Visual Perception Enabled Industry Intelligence: State of the Art, Challenges and Prospects

Jiachen Yang, *Member, IEEE,* Chenguang Wang, Bin Jiang, Houbing Song, *Senior Member, IEEE* and Qinggang Meng, *Senior Member, IEEE*

***Abstract*—Visual perception refers to the process of organizing, identifying and interpreting visual information in environmental awareness and understanding. With the rapid progress of multimedia acquisition technology, research on visual perception has been a hot topic in the academical field and industrial applications. Especially after the introduction of artificial intelligence theory, intelligent visual perception has been widely used to promote the development of industrial production towards intelligence. In this paper, we review the previous research and application of visual perception in different industrial fields such as product surface defect detection, intelligent agricultural production, intelligent driving, image synthesis and event reconstruction. The applications basically cover most of the intelligent visual perception processing technologies. Through this survey, it will provide a comprehensive reference for research on this direction. Finally, this paper also summarizes the current challenges of visual perception and predicts its future development trends.**

***Index Terms*—Visual perception, industrial application, artificial intelligence**

I. INTRODUCTION

AMONG the five senses, vision provides a wealth of information for humans to observe and understand the world. Visual perception is an intuitive and internal observation and understanding process. Based on the feature-integration theory proposed by Treisman and Gelade, human visual perception is divided into feature registration stage and integration stage [1]. The first stage is that the visual system performs parallel and automated processing of features such as color, brightness, orientation, and size from the light stimulation mode. In the feature integration stage, the visual system locates the feature representations that are separate from each other. Through concentrated attention, just like glue, the original and separate features are integrated into a single object to complete visual perception. People always imitate the human visual perception mode, hope that the machine can convert the real-world three-dimensional information into pictures and videos through visual acquisition devices (CCD cameras, CMOS cameras, etc.), and then process, identify and explain these pictures or videos to understand the real environment, so that the machine has the ability to replace humans to complete various tasks. In industrial applications, this process is also called machine vision [2].

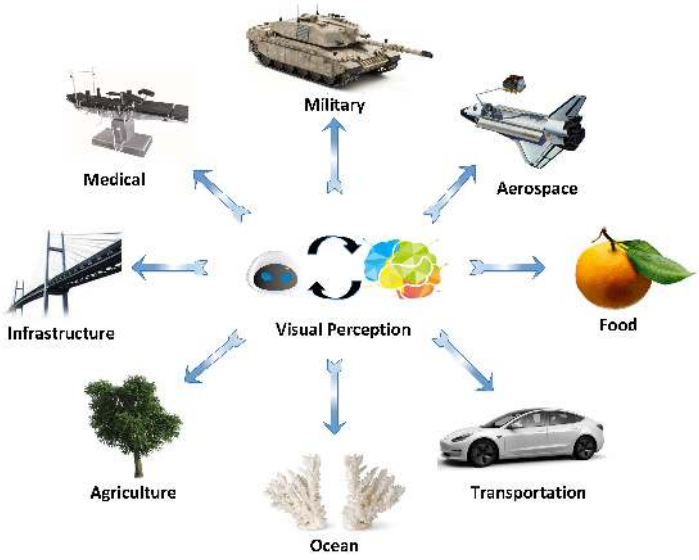

Fig. 1: Typical application fields of visual perception.

At present, although there is a large gap between the machine's visual perception comprehension ability and human visual perception level, it has a wide range of observations, which can not be observed by the human eye, such as infrared [3], microwave [4], ultrasound [5], etc. In addition, it is non-contact and can be widely used in long and harsh working environments [6], so the theoretical research and practical application of industrial visual information perception technology has been a research hotspot in various industrial fields. Especially in the era of Industry 4.0, visual perception technology is destined to become the leading technology [7].

With the rapid development of artificial intelligence technology, vision perception have become essential research topics pursued in the field of artificial intelligence. Computer vision is usually the pioneer and it is gradually used and developed in industry to promote automation and informatization of industrial production, enable the machine to autonomously perform intelligent activities such as analysis, reasoning, judgment, conception and decision-making [8] to save manpower, improve efficiency and reduce risk.

At present, intelligent visual perception technology has

This work was partially supported by National Natural Science Foundation of China (No. 61871283), the Foundation of Pre-Research on Equipment of China (No.61400010304) and Major Civil-Military Integration Project in Tianjin, China (No.18ZXJMTG00170).

J. Yang, C. Wang and B. Jiang are with School of Electrical and Information Engineering, Tianjin University, Tianjin, P.R. China. (e-mail: yangjiachen@tju.edu.cn; wcg5262@tju.edu.cn; jiangbin@tju.edu.cn).

H. Song is with the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, USA. (e-mail: h.song@ieee.org).

Q. Meng is with the Department of Computer Science, Loughborough University, Loughborough, UK (email: q.meng@lboro.ac.uk).

developed into an interdisciplinary discipline involving many fields such as artificial intelligence, neurobiology, psychophysics, computer science, pattern recognition, etc. Its technology mainly includes image and video generation [9], processing [10], quality evaluation [11] [12], and three-dimensional vision [13], object location, recognition, detection [14] and ranging [15], etc. Nowadays, visual perception has been widely used in aerospace [16], military [17], ocean [18], medical [19], infrastructure [20], agriculture [21], transportation [22], food [23] and other fields, as shown in Fig.1. Regarding the application fields of visual perception technology, there have been many review articles [24] [23] [22] [14], but most of them summarize the single application field or technology field and there has not been an review literature on the overall introduction of visual perception. In this survey, we introduce some applications and corresponding technologies of intelligent visual perception from a macro perspective, showing its advantages and its progressive impact on human production and life.

The outline of the contributions of this paper can be summarized as follows:

- Compared with other survey papers in the fields of visual perception, this survey summarizes the latest industrial applications of visual perception for the first time and not only does research on a single field including product surface defect detection, agricultural production intelligence, intelligent driving, image synthesis, event reconstruction and object pose measurement, these applications cover the aerospace, military, ocean, medical, infrastructure, agriculture, transportation, food fields shown in Fig. 1. At the same time, these applications also basically cover the latest technologies used in current visual perception, such as image and video processing, generation, object location, recognition, detection and 3D reconstruction, etc. Through these, reader can clearly understand the application field and technical composition of intelligent visual perception.
- We have summarized some of the limitations and introduced the main challenges faced by current visual perception from the perspective of users and researchers.
- We also put forward some visible development prospects of visual perception to reflect the theme of the paper, for which researchers in this field have pointed out the direction of scientific research.

The rest of this paper is organized as follows. Section II introduces some different industrial applications of vision perception and corresponding technologies, Section III gives some challenges faced by visual perception we can see now, Section IV illustrates its development prospects and trends, and Section V concludes this paper.

II. Industrial Applications

As visual perception becomes more intelligent and information-based, its applications have penetrated into many aspects of industry [25]. In this section, we introduce some applications of intelligent vision, including product surface defect detection, intelligent agricultural production, intelligent driving, image synthesis, event reconstruction and pose measurement. These applications often affect people's production and life that are hot spots for researchers and basically cover different intelligent visual perception technologies.

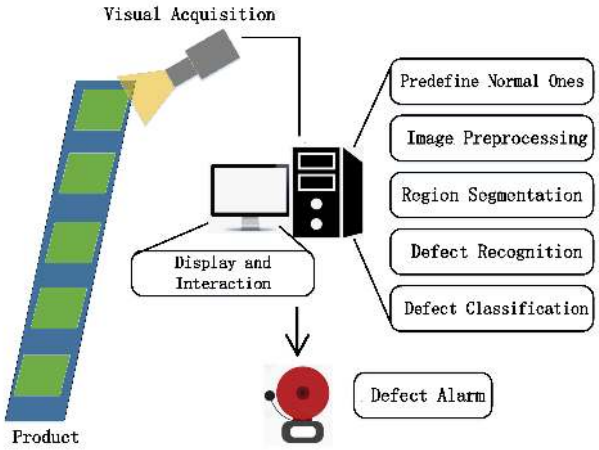

Fig. 2: General process of product surface defect detection.

A. Product Surface Defect Detection

A high-quality industrial product not only meets the requirements of clients and production enterprises in terms of performance, but also has good aesthetics and safety in appearance. The surface defects of some products will seriously affect the use of the products and even cause serious consequences. Different products have different definitions and types. Generally, surface defects are areas with uneven physical or chemical properties on the product surface, such as scratches, spots, holes, glass on metal surfaces, and inclusions, stains and damage on non-metal surfaces, etc. In the process of manufacturing products, the occurrence of surface defects is often unavoidable. Therefore, in the industrial field, the detection of surface defects has been paid much attention. Vision-based manual detection is the traditional detection method for product surface defects which has low sampling rate, low accuracy, poor real-time performance, low efficiency, high labor intensity, and is greatly affected by artificial experience and subjective factors. Detection based on machine vision methods can largely overcome the above drawbacks [26]. The usual detection process includes predefine nomal products, image preprocessing, target region segmentation, defect recognition and classification as shown in Fig.2.

Machine vision has been used to carry out a lot of work on the automatic detection and classification of textile defects.

1) Textile defects detection: For textile defects detection, there are mainly statistical-based methods, transform-domain-based methods, and model-based methods [27] [28]. Ngan *et al.* [29] proposed a method based on pattern primitives to detect defects in patterned textured fabrics, and used the symmetry of the primitives to calculate the moving energy variance between different primitives. The distribution of values is learned, the boundary conditions are determined, and then defects are identified. Chandra *et al.* [30] believed that because basic morphological operations were difficult to select

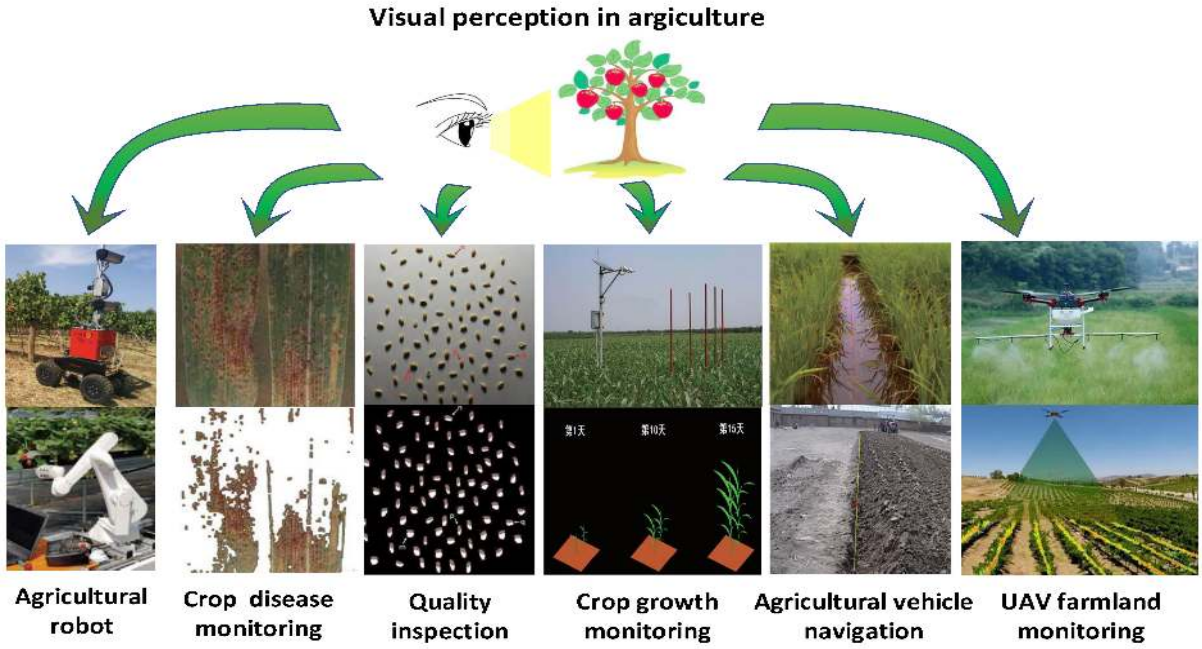

Fig. 3: Some applications of visual perception in intelligent agricultural production.

structural elements expediently, it was not easy to detect all kinds of defects appearing on woven fabrics. He proposed that utilizes artificial neural networks (ANN) to obtain structural elements and perform morphological reconstruction of the binary image of the fabric to detect defects. Chan *et al.* [31] used the simulated fabric model to get the relationship between the fabric structure in image space and frequency space and defined the two central space spectra in the three-dimensional spectrum, then used the difference between the simulation model and the real samples to analyze fabric defects.

2) Textile defects classification: For machine vision-based defect classification methods, there are some studies that use bayesian classifiers to classify fabric defects [32]. The gray level co-occurrence matrix method was used to extract the features of defects in the image, and then the k-nearest neighbor algorithm was used to classify the defects [33]. Support vector machine(SVM) was also often used in fabric defect classification [34]. There were also some studies that use neural networks to complete the task of classifying fabric defects in images or videos [35]. Although traditional image processing-based fabric detection and classification methods have achieved good results, most methods require manual feature extraction, which often consumes a lot of computing time. As a result, the processing process is not intelligent and robust. Until deep learning was applied to the field of image processing, the convolutional neural network(CNN) [36] has made remarkable progress in the field of image processing, and it has also been widely used in textile defect detection and classification. Jing *et al.* [37] improved that AlexNet extracted the characteristics of defective fabrics, and realized the classification of yarn-dyed fabric defects. In [38], authors studied the combination of compression sampling theorem and CNN in the case of few-shot and applies it to the classification of fabric texture defects, which achieves good results. Recently, Zhao *et al.* [39] inspired by human visual perception and memory mechanism proposed a CNN model based on visual long-term and short-term memory, which greatly improved the classification of fabric defects.

Similarly, machine vision-based product surface defect detection has applications in many industries, such as steel plates [40], glass [41], printing [42], parts [43], rail [44], fruit [45] and so on.

B. Intelligent Agricultural Production

Agricultural production is an important part of the global economy. As the global population continues to grow, urbanization will lead to a continuous reduction in the area of arable land and the number of farmers. The agricultural production system faces many challenges [46], so we must seek to some efficient intelligent and information agricultural technologies which save manpower and material resources to promote high-quality and high-yield agricultural development [47]. The application research of machine vision technology in the agricultural field started in the 1970s when most of the initial researches were on the feasibility of machine vision in agricultural applications and the development of image processing and analysis algorithm. With the rapid development of computer software and hardware, image acquisition and processing devices, and image processing technology, the application of machine vision technology in agriculture continuously expands. At present, some countries have begun to apply machine vision systems to various stages of agricultural production to solve the problem of increasing population aging and labor shortage [48]. This section mainly introduces two

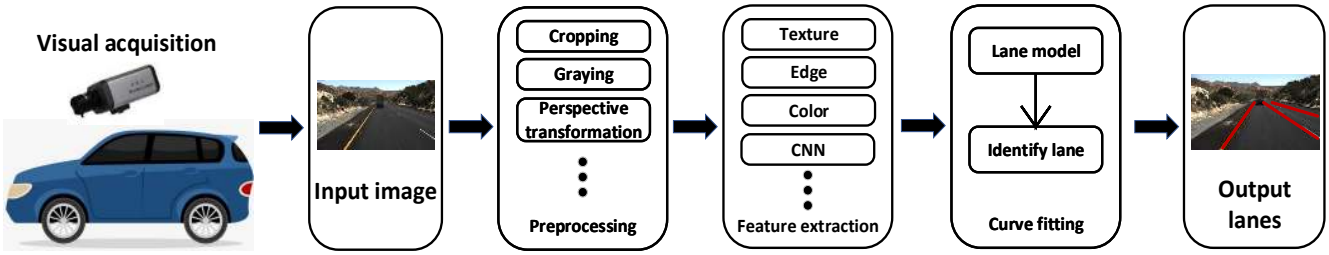

Fig. 4: General procedure of lane detection.

applications of machine vision technology in the field of agricultural production: agricultural robot and crop pest and disease monitoring.

1) Agricultural robot: Agricultural robot is an automated or semi-automated equipment to identify targets, collaborate, and identify color, texture, and odor characteristics [49], which can not only greatly increase labor productivity and reduce labor costs, but also reduce the damage of pesticides to the natural environment such as soil and water resources [50].

Research on agricultural robots began in Japan where picking robots began in the early 1980s. Kondo *et al.* [51] developed a cherry tomato picking robot, which used color cameras to collect images, through thresholding, filtering and other steps to segment fruits from the image background and identify the number of fruits, locating fruit three-dimensional information by stereoscopic vision. However, due to environmental influences, it was unable to complete obstacle-free picking and it was difficult to harvest short and hard fruits with inflorescences. Yaguchi *et al.* [52] used an electric wheeled omnidirectional chassis, a robotic arm, a binocular stereo camera, and a two-degree-of-freedom twisting actuator to form a tomato picking robot, which could complete picking operations in the shallow passage of the greenhouse under natural light.

In recent years, in order to realize the automatic recognition of cherries in the natural environment, Zhang *et al.* [53] has designed a method to implement robot vision using median filtering preprocessing, otsu algorithm threshold segmentation, and region threshold denoising by which the cherry recognition success rate was over 96 percet, and the picking efficiency was improved. In order to make the robot more efficient in picking mature apples, and has the ability to continuously recognize and operate at night, Ji *et al.* [54] proposed a Retinex algorithm based on pilot filtering to enhance nighttime images. The improvement of the vision-based control system has expanded the application scope of agricultural robots in greenhouses and orchards, etc. [55] [56], and has reduced workload and labor intensity.

2) Crop pest and disease monitoring: The control of crop diseases and insect pests and weeds is the key to achieving high quality, pollution-free and high yield in agricultural production. Traditional large-scale spraying of pesticides not only wastes resources, but also causes pollution and damage to the environment. The development of intelligent vision technology makes crop disease and pest diagnosis and weed identification faster, cheaper and non-destructive [46].

In early years, Pydipati *et al.* [57] proposed a method for identifying citrus diseased leaves and normal leaves. This method used color and texture to represent the features of the image, combined with the designed feature extraction and classification algorithm. Finally, the detection of citrus leaf disease was realized. Mayo *et al.* [58] described the features in moth images, used various classifiers and datasets for various experiments, and applied them to the automatic recognition of living moths.

In recent years, the literature [59] proposed the method of using region descriptors to simplify images containing aphids, and then used the histogram's directed gradient feature and support vector machine (SVM) to build models to realize the identification and population monitoring of aphids. Simple and easy to use, it could be used to investigate aphid infection in wheat field. Liu *et al.* [60] believed that traditional machine vision was limited by laboratories or pest traps in counting and identification, and has developed a vision-based multispectral detector for detecting 12 species on crops. Recently, research has proposed a method [61] for long-term pest behavior observation and integrated pest management. This work proposed a sensor network based on an integrated camera module and an embedded system that could simultaneously perform automatic detection and counting of sticky trap pests and other tasks, achieving integrated pest monitoring.

In this section, we introduced the advantages of visual perception in agricultural intelligence, and introduced agricultural robots, crop pest and disease monitoring in detail which are more representative in the application of visual perception technology. In addition to these two parts, the application of machine vision in agricultural intelligence also includes agricultural product quality inspection [62], crop healthy growth monitoring [63], agricultural vehicle visual navigation [64], and UAV farmland information monitoring [65], which are shown in Fig.3.

C. Intelligent Driving

Vision is the main source of information for humans in the course of various types of traffic [66]. Since the successful attempt of autonomous driving in the 1980s [67], people have paid attention to autonomous driving, At present, research on autonomous driving has attracted a large number of researchers and investors, and a large number of review papers have appeared [68] [69] [70]. Autonomous driving refers to the

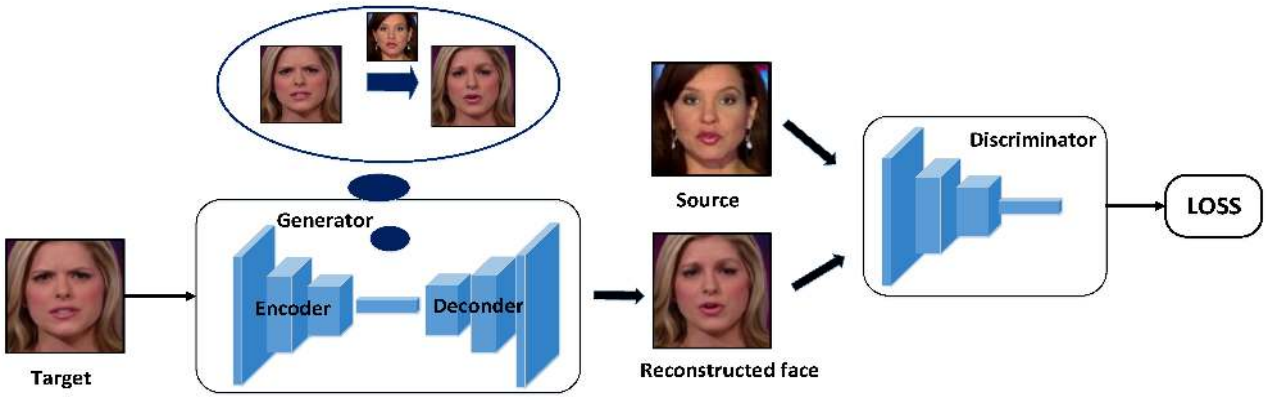

Fig. 5: Basic process of face synthesis based on GAN.

process of autonomously completing environmental perception and action execution for which the visual based environmental perception is an important source of information. Its main technologies are: detection of vehicles, pedestrians, and non-motorized vehicles on the road [71] [72] , traffic sign detection [73], lane detection [74], departure warning [75], drivable area detection [76],3D detection [77] [78], map 3D reconstruction [79], and object ranging [80], etc. Among them, lane detection is an important link to realize autonomous driving. In this section, we mainly introduce the application and development of visual perception technology in lane detection.

Most lane detection mainly have three steps: image preprocessing, feature extraction and parameter curve fitting [81], as shown in Fig.4.

1) Image preprocessing: The main purpose of image preprocessing is to enhance the image features and robustness, so that the detection can adapt to a variety of weather conditions, such as day, night, sunny, rain, etc. Usual preprocessing methods include color image gray processing [82], gradient enhancement Image 6507318, low-contrast image enhanced by histogram equalization [83], image binarization by edge detector [84], and cropped image by region of interest [85], etc.

2) Feature extraction: Feature extraction is a key step in detecting lanes. Liu *et al.* [86] obtaind the light intensity and width characteristics of the lane, and used the local threshold segmentation algorithm and morphological operations to accurately identify the lane. Others were different, Gopalan *et al.* [87] used the edge and texture features of the lane to achieve detection. Abramov *et al.* [88] proposed to use multi-source sensors to collect information, and used Graph simultaneous localization and mapping (SLAM) to fuse multi-source features obtained from various sensors, and finally through the fusion results to perceive multiple lanes in real time.

3) Parameter curve fitting: The detected lane points usually need to form lane lines in the graph by curve fitting. General algorithms usually use approximate clothoid curve models, such as quadratic curve, cubic curve, hyperbolic polynomial curve, parabola, B-spline, straight line [89], etc. Some algorithms do not use a fixed curve model, and are mainly used in scenarios without clear lanes, such as deserts. Broggi *et al.* [90] proposed an ant colony optimization method, which was a road boundary determination method based on reinforcement learning.

The detection algorithms mentioned above are mostly manual feature extraction at the feature extraction stage which usually has disadvantages such as low detection efficiency, poor robustness, and poor curve detection effect, etc. In recent years, lane detection methods using convolutional neural networks (CNN) have become popular, making lane detection easier to implement and highly accurate. Lee *et al.* [91] proposed an end-to-end multi-tasking network that utilized vanishing point information to simultaneously identify lane and road markings in extreme weather conditions, and solved rainy and low-light conditions for the first time. Pan *et al.* [92] proposed a new network structure Spatial CNN by converting the traditional convolutional layer-by-layer connection form into the form of slice-by-slice in a continuous volume which enabled information to be transmitted between rows and columns in a pixel, and enhances the ability of CNN to obtain semantic information of long continuous shape structures or large objects, such as lane lines, telephone poles, etc. In the detection phase, a network branch was added to enable the network to directly distinguish between different lanes and improve robustness. Recently, Hou *et al.* [93] introduced a method named Self Attention Distillation(SAD). The CNN model can learn by itself without labels and achieve substantial improvements with SAD. This method not only has a good detection effect, but also runs fast and has fewer model parameters.

D. Image Synthesis

Generative Adversarial Networks(GAN), proposed in 2014, is an emerging technology in the field of neural networks in recent years whose basic idea is derived from the zero-sum game. It regards the generation problem as the confrontation and game between the two networks, the generator and the discriminator. The former tries to produce data closer to the real and the latter tries to distinguish between real data and generated data more perfectly [94]. GAN can be applied to different types of signal processing. Visual image or video

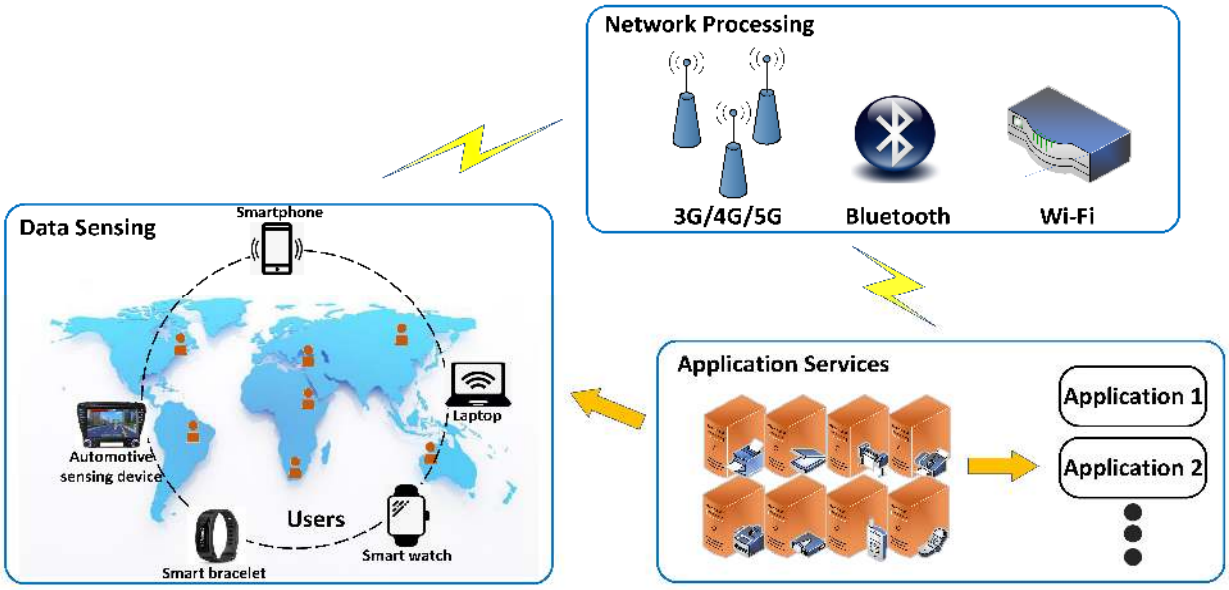

Fig. 6: A typical MCS architecture.

processing is one of its important application fields, such as image generation [95], video generation and prediction [96], object detection [97], image translation [98], image editing [99], image restoration [100], style migration [101], super-resolution reconstruction [102], etc.

Face synthesis has been a research hotspot in the field of computer vision, and has achieved many remarkable results [103] [104]. With the success of GANs in image and video generation, this end-to-end approach has attracted many researchers to start using GANs for face synthesis. Until now, GAN-based face synthesis technology has been impossible for human eyes to recognize as shown in Fig.5, and it is expected to be used in movies, animations, games, virtual reality, etc. But at the same time, it has caused a lot of public discussions about the dangers of this technology and many techniques for identifying computer-generated fake faces have also been developed [105]. In this section, we introduce a new application of visual perception: GAN-based face synthesis and corresponding fake face recognition techniques.

1) Face synthesis: In 2016, Isola *et al.* [106] proposed a method that changed the objective function of the Conditional Generative Adversarial Network(CGAN) [107], the network structure of the generator and the discriminator's discrimination, so that the network could learn the mapping relationship between the input and the output image and the loss function during training, providing a new framework for pixel-to-pixel conversion. On this basis, Wang *et al.* [108] used a multi-scale generator and discriminator to implement an interactive semantic editing for generating high-resolution images. Based on the principle of pix2pix [106], a face swapping software called Face2Face has aroused people's interest which could control the facial expressions and movements of people in TV or videos through cameras and face tracking software. Shen *et al.* [109] learned a symmetrical triad GAN to ease the training difficulty of the GAN, which could generate faces with multiple perspectives and expressions and retained the identity of this person. For video-to-video synthesis, Wang *et al.* [110] added optical flow constraints to the generator and discriminator, and designed a spatio-temporal objective function to focus on the inconsistency in front and back frames during video-to-video conversion applied to face swapping well. Recently, in order to solve the previous model's inability to work on few-shot during the training process and need to consume a lot of data resources. Wang *et al.* [111] introduced video-to-video synthesis under the condition of few-shot by adding attention mechanism in the network. In addition to the techniques discussed above, there were some studies that changed specific facial features, such as aging [112], makeup [113], complexion [114], etc.

2) Forgery detection: With the continuous upgrading of faking face technology, the public's voice for identifying fake is getting higher and higher, and more and more researchers have begun to study forgery detection methods. Two scientists from the Idiap Institute in Switzerland conducted a comprehensive evaluation of the effectiveness of face recognition methods in detecting DeepFake [115] which is also a popular face swapping software, and they found that general face recognition algorithms, such as FaceNet [116], identifying face generated by GAN was extremely poor. It was proposed that only image-based methods could effectively detect DeepFake videos. At present, many detection algorithms are proposed specifically for detecting forged images. In [117], the authors used CNN to complete this task that the performance was significantly improved compared to traditional detectors. Later, many CNN-based image forgery detection technologies are developed, such as [118] [119]. In [120], the authors evaluated the performance of related technologies in face forgery detection. Aiming at the problem that many people apply face fraud technology to national leaders, Agarwal *et al.* [105] studied that the facial expressions and movements of people when

they were speaking from which they used the correlation to identify real and fake faces, the probability of identifying fake videos reached 92 percent, and they said that the next study would be made on the rhythm and characteristics of people's speaking voices to further improve the accuracy of forgery detection.

E. Event Reconstruction

The development of the world's information industry has experienced two major trends: Computer and Internet. With the rapid development of mobile communication and perceptive technology, a large number of innovative applications and services have emerged, which has quickly brought us into the third information industry revolution — the Internet of Things (IoT) [121]. In the era of the IoT, people are increasingly using mobile smart terminals with cameras and various sensors, such as laptops, smartphones, GPS, smart bracelets, automotive sensing devices, smart watches, etc. A large amount of data obtained by using mobile terminals will be connected together through the network (Wi-Fi, 3G/4G/5G, Bluetooth, etc.) to form a group-aware network, which enables us to more comprehensively and large-scale perception of various physical objects and environmental conditions in the real world [122] [123]. They greatly expand the dimensions of human perception of the world, change the way people perceive the world, and open up a new field of mobile Internet — Mobile Crowd Sensing (MCS) [124], whose architecture is shown in Fig.6. At present, MCS has entered a stage of rapid and deep development and has penetrated deeply into all aspects of society, such as intelligent transportation [125], infrastructure and municipal management services [126], environmental monitoring and early warning [127], social relations and public safety services [128], etc.

In MCS, using the built-in camera of the mobile device to perceive is still an extremely important way. In this field, related research on vision-based MCS has also attracted a large number of researchers. In this regard, Guo *et al.* [129] puts forward the concept of visual crowdsensing(VCS), and summarizes the task models, characteristics, important technologies and applications of VCS in recent years. According to the summaryof VCS [129], its application scope can be divided into: floor plan generation [130], scene reconstruction [131], event reconstruction [132], indoor localization [133], indoor navigation [134], personal wellness and health [135], disaster relief [136], and city awareness [137]. In most cases, MCS is better than traditional visual perception methods that rely on fixed visual perception devices for monitoring. Event reconstruction is closest to people's daily life, so it has high research value. In this section, we discuss the development and significance of event reconstruction related technologies based on VCS.

With the popularity of wireless internet and smart phones by which people can record events in their lives in the form of pictures or videos and share them with others via the Internet, such as the popular short video platforms Vine, Instagram, Douyin, etc. Thousands of users record events in all corners of the world in this way, which not only broadens people's horizons, but also provides sufficient data for researchers in various fields [138]. Bao *et al.* [132] proposed a smartphone-based on-demand system MoVi, which used smartphones to cooperatively sense the surrounding environment, and performed video recording based on event trigger points (laughter, etc.). Videos recorded on different phones would be spliced into video highlights to provide users with key social information. Giridhar *et al.* [139] introduced an adaptive positioning algorithm that utilized image information in the social network Instagram to locate events that occured in cities in time and space. People in other cities could also experience the current event remotely from the sight of a witness. Bano *et al.* [138] proposed a framework that could match and cluster user-generated videos at the same time and space which automatically grouped and aligned videos captured by multiple user devices from different locations simultaneously, completing event reconstruction. Participants can review the entire event from different perspectives through information provided by others. Bohez *et al.* [140] introduced an integrated framework to mix users phones shooting perspectives with professional camera lenses and displayed during the event. The framework could transmit, process, and display hundreds of user videos in real time in an ultra-dense Wi-Fi environment.

Some studies focus on user feedback on videos on the network to evaluate video quality and classification, and then feed them back to users. Singhal *et al.* [141] analyzed the emotions of multiple users watching the same video through Electroencephalogram(EEG) signals, including sadness, happiness, and neutrality, and combined the video with various emotions. Then he adopted crowdsourcing mode [142] to summarize and evaluate the video, extracted the video summary, let users better understand it. Event reconstruction based on VCS will also have a certain impact on ecommerce. Recently, Diwanji *et al.* [143] discussed the users review videos of the product after purchasing. He found that this user-generated video greatly affected other consumers 'perceptions, attitudes and purchase intentions of the product, and provided an important management reference for online sellers.

F. Pose Measurement

Object pose measurement is also one of the important application directions of visual perception, refering to obtaining three position parameters and three attitude parameters of the target in a specific coordinate system, which can be the world coordinate system, object coordinate system or camera coordinate system. Object pose measurement has very important applications in the fields of robots [146], aerospace [145], industrial production [147], rotorcraft [148], vehicles [149], and ocean [150]. For example, in the space docking between a spacecraft and a target spacecraft, it is indispensable to accurately measure the relative position and attitude parameters between the spacecraft and the target spacecraft. The same is true in industrial production. Only by accurately measuring the pose of the accessory can the industrial robot grasp the object in a prescribed posture and align it for installation. Vision-based pose measurement has the characteristics of non-contact, high accuracy, good stability [162], which is of great

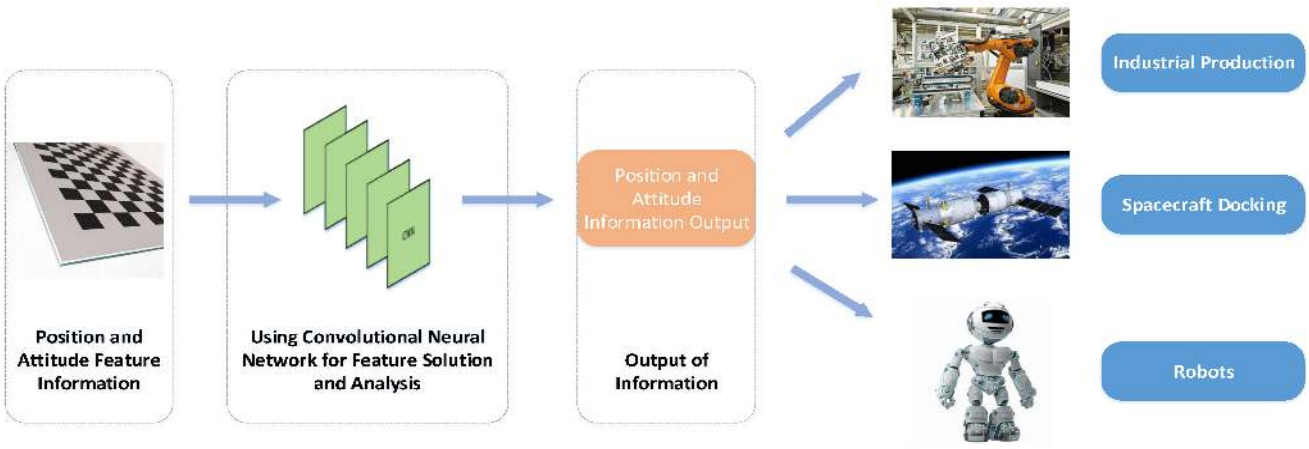

Fig. 7: CNN-based pose measurement and application.

significance for improving industrial production efficiency. Among them, the method based on monocular vision is the mainstream of pose measurement, and its biggest advantage is that the equipment is simple and easy to implement [151]. There is also a method based on binocular vision, which adds auxiliary depth information to the RGB image to help improve the measurement accuracy [152].

Most of the traditional pose measurement methods are based on geometric features. These methods have a certain dependence on the texture of the target surface and are susceptible to factors such as lighting, occlusion, and complex backgrounds. Later, the pose measurement mostly used feature descriptor-based methods to train classifiers by constructing distinguishing feature descriptors around the feature points of objects [153] [154]. Gee *et al.* [155] proposed a method for estimating the 6D pose of the camera using RGB-D information. This method extracted points of interest from the image based on sparse features, and described these points of interest with local descriptors then matching to the database. The sparse feature-based method and the traditional geometry-based method have some similarities, both of which are more difficult to recognize objects with less texture. Some studies used dense feature-based methods to predict the desired result with each pixel. Brachmann *et al.* [156] introduced a method for estimating the 6D pose of a specific target from a single frame of RGB-D images by using a new representation that combined dense 3D target coordinates and object class labels. This method could flexibly deal with textured or non-textured targets, and was robust under different lighting conditions. Later, the author improved on the basis of previous work [157]. He proposed a method to estimate the 6D pose using only a single RGB image by marginalizing the weight of the depth image and using only color to obtain the pose. There are also some studies that use template-based matching methods to scan pictures with a fixed template to find the best match. In the paper [158], the author sampled the object to be detected sufficiently by rendering in the possible SE3 space, extractd a sufficiently robust template, and then matched the template to estimate the pose.

In recent years, CNNs have also been widely used for vision-based object pose estimation in the field of industrial production, spacecraft docking and robots, etc. This process can be simply summarized as Fig. 7. Based on the literature [158], Wohlhart *et al.* [159]trained object types and object view templates together by CNN to learn descriptors representing object types and poses to detecte low textures. The method had a certain effect. Kehl *et al.* [160] used a 2D detector SSD to achieve 3D object detection and full 6D pose estimation only by RGB data and training the data of the synthetic model. For each 2D detection result, the most likely perspective and in-plane rotation were analyzed, and then a series of 6D hypotheses were established to select an optimal one as the result. Recently, Yang *et al.* [161] proposed a method of target pose measurement using CNN. This method directly returned the 6D attitude information of the object, eliminating the template used by the previous methods, which was simpler, faster speed and higher accuracy.

In order to more intuitively compare the applications of visual perception we have listed, we summarize the relevant fields and technologies of each application which are shown in Table 1.

III. The Serious Challenges Faced by Visual Perception

With the development of software and hardware technologies such as parallel computing, cloud computing, and machine learning, related technologies of visual perception have been greatly improved whose applications have also taken root in various fields. However, there are also many problems with current visual perception. The technology and application in many aspects are not mature enough, and even cannot be applied to actual production and life. In this section, we analyze the challenges faced by current visual perception.

A. Vision acquisition

Most of the existing methods of vision acquisition use various sensors to convert perceptual information into images

TABLE I: Summary of related fields and technologies of visual perception applications introduced in this survey

Application	Description	Related fields	Related technologies
Product Surface Defect Detection	The detection method based on machine vision can detect the surface defect areas that occur in the production process of the product. Compared with manual, it has the advantages of high sampling rate, high accuracy, strong real-time, high efficiency, and labor saving.	Textile [27], [29]–[31], [33]–[35], [37]–[39] Transportation [44] Food [45] Printing [42] Industrial manufacturing [40], [41], [43]	Object detection [29]–[31] Object classification [32]–[35] CNN-based object detection and classification [37]–[39]
Agricultural Production Intelligence	Machine vision perception can be applied to all aspects in agricultural production such as planting, monitoring, prevention, and picking, which is conducive to solving the problems of increasing population aging and lack of labor.	Agricultural robot [49]–[56] Disease and pest monitoring [46], [57]–[61] Agricultural product quality inspection [62] Crop healthy growth monitoring [63] Agricultural vehicle visual navigation [64] UAV farmland information monitoring [65]	Object segmentation [51] Binocular-based 3D information acquisition [51] Object recognition [53], [57] Image enhancement [53] Improved vision sensor [60], [61]
Intelligent Driving	Vision-based environmental perception is an important source of information for autonomous driving and provides strong support for the realization of autonomous driving.	Transportation [70] Military Agriculture [64]	Object segmentation [82]–[87], [144] Multi-source information fusion [88] SLAM [88] Reinforcement learning [90] CNN-based object detection [71]–[73], 3D object detection [77], [78], object segmentation [74], [76], [91]–[93] Deviation warning [75] Map 3D reconstruction [79] Object ranging [80]
Image Synthesis	Generative adversarial networks (GAN) have made great progress in generating images. It can learn its distribution based on target data and does not need to infer hidden variables during training.	Face synthesis and identification [105], [106], [108]–[111], [115]–[117], [120] Movies, animations, games, virtual reality.	GAN-based image generation, video generation and prediction [96], object detection [97], image translation [98], image editing [99], image restoration [100], style migration [101], image super-resolution [102] CNN-based image forgery detection [105], [118]–[120]
Event Reconstruction	The information collected by the multi-person visual device is integrated through the Mobile Internet, allowing people to perceive a variety of physical objects and environmental conditions in the real world more comprehensively and on a larger scale.	Social [139] Health [135] Smart City [131], [137] Architecture [130] Navigation [131], [133], [134] Rescue [136]	Image or video information fusion [132], [140] Event location [139] Event matching and clustering [138] Crowdsourcing analysis [141]
Object Pose Measurement	The three position parameters and three pose parameters of the object are obtained in a specific coordinate system. The advantages of vision-based pose measurement are: non-contact, high accuracy, and good stability.	Aerospace [145] Robot [146] Industrial production [147] Aircraft [148] Vehicle [149] Ocean [150]	Monocular-based pose measurement [151] Binocular-based pose measurement [152] Feature descriptors-based pose measurement [153]–[157] Template matching-based pose measurement [158] CNN-based pose measurement [159]–[161]

or videos. For example, the most common CCD and CMOS cameras are converted into electronic signals according to different light. The quality of vision acquisition and imaging technology directly affects the authenticity of information and is an important basis for visual information processing. Although the existing vision acquisition equipment and imaging technology have made significant progress, such as high dynamic range (HDR), global shutter, near infrared enhancement (NIR+), RGB-IR, power scalability and so on. However, under the influence of changes in real-world lighting and lens distortion, current vision acquisition and imaging technologies sometimes do not accurately reflect the real world. Backward vision acquisition equipment and imaging technology may become an obstacle to the development of visual perception technology.

B. Information Security

With the combination of artificial intelligence and visual perception technology, there are endless examples of perception is not true, so it is especially important to think about the security of visual information. For example, the GAN-based face synthesis technology mentioned earlier. Now some criminals use AI face swapping to pretend to be a national leader to make a bad speech and interfere with the presidential election. If the society cannot detect it in time, there may be serious consequences [105]. The security of visual perception is a key issue that researchers must attach great importance to during its rapid development.

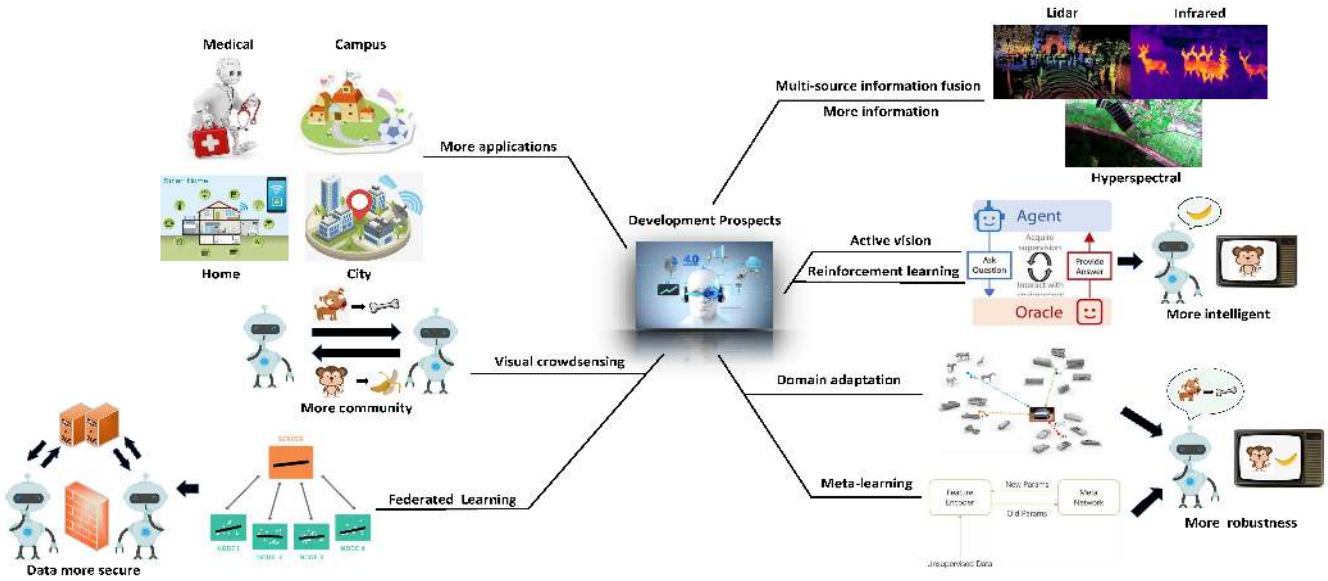

Fig. 8: Development Prospects of Visual Perception.

C. Speed, accuracy, and robustness

The trade-off between speed and accuracy has always been an important issue in the field of visual perception, especially in the field of computer vision [163]. Increasing the processing speed will inevitably reduce the information acquisition and analysis capabilities of deep networks, and vice versa. Its importance is self-evident. For example, in the field of automatic driving, the speed of detecting obstacles cannot achieve real-time or insufficient accuracy of recognition will hinder the realization of automatic driving [70]. And there is currently no machine vision technology that can achieve batch detection in the true sense while ensuring extremely high accuracy, minimal false detection rates, and eliminating missed detections. This goal cannot be achieved, reducing the application expectations of machine vision.

Due to the variability of the real world, the visual information collected by people is also diverse, and current visual perception and processing technologies in various fields often cannot adapt to such changing visual conditions, such as light intensity and shadows. The low robustness of the algorithm is also a universal problem in this field.

D. Construction in deep learning

CNN under deep learning is currently widely used in the processing of visual images or videos. Its theoretical problems are mainly reflected in statistics and computing. For any nonlinear function, a shallow network and a deep network can be found to represent it. The deep model has better performance for nonlinear functions than the shallow model. But the representability of deep networks does not represent learnability [164]. That is to say, deep learning is not intelligent enough, often accompanied by over-fitting and under-fitting problems [165], and requires the support of big data, but humans do not complete a large number of calculations to achieve related functions. Therefore, deep learning cannot be used as the main idea for the development of intelligent vision. Whether in terms of learning or implementation, the intelligence of visual perception is still a severe test.

E. Computing power and device volume

The success of computer vision depends not only on deep learning and large-scale data, but also on the computing carriers it implements, such as Central Processing Unit (CPU), Graphics Processing Unit (GPU), Application Specific Integrated Circuit (ASIC), Field Programmable Gate Array (FPGA) [166]. In the future, visual perception technology will also be inseparable from these computing units. Insufficient and slow computing power will also restrict the development of visual perception. The volume of integrated computing devices is also an important factor. At present, many companies are making such high-performance development boards for edge computing, such as Jetson TX2, Jetson AGX Xavier and so on. Small computing devices are of great significance to the practical application of the algorithm, but now there are still problems such as slow speed, insufficient computing power, and small memory.

F. The combination of software and hardware

The convergence of hardware and software has reached a turning point, and the two are no longer independent of each other, but are increasingly showing a mirror dependency. However, since software and hardware are two completely different fields, in the application of visual perception, many researchers have failed to implement the hardware well after proposing excellent visual perception algorithms, so the problem of combining software and hardware is also a challenge in this field.

IV. Development Prospects of Visual Perception

Vision is the most important source of information for humans to understand the world. The research on visual perception and processing will always accompany human scientific steps. In this section, we introduce its future development directions and trends based on the current challenges of visual perception, as shown in Fig.8.

A) Multi-source information fusion technology will become a hot research topic in the future. A single vision sensor has a specific range of use, and there are shortcomings such as less information and less accuracy. Different visual sensors have specific advantages. For example, ordinary visible light camera is good at acquiring color and shape information, lidar can obtain more depth information and point cloud information, infrared detectors can sense ambient temperature information, hyperspectral sensors can improve the ability to detect the attribute information of ground objects, etc. Multi-source information fusion technology has always been an effective method to maximize the amount of information. In the future, it will still be an important research direction. On the one hand, researchers can focus their research on sensors and hardware devices that can simultaneously acquire more visual information to improve ability to acquire visual information and compute big data. On the other hand, in terms of software, the fusion algorithm with high precision, low latency and less calculation will be further upgraded to achieve more reliable and accurate results for specific visual perception tasks.

B) Active vision and visual question answering is a hotspot in the field of computer vision and machine vision research today, and will be an important direction for solving current visual perception problems. Here the vision system can actively sense the environment, and according to certain rules, let the computer actively extract the required image features and answer questions about the picture. In active vision, multiple artificial intelligence methods may be integrated, such as reinforcement learning and other unsupervised, weakly supervised learning, which may help solve the current state of research that relies too much on mathematical modeling and mathematical calculations to meet the requirements of system speed and intelligence.

C) Visual perception will develop towards higher adaptability and robustness in the face of different tasks, which may include domain adaptation and meta-learning. Domain adaptation is a sub-discipline of machine learning that deals with the use of models trained on information source distributions in the context of different target distributions. According to the amount of training data required for a new specific computer vision task, the performance of the function of deep domain adaptation is closer to human intelligence. Progress in this field is critical to the entire field of computer vision, and deep-domain adaptation can ultimately lead people to reuse effective and simple knowledge in vision tasks. Similarly, meta-learning is intended to allow machines to learn to learn. When the machine has the ability to learn, it can quickly adapt to different tasks. Meta-learning is also an important direction for improving the robustness of future visual perception.

D) Visual crowdsensing is a technological idea that conforms to the trend of world development. As humans enter the age of the Internet of Things, valuable data is gradually being socialized, shared, and experiential. In VCS, pictures and videos can contain richer information, and they are more closely related to the environment and others. The volume of data items is larger, and conform to the development idea of IoT, it may become a mainstream technology in visual perception. Similarly, Federated Machine Learning [167] is an emerging artificial intelligence basic technology, which is proposed in order to solve the problem of data islands and strengthen data security. In recent years, research on federal learning has continuously emerged, and will lead the next wave of commercialization of machine learning technology. Federal learning is also a new road for the development of visual perception under the tide of the IoT.

E) The global Internet and semiconductor giants have laid out, showing that intelligent image and video processing will be the next arena, which may mean that vision technology is ushering in a golden period of development. In the future, visual perception will continue to make breakthroughs in applications such as unmanned aerial vehicle (UAV), autonomous driving, smart doctors, smart security, and smart cities, etc. Exploring new technical support and application areas is always the trend of visual perception development.

V. Conclusion

Overall, in this paper we have reviewed and analyzed several major application fields of visual perception, including industrial quality inspection, agricultural production, autonomous driving, visual fraud and crowd sensing. Specifically, we introduced textile defect detection in product surface inspection, agricultural robots, agricultural pest and disease monitoring in intelligent agricultural production, lane detection in autonomous driving, image synthesis and forgery detection in visual fraud and event reconstruction in crowd sensing and object measurement. These applications basically cover the popular visual perception research directions in recent years, including classification in image or video, segmentation, object detection, tracking, image or video generation, forgery detection, 3D reconstruction and multi-source information fusion. We can conclude that most of the current visual perception technologies and applications are combined with artificial intelligence, which is helpful to human production and life, and has the advantages of low cost, high precision and high efficiency.

In addition, based on the status quo, we analyze the current challenges faced by humans when using visual perception technology, including vision acquisition, computing power, device volume, technology security, speed, accuracy, robustness, and intelligence, software and hardware combination, etc. Based on these challenges, we have made predictions about the development prospects of visual perception. In the future,

visual perception will be more closely integrated with artificial intelligence, and will move towards multi-source information fusion, active vision, domain adaptation, meta-learning, reinforcement learning, federal learning, crowd sensing and other directions, and more fields will be applied to visual perception technology. With the continuous development and intelligentization of visual perception technology, human production efficiency and quality will continue to improve, which will be one of the important driving forces for human social progress.

REFERENCES

[1] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97 – 136, 1980.

[2] A. O. Fernandes, L. F. E. Moreira, and J. M. Mata, "Machine vision applications and development aspects," *2011 9th IEEE International Conference on Control and Automation (ICCA)*, pp. 1274–1278, 2011.

[3] R. Alfredo Osornio-Rios, J. A. Antonino-Daviu, and R. de Jesus Romero-Troncoso, "Recent industrial applications of infrared thermography: A review," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 615–625, 2019.

[4] C. Ding, X. Qiu, F. Xu, X. Liang, Z. Jiao, and F. Zhang, "Synthetic aperture radar three-dimensional imaging – from tomosar and array insar to microwave vision," *Journal of Radars*, vol. 8, p. 1, 2019.

[5] T. C. Lei and L. X. Dong, "Ultrasonic phased array length measurement of internal detects in butt weld," *2016 23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pp. 1–7, 2016.

[6] B. Prasad, K. Prabha, and P. Kumar, "Condition monitoring of turning process using infrared thermography technique – an experimental approach," *Infrared Physics and Technology*, vol. 81, pp. 137–147, 2017.

[7] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zuhlke, "Human-machine-interaction in the industry 4.0 era," *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, pp. 289–294, 2014.

[8] L. Fu, Y. Zhang, Q. Huang, and X. Chen, "Research and application of machine vision in intelligent manufacturing," *2016 Chinese Control and Decision Conference (CCDC)*, pp. 1126–1131, 2016.

[9] A. Clark, J. Donahue, and K. Simonyan, "Adversarial video generation on complex datasets," *arXiv preprint arXiv:1907.06571*, 2019.

[10] A. Mohan and S. Poobal, "Crack detection using image processing: A critical review and analysis," *Alexandria Engineering Journal*, vol. 57, no. 2, pp. 787 – 798, 2018.

[11] J. Yang, K. Sim, X. Gao, W. Lu, Q. Meng, and B. Li, "A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1314–1328, 2019.

[12] B. Jiang, J. Yang, Q. Meng, B. Li, and W. Lu, "A deep evaluator for image retargeting quality by geometrical and contextual interaction," *IEEE Transactions on Cybernetics*, vol. 50, no. 1, pp. 87–99, 2020.

[13] B. Wandt and B. Rosenhahn, "Repnet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[14] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *arXiv preprint arXiv:1905.05055*, 2019.

[15] B. Xu, S. Zhao, X. Sui, and C. Hua, "High-speed stereo matching algorithm for ultra-high resolution binocular image," *2018 IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pp. 87–90, 2018.

[16] Z. Tang, R. Cunha, T. Hamel, and C. Silvestre, "Aircraft landing using dynamic two-dimensional image-based guidance control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2104–2117, 2019.

[17] X. Xiaozhu and H. Cheng, "Object detection of armored vehicles based on deep learning in battlefield environment," *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 1568–1570, 2017.

[18] C. Xie, M. Li, H. Wang, and J. Dong, "A survey on visual analysis of ocean data," *Visual Informatics*, vol. 3, no. 3, pp. 113 – 128, 2019.

[19] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[20] W. Fang, L. Ding, P. E. Love, H. Luo, H. Li, F. Pena-Mora, B. Zhong, and C. Zhou, "Computer vision applications in construction safety assurance," *Automation in Construction*, vol. 110, p. 103013, 2020.

[21] D. I. Patricio and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Computers and Electronics in Agriculture*, vol. 153, pp. 69 – 81, 2018.

[22] W. Shi, M. B. Alawieh, X. Li, and H. Yu, "Algorithm and hardware implementation for visual perception system in autonomous vehicle: A survey," *Integration, the VLSI Journal*, vol. 59, 2017.

[23] I. Witus, C. On, R. Alfred, A. A. Ag Ibrahim, T. G. Tan, and P. Anthony, "A review of computer vision methods for fruit recognition," *Advanced Science Letters*, vol. 24, pp. 1538–1542, 02 2018.

[24] C. S. Pereira, R. Morais, and M. J. C. S. Reis, "Recent advances in image processing techniques for automated harvesting purposes: A review," *2017 Intelligent Systems Conference (IntelliSys)*, pp. 566–575, 2017.

[25] P. Ranky, "Advanced machine vision systems and application examples," *Sensor Review - SENS REV*, vol. 23, pp. 242–245, 2003.

[26] B. Tang, J. Kong, and S. Wu, "Review of surface defect detection based on machine vision," *Journal of Image and Graphics*, vol. 22, no. 12, pp. 1640–1663, 2017.

[27] A. Kumar, "Computer-vision-based fabric defect detection: A survey," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 1, pp. 348–363, 2008.

[28] H. Y. Ngan, G. K. Pang, and N. H. Yung, "Automated fabric defect detection–a review," *Image and Vision Computing*, vol. 29, no. 7, pp. 442 – 458, 2011.

[29] ——, "Motif-based defect detection for patterned fabric," *Pattern Recognition*, vol. 41, no. 6, pp. 1878 – 1894, 2008.

[30] J. K. Chandra, P. K. Banerjee, and A. K. Datta, "Neural network trained morphological processing for the detection of defects in woven fabric," *The Journal of The Textile Institute*, vol. 101, no. 8, pp. 699–706, 2010.

[31] Chi-Ho Chan and G. K. H. Pang, "Fabric defect detection by fourier analysis," *IEEE Transactions on Industry Applications*, vol. 36, no. 5, pp. 1267–1276, 2000.

[32] M. T. Habib, S. B. Shuvo, M. S. Uddin, and F. Ahmed, "Automated textile defect classification by bayesian classifier based on statistical features," *2016 International Workshop on Computational Intelligence (IWCI)*, pp. 101–105, 2016.

[33] K. Yildiz, A. Buldu, and M. Demetgul, "A thermal-based defect classification method in textile fabrics with k-nearest neighbor algorithm," *Journal of Industrial Textiles*, vol. 45, no. 5, pp. 780–795, 2016.

[34] W. Li and L. Cheng, "Yarn-dyed woven defect characterization and classification using combined features and support vector machine," *The Journal of The Textile Institute*, vol. 105, no. 2, pp. 163–174, 2014.

[35] H. Celik, L. Dulger, and M. Topalbekiroglu, "Development of a machine vision system: real-time fabric defect detection and classification with neural networks," *The Journal of The Textile Institute*, vol. 105, no. 6, pp. 575–585, 2014.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems 25*, pp. 1097–1105, 2012.

[37] J. Jing, A. Dong, P. Li, and K. Zhang, "Yarn-dyed fabric defect classification based on convolutional neural network," *Optical Engineering*, vol. 56, no. 9, pp. 1–9, 2017.

[38] B. Wei, K. Hao, X. song Tang, and Y. Ding, "A new method using the convolutional neural network with compressive sensing for fabric defect classification based on small sample sizes," *Textile Research Journal*, vol. 89, no. 17, pp. 3539–3555, 2019.

[39] Y. Zhao, K. Hao, H. He, X. Tang, and B. Wei, "A visual long-short-term memory based integrated cnn model for fabric defect image classification," *Neurocomputing*, 2019.

[40] D. K. H. Singhka, N. Neogi, and D. Mohanta, "Surface defect classification of steel strip based on machine vision," *International Conference on Computing and Communication Technologies*, pp. 1–5, 2014.

[41] J. George, S. Janardhana, J. Jaya, and K. J. Sabareesaan, "Automatic defect detection inspectacles and glass bottles based on fuzzy c means clustering," *2013 International Conference on Current Trends in Engineering and Technology (ICCTET)*, pp. 8–12, 2013.

[42] H. Kalviainen, "Machine vision based quality control from pulping to papermaking for printing," *Pattern Recognition and Image Analysis*, vol. 21, pp. 486–490, 2011.

[43] P. Kunakornvong and P. Sooraksa, "Machine vision for defect detection on the air bearing surface," *2016 International Symposium on Computer, Consumer and Control (IS3C)*, pp. 37–40, 2016.

[44] Z. Liu, W. Wang, and P. Wang, "Design of machine vision system for inspection of rail surface defects," *JOURNAL OF ELECTRONIC MEASUREMENT AND INSTRUMENT*, vol. 24, pp. 1012–1017, 2010.

[45] D. Rong, X. Rao, and Y. Ying, "Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm," *Computers and Electronics in Agriculture*, vol. 137, pp. 59 – 68, 2017.

[46] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation–a review," *Information Processing in Agriculture*, 2019.

[47] Seema, A. Kumar, and G. S. Gill, "Automatic fruit grading and classification system using computer vision: A review," 2015, pp. 598–603.

[48] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1 – 12, 2019.

[49] K. Tanigaki, T. Fujiura, A. Akase, and J. Imagawa, "Cherry-harvesting robot," *Computers and Electronics in Agriculture*, vol. 63, no. 1, pp. 65 – 72, 2008, special issue on bio-robotics.

[50] R. Shamshiri, C. Weltzien, I. Hameed, I. Yule, T. Grift, and e. a. Balasundram, S.K., "Research and development in agricultural robotics: A perspective of digital farming," *Int J Agric and Biol Eng*, vol. 11, no. 4, pp. 1–14, 2018.

[51] N. Kondo, M. Monta, and T. Fujiura, "Fruit harvesting robots in japan," *Advances in Space Research*, vol. 18, no. 1, pp. 181 – 184, 1996, physical, Chemical, Biochemical and Biological Techniques and Processes.

[52] H. Yaguchi, K. Nagahama, T. Hasegawa, and M. Inaba, "Development of an autonomous tomato harvesting robot with rotational plucking gripper," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 652–657, 2016.

[53] Q. Zhang, S. Chen, T. Yu, and Y. Wang, "Cherry recognition in natural environment based on the vision of picking robot," *IOP Conference Series: Earth and Environmental Science*, vol. 61, p. 012021, 2017.

[54] J. Wei, Q. Zhijie, X. Bo, and Z. Dean, "A nighttime image enhancement method based on retinex and guided filter for object recognition of apple harvesting robot," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, p. 1729881417753871, 2018.

[55] Y. Zhao, L. Gong, Y. Huang, and C. Liu, "A review of key techniques of vision-based control for harvesting robot," *Computers and Electronics in Agriculture*, vol. 127, pp. 311 – 323, 2016.

[56] S. Mehta, W. MacKunis, and T. Burks, "Robust visual servo control in the presence of fruit motion for robotic citrus harvesting," *Computers and Electronics in Agriculture*, vol. 123, pp. 362 – 375, 2016.

[57] R. Pydipati, T. Burks, and W. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Computers and Electronics in Agriculture*, vol. 52, no. 1, pp. 49 – 59, 2006.

[58] M. Mayo and A. T. Watson, "Automatic species identification of live moths," *Knowledge-Based Systems*, vol. 20, no. 2, pp. 195 – 202, 2007, aI 2006.

[59] T. Liu, W. Chen, W. Wu, C. Sun, W. Guo, and X. Zhu, "Detection of aphids in wheat fields using a computer vision technique," *Biosystems Engineering*, vol. 141, pp. 82 – 93, 2016.

[60] H. Liu and J. S. Chahl, "A multispectral machine vision system for invertebrate detection on green leaves," *Computers and Electronics in Agriculture*, vol. 150, pp. 279 – 288, 2018.

[61] "Application of an image and environmental sensor network for automated greenhouse insect pest monitoring," *Journal of Asia-Pacific Entomology*, vol. 23, no. 1, pp. 17 – 28, 2020.

[62] B. Zhang, W. Huang, J. Li, C. Zhao, S. Fan, J. Wu, and C. Liu, "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," *Food Research International*, vol. 62, pp. 326 – 343, 2014.

[63] M. Rico-Fernandez, R. Rios-Cabrera, M. Castelan, H.-I. Guerrero-Reyes, and A. Juarez-Maldonado, "A contextualized approach for segmentation of foliage in different crop species," *Computers and Electronics in Agriculture*, vol. 156, pp. 378 – 386, 2019.

[64] M. H. Jones, J. Bell, D. Dredge, M. Seabright, A. Scarfe, M. Duke, and B. MacDonald, "Design and testing of a heavy-duty platform for autonomous navigation in kiwifruit orchards," *Biosystems Engineering*, vol. 187, pp. 129 – 146, 2019.

[65] Y. Niu, L. Zhang, H. Zhang, W. Han, and X. Peng, "Estimating above-ground biomass of maize using features derived from uav-based rgb imagery," *Remote Sensing*, vol. 11, no. 11, 2019.

[66] B. Ranft and C. Stiller, "The role of machine vision for intelligent vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 8–19, 2016.

[67] J. Janai, F. Guney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *arXiv preprint arXiv:1704.05519*, 2017.

[68] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," *IATSS Research*, 2019.

[69] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3d object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3782–3795, 2019.

[70] W. Shi, M. B. Alawieh, X. Li, and H. Yu, "Algorithm and hardware implementation for visual perception system in autonomous vehicle: A survey," *Integration*, vol. 59, pp. 148 – 156, 2017.

[71] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems 28*, pp. 91–99, 2015.

[72] P. Yu, Y. Zhao, J. Zhang, and X. Xie, "Pedestrian detection using multi-channel visual feature fusion by learning deep quality model," *Journal of Visual Communication and Image Representation*, vol. 63, p. 102579, 2019.

[73] T. Yang, X. Long, A. K. Sangaiah, Z. Zheng, and C. Tong, "Deep detection network for real-life traffic sign in vehicular networks," *Computer Networks*, vol. 136, pp. 95 – 104, 2018.

[74] J. Liu, "Learning full-reference quality-guided discriminative gradient cues for lane detection based on neural networks," *Journal of Visual Communication and Image Representation*, vol. 65, p. 102675, 2019.

[75] I. Gamal, A. Badawy, A. M. Al-Habal, M. E. Adawy, K. K. Khalil, M. A. El-Moursy, and A. Khattab, "A robust, real-time and calibration-free lane departure warning system," *Microprocessors and Microsystems*, vol. 71, p. 102874, 2019.

[76] Z. Liu, S. Yu, and N. Zheng, "A co-point mapping-based approach to drivable area detection for self-driving cars," *Engineering*, vol. 4, no. 4, pp. 479 – 490, 2018.

[77] P. Li, X. Chen, and S. Shen, "Stereo r-cnn based 3d object detection for autonomous driving," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[78] L. Wang, X. Fan, J. Chen, J. Cheng, J. Tan, and X. Ma, "3d object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustainable Cities and Society*, p. 102002, 2019.

[79] C. Hane, L. Heng, G. H. Lee, F. Fraundorfer, P. Furgale, T. Sattler, and M. Pollefeys, "3d visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection," *Image and Vision Computing*, vol. 68, pp. 14 – 27, 2017, automotive Vision: Challenges, Trends, Technologies and Systems for Vision-Based Intelligent Vehicles.

[80] X. Sun, Y. Jiang, Y. Ji, W. Fu, S. Yan, Q. Chen, B. Yu, and X. Gan, "Distance measurement system based on binocular stereo vision," *IOP Conference Series: Earth and Environmental Science*, vol. 252, no. 5, p. 052051, 2019.

[81] H. Zhu, K. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2584–2601, 2017.

[82] Z. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 16–26, 2008.

[83] U. Meis, W. Klein, and C. Wiedemann, "A new method for robust far-distance road course estimation in advanced driver assistance systems,"

13th International IEEE Conference on Intelligent Transportation Systems, pp. 1357–1362, 2010.

[84] J. Wang, Y. Wu, Z. Liang, and Y. Xi, "Lane detection based on random hough transform on region of interesting," *The 2010 IEEE International Conference on Information and Automation*, pp. 1735–1740, 2010.

[85] J. Son, H. Yoo, S. Kim, and K. Sohn, "Real-time illumination invariant lane detection for lane departure warning system," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816 – 1824, 2015.

[86] G. Liu, S. Li, and W. Liu, "Lane detection algorithm based on local feature extraction," *2013 Chinese Automation Congress*, pp. 59–64, 2013.

[87] R. Gopalan, T. Hong, M. Shneier, and R. Chellappa, "A learning approach towards detection and tracking of lane markings," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1088–1098, 2012.

[88] A. Abramov, C. Bayer, C. Heller, and C. Loy, "Multi-lane perception using feature fusion based on graphslam," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3108–3115, 2016.

[89] S. Jung, J. Youn, and S. Sull, "Efficient lane detection based on spatiotemporal images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 289–295, 2016.

[90] A. Broggi and S. Cattani, "An agent based evolutionary approach to path detection for off-road vehicle guidance," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1164 – 1173, 2006, evolutionary Computer Vision and Image Understanding.

[91] S. Lee, J. Kim, J. Shin Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. Seok Hong, S.-H. Han, and I. So Kweon, "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition," *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[92] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[93] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," 2019.

[94] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems 27*, pp. 2672–2680, 2014.

[95] E. L. Denton, S. Chintala, a. szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," *Advances in Neural Information Processing Systems 28*, pp. 1486–1494, 2015.

[96] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[97] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[98] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[99] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2019.

[100] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," *The European Conference on Computer Vision (ECCV)*, 2018.

[101] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," *Computer Vision–ECCV 2016*, pp. 702–716, 2016.

[102] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

[103] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, "Face2face: Real-time face capture and reenactment of rgb videos," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[104] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. NieBssner, "Facevr: Real-time gaze-aware facial reenactment in virtual reality," *ACM Trans. Graph.*, vol. 37, no. 2, pp. 25:1–25:15, Jun. 2018.

[105] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

[106] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016.

[107] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[108] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[109] Y. Shen, P. Luo, J. Yan, X. Wang, and X. Tang, "Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[110] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," *arXiv preprint arXiv:1808.06601*, 2018.

[111] T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, B. Catanzaro, and J. Kautz, "Few-shot video-to-video synthesis," *Advances in Neural Information Processing Systems 32*, pp. 5014–5025, 2019.

[112] G. Antipov, M. Baccouche, and J. Dugelay, "Face aging with conditional generative adversarial networks," *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2089–2093, 2017.

[113] Y. Li, L. Song, X. Wu, R. He, and T. Tan, "Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification," *Pattern Recognition*, vol. 90, pp. 99–108, 2019.

[114] Y. Lu, Y.-W. Tai, and C.-K. Tang, "Conditional cyclegan for attribute guided face image generation," *European Conference on Computer Vision*, 2014.

[115] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," *arXiv preprint arXiv:1812.08685*, 2018.

[116] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[117] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection," *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, pp. 159–164, 2017.

[118] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, 2017.

[119] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018.

[120] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. NieBner, "Faceforensics++: Learning to detect manipulated facial images," *arXiv preprint arXiv:1901.08971*, 2019.

[121] H.-D. Ma, "Internet of things: Objectives and scientific challenges," *Journal of Computer Science and Technology*, vol. 26, no. 6, pp. 919–924, 2011.

[122] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, 2014.

[123] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," pp. 593–598, 2014.

[124] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.

[125] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, and K. Zhou, "Mobile crowd sensing for traffic prediction in internet of vehicles," *Sensors*, vol. 16, no. 1, 2016.

[126] Y. Wang, Q. Chen, L. Liu, X. Li, A. K. Sangaiah, and K. Li, "Systematic comparison of power line classification methods from als and mls point cloud data," *Remote Sensing*, vol. 10, no. 8, 2018.

[127] A. Antonic, V. Bilas, M. Marjanovic, M. Matijasevic, D. Oletic, M. Pavelic, I. P. Zarko, K. Pripuzic, and L. Skorin-Kapov, "Urban crowd sensing demonstrator: Sense the zagreb air," *2014 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 423–424, 2014.

[128] T. Ludwig, C. Reuter, T. Siebigteroth, and V. Pipek, "Crowdmonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4083–4092, 2015.

[129] B. Guo, Q. Han, H. Chen, L. Shangguan, Z. Zhou, and Z. Yu, "The emergence of visual crowdsensing: Challenges and opportunities," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2526–2543, 2017.
[130] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, pp. 249–260, 2014.
[131] Z. Peng, S. Gao, B. Xiao, S. Guo, and Y. Yang, "Crowdgis: Updating digital maps via mobile crowdsensing," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 369–380, 2018.
[132] X. Bao and R. Roy Choudhury, "Movi: Mobile phone based video highlights via collaborative sensing," *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, pp. 357–370, 2010.
[133] H. Xu, Z. Yang, Z. Zhou, L. Shangguan, K. Yi, and Y. Liu, "Enhancing wifi-based localization with visual clues," *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 963–974, 2015.
[134] E. Dong, J. Xu, C. Wu, Y. Liu, and Z. Yang, "Pair-navi: Peer-to-peer indoor navigation with mobile visual slam," *IEEE INFOCOM 2019 – IEEE Conference on Computer Communications*, pp. 1189–1197, 2019.
[135] J. Lee, A. Banerjee, and S. K. S. Gupta, "Mt-diet: Automated smartphone based diet assessment with infrared images," *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–6, 2016.
[136] T. Dao, A. K. Roy-Chowdhury, H. V. Madhyastha, S. V. Krishnamurthy, and T. La Porta, "Managing redundant content in bandwidth constrained wireless networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 2, pp. 988–1003, 2017.
[137] Y. Li, F. Xue, X. Fan, Z. Qu, and G. Zhou, "Pedestrian walking safety system based on smartphone built-in sensors," *IET Communications*, vol. 12, no. 6, pp. 751–758, 2018.
[138] S. Bano and A. Cavallaro, "Discovery and organization of multi-camera user-generated videos of the same event," *Information Sciences*, vol. 302, pp. 108 – 121, 2015.
[139] P. Giridhar, S. Wang, T. Abdelzaher, R. Ganti, L. Kaplan, and J. George, "On localizing urban events with instagram," *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, 2017.
[140] S. Bohez, , G. Daneels, L. Van Herzeele, , N. Van Kets, S. Decrock, M. D. Geyter, G. Van Wallendael, P. Lambert, B. Dhoedt, P. Simoens, S. Latre, and J. Famaey, "The crowd as a cameraman: on-stage display of crowdsourced mobile video at large-scale events," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 597–629, 2018.
[141] A. Singhal, P. Kumar, R. Saini, P. P. Roy, D. P. Dogra, and B.-G. Kim, "Summarization of videos by analyzing affective state of the user through crowdsource," *Cognitive Systems Research*, vol. 52, pp. 917 – 930, 2018.
[142] D. C. Brabham, "Crowdsourcing as a model for problem solving: An introduction and cases," *Convergence*, vol. 14, no. 1, pp. 75–90, 2008.
[143] V. S. Diwanji and J. Cortese, "Contrasting user generated videos versus brand generated videos in ecommerce," *Journal of Retailing and Consumer Services*, vol. 54, p. 102024, 2020.
[144] H. Yoo, U. Yang, and K. Sohn, "Gradient-enhancing conversion for illumination-robust lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1083–1094, 2013.
[145] H. Li and H. Duan, "Verification of monocular and binocular pose estimation algorithms in vision-based uavs autonomous aerial refueling system," *Science China Technological Sciences*, vol. 59, no. 11, pp. 1730–1738, 2016.
[146] T. Lemaire, C. Berger, I.-K. Jung, and S. Lacroix, "Vision-based slam: Stereo and monocular approaches," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 343–364, 2007.
[147] D. WU and F. DU, "A multi-constraints based pose coordination model for large volume components assembly," *Chinese Journal of Aeronautics*, 2019.
[148] T. P. Nascimento and M. Saska, "Position and attitude control of multi-rotor aerial vehicles: A survey," *Annual Reviews in Control*, vol. 48, pp. 129 – 146, 2019.
[149] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and A. Broggi, "Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2364–2375, 2006.
[150] G. Wang, W. Wang, and C. Wang, "The method and error analysis of deep-sea pose measurement system," *Measurement*, vol. 98, pp. 276 – 282, 2017.
[151] Z. He, Z. Jiang, X. Zhao, S. Zhang, and C. Wu, "Sparse template-based 6-d pose estimation of metal parts using a monocular camera," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 390–401, 2020.
[152] J. Li, "Relative pose measurement of moving rigid bodies based on binocular vision," *Optik*, vol. 180, pp. 159 – 165, 2019.
[153] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer Vision – ECCV 2006*, pp. 404–417, 2006.
[154] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1465–1479, 2006.
[155] A. P. Gee and W. Mayol-cuevas, "6d relocalisation for rgbd cameras using synthetic view regression," *BMVC*, 2012.
[156] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother, "Learning 6d object pose estimation using 3d object coordinates," *Computer Vision – ECCV 2014*, pp. 536–551, 2014.
[157] E. Brachmann, F. Michel, A. Krull, M. Ying Yang, S. Gumhold, and c. Rother, "Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[158] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes," *Computer Vision – ACCV 2012*, pp. 548–562, 2013.
[159] P. Wohlhart and V. Lepetit, "Learning descriptors for object recognition and 3d pose estimation," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[160] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again," *The IEEE International Conference on Computer Vision (ICCV)*, 2017.
[161] J. Yang, J. Man, M. Xi, X. Gao, W. Lu, and Q. Meng, "Precise measurement of position and attitude based on convolutional neural network and visual correspondence relationship," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2019.
[162] P. Ferrara, A. Piva, F. Argenti, J. Kusuno, M. Niccolini, M. Ragaglia, and F. Uccheddu, "Wide-angle and long-range real time pose estimation: A comparison between monocular and stereo vision systems," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 159 – 168, 2017.
[163] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[164] G. Marcus, "Deep learning: A critical appraisal," *arXiv preprint arXiv:1801.00631*, 2018.
[165] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[166] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, "Computer vision algorithms and hardware implementations: A survey," *Integration*, vol. 69, pp. 309 – 320, 2019.
[167] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, 2019.

Jiachen Yang (Member, IEEE) received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2005 and 2009, respectively. He is currently a professor at Tianjin University. He was also a visiting scholar with the Department of Computer Science, School of Science, Loughborough University, U.K and the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, U.S. His research interests include image quality evaluation stereo vision research, pattern recognition and virtual reality.

Chenguang Wang received the B.S degree in communication and information engineering from Yanshan University, Hebei, China, in 2018. He is currently pursuing the M.S. degree at school of information and communication engineering, Tianjin University, Tianjin, China. His research interests include object detection, computer vision and pattern recognition.

Bin Jiang received the B.S. and M.S. degree in communication and information engineering from Tianjin University, Tianjin, China, in 2013 and 2016. He is currently pursuing the Ph.D. degree at the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. He is also a visiting scholar in Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, US, where he is a member of Security and Optimization for Networked Globe Laboratory. His research interests lie in multimedia processing and cyber-physical systems.

Houbing Song (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012, and the M.S. degree in civil engineering from the University of Texas, El Paso, TX, in December 2006.

In August 2017, he joined the Department of Electrical Engineering & Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, FL, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). He served on the faculty of West Virginia University from August 2012 to August 2017. In 2007 he was an Engineering Research Associate with the Texas A&M Transportation Institute. He has served as an Associate Technical Editor for IEEE Communications Magazine (2017-present), an Associate Editor for IEEE Internet of Things Journal (2020-present) and a Guest Editor for IEEE Journal on Selected Areas in Communications (J-SAC), IEEE Internet of Things Journal, IEEE Transactions on Industrial Informatics, IEEE Sensors Journal, IEEE Transactions on Intelligent Transportation Systems, and IEEE Network. He is the editor of six books, including Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things, Elsevier, 2019, Smart Cities: Foundations, Principles and Applications, Hoboken, NJ: Wiley, 2017, Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications, Chichester, UK: Wiley-IEEE Press, 2017, Cyber-Physical Systems: Foundations, Principles and Applications, Boston, MA: Academic Press, 2016, and Industrial Internet of Things: Cybermanufacturing Systems, Cham, Switzerland: Springer, 2016. He is the author of more than 100 articles. His research interests include cyber-physical systems, cybersecurity and privacy, internet of things, edge computing, AI/machine learning, big data analytics, unmanned aircraft systems, connected vehicle, smart and connected health, and wireless communications and networking. His research has been featured by popular news media outlets, including IEEE GlobalSpec's Engineering360, USA Today, U.S. News & World Report, Fox News, Association for Unmanned Vehicle Systems International (AUVSI), Forbes, WFTV, and New Atlas.

Dr. Song is a senior member of ACM. Dr. Song was a recipient of the Best Paper Award from the 12th IEEE International Conference on Cyber, Physical and Social Computing (CPSCom-2019), the Best Paper Award from the 2nd IEEE International Conference on Industrial Internet (ICII 2019), and the Best Paper Award from the 19th Integrated Communication, Navigation and Surveillance technologies (ICNS 2019) Conference.

Qinggang Meng (Senior Member, IEEE) received his B.S. and M.S. degrees in electronic engineering from Tianjin University, Tianjin, China, and the Ph.D. degree in intelligent robotics from the Department of Computer Science at Aberystwyth University, Aberystwyth, U.K. He is currently a Professor with the Department of Computer Science, Loughborough University, Loughborough, U.K. His current research interests include biologically inspired learning algorithms and developmental robotics, service robotics, robot learning and adaptation, multi-UAV cooperation, human motion analysis and activity recognition, activity pattern detection, pattern recognition, artificial intelligence, and computer vision. Prof. Meng is on the editorial boards of several journals including IEEE Transactions on Cybernetics.