







Review

A Comprehensive Review of Vision-Based Robotic Applications: Current State, Components, Approaches, Barriers, and Potential Solutions

Md Tanzil Shahria ^{1,*}, Md Samiul Haque Sunny ¹, Md Ishrak Islam Zarif ², Jawhar Ghommam ³,
Sheikh Iqbal Ahamed ² and Mohammad H Rahman ^{1,4}

¹ Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

² Computer Science, Marquette University, Milwaukee, WI 53233, USA

³ Electrical and Computer Engineering, Sultan Qaboos University, Al Seeb, Muscat 123, Oman

⁴ Mechanical Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

* Correspondence: mshahria@uwm.edu

† Current address: Biorobotics Lab, University of Wisconsin-Milwaukee, 115 East Reindl Way, USR 281, Milwaukee, WI 53212, USA.

Abstract: Being an emerging technology, robotic manipulation has encountered tremendous advancements due to technological developments starting from using sensors to artificial intelligence. Over the decades, robotic manipulation has advanced in terms of the versatility and flexibility of mobile robot platforms. Thus, robots are now capable of interacting with the world around them. To interact with the real world, robots require various sensory inputs from their surroundings, and the use of vision is rapidly increasing nowadays, as vision is unquestionably a rich source of information for a robotic system. In recent years, robotic manipulators have made significant progress towards achieving human-like abilities. There is still a large gap between human and robot dexterity, especially when it comes to executing complex and long-lasting manipulations. This paper comprehensively investigates the state-of-the-art development of vision-based robotic application, which includes the current state, components, and approaches used along with the algorithms with respect to the control and application of robots. Furthermore, a comprehensive analysis of those vision-based applied algorithms, their effectiveness, and their complexity has been enlightened here. To conclude, there is a discussion over the constraints while performing the research and potential solutions to develop a robust and accurate vision-based robot manipulation.

Keywords: computer vision; machine learning; robot manipulation; sensors; vision-based control



Citation: Shahria, M.T.; Sunny, M.S.H.; Zarif, M.I.I.; Ghommam, J.; Ahamed, S.I.; Rahman, M.H. A Comprehensive Review of Vision-Based Robotic Applications: Current State, Components, Approaches, Barriers, and Potential Solutions. *Robotics* **2022**, *11*, 139. <https://doi.org/10.3390/robotics11060139>

Academic Editor: Connie Man-Ching YUEN

Received: 3 November 2022

Accepted: 29 November 2022

Published: 2 December 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

Robotic manipulation alludes to the manner in which robots directly and indirectly interact with surrounding objects. Such interaction includes picking and grasping objects [1–3], moving objects from place to place [4,5], folding laundry [6], packing boxes [7], operating as per user requirement, etc. Object manipulation is considered the pivotal role of robotics. Over time, robot manipulation has encountered considerable changes that cause technological development in both industry and academia.

Manual robot manipulation was one of the initial steps of automation [8,9]. A manual robot refers to a manipulation system that requires continuous human involvement to operate [10]. In the beginning, spatial algebra [11], forward kinematics [12–14], differential kinematics [15–17], inverse kinematics [18–22], etc. were explored by researchers for pick and place tasks, which is not the only application of robotic manipulation systems but the stepping-stone for a wide range of possibilities [23]. The capability of gripping, holding, and manipulating objects requires dexterity, perception of touch, and response from eyes and muscles; mimicking all these attributes is a complex and tedious task [24]. Thus, researchers have explored a wide range of algorithms to adopt and design more efficient and

appropriate models for this task. Through time, manual manipulators got advanced and had individual control systems according to their specification and application [25,26].

Along with the individual use of robotic manipulation systems, it has a wide range of industrial applications nowadays as it can be applied to complex and diverse tasks [27]. Hence, typical manipulative devices have become less suited in these times [28]. Different kinds of new technologies, such as wireless communication, augmented reality [29], etc., are being adopted and applied in manipulation systems to uncover the most suitable and friendly human–robot collaboration model for specific tasks [30]. To make the process more efficient and productive and to obtain successful execution, researchers have introduced automation in this field [31].

To habituate to the automated system, researchers first introduced automation in the motion planning technique [3,32], which eventually contributed to the automated robotic manipulation system. Automated and semi-automated manipulation systems not only boost the performance of industrial robots but also contribute to other fields of robotics such as mobile robots [33], assistive robots [34], swarm robots [35], etc. While designing the automated system, the utilization of vision is increasing rapidly as vision is undoubtedly a loaded source of information [36–38]. By properly utilizing vision-based data, a robot can identify, map, localize, and calculate various measurements of any object and respond accordingly to complete its tasks [39–42]. Various studies confirm that vision-based approaches are more appropriate in different fields of robotics such as swarm robotics [35], fruit-picking robots [1], robotic grasping [43], mobile robots [33,44,45], aerial robotics [46], surgical robots [47], etc. To process the vision-based data, different approaches are being introduced by the researchers. However, learning-based approaches are at the center of such autonomous approaches, as in the real world, there are too many deviations and learning algorithms that help the robot gain knowledge from its experience with the environment [48–50]. Among different learning methods, various neural network-based models [51–54], deep learning-based models [49,50,54–56], and transfer learning models [57–60] are mostly exercised by the experts of manipulation systems, whereas different filter-based approaches are also popular among researchers [61–63].

This paper presents some recent notable works on robotic manipulation systems, specifically focused on vision-based approaches. Moreover, the current state, the issues researchers addressed throughout the experiments, their approaches, and the proper applications of such models have also been analyzed here. Researchers use a variety of control tactics to manipulate robots, but this study will focus exclusively on vision-based decision making in robotic applications. The control techniques for manipulating robots are beyond the scope of this study. The primary contributions of this study are four-fold:

- Presenting the current state of the vision-based robotic system with a chronological progression until now.
- Reviewing algorithmic highlights of various approaches, including used components and applied vision-based control theory. We scrutinize all the proposed methods and identify the most adopted ones in this field.
- Generalizing the focused application. We review all the approaches and narrow down the essential applications.
- Summarizing the barriers. We sum up all the mentioned studies and present the barriers as well as potential solutions here.

The rest of the paper is structured as follows: Section 2 briefly represents the inclusion and exclusion criteria of the studies, Section 3 discusses the current state to date, Section 4 presents some brief information about the components used in these studies, Section 5 summarizes the experimental environments, Section 6 represents the control theories used in the selected publications, Section 7 discusses the focus applications, Section 8 discloses the challenges and potential solutions for the vision-based approaches in the robotic manipulation systems, Section 9 contains discussion, and finally, Section 10 portrays the conclusion.

2. Journal Selection

Studies were chosen by performing a systematic electronic search of a handful of databases as of 1 August 2022, by the authors who specialized in computer vision and eye-gaze control areas. The timeline of the studies was limited to the last seven years to focus on the recent advancement in this field. The search was performed in the following databases; IEEE Xplore, Elsevier B.V., arXiv, Springer, Hindawi, MDPI, and Wiley. While searching, the following keywords were used; “vision-based robotic manipulation”, “vision-based telerobot review”, “computer vision”, “vision-based surgical robots”, and “vision for robust robot manipulation”. Figure 1 illustrates the overall inclusion and exclusion criteria of the selected studies.

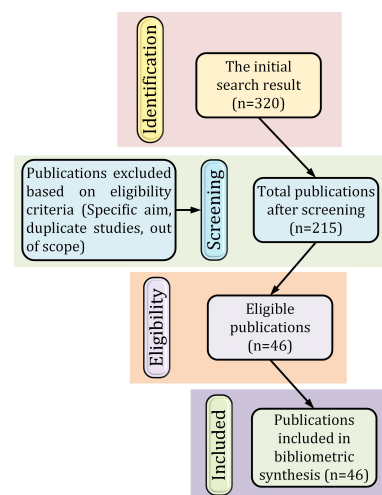


Figure 1. Inclusion and exclusion criteria of the selected studies.

This search identified approximately 320 relevant publications for consideration. However, after initial screening (keyword mismatch, out-of-scope papers resulting during keyword search), 215 studies were shortlisted for review. After that, some of the publications were excluded based on selection criteria (specific aim, duplicate studies, review, etc.), and 46 studies were narrowed down (aligned with manuscript scope) for the in-depth review. The authors went through all the relevant sections of the studies, including the abstract, introduction, methodology, experiments, conclusion, and future work sections of all the selected papers ($n = 46$) to identify the other significant information such as the addressed problem, contribution, approach, control theory, experiment setup, complexity, and communication protocol.

3. Current State

A common structural assumption for manipulative tasks for a robot is that an object or set of objects in the environment is what the robot is trying to manipulate. Because of this, generalization via objects—both across different objects and between similar (or identical) objects in different task instances—is an important aspect of learning to manipulate.

Commonly used object-centric manipulation skills and task model representations are often sufficient to generalize across tasks and objects, but adapting to differences in shape, properties, and appearance is required. A wide range of robotic manipulation problems can be solved using the vision-based approach as it works as a better sensory source for the system. Because of that and the availability of a fast processing power, vision-based approaches have become very popular among researchers who are working on robotic manipulation-based problems. A chronological observation depicting the contributions of the researchers based on the addressed problems and their outcomes is compiled in Table 1.

Table 1. Chronological progression of the vision-based approach.

Year	Addressed Problems	Contributions	Outcomes
2016 [64–66]	Manipulation [64] and grasping control strategies [66] using eye-tracking and sensory-motor fusion [65].	Object detection, path planning, and navigation [64]; Control of an endoscopic manipulator [65]; Sensory-motor fusion-based manipulation and grasping control strategy for a robotic hand–eye system [66].	The proposed approach has improved performance in calibration, task completion, and navigation [64]; Shows better a performance than endoscope manipulation by an assistant [64]; Demonstrates responsiveness and flexibility [66].
2017 [67–73]	Following human user with robotic blimp [70]; Deformable object manipulation [67]; Tracking and navigation for aerial vehicles [68,73]; Object detection without GPU support [69]; Automated object recognition for assistive robots [71]; Path-finding for humanoid robot [72].	Robotic rope manipulation using vision-based learning model [67]; Robust vision-based tracking system for a UAV [68]; Real-time robotic object detection and recognition model [69]; Behavioral stability in humanoid robots and path-finding algorithms [72]; Robust real-time navigation [70] and long-range object tracking system [70,73].	Robot successfully manipulates a rope [67]; System achieves robust tracking in real-time [68] and proved to be efficient in object detection [69]; Robotic blimp can follow humans [70]; System was able to detect and recognize objects [71]; Algorithm successfully able to find a path to guide the robot [72]; System arrived at an operational stage for lighting and weather conditions [73].
2018 [74–81]	Real-time mobile robot controller [74]; Target detection for safe UAV landing [75]; Vision-based grasping [76], object sorting [79], and dynamic manipulation [77]; Multi-task learning [78]; Learn complex skills from raw sensory inputs [80]; Autonomous landing of a quadrotor on moving targets [81].	Sensor-independent controller for real-time mobile robots [74]; Detection and landing system for drones [75]; GDRL-based grasping benchmark [76]; Effective robotic framework for extensible RL [77]; Complete controller for generating robot arm trajectories [78]; Successfully inaugurate a camera-robot system [79]; Successful framework to learn a deep dynamics model on images [80]; Autonomous NN-based landing controller of UAVs on moving targets in search and secure applications [81].	The mobile robot reaches its goal [74]; The system finds targets and lands safely [75]; System grasps better than other algorithms [76]; Real-world reinforcement learning can handle large datasets and models [77]; Method is a versatile manipulator that can accurately correct errors [78]; Placement of objects by the robot gripper [79]; Generalization to a wide range of tasks [80]; Successful autonomous quadrotor landing on fixed and moving platforms [81].
2019 [82–89]	Nonlinear approximation for mobile robots [82]; Control of cable-driven robots [83]; Leader–follower formation control [84]; Motion control for a free-floating robot [85]; Control of soft robots [86]; Approach an object when obstacles are present [87]; Needle-based percutaneous using robotic technologies [88]; Natural interaction control of surgical robots [89].	Effective recurrent neural network-based controller for robots [82]; Robust method for analyzing the stability of the cable-driven robots [83]; Effective formation control for a multi-agent system [84]; Efficient vision-based system for a free-floating robot [85]; Stable framework for soft robots [86]; Useful system to increase the autonomy of people with upper-body disabilities [87]; Accurate system to identify the needle position and orientation [88]; Smooth model to use eye movements to control a robot [89].	System outperforms existing ones [82]; Vision-based control is a good alternative to model-based control [83]; Control protocol completes formation tasks with visibility constraints [84]; Method eliminates false targets and improves positioning precision [85]; System maintained an acceptable accuracy and stability [86]; A person successfully controlled the robotic arm using the system [87]; Framework shows the proposed robotic hardware’s efficiency [88]; movement was feasible and convenient [89].
2020 [90–95]	Grasping under occlusion [90]; Recognition and manipulation of objects [91]; Controllers for decentralized robot swarms [92]; Robot manipulation via human demonstration [93]; Robot manipulator using Iris tracking [94]; Object tracking of visual servoing [95].	Robust grasping method for a robotic system [90]; Effective stereo algorithm for manipulation of objects [91]; Successful framework to control decentralized robot swarms [92]; Generalized framework for activity recognition from human demonstrations [93]; Real-time iris tracking method for the ophthalmic robotic system [94]; Successful method for conventional template matching [95].	Method’s effectiveness validated through experiments [90]; R-CNN method is very stable [91]; Architecture shows promising performance for large-sized swarms [92]; Proposed approach achieves good generalized performance [93]; Tracker is suitable for the ophthalmic robotic system [94]; Control system demonstrates significant improvement to feature tracking and robot motion [95].

Table 1. Cont.

Year	Addressed Problems	Contributions	Outcomes
2021 [96–102]	Human–robot handover applications [96]; Imitation learning for robotic manipulation [97]; Reaching and grasping objects using a robotic arm [98]; Integration of libraries for real-time computer vision [99]; Mobility and key challenges for various construction applications [100]; Obtaining the spatial information of operated target [101]; Training actor–critic methods is RL [102].	Efficient human–robot hand-over control strategy [96]; Intelligent vision-guided imitation learning framework for robotic exactitude manipulation [97]; Robotic hand–eye coordination system to achieve robust reaching ability [98]; Upgraded vision of a real-time computer vision system [99]; Mobile robotic system for object manipulation using autonomous navigation and object grasping [100]; Calibration-free monocular vision-based robot manipulation [101]; Attention-driven robot manipulation for discretization of the translation space [102]	Control shows promising and effective results [96]; Object can reach the goal positions smoothly and intelligently using the framework [97]; Dual neural-network-based controller leads to higher success rate and better control performance [98]; Successfully implemented and tested on the latest technologies [99]; UGV autonomously navigates toward a selected location [100]; Performance of the method has been successfully evaluated [101]; Algorithm achieves state-of-the-art performance on several difficult robotics tasks [102].
2022 [103–109]	Micro-manipulation on cells [103]; Collision-free navigation [104]; Highly nonlinear continuum manipulation [105]; Complexity of RL in broad range of robotic manipulation task [106]; Uncertainty in DNN-based prediction for robotic grasping [107]; Path planning for a robotic arm in a 3D workspace [108]; Object tracking and control of a robotic arm in real-time [109].	Path planning for magnetic micro-robots [103]; Neural radiance fields (NeRFs) for navigation in 3D environment [104]; Aerial continuum manipulation systems (ACMSs) [105]; Attention-driven robotic manipulation [106]; Robotic grabbing in distorted RGB-D data [107]; Real-time path generation with lower computational cost [108]; Real-time object tracking with reduced stress load and a high rate of success. [109].	Magnetic micro-robots performed accurately in complex environment [103]; NeRFs outperforms the dynamically informed INeRF baseline [104]; simulation demonstrates good results [105]; ARM was successful on a range of RL Bench tasks [106]; System performs better than end-to-end networks in difficult conditions [107]; System significantly eased the limitations of prior research [108]; System effectively locates the robotic arm in the desired location with very high accuracy [109].

Figure 2 represents the basic categorization of the problems addressed by the researchers. The problems are primarily divided into two categories: control-based problems and application-based problems. Each of these problems is further categorized into several sub-categories. While dealing with control-based problems such as human demonstration-based control [78,93,97], vision (raw images)-based control [74,82,83,85,86], multi-agent system control [84,92,100,105], etc., researchers have tried and succeeded to solve them by adopting vision-based approaches. The addressed control-based problems are designing a vision-based real-time mobile robot controller [74], multi-task learning from demonstration [78,102,106], nonlinear approximation in the control and monitoring of mobile robots [82], control of cable-driven robots [83], leader–follower formation control [84], motion control for a free-floating robot [85], control of soft robots [86], controllers for decentralized robot swarms [92], robot manipulation via human demonstrations [93], and imitation learning for robotic manipulation [97].

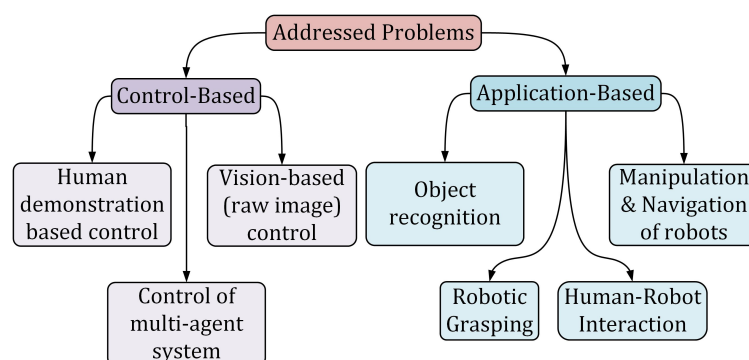


Figure 2. Categorization of problems addressed by the researchers.

Similarly, while solving application-based problems such as object recognition and manipulation [67,69–71,77,79,80,91,101], navigation of robots [68,72,73,75,99,104,108,109], robotic grasping [76,90,107], human–robot interaction [96], etc., researchers successfully applied vision-based approaches and obtained very promising results. The addressed application-based problems are the manipulations of the deformable objects, such as ropes [67], a vision-based tracking system for aerial vehicles [68], object detection without a graphics processing unit (GPU) support for robotic applications [69], detecting and following the human user with a robotic blimp [70], object detection and recognition for autonomous assistive robots [71], path-finding for a humanoid robot [72] or robotic arms [108], navigation of an unmanned surface vehicle [73], vision-based target detection for the safe landing of UAV in both fixed [75] and moving platforms [81], vision-based grasping for robots [76], vision-based dynamic manipulation [77], vision-based object sorting robot manipulator [79], learning complex robotic skills from raw sensory inputs [80], grasping under occlusion for manipulating a robotic system [90], recognition and manipulation of objects [91], human–robot handover applications [96], targeted drug delivery in biological research [103], uncertainty in DNN-based robotic grasping [107], and object tracking via a robotic arm in a real-time 3D environment [109].

4. Components

The main component of any vision-based system is the sensory input devices and the primary sensory input source of vision-based manipulation systems is cameras to perceive the 3D physical world. Researchers used a variety of cameras during their research and tested the system’s performance accordingly.

While dealing with object detection and robot manipulation tasks, most researchers use basic RGB cameras [69,70,72,74–83,85,88,89,106,108] and applied their proposed model to design different detection and control systems. However, for the systems that interact with humans for example, human–robot handover applications [96], wheelchair navigation [64], control of autonomous assistive robots [71], robot manipulation learning via hu-

man demonstrations [93], researchers preferred a depth camera for getting 3D information about the surrounding. Among different commercially available depth cameras, Microsoft Kinect [64,67,90,96,98], RGB depth camera [71,84,87,107], and RealSense [93,99,102] are mostly used and utilized by the researchers. For various robot navigation tasks, some researchers preferred stereo cameras [66,73,91,100,109] to handle the issue of depth perception as it is similar to 3D perception in human vision. Similarly, while developing the control system of soft robots, researchers found the endoscopic [86,90] and microscopic [97,103] cameras very useful because of their random window reading ability. Monocular cameras [68,101,104] and eye trackers [65] have also been used by several experimenters in different studies.

On the other hand, researchers used more than one camera [72,73,80,91,92,96,97,108,109] to achieve a better performance and complete the tasks properly. In several cases, LiDAR [75] is also used along with the camera setup. Attaching gimbal [68,75] with the camera is also popular among researchers. Figure 3 and Table 2 represents different sensory input for vision-based systems and applications along with their advantages. Even with today's cutting-edge technologies, vision-based systems researchers still face issues including reflected patterns, drift, accumulation error, low spatial resolution, line-of-sight obstruction, and ambient light saturation. According to recent studies, photometric approaches are gaining popularity over geometric approaches, as are multi-depth 3D cameras.

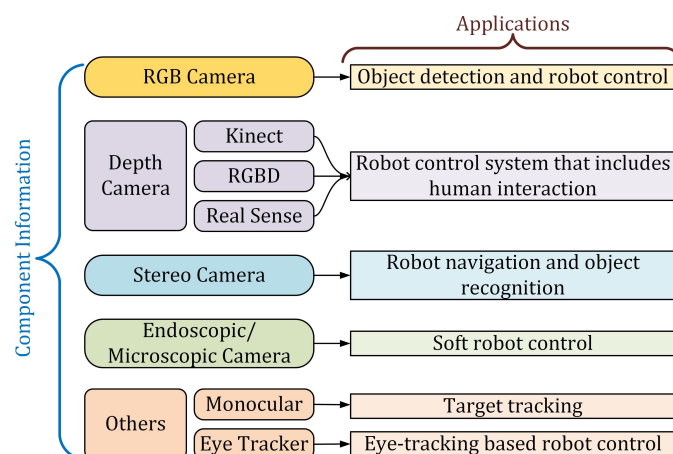


Figure 3. Different input components and their applications.

Table 2. Component (sensory input) used in different vision-based manipulation systems.

Input Category	Advantages	Potential Application
RGB Camera	Capture real-time images with a wider range of colors.	Object detection and robot manipulation/control.
Depth Camera (Kinect, RGBD, RealSense)	Sense depth of different objects and associate with an RGB camera for real-time image capture.	Robot control system that includes human interaction.
Stereo Camera	Mimic human binocular vision by using multiple lenses.	Robot navigation and object recognition.
Endoscopic/microscopic Camera	Capture images with the higher resolution magnification capability.	Soft robot control.
Monocular Camera	Ensure true field of view in low cost and lighter weight.	Target tracking.
Eye-Tracker	Track the movement of eyes in real time.	Eye-tracking-based robot control.

5. Experimental Environments

Researchers examined the models and compared their performances in order to validate the performance of the proposed models. Real-time experiments and simulated platforms were both used by the scientists for testing the performance. Real-time experiments were the most popular among researchers when solving vision-based robotic manipulation problems [64–67,70,71,73,74,78,83,86,87,90,99–101,103,107,109].

On the other hand, a lot of studies preferred exploring both real-time and simulated platforms at the same time and had presented both the results [68,72,75,77,79–82,84,89,95–98,102,108]. For efficiency, few researchers took the best values of parameters from the simulated experiments and applied them to the actual experiments [91,97]. Only Simulated experiments were only exercised by few researchers [69,76,85,88,92,94,104–106]. Figure 4 presents two experiments where the implementations were performed on a simulated platform and also applied to an actual robot.

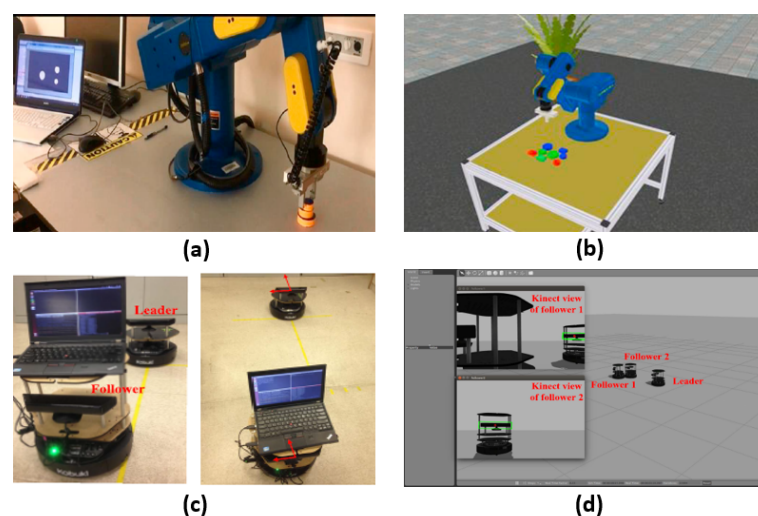


Figure 4. Experiments performed both in simulated platforms and actual robots. (a,b) Demonstration of object sorting robot manipulation [79], and (c,d) demonstration of leader–follower formation control both in real experiments and in simulated platforms [84].

6. Control Theories

While implementing vision-based robot manipulation systems, researchers have offered a number of methods and approaches, although learning-based approaches were the most popular. Different machine learning and deep learning models have been used by researchers for processing image or video data for making decisions by the system and robust output.

While designing the system architecture, to process input data, researchers used deep learning techniques and adopted a variety of modified neural networks into their system such as deep neural networks (DNNs) [69,81,85,108], deep reinforcement learning (DRL) [101], deep Q-learning [102], graph neural networks (GNNs) [92], neural network-based brain emotional nesting network (BENN) [98], probabilistic neural network (PNN) [65], etc. Different types of convolutional networks are the most popular among scientists as they have some unique features that work really well with image data. Thus, researchers applied convolutional neural networks [67,69,88–90,92,94,97], action primitive convolutional neural networks (AP-CNN) [93], and regions with convolutional neural networks (R-CNNs) [91,93], and achieved robust and generalized performance from the systems. Recurrent neural networks (RNNs), another variation of artificial neural networks also successfully examined by the experts in some systems [78,80,82]. Different pre-trained models are also gaining notable popularity among researchers especially for the data-processing tasks as these models are trained over millions of data, well known for their remarkable

For various detection tasks, different filter and feedback-based approaches were also explored by some researchers. Window filters [93], Kalman filter [96,103], pose filter [104], and other filters were applied successfully in some study and effective performances were achieved by the systems. Similarly, color, motion, and shape-based cues [71], A* method [72,87], image-processing toolbox in Matlab [79], cerebellar model articulation method [82], color quantization method, adaptive control [105], and mask function [85], refinement module [97] were also applied by researchers while solving robotic manipulation problems.

For the control segment, Gaussian and regression evaluation [68,74,86], eye-in-hand visual servo control framework [83,86,91], PID controller [70,75,81], and leap motion controller [78] were most popular among the scientists. Although other algorithms such as random actions sample- and planning-based control method [80], geometry methods [85,109], principal component analysis [93], etc. were successfully explored by the researchers as well for controlling the action of the robotic systems. Table 4 presents different deterministic methods applied in the vision-based manipulation and their accuracy.

Table 4. Deterministic methods applied in the vision-based manipulation along with accuracy.

Study	Filters	A* Method	Gaussian	Eye-in-Hand	PID	Instance/Image Segmentation	Geometry Method	Others	Accuracy (%)
[68]			✓						-
[70]					✓				-
[71]								Phase-based representation	96.1
[72]		✓							-
[74]			✓						97.75
[75]					✓				-
[78]								Leap motion controller	88
[79]								Matlab toolbox	-
[80]								Random actions sampled method	83
[81]					✓				-
[82]								Cerebellar model articulation controller	-
[83]				✓					-
[85]	✓						✓		99.8
[86]			✓	✓					-
[91]				✓					99.22
[93]	✓							Bayesian probability model	90
[96]	✓								-
[97]						✓		Refinement method	89
[87]		✓							92
[107]						✓		Plane extraction	97
[109]							✓		99.18

Researchers have shown their innovation and creativity while designing the system but we still can generalize and categorize the basic structure of those systems into two

categories: learning-based models, and filter/mask-based approaches. In both designs, the model takes image or video data as input, pre-processes them, and sends them through a network. In the learning-based model, the network is usually a neural network where the processed data pass through different types of layers, such as the convolution layer, recurrent layer, etc., for extracting different kinds of information from it. Similarly, in the filter/mask-based design, the network is generally a filtering/masking network where the processed data pass through different types of filters and masks for extracting different kinds of information from it. Then, both networks detect and identify different features and pass them to the action planning section of the network. In this section, the system processes the information gathered from the previous network and plans actions accordingly. Finally, the system executes the action planned by the network. Figure 5 illustrates the generalized system architecture exercised by the researchers while solving the vision-based robotic manipulation problems.

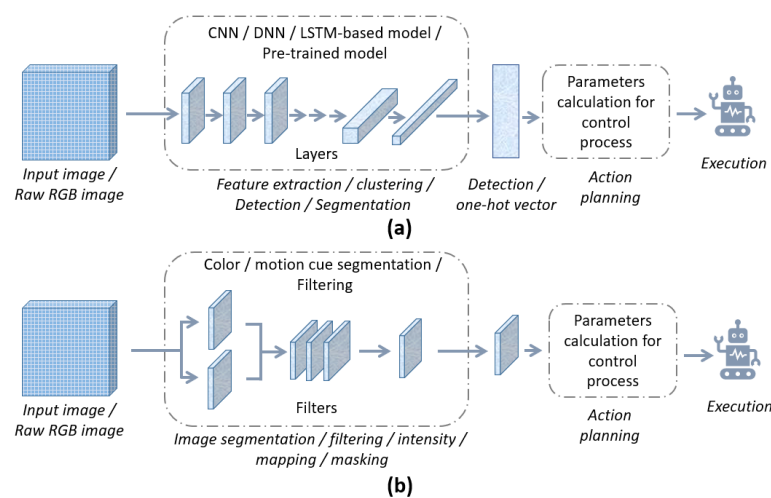


Figure 5. Generalized structure of the systems: (a) shows the generalized architecture of learning based models; and (b) shows the generalized architecture of filter/masking based approaches for the robotic manipulation problems.

7. Applications

Vision-based autonomous robot manipulation for various applications has received a lot of attention in the recent decade. Manipulation based on vision occurs when a robot manipulates an item utilizing computer vision with the feedback from the data of one or more camera sensors. The increased complexity of jobs performed by fully autonomous robots has resulted from advances in computer vision and artificial intelligence. A lot of research is going on in the computer vision field, and it may be able to provide us with more natural, non-contact solutions in the future. Human intelligence is also required for robot decision-making and control in situations in which the environment is mainly unstructured, the objects are unfamiliar, and the motions are unknown. A human–robot interface is a fundamental approach for teleoperation solutions because it serves as a link between the human intellect and the actual motions of the remote robot. The current approach of robot-manipulator in teleoperation, which makes use of vision-based tracking, allows the communication of tasks to the robot manipulator in a natural way, often utilizing the same hand gestures that would ordinarily be used for a task. The use of direct position control of the robot end-effector in vision-based robot manipulation allows for greater precision in manipulating robots.

Manipulation of deformable objects, autonomous vision-based tracking systems, tracking moving objects of interest, visual-based real-time robot control, vision-based target detection as well as object recognition, multi-agent system leader–follower formation control using a vision-based tracking scheme and vision-based grasping method to grasp the

target object for manipulation are some of the well-known applications in vision-based robot manipulation. We have classified the application of vision-based works into six categories: manipulation of the object, vision-based tracking, object detection, pathfinding/navigation, real-time remote control, and robotic arm/grasping. The summary of recent vision-based applications are mentioned in Table 5.

Table 5. Application of vision-based works.

Study	Manipulation of Object	Vision-Based Tracking	Object Detection	Path Finding/Navigation	Real-Time Remote Control	Robotic Arm/Grasping
[67]	✓					
[68]		✓				
[69]			✓			
[70]		✓	✓			
[71]			✓			
[72]			✓	✓		
[73]		✓		✓		
[74]			✓		✓	
[75]		✓	✓			
[76]						✓
[77]	✓	✓				✓
[78]	✓	✓	✓			✓
[79]	✓	✓	✓			✓
[80]			✓			✓
[81]		✓		✓		
[82]		✓			✓	
[83]			✓		✓	
[84]		✓		✓		
[85]			✓	✓		
[86]	✓				✓	✓
[90]	✓					✓
[91]	✓		✓			
[92]		✓		✓	✓	
[93]	✓		✓			
[96]		✓				✓
[97]			✓			✓
[64]			✓	✓		
[94]		✓			✓	
[65]		✓			✓	
[87]		✓	✓	✓		✓
[95]		✓		✓		
[88]		✓			✓	
[89]	✓	✓			✓	✓
[66]	✓					✓
[98]		✓	✓			✓
[99]				✓		
[100]	✓					✓
[101]		✓	✓			
[102]	✓					

Table 5. Cont.

Study	Manipulation of Object	Vision-Based Tracking	Object Detection	Path Finding/Navigation	Real-Time Remote Control	Robotic Arm/Grasping
[103]		✓		✓		
[104]				✓		
[105]	✓					
[106]	✓					
[107]						✓
[108]				✓		
[109]		✓		✓		

8. Challenges and Potential Solutions

Researchers from diverse academic fields have effectively implemented vision-based techniques in robotic manipulation tasks. As a result, these methods have emerged as one of the most promising means currently available. Even though the performances of those systems were exceptional, there is still some potential for development in virtually all studies, as well as a lot of obstacles to overcome.

To begin with, not all suggested systems were subjected to real-world testing; instead, trials were only conducted on simulated platforms [69,76,85,88,92,94,100,102–104,106]. There is a significant possibility that the system may not perform as well in the actual world, although the experimental results were impressive. Therefore, additional real-world trials should be included in future research.

The most prevalent function of a robot vision system is to identify the position and orientation of a known object. Consequently, the challenges associated with both have typically been resolved in most integrated vision solutions. However, still, the deformation of the object caused by force or movable joints, background, incorrect camera placement, and occlusion can cause considerable problems for robotic vision techniques.

While the majority of studies covered numerous experiments, in the majority of instances, the experiments were conducted under various assumptions, such as that humans will not move excessively fast was assumed in a study [70], and exploring new areas was left for future research [67,69,71,74,76,77,80,81,83,93,106–108]. Few research also failed to handle dynamic tasks [72,86,99] and noted the need for new data [67,78] to enhance the performance of their suggested solutions. Consequently, robustness is yet to be determined by this research. Therefore, researchers should concentrate on developing a more reliable control system for vision-based robot manipulation system.

The human eye is more adaptable and sensitive than imaging sensors. A vision sensor will be unable to detect objects reliably if it is exposed to improper lighting. There are numerous solutions to the lighting problem. Active lighting can be incorporated into the vision sensor itself. Other solutions include infrared lighting, environment-fixed lighting, and technologies that employ other forms of light, such as lasers.

When offering learning-based methodologies employing visual input, several researchers attempted to add autonomy into their system but failed and reserved them for future study [68,71]. Additionally, hardware components and sensors/camera upgrades throughout time were maintained for future research and development [96,108,110], as well as the examination of alternative potential system architectures [65,71,72,74,77]. To achieve versatility in this subject, further research has to be conducted.

The amount of time and space required for computing is still an unresolved issue for vision-based systems. For the suggested solutions to be implemented in real time, the models need to analyze the input and respond appropriately quickly. Therefore, further research must be done to reduce the time needed for processing, and the complexity of the computations reference [96]. Table 6 summarizes the categorization of the vision-based methods based on computational complexity.

Table 6. Computational complexity of the vision-based approach.

Studies	Complexity	Proposed Solution
[68,71,73]	Low	Frame-difference and machine learning-based approach; Color, motion, and shape cues, and phase-based object representation; Haar-like feature, line detection, fine-tuned object detection, and template matching-based approach.
[64,70,72,74,75,79,86,87,95,99,100,107,109]	Moderate	Haar feature-based design, Kanade Lucas Tomasi method, 3D localization, and PID controller-based system; SVM and A* algorithm-based method; Thresholding colors, mask function, and color quantization-based process; Color-based image processing using OpenCV, fuzzy logic, and PID-based control; Image processing tool in Matlab and visual basic-based system; Eye-in-hand visual servo and a local Gaussian-based process; K-means algorithm and Voronoi diagram-based path planning model; Cube decomposition and A* algorithm-based system; Model-free feedback controller; Planar extraction and clustering-based instance segmentation and grasping pose estimation; Combination of triangulation and the CAMSHIFT algorithm for tracking a target object.
[65–67,69,76–78,80–85,88–94,96–98,101–106,108]	High	Deep CNN model; Proposal layer and CNN-based method; Deep RL algorithm; RL and DNN-based framework; leap motion, Playstation, and NN-based controller; Random action sampled method, RL model with RNN-based video prediction, and planning-based control; Vision-based RNN, an emotional network, and a recurrent loop-based structure; Kinematics, Lyapunov analysis, and vision models; Combination of DNN and PID for landing of UAV; CamShift algorithm-based architecture; Deep learning, reference marker, and geometry methods-based approach; ResNet-50-based object detection method, an image recognition network, and a deep grasping guidance network-based framework; R-CNN-based model and an eye-to-hand stereo camera configuration; CNN- and GNN-based architecture; Action primitive CNN, window filter, R-CNN, principal component analysis, and action planner-based framework; Kalman filter, Wiener forthcoming human hand position estimation, and a local path planning algorithm-based architecture; U-Net-based CNN model, an image segmentation method, a policy module, and a refinement module-based system; CNN-based network; SVM and PNN-based model; Face-detector module in OpenCV, Deep CNN model, and ROS master robot arm controller-based system; Vision servoing, AdaBoost-SVM, and hybrid force and motion optimization-based method; Rough reaching movement controller (pre-trained RBF), inverse kinematics, brain emotional nesting network (BENN), and adaptive laws-based controller; Combination of Q-learning, computer vision, and neural networks for robotic path planning.

Some challenges relate more to the approach to vision setups than the technical aspects of vision algorithms. A common pitfall is having overly optimistic expectations for a computer's visual capabilities. The best results can be achieved from a technological tool by making sure that one's expectations are in line with the technology's capabilities.

9. Discussions

Vision sensors offer a large amounts of information about the environment in which robots operate. As a result, vision is critical for robots that operate in unstructured settings. In structured settings, vision is also important to provide some flexibility or looseness in order to consolidate workplace conditions. As a result, a significant amount of research has been conducted in order to build vision-based robot controllers. In the early 1970s, the first vision-based robot control system was described [111]. Progress in vision-based control has been sluggish since then, owing to the need for specialized and costly pipelined pixel-processing hardware. However, as processing power and sensor technologies improve, we may expect to see more scientific studies in the vision-based control field. The constraints of vision sensors, long-time image processing, picture resolution, and frame rate are critical issues with vision-based robot control. Despite the existence of relatively fast cameras and updated algorithms, and visual measurement sampling frequencies are still lower than the frequency of positioning encoders and angular position sensors. Visual measurements are

frequently required in the context of dynamic robotics to offer feedback for controlling or estimating the dynamic state variables of the system. In a feedback-based robot control system, the sample rate must be high enough, and the sensor latency must be minimized to achieve the controller stability and robustness. When calculating the state variables parameters of the system for state feedback robot control, the sensor latency of the visual observations must be taken into account. When visual measurements are combined with high-frequency position data, the feedback controller can be executed at a higher frequency, resulting in improved stability and quick convergence. In addition, when integrating the measurements, the sensor latency of the visual measurement techniques must be considered appropriately. Otherwise, when the end-effector of a robot is moving, vision provides incorrect information. The visual observations are unreliable due to the low sampling frequency and sensor latency. The images have faults due to the camera sensor's low resolution, motion blur, and the inclusion of noise in the picture. When taking a single assessment at each instant, as is customary in a vision-based robot control, three-dimensional flexible robot motion dynamics and control become more challenging. In three-dimensional vision-based robot control, more than one camera can be used efficiently. Visual measurements that are subject to ambiguity may cause undesired oscillations and a decrease in accuracy. Still, by comparing a single image to several measures, more precise estimations of target motion can be generated.

The manipulation of real-world things is one of machine intelligence's most significant constraints. Robots can learn complicated manipulation tasks using vision-based learning techniques. Improving vision-based robot manipulation performance might be a good trend in tackling the problems as robotic technology and sensor technologies advance. The goal of this study is to provide a general review of the evolution of vision-based control technologies for robot manipulation. More information will be accessible to manipulate robots as existing imaging technologies progress and new control algorithms are developed. As a result, we hope that our review study will help develop autonomous robotic systems with human-like manipulation abilities.

10. Conclusions

This paper delivers a comprehensive study on vision-based approaches in the field of robotic manipulation systems. Different innovative and exceptional manipulation techniques have recently been introduced by researchers. Nonetheless, vision-based approaches have found their popularity among researchers because of their accuracy and promising performance. Forty-six recent papers have been accumulated together by prioritizing both control and application-based problems and analyzed for this study. After summing up all the studies, we can state that among all the mentioned methods, different deep neural networks and deep convolutional network-based approaches are the most popular ones; contrarily, different conventional methods are becoming less popular among researchers nowadays. While designing the system architecture, researchers have mostly followed two types of structure, learning-based models, deterministic and filter/mask-based approaches. Both the simulated platform and the real-world environment were equally explored by the researchers during the testing of the proposed models. Basic RGB cameras and USB cameras were mostly used by them; nevertheless, for the systems that interact with humans, researchers preferred depth cameras for getting more information about the surrounding. Additionally, further exploration is needed to resolve the addressed open challenges so that vision-based approaches can have more efficient and practical applications in the field of robotic manipulation.

Author Contributions: Conceptualization, methodology, investigation, validation, formal analysis, data curation M.T.S.; resources, supervision, project administration, M.H.R., S.I.A., and J.G.; visualization, writing—original draft preparation, writing—review and editing M.T.S., M.S.H.S., M.I.I.Z., and M.H.R. All authors have read and agreed to the published version of the manuscript.

Funding: The contents of this Journal were partially supported by a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90DPGE0018-01-00). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this journal do not necessarily represent the policy of NIDILRR, ACL, or HHS, and it should not assume endorsement by the Federal Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

3D	Three-Dimensional
ACMs	Aerial Continuum Manipulation System
APCNN	Action Primitive Convolutional Neural Networks
ARM	Attention-Driven Robot Manipulation
BENN	Brain Emotional Nesting Network
CNN	Convolutional Neural Networks
DCNN	Deep Convolutional Neural Networks
DN	DenseNet
DNN	Deep Neural Networks
DRL	Deep Reinforcement Learning
GDRL	Generalized Deep Reinforcement Learning
GNN	Graph Neural Networks
GPU	Graphics Processing Unit
INeRF	Inverting Neural Radiance Field
ML	Machine Learning
NeRF	Neural Radiance Field
NN	Neural Network
PID	Proportional–Integral–Derivative Control
PNN	Probabilistic Neural Network
RCNN	Regions with Convolutional Neural Networks
RBF	Radial Basis Function
RL	Reinforcement Learning
RNN	Recurrent Neural Networks
ROS	Robotic Operating System
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
USV	Unmanned Surface Vehicles

References

1. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and localization methods for vision-based fruit picking robots: A review. *Front. Plant Sci.* **2020**, *11*, 510. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, H.; Zhou, X.; Lan, X.; Li, J.; Tian, Z.; Zheng, N. A real-time robotic grasping approach with oriented anchor box. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *51*, 3014–3025. [[CrossRef](#)]
3. Bertolucci, R.; Capitanelli, A.; Maratea, M.; Mastrogiiovanni, F.; Vallati, M. Automated planning encodings for the manipulation of articulated objects in 3d with gravity. In Proceedings of the International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, 19–22 November; Springer: Berlin/Heidelberg, Germany, 2019; pp. 135–150.
4. Marino, H.; Ferrati, M.; Settini, A.; Rosales, C.; Gabicini, M. On the problem of moving objects with autonomous robots: A unifying high-level planning approach. *IEEE Robot. Autom. Lett.* **2016**, *1*, 469–476. [[CrossRef](#)]
5. Kong, S.; Tian, M.; Qiu, C.; Wu, Z.; Yu, J. IWSCR: An intelligent water surface cleaner robot for collecting floating garbage. *IEEE Trans. Syst. Man, Cybern. Syst.* **2020**, *51*, 6358–6368. [[CrossRef](#)]
6. Miller, S.; Van Den Berg, J.; Fritz, M.; Darrell, T.; Goldberg, K.; Abbeel, P. A geometric approach to robotic laundry folding. *Int. J. Robot. Res.* **2012**, *31*, 249–267. [[CrossRef](#)]

7. Do, H.M.; Choi, T.; Park, D.; Kyung, J. Automatic cell production for cellular phone packing using two dual-arm robots. In Proceedings of the 2015 15th International Conference on Control, Automation and Systems (ICCAS), Busan, Republic of Korea, 13–16 October 2015; IEEE: Piscataway Township, NJ, USA, 2015; pp. 2083–2086.
8. Kemp, C.C.; Edsinger, A. Robot manipulation of human tools: Autonomous detection and control of task relevant features. In Proceedings of the Fifth International Conference on Development and Learning, Bloomington, IN, USA, 31 May–3 June 2006; Volume 42.
9. Edsinger, A. *Robot Manipulation in Human Environments*; CSAIL Technical Reports: Cambridge, MA, USA, 2007.
10. *What Are Manual Robots?* Bright Hub Engineering: Albany, NY, USA, 11 December 2009.
11. Van Pham, H.; Asadi, F.; Abut, N.; Kandilli, I. Hybrid spiral STC-hedge algebras model in knowledge reasonings for robot coverage path planning and its applications. *Appl. Sci.* **2019**, *9*, 1909. [\[CrossRef\]](#)
12. Merlet, J.P. Solving the forward kinematics of a Gough-type parallel manipulator with interval analysis. *Int. J. Robot. Res.* **2004**, *23*, 221–235. [\[CrossRef\]](#)
13. Kucuk, S.; Bingul, Z. *Robot Kinematics: Forward and Inverse Kinematics*; INTECH Open Access Publisher: London, UK, 2006.
14. Seng Yee, C.; Lim, K.B. Forward kinematics solution of Stewart platform using neural networks. *Neurocomputing* **1997**, *16*, 333–349.
15. Lee, B.J. Geometrical derivation of differential kinematics to calibrate model parameters of flexible manipulator. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 106. [\[CrossRef\]](#)
16. Ye, S.; Wang, Y.; Ren, Y.; Li, D. Robot calibration using iteration and differential kinematics. *J. Phys. Conf. Ser.* **2006**, *48*, 1. [\[CrossRef\]](#)
17. Park, I.W.; Lee, B.J.; Cho, S.H.; Hong, Y.D.; Kim, J.H. Laser-based kinematic calibration of robot manipulator using differential kinematics. *IEEE/ASME Trans. Mechatron.* **2011**, *17*, 1059–1067. [\[CrossRef\]](#)
18. D'Souza, A.; Vijayakumar, S.; Schaal, S. Learning inverse kinematics. In Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180), Maui, HI, USA, 29 October–3 November, 2001; IEEE: Piscataway Township, NJ, USA, 2001; Volume 1, pp. 298–303.
19. Grochow, K.; Martin, S.L.; Hertzmann, A.; Popović, Z. Style-based inverse kinematics. In Proceedings of the ACM SIGGRAPH 2004 Papers, Los Angeles, CA, USA, 8–12 August 2004; AMC: New York, New York, USA, 2004; pp. 522–531.
20. Manocha, D.; Canny, J.F. Efficient inverse kinematics for general 6R manipulators. *IEEE Trans. Robot. Autom.* **1994**, *10*, 648–657. [\[CrossRef\]](#)
21. Goldenberg, A.; Benhabib, B.; Fenton, R. A complete generalized solution to the inverse kinematics of robots. *IEEE J. Robot. Autom.* **1985**, *1*, 14–20. [\[CrossRef\]](#)
22. Wang, L.C.; Chen, C.C. A combined optimization method for solving the inverse kinematics problems of mechanical manipulators. *IEEE Trans. Robot. Autom.* **1991**, *7*, 489–499. [\[CrossRef\]](#)
23. Tedrake, R. Robotic Manipulation Course Notes for MIT 6.4210. 2022. Available online: <https://manipulation.csail.mit.edu/> (accessed on 2 November 2022).
24. Billard, A.; Kragic, D. Trends and challenges in robot manipulation. *Science* **2019**, *364*, eaat8414. [\[CrossRef\]](#)
25. Harvey, I.; Husband, P.; Cliff, D.; Thompson, A.; Jakobi, N. Evolutionary robotics: The Sussex approach. *Robot. Auton. Syst.* **1997**, *20*, 205–224. [\[CrossRef\]](#)
26. Belta, C.; Kumar, V. Abstraction and control for groups of robots. *IEEE Trans. Robot.* **2004**, *20*, 865–875. [\[CrossRef\]](#)
27. Arents, J.; Greitans, M. Smart industrial robot control trends, challenges and opportunities within manufacturing. *Appl. Sci.* **2022**, *12*, 937. [\[CrossRef\]](#)
28. Su, Y.H.; Young, K.Y. Effective manipulation for industrial robot manipulators based on tablet PC. *J. Chin. Inst. Eng.* **2018**, *41*, 286–296. [\[CrossRef\]](#)
29. Su, Y.; Liao, C.; Ko, C.; Cheng, S.; Young, K.Y. An AR-based manipulation system for industrial robots. In Proceedings of the 2017 11th Asian Control Conference (ASCC), Gold Coast, Australia, 17–20 December 2017; IEEE: Piscataway Township, NJ, USA, 2017; pp. 1282–1285.
30. Inkulu, A.K.; Bahubalendruni, M.R.; Dara, A.; SankaranarayanaSamy, K. Challenges and opportunities in human robot collaboration context of Industry 4.0-a state of the art review. In *Industrial Robot: The International Journal of Robotics Research and Application*; Emerald Publishing: Bingley, UK, 2021.
31. Balaguer, C.; Abderrahim, M. *Robotics and Automation in Construction*; BoD—Books on Demand: Elizabeth, NJ, USA, 2008.
32. Capitanielli, A.; Maratea, M.; Mastrogiorganni, F.; Vallati, M. Automated planning techniques for robot manipulation tasks involving articulated objects. In Proceedings of the Conference of the Italian Association for Artificial Intelligence, Bari, Italy, 14–17 November, 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 483–497.
33. Finžgar, M.; Podržaj, P. Machine-vision-based human-oriented mobile robots: A review. *Stroj. Vestn. J. Mech. Eng.* **2017**, *63*, 331–348 [\[CrossRef\]](#)
34. Zlatintsi, A.; Dometios, A.; Kardaris, N.; Rodomagoulakis, I.; Koutras, P.; Papageorgiou, X.; Maragos, P.; Tzafestas, C.S.; Vartholomeos, P.; Hauer, K.; et al. I-Support: A robotic platform of an assistive bathing robot for the elderly population. *Robot. Auton. Syst.* **2020**, *126*, 103451. [\[CrossRef\]](#)
35. Shahria, M.T.; Iftekhar, L.; Rahman, M.H. Learning-Based Approaches in Swarm Robotics: In A Nutshell. In Proceedings of the International Conference on Mechanical, Industrial and Energy Engineering 2020, Khulna, Bangladesh, 19–21 December 2020.
36. Martinez-Martin, E.; Del Pobil, A.P. Vision for robust robot manipulation. *Sensors* **2019**, *19*, 1648. [\[CrossRef\]](#) [\[PubMed\]](#)

37. Budge, B. 1.1 Computer Vision in Robotics. In *Deep Learning Approaches for 3D Inference from Monocular Vision*; Queensland University of Technology: Brisbane, Australia, 2020; p. 4.
38. Wang, X.; Wang, X.L.; Wilkes, D.M. An automated vision based on-line novel percept detection method for a mobile robot. *Robot. Auton. Syst.* **2012**, *60*, 1279–1294. [\[CrossRef\]](#)
39. *Robot Vision-Sensor Solutions for Robotics*; SICK: Minneapolis, MN, USA. Available online: <https://www.sick.com/cl/en/robot-vision-sensor-solutions-for-robotics/w/robotics-robot-vision/> (accessed on 2 November 2022).
40. Gao, Y.; Spiteri, C.; Pham, M.T.; Al-Milli, S. A survey on recent object detection techniques useful for monocular vision-based planetary terrain classification. *Robot. Auton. Syst.* **2014**, *62*, 151–167. [\[CrossRef\]](#)
41. Baerveldt, A.J. A vision system for object verification and localization based on local features. *Robot. Auton. Syst.* **2001**, *34*, 83–92. [\[CrossRef\]](#)
42. Garcia-Fidalgo, E.; Ortiz, A. Vision-based topological mapping and localization methods: A survey. *Robot. Auton. Syst.* **2015**, *64*, 1–20. [\[CrossRef\]](#)
43. Du, G.; Wang, K.; Lian, S.; Zhao, K. Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: A review. *Artif. Intell. Rev.* **2021**, *54*, 1677–1734. [\[CrossRef\]](#)
44. Gupta, M.; Kumar, S.; Behera, L.; Subramanian, V.K. A novel vision-based tracking algorithm for a human-following mobile robot. *IEEE Trans. Syst. Man, Cybern. Syst.* **2016**, *47*, 1415–1427. [\[CrossRef\]](#)
45. Zhang, K.; Chen, J.; Yu, G.; Zhang, X.; Li, Z. Visual trajectory tracking of wheeled mobile robots with uncalibrated camera extrinsic parameters. *IEEE Trans. Syst. Man, Cybern. Syst.* **2020**, *51*, 7191–7200. [\[CrossRef\]](#)
46. Lin, L.; Yang, Y.; Cheng, H.; Chen, X. Autonomous vision-based aerial grasping for rotorcraft unmanned aerial vehicles. *Sensors* **2019**, *19*, 3410. [\[CrossRef\]](#)
47. Yang, L.; Etsuko, K. Review on vision-based tracking in surgical navigation. *IET Cyber-Syst. Robot.* **2020**, *2*, 107–121. [\[CrossRef\]](#)
48. Kroemer, O.; Niekum, S.; Konidaris, G. A Review of Robot Learning for Manipulation: Challenges, Representations, and Algorithms. *J. Mach. Learn. Res.* **2021**, *22*, 30–31.
49. Ruiz-del Solar, J.; Loncomilla, P. Applications of deep learning in robot vision. In *Deep Learning in Computer Vision*; CRC Press: Boca Raton, FL, USA, 2020; pp. 211–232.
50. Watt, N. *Deep Neural Networks for Robot Vision in Evolutionary Robotics*; Nelson Mandela University: Gqeberha, South Africa, 2021.
51. Jiang, Y.; Yang, C.; Na, J.; Li, G.; Li, Y.; Zhong, J. A brief review of neural networks based learning and control and their applications for robots. *Complexity* **2017**, *2017*, 1895897. [\[CrossRef\]](#)
52. Vemuri, A.T.; Polycarpou, M.M. Neural-network-based robust fault diagnosis in robotic systems. *IEEE Trans. Neural Netw.* **1997**, *8*, 1410–1420. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Prabhu, S.M.; Garg, D.P. Artificial neural network based robot control: An overview. *J. Intell. Robot. Syst.* **1996**, *15*, 333–365. [\[CrossRef\]](#)
54. Köker, R.; Öz, C.; Çakar, T.; Ekiz, H. A study of neural network based inverse kinematics solution for a three-joint robot. *Robot. Auton. Syst.* **2004**, *49*, 227–234. [\[CrossRef\]](#)
55. Pierson, H.A.; Gashler, M.S. Deep learning in robotics: A review of recent research. *Adv. Robot.* **2017**, *31*, 821–835. [\[CrossRef\]](#)
56. Liu, H.; Fang, T.; Zhou, T.; Wang, Y.; Wang, L. Deep learning-based multimodal control interface for human–robot collaboration. *Procedia CIRP* **2018**, *72*, 3–8. [\[CrossRef\]](#)
57. Liu, Y.; Li, Z.; Liu, H.; Kan, Z. Skill transfer learning for autonomous robots and human–robot cooperation: A survey. *Robot. Auton. Syst.* **2020**, *128*, 103515. [\[CrossRef\]](#)
58. Nakashima, K.; Nagata, F.; Ochi, H.; Otsuka, A.; Ikeda, T.; Watanabe, K.; Habib, M.K. Detection of minute defects using transfer learning-based CNN models. *Artif. Life Robot.* **2021**, *26*, 35–41. [\[CrossRef\]](#)
59. Tanaka, K.; Yonetani, R.; Hamaya, M.; Lee, R.; von Drigalski, F.; Ijiri, Y. Trans-am: Transfer learning by aggregating dynamics models for soft robotic assembly. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 4627–4633.
60. Zarif, M.I.I.; Shahria, M.T.; Sunny, M.S.H.; Rahaman, M.M. A Vision-based Object Detection and Localization System in 3D Environment for Assistive Robots' Manipulation. In Proceedings of the 9th International Conference of Control Systems, and Robotics (CDSR'22), Niagara Falls, Canada, 02–04 June 2022.
61. Janabi-Sharifi, F.; Marey, M. A kalman-filter-based method for pose estimation in visual servoing. *IEEE Trans. Robot.* **2010**, *26*, 939–947. [\[CrossRef\]](#)
62. Wu, B.F.; Jen, C.L. Particle-filter-based radio localization for mobile robots in the environments with low-density WLAN APs. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6860–6870. [\[CrossRef\]](#)
63. Zhao, D.; Li, C.; Zhu, Q. Low-pass-filter-based position synchronization sliding mode control for multiple robotic manipulator systems. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2011**, *225*, 1136–1148. [\[CrossRef\]](#)
64. Eid, M.A.; Giakoumidis, N.; El Saddik, A. A novel eye-gaze-controlled wheelchair system for navigating unknown environments: Case study with a person with ALS. *IEEE Access* **2016**, *4*, 558–573. [\[CrossRef\]](#)
65. Cao, Y.; Miura, S.; Kobayashi, Y.; Kawamura, K.; Sugano, S.; Fujie, M.G. Pupil variation applied to the eye tracking control of an endoscopic manipulator. *IEEE Robot. Autom. Lett.* **2016**, *1*, 531–538. [\[CrossRef\]](#)
66. Hu, Y.; Li, Z.; Li, G.; Yuan, P.; Yang, C.; Song, R. Development of sensory-motor fusion-based manipulation and grasping control for a robotic hand-eye system. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 1169–1180. [\[CrossRef\]](#)

67. Nair, A.; Chen, D.; Agrawal, P.; Isola, P.; Abbeel, P.; Malik, J.; Levine, S. Combining self-supervised learning and imitation for vision-based rope manipulation. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May 2017; IEEE: Piscataway Township, NJ, USA, 2017; pp. 2146–2153.
68. Cheng, H.; Lin, L.; Zheng, Z.; Guan, Y.; Liu, Z. An autonomous vision-based target tracking system for rotorcraft unmanned aerial vehicles. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Exeter, UK, 21–23 June 2017; IEEE: Piscataway Township, NJ, USA, 2017; pp. 1732–1738.
69. Lu, K.; An, X.; Li, J.; He, H. Efficient deep network for vision-based object detection in robotic applications. *Neurocomputing* **2017**, *245*, 31–45. [\[CrossRef\]](#)
70. Yao, N.; Anaya, E.; Tao, Q.; Cho, S.; Zheng, H.; Zhang, F. Monocular vision-based human following on miniature robotic blimp. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Vancouver, BC, Canada, 24–28 September 2017; IEEE: Piscataway Township, NJ, USA, 2017; pp. 3244–3249.
71. Martinez-Martin, E.; Del Pobil, A.P. Object detection and recognition for assistive robots: Experimentation and implementation. *IEEE Robot. Autom. Mag.* **2017**, *24*, 123–138. [\[CrossRef\]](#)
72. Abiyev, R.H.; Arslan, M.; Gunsul, I.; Cagman, A. Robot pathfinding using vision based obstacle detection. In Proceedings of the 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), Exeter, UK, 21–23 June 2017; IEEE: Piscataway Township, NJ, USA, 2017; pp. 1–6.
73. Shin, B.S.; Mou, X.; Mou, W.; Wang, H. Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities. *Mach. Vis. Appl.* **2018**, *29*, 95–112. [\[CrossRef\]](#)
74. Dönmez, E.; Kocamaz, A.F.; Dirik, M. A vision-based real-time mobile robot controller design based on gaussian function for indoor environment. *Arab. J. Sci. Eng.* **2018**, *43*, 7127–7142. [\[CrossRef\]](#)
75. Rabah, M.; Rohan, A.; Talha, M.; Nam, K.H.; Kim, S.H. Autonomous vision-based target detection and safe landing for UAV. *Int. J. Control. Autom. Syst.* **2018**, *16*, 3013–3025. [\[CrossRef\]](#)
76. Quillen, D.; Jang, E.; Nachum, O.; Finn, C.; Ibarz, J.; Levine, S. Deep reinforcement learning for vision-based robotic grasping: A simulated comparative evaluation of off-policy methods. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 6284–6291.
77. Kalashnikov, D.; Irpan, A.; Pastor, P.; Ibarz, J.; Herzog, A.; Jang, E.; Quillen, D.; Holly, E.; Kalakrishnan, M.; Vanhoucke, V.; et al. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. *arXiv* **2018**, arXiv:1806.10293.
78. Rahmatizadeh, R.; Abolghasemi, P.; Bölöni, L.; Levine, S. Vision-based multi-task manipulation for inexpensive robots using end-to-end learning from demonstration. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 3758–3765.
79. Ali, M.H.; Aizat, K.; Yerkhan, K.; Zhandos, T.; Anuar, O. Vision-based robot manipulator for industrial applications. *Procedia Comput. Sci.* **2018**, *133*, 205–212. [\[CrossRef\]](#)
80. Ebert, F.; Finn, C.; Dasari, S.; Xie, A.; Lee, A.; Levine, S. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control. *arXiv* **2018**, arXiv:1812.00568.
81. Almeshal, A.M.; Alenezi, M.R. A vision-based neural network controller for the autonomous landing of a quadrotor on moving targets. *Robotics* **2018**, *7*, 71. [\[CrossRef\]](#)
82. Fang, W.; Chao, F.; Yang, L.; Lin, C.M.; Shang, C.; Zhou, C.; Shen, Q. A recurrent emotional CMAC neural network controller for vision-based mobile robots. *Neurocomputing* **2019**, *334*, 227–238. [\[CrossRef\]](#)
83. Zake, Z.; Chaumette, F.; Pedemonte, N.; Caro, S. Vision-based control and stability analysis of a cable-driven parallel robot. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1029–1036. [\[CrossRef\]](#)
84. Liu, X.; Ge, S.S.; Goh, C.H. Vision-based leader–follower formation control of multiagents with visibility constraints. *IEEE Trans. Control Syst. Technol.* **2018**, *27*, 1326–1333. [\[CrossRef\]](#)
85. Shangguan, Z.; Wang, L.; Zhang, J.; Dong, W. Vision-based object recognition and precise localization for space body control. *Int. J. Aerosp. Eng.* **2019**, *2019*, 7050915. [\[CrossRef\]](#)
86. Fang, G.; Wang, X.; Wang, K.; Lee, K.H.; Ho, J.D.; Fu, H.C.; Fu, D.K.C.; Kwok, K.W. Vision-based online learning kinematic control for soft robots using local gaussian process regression. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1194–1201. [\[CrossRef\]](#)
87. Cio, Y.S.L.K.; Raison, M.; Ménard, C.L.; Achiche, S. Proof of concept of an assistive robotic arm control using artificial stereovision and eye-tracking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 2344–2352. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Guo, J.; Liu, Y.; Qiu, Q.; Huang, J.; Liu, C.; Cao, Z.; Chen, Y. A novel robotic guidance system with eye gaze tracking control for needle based interventions. *IEEE Trans. Cogn. Dev. Syst.* **2019**, *13*, 178–188. [\[CrossRef\]](#)
89. Li, P.; Hou, X.; Duan, X.; Yip, H.; Song, G.; Liu, Y. Appearance-based gaze estimator for natural interaction control of surgical robots. *IEEE Access* **2019**, *7*, 25095–25110. [\[CrossRef\]](#)
90. Yu, Y.; Cao, Z.; Liang, S.; Geng, W.; Yu, J. A novel vision-based grasping method under occlusion for manipulating robotic system. *IEEE Sens. J.* **2020**, *20*, 10996–11006. [\[CrossRef\]](#)
91. Du, Y.C.; Muslikhin, M.; Hsieh, T.H.; Wang, M.S. Stereo vision-based object recognition and manipulation by regions with convolutional neural network. *Electronics* **2020**, *9*, 210. [\[CrossRef\]](#)
92. Hu, T.K.; Gama, F.; Chen, T.; Wang, Z.; Ribeiro, A.; Sadler, B.M. VGAI: End-to-end learning of vision-based decentralized controllers for robot swarms. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 4900–4904.

93. Jia, Z.; Lin, M.; Chen, Z.; Jian, S. Vision-based robot manipulation learning via human demonstrations. *arXiv* **2020**, arXiv:2003.00385.
94. Qiu, H.; Li, Z.; Yang, Y.; Xin, C.; Bian, G.B. Real-Time Iris Tracking Using Deep Regression Networks for Robotic Ophthalmic Surgery. *IEEE Access* **2020**, *8*, 50648–50658. [\[CrossRef\]](#)
95. Wang, X.; Fang, G.; Wang, K.; Xie, X.; Lee, K.H.; Ho, J.D.; Tang, W.L.; Lam, J.; Kwok, K.W. Eye-in-hand visual servoing enhanced with sparse strain measurement for soft continuum robots. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2161–2168. [\[CrossRef\]](#)
96. Melchiorre, M.; Scimmi, L.S.; Mauro, S.; Pastorelli, S.P. Vision-based control architecture for human–robot hand-over applications. *Asian J. Control* **2021**, *23*, 105–117. [\[CrossRef\]](#)
97. Li, Y.; Qin, F.; Du, S.; Xu, D.; Zhang, J. Vision-Based Imitation Learning of Needle Reaching Skill for Robotic Precision Manipulation. *J. Intell. Robot. Syst.* **2021**, *101*, 1–13. [\[CrossRef\]](#)
98. Fang, W.; Chao, F.; Lin, C.M.; Zhou, D.; Yang, L.; Chang, X.; Shen, Q.; Shang, C. Visual-Guided Robotic Object Grasping Using Dual Neural Network Controllers. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2282–2291. [\[CrossRef\]](#)
99. Roland, C.; Choi, D.; Kim, M.; Jang, J. Implementation of Enhanced Vision for an Autonomous Map-based Robot Navigation. In Proceedings of the Korean Institute of Information and Commucation Sciences Conference, Republic of Korea, 3 October, 2021; Yeosu, Republic of Korea, 3 October 2021; The Korea Institute of Information and Commucation Engineering: Seoul, Republic of Korea, 2021; pp. 41–43.
100. Asadi, K.; Haritsa, V.R.; Han, K.; Ore, J.P. Automated object manipulation using vision-based mobile robotic system for construction applications. *J. Comput. Civ. Eng.* **2021**, *35*, 04020058. [\[CrossRef\]](#)
101. Luo, Y.; Dong, K.; Zhao, L.; Sun, Z.; Cheng, E.; Kan, H.; Zhou, C.; Song, B. Calibration-free monocular vision-based robot manipulations with occlusion awareness. *IEEE Access* **2021**, *9*, 85265–85276. [\[CrossRef\]](#)
102. James, S.; Wada, K.; Laidlow, T.; Davison, A.J. Coarse-to-Fine Q-attention: Efficient Learning for Visual Robotic Manipulation via Discretisation. *arXiv* **2021**, arXiv:2106.12534.
103. Tang, X.; Li, Y.; Liu, X.; Liu, D.; Chen, Z.; Arai, T. Vision-Based Automated Control of Magnetic Microrobots. *Micromachines* **2022**, *13*, 337. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Adamkiewicz, M.; Chen, T.; Caccavale, A.; Gardner, R.; Culbertson, P.; Bohg, J.; Schwager, M. Vision-only robot navigation in a neural radiance world. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4606–4613. [\[CrossRef\]](#)
105. Samadikhoshkho, Z.; Ghorbani, S.; Janabi-Sharifi, F. Vision-based reduced-order adaptive control of aerial continuum manipulation systems. *Aerosp. Sci. Technol.* **2022**, *121*, 107322. [\[CrossRef\]](#)
106. James, S.; Davison, A.J. Q-attention: Enabling Efficient Learning for Vision-based Robotic Manipulation. *IEEE Robot. Autom. Lett.* **2022**, *7*, 1612–1619. [\[CrossRef\]](#)
107. Yin, R.; Wu, H.; Li, M.; Cheng, Y.; Song, Y.; Handroos, H. RGB-D-Based Robotic Grasping in Fusion Application Environments. *Appl. Sci.* **2022**, *12*, 7573. [\[CrossRef\]](#)
108. Abdi, A.; Ranjbar, M.H.; Park, J.H. Computer vision-based path planning for robot arms in three-dimensional workspaces using Q-learning and neural networks. *Sensors* **2022**, *22*, 1697. [\[CrossRef\]](#)
109. Montoya Angulo, A.; Pari Pinto, L.; Sulla Espinoza, E.; Silva Vidal, Y.; Supo Colquehuanca, E. Assisted Operation of a Robotic Arm Based on Stereo Vision for Positioning Near an Explosive Device. *Robotics* **2022**, *11*, 100. [\[CrossRef\]](#)
110. Hussein, M. A review on vision-based control of flexible manipulators. *Adv. Robot.* **2015**, *29*, 1575–1585. [\[CrossRef\]](#)
111. Shirai, Y.; Inoue, H. Guiding a robot by visual feedback in assembling tasks. *Pattern Recognit.* **1973**, *5*, 99–108. [\[CrossRef\]](#)