

Robot Bionic Vision Technologies: A Review

Hongxin Zhang ¹  and Suan Lee ^{2,*} 

¹ Jewxon Intelligent Technology Co., Ltd., No. 293, Gongren East Road, Jiaojiang District, Taizhou 318000, China

² School of Computer Science, Semyung University, 65 Semyung-ro, Jecheon-si 27136, Chungcheongbuk-do, Korea

* Correspondence: suanlee@semyung.ac.kr; Tel.: +82-43-649-1273

Abstract: The visual organ is important for animals to obtain information and understand the outside world; however, robots cannot do so without a visual system. At present, the vision technology of artificial intelligence has achieved automation and relatively simple intelligence; however, bionic vision equipment is not as dexterous and intelligent as the human eye. At present, robots can function as smartly as human beings; however, existing reviews of robot bionic vision are still limited. Robot bionic vision has been explored in view of humans and animals' visual principles and motion characteristics. In this study, the development history of robot bionic vision equipment and related technologies are discussed, the most representative binocular bionic and multi-eye compound eye bionic vision technologies are selected, and the existing technologies are reviewed; their prospects are discussed from the perspective of visual bionic control. This comprehensive study will serve as the most up-to-date source of information regarding developments in the field of robot bionic vision technology.

Keywords: artificial intelligence; robot bionic vision; optical devices; bionic eye; intelligent camera



Citation: Zhang, H.; Lee, S. Robot Bionic Vision Technologies: A Review. *Appl. Sci.* **2022**, *12*, 7970. <https://doi.org/10.3390/app12167970>

Academic Editor: Anton Civit

Received: 29 April 2022

Accepted: 3 August 2022

Published: 9 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Creatures and animals appeared on the earth as early as the Cambrian period. The visual organs of animals first appeared as photosensitive spots and evolved into more complex forms [1,2]. Among these, the most representative are the binocular stereoscopic visual organs of vertebrates and compound eye panoramic visual organs of insects [3–5]. Humans are the most advanced animals in nature. The human eye is an important organ for human beings to recognize and understand the external world. Humans' intellect is based on their sharp eyes and smart brains [6]. The human visual system is mainly composed of the eyeball, optic nerve, lateral geniculate body, and visual cortex of the brain [7]. The human eye possesses not only a sensing system but also a motor and analysis function [8]. Humans have always been curious about their eyes and brains [9], and Carlson [10] began studying the human brain in 1870. Berson proposed that “good camera equipment can provide stable images, and the movement of the retina and the vestibular system are the cameras of our eyes” [11–13]. For many years, most robot vision research has focused on the surface level of optical structures and vision algorithms in bionics. Currently, robot vision can only imitate one or a few physiological functions of the human eye and cannot gather, process, and evaluate images in complex and changing settings as the human eye does. With advancements in computer vision technology, many researchers have focused on the study of associated technologies [14–20]. Through iterative technology, they intend to achieve bionic vision, proceed to understand and model bionic brain processing, and finally, produce robots that are as flexible and intelligent as humans. This paper reviews the development of machine vision equipment and related technologies, selects the most representative binocular bionic vision technology and multi-eye bionic vision technology to display, and reviews the prospects of the existing technology from the perspective of

visual bionic control. This lays a foundation for further understanding the human visual system and researching humanoid visual electronic systems, Table 1 The history of related techniques' research is presented.

Table 1. Representative related research.

Year	Survey Title	Reference	Focus
1859	THE ORIGIN OF SPECIES	Darwin, C	Established the theory of biological evolution and studied the origin of everything [2].
1870	The excitable cerebral cortex	Chad Carlson · · ·	Pioneered the study of the human cortex [10].
1963	Variable Feedback Experiments Testing a Sampled Data Model for Eye Tracking Movements	Young, L R Stark, L.	Created models of machines that imitate human vision, saccades, etc. [21].
1968	The Oculomotor Control System: A Review	DAVID A. ROBINSON.	The saccade, smooth movement, convergence, and control systems of eye movement were studied [22].
1970	Oculomotor Unit Behavior in the Monkey	Robinson D A.	The relationship between the firing rate of motor neurons and eye position and movement was revealed by recording the oculomotor nerves of awake monkeys [23].
1986	A model of the smooth pursuit eye movement system	D. A. Robinson, J. L. Gordon *, and S. E, Gordon.	Through research on macaques, a smooth-tracking-movement model of human eyes was created [24].
1987	Visual motion processing and sensory-motor integration for the smooth pursuit of eye movements	Lisberger S G, Morris E J Tychsen	The proposed smooth-tracking model improves the visual tracking effect and overcomes the contradiction between high gain and large delay [25].
1989	Dynamical neural network organization of the visual pursuit system	D.C. Deno, E.L. Keller, W.F, Crandall.	The dynamic neural network model was extended to a smooth-tracking system [26].
1998	Neural adaptive predictor for visual tracking system	Lunghi, F Lazzari, S Magenes, G	An adaptive predictor was designed to simulate the prediction mechanism of the brain for visual tracking [27].
1992	Adaptive feedback control models of the vestibulocerebellum and spinocerebellum	GOMI, H.; KAWATO, M.	An adaptive feedback control model was proposed, which is helpful for the vestibular eye reflex, the lobule and adaptive correction of eye movement [28].
1998	Eye Finding via Face Detection for a Foveated, Active Vision System	Scassellati, B.	For the first time, robot eye interaction, image acquisition, and recognition functions were realized [29].
2002	Quantitative Analysis of Catch-Up Saccades During Sustained Pursuit	De Brouwer S, Missal M Barnes G	A target-tracking experiment was carried out on macaques. Two conclusions were drawn: (1) there is a continuous overlap between saccade and smooth tracking; (2) the retinal sliding signal is shared by the two movements, which is different from the conclusion that the two traditional systems are completely separated [30].
2005	Vestibular Perception and Action Employ Qualitatively Different Mechanisms. I. Frequency Response of VOR and Perceptual Responses During Translation and Tilt	Merfeld D M, Park S, Gianna-Poulin C, et al.	Established tVOR and rVOR models that simulate humans and studied the VOR and OKR vestibular-reflex eye-movement models through the ICub robot, which proved the important role of the cerebellum in image stabilization [31].

Table 1. Cont.

Year	Survey Title	Reference	Focus
2006	An Object Tracking System Based on Human Neural Pathways of Binocular Motor System	Xiaolin Zhang	A binocular motor system model based on the neural pathways of the human binocular motor system is proposed. Using this model, an active camera control system was constructed [32].
2006	Design of a Humanoid Robot Eye: Models and Experiments	Cannata, G.; D'Andrea, M.; Maggiali, M.	By quantitatively comparing the performance of the robot's eyes with the physiological data of humans and primates during saccades, the hypothesis that the geometry of the eyes and their driving system (extraocular muscles) are closely related was verified [33].
2014	Binocular Initial Location and Extrinsic Parameters Real-time Calculation for Bionic Eye System	Qingbin Wang, Wei Zou, Feng Zhang and De Xu	A simple binocular vision device was designed, using hand-eye calibration and an algorithm model to ensure the depth perception of the binocular vision system [34].
2017	Design of Anthropomorphic Robot Bionic Eyes	Di Fan, Xiaopeng Chen, Taoran Zhang, Xu Chen, Guilin Liu . . .	A new type of anthropomorphic robot with bionic eyes was proposed. A compact series-parallel eye mechanism with 3 degrees of freedom was designed to realize the eyeball movement without eccentricity [35].
2020	Real-Time Robust Stereo Visual SLAM System Based on Bionic Eyes	Yanqing Liu, Dongchen Zhu, Xiaolin Zhang,	A real-time stereo vision SLAM system based on bionic eyes was designed that performed all movements of the human eye [36].
2021	Panoramic Stereo Imaging of a Bionic Compound-Eye Based on Binocular Vision	Wang, X.; Li, D.; Zhang, G.	The optical optimization design scheme and algorithm for panoramic imaging based on binocular stereo vision were proposed, and a panoramic stereo real-time imaging system was developed [37].

* Source: Web of Science and Google Science.

2. Literature Review

Over the years, many academics have investigated the principles, structure, and characteristics of the animal and human visual systems, particularly the bionic vision system. For example, Carlson [10] mimicked both human and animal visual processing and eye-movement properties, creating a model that adequately mimicked human eye properties. Young [38–42] was a pioneer in the study of the human brain cortex in the twentieth century. Robinson [22,43–45] conducted an extensive study on human vestibular movements at MIT and developed a motion model for the human eye. Hubel [46,47] investigated the structure of the visual cortex through his work on eye-movement control and visual neural networks and made seminal contributions to the bionic study of human vision and eye movement that ultimately earned him the Nobel Prize. With the advancement of science and technology, several research institutes, such as MIT, Tokyo Institute of Technology, and the Chinese Academy of Sciences Bionic Machine Vision Laboratory have contributed relevant research. So far, robot eyes have been able to mimic some of the traits and functions of the human eye.

2.1. Human Visual System

Humans and other primates have distinct visual systems. How do you make a robot's eyes as flexible as a human's so that it can understand things quickly and efficiently? Through the study of the human and primate visual systems, we found that the human visual system has the following characteristics. Figure 1 depicts the human visual system's structure.

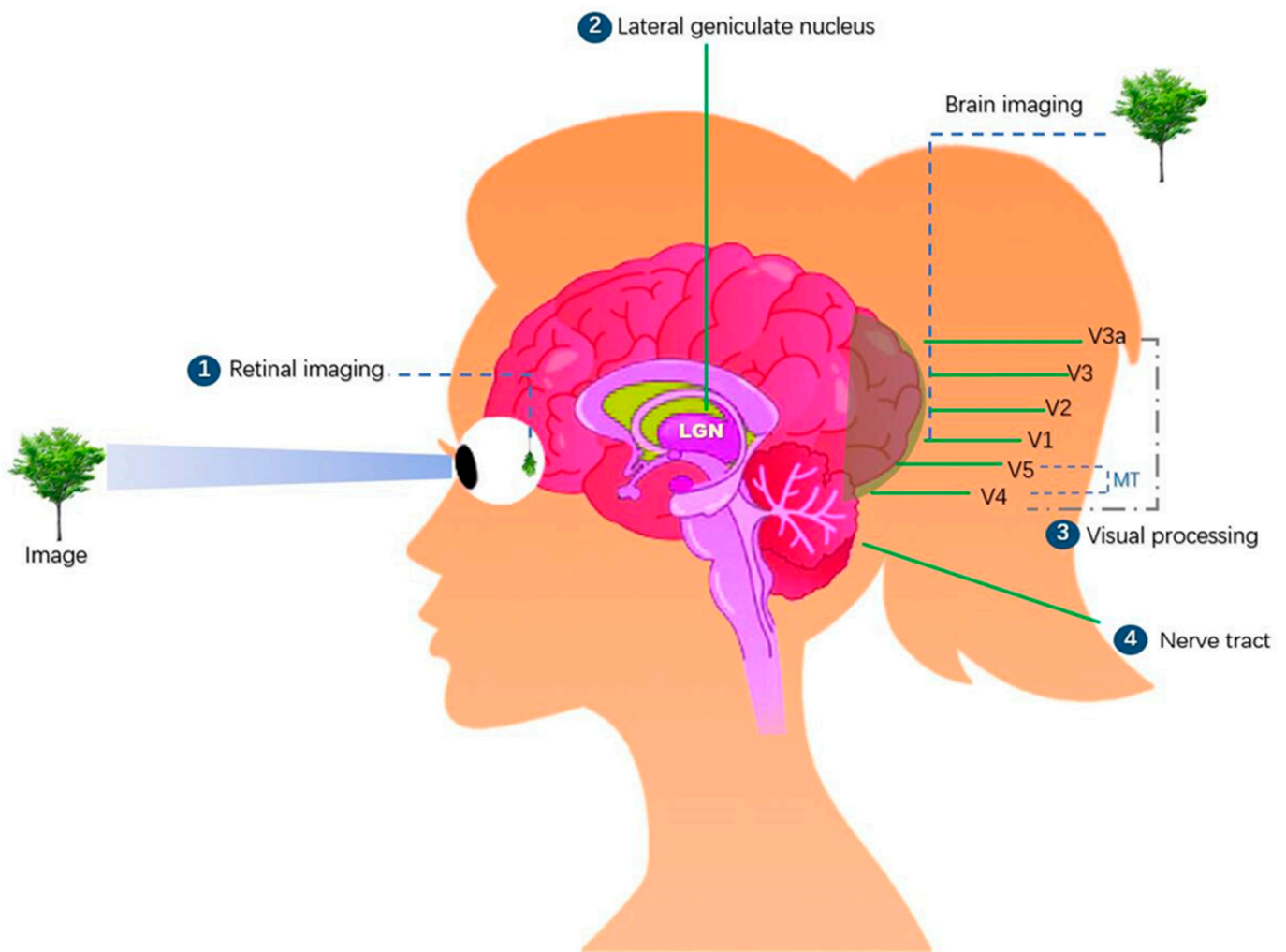


Figure 1. In the human visual system, “Retinal imaging” (①) is the first stage for processing an image when it enters the eye’s visual field. The retina transmits pixels, shade, and color through the chiasma to the brain stem’s “lateral geniculate nucleus” (LGN) (②). Visual information is further processed here, important visual information is extracted, useless information is discarded, and critical visual information is processed and transmitted to the primary “visual cortex” (③) V1 and then via V2 and V3 to V4, V5 (MT area), and higher brain areas [48,49]. The visual flow is then processed layer by layer and transmitted to more areas on the surface of the brain, where the brain identifies the object in the eye’s view [50]. The “nerve tract” (④) runs through the brain and spinal cord throughout the body, and its main function is to control body movement [51].

In nature, the human eye is not the strongest. Animals such as eagles and octopi have eyes more developed than those of humans. What makes humans the most powerful animals in nature? This is the fact that the human brain is highly powerful [52–54]. To understand the deeper principles of the human visual processing system, let us examine the principles of human visual pathways. As shown in Figure 2, the human brain is divided into the left and right brains. The left brain is responsible for the visual field information of the right eye, while the right brain is responsible for the visual field of the left eye. Light waves are projected onto the retina at the bottom of the eyeball through the pupil, lens, and vitreous of the human eye, thereby forming a light path for vision. The optic nerve carries information from both the eyes to the suprachiasmatic nucleus. It then passes to the lateral geniculate nucleus (LGN), where visual information is broken down, twisted, assembled, packaged, and transmitted via optic radiation to the primary visual cortex V1, which detects information regarding the image’s orientation, color, the direction of

motion, and position in the field of vision. The image information is then projected onto the ventral stream, also known as the “what” pathway, and the dorsal stream, also known as the “where” pathway, which is responsible for object recognition. The dorsal pathway is responsible for processing information such as movement and spatial orientation [55–58].

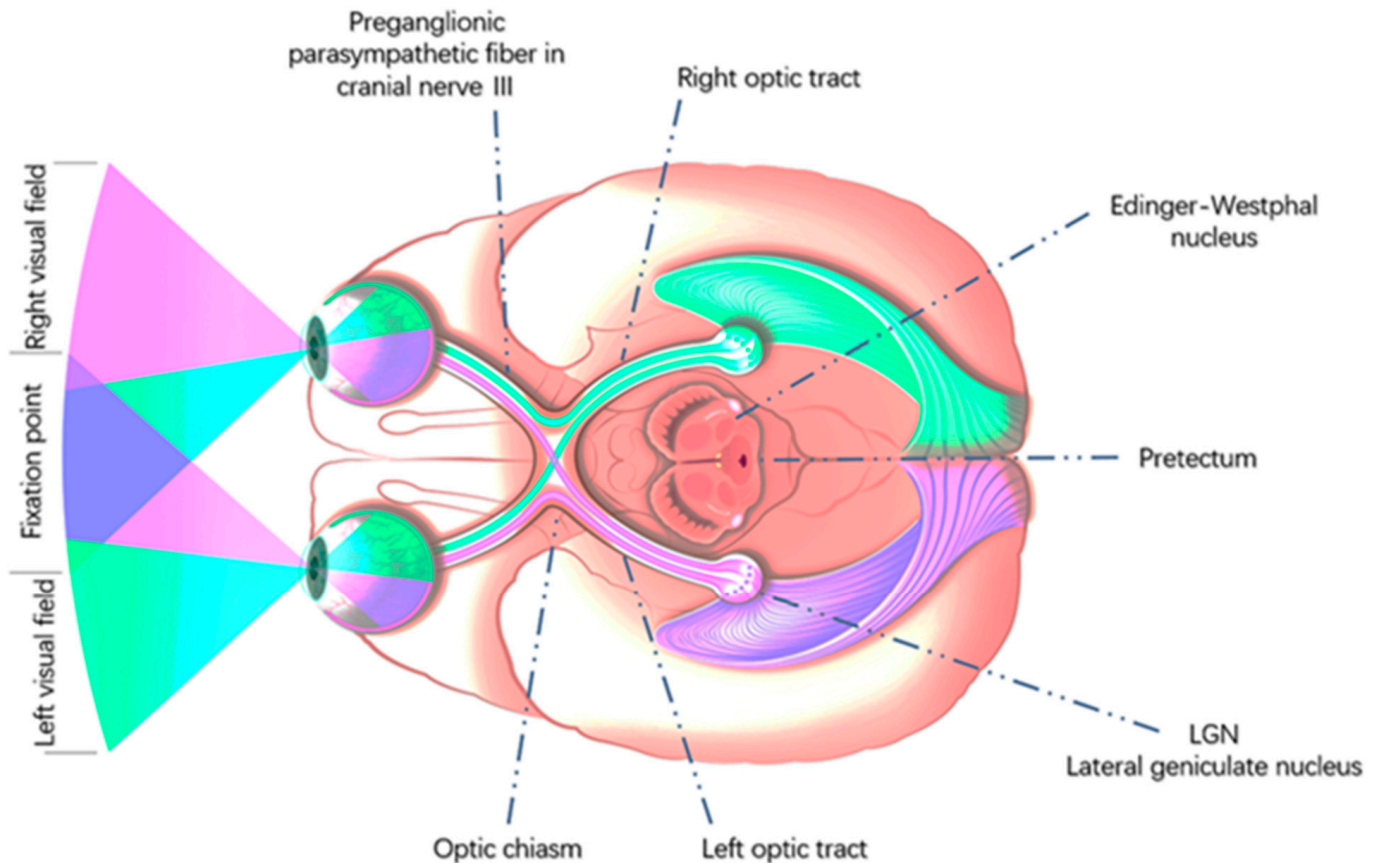


Figure 2. Human visual pathways. Copyright citation: This image is licensed by Eduards Normaaals.

The superior colliculus controls the unconscious eye and head movements, automatically aiming the eye at the target of interest in the field of vision and making it shoot into the fovea of the retina of the human eye. This area contains approximately 10 percent of the neurons in the human brain and is the primary area for visual projection in mammals and vertebrates. The anterior tectum controls iris activity to regulate pupil size [59]. As shown in Figure 2, light information first enters the Edinger–Westphal nucleus through the anterior tectum. The Edinger–Westphal nucleus has the following functions: (i) controlling the ciliary muscle to regulate the shape of the lens; (ii) controlling the ciliary ganglion; and (iii) regulating pupil size via cranial nerve III [60].

Borra conducted anatomical experiments and studies on visual control in the rhesus-monkey brain. Stimuli and scans of the visual cortex of awake rhesus-monkey brains revealed that humans and nonhuman primates have similar visual control systems. However, the human brain is more complicated. Therefore, the similarity between the brain visual systems of rhesus monkeys and humans provides an important basis for humans to understand the principles of their eye movements and conduct experimental research [61,62]. The visual cortex regions of the human and macaque brains are depicted in Figure 3.

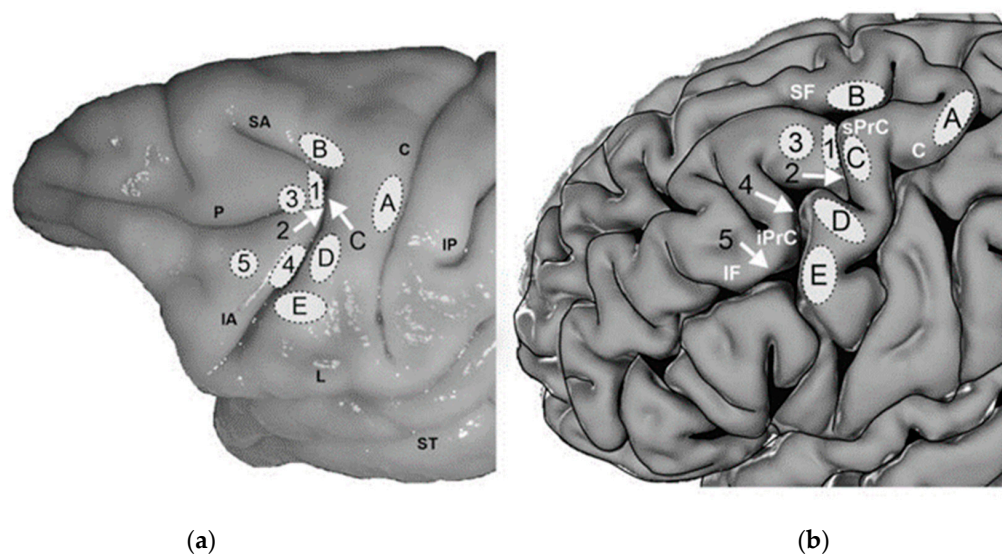


Figure 3. A region at the junction of the prefrontal and premotor cortex in (a) rhesus monkeys and primates, (b) human brain. Known as the anterior field of vision (FEF), it is primarily involved in controlling eye-movement behavior and spatial attention. Humans have at least two types of forwarding fields of vision. After comparing the letters in the visual nerve regions of the human brain and monkey brain, it was found that the monkey visual cortex corresponds to the human visual cortex; and the principles of spatial attention, dynamic vision control, and attention orientation control are very similar [63]. The letters A, B, C, D, and E in the figure represent the visual cortex regions of the human and monkey brains. The image has been licensed by ELSEVIER.

Most mammals have two eyes that allow them to perform complex movements. Each eye is coordinated by six extraocular muscles according to [64]: the inner, outer, upper, and lower rectus muscles and the upper and lower oblique muscles. These muscles coordinate movements, allowing the eye to move freely in all directions and to change the line of sight as needed. The function of the internal and external rectus muscles is to turn the eyes inward and outward, respectively. Because the superior and inferior rectus muscles have an angle of 23° with the optic axis and the superior and inferior oblique muscles have an angle of 51° with the optic axis, they play a secondary role in addition to their main roles. The main function of the superior rectus muscle is upward rotation, whereas its secondary function is internal rotation.

The main function of the inferior rectus muscle is downward rotation, and its secondary functions are internal and external rotation. The main function of the superior oblique is internal rotation, and the secondary functions are external and downward rotation. The main function of the inferior oblique is external rotation, and the secondary functions are external and upward rotation. In addition, the movement of the extraocular muscles is restricted to prevent the eyes from going beyond the range of motion, and mutually restricted extraocular muscles are called antagonistic muscles. Synergistic muscles are cooperative in one direction and antagonistic in another. Movement of the eyeball in all directions requires fine coordination of the external eye muscles, and no eye muscle activity is isolated. Regardless of the direction of the eye movement, 12 extraocular muscles are involved. When a certain extraocular muscle contracts, its synergistic, antagonistic, and partner muscles must act simultaneously and equally to ensure that the image is accurately projected onto the corresponding parts of the retinas of the two eyes [65,66].

The extraocular muscle is driven and controlled by the cranial nerve, which directs the muscle to contract or relax. Eye movements occur when one part of the extraocular muscle contracts and the other relaxes. The motion features of the human eyes mainly include conjugate movement, vergence or disjunctive movement, vestibulo-ocular reflex (VOR), and optokinetic reflex (OKR) [63]. Specifically, the human eye also exhibits the following

motion characteristics [67–73]. The saccade, the most common feature, is a movement occurring rapidly and in parallel from one fixation point to another during which the human eye freely observes its surroundings, lasting between 10 and 80 ms. Mainly used by the human eye to actively adjust the direction of gaze, large saccades require head and neck movement. Smooth pursuit is the movement of the eye tracking a moving object, which is characterized by smooth tracking of the target. The angular velocity of the eye movement corresponds to the velocity of the moving object when the eye is observing a low-speed object. The adult eye can follow objects with 100% smoothness. When the target angular velocity is greater than $150^\circ/\text{s}$, the human eye cannot track the target. When the target object appears in the visual field of the human eye and begins to move at a high speed, the human eye tracks the target. The eyes have rhythmic horizontal nystagmus and maintain the same direction as the movement of the object. When the direction and speed of the human eye are consistent with those of the object, the human eye can clearly see the object. For example, one can sit in a fast-moving car and see the scenery outside the window. When the head moves, the eyes move in the opposite direction; thus, the image maintains steady reflective movement on the retina. For example, when the head moves to the right, the eyes move to the left, and vice versa. The primary function of OVR is motion stabilization, which is characterized by the coordination of head movements to obtain a more stable image. OVR has been used in camera anti-shake mechanisms. Conjugate motion is the simultaneous movement of the eyes in the same direction so that diplopia does not occur. Its characteristics can be divided into saccades and smooth tracking motions. When a target approaches, the visual axes of the two eyes intersect inward, presenting a convergent motion state, whereas when the target moves far away, the visual axis of the two eyes diverges outward, presenting a state of divergent motion. The hold time of the convergent motion is approximately 180 ms, and that of the divergent motion is approximately 200 ms. The process of keeping one's eyes on the target for a long time is called fixation. The eyeball quivers during the gaze, keeping the image refreshed and achieving a clearer visual effect. The eyes of most mammals vibrate slightly when focusing on stationary objects. The mean amplitude of a saccade is 0.2 ms (millisecond), which is half of the cone cell diameter. The vibration frequency reaches 80–100 Hz. Movement can produce light and light effects, and its purpose is to keep the visual center and visual cells in a state of excitement and maintain a high degree of visual sensitivity.

When we observe a scene through our eyes, our eyes do not perform a single motion feature but combine the various eye movement features mentioned above to achieve high-definition, wide-dynamic-range, high-speed, and flexible-target visual tracking. Making machines have the same advantages as humans and animals is the research goal of future robots.

2.2. Differences and Similarities between Human Vision and Bionic Vision

The human visual system operates on a network of interconnected cortical cells and organic neurons. In contrast, a bionic vision system runs on an electronic chip composed of transistors, mainly through a camera with CMOS and CCD sensors, image acquisition, and sending the image to a special image-processing system to obtain the form of the recorded target information and encode it according to pixel, brightness, color, and other information. The image-processing system performs various operations on these codes to extract the characteristics of the target. Through specific equipment, it simulates the visual ability of human beings to make corresponding decisions or execute on these decisions [74]. Figure 4 depicts the human vision processing and representative machine vision processing processes.

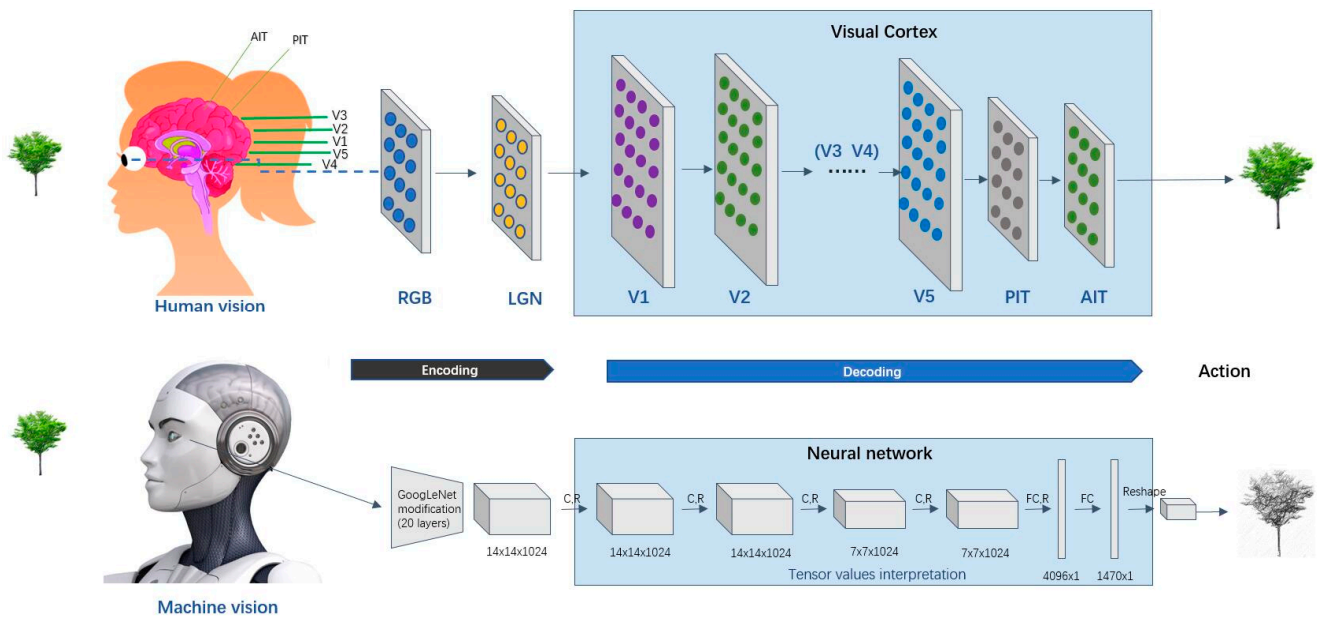


Figure 4. Human vision processing and Representative machine vision processing processes.

Human visual path: RGB light wave → cornea → pupil → lens (light refraction) → vitreous (supporting and fixing the eyeball) → retina (forming object image) → optic nerve (conducting visual information) → lateral geniculate body → optic chiasma → primary visual cortex V1 → other visual cortices V2, V3, V4, V5, etc. → the brain understands and executes → other nerves are innervated to perform corresponding actions.

Bionic vision path: RGB color image → image sensor (image coding) → image processing (image decoding) → image operator (image gray processing, image correction, image segmentation, image recognition, etc.) → output results and execution.

The information in the brain moves in multiple directions. Light signals move from the retina to the inferior temporal cortex, where they are transmitted to V1, V2, and other layers of the visual cortex. Simultaneously, each layer provides feedback on the previous layer. In each layer, neurons interact with each other and communicate information, and all interactions and connections essentially help the brain to fill in gaps in visual input and make inferences when information is incomplete.

The machine mainly converts the image into a digital electrical signal through the image acquisition device and transmits it to the image processor. There are plenty of image processors, and the most common one is GPU. With professional AI computing requirements, ASIC (application-specific integrated circuit) emerged. ASIC performs better than CPU, GPU, and other chips, with higher processing speed and lower power consumption. However, ASIC is very expensive to produce [75–77]. For example, Google has developed an ASIC-based TPU that is dedicated to accelerating the computing power of deep neural networks [78]. In addition, there are also NPU, DPU, etc. [79–81]. FPGA is a processor for developers to customize; usually, an image processor integrates machine vision perception and recognition algorithms. There are diverse algorithm models, such as CNN, RNN, R-CNN, Fast R-CNN, Mask R-CNN, YOLO, and SSD [82–84], which vary in terms of performance and processing speed. The most classical and widely used algorithm is the convolutional neural network (CNN) algorithm. Specifically, the convolutional layer first extracts the initial features, then the pooling layer extracts the main features, and finally, the fully connected layer summarizes all features to provide a classifier for prediction and recognition. The convolutional and pooling layers are sufficient to identify the complex image contents, such as faces and car license plates. Data in artificial neural networks often move in a one-way manner. CNN is a “feedforward network”, where the information moves step by step from the input layer to the next layer and the output layer. This network

mainly makes use of the visual center of human beings, and its disadvantage is that it receives all continuous sequences. Utilizing CNN is equivalent to separating the front and the back connections. Therefore, the recursive neural network (RNN) that is dedicated to processing sequences emerges. In addition to entering the current information, each neuron has previously generated memory information that retained the sequence-dependent type. In the visual cortex, the information moves in multiple directions. The brain neurons are equipped with “complex temporal integration capabilities that are lacking in existing networks” [85]. Beyond CNN, Geoffrey Hinton et al. have been developing the “capsule network”. Similar to the neuronal pathway, this new network is much better than the convolutional networks at identifying overlapping numbers [86].

In general, a high-performance image-processor platform is required for a high-precision image recognition algorithm. Although the great leaps in the fields of GPU rendering, hardware architecture, and computer graphics have enabled realistic scene and object rendering, these methods are demanding in terms of both time and computational resources [87]. As most robots are mobile, most bionic vision systems adopt a mobile terminal processor, and their performance cannot match a graphics server. For example, one of the features of the iCub robot is that it is equipped with a stereo camera rig in the head and a 320×240 resolution color camera in each eye. This setup consists of three actuated DoF in the neck of the robot to grant the roll, pitch, and yaw capabilities to the head, as well as three actuated DoF to model the oculomotor system of a human being (tilt, version, and vergence) [88]. Furthermore, F. Bottarel [89] made a comparison of the Mask R-CNN and Faster R-CNN algorithms using the bionic vision hardware of the iCub robot [90] and discovered that Mask R-CNN is superior to Faster R-CNN. Currently, many computer vision algorithm workers are constantly optimizing and compressing their algorithms to minimize the loss of accuracy and match the processing capacity of the end-load system. As shown in Table 2, the volume and accuracy of Mobilenet-Yolov3 surpass Mobilenet-SSD [91]. Moreover, Yuan, L., et al. designed the Florence algorithm, which performed well in ImageNet-1K zero-shot classification, with the top-1 accuracy of 83.74, the top-5 accuracy of 97.18, 62.4 mAP on COCO fine-tuning, 80.36 on VQA, and 87.8 on Kinetics-600 [92].

Table 2. Mobile-net and YOLO algorithm performance test data.

Network	VOC mAP(0.5)	COCO mAP(0.5)	Resolution	Inference Time (NCNN/Kirin 990)	Inference Time (MNN arm82/Kirin 990)	FLOPS	Weight Size
MobileNetV2- YOLOv3-Lite(our)	73.26	37.44	320	28.42 ms	18 ms	1.8 BFlops	8.0 MB
MobileNetV2- YOLOv3- Nano(our)	65.27	30.13	320	10.16 ms	5 ms	0.5 BFlops	3.0 MB
MobileNetV2- YOLOv3	70.7	&	352	32.15 ms	& ms	2.44 BFlops	14.4 MB
MobileNet-SSD	72.7	&	300	26.37 ms	& ms	& BFlops	23.1 MB
YOLOv5s	&	56.2	416	150.5 ms	& ms	13.2 BFlops	28.1 MB
YOLOv3-Tiny-Prn	&	33.1	416	36.6 ms	& ms	3.5 BFlops	18.8 MB
YOLOv4-Tiny	&	40.2	416	44.6 ms	& ms	6.9 BFlops	23.1 MB
YOLO-Nano	69.1	&	416	& ms	& ms	4.57 BFlops	4.0 MB

Data sources: <https://github.com/dog-qiuqiu/MobileNet-Yolo>, accessed on 1 January 2021. Test platform: Mobile inference frameworks benchmark (4*ARM_CPU). The platform system is based on: NCNN and MNN. Conclusion: With a volume of 8.0 MB (15.1 MB less than SSD), Mobilenet-Yolov3 achieves 73.26% mAP (0.56% more than SSD).

Therefore, compared with machine vision, not only is the human vision system processing power great, but the human visual system also can dynamically modify the attention sensitivity in response to various targets. However, this flexibility is difficult to achieve in computer vision systems. Currently, computer vision systems are mostly intended for single purposes such as object classification, object placement, image region partitioning by

object, image content description, and fresh image production. There is still a gap between the high-level architecture of artificial neural networks and the functioning of the human visual brain [93].

Perception and recognition technology is the most critical technology for intelligent robots [94,95]. Perception technology based on visual sensors has been the most researched technology [96]. People use a variety of computer-vision-based methods to build a visual system with an initial “vision” function. However, the functions of the “eyes” of intelligent robots are still relatively low-level, especially in terms of binocular coordination, tracking of sudden changes or unknown moving targets, the contradiction between a large field of view and accurate tracking, and compensation for vision deviation caused by vibration. After millions of years of evolution, biological vision has developed and perfected the ability to adapt to internal and external environments. According to the different functional performances and applications of simulated biological vision, the research on visual bionics can be summarized into two aspects: one is research from the perspective of visual perception and cognition. The second is research from the perspective of eye movement and sight control. The former mainly studies the visual perception mechanism model, information feature extraction and processing mechanisms, and target search in complex scenes. The latter, based on the eye-movement control mechanism of humans and other primates, attempts to build the “eyes” of intelligent robots to achieve a variety of excellent biological eye functions [97–99].

2.3. Advantages and Disadvantages of Robot Bionic Vision

Compared to human vision, robot bionic vision has the following advantages [100–103]:

1. High accuracy: Human vision is 64 gray-levels, and the resolution of small targets is low [104]. Machine vision can identify significantly more gray levels and resolve micron-scale targets. Human visual adaptability is considerably strong and can identify a target in a complex and changing environment. However, color identification is easily influenced by a person’s psychology. Humans cannot identify color quantitatively, and their gray-level identification can also be considered poor, normally seeing only 64 grayscale levels. Their ability to resolve and identify tiny objects is weak.
2. Fast: According to Potter [105], the human brain can process images seen by the human eye within 13 ms, which is converted into approximately 75 frames per second. The results extend beyond the 100 ms recognized in earlier studies [106]. Bionic vision can use a 1000 frames per second frame rate or higher to realize rapid recognition in high-speed image movement, which is impossible for human beings.
3. High stability: Bionic vision detection equipment does not have fatigue problems, and there are no emotional fluctuations. Bionic vision is carefully executed according to the algorithm and requirements every time with high efficiency and stability. However, for large volumes of image detection or in the cases of high-speed or small-object detection, the human eye performs poorly. There is a relatively high rate of missed detections owing to fatigue or inability.
4. Information integration and retention: The amount of information obtained by bionic vision is comprehensive and traceable, and the relevant information can be easily retained and integrated.

Robot bionic vision also has disadvantages such as a low level of intelligence, for it is unable to make subjective judgments, has poor adaptability, and has large initial investment cost. Furthermore, by incorporating non-visible light filling and photosensitive technology, bionic vision achieves night vision beyond the sensitivity range of the human eye. Bionic vision is not affected by severe environments, is not fatigued, can operate frequently, and has a low cost of continuous use. It simply requires an investment at the beginning to set up. Continued operation incurs only energy and maintenance costs.

2.4. The Development Process of Bionic Vision

If light from a source object is shone into a dark space through a small hole, an inverted image of the object can be projected onto a screen, such as the opposite wall. This phenomenon is called "pinhole imaging". In 1839, Daguerre of France invented the camera [107], and thus began the era of optical imaging technology. The stereo prism was invented in 1838. By 1880, Miguel, a handmade camera manufacturer in London, UK, had produced dry stereo cameras. At that time, people had early stereo vision devices, but these devices could only store images in films. In 1969, Willard S. Boyle and George E. Smith of Bell Laboratories invented a charge-coupled device (CCD), which is an important component of the camera [108]. A CCD is a photosensitive semiconductor. They can convert optical images into electrical signals. A tiny photosensitive material implanted in the CCD is called a pixel. The more pixels a CCD contains, the higher its picture resolution. However, the function of the CCD is similar to that of a film: it converts the optical signal into a charge signal. The current signal is amplified and converted into a digital signal to realize the acquisition, storage, transmission, processing, and reproduction of the image. In 1975, camera manufacturer Kodak invented the world's first digital camera using photosensitive CCD elements six years earlier than Sony [109]. At that time, the camera had 10,000 pixels, but the digital camera was still black and white. Surprisingly, in 2000, Sharp's J-SH04 mobile phone was equipped with a 110,000 pixel micro CCD image sensor. Thus began the age of phone cameras. People can now carry cameras in their pockets [110,111]. The development of vision equipment and bionic vision technology in Figure 5.

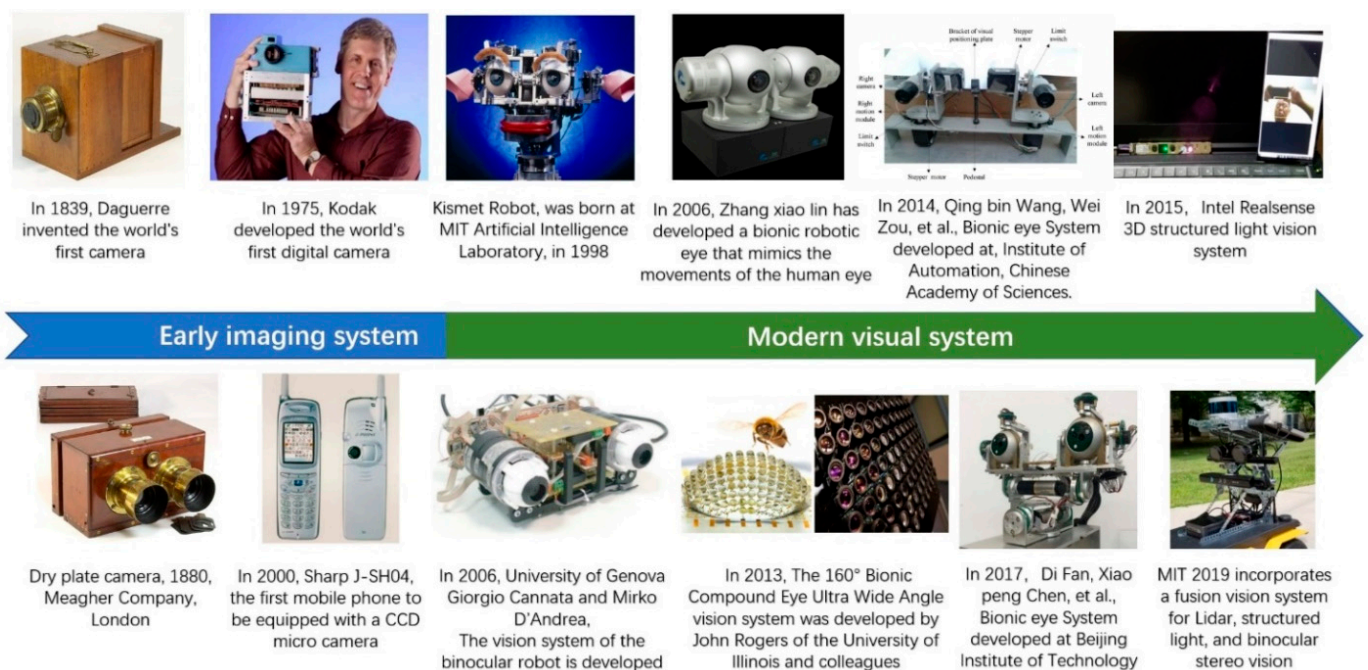


Figure 5. The development of vision equipment and bionic vision technology.

Early image systems laid a solid foundation for modern vision systems and the future development of imaging techniques and processing. By consulting and searching through vast and relevant technical literature, it was found that the current industry mainly includes two technical schools: the bionic binocular vision system and the bionic compound eye vision system. The main research scholars are B. Scassellati, G. Canata, Z. Wei, Z. Xiaolin, and W. Xinhua.

According to WHO statistics, 285 million people worldwide are visually impaired. Of these, 39 million are classified as blind [112]. Scientists have been working to develop bionic eye devices that can restore vision. Arthur Lowery has studied this topic for several decades. His team successfully developed a bionic vision system called Gennaris [113].

Because the optic nerve of a blind person is damaged, signals from the retina cannot be transmitted to the “visual center” of the brain. The Gennaris bionic vision system can bypass damaged optic nerves. The system includes a “helmet” with a camera, wireless transmitter, processor, and 9×9 mm patch implanted into the brain. The camera on the helmet sends captured images to the visual processor, where they are processed into useful information. The processed data are wirelessly transmitted to the patch implanted into the brain and then converted into electrical pulses. Through microelectrode stimulation, images are generated in the visual cortex of the brain. Professor Fan Zhuo of the Hong Kong University of Science and Technology and his team successfully developed a bionic eye based on the hemispherical retina of a perovskite nanowire array. This bionic eye can achieve a high level of image resolution and is expected to be used in robots and scientific instruments [114,115].

In 1998, the Artificial Intelligence Laboratory at MIT created a vision system for the Kismet robot. However, at that time, robot vision could only achieve a considerably simple blinking function and image acquisition. Since then, significant progress has been made in binocular robot vision systems. In 2006, Cannata [33] from the University of Genoa designed a binocular robot eye by imitating the human eye and adding a simple eye-movement algorithm. At that time, the eyes of the robot could move rigidly up, down, left, and right. Subsequently, it was recognized by the Italian Institute of Technology (IIT) and applied to the iCub robot after continuous technical iterations [116]. In 2000, Honda developed the ASIMO robot [117]. Currently, the latest generation of ASIMO is equipped with binocular vision devices [118]. Thus far, most robots, including HRP-4C [119] and the robots being developed at Boston Dynamics, combine 3D LiDAR with a variety of recognition technologies to improve the reliability of robot vision [120]. Robots that have bionic vision in Figure 6.



Figure 6. Robots that have bionic vision. Copyright: Robot licensed by Maximalfocus US-L.

Machine vision processing technology is also being developed. As early as the 1980s, the Artificial Intelligence Laboratory at MIT began conducting relevant research. In the 1990s, bionic vision technology began to perform feature analysis and image extraction. In June 2000, after the first open-source version of OpenCV Alpha 3 was released [121], with the emergence of CNN, RNN, YOLO, ImageNet, and other technologies, bionic vision entered a stage of rapid development using artificial intelligence [122,123].

2.5. Common Bionic Vision Techniques

Bionic vision technology uses a camera equipped with a CCD or CMOS image sensor to obtain the depth information for an object or space. It is the most used technology

in intelligent devices such as robot vision systems and automatic vehicle driving. For example, NASA carried a binocular vision system on the Perseverance Mars Rover in 2020 [124]. In 2021, the Astronomy-1 Lunar Rover developed by China Aerospace also used a binocular terrain-navigation vision system [125]. At present, the mainstream technologies of binocular bionic vision include stereo vision, structured light, time of flight (TOF), etc. [96,126–140].

2.5.1. Binocular Stereo Vision

This technology was first introduced in the mid-1960s. It is a method for obtaining three-dimensional geometric information about an object by calculating the position deviation between the corresponding points on its images based on the parallax principle and using imaging equipment to obtain two images of the measured object from different positions. A binocular camera does not actively emit light; therefore, it is called a “passive depth camera”. At present, it is used in the field of scientific research because it does not require a transmitter and receiver of structured light and TOF; thus, the hardware cost is low. Because it depends on natural light, it can be used indoors and outdoors, but strong and dark light can affect the quality of the image or the depth measurement of the object. This is a known disadvantage of the binocular camera. For example, for a solid color wall, because the binocular camera matches the image according to visual characteristics, a single background color causes a matching failure, and the visual matching algorithm is complex. The main representative products include Leap Motion and robotic eyes.

2.5.2. Structured Light





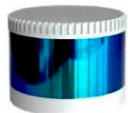
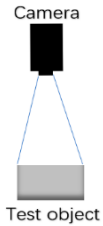
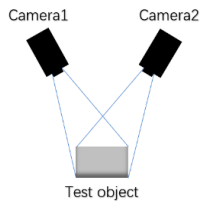
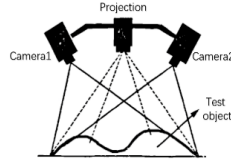
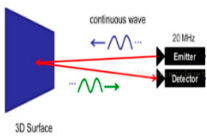
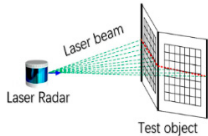
The basic principle of this technology is that the main hardware includes a projector and camera. The projector actively emits (therefore called active measurement) invisible infrared light onto the surface of the measured object, then captures pictures of the object through one or more cameras, collects structured light images, sends the data to a calculation unit, and calculates and obtains position and depth information through the triangulation principle to realize 3D reconstruction. Therefore, structured light is used to determine the structure of the object. There are many projection patterns, such as the sinusoidal fringe phase-shift method, binary-coded gray code, and phase-shift method + gray code. The advantages of this method include mature technology, low power consumption, low cost, active projection, suitability for weak lighting, high precision within a close range (within 1 m), and millimeter-level accuracy. A major disadvantage is poor long-distance accuracy. With the extension of the distance, the projection pattern becomes larger, and the accuracy worsens. It is not suitable for strong outdoor light, which can easily interfere with projection light. Representative products include Kinect, Primesense, and Intel Realsense.

2.5.3. TOF (Time of Flight)

This technology is a 3D imaging technology that has been applied to mobile phone cameras since 2018. Its principle is to emit continuous pulsed infrared light of a specific wavelength onto the target and then receive the optical signal returned by the target object via a specific sensor. The round-trip flight time or phase difference of the light is calculated to obtain depth information for the target object. The TOF lens is mainly composed of a light-emitting unit, optical lens, and image sensor. The TOF performs well in terms of close measurements and recognition. Its recognition distance can reach 0.4 m to 5 m. Unlike binocular cameras and structured light, which require algorithmic processing to output 3D data, TOF can directly output the 3D data of the target object. Although structured light technology is best suited for static scenes, TOF is more suitable for dynamic scenes. The main disadvantage is that its accuracy is poor, and it is difficult to achieve millimeter-level accuracy. Because of its requirements for time measurement equipment, it is not suitable for short-range (within 1 m) applications compared to structured light, which performs well for short- and high-precision applications. It cannot operate normally under strong light interference. Representative products that use this technology include TI-Opt and ST-V153.

It can be seen that Table 3 presents the following advantages and disadvantages: (1) monocular technology has a low cost and high speed, but its accuracy is limited, and it is unable to obtain depth information; (2) binocular ranging is the closest to human vision, without a passive light source, and the working efficiency is acceptable, but it is difficult to realize a high-precision algorithm, and it is easily disturbed by environmental factors; (3) a structured light scheme is more suitable for short-range measurements but has high power consumption; and (4) TOF and structured light have the same disadvantages. The miniaturization of the sensor has a significant impact on resolution. (5) Although LiDAR has high precision, even for long distances, its resolution is low. The machine vision measurement technologies have both advantages and disadvantages. Therefore, in recent years, many well-known robot manufacturers have adopted technology integration to improve the accuracy and practicality of robotic vision. For example, Atlas [141] of Boston Power adopts a vision combination of binocular and structured light + LiDAR.

Table 3. Performance comparison table of mainstream visual technology.

Technology Category	Monocular Vision	Binocular Stereo Vision	Structured Light	TOF	Optical Laser Radar
Product pictures					
Technology principle					
Principle of work	Single camera	Dual camera	Camera and infrared projection patterns	Infrared reflection time difference	Time difference of laser pulse reflection
Response time	Fast	Medium	Slow	Medium	Medium
Weak light	Weak	Weak	Good	Good	Good
Bright light	Good	Good	Weak	Medium	Medium
Identification precision	Low	Low	Medium	Low	Medium
Resolving capability	High	High	Medium	Low	Low
Identification distance	Medium	Medium	Very short	Short	Far
Operation difficulty	Low	High	Medium	Low	High
Cost	Low	Medium	High	Medium	High
Power consumption	Low	Low	Medium	Low	High
Disadvantages	Low recognition accuracy, poor dark light	Dark light features are not obvious	High requirements for ambient light, short recognition distance	Low resolution, short recognition distance, limited by light intensity	Cloudy and rainy days, fog, and other weather interference have effects
Representative company	Cognex, Honda, Keyence	LeapMoTion, iit	Intel, Microsoft, PrimeSense	Intel, TI, ST, Pmd	Velodyne, Boston Dynamics

Data sources: Official websites of related products and patent databases.

2.6. Robot Bionic Vision

Since the development of the first autonomous mobile robot, Shakey [142], by the Stanford Institute in the late 1960s, various robots have been developed. An important

device for robots to obtain external information is the robot vision system, which is an evolved version of bionic vision. It integrates bionic kinematics, visual algorithms, and adaptive deep-learning technology. The goal of robot vision is to realize a combination of the flexible movement of the human eye, understanding and memory of the human visual nervous system, and rapid image processing, which are mainly divided into two research directions:

2.6.1. Early Robot Vision System Model

In 1998, Breazeal created the Kismet robot at the Artificial Intelligence Laboratory of MIT. As shown in Figure 7, Kismet is an autonomous robot designed for social interactions with humans. The robot vision system consists of four-color CCD cameras installed on an active-vision head. Two wide field-of-view (FOV) cameras are mounted at the center and moved relative to the head. These are 0.25-inch CCD lipstick cameras with a 2.2 mm lens produced by Elmo. They are used to determine what the robot should pay attention to and for distance estimation. A camera is installed on the pupil of each eye. These are 0.5-inch CCD concave cameras with an 8 mm focal length lens for higher-resolution attention post-processing, such as eye detection. Kismet has three degrees of freedom to control the direction of sight and three degrees of freedom to control its neck. This motion is enabled using a Maxon DC servo motor. It is equipped with a high-resolution optical encoder to realize accurate position control. This enables robots to move and adjust the direction of their eyes like humans and participate in various human visual behaviors. Kismet can express emotions such as calm, interest, anger, happiness, depression, surprise, and nausea. The robot behaves like a human through facial expressions [143].

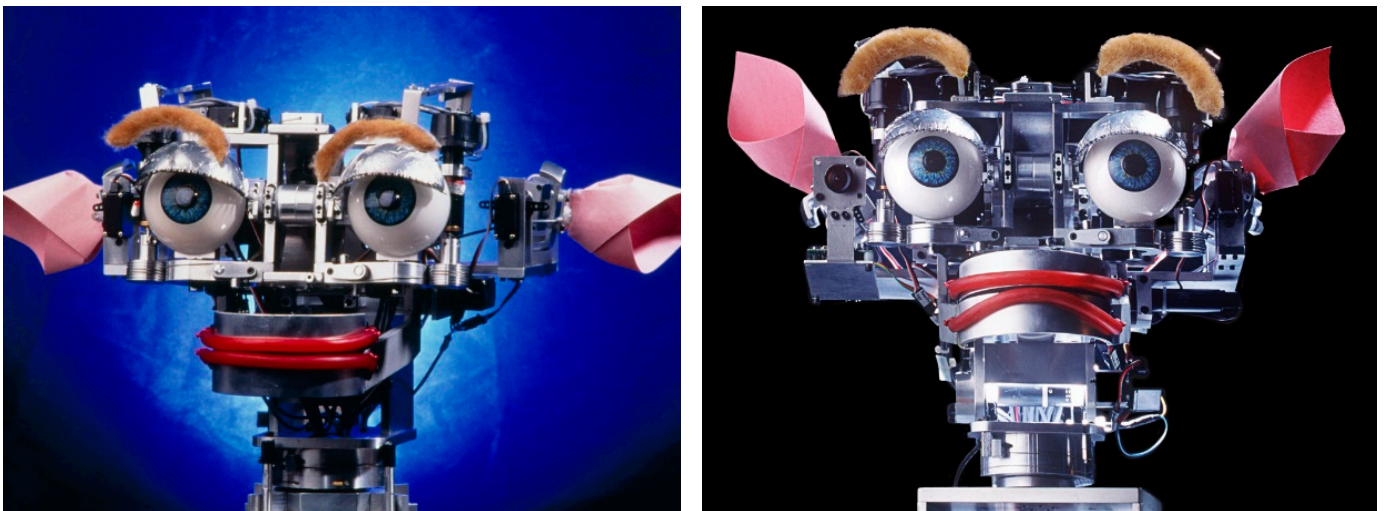


Figure 7. “Kismet” robot expression interactive vision system developed by MIT Artificial Intelligence Laboratory; Copyright citation: “MIT Museum” by Angela N. License CC (Attribution 2.0).

In 2006, Cannata [33] of the University of Genoa developed a robot bionic eye and proposed an eye-motion model and human eye muscle motion algorithm. As shown in Figure 8, It aimed to simulate the actual scanning and smooth tracking motions performed by the human eye. The system consists of a sphere (eyeball) driven by four independent tendons, which, in turn, are driven by four DC motors integrated into the system. The eyeball is equipped with a miniature CMOS color camera and is fixed by low-friction support that allows three degrees of freedom of rotation. Optical sensors provide feedback to control tendon tension during applications such as surgery. The control of the eyes was performed using an embedded controller connected to the host using the CAN bus. The proposed model was used for tendon-driven humanoid robot eyes [116]. Subsequently, this technology was applied to the neuron morphological control module of the real-time radiant eye movement of the iCub robot head.

and electronic compensation was proposed. In the electronic image stabilization stage, the IMU generates the rotational motion of the three-dimensional camera and detects the two-dimensional image movement using the image-movement-detection algorithm. These motions are further filtered, and the image sequence is re-bent for a second time to remove the rotation jitter and translation jitter. The algorithm not only ensures real-time performance but also improves the fidelity of the inter-frame transformation [157].

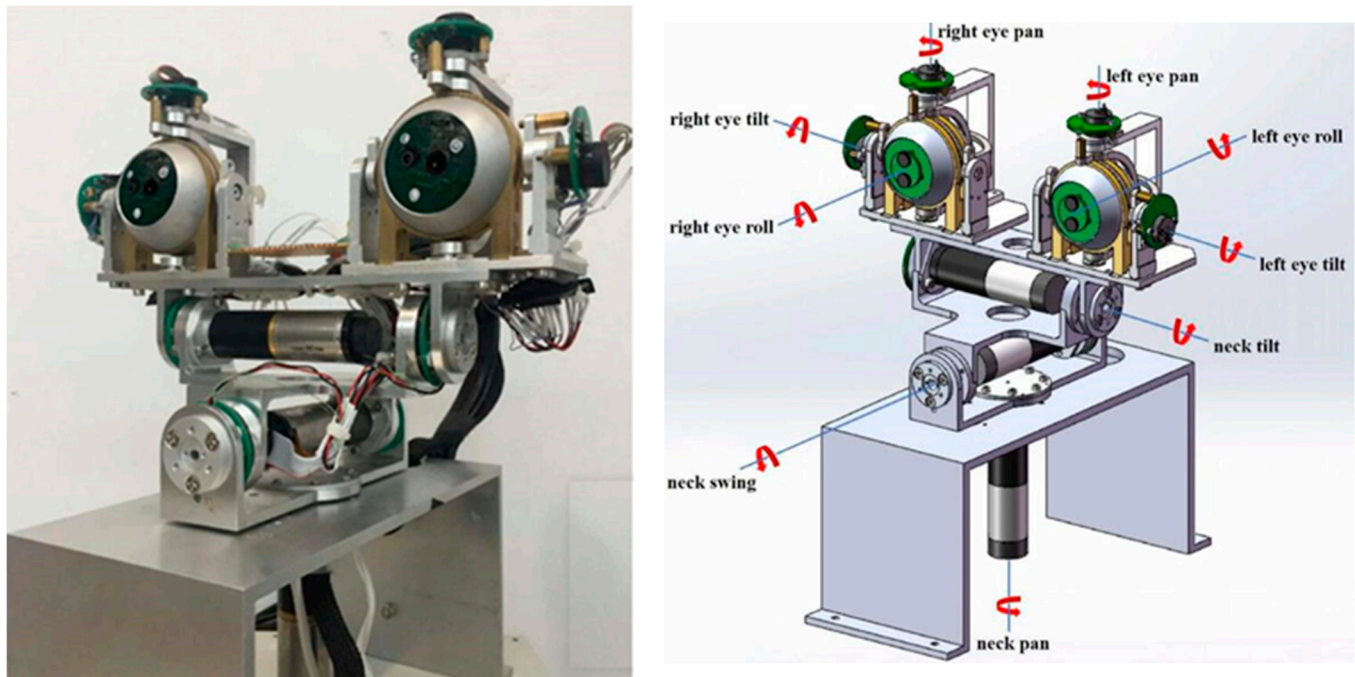


Figure 11. Bionic eye Vision Platform and Algorithm system. Copyright citation: This image is licensed by IEEE.

2.7. A Panoramic Imaging System Based on Binocular Stereo Vision

Compound eye technology has entered the field of mobile cameras. Currently, many new mobile phones are equipped with multiple cameras to achieve higher image quality through the combination of long-focus, medium-focus, and short-focus cameras. Then, Wang Xinhua and others at the School of Computer Science of Northeast Electric Power University and the State Key Laboratory of Applied Optics at the Changchun Institute of Optics, Precision Mechanics, and Physics, within the Chinese Academy of Sciences, refined the technology and designed a panoramic imaging system based on binocular stereo vision. This technology simulates the panoramic vision of the left and right eyes using an array camera and realizes the binocular stereo vision effect of a large field of view. The imaging principle of stereo vision is discussed based on a geometric model of binocular stereo vision. An optical optimization design scheme for panoramic imaging based on binocular stereo vision is proposed, and a spatial coordinate calibration platform for an ultra-high-precision panoramic camera based on the theodolite angle compensation function is constructed. The projection matrix of adjacent edges is obtained by solving the imaging principle of binocular stereo vision. Then, a real-time registration algorithm of multidetector stitched images and the Lucas Kanade optical-flow method based on image segmentation is proposed to realize stereo matching and depth-information estimation of panoramic images, and the estimation results are effectively analyzed. The experimental results show that the binocular panoramic stereo imaging system can satisfy the requirements for real-time and accurate image processing in virtual reality engineering applications.

2.8. Robot Bionic Vision Development Based on Human Vision

The human eye changes the shape of the lens through the contraction or relaxation of the ciliary muscle, thereby realizing zooming to obtain clear images of objects at different distances [54]. Referring to the zoom mechanism of the human eye, researchers have designed flexible zoom lenses with different driving modes and physical states [9]. Although the zooming process is different and the zooming effect is slightly different, most zooming processes achieve the zooming purpose by changing the curvature of the lens surface, and a few achieve the zooming purpose by changing the refractive index of the lens constituent material. The human eye changes the aperture of the pupil through the expansion and contraction of the smooth muscle on the iris (shrinks in bright light and expands in dark light), thereby controlling the amount of light entering the pupil to adapt to various lighting environment changes [8]. To enhance the imaging performance of the bionic visual optical system, in recent years, researchers have simulated the iris adjustment mechanism of the human eye, and have carried out many studies related to iris bionic technology. According to the differing focus of research, it can be roughly divided into two categories: the first category focuses on exploring the physiological model of pupil change, studies the relationship between human pupil diameter and brightness, and proposes a mathematical model of pupil diameter change based on adaptive adjustment of brightness. The other type focuses on the application of the iris driving mechanism in engineering. By simulating the telescopic motion mechanism of the smooth muscle on the iris, a freely adjustable optofluidic variable aperture optomechanical device is developed [59].

There are two different types of photoreceptors in the retina of the human eye: cone and rod cells. The two types of photoreceptor cells contribute differently to vision [65]. Cone cells sense strong light and color, have a strong resolution for object details and colors, and are mainly responsible for producing photopic vision in a brighter environment. Rod cells are sensitive to weak light and can only distinguish between light and dark but cannot distinguish the details and colors of objects. Rod cells are mainly responsible for producing scotopic vision in dim environments. These two types of photoreceptor cells are unevenly distributed in the retina [12]. There are approximately 7 million cone cells that are mainly distributed in the fovea area of the retina. The closer to the edge of the retina, the lower the density of cone cells, and the greater the density of rod cells. Based on the distribution and physiological characteristics of the above-mentioned retinal cone cells and rod cells, the details and color resolution of the human retina change with the change in spatial position; therefore, the retina can process information in a large field of view, high resolution, and real-time [13]. A balance between them can be achieved to meet the application requirements on different occasions. Inspired by this, in recent years, researchers at home and abroad have started to work on bionic research on retinal photoreceptors and imaging mechanisms. At present, research related to retinal bionic technology can be divided into two categories: the distribution characteristics of cone and rod cells and the retina's physiological characteristics [51,52].

According to the uneven distribution of retinal photoreceptor cells, only the target image in the fovea area has the characteristic of high resolution. However, the objects we observe are usually not static, and the scene changes. Therefore, to consider the need of a large field of view and high resolution, the entire eyeball needs to keep rotating, and the head and body movements also assist with changes in eyesight. Many physiological studies have shown that eye movement is a complex physiological process that includes the cooperation of brain nerve centers and multiple extraocular muscles to achieve the function of human eyesight shifting. To understand the eye-movement mechanism, simplify the human eye model, and make the structure and bionic kinematics of the human eye easy to realize, researchers proposed a hypothetical eye-movement center, eye-rotation axis, and Listing's law.

Some bionic vision optomechanical devices with anthropomorphic vision functions, intelligent image processing algorithms, and servo motion control models have been developed. However, after millions of years of continuous evolution, the human eye has

become a complete visual system that integrates image acquisition, information processing, and feedback, including the cornea, iris, lens, retina, optic neuron, and cerebral cortex visual center [135,136]. At present, the image acquisition, processing speed, and imaging performance of optomechanical devices, intelligent algorithms, and servo motion control models are far from those of the complex human visual system. This limits the application and development of intelligent robots in complex dynamic environments.

3. Challenges and Countermeasures for Robot Bionic Vision Development

3.1. Challenges That Restrict Robot Bionic Vision Development

The eye-movement control system is not only related to eye activity but also closely related to neural activity in the brain, cerebellum, and brain stem. From the medical perspective, this is an interdisciplinary and complex system. Therefore, the establishment of a machine eye-movement control system based on a biological vision mechanism is a research topic of interdisciplinary integration [158]. It is not only related to neurophysiology and medicine but also closely related to control science, bionics, biomedical engineering, and computer science. Although scholars at home and abroad have conducted many simulations and bionic studies on biological eye movement from different angles and for different purposes, there are still many problems to be further studied and discussed, mainly the following aspects:

- Although many scholars have conducted in-depth research on human vision, bionic vision still requires further research.
- At present, most robot vision systems use only left and right cameras to obtain and process target images. Left- and right-camera-driven systems have poor coordination, slow target tracking, and easy target loss. It is difficult to compensate for the line-of-sight deviation caused by image jitter in a complex motion environment. At present, researchers have constructed a head–eye motion control model of a bionic robot, which cannot fundamentally solve the problems faced by this version of this bionic machine [159].
- Existing bionic eye models usually only realize one or two movements of the human eye, and most are one-dimensional horizontal monocular or binocular movements. The separation mechanism of saccade motion and smooth tracking motion is widely used; however, its ability to realize multi-eye motion is poor.
- Binocular linkage control and head–eye coordination control in the line-of-sight transfer process are key technologies of bionic robot eyes. At present, research on binocular head–eye coordinated movement mostly remains at the level of biological neural control mechanisms and physiological models. Research on the construction of bionic binocular head-and-neck coordinated control is still in its infancy.
- The advanced mechanisms of human vision have not been fully explored, and some mechanisms remain controversial and undiscussed. Despite research and analysis from the perspectives of neurophysiology and anatomy, there has been no breakthrough in head–eye coordinated motion control technology. Therefore, it is necessary to establish new models, theories, and methods to make a machine’s eyes more intelligent, dexterous, and realistic.

At present, most research on robot eyes is based on engineering methods. The left and right cameras are used to obtain the target images for processing. The left and right eyes and the head and neck lack a coordinated linkage mechanism. There are many technical obstacles, and the adoption of bionic technology is an important method to solve these problems. Research on bionic robotic eyes is multidisciplinary, involving neurophysiology, bionics, control science, and computer science. There are still many problems in the research on robot bionic vision that need to be explored further.

3.2. Countermeasures That Can Promote Robot Bionic Vision Development

3.2.1. Building a Complete Bionic Eye Model

Most foreign research is conducted by neurophysiologists. Through physiological experiments and data analysis, the neural mechanisms and pathological characteristics of eye movements are explored to provide technical support for clinical medicine. Most of these research results are limited to physiological models, and their engineering implementation and application have rarely been considered. Based on the research results of physiology, a combination of engineering bionics and control theory is used to transform the physiological model of vision control into an engineering technology model, thereby realizing the excellent performance of biological vision. This fundamentally solves the technical problems faced by machine vision and is an important direction for intelligent robot research. The bionic eye models in the current literature only simulate one or two movements of the human eye, and most of them are one-dimensional horizontal movements of one or both eyes. The separation mechanism of saccade and smooth tracking is generally used, and it is difficult to simultaneously achieve multiple eye movements. Further study of the internal correlation and neural mechanism between various eye movements, the establishment of a multi-degree-of-freedom nonlinear bionic binocular movement model, and realization of various eye movements such as saccades, smooth tracking, anisotropic movement, and reflex movement are required.

3.2.2. Achieving Robot Head–Eye Coordination through a Biological Neural Control Mechanism

The joint redundancy of the robot is required to effectively coordinate the head and binocular movements when performing large-scale gaze shifts. Most of the currently proposed head–eye coordination motion systems and learning algorithms are based on two-dimensional head–eye systems in the horizontal direction. Although some scholars have proposed 3D head–eye coordination motion control algorithms, they use binocular focusing first and then control the head rotation. The “time-sharing” and “segmented” execution methods of eye compensation do not realize coordinated control in the simultaneous turning of the head and eye to the target. Currently, 3D head–eye coordination remains a technical challenge. It is an important research objective to obtain inspiration from the coordinated movement of the biological head and eye, study the neural control mechanism of the coordinated combination of the head, eye, and body of primates to complete the sight shift, and design a control algorithm for the coordinated movement of the head and neck of the binocular.

3.2.3. Improving the Speed and Accuracy of Random-Moving-Target Tracking

One of the important functions of the “eye” of a robot is visual tracking. Currently, a visual servo feedback control method is often used. However, for sudden, fast-changing, and erratic targets, target loss and tracking failure often occur. The extraordinary ability of the human eye to track changing targets stems from the coordination and real-time switching between rapid saccade and the slow and smooth follow-up of the eye in its visual system. According to an in-depth study, the neurophysiological mechanism of automatic switching of human eye-movement mode to track the target quickly and accurately, the switching of saccade and smooth pursuit modes in visual tracking, is regarded as the most adaptive hybrid system. The optimal control problem is studied to solve the problem of the rapidity and accuracy of random-moving-target tracking. The optimal switching timing between the two modes and the prediction algorithm are the key technologies of this research.

3.2.4. Solving the Problem of Large-Scale Line-of-Sight Deviation Compensation

When the robot operates in a complex, unstructured environment, the vibration of its own body or the change in its posture causes a large line-of-sight deviation. Usually, image processing is used to adjust the servo mechanical head; however, the compensation range is small, the image stability is poor, and there is no way to solve the large line-

of-sight deviation. The human eye has strong self-adaptive and self-adjusting functions. When the head and body posture changes or the background changes dynamically, it can still clearly gaze at and track the target because of its vestibule–oculomotor reflex (VOR) and optokinetic reflex (OKR) functions. By studying the mechanism of human reflex eye movement, a feedback controller based on retinal sliding information and an adaptive VOR-OKR model can be established, which can actively compensate for the visual error caused by the change in the robot's body position and then solve the large-scale visual deviation compensation problem of the robot.

4. Conclusions

In today's smart era, humans desperately require more adaptable and intelligent machines to serve humanity. Bionic vision systems have a wide range of applications. However, robot vision remains in the early stages of bionics. Although exploration of the human visual system has laid a solid foundation for artificial-intelligence vision technology, at the current stage of human scientific exploration, machine realization of the human visual system remains in the V1 primary visual cortex. The specific mechanism of action of the advanced visual cortex requires further investigation. To date, robot vision has not been able to fully actualize all the capabilities of the human eye.

At present, many scientific research achievements remain in human eye physiological characteristics, eye-movement principles, and bionic eye prostheses, and research on the robot bionic eye is still in its primary stage. This paper summarized the research progress of artificial intelligence bionic vision technology at home and abroad by citing the relevant literature. By summarizing the visual principles and motion characteristics of humans and animals, robot vision was explored. This study meticulously examined the development history of bionic vision equipment and related technologies, discussed existing bionic vision technologies, and selected the most representative binocular, bionic, and multi-eye compound bionic vision technologies. Prospects of existing technologies from the standpoint of visual bionic control were discussed. Although this study showed that some research progress has been made in using biological vision mechanisms as a reference to study bionic vision control methods that mimic highly accurate and complex human vision, the current technology is only a preliminary attempt, and there are still many problems to be addressed.

Biological head–eye movement and visual control are not only related to the activities of the head and eyes but also related to the activities of the brain and central nervous system. The behavioral expression of brain control strategies is an interdisciplinary and highly complex research subject. Humans are familiar with and yet unfamiliar with their cognition. Perhaps soon, humans will fully understand the working principles of the entire visual cortex and uncover the mysteries of the brain. That will be the era of robots that can perform sophisticated human tasks.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z.; validation, S.L.; formal analysis, H.Z.; investigation, H.Z.; resources, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z.; supervision, S.L.; project administration, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

VOR	Vestibular ocular reflex
OKR	Objectives and key results
LGN	Lateral geniculate nucleus
RGB	Red–green–blue
FOV	Field of view
OpenCV	Open source computer vision library
CNN	Convolutional neural network
RNN	Recurrent neural network
NCNN	High-performance neural network inference computing framework
R-CNN	Region-Convolutional neural network
Fast R-CNN	Fast region-Convolutional neural network
Mask R-CNN	Mask region-Convolutional neural network
SSD	Single Shot multibox detector algorithm
YOLO	YOLO algorithm
MS	Millisecond
CCD	Charge-coupled device
COMS	Complementary metal-oxide semiconductor
TOF	Time of flight
DC	Direct current
CAN	Controller area network
DOF	Degree of freedom
IMU	Inertial measurement unit
CPU	Central processing unit
GPU	Graphics processing unit
DPU	Deep learning processing unit
NPU	Neural network processing unit
TPU	Tensor processing unit
FPGA	Field-programmable gate array
ASIC	Application-specific integrated circuit

References

- Gorodilov, Y. About the origin of the “Cambrian Explosion” phenomenon and the origin of animal types. *Proc. Zool. Inst. RAS* **2019**, *323*, 1–125. [[CrossRef](#)]
- Charles. *On The Origin Of Species, 1859*; New York University Press: New York, NY, USA, 2010.
- Nilsson, D. 1.07—Eye Evolution in Animals. In *The Senses: A Comprehensive Reference*, 2nd ed.; Fritzsche, B., Ed.; Elsevier: Oxford, UK, 2020; pp. 96–121.
- Nityananda, V.; Read, J. Stereopsis in animals: Evolution, function, and mechanisms. *J. Exp. Biol.* **2017**, *220*, 2502–2512. [[CrossRef](#)] [[PubMed](#)]
- Nilsson, D. The Evolution of Visual Roles—Ancient Vision Versus Object Vision. *Front. Neuroanat.* **2022**, *16*, 789375. [[CrossRef](#)]
- Tan, Y.; Shi, Y.; Tang, Q. Brain Storm Optimization Algorithm. In *Advances in Swarm Intelligence*; Lecture Notes in Computer Science Series; Springer: Berlin/Heidelberg, Germany, 2011.
- Joukal, M. *Anatomy of the Human Visual Pathway*; Springer: Cham, Switzerland, 2017.
- Alipour, H.; Namazi, H.; Azarnoush, H.; Jafari, S. Fractal-based analysis of the influence of color tonality on human eye movements. *Fractals* **2019**, *27*, 403. [[CrossRef](#)]
- Sebastian, E.T. The Complexity and Origins of the Human Eye: A Brief Study on the Anatomy, Physiology, and Origin of the Eye. Senior Honors Thesis, Liberty University, Lynchburg, VA, USA, 2010.
- Fritsch, G.; Hitzig, E. The excitable cerebral cortex. Fritsch, G., Hitzig, E. Über die elektrische Erregbarkeit des Grosshirns. *Arch. Anat. Physiol. Wissen.* **1870**, *37*, 300–332.
- Sabbah, S.; Gemmer, J.A.; Bhatia-Lin, A.; Manoff, G.; Castro, G.; Siegel, J.K.; Jeffery, N.; Berson, D.M. A retinal code for motion along the gravitational and body axes. *Nature* **2017**, *546*, 492–497. [[CrossRef](#)] [[PubMed](#)]
- Berson, D.M. 1.03—The Sensory Organ: Eye, Receptors, Retina. In *The Senses: A Comprehensive Reference*, 2nd ed.; Fritzsche, B., Ed.; Elsevier: Oxford, UK, 2020; pp. 31–35.
- Berson, D. Keep both eyes on the prize: Hunting mice use binocular vision and specialized retinal neurons to capture prey. *Neuron* **2021**, *109*, 1418–1420. [[CrossRef](#)] [[PubMed](#)]
- Carlson, C.; Devinsky, O. The excitable cerebral cortex Fritsch G, Hitzig E. Über die elektrische Erregbarkeit des Grosshirns. *Arch Anat Physiol Wissen 1870;37:300-32. Epilepsy Behav.* **2009**, *15*, 131–132. [[CrossRef](#)]

15. Crawford, B.H.; Ikeda, H. The physics of vision in vertebrates. *Contemp. Phys.* **1971**, *12*, 75–97. [[CrossRef](#)]
16. Elizabeth Barsotti, A.C.A.C. Neural architectures in the light of comparative connectomics. *Curr. Opin. Neurobiol.* **2021**, *71*, 139–149. [[CrossRef](#)]
17. Liu, S.C.; Delbruck, T. Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* **2010**, *20*, 288–295. [[CrossRef](#)]
18. Nilsson, D. The Diversity of Eyes and Vision. *Annu. Rev. Vis. Sci.* **2021**, *7*, 19–41. [[CrossRef](#)]
19. Zhang, X.S.D. Causes and consequences of the Cambrian explosion. *China Earth Sci.* **2014**, *57*, 930–942. [[CrossRef](#)]
20. Zhang, X.; Shu, D. Current understanding on the Cambrian Explosion: Questions and answers. *Paläontologische Z.* **2021**, *95*, 641–660. [[CrossRef](#)]
21. Young, L.R.; Stark, L. Variable Feedback Experiments Testing a Sampled Data Model for Eye Tracking Movements. *IEEE Trans. Hum. Factors Electron.* **1963**, *HFE-4*, 38–51. [[CrossRef](#)]
22. Robinson, D.A. The oculomotor control system: A review. *Proc. IEEE* **1968**, *56*, 1032–1049. [[CrossRef](#)]
23. Robinson, D.A. Oculomotor unit behavior in the monkey. *J. Neurophysiol.* **1970**, *33*, 393. [[CrossRef](#)]
24. Robinson, D.A.; Gordon, J.L.; Gordon, S.E. A model of the smooth pursuit eye movement system. *Biol. Cybern.* **1986**, *55*, 43–57. [[CrossRef](#)]
25. Lisberger, S.G.; Morris, E.J.; Tychsen, J. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annu. Rev. Neurosci.* **1987**, *10*, 97–129. [[CrossRef](#)]
26. Deno, D.C.; Keller, E.L.; Crandall, W.F. Dynamical neural network organization of the visual pursuit system. *IEEE Trans. Biomed. Eng.* **1989**, *36*, 85–92. [[CrossRef](#)]
27. Lungchi, F.; Lazzari, S.; Magenes, G. Neural adaptive predictor for visual tracking system. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, China, 1 November 1998.
28. Gomi, H.; Kawato, M. Adaptive feedback control models of the vestibulocerebellum and spinocerebellum. *Biol. Cybern.* **1992**, *68*, 105–114. [[CrossRef](#)]
29. Scassellati, B. Eye finding via face detection for a foveated active vision system. In Proceedings of the 15th National Conference on Artificial Intelligence AAAI/IAAI, Menlo Park, CA, USA, 26–30 July 1998.
30. de Brouwer, S.; Missal, M.; Barnes, G.; Lefèvre, P. Quantitative Analysis of Catch-Up Saccades During Sustained Pursuit. *J. Neurophysiol.* **2002**, *87*, 1772–1780. [[CrossRef](#)]
31. Merfeld, D.M.; Park, S.; Gianna-Poulin, C.; Black, F.O.; Wood, S. Vestibular Perception and Action Employ Qualitatively Different Mechanisms. I. Frequency Response of VOR and Perceptual Responses During Translation and Tilt. *J. Neurophysiol.* **2005**, *94*, 186–198. [[CrossRef](#)]
32. Zhang, X. An object tracking system based on human neural pathways of binocular motor system. In Proceedings of the 2006 9th International Conference on Control, Automation, Robotics and Vision, Singapore, 5–8 December 2006.
33. Cannata, G.; D’Andrea, M.; Maggiali, M. Design of a Humanoid Robot Eye: Models and Experiments. In Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy, 4–6 December 2006.
34. Wang, Q.; Zou, W.; Zhang, F.; Xu, D. Binocular initial location and extrinsic parameters real-time calculation for bionic eye system. In Proceedings of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 29 June–4 July 2014.
35. Fan, D.; Chen, X.; Zhang, T.; Chen, X.; Liu, G.; Owais, H.M.; Kim, H.; Tian, Y.; Zhang, W.; Huang, Q. Design of anthropomorphic robot bionic eyes. In Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), Macao, China, 5–8 December 2017.
36. Liu, Y.; Zhu, D.; Peng, J.; Wang, X.; Wang, L.; Chen, L.; Li, J.; Zhang, X. Real-Time Robust Stereo Visual SLAM System Based on Bionic Eyes. *IEEE Trans. Med. Robot. Bionics* **2020**, *2*, 391–398. [[CrossRef](#)]
37. Wang, X.; Li, D.; Zhang, G. Panoramic Stereo Imaging of a Bionic Compound-Eye Based on Binocular Vision. *Sensors* **2021**, *21*, 1944. [[CrossRef](#)] [[PubMed](#)]
38. Berthoz, A.; Pavard, B.; Young, L.R. Perception of linear horizontal self-motion induced by peripheral vision (linearvection) basic characteristics and visual-vestibular interactions. *Exp. Brain Res.* **1975**, *23*, 471–489. [[CrossRef](#)] [[PubMed](#)]
39. Zacharias, G.L.; Young, L.R. Influence of combined visual and vestibular cues on human perception and control of horizontal rotation. *Exp. Brain Res.* **1981**, *41*, 159–171. [[CrossRef](#)]
40. Huang, J.; Young, L.R. Sensation of rotation about a vertical axis with a fixed visual field in different illuminations and in the dark. *Exp. Brain Res.* **1981**, *41*, 172–183. [[CrossRef](#)]
41. Lichtenberg, B.K.; Young, L.R.; Arrott, A.P. Human ocular counterrolling induced by varying linear accelerations. *Exp. Brain Res.* **1982**, *48*, 127–136. [[CrossRef](#)]
42. Young, L.R.; Oman, C.M.; Watt, D.G.D.; Money, K.E.; Lichtenberg, B.K.; Kenyon, R.V.; Arrott, A.P. M.I.T./Canadian vestibular experiments on the Spacelab-1 mission: 1. Sensory adaptation to weightlessness and readaptation to one-g: An overview. *Exp. Brain Res.* **1986**, *64*, 291–298. [[CrossRef](#)]
43. Arnold, D.B.; Robinson, D.A. A neural network that learns to integrate oculomotor signals. In Proceedings of the 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 17–21 June 1990.
44. Robinson, D.A.; Fuchs, A.F. Eye movements evoked by stimulation of frontal eye fields. *J. Neurophysiol.* **1969**, *32*, 637–648. [[CrossRef](#)]
45. Robinson, A.D. Real neural networks in movement control. In Proceedings of the 1994 IEEE American Control Conference, Baltimore, MD, USA, 29 June–1 July 1994.
46. Hubel, D.H. Exploration of the primary visual cortex, 1955–1978. *Nature* **1982**, *299*, 515–524. [[CrossRef](#)]

47. Hubel, D.H. Evolution of ideas on the primary visual cortex, 1955–1978: A biased historical account. *Biosci. Rep.* **1982**, *2*, 435–469. [[CrossRef](#)]
48. Yau, J.M.; Pasupathy, A.; Brincat, S.L.; Connor, C.E. Curvature Processing Dynamics in Macaque Area V4. *Cereb. Cortex* **2013**, *23*, 198–209. [[CrossRef](#)]
49. Grill-Spector, K.; Malach, R. The human visual cortex. *Annu. Rev. Neurosci.* **2004**, *27*, 649–677. [[CrossRef](#)]
50. Hari, R.; Kujala, M.V. Brain Basis of Human Social Interaction: From Concepts to Brain Imaging. *Physiol. Rev.* **2009**, *89*, 453–479. [[CrossRef](#)]
51. Hao, Q.; Tao, Y.; Cao, J.; Tang, M.; Cheng, Y.; Zhou, D.; Ning, Y.; Bao, C.; Cui, H. Retina-like Imaging and Its Applications: A Brief Review. *Appl. Sci.* **2021**, *11*, 7058. [[CrossRef](#)]
52. Ayoub, G. On the Design of the Vertebrate Retina. *Orig. Des.* **1996**, *17*, 1.
53. Williams, G.C. *Natural Selection: Domains, Levels, and Challenges*; Oxford University Press: Oxford, UK, 1992; Volume 72.
54. Navarro, R. The Optical Design of the Human Eye: A Critical Review. *J. Optom.* **2009**, *2*, 3–18. [[CrossRef](#)]
55. Schiefer, U.; Hart, W. *Functional Anatomy of the Human Visual Pathway*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 19–28.
56. Horton, J.C.; Hedley-Whyte, E.T. Mapping of cytochrome oxidase patches and ocular dominance columns in human visual cortex. *Philos. Trans. R. Soc. B Biol. Sci.* **1984**, *304*, 255–272.
57. Choi, S.; Jeong, G.; Kim, Y.; Cho, Z. Proposal for human visual pathway in the extrastriate cortex by fiber tracking method using diffusion-weighted MRI. *Neuroimage* **2020**, *220*, 117145. [[CrossRef](#)]
58. Welsh, D.K.; Takahashi, J.S.; Kay, S.A. Suprachiasmatic nucleus: Cell autonomy and network properties. *Annu. Rev. Physiol.* **2010**, *72*, 551–577. [[CrossRef](#)]
59. Ottes, F.P.; Van Gisbergen, J.A.M.; Eggermont, J.J. Visuomotor fields of the superior colliculus: A quantitative model. *Vis. Res.* **1986**, *26*, 857–873. [[CrossRef](#)]
60. Lipari, A. Somatotopic Organization of the Cranial Nerve Nuclei Involved in Eye Movements: III, IV, VI. *Euromediterr. Biomed. J.* **2017**, *12*, 6–9.
61. Deangelis, G.C.; Newsome, W.T. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* **1999**, *19*, 1398–1415. [[CrossRef](#)]
62. Larsson, J.; Heeger, D.J. Two Retinotopic Visual Areas in Human Lateral Occipital Cortex. *J. Neurosci.* **2006**, *26*, 13128–13142. [[CrossRef](#)]
63. Borra, E.; Luppino, G. Comparative anatomy of the macaque and the human frontal oculomotor domain. *Neurosci. Biobehav. Rev.* **2021**, *126*, 43–56. [[CrossRef](#)]
64. França De Barros, F.; Bacqué-Cazenave, J.; Taillebuis, C.; Courtand, G.; Manuel, M.; Bras, H.; Tagliabue, M.; Combes, D.; Lambert, F.M.; Beraneck, M. Conservation of locomotion-induced oculomotor activity through evolution in mammals. *Curr. Biol.* **2022**, *32*, 453–461. [[CrossRef](#)]
65. McLoon, L.K.; Willoughby, C.L.; Andrade, F.H. *Extraocular Muscle Structure and Function*; McLoon, L., Andrade, F., Eds.; Craniofacial Muscles; Springer: New York, NY, USA, 2012.
66. Horn, A.K.E.; Straka, H. Functional Organization of Extraocular Motoneurons and Eye Muscles. *Annu. Rev. Vis. Sci.* **2021**, *7*, 793–825. [[CrossRef](#)]
67. Adler, F.H.; Fliegelman, M. Influence of Fixation on the Visual Acuity. *Arch. Ophthalmol.* **1934**, *12*, 475–483. [[CrossRef](#)]
68. Carter, B.T.; Luke, S.G. Best practices in eye tracking research. *Int. J. Psychophysiol.* **2020**, *155*, 49–62. [[CrossRef](#)] [[PubMed](#)]
69. Cazzato, D.; Leo, M.; Distante, C.; Voos, H. When I Look into Your Eyes: A Survey on Computer Vision Contributions for Human Gaze Estimation and Tracking. *Sensors* **2020**, *20*, 3739. [[CrossRef](#)] [[PubMed](#)]
70. Kreiman, G.; Serre, T. Beyond the feedforward sweep: Feedback computations in the visual cortex. *Ann. N. Y. Acad. Sci.* **2020**, *1464*, 222–241. [[CrossRef](#)] [[PubMed](#)]
71. Müller, T.J. Augenbewegungen und Nystagmus: Grundlagen und klinische Diagnostik. *HNO* **2020**, *68*, 313–323. [[CrossRef](#)]
72. Golomb, J.D.; Mazer, J.A. Visual Remapping. *Annu. Rev. Vis. Sci.* **2021**, *7*, 257–277. [[CrossRef](#)]
73. Tzvi, E.; Koeth, F.; Karabanov, A.N.; Siebner, H.R.; Krämer, U.M. Cerebellar—Premotor cortex interactions underlying visuomotor adaptation. *NeuroImage* **2020**, *220*, 117142. [[CrossRef](#)]
74. Banks, M.S.; Read, J.C.; Allison, R.S.; Watt, S.J. Stereoscopy and the Human Visual System. *SMPTE Motion Imaging J.* **2012**, *121*, 24–43. [[CrossRef](#)]
75. Einspruch, N. (Ed.) *Application Specific Integrated Circuit (ASIC) Technology*; Academic Press: Cambridge, MA, USA, 2012; Volume 23.
76. Verri Lucca, A.; Mariano Sborz, G.A.; Leithardt, V.R.Q.; Beko, M.; Albenes Zeferino, C.; Parreira, W.D. A Review of Techniques for Implementing Elliptic Curve Point Multiplication on Hardware. *J. Sens. Actuator Netw.* **2021**, *10*, 3. [[CrossRef](#)]
77. Zhao, C.; Yan, Y.; Li, W. Analytical Evaluation of VCO-ADC Quantization Noise Spectrum Using Pulse Frequency Modulation. *IEEE Signal Proc. Lett.* **2014**, *11*, 249–253.
78. Jouppi, N.P.; Young, C.; Patil, N.; Patterson, D.; Agrawal, G.; Bajwa, R.; Bates, S.; Bhatia, S.; Boden, N.; Borchers, A.; et al. In-datacenter performance analysis of a tensor processing unit. In Proceedings of the ISCA '17: Proceedings of the 44th Annual International Symposium on Computer Architecture, Toronto, ON, Canada, 24–28 June 2017.
79. Huang, H.; Liu, Y.; Hou, Y.; Chen, R.C.-J.; Lee, C.; Chao, Y.; Hsu, P.; Chen, C.; Guo, W.; Yang, W.; et al. 45nm High-k/Metal-Gate CMOS Technology for GPU/NPU Applications with Highest PFET Performance. In Proceedings of the 2007 IEEE International Electron Devices Meeting, Washington, DC, USA, 10–12 December 2007.

80. Kim, S.; Oh, S.; Yi, Y. Minimizing GPU Kernel Launch Overhead in Deep Learning Inference on Mobile GPUs. In Proceedings of the HotMobile '21: The 22nd International Workshop on Mobile Computing Systems and Applications, Virtual, 24–26 February 2021.
81. Shah, N.; Olascoaga, L.I.G.; Zhao, S.; Meert, W.; Verhelst, M. DPU: DAG Processing Unit for Irregular Graphs With Precision-Scalable Posit Arithmetic in 28 nm. *IEEE J. Solid-state Circuits* **2022**, *57*, 2586–2596. [[CrossRef](#)]
82. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
83. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
84. Ren, J.; Wang, Y. Overview of Object Detection Algorithms Using Convolutional Neural Networks. *J. Comput. Commun.* **2022**, *10*, 115–132.
85. Kreiman, G. *Biological and Computer Vision*; Cambridge University Press: Oxford, UK, 2021.
86. Sabour, S.; Frosst, N.; Geoffrey, E. Hinton Dynamic routing between capsules. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Curran Associates Inc., Red Hook, NY, USA, 4–9 December 2017; pp. 3859–3869.
87. Schwarz, M.; Behnke, S. Stillleben: Realistic Scene Synthesis for Deep Learning in Robotics. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 10502–10508, ISSN: 2577-087X.
88. Piga, N.A.; Bottarel, F.; Fantacci, C.; Vezzani, G.; Pattacini, U.; Natale, L. MaskUKF: An Instance Segmentation Aided Unscented Kalman Filter for 6D Object Pose and Velocity Tracking. *Front. Robot. AI* **2021**, *8*, 594593. [[CrossRef](#)]
89. Bottarel, F.; Vezzani, G.; Pattacini, U.; Natale, L. GRASPA 1.0: GRASPA is a Robot Arm grasping Performance Benchmark. *IEEE Robot. Autom. Lett.* **2020**, *5*, 836–843. [[CrossRef](#)]
90. Bottarel, F. *Where's My Mesh? An Exploratory Study on Model-Free Grasp Planning*; University of Genova: Genova, Italy, 2021.
91. Dog-qiuqiu. MobileNet-YOLO That Works Better Than SSD. GitHub. 2021. Available online: <https://github.com/dog-qiuqiu/MobileNet-Yolo> (accessed on 6 February 2021).
92. Yuan, L.; Chen, D.; Chen, Y.L.; Codella, N.; Dai, X.; Gao, J.; Hu, H.; Huang, X.; Li, B.; Li, C.; et al. Florence: A New Foundation Model for Computer Vision. *arXiv* **2021**, arXiv:2111.11432.
93. Yamins, D.L.K.; Dicarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **2016**, *19*, 356–365. [[CrossRef](#)]
94. Breazeal, C. Socially intelligent robots: Research, development, and applications. In Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace, Tucson, AZ, USA, 7–10 October 2001.
95. Shaw-Garlock, G. Looking Forward to Sociable Robots. *Int. J. Soc. Robot.* **2009**, *1*, 249–260. [[CrossRef](#)]
96. Gokturk, S.B.; Yalcin, H.; Bamji, C. A Time-Of-Flight Depth Sensor—System Description, Issues and Solutions. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004.
97. Gibaldi, A.; Canessa, A.; Chessa, M.; Sabatini, S.P.; Solari, F. A neuromorphic control module for real-time vergence eye movements on the iCub robot head. In Proceedings of the 2011 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia, 26–28 October 2011.
98. Xiaolin, Z. A Novel Methodology for High Accuracy Fixational Eye Movements Detection. In Proceedings of the 4th International Conference on Bioinformatics and Biomedical Technology, Singapore, 26–28 February 2012.
99. Song, Y.; Xiaolin, Z. An Integrated System for Basic Eye Movements. *J. Inst. Image Inf. Telev. Eng.* **2012**, *66*, J453–J460.
100. Dorrington, A.A.; Kelly, C.D.B.; McClure, S.H.; Payne, A.D.; Cree, M.J. Advantages of 3D time-of-flight range imaging cameras in machine vision applications. In Proceedings of the 16th Electronics New Zealand Conference (ENZCon), Dunedin, New Zealand, 18–20 November 2009.
101. Jain, A.K.; Dorai, C. Practicing vision: Integration, evaluation and applications. *Pattern Recogn.* **1997**, *30*, 183–196. [[CrossRef](#)]
102. Ma, Z.; Ling, H.; Song, Y.; Hospedales, T.; Jia, W.; Peng, Y.; Han, A. IEEE Access Special Section Editorial: Recent Advantages of Computer Vision. *IEEE Access* **2018**, *6*, 31481–31485.
103. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. Events-to-Video: Bringing Modern Computer Vision to Event Cameras. *arXiv* **2019**, arXiv:1904.08298.
104. Silva, A.E.; Chubb, C. The 3-dimensional, 4-channel model of human visual sensitivity to grayscale scrambles. *Vis. Res.* **2014**, *101*, 94–107. [[CrossRef](#)] [[PubMed](#)]
105. Potter, M.C.; Wyble, B.; Haggmann, C.E.; Mccourt, E.S. Detecting meaning in RSVP at 13 ms per picture. *Atten. Percept. Psychophys.* **2013**, *76*, 270–279. [[CrossRef](#)] [[PubMed](#)]
106. Al-Rahayfeh, A.; Faezipour, M. Enhanced frame rate for real-time eye tracking using circular hough transform. In Proceedings of the 2013 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, USA, 3 May 2013.
107. Wilder, K. *Photography and Science*; Reaktion Books: London, UK, 2009.
108. Smith, G.E. The invention and early history of the CCD. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2009**, *607*, 1–6. [[CrossRef](#)]
109. Marcandali, S.; Marar, J.F.; de Oliveira Silva, E. Através da Imagem: A Evolução da Fotografia e a Democratização Profissional com a Ascensão Tecnológica. In *Perspectivas Imagéticas*; Ria Editorial: Aveiro, Portugal, 2019.
110. Kucera, T.E.; Barret, R.H. A History of Camera Trapping. In *Camera Traps in Animal Ecology*; Springer: Tokyo, Japan, 2011.
111. Boyle, W.S. CCD—An extension of man's view. *Rev. Mod. Phys.* **2010**, *85*, 2305. [[CrossRef](#)]

112. Sabel, A.B.; Flammer, J.; Merabet, L.B. Residual Vision Activation and the Brain-eye-vascular Triad: Dysregulation, Plasticity and Restoration in Low Vision and Blindness—A Review. *Restor. Neurol. Neurosci.* **2018**, *36*, 767–791. [[CrossRef](#)]
113. Rosenfeld, J.V.; Wong, Y.T.; Yan, E.; Szlawski, J.; Mohan, A.; Clark, J.C.; Rosa, M.; Lowery, A. Tissue response to a chronically implantable wireless intracortical visual prosthesis (Gennaris array). *J. Neural Eng.* **2020**, *17*, 46001. [[CrossRef](#)]
114. Gu, L.; Poddar, S.; Lin, Y.; Long, Z.; Zhang, D.; Zhang, Q.; Shu, L.; Qiu, X.; Kam, M.; Javey, A.; et al. A biomimetic eye with a hemispherical perovskite nanowire array retina. *Nature* **2020**, *581*, 278–282. [[CrossRef](#)]
115. Gu, L.; Poddar, S.; Lin, Y.; Long, Z.; Zhang, D.; Zhang, Q.; Shu, L.; Qiu, X.; Kam, M.; Fan, Z. Bionic Eye with Perovskite Nanowire Array Retina. In Proceedings of the 2021 5th IEEE Electron Devices Technology & Manufacturing Conference (EDTM), Chengdu, China, 8–11 April 2021.
116. Cannata, G.; Maggiali, M. Models for the Design of a Tendon Driven Robot Eye. In Proceedings of the IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; pp. 10–14.
117. Hirai, K. Current and future perspective of Honda humanoid robot. In Proceedings of the Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems IROS '97, Grenoble, France, 11 September 1997.
118. Goswami, A.V.P. *ASIMO and Humanoid Robot Research at Honda*; Springer: Berlin/Heidelberg, Germany, 2007.
119. Kajita, S.; Kaneko, K.; Kaneiro, F.; Harada, K.; Morisawa, M.; Nakaoka, S.I.; Miura, K.; Fujiwara, K.; Neo, E.S.; Hara, I.; et al. *Cybernetic Human HRP-4C: A Humanoid Robot with Human-Like Proportions*; Springer: Berlin, Heidelberg, 2011; pp. 301–314.
120. Faraji, S.; Pouya, S.; Atkeson, C.G.; Ijspeert, A.J. Versatile and robust 3D walking with a simulated humanoid robot (Atlas): A model predictive control approach. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014.
121. Kaehler, A.; Bradski, G. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016.
122. Karve, P.; Thorat, S.; Mistary, P.; Belote, O. Conversational Image Captioning Using LSTM and YOLO for Visually Impaired. In *Proceedings of 3rd International Conference on Communication, Computing and Electronics Systems*; Springer: Berlin/Heidelberg, Germany, 2022.
123. Kornblith, S.; Shlens, J.; Quoc, L.V. Do better imagenet models transfer better? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
124. Jacobstein, N. NASA's Perseverance: Robot laboratory on Mars. *Sci. Robot.* **2021**, *6*, eabh3167. [[CrossRef](#)]
125. Zou, Y.; Zhu, Y.; Bai, Y.; Wang, L.; Jia, Y.; Shen, W.; Fan, Y.; Liu, Y.; Wang, C.; Zhang, A.; et al. Scientific objectives and payloads of Tianwen-1, China's first Mars exploration mission. *Adv. Space Res.* **2021**, *67*, 812–823. [[CrossRef](#)]
126. Jung, B.; Sukhatme, G.S. Real-time motion tracking from a mobile robot. *Int. J. Soc. Robot.* **2010**, *2*, 63–78. [[CrossRef](#)]
127. Brown, J.; Hughes, C.; DeBrunner, L. Real-time hardware design for improving laser detection and ranging accuracy. In Proceedings of the Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 4–7 November 2012; pp. 1115–1119.
128. Min Seok Oh, H.J.K.T. Development and analysis of a photon-counting three-dimensional imaging laser detection and ranging (LADAR) system. *Opt. Soc. Am.* **2011**, *28*, 759–765.
129. Rezaei, M.; Yazdani, M.; Jafari, M.; Saadati, M. Gender differences in the use of ADAS technologies: A systematic review. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *78*, 1–15. [[CrossRef](#)]
130. Maybank, S.J.; Faugeras, O.D. A theory of self-calibration of a moving camera. *Int. J. Comput. Vis.* **1992**, *8*, 123–151. [[CrossRef](#)]
131. Murphy-Chutorian, E.; Trivedi, M.M. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans. Pattern Anal.* **2009**, *31*, 607–626. [[CrossRef](#)]
132. Jarvis, R.A. A Perspective on Range Finding Techniques for Computer Vision. *IEEE Trans. Pattern Anal.* **1983**, *PAMI-5*, 122–139. [[CrossRef](#)]
133. Grosso, E. *On Perceptual Advantages of Eye-Head Active Control*; Springer: Berlin, Heidelberg, 2005; pp. 123–128.
134. Binh Do, P.N.; Chi Nguyen, Q. A Review of Stereo-Photogrammetry Method for 3-D Reconstruction in Computer Vision. In Proceedings of the 19th International Symposium on Communications and Information Technologies (ISCIT), Ho Chi Minh City, Vietnam, 25–27 September 2019.
135. Mottocchia, S. *Stereo Vision Algorithms Suited to Constrained FPGA Cameras*; Springer International Publishing: Cham, Switzerland, 2014; pp. 109–134.
136. Mottocchia, S. Stereo Vision Algorithms for FPGAs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013.
137. Park, J.; Kim, H.; Tai, Y.W.; Brown, M.S.; Kweon, I. High quality depth map upsampling for 3D-TOF cameras. *Int. Conf. Comput. Vis.* **2011**, 1623–1630.
138. Foix, S.; Alenya, G.; Torras, C. Lock-in Time-of-Flight (ToF) Cameras: A Survey. *IEEE Sensors J.* **2011**, *11*, 1917–1926. [[CrossRef](#)]
139. Li, J.; Yu, L.; Wang, J.; Yan, M. Obstacle information detection based on fusion of 3D LADAR and camera; Technical Committee on Control Theory, CAA. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017.
140. Gill, T.; Keller, J.M.; Anderson, D.T.; Luke, R.H. A system for change detection and human recognition in voxel space using the Microsoft Kinect sensor. In Proceedings of the 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 11–13 October 2011.
141. Atmeh, G.M.; Ranatunga, I.; Popa, D.O.; Subbarao, K.; Lewis, F.; Rowe, P. *Implementation of an Adaptive, Model Free, Learning Controller on the Atlas Robot*; American Automatic Control Council: New York, NY, USA, 2014.

142. Nilsson, N.J. *Shakey the Robot*; Tech. Rep. TR223; SRI Int.: Menlo Park, CA, USA, 1984.
143. Burnham, T.C.; Hare, B. Engineering Human Cooperation. *Hum. Nat.* **2007**, *18*, 88–108. [[CrossRef](#)]
144. Liu, Y.; Zhu, D.; Peng, J.; Wang, X.; Wang, L.; Chen, L.; Li, J.; Zhang, X. Robust Active Visual SLAM System Based on Bionic Eyes. In Proceedings of the 2019 IEEE International Conference on Cyborg and Bionic Systems (CBS), Munich, Germany, 18–20 September 2019.
145. Li, B.; Xiaolin, Z.; Sato, M. Pitch angle estimation using a Vehicle-Mounted monocular camera for range measurement. In Proceedings of the International Conference on Signal Processing, Hangzhou, China, 19–23 October 2014.
146. Xiaolin, Z. Novel Human Fixational Eye Movements Detection using Sclera Images of the Eyeball. *Jpn. J. Appl. Physiol.* **2012**, *42*, 143–152.
147. Xiaolin, Z. Wide Area Tracking System Using Three Zoom Cameras. Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 2011.
148. Xiaolin, Z. A Binocular Camera System for Wide Area Surveillance. *J. Inst. Image Inf. Telev. Eng.* **2009**. [[CrossRef](#)]
149. Xiaolin, Z. A Mathematical Model of a Neuron with Synapses based on Physiology. *Nat. Preced.* **2008**. [[CrossRef](#)]
150. Xiaolin, Z. Cooperative Movements of Binocular Motor System. In Proceedings of the 2008 IEEE International Conference on Automation Science and Engineering, Arlington, VA, USA, 23–26 August 2008.
151. Zhang, X.K.H.A. Image Segmentation through Region Fusion Based on Watershed. *J. Comput. Inf. Syst.* **2014**, *19*, 8231–8236.
152. Wang, Q.; Yin, Y.; Zou, W.; Xu, D. Measurement error analysis of binocular stereo vision: Effective guidelines for bionic eyes. *IET Sci. Meas. Technol.* **2017**, *11*, 829–838. [[CrossRef](#)]
153. Wang, Q.; Zou, W.; Xu, D.; Zhu, Z. Motion Control in Saccade and Smooth Pursuit for Bionic Eye Based on Three-dimensional Coordinates. *J. Bionic Eng.* **2017**, *14*, 336–347. [[CrossRef](#)]
154. Wang, Q.; Zou, W.; Xu, D.; Zhang, F. 3D Perception of Biomimetic Eye Based on Motion Vision and Stereo Vision. *Robot* **2015**, *37*, 760–768.
155. Zheng Zhu, W.Z.Q.W. A Velocity Compensation Visual Servo Method for Oculomotor Control Of Bionic Eyes. *Int. J. Robot. Autom.* **2018**, *33*, 33–44.
156. Zhu, Z.; Wang, Q.; Zou, W.; Zhang, F. Motion Control on Bionic Eyes: A Comprehensive Review. *arXiv* **2019**, arXiv:1901.01494.
157. Chen, X.; Wang, C.; Zhang, T.; Hua, C.; Fu, S.; Huang, Q. Hybrid Image Stabilization of Robotic Bionic Eyes. In Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO), Kuala Lumpur, Malaysia, 12–15 December 2018.
158. Kardamakis, A.A.; Grantyn, A.; Moschovakis, A.K. Neural network simulations of the primate oculomotor system. v. eye–head gaze shifts. *Biol. Cybern.* **2010**, *102*, 209–225. [[CrossRef](#)] [[PubMed](#)]
159. Aragon-Camarasa, G.; Fattah, H.; Siebert, J.P. Towards a unified visual framework in a binocular active robot vision system. *Robot. Auton. Syst.* **2010**, *58*, 276–286. [[CrossRef](#)]