problem. In the branch of surface-based reconstruction, Tsai et al. [134] obtained the surface parameters of the scene by calculating the derivative relative to NLOS geometry and measuring the reflectivity. Using the occlusions [135] or edges [136] in the scene also helps to retrieve the geometric structure. Recently, the polar signal [137], CS [138], and deep learning [145] have also been used in NLOS to improve the performance. Most recently, NLOS has achieved a range of kilometers [139]. Other than optical imaging, Maeda et al. [140] conducted thermal NLOS imaging to estimate the position and pose of the occluded object, and Lindell et al. [141] implemented acoustic NLOS imaging system to enlarge the distance and reduce the cost.

At present, NLOS has been used in human pose classification through scattering media [142], 3-D multihuman pose estimation [143], and movement-based object tracking [144].

In the future, NLOS has potential applications in self-driving vehicles and military reconnaissance.

G. Radar Imaging

As a widely used imaging technique in remote sensing, synthetic aperture radar (SAR) is an indirect measurement method that uses a small-aperture antenna to produce a virtual large-aperture radar via motion and mathematical calculations. It requires the joint numerical calculation of radar signals at multiple locations to obtain a large FOV and high resolution.

Moses et al. [172] proposed a wide-angle SAR imaging method to improve the image quality by combining GPS and the UAV technology. On the basis of traditional SAR, interferometric SAR, polarimetric SAR, and other technologies have further improved the performance of remote sensing and helped implement the functions of 3-D positioning, classification, and recognition.

Recently, other radar techniques have also been used for imaging. For example, multiple-input–multiple-output radar has been widely used in mobile intelligent platforms due to its high resolution, low cost, and small size, which synthesizes a large-aperture virtual array with only a small number of antennas to achieve a high angular resolution [173]. Many efforts have been made to improve the imaging accuracy of radar on mobile platforms. In the 2-D radar imaging model, it is assumed that a single reflector exists on each grid point in the region of interest (ROI). For the signal from a transmitting antenna at position $r$, the signal received by the antenna at position $r'$ at all frequencies can be written in vector form as

$$y(r, r') = A(r, r')x + n(r, r')$$

(14)

where $A(r, r')$ denotes the radar operator containing the term of magnitude attenuation and phase change, $x$ is the reflectivity vector at each grid point, and $n(r, r')$ is the noise. Theoretically, the operator $A(r, r')$ can be provided if the positions $r$ and $r'$ are exactly known; thus, the reflectivity distribution $x$ can be derived by methods similar to (4). However, the uncertainty of positions remains due to the inaccuracy of global positioning systems. Under this circumstance, Mansour et al. [174] proposed a method to map the position error of each antenna to the spatial shift operator in the image domain and perform the multichannel deconvolution to achieve superior autofocus performance. Sun and Zhang [175] used random sparse step-frequency waveform to suppress high sidelobes in azimuth and elevation and to improve the capability of weak target detection. In recent years, through-the-wall radar imaging (TWRI) has attracted extensive attention because of its applications in urban monitoring and military rescue [176]. Amin et al. [177] analyzed radar images by using short-time Fourier transform and wavelet transform to perform anomaly detection (such as falling) in elder care. At present, machine learning algorithms and deep neural networks are also applied to object classification and detection in Radar imaging. Wu et al. [178], [179] used the clustering properties of sparse scenes to reconstruct high-resolution radar images and further developed a practical subband scattering model to solve the multitask sparse signal recovery problem with Bayesian CS, respectively. Wang et al. [180] used the stacked recurrent neural network (RNN) with long-short-term-memory (LSTM) units with the spectrum of the original radar data to successfully classify human motions into six types. In addition, radar has also been applied to hand gesture recognition [181], [182], [183], and Wi-Fi has been used for localization [184], [185], [186], [190].

Based on its unique frequency, depending on platforms such as satellites, drones, and mobile devices, radar has its advantageous applications in the military, agriculture, remote sensing, and elder care.
H. SR Imaging

High-resolution imaging has been a long-sought goal, which is crucial for successive analysis. As hardware solutions such as reducing to finer pixel size or increasing chip size lead to higher noise and growing costs, various methods have been explored to generate an SR image from an LR input or a set of LR inputs acquired from different perspectives at the same scene [191], [192].

The resolution degeneration might arise from the low density of sensor elements, insufficient spatial bandwidth product of the camera lens (i.e., optical aberrations), and even camera motion [191]. Most SR algorithms follow a general imaging model:

$$y_k = H_k x + e_k$$  \hspace{1cm} (15)

which has a similar form to (2). Assuming that the upsampling rate of SR is $L_1$ and $L_2$ in height and width dimensions, respectively, the vector form of the $k$th measurement and target data then should be $y_k \in \mathbb{R}^{H \times W}$ and $x \in \mathbb{R}^{L_1 \times L_2 \times W}$, respectively, and the sensing matrix $H_k \in \mathbb{R}^{H \times W} \times L_1 \times L_2 \times W$ represents the contribution of HR pixels in LR pixels under the effect of blurring, warping, and downsampling [191], [193]. When extended to videos, we can simply super-resolve the input in a framewise manner or additionally introduce temporal priors [192].

As for the reconstruction, inferring $x$ from $p$ LR inputs \{ $y_k, k = 1, 2, \ldots, p$ \}, the field has witnessed a long-standing development in the past decades [191], [192], [194], and various algorithms have been proposed to infer the latent HR image via inverting the linear system in frequency domain [195], [196] and spatial interpolation [197], [198], introducing regularization into an optimization framework [193], [199], projecting onto convex sets [200], [201] or jointly maximizing the likelihood and posterior probability [202], and learning in data-driven manner [203], [204]. Researchers can choose appropriate algorithms according to their pros and cons, as well as the demands of specific applications. Despite the differences, most algorithms follow a similar pattern to (4) and optimize

$$\hat{x} = \arg \min_{x} \sum_{k=1}^{p} \| y_k - H_k x \|^2 + \lambda R(x)$$  \hspace{1cm} (16)

where $\lambda$ and $R(x)$ are defined similar to (4).

In recent years, the vast development of deep learning has greatly advanced the performance of SR [205]. In the regime of computational reconstruction, multiple deep learning architectures, such as SR convolutional neural network (SRCNN) and deep residual network for SR [very deep convolutional networks for SR (VDSR)], have been proposed for the natural image sets [203], [206], [207], [208], [209], [210]. Distinctively, the SRCNN proposed by Dong et al. [203] is designed with high simplicity and efficiency, demonstrating the important role of deep learning in the SR task. There is also some new work addressing the resolution degeneration by the imaging lens via designing a CI system with an engineered point spread function (PSF). For example, Zhang et al. [212] introduced deep optics into a temporal CS system to encode fast HR motions into an LR snapshot and developed an end-to-end deep neural network to optimize the PSF and reconstruction algorithm simultaneously.

Applications of SR imaging can extend to many fields concerning high-quality images. For example, Wang et al. [211] summarized the big progress of remote sensing image SR brought by deep learning techniques and presented a new high-resolution dataset from different satellites of various landscapes, targeting for benchmarking and further promoting the advancement in this field. Hyperlenses and metalenses could also retrieve the SR information and project details about the images [215]. Future development prospects of SR imaging include the acceleration of the models, the extensive comprehension, and the criteria for designing and evaluating the functions [204].

I. Single-Pixel Imaging

Single-pixel cameras capture the target scene with a single bucket detector by recording its correlated measurements synchronized with time-varying spatial patterns, which are usually implemented with an SLM. Such an imaging scheme is advantageous in nonvisible spectrum-lacking mature array sensors and photon-starved scenarios or cases where light scattering exists. Unlike conventional imaging with silicon array sensors, single-pixel imaging needs to computationally retrieve the scene from the coded measurements [216], [217].

The single-pixel imaging model can be formulated as

$$y = Px$$  \hspace{1cm} (17)

where $x \in \mathbb{R}^{H \times W}$ is the vector form of intensity distribution of the target scene, $P \in \mathbb{R}^{M \times H \times W}$ is a $M$ binary patterns that sequentially encode the scene, and $y \in \mathbb{R}^{M}$ is the vector of the recorded correlated measurements [217]. Since the target scene is statistically redundant, one can record a small portion of measurements and reconstruct the scene via solving this ill-conditioned problem after imposing some regularizations, similar to (4). To address the high computational complexity, which hampers high-resolution imaging, Zhang et al. [218] proposed to use sinusoidal spatial patterns and retrieve the high-resolution 2-D image with high quality in the Fourier domain. Later, they further proposed to address the low refresh rate of the sinusoidal spatial light modulation [219].

In addition to computational reconstruction in the CS regime, one can also conduct fast reconstruction by exploiting the correlation between the known random patterns and measured signals [220]. Denoting the sequential...
patterns as $\mathbf{P}_i$ ($i = 1, 2, \ldots, N$) and the $i$th correlated measurement as $\mathbf{y}_i$, the scene can be reconstructed by

$$
\mathbf{O}(x, y) = \langle (\mathbf{y}_i - \langle \mathbf{y}_i \rangle) (\mathbf{P}_i(x, y) - \langle \mathbf{P}_i(x, y) \rangle) \rangle
$$

where $\langle \cdot \rangle$ takes average over all the entries [220], [222]. Despite the higher efficiency in reconstruction, high-quality results demand largely redundant measurements. Lyu et al. [224] proposed a new framework to utilize deep learning techniques to promote the imaging quality at an extremely low sampling rate, as deep neural networks can learn features from the noisy measurement for prediction [223]. Recently, Wang et al. [225] achieved better performance in terms of robustness and fidelity by blending a physics-informed layer and a model-driven fine-tuning process.

Single-pixel cameras can potentially be used for low-cost imaging outside the visible spectrum or in harsh illumination scenarios and achieve time-resolved imaging, which allows visualizing gas leaks as well as 3-D ranging for autonomous vehicles [217]. Machine learning techniques have also demonstrated their potential in “image-free” detection and classification [226] under the single-pixel imaging architecture.

J. Imaging Through the Fog

Light propagating through the fog or clouds is scattered by airborne particles and degenerates the image quality; therefore, CI is required to reconstruct the latent scene from the degraded measurement.

In the early years, Bissonnette [227] proposed a method to calculate the forward-scattering effect on the point spread and also some modulation functions, which was demonstrated to be effective within certain ranges. Jaruwatanadilok et al. [228] looked further into the impact of fog and clouds in imaging under optical wavelengths and presented a system applicable to any optical and any polarization state based on the point-source radiative transfer theory. In recent years, more approaches have been applied for better performance. For example, Satat et al. [229] demonstrated that the time profiles of light scattered from fog and occluded objects have different distributions (Gamma and Gaussian, respectively), which can be utilized to distinguish the fog properties from the measurement and improve the imaging quality. In contrast to the aforementioned methods, Guan et al. [230] adopted the millimeter wave (mmWave), which has favorable propagation characteristics despite the very LR, peculiarity, and noise artifacts. The introduced system is able to retrieve high-frequency shapes from raw mmWave signals and a data synthesizer is developed to aid the dataset generation, thus greatly improving the performance of mmWave radars in low visibility conditions.

In addition to modeling PSF or modifying the imaging system, various methods have also been proposed to directly reconstruct clear images from recording through the fog, dubbed image dehazing [231], [232], [233]. The commonly used image degradation model in dehazing can be described as [234]

$$
\mathbf{I}(x, y) = \mathbf{J}(x, y)e^{-r d(x, y) J} + A(1 - e^{-r d(x, y) J})
$$

where $\mathbf{I}(x, y)$ is the observed image, $\mathbf{J}(x, y)$ is the latent counterpart, $r$ is the atmospheric scattering coefficient, $d(x, y)$ is the scene depth, and $A$ is the global constant independent of the spatial coordinate. Let the transmittance $t(x, y) = e^{-r d(x, y)}$, and the “fog-free” image $\mathbf{J}(x, y)$ can be inferred by first estimating $A$ from $\mathbf{I}(x, y)$ via imposing different priors (e.g., dark channel prior [235], [236], [237], [238], patchwise regression [239], and maximum brightness [237]) and the transmittance

$$
t(x, y) = 1 - w \frac{\mathbf{I}(x, y)}{A}
$$

where $w$ is the adjustment parameter.

Deep learning is also employed to promote the quality of looking through the fog, where Tahir et al. [240] proposed an adaptive learning framework, termed dynamic synthesis network (DSN) to adapt to different scattering conditions, which is achieved by a “mixture of experts” architecture and used for holographic 3-D particle imaging. The positive experimental results validate its effectiveness and prospects in this field, and the progress has wide potential in industries, such as self-driving cars, augmented driving, airplanes, helicopters, UAVs, and trains [229].

K. Low-Light Imaging

The low-light imaging technology refers to enhancing the visibility in circumstances with weak illumination when the latent contents might be buried by severe sensor noise [247], [248]. This target can be implemented by developing either higher end camera sensors or advanced algorithms. Improvements in the hardware lead to high sensitivity. Moomaw [249] reviewed four camera technologies for low-light imaging, including a cooled charge-coupled device, cooled scientific CMOS, intensified camera, and electron multiplier camera. More development potentials, however, lie in the methodologies of low-light image enhancement. As reviewed by Kim [248], there are two representative approaches—handcrafted feature-based and learning-based. The former is further classified into two submodels, i.e., statistical and decomposition models, while the latter consists of reference-based and no-reference-based models.

In the conventional handcrafted approaches, the very first enhancement techniques include the histogram equalization and gamma correction [250], [251], [252]. In order to better suppress noise while preserving local details, the Retinex model is proposed [253], [254], modeling the captured image $\mathbf{S}$ as the elementwise product of
the illuminance map $L$ and the reflectance map $R$, i.e.,

$$S(x, y) = R(x, y)L(x, y). \quad (21)$$

To reveal the intrinsic properties of the scene, multiple algorithms have been proposed to retrieve $R$ \cite{255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265}. The representative single-scale Retinex algorithm calculates the reflectance map by

$$r(x, y) = \log R(x, y) = \log S(x, y) - \log [F(x, y) * S(x, y)] \quad (22)$$

where $*$ is the convolution operator and

$$F(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (23)$$

and $\sigma$ is the Gaussian surround space constant \cite{266}. The Retinex model also inspires some deep learning methods \cite{267, 268, 269, 270, 271, 272, 273, 274, 275}

Generally, the learning-based approaches rely on the exquisitely designed dataset as well as elegant neural network architectures \cite{276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289}. To this end, multiple low-light image datasets have been constructed \cite{247, 256, 259, 267, 290, 291, 292}. Xu et al. \cite{294} prepared a new dataset with real noise that disrupts the image content, where a novel observation is also proposed that it is much easier to detect noise in the low-frequency layer than in the high one, which serves as the fundamental theory to set up a frequency-based decomposition-and-enhancement model and a novel network. Another method by Lv et al. \cite{295} is based on a multibranch convolutional neural network (CNN), where the training dataset is much larger and more diverse and two attention maps are learned. Despite the progress, it should be noted that current exploration still calls for more dedicated datasets and network design \cite{293}. On the one hand, the existing dataset lacks precise calibration, limited scale, and diversity. On the other hand, recent methods might be unsatisfying in some cases where details about small objects \cite{294}, human faces \cite{295}, or nearly black regions \cite{297} are required. Apart from those approaches, low-light image enhancement can be implemented using a deep autoencoder. According to Lore et al. \cite{296}, a variant of the stacked-sparse denoising autoencoder can be learned from synthetic noisy dark images with high efficiency.

L. Summary

All these systems described above involve computation in the image acquisition process. Toward this end, diverse signal processing algorithms have also been developed to handle the inverse reconstructing problem represented by \eqref{eq:inverse}. In the conventional framework based on convex optimization solutions, high interpretability and flexibility have been seen in various designed priors, which are able to achieve competitive performance after sufficient iterative steps. In addition, the adopted priors can also provide guidance for CI schemes to some extent, such as the designing of masks. However, the severe dependence of these classical methods on prior selection and design may also lead to shortcomings, i.e., excessive operation time and empirical parameter setting due to the neglect of implicit features of the data itself. During the past decade, the emergence of AI, especially advances in deep neural networks, has dramatically improved the efficacy of the algorithm design. By applying the flourishing AI methods to the inverse model in an end-to-end manner \cite{299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311} or to some procedure in the iterative process \cite{312, 313, 314, 315}, it has been proved that the data-driven methods are capable for greatly enhancing the computing efficiency while maintaining the performance by virtue of their strong ability to explore data features. Currently, there is a general trend to apply AI to CI design and processing.

Again, we want to mention that the above representative list is by no means exhaustive. Some other important CI research topics, such as raindrop removal, scattering, diffusion imaging, and phaseless imaging, potentially also have a high impact on mobile vision.

III. AI FOR CI: CURRENT STATUS

The CI systems introduce computing into the design of imaging setup. They are generally used in three cases: 1) it is difficult to directly record the required information, such as the depth and large FOV; 2) there exists a dimensional mismatch between the sensor and the target visual information, such as light field, hyperspectral data, and tomography; and 3) the imaging conditions are too harsh for conventional cameras, e.g., high-speed imaging and HDR imaging.

CI obtains the advantage over conventional cameras in terms of the imaging quality or the amount of task-specific information. These advantages come at the expense of high computing costs or limited imaging quality since computational reconstruction is required to decode the information from the raw measurements. The introduction of AI provided early CI with many revolutionary advantages, mainly improving the quality (higher signal-to-noise ratio, improved fidelity, fewer artifacts, and so on) and efficiency of computational reconstruction. In some applications, with the help of AI, CI can break performance barriers in photography \cite{316, 317}. Most recently, AI has been employed into the design of CI systems to achieve adaptivity and flexibility as well as to improve performance in high-level mobile vision tasks with a schematic shown in Fig. 2 and the representative works summarized in Table 3. In the following, we review the four typical schemes...
incorporating AI techniques for advances in CI systems and the key optoelectronic elements for implementing such schemes. Finally, we review some typical examples of AI-guided CI system designs in the representative CI fields referred to in Section II, adopting the aforementioned integration strategies and coding elements.

A. AI Improves Quality and Efficiency of CI System

Using AI to enhance the quality and efficiency of CI systems is the most direct and commonly used approach, as shown in the first plot in Fig. 2. In this section, we illustrate the ability of AI to enhance the reconstruction quality of CI using compressive imaging as an example. By incorporating spatial light modulation into the imaging system, compressive imaging can encompass high-dimensional visual information (video, hyperspectral data cube, light field, and so on) into one or more snapshot(s). Such encoding schemes make extensive use of the intrinsic redundancy of visual information and can reconstruct the desired high-throughput data computationally. Thus, compressive imaging greatly reduces the recording and transmission bandwidth of high-dimensional imaging systems. Since the decoding is an ill-posed problem, complex algorithms are required for high-quality reconstruction. Researchers have spent great efforts on algorithm development (such as optimization methods using total variation (TV) [318] and sparsity priors [82], [108] into generalized alternating projection (GAP) [319], [320] and two-step iterative shrinkage/thresholding (TwIST) [321], and new algorithms such as Gaussian mixture model (GMM) [322], [323] and decompress SCI (DeSCI) [324]) and the performance is improved continuously during the past years. However, these algorithms are inferred under the optimization framework and are usually either of limited accuracy or time-consuming.

Instead, the pretrained deep networks can learn the image priors quite well and are of high computing efficiency. Taking the SCI [22] as an example, SCI conducts specific encoding and acquisition from the scene and takes the encoded single shot as the input of the decoding algorithm to reconstruct multiple interrelated images at the same time, such as continuous video frames or multispectral images from the same scene, with high compression ratio. At present, a variety of high-precision CNN structures have been proposed for SCI. Ma et al. [325] proposed a tensor-based deep neural network and achieved promising results. Similarly, Qiao et al. [299] built a video SCI system using a digital micromirror device and developed both an end-to-end CNN (E2E-CNN) and a plug-and-play (PnP) algorithm for video reconstruction. The PnP framework [312], [313] can serve as a baseline in video SCI reconstruction considering the tradeoff among speed, accuracy, and flexibility. PnP has also been used in other CI systems [326], [327]. With the development of deep learning, more optimization of the speed and logic of the SCI networks are proposed. By embedding Anderson acceleration into the network unit, a deep unrolling algorithm has been developed for SCI reconstruction [328]. Considering the temporal correlation in video frames, the RNN has also been a tool [300], [329] for video SCI. For large-scale SCI reconstruction, the memory-efficient network and meta-learning [301], [302] pave the way. To overcome the challenge of limited training data, untrained neural networks have been proposed most recently [314], [330], which also achieve high performance with the requirement of less data. Based on the optimized neural network, the application-driven SCI meets the specific needs and achieves real-time and dynamic imaging. For example, neural network-based reconstruction methods have been developed for spectral SCI systems [303], [304], [331]. By introducing a diffractive optical element in front of the conventional image sensor, a compact hyperspectral imaging system [332] has been built with CNN for
Table 3 Reference List of AI for CI

| AI for CI | Optimization Method | References |
|-----------|---------------------|------------|
| AI improves quality and efficiency of CI system | Use deep learning to decode the high-dimensional data from multiple coded measurements | [333] [334] |
| | Propose a deep neural network based on a standard tensor ADMM algorithm | [325] |
| | Build a video CI system using a digital micromirror device and develop end-to-end convolutional neural networks for reconstruction | [299] |
| | Build a multispectral endomicroscopy CI system using coded aperture plus disperser and develop end-to-end convolutional neural networks for reconstruction | [109] |
| | Propose a new auto-focus mechanisms based on reinforcement learning | [336] |
| AI optimizes structure and design of CI system | Assist in designing the physical layout of imaging system for compact imaging | [342] |
| | Perform end-to-end optimization of an optical system | [56] [115] [343]–[346] |
| | Introduce special optical elements, or replace the original ones for lightweight system | [347]–[350] |
| | Show better results by novel end-to-end network frameworks via optimization | [332] [351]–[354] |
| AI promotes scene adaptive CI system | Propose the concept of pre-deep-learning that uses random projections for information encoding, and reduces the number of required measurements to achieve high quality imaging for specific tasks | [355]–[358] |
| | Provide powerful means for scenes adaptive acquisition by upper-deep-learning combined with public database and open source | [359] [360] |
| AI guides high-level task of CI system | Use scene’s random-projections as coded measurements and embed the random features into a bag-of-words model | [361] |
| | Introduce multiscale binary descriptor for texture classification and obtain improved robustness | [362] |
| | Other application scenarios of task-oriented CI | [363] [364] |

high-quality reconstruction, performing well in terms of spectral accuracy and spatial resolution.

In addition to reconstructing multiple images with a single measurement, deep learning has also been used to decode the high-dimensional data from multiple coded measurements and achieve satisfying results [333], [334], [335]. A reinforcement learning-based autofocus method has been proposed in [336]. Taking one step further, the sampling in video compressive imaging can be optimized to achieve sufficient spatiotemporal sampling of sequential frames at the maximal capturing speed [337], and learning the sensing matrix to optimize the mask can further improve the quality of the CS reconstruction algorithm [338].

In addition to SCI, deep-learning-based reconstruction has recently been widely used in other CI systems, such as 3-D imaging with deep sensor fusion [339] and lensless imaging [340], [341].

### B. AI Optimizes Structure and Design of CI System

Instead of designing the system in an ad hoc manner and adopting AI for high-quality or efficient reconstruction (Section III-A), recently, some researchers proposed closer integration between CI and AI. In particular, as shown in the second plot of Fig. 2, by optimizing the encoding and decoding jointly, AI can assist in designing the physical layout of the system [342] for compact and high-performance imaging, such as extended depth of field, auto focus, and HDR.

One method is to perform end-to-end optimization [56] of an optical system using a deep neural network. A completely differentiable simulation model that maps the real source image to the (identical) reconstructed counterpart was proposed in [343]. This model jointly optimizes the optical parameters (parameters of the lens and the diffractive optical elements) and image-processing algorithm using automatic differentiation, which achieves achromatic extended depth of field and snapshot SR. By collecting feature vectors consisting of focus value increment ratio and comparing them with the input feature vector, Han et al. [345] proposed a training-based autofocus method to efficiently adjust lens position. Another CI method achieves HDR capture by jointly training an optical encoder and algorithmic decoder [115]. Here, the encoder is parameterized as the PSF of the lens and the decoder is a CNN. Similar ideas can also be applied to a variety of devices. For instance, a deep adaptive sampling LiDAR was developed by Bergman et al. [346].

In contrast to the above systems employing existing CNNs, improved deep networks have also been developed to achieve better results in some systems. For example, based on CNN, Wang et al. [351] proposed a unified framework for coded aperture optimization and image reconstruction. Zhang et al. [352] proposed an optimization-inspired explainable deep network composed of a sampling subnet, initialization subnet, and recovery subnet. In this framework, all the parameters are learned in an end-to-end manner. Compared with existing state-of-the-art network-based methods, the OPINE-Net not only achieves high quality but also requires fewer parameters and less storage space while maintaining a real-time running speed. This method also works in reconstructing a high-quality light field through a coded aperture camera [353] and depth of field extension [354].

The booming development of inter-disciplines, such as material and computer science, has seen much effort to replace the original optical elements with lower weight implementation. Using only a single thin-plate lens element, a novel lens design [347] and learned reconstruction architecture achieve a large FOV. Banerji et al. [348] proposed a multilevel diffractive lens design based on deep learning, which drastically enhances the depth of focus.
In addition, the design of the aperture pattern plays an essential role in imaging systems. In view of this situation, a data-driven approach has been proposed to learn the optimal aperture pattern where coded aperture images are simulated from a training dataset of all-focus images and depth maps [349].

C. AI Promotes Scene Adaptive CI System

Taking one step further, imaging systems should adapt to different scenes, where AI can play significant roles as shown in the third plot in Fig. 2. Conventional imaging systems use a uniform acquisition for varying environments and target tasks. This usually ignores the diverse results in improper camera settings or records a large amount of data irrelevant to the task, which is discarded in successive processing. Bearing this in mind, the imaging setting of an ideal CI system should actively adapt to the target scene and exploit the task-relevant information [365].

Pre-deep learning studies in this area frequently make use of optimized random projections for information encoding and attempt to reduce the number of required measurements for specific tasks. Rao et al. [355] proposed a compressive optical foveated architecture that adapts the dictionary structure and compressive measurements to the target signal, by reducing the mutual coherence between the measurement and the dictionary and increasing the sparsity of representation coefficients. Compared to conventional Nyquist sampling and CS-based approaches, this method adaptively extracts task-relevant ROIs and thus reduces meaningless measurements. Another task-specific compressive imaging system employs generalized Fisher discriminant projection bases to achieve optimized performance [356]. It analyzes the irrelevant performance of target detection tasks and achieves optimized task-specific performance. In addition, adaptive sensing has also been applied to ghost imaging [357] and iris recognition [358].

Inspired by deep learning, machine vision now has become a more powerful approach for scene adaptive acquisition. As camera parameters and scene composition play a vital role in the esthetics of a captured image, researchers can now use publicly available photo databases, social media tips, and open-source information for reinforcement learning to improve the user experience [359]. ClickSmart [360], a viewpoint recommendation system, can assist users in capturing high-quality photographs at well-known tourist locations, based on the preview at the user’s camera, current time, and geolocation. Specifically, composition learning is defined according to factors affecting the exposure, geographic location, environmental conditions, and image type, and then provides adaptive viewpoint recommendation by learning from the large high-quality database. Moreover, the recommendation system also provides camera motion guidance for pan, tilt, and zoom to the user to improve scene composition. An end-to-end optimization of optics and image processing for achromatic extended depth of field and SR imaging framework is proposed in [343]. Toward correcting large-scale distortions in computational cameras, an adaptive optics approach is proposed in [366].

D. AI Guides High-Level Task of CI System

As mentioned in Section I, instead of capturing more data, CI systems will focus on performing specific tasks and AI will help to maximize the performance, which is shown at the bottom of Fig. 2. Specifically, CI optimizes the design of the imaging system by considering the subsequent processing. Toward this end, one can build systems optimized for specific high-level vision tasks and conduct analysis directly from the coded measurements [367]. Such a method can capture information tailored for the desired task and thus has improved performance and largely reduced bandwidth and memory requirements. Taking the properties of the target scene into system design, we can also obtain high flexibility in the environments, which is important for mobile platforms.

Task-oriented CI is an emerging ongoing direction and there only exist some preliminary studies during the writing of this article. Texture is an important sensor property of nature scenes and provides informative visual cues for high-level tasks. Traditional texture classification algorithms for 2-D images have been widely studied, but they occupy a large bandwidth since object with a complicated texture is difficult to compress. To achieve texture classification from a small number of measurements, an approach is proposed in [361] to use the scene’s random projections as coded measurements and embed the random features into a bag-of-words model to perform texture classification. This approach results in significant improvements in classification accuracy and reductions in feature dimensionality. On the contrary, Liu et al. [362] introduced a multiscale local binary pattern (LBP) descriptor for texture classification and obtained improved robustness. The classification from coded measurements is also proved to be applicable for other kinds of data [363], such as hyperspectral images. Besides texture classification, Zhang et al. [364] proposed an appearance model tracking algorithm based on extracted multiscale image features, which performs favorably against state-of-the-art methods in terms of efficiency, accuracy, and robustness in target tracking.

E. Implementations of AI-Optimized CI Systems

After obtaining the AI-optimized imaging mechanisms, special optical elements must be introduced into the light path for better image/video capture. Researchers have utilized various optoelectric or optical elements at different locations in the light path for advanced task-oriented system design. Hereby, we provide a brief description of the modulation elements widely used in CI systems.

1) Spatial Light Modulator: In a broad sense, any device that imposes spatially varying modulation on a beam of
light can be named an SLM, such as printed or photoetched mask, but in CI systems, the term SLM usually refers to programmable ones, i.e., the modulation can be controlled by a computer [368]. To this end, researchers have invented various SLM principles [369], among which the two most widely used are liquid crystal SLM (LC-SLM) and DMD. The former utilizes its optical rotation effect to change the polarization direction of the incident light with the voltage, i.e., modulating the polarization state. Correspondingly, the amplitude can be modulated by controlling the direction change to adjust the proportion of outgoing light and phase by changing the refractive index of the birefringent crystal through its rotation [370], [371]. The latter builds on an array of micromirrors factorized at the micrometer level. Each micromirror can change the angle of the incoming light beam by flipping up and down quickly to control the light flux transmitted into the aperture and thus achieve amplitude modulation. When not perpendicular to the optical axis, the DMD can perform spatially varying phase modulation by controlling the micromirror flip and achieve phase modulation [372], [373]. DMD has been widely used and is developing rapidly toward modularization, miniaturization, and specialization, which promises broader applications in the future. In addition to the two aforementioned categories, various SLMs have also been developed based on deformable membranes [374], [375], thermoplastic deformation [376], magnetic deformation [377], [378], [379], acousto-optic effect [380], Pockels effect [381], photorefractive effect [382], [383], microchannel plate [384], [385], and so on.

Researchers are also attempting to increase SLM’s pixel count, which is highly related to the modulation accuracy and even the resolution of the whole imaging system. High-definition SLM is enabled by the progress of semiconductor technology, e.g., optical units at micrometer or even submicrometer scale [379], [386], and driven by the demanding high-definition display [371], [387], while the high cost is still a big challenge. To sum up, SLM is a key device in CI setups and expected to have broad application prospects in optical imaging [86], [93], [108], [220], [388], [389], phase modulation [390], [391], laser beam shaping [392], wavefront modulation [393], [394], optical tweezers [395], [396], [397], holographic projection [398], [399], and so on.

2) Ferroelectric Liquid Crystal: As a special type of liquid crystal, the arrangement of the ferroelectric liquid crystal (FLC) molecules can be controlled by the imposed voltage and rotate the polarization state of the linear polarized incident light by 0° or 90°, which determines whether the light is transmitted or blocked together with a linear polarizer, i.e., acting as a switching control of light [400], [401]. Based on this property, FLC can be placed in the light path to serve as an electrooptic shutter [402] with customized exposure pattern, which is an ideal option for coded exposure photography [403], [404]. Some specially designed FLC-SLM can also achieve pixelwise fast switching control on the incident light [405], [406], serving as a binary SLM. At present, ferroelectric materials are being developed toward a lightweight and thin scale [407], which facilitates integration into compact CI systems.

3) Structured Light Source: Structured light is a CI technology that projects specific textures onto the target scene to infer the 3-D structure [408], [409]. Primary studies use point or line structured light [410], [411], [412] that can be implemented using either a scanning laser or light-emitting diode (LED) array but suffer from low efficiency. Hence, 2-D-patterned structured light is more commonly used [413], [414]. One can sequentially project a series of simple patterns to encode each pixel uniquely [408], [415], [416], [417], [418] or a single pattern being nonrepetitive in a certain neighborhood [419], [420]. In comparison, the former generally works on static scenes but has higher accuracy and at lower manufacturing costs [413], while the latter is applicable for dynamic scenes. In terms of physical implementation, SLM [421] and geometric phase elements [422] have been widely used to generate structured light in CI setups but often lead to bulky setups. To address this issue, researchers have made great efforts to generate light sources emitting structured light. Various lasers were invented to produce high-quality structured light and even achieve dynamic control of patterns [422]. On the contrary, the LED can also generate grids, dots, lines, and other simple patterns through the photolithographic raster slice, at a lower cost than lasers [423], [424]. Although the idea of structured light was proposed about two centuries ago, new pattern designs have been emerging, such as 3-D structured light by constructing the initial 2-D field and 4-D structured light by integrating the space–time joint design [425], [426].

4) Spectral Modulators: Optical elements being able to conduct spectrum-dependent modulation are indispensable for computational spectral cameras. Among the existing multispectral imaging systems, the filter array is an intuitive modulator [427] but conducts spectral imaging at the expense of temporal or spatial resolution [428], [429], [430]. The hyperspectral sensor with tunable filter addresses these challenges properly [431], with liquid crystal tunable filter (LCTF) and AOTF being the two most widely used components. The former leverages the birefringence effect of crystal to adjust the transmittance spectrum by the imposed voltage, and the latter uses anisotropic acousto-optic medium [432]. Moreover, the development of material science and chemistry enables manufacturing of elements with spectral responses similar to conventional optical modulators, thus promoting the innovation of lightweight hyperspectral sensors, such as metasurface [433], [434] and quantum dots [435].

5) Programmable Sensor: Due to space limitations, pixelwise modulation usually needs additional optics to relay the modulation pattern at the image plane onto the sensor, which is bulky and fragile. Recently, the
emerging programmable sensor (or focal-plane sensor) has addressed this issue by integrating sensing and processing on a single chip in a fine-grain parallel manner [436]. A representative reconfigurable focal-plane sensor consists of an array of pixel units composed of both photosensitive and programmable processing elements permitting low-level processing during acquisition [437]. The novel sensor has demonstrated its superior performance in many CI fields, such as HDR imaging [123] and motion deblurring [438], and is expected to participate in video compression [123], hyperspectral imaging [439], feature classification [440], and other fields with higher flexibility in the future.

6) Streak Camera: As an optoelectronic instrument integrating optics, electronics, and semiconductor technologies, the streak camera glances the transient process by spatially shearing the arriving photoelectrons at different instants to specific positions on the detector, thus obtaining ultrahigh time resolution [87], [90]. The very first version of the streak camera was invented in 1971 [441]. After decades of development, the current manufactured streak camera can achieve a time resolution of less than 1 ps. However, streak cameras can only capture transient events across a line, and researchers use streak cameras together with transmissive masks to conduct transient modulation of 2-D or 3-D scenes [87], [442], [443], achieving optical streaking with a frame rate up to $10^{11}$–$10^{12}$ frames/s [444]. The streak camera is also employed in the NLOS imaging [126]. The future development of streak cameras is expected to provide a new perspective of transient dynamics and might advance imaging setups looking around corners or through scattering medium.

7) Single-Photon Avalanche Diode: Distinguished from ordinary avalanche photodiodes, a single photon hitting the electron can induce an avalanche in SPAD, which is thus empowered with photon counting sensitivity at picosecond temporal resolution. With its ability to sense transient processes, e.g., light transportation, SPAD sensors are widely used in the CI field to infer scene structures. For instance, combined with mechanical scanning and single-pixel imaging, SPAD can assist in the reconstruction of 3-D objects in NLOS [130]. The mechanism can be further extended to single-shot 3-D NLOS using an SPAD array [127], [445], with the same spirit as LiDAR [446].

The emerging modulators with a compact form, programmable modulation, and high light efficiency would inspire more researchers to design advanced CI systems using emerging AI tools and lead the trend of integrating AI in compact CI devices.

F Typical Examples of AI-Optimized CI Systems

A CI system optically encodes the scene radiance and the codes are determined by the introduced photoelectric elements that provide more advanced imaging capabilities than a conventional camera. In primary CI system designs, the encoding scheme is often designed in an ad hoc manner, based on the target imaging performance, functions of coding elements, optical imaging rules, and theories in signal processing. Moreover, the computational decoding procedure often receives the calibrated encoding patterns as part of the input in a “passive” manner. Therefore, the final performance largely depends on experience and often does not generalize well. Recently, AI has begun to participate actively in the end-to-end design of CI systems, where the optical coding and computational decoding are jointly optimized by AI algorithms [447]. Combined with a variety of widely used facilities described in Section III-E, this AI-assisted joint design has demonstrated emerging prospects in many CI fields.

One example is computational depth imaging, in which the end-to-end optimization scheme proposed by Chang and Wetzstein [56] has proven that the lens with freeform phase masks or chromatic aberrations brings higher depth accuracy, where the phase mask can be optimized for depth estimation from a single image through deep learning [57], [448]. For high-speed CI, Dong et al. [449] proposed to employ the temporally coded snapshot imaging implemented with the aforementioned liquid crystal electrooptic shutter, retrieving the motion information of objects depicted by the sequential bounding box from a single image, where the pattern and length of the binary encoding sequence used for the realistic measurement can be optimized from the simulation performed by the adopted network model. The proposed scheme has also been proven to be effective in low-light imaging [450]. In the field of multicolor/spectral imaging, Chakrabarti [451] proposed a color camera that utilizes the learnable color multiplexing pattern instead of the traditional Bayer pattern, and the color channel acquired by each pixel is selected by a learnable weight. Peng et al. [347] designed an AI-guided wide-FOV computational camera, which is implemented by fabricating a thin-plate lens element with learned filters to compensate for aberrations and achieve a spatially invariant PSF. For HDR imaging, Metzler et al. [115] produced a learned diffractive optical element whose performance can be modeled as a function related to the shape of the surface, and thus, it can be jointly optimized by a CNN to adapt to the single-shot HDR imaging. In the field of SR imaging, Sitzmann et al. [343] associated the raw measurement and reconstructed image through a fully differentiable model, where the parameters of optical propagation can be automatically optimized to achieve extended depth of field and spatial resolution. For single-photon imaging, the SPAD array demonstrates impressive sensitivity, but the low pixel count and low filling factor remain a challenge, Sun et al. [452], [453] proposed an end-to-end framework to optimize the optical design of the sensor in order to enhance SPAD’s applications in high-speed imaging and transient imaging. Such AI-assisted joint design also applies to lensless imaging [454] and phase imaging [455]. In sum, the diverse CI
fields have witnessed the deep integration of AI and CI in novel designs, and the connection is believed to be further promoted with the development of programmable optical devices and advances in fabrication technologies.

IV. CI + AI + X: WIN-WIN FUTURE BY COLLABORATION

In Section III, we have shown the different preliminary integration approaches of CI and AI, and the advantages over systems using conventional cameras. Smartphones have become compact carriers that integrate CI and AI and play a crucial role in daily life. However, due to the limitations of load capacity, space, and budget, the current CI + AI methods still cannot provide ready machine vision technologies for other mobile platforms such as self-driving cars and UAVs. Considering that many mobile vision platforms require a complex CI system and AI algorithms involving heavy calculations, integrating CI and AI compactly to form a general and computing-efficient device has become an urgent challenge. Fortunately, with the recent advances in mobile communication (5G), edge and cloud computing, big data analysis, and lightweight algorithm design, as shown in Fig. 3, it is now feasible to tightly integrate CI and AI in an unprecedented manner. For the future of mobile vision systems, it is as important as strengthening “CI + AI” integration by the aforementioned technologies to extend the integration to new fields for a win-win collaboration. On the one hand, a broad range of disciplines would benefit from CI acquisition, AI processing, and their integration. For example, the way that biological vision systems record and process visual information is very delicate. Future mobile vision can draw on the principles of biological vision and brain science to meet the developing trend of high-precision, high-speed, and low-cost next-generation imaging systems (Fig. 4). The hardware development in chips, sensors, and optical elements would promote compact lightweight implementations. On the other hand, mobile vision systems have wide applications and can serve as a testing bed for achieving strong AI [456] in specific tasks.

A. Technologies Supporting Applications of CI + AI on Mobile Vision Platform

For a mobile vision platform integrating CI and AI, the CI setup extracts effective visual information from the environment by encoded illumination and acquisition. Next, the AI methods, in which deep learning now become the mainstream, are applied to process the collected data to retrieve semantic information, with high efficiency and accuracy. Subsequently, semantic information assists in the decision-making and actions of intelligent mobile platforms.

As mentioned before, the advantages of CI + AI systems come at the expense of complex (sometimes heavy) setups and high computational costs, which impose multiple challenges on mobile platforms such as cell phones and local vehicles. First, since the mobile platform is usually of limited size and resources, the imaging systems on it need to be compact, lightweight, and of low power consumption. Second, the platform needs to process a large amount of data from computational cameras and other sensors, and combine the multisource information for decision-making and action. Taking the self-driving vehicle as an example, for reliable depth sensing, it is typical to adopt multimodal strategies, e.g., “image + LiDAR” or “image + radar,” and the required data processing at least includes inpainting the missing values in the sparse LiDAR point cloud [457], [458], [459], [460], [461], increasing the signal-to-noise ratio of radar measurements [462], [463], and associating radar [464], [465] or LiDAR acquisition with RGB videos [466], [467], [468]. The advantages of multimodal CI systems come at the expense of demanding power supply and GPU memory that might be infeasible on low-capacity mobile platforms. Third, some mobile platforms have high requirements to conduct real-time operation and thus build the rapid observation-orientation-decision-action loop. Therefore, efficient computing is required to meet the demands of system scale, power consumption, and response time. In addition, a large amount of data must be stored for pattern recognition and learning, which demands a large amount of storage. In summary, the mobile platform needs to be equipped with a miniaturized imaging setup, high-performance processing units, and large storage, which might either exceed the load budget or take too long a response time, which is fatal for mobile platforms. For effective computing on mobile vision platforms, we need to reduce computing costs by designing lightweight algorithms, deploying big data analysis in the cloud, or adopting distributed edge computing techniques.

1) Lightweight Algorithm Design: To adapt the computational reconstruction and analysis algorithms to available computing resources on mobile platforms, various approaches have been proposed. Some methods build customized lightweight neural networks to reduce resource requirements. Among the prominent representatives are MobileNet and ShuffleNet, which enable running neural network models on mobile platforms or embedded
devices. MobileNet consists of a smaller parameter scale by using depthwise separable convolution instead of standard convolution to achieve a tradeoff between latency and accuracy [469], [470], [471]. At the same time, it can realize many applications on mobile terminals such as target detection, target classification, and face recognition. ShuffleNet is mainly characterized by high running speed, which improves the efficiency of implementing this model on hardware [472], [473]. On the contrary, some methods achieve smaller sizes by trimming the existing network structure. In this research line, taking object detection as an example, Tiny YOLO [474], [475] is designed as a slim implementation of YOLO series and can run on low-capacity drones for real-time detection. Such lightweight implementations play big roles in carrying out various tasks efficiently on mobile vision platforms.

2) Big Data Analysis in Cloud: It is also practical to reduce the calculation burden on the mobile platform by characterizing and understanding real-life phenomena from large-scale visual data in the cloud, including individual traits and interaction patterns. Big data analysis, as one of the most promising technologies for breaking the restrictions in current data management, can help in the efficient implementation of algorithms deployed on mobile platforms from the following aspects [476].

1) Process massive amounts of data efficiently for internal operations and interactions of mobile platforms involving large amounts of data traffic.
2) Accelerate decision-making procedure to reduce the risk level and promote the users’ experience, which is crucial for reliable and safe decisions of mobile platforms.
3) Combine with machine learning or data mining techniques to reveal new patterns and values in the data, thereby helping to predict and prepare strategies for possible future events.
4) Integrate and compactly code the large variety of data sources, including mobile sensor data, and the database is still being updated online.

As an example of applying big data analysis to low-bandwidth data sharing and transmission among a mobile platform network, an image encoding approach is designed to reduce the size of the captured visual data based on cloud computing. Specifically, the features extracted from the input image can be used for referencing the correlated counterparts in the cloud and achieving high fidelity at thousands to one compression ratio [477]. This new encoding scheme can achieve higher performance than conventional compressors and is generally applicable for various visual data from CI setups equipped on mobile platforms, for example, image, video, 3-D point cloud, and spatial–spectral volume. In addition, big data analysis is expected to be employed in emerging 5G networks to improve the performance of mobile networks [478].

3) Distributed Edge Computing: Designed for effective computations and resource distributions under energy and resource constraints [479], distributed computing is another way to lessen the computing pressure of mobile platforms. As some mobile platforms are required to cooperate with other similar platforms, such as drone swarms, joint computing and decision-making pose a more complex challenge to the integration of CI and AI on intelligent mobile platforms. To this end, the rapid development of communication and edge computing technologies provides a promising combination. With a top-down architecture, most of the data are consumed at the edge of the mesh structure, and network traffic is distributed [480] and can be performed in parallel.

At the level of edge cloud, which can be a roadside unit or a cellular tower having more power than the single mobile platform, the computers execute communications that need to be dealt with immediately, such as real-time location and mapping [481], or some resource-demanding calculation infeasible on a single mobile device. In addition, edge servers will also conduct cross-detection and verification for the acquisition and decision on a single mobile device. At the level of central cloud, cloud computing technology provides a large amount of computing power, storage resources, and databases to assist complex large-scale computing tasks and make decisions. This will help the high-level management and large-scale planning. Moreover, it is robust to recovery from the failure of some edge servers or single mobile platforms.

In summary, cloud computing and edge computing play roles in a manifold way in addition to promoting information processing and mechanical control through streaming with local processing platforms. First, it can be used to control the settings of computational cameras for better acquisition [482] as well as to manage large-scale multimodality camera networks to reduce the data deluge [483]. Second, the modularized software framework composed of cloud computing and edge computing can effectively hand out multiple image-processing tasks to distributed computing resources. For example, one can only transmit the preprocessed image information to avoid insufficiency in storage, bandwidth, and computing power of mobile platforms [484]. Third, at the decision-making stage, cloud and edge computing technologies have changed the ways of communication and underlying management of mobile platform networks, by sharing traffic resources via networks for more effective management and planning. This has wide applications in the field of intelligent transportation [485]. For example, various algorithms can be performed on the cloud for mobile-platform-cluster scheduling in either centralized [486], [487], [488], [489], [490], [491], [492] or distributed [493], [494], [495], [496], [497], [498], [499], [500] manner. Similarly, cloud (and edge) computing should be applied to control drone fleet formation to improve real-time performance.

In conclusion, the novel algorithms represented by lightweight deep neural networks enable efficient handling of CI data on mobile vision platforms. For platforms demanding computing power beyond its load, 5G communication technology, big data analysis, and distributed
computing will further reduce the calculation burden to assist high-quality imaging, complex analysis (such as detection, tracking, and navigation), timely decision-making, and scheduling. These aforementioned strategies would promote practical applications of CI + AI-based mobile vision: self-driving cars are intelligent transportation vehicles that will most likely be put into use in the near-future; meanwhile, UAV and unmanned underwater vehicles can potentially be used in both military and civilian workplaces.

**B. Technologies for Development of CI + AI**

1) Biological Vision for CI + AI: After millions of years of evolution, biological vision systems have achieved delicate and high performance, especially in terms of semantic perception. Thus, reverse engineering of biological vision systems or biomimetic intelligent imaging is important. As an inspiration, current research on CI draws on the mechanisms of biological vision systems, including humans.

For example, borrowing the “ROI” strategy in the human visual system, computational visual attention systems selectively detect ROIs in the target scene and have great applications in fields such as computer vision, cognitive systems, and mobile robotics [501]. The event camera draws inspiration from the transient information pathway of “photoreceptor cells–bilevel cells–ganglion cells” and records the event of abrupt intensity changes instead of conducting regular sampling in space and time domains. By directly performing detection from the event streams [502], such sensing mechanism obtains qualitative improvement in both perception speed and sensitivity [503], [504], especially under low-light conditions. As a silicon retina, an event stream from an event camera can be used to track accurate camera rotation while building a persistent and high-quality mosaic of a scene that is super-resolved and of an HDR [503]. Using only an event camera, depth estimation of features and six degrees-of-freedom pose estimation can be realized, thus enhancing the 3-D dynamic reconstruction [505].

In addition, evidence from neuroscience suggests that the human visual system uses a segmentation strategy based on identifying discontinuities and grouping them into contours and boundaries. This strategy provides better ideas for image segmentation, and some studies have shown that the perceptual performance of visual contour grouping can be improved through perceptual learning [506]. Moreover, the information about the object and self-motion in the human brain has realized a fast, feedforward hierarchical structure. The visual processing unit integrates the motion to calculate the local moving direction and speed, which has become an important reference for optical flow research [506].

Due to the incomplete understanding of biological vision, adopting it to machine vision is still limited to individual optical elements and the overall imaging mechanism still follows the traditional imaging scheme; thus, the system is of high complexity, large size, and high power consumption, which limits the mobile applications. To this end, one can draw inspiration from the working mechanism of animals’ eyes and visual cortex to build new imaging systems or develop task-oriented AI algorithms. For example, the visual system of mantis shrimp is sensitive to up to 12 wavebands [507], which may motivate the invention of novel bionic multispectral imaging sensors. As another illustration, consider the frog that does not develop any specialized muscles in the eyes to regulate its crystalline lens but does excel at observing fast-moving objects, which benefits from the ability to detect different types of features by its four layers of visual cortex cells. The current electronic frog eye device has borrowed this scheme but still requires multiple sensors for different feature detection, which might be bypassed by integrating the sensors into one hierarchical architecture with the aid of new materials. In addition, understanding how frogs jointly process the features to generate quick decisions or strategies can also inspire the building of compact deep neural networks for high-efficiency motion feature extraction. Based on the structural and functional analysis of the biological vision system, brain-like intelligence for the machine vision system might be a good research direction for the next generation of mobile vision.

2) Brain Science for CI + AI: The information path of the human visual system eventually leads to the visual cortex in the brain. Since Hubel and Wiesel’s groundbreaking study [508] of information processing in the primary visual system of cats in 1959, researchers have initially understood the general characteristics of neuron responses at various levels from the retina to the higher-order visual cortex, establishing effective visual computing models of the brain. Given the human brain’s enormous advantages over other systems in terms of response time, energy efficiency, bandwidth utilization, and accuracy, understanding how information is processed and drawing inspiration from the structure of the brain could accelerate the development of AI.
The available knowledge is that biological neurons have complex dendritic structures [509], which is the structural basis for their ability to efficiently process visual information. Existing artificial neural networks, including deep learning, can be recognized as an oversimplification of biological neural networks [510]. At the level of the neural circuit, each neuron is interconnected with other neurons through thousands or even tens of thousands of synapses, where procedures similar to the backpropagation in deep learning are performed to modify the synapses to improve the behaviors [511], and the number of inhibitory synapses often significantly exceeds that of excitatory synapses. This forms a strong nonlinearity as the physical foundation on which the brain can efficiently complete basic functions such as target detection, tracking, and recognition, and advanced functions such as visual reasoning and imagination. In addition to revealing the neural architecture and organization for processing visual information, a functional brain atlas is also a significant tool for associating animal behaviors with neuron dynamics under different visual stimuli or environments [512]. In this case, the functional imaging depicted by wearable devices in animal behavior experiments may serve as a reference for the development of multitarget low-power AI algorithms. It is possible to borrow from the brain's functional specialization and organization of different behaviors, from either humans or other animals, to develop task-specific AI algorithms or design deep neural networks that share modules among multiple tasks. Using neural networks to simulate the information processing strategy of biological neurons and synapses, brain-like computing has become a new generation of low-cost computing.

3) Brain-Like Computing for CI + AI: In the next generation of mobile vision, the major function of computers is to transform from calculation to intelligent processing and information retrieval. On the one hand, intelligent CI systems are no longer for recording, but to provide robust, reliable guidance for intelligent applications. This processing will be computing intensive and of huge power consumption [513]. On the other hand, mobile vision systems must meet the demanding requirements of real applications, such as high throughput, high frame rate, and weak illumination, and low computation cost to work under a limited lightweight budget. Therefore, optoelectronic computing has become a hot topic due to its stronger computing power and potentially lower power consumption.

For common CNNs, the number of parameters and nodes in CNNs increases dramatically with the performance requirement, and thus, the power consumption and memory demands grow correspondingly. Recently, the hybrid architecture combining optoelectronic computing and traditional neural networks has been used to improve current network performance in terms of speed, accuracy, and energy consumption. Benefiting from the broad bandwidth, high computing speed, high interconnectivity, and inherent parallel processing characteristics of optical computing, the optical neural network (ONN) performs expensive matrix multiplication of a fully connected layer by optics to increase the computation speed [514]. Furthermore, by placing an optical convolution layer for CI systems at the front end of the CNN, the network performance can be maintained with significantly reduced energy consumption [515]. In addition to locally improving the computational performance of deep networks, a photonic neurosynaptic network based on wavelength division multiplexing technology has recently been proposed [516]. Unlike traditional computing architectures, this all-optical network processes information in a more brain-like manner, which enables fast, efficient, and low-energy computing.

Recently, a new type of brain-like computing architecture, hybrid Tianjic chip architecture that draws on the basic principles of brain science, has been proposed. As a brain-like computing chip that integrates different structures, Tianjic can support neural network models in both computer science and neuroscience [517], such as artificial neural networks and spiking neural networks [518], [519], to demonstrate their respective advantages. Compared with the current world-leading chips, Tianjic is flexible and expandable while enjoying a high integration level, high speed, and large bandwidth. Inspired by this, brain-like computing technologies might significantly aid self-driving, UAVs, and intelligent robots and also provide higher computing power and low-latency computing systems for the Internet industry.

4) New Materials for CI + AI: Using coded acquisition schemes, CI systems often involve complex optical designs and modulations. Therefore, it is desirable to adopt lightweight modules on mobile platforms. For instance, metasurfaces are surfaces covered by ultrathin plasmonic structures [520]. Such emerging materials have the advantage of being thin but powerful in controlling the phase of light and have been widely used in optical imaging. Ni et al. [521] proved that the ultrathin metasurface hologram with a thickness of 30 nm can work in the visible light range. This technology can be used to specifically modulate the amplitude and phase to produce high-resolution low-noise images, which serve as the basis for new ultrathin optical devices. Furthermore, Zheng et al. [522] demonstrated that the geometric metasurfaces consisting of an array of plasmonic nanorods with spatially varying orientations can further improve the efficiency of holograms. In addition, the gradient metasurface based on this structure can serve as a 2-D optical element, such as ultrathin gratings and lenses [523]. The combination of semiconductor manufacturing technology and metasurfaces can effectively improve the quality and efficiency of imaging and significantly save space. For example, by using metasurface, Holsteen et al. [524] improved the spatial resolution of the 3-D information collected by ordinary microscopes. The ultrathin nature of metasurfaces also provides other advantages such as higher energy conversion efficiency [525]. In summary,
the flat and highly integrated optical elements made of metasurface can achieve local control of phase, amplitude, and polarization, and have the properties of high efficiency and small size. In addition to optical modulation, it is feasible to introduce new fabrication techniques to circumvent miniaturization issues in conventional optical imaging systems or produce new imaging schemes. For example, we can replace bulky and fragile relay optics with novel thin optical elements with nanoscale microstructures, and utilize nanofabricated thin films to directly translate the subtle density changes in a transparent medium into color.

These devices are expected to replace traditional optical devices in many fields, including mobile vision or wearable optics, to overcome some distortions or noise, as well as to improve imaging quality, efficiency, and compactness.

The fields listed above will play hybrid (sometimes complementary) roles to enhance future mobile vision systems with CI and AI. For example, the research on biological vision and brain science promotes the recognition of a biological system, thus helping the development of brain-like computing, which will eventually be used in mobile vision systems to improve performance.

V. CONCLUSION
In the current digital era, mobile vision plays an important role. The explosively growing visual data from mobile vision platforms are imposing challenges to storage, bandwidth, and analysis, which call for a new-generation motion vision. Inspired by vast applications in computer vision and machine intelligence, CI systems have begun to change the way information is captured. In contrast to the style of “capturing images first and processing afterward” used in conventional cameras, in CI, more information can be captured indirectly. By integrating AI into CI, as well as the new techniques of 5G and beyond plus edge computing, we believe that a new revolution in mobile vision is around the corner.

This study reviewed the history of mobile vision and presented a survey of using AI for CI. Mobile vision has been developed for more than three decades, supported by the invention of digital cameras, the Internet, communication networks, intelligent terminals, and so on. Recently, mobile platforms have been seeking new imaging schemes capable of bypassing the grand challenges of high-throughput visual data. CI was born from a variety of practical demands in photography and is in line with the demands of mobile vision systems. Some CI topics, such as hyperspectral imaging, high-speed imaging, and SR imaging, are committed to improving the imaging performance along certain dimension(s) of visual information. On the contrary, some other topics focus on extending the imaging capability in harsh environments, such as low-light, NLOS, or HDR imaging. The CI techniques are at the convergence of multiple disciplines, including optical engineering, signal processing, AI, and computer vision. The core concept in CI is to capture the visual information in an encoded manner and subsequently decode the target information from the encoded recordings. The former is implemented with various photoelectric modulation devices, while the latter step involves corresponding algorithm design. Both stages significantly benefit from the rapidly developing AI techniques.

This combination of CI and AI inspires new designs, capabilities, and applications for imaging systems. With the development of both fields, their integration has shown an increasingly integrative trend. The most intuitive idea is to employ AI techniques to solve the inverse problem of decoding, where various deep neural networks have been designed and achieved state-of-the-art performance in most CI tasks. However, this combination, which solely uses AI as a solver, does not truly incorporate AI into CI design. On the contrary, the joint optimization of the optical encoding parameters/patterns and computational decoding algorithm in a CI system produces imaging designs guided by AI, which largely changes the role of AI that only processes the raw acquisition of CI. Furthermore, there have been CI systems designed for specific scenarios or tasks, where AI techniques for representing specific scenarios or even semantic analysis can serve as potential guidance to substitute the handcrafted or empirical designs. In the future, the deep connection between CI and AI can be advanced by a broad range of novel technologies such as big data analysis, distributed computation, and new progress in other disciplines.

On mobile vision platforms, the connections between CI and AI have exhibited great promise and put forward some inherent requirements for low data burden, high response speed, and compact hardware when implementing intelligent imaging methods. Specifically, we take autonomous driving as a typical example, where joint multimodal sensing has become the mainstream and synchronous processing of data from multiple sensors places a heavy computation burden, to show the challenges and potential solutions. On the one hand, mining features from multimodal input and CI cameras in the cloud is expected to become an effective strategy to reduce the offline data burden, where the emerging 5G technology serves as a key tool to manage and associate the data and resources. On the other hand, deep learning algorithms in computer vision are marching toward higher performance and updating continuously, but it unfortunately means more demanding graphic computation, so lightweight algorithms for specific tasks and resource conditions effectively cope with the challenge, and cloud/edge computing is capable of offloading the pressure as well. Moreover, more lightweight CI systems and algorithms are expected for low-capacity mobile platforms such as drones, smartphones, and wearable devices.

In the future, interdisciplinary integration has become a trend toward higher performance, faster computing speed, and more compact systems. Determining the mechanism by which biological visual systems capture and analyze
visual information will further benefit the advancement of CI schemes and AI algorithms. The continuous emergence of new materials will also contribute to the birth of a new generation of miniaturized and refined optical imaging systems. We expect and believe that the near future will witness a closer integration between CI and AI, as well as advances in next-revolution mobile vision systems, which would facilitate or even change our lives.

**REFERENCES**

[1] W. S. Boyle and G. E. Smith, "Charge coupled semiconductor devices," Bell Syst. Tech. J., vol. 49, no. 4, pp. 587–593, Apr. 1970.

[2] A. Chidester, J. H. Kunz, T. C. Mercer, and K. S. Schulz, "Recording automotive crash event data," in Proc. Trumpl. Recording, 2000 Beyond, Int. Symp. Trumpl. Recorders, 1999, pp. 85–98.

[3] J. Callahan, "The first camera phone was sold 20 years ago, and it’s not what you might expect," Android Authority, II, USA, 2021. [Online]. Available: https://www.androidauthority.com/first-camera-phone-anniversary-993492/.

[4] J. Dalympie, "Apple releases the MacBook," Macworld, vol. 23, no. 7, pp. 20–21, 2006.

[5] S. Chen, J. Zhao, and Y. Peng, "The development of TD-SCDMA 3G to TD-LTE-advanced 4G from 1998 to 2013," IEEE Wireless Commun., vol. 21, no. 6, pp. 167–175, Dec. 2014.

[6] N. J. Nilsson, "Shakey the robot," SRI Int., Menlo Park, CA, USA, Tech. Rep. 323, 1984. [Online]. Available: https://seanborman.com/publications/SRIViews.pdf

[7] T. Jochem, D. Pomerleau, B. Kumar, and J. Armstrong, "BANS: A portable navigation platform," in Proc. Intell. Vehicles Symp., 1995, pp. 107–112.

[8] V. Altman, S. McLaughlin, M. J. Pudged, V. K. Goyal, A. Hero, and P. Fiacco, "Quantum-inspired computational imaging," Science, vol. 361, no. 6403, pp. eaat2289, Aug. 2018.

[9] N. S. Majumdar, G. W. Euilis, and R. A. Achale, "Computational imaging," Adv. Opt. Photon., vol. 10, no. 2, pp. 409–483, Jun. 2018.

[10] D. J. Brady et al., "Smart cameras," 2020, arXiv:2002.04705.

[11] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," IEEE Signal Process. Mag., vol. 35, no. 1, pp. 20–36, Jan. 2018.

[12] J. H. R. Chang, C.-L. Li, B. Pizcero, B. V. K. Vijaya Kumar, and A. C. Sankaranarayanan, "One network for all—Solving linear inverse problems using deep projection models," in Proc. IEEE Int. Conf. Civ. Vis. (ICCV), Oct. 2017, pp. 5889–5898.

[13] G. Govig, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," IEEE J. Sel. Areas Inf. Theory, vol. 1, no. 1, pp. 39–56, May 2020.

[14] G. D. Sharp, "Google Pixel: The Complete Beginner’s Guide, vol. 1. North Charleston, SC, USA: CreateSpace Independent Publishing Platform, 2017.

[15] N. Wadlawa et al., "Synthetic depth-of-field with a single-camera mobile phone," ACM Trans. Graph., vol. 37, no. 4, pp. 1–13, Aug. 2018.

[16] V. Bavarevsky, Y. Kartynnik, A. Vakunov, R. A. Athale, and K. S. Schulz, "Recording automotive crash event data," in Proc. Trumpl. Recording, 2000 Beyond, Int. Symp. Trumpl. Recorders, 1999, pp. 85–98.

[17] J. Lehtinen, T. Aila, J. Chen, S. Laine, and F. Heide, "Doppler time-of-flight imaging," in Proc. IEEE Conf. Civ. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2289–2297.

[18] S. Su, F. Heide, G. Wetzstein, and H. Weidrich, "Deep end-to-end time-of-flight imaging," in Proc. IEEE Conf. Civ. Vis. Pattern Recognit., Jun. 2018, pp. 6383–6392.

[19] W. Guo, C. Zhao, H. Yu, M. Chen, W. Xu, and S. Han, "Three-dimensional gimbal ghost imaging LiDAR via sparsity constraint," Sci. Rep., vol. 6, no. 1, May 2016.

[20] A. Kirmani et al., "First-photon imaging," Science, vol. 343, no. 6166, pp. 58–61, 2014.

[21] M. O’Toole, F. Heide, D. B. Lindell, K. S. Han, and G. Wetzstein, "FramedToL: Structured light 3D imaging using single-photon sensors," in Proc. IEEE Conf. Civ. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 10192–10201.

[22] D. B. Lindell, M. O’Toole, and G. Wetzstein, "Towards transmittance imaging with interactive rates with single-photon detectors," in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP), May 2018, pp. 1–8.

[23] F. Heide, S. Diamond, D. B. Lindell, and G. Wetzstein, "Sub-picosecond photon-efficient 3D imaging using single photon sensors," Sci. Rep., vol. 8, no. 1, pp. 1–8, Dec. 2018.

[24] J. Wu, C. Zhang, Z. Zhang, Z. Zhang, W. T. Freeman, and J. B. Tenenbaum, "Learning shape priors for single-view 3D completion and reconstruction," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 646–662.

[25] Z. Sun, D. B. Lindell, O. Solgaard, and G. Wetzstein, "SPADine: Deep RGB-SPAD sensor fusion assisted by monocular depth estimation," Opt. Exp., vol. 28, no. 10, pp. 14948–14962, May 2020.

[26] J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3D object detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 10192–10201.

[27] Y. Wu, V. Boomminathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "PulseCam3D—Learning phase masks for passive single-view depth estimation," in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP), May 2019, pp. 1–12.

[28] Y. Ba et al., "Deep shape from polarization," in
K. Tomiyasu, "Tutorial review of...

S. P. Kim, N. K. Bose, and H. M. Valenzuela, "...

J. Tian and K.-K. Ma, "A survey on...

M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and J. Holloway, Y. Wu, M. K. Sharma, O. Cossairt, and R. M. Willett, M. F. Duarte, M. A. Davenport, and M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Robust gesture...

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, and M. Zhou, "Fast fourier single pixel imaging via binary illumination," Sci. Rep., vol. 7, no. 1, 12097, Sep. 2017.

M. Abbas, M. Elhamshary , H. Rizk, M. Torki, and J. Holloway, Y. Wu, M. K. Sharma, O. Cossairt, and R. M. Willett, M. F. Duarte, M. A. Davenport, and M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Robust gesture...

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.

Z. Zhang, Z. Tian, Y. Zhang, M. Zhou, and Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic subpixel shifted pictures," CVGIP, Graph. Models

IEEE Trans. Image Process., vol. 66, no. 1, pp. 763–776, Jan. 2017.
Suo et al.: CI and AI: The Next Revolution of Mobile Vision

1634 PROCEEDINGS OF THE IEEE | Vol. 111, No. 12, December 2023

Authorized licensed use limited to the terms of the applicable license agreement between a subscriber and IEEE. Restrictions apply.
Z. Meng, J. Ma, and X. Yuan, “End-to-end low cost compressive spectral imaging with spatial–spectral self-attenuation,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 187–204.

T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, “Deep circularly symmetric autoregressive prior for spectral compressive imaging,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 16211–16220.

C. Chen et al., “Recurrent neural networks for snapshot compressive imaging,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 2, pp. 2264–2281, Feb. 2023.

Y. Cai et al., “Machine learning and spatial–spectral self-attention for efficient hyperspectral image reconstruction,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 112–125.

Y. Cai et al., “Coarse-to-fine sparse transformer for hyperspectral image reconstruction,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2022, pp. 663–670.

Y. Cai et al., “Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging,” in IEEE Trans. Adv. Inf. Process. Syst., vol. 35, 2022, pp. 37749–37761.

L. Wang, Z. Wu, X. Yuan, and J. Suo, “Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer,” Photon. Res., vol. 10, no. 8, pp. 1848–1858, 2022.

Y. Wang, X. Su, Z. Meng, and Z. Tao, “Modeling mask uncertainty in hyperspectral image reconstruction,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2022, pp. 112–125.

L. Wang, M. Cao, Y. Zhong, and X. Yuan, “Spatial–temporal transformer for video snapshot compressive imaging,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 10, pp. 7093–7111, Sep. 2019.

X. Yuan, L. Liu, J. Suo, and Q. Dai, “Plug-and-play algorithms for large-scale snapshot compressive imaging,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 1444–1454.

X. Yuan, L. Liu, J. Suo, S. Zheng, and Q. Dai, “Spatial–temporal transformer for video snapshot compressive imaging,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2022, pp. 112–125.

L. Wang, M. Cao, Y. Zhong, and X. Yuan, “Modeling mask uncertainty in hyperspectral image reconstruction,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2022, pp. 112–125.

Y. Cai et al., “Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging,” in IEEE Trans. Adv. Inf. Process. Syst., vol. 35, 2022, pp. 37749–37761.

L. Wang, Z. Wu, X. Yuan, and J. Suo, “Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer,” Photon. Res., vol. 10, no. 8, pp. 1848–1858, 2022.

Y. Wang, X. Su, Z. Meng, and Z. Tao, “Modeling mask uncertainty in hyperspectral image reconstruction,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2022, pp. 112–125.

L. Wang, M. Cao, Y. Zhong, and X. Yuan, “Spatial–temporal transformer for video snapshot compressive imaging,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 10, pp. 7093–7111, Sep. 2019.

Z. Meng, Z. Yu, K. Xu, and X. Yuan, “Self-updating neural networks for spectral snapshot compressive imaging,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 2662–2611.

Z. Wu, C. Yang, X. Su, and X. Yuan, “Adaptive deep PnP algorithm for video snapshot compressive imaging,” Int. J. Comput. Vis., vol. 131, no. 7, pp. 1662–1679, Mar. 2023.

K. Mirafzal, “Performance limits for computational photography,” in Proc. Fridge. Berlin, Germany: Springer, 2014, pp. 663–670.

X. Yuan and S. Han, “Single-pixel neuron imaging with artificial intelligence: Breaking the barrier in multi-parameter imaging, sensitivity and spatial resolution,” Innovation, vol. 2, no. 2, 2021, Art. no. 100100.

I. L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D, Nonlinear Phenomena, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

X. Luo, H. Li, and L. Carin, “Generalized alternating projection for weighted-l_1 minimization with applications to model-based compressed sensing,” SIAM J. Imaging Sci., vol. 7, no. 2, pp. 797–823, Jan. 2014.

X. Yuan, “Generalized alternating projection based total minimization for compressive sensing,” Dissertation, IMTech, SUA, arXiv:1910.07928.

J. Y. Cai et al., “Performance limits for computational photography,” in Proc. Fridge. Berlin, Germany: Springer, 2014, pp. 663–670.

X. Yuan and S. Han, “Single-pixel neuron imaging with artificial intelligence: Breaking the barrier in multi-parameter imaging, sensitivity and spatial resolution,” Innovation, vol. 2, no. 2, 2021, Art. no. 100100.

I. L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D, Nonlinear Phenomena, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

X. Luo, H. Li, and L. Carin, “Generalized alternating projection for weighted-l_1 minimization with applications to model-based compressed sensing,” SIAM J. Imaging Sci., vol. 7, no. 2, pp. 797–823, Jan. 2014.

X. Yuan, “Generalized alternating projection based total minimization for compressive sensing,” Dissertation, IMTech, SUA, arXiv:1910.07928.

J. Y. Cai et al., “Performance limits for computational photography,” in Proc. Fridge. Berlin, Germany: Springer, 2014, pp. 663–670.

X. Yuan and S. Han, “Single-pixel neuron imaging with artificial intelligence: Breaking the barrier in multi-parameter imaging, sensitivity and spatial resolution,” Innovation, vol. 2, no. 2, 2021, Art. no. 100100.

I. L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Phys. D, Nonlinear Phenomena, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

X. Luo, H. Li, and L. Carin, “Generalized alternating projection for weighted-l_1 minimization with applications to model-based compressed sensing,” SIAM J. Imaging Sci., vol. 7, no. 2, pp. 797–823, Jan. 2014.

X. Yuan, “Generalized alternating projection based total minimization for compressive sensing,” Dissertation, IMTech, SUA, arXiv:1910.07928.

J. Y. Cai et al., “Performance limits for computational photography,” in Proc. Fridge. Berlin, Germany: Springer, 2014, pp. 663–670.
Access, vol. 8, pp. 115448-115460, 2020.

[48] J. Yin, Z. Qian, and W. Yang, “UAV cluster-based video surveillance system optimization in heterogeneous communication of smart cities,” IEEE Access, vol. 8, pp. 55654-55664, 2020.

[49] H. Dai, H. Zhang, M. Hua, C. Li, Y. Huang, and B. Wang, “How to deploy multiple UAVs for providing communication service in an unknown region?” IEEE Wireless Commun. Lett., vol. 8, no. 4, pp. 1276–1279, Aug. 2019.

[50] Ç. Tanlı, C. Warty, and E. Obiedat, “Collaborative mission planning for UAV cluster to optimize relay distance,” in Proc. IEEE Aerosp. Conf., Mar. 2013, pp. 1–11.

[51] T. Shima, S. J. Rasmussen, A. G. Sparks, and Ç. Tanil, “Simultaneous mosaicing and tracking with an event camera,” in Proc. Brit. Mach. Vis. Conf., 2014, pp. 566–576.

[52] A. Meier, H. Berbeglia, G. Gallego, T. Delbruck, and D. Scaramuzza, “The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM,” Int. J. Robot. Res., vol. 36, no. 2, pp. 142–149, Feb. 2017.

[53] H. Kim, S. Keuninger, and A. Davison, “Real-time 3D reconstruction and 6-DoF tracking with an event camera,” in Proc. EAR Conf. Comput. Vis. Comput. Imaging, 2016, pp. 349–364.

[54] N. K. Medhati, H. Neumann, G. S. Masson, and P. Korporst, “Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision,” Comput. Vis. Image Understand., vol. 150, pp. 1–30, Sep. 2016.

[55] J. Marshall and J. Oberwinkler, “The colourful world of the mantis shrimp,” Nature, vol. 401, no. 6756, pp. 473–474, Oct. 1999.

[56] D. H. Hubel and T. N. Wiesel, “Early exploration of the visual cortex,” Neuroscientist, vol. 20, no. 3, pp. 401–412, Mar. 1998.

[57] A. Kumar, R. B. Hetrick, and A. Artenstein, “Spiking activity propagation in neuronal networks: Reconciling different perspectives on neural coding,” Nature Rev. Neurosci., vol. 11, no. 9, pp. 613–627, Sep. 2010.

[58] J. L. Sanchez-Lopez, C. Sampedro, D. Cazzato, and H. Voos, “Deep learning based semantic situation awareness system for multilotor aerial robots using LiDAR,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), May 2014, pp. 584–595.

[59] M. Saska, V. Vonaik, J. Chudoba, J. Thomas, G. Letanov, and V. Kumar, “Swarm distribution and deployment for cooperative surveillance by micro-air vehicles,” J. Intell. Robot. Syst., vol. 84, nos. 1–4, pp. 469–492, Dec. 2016.

[60] J. L. Sanchez-Lopez, M. Castillo-Lopez, and H. Voos, “Semantic situation awareness of ellipse shapes via deep learning for multilotor aerial robots with a 2D LiDAR,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), Sep. 2020, pp. 1014–1023.

[61] J. L. Sanchez-Lopez, C. Sampedro, D. Cazzato, and H. Voos, “Deep learning based semantic situation awareness system for multilotor aerial robots using LiDAR,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), Jun. 2019, pp. 899–908.

[62] M. Saska, J. Vakula, and L. Přeučil, “UAV cluster-based video surveillance system optimization in heterogeneous communication of smart cities,” IEEE Access, vol. 8, pp. 115448-115460, 2020.

[63] H. Dai, H. Zhang, M. Hua, C. Li, Y. Huang, and B. Wang, “How to deploy multiple UAVs for providing communication service in an unknown region?” IEEE Wireless Commun. Lett., vol. 8, no. 4, pp. 1276–1279, Aug. 2019.

[64] Ç. Tanlı, C. Warty, and E. Obiedat, “Collaborative mission planning for UAV cluster to optimize relay distance,” in Proc. IEEE Aerosp. Conf., Mar. 2013, pp. 1–11.

[65] T. Shima, S. J. Rasmussen, A. G. Sparks, and K. M. Passino, “Multiple task assignments for cooperating uninhabited aerial vehicles using genetic algorithms,” Comput. Oper. Res., vol. 33, no. 11, pp. 3252–3269, Nov. 2006.

[66] R. K. Sharma and D. Ghoose, “Collision avoidance between UAV clusters using swarm intelligence techniques,” Int. J. Syst. Sci., vol. 40, no. 5, pp. 521–538, May 2009.

[67] R. A. Clark et al., “Autonomous and scalable control for remote inspection with multiple aerial vehicles,” Robot. Auto. Syst., vol. 87, pp. 258–268, Jan. 2017.

[68] M. Saska et al., “Autonomous deployment of swarms of micro-aerial vehicles in cooperative surveillance,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), May 2014, pp. 584–595.

[69] M. Saska, V. Vonaik, J. Chudoba, J. Thomas, G. Letanov, and V. Kumar, “Swarm distribution and deployment for cooperative surveillance by micro-air vehicles,” J. Intell. Robot. Syst., vol. 84, nos. 1–4, pp. 469–492, Dec. 2016.

[70] J. L. Sanchez-Lopez, M. Castillo-Lopez, and H. Voos, “Semantic situation awareness of ellipse shapes via deep learning for multilotor aerial robots with a 2D LiDAR,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), Sep. 2020, pp. 1014–1023.

[71] J. L. Sanchez-Lopez, C. Sampedro, D. Cazzato, and H. Voos, “Deep learning based semantic situation awareness system for multilotor aerial robots using LiDAR,” in Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS), Jun. 2019, pp. 899–908.

[72] M. Saska, J. Vakula, and L. Přeučil, “UAV cluster-based video surveillance system optimization in heterogeneous communication of smart cities,” IEEE Access, vol. 8, pp. 115448-115460, 2020.

[73] H. Dai, H. Zhang, M. Hua, C. Li, Y. Huang, and B. Wang, “How to deploy multiple UAVs for providing communication service in an unknown region?” IEEE Wireless Commun. Lett., vol. 8, no. 4, pp. 1276–1279, Aug. 2019.

[74] Ç. Tanlı, C. Warty, and E. Obiedat, “Collaborative mission planning for UAV cluster to optimize relay distance,” in Proc. IEEE Aerosp. Conf., Mar. 2013, pp. 1–11.

[75] T. Shima, S. J. Rasmussen, A. G. Sparks, and K. M. Passino, “Multiple task assignments for cooperating uninhabited aerial vehicles using genetic algorithms,” Comput. Oper. Res., vol. 33, no. 11, pp. 3252–3269, Nov. 2006.
with the Department of Electrical and Computer Engineering, Duke University, from 2012 to 2015, where he was working on compressive sensing and machine learning. In 2021, he joined Westlake University, Hangzhou, China, as an Associate Professor.

Dr. Yuan has been an Associate Editor of Pattern Recognition since 2019, International Journal of Pattern Recognition and Artificial Intelligence since 2020, and Chinese Optics Letters since 2021. He led the Special Issue of “Deep Learning for High Dimensional Sensing” in the IEEE Journal of Selected Topics in Signal Processing in 2022.

David J. Brady (Fellow, IEEE) received the B.A. degree in physics and math from the Macalester College, St. Paul, MN, USA, in 1984, and the M.S. and Ph.D. degrees in applied physics from the California Institute of Technology, Pasadena, CA, USA, in 1986 and 1990, respectively.

He was on the faculty of the University of Illinois, Urbana, IL, USA, from 1990 until moving to Duke University, Durham, NC, USA, in 2001. He was a Professor of electrical and computer engineering at Duke University and Duke Kunshan University, Kunshan, China, where he led the Duke Imaging and Spectroscopy Program (DISP) and the Computer Laboratory, respectively, from 2001 to 2020. In 2021, he joined the Wyant College of Optical Sciences, The University of Arizona, Tucson, AZ, USA, as the J. W. and H. M. Goodman Endowed Chair Professor. He is the author of the book Optical Imaging and Spectroscopy (Hoboken, NJ, USA: Wiley). His research interests include computational photography and microscopy, computer vision and graphics, brain science, and video communication.

Dr. Brady is a Fellow of the Optical Society of America (OSA) and Society of Photo-Optical Instrumentation Engineers (SPIE). He won the 2013 SPIE Dennis Gabor Award for his work on compressive holography.

Qionghai Dai (Senior Member, IEEE) received the M.E. and Ph.D. degrees in computer science and automation from Northeastern University, Shenyang, China, in 1994 and 1996, respectively. He has been a Faculty Member at Tsinghua University, Beijing, China, since 1997, where he is currently a Professor at the Department of Automation and an Adjunct Professor at the School of Life Science. His research areas include computational photography and microscopy, computer vision and graphics, brain science, and video communication.

Dr. Dai is an Associate Editor of the Journal of Visual Communication and Image Representation, IEEE Transactions on Neural Networks and Learning Systems, and IEEE Transactions on Image Processing. He is also an Academician of the Chinese Academy of Engineering.