

Vision-Based Autonomous Landing for the UAV: A Review

Long Xin ¹, Zimu Tang ², Weiqi Gai ^{2,*} and Haobo Liu ^{2,*}

¹ Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China

² School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China

* Correspondence: gwq1926038838@buaa.edu.cn (W.G.); liuhaobo@buaa.edu.cn (H.L.)

Abstract: With the rapid development of the UAV, it is widely used in rescue and disaster relief, where autonomous landing is the key technology. Vision-based autonomous landing has the advantages of strong autonomy, low cost, and strong anti-interference ability. Moreover, vision navigation has higher guidance and positioning accuracy combined with other navigation methods, such as GPS/INS. This paper summarizes the research results in the field of vision-based autonomous landing for the UAV, and divides it into static, dynamic, and complex scenarios according to the type of landing destination. Among them, the static scenario includes two categories: cooperative targets and natural landmarks; the dynamic scenario is divided into two categories: vehicle-based autonomous landing and ship-based autonomous landing. The key technologies are summarized, compared, and analyzed and the future development trends are pointed out, which can provide a reference for the research on vision-based autonomous landing of UAVs.

Keywords: autonomous landing; computer vision; UAV



Citation: Xin, L.; Tang, Z.; Gai, W.; Liu, H. Vision-Based Autonomous Landing for the UAV: A Review. *Aerospace* **2022**, *9*, 634. <https://doi.org/10.3390/aerospace9110634>

Academic Editor: Konstantinos Kontis

Received: 31 August 2022

Accepted: 29 September 2022

Published: 22 October 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

The UAV platform has the characteristics of low cost, minimal risk and high efficiency, and has been paid more and more attention by researchers. In addition to military use, UAVs have been widely used in environmental monitoring, natural exploration, maintaining social security, and disaster relief. Among them, many tasks are related to the autonomous landing of UAVs. At the same time, the landing stage is a high-incidence stage for the UAV, and the accuracy and success rate of autonomous landing often decides the success of the mission.

A lot of research has been conducted on the autonomous landing of UAVs due to the following characteristics [1]:

(1) Real-time massive information processing:

The autonomous landing of the UAV needs to comprehensively consider both the environment and the UAV itself, so as to recognize the target and calculate the relative pose in real-time in the case of unstable landing platforms and complex landing situations. During the period, the huge amount of information, the complex calculation, and the extreme requirement of high real-time performance and stability all bring difficulties to autonomous landing.

(2) Limited onboard resources:

Due to the huge amount of information, requirements for the airborne computing system are high, especially on the computer vision-oriented computing platform, so the problem of limited computing power cannot be ignored. Moreover, the complex vision algorithms are time consuming in image processing, which is very unfavorable for the real-time navigation of the UAV system and makes it more difficult to fulfill the needs of autonomous landing when the landing platform is unstable or in various complex situations.

(3) High maneuverability of UAV platforms:

The high maneuverability of the UAV itself brings higher requirements to the control system, which means it is necessary to give more rapid and accurate feedback results for the pose estimation and motion state.

(4) Limitations of traditional image processing algorithms:

At present, the target detection algorithms carried on UAV embedded systems are basically based on traditional image processing algorithms. Usually, different detection icons need to be designed in different scenarios. The detection algorithm is limited by specific geometric icons, so different detection icons need to be designed. According to the situation above, it is difficult to easily migrate the feature extraction algorithm from one scenario to another, so the stability and robustness need to be improved.

In recent years, research and literatures on the autonomous landing of UAVs continue to emerge. Among them, scholars have reviewed general target detection, UAV target detection, UAV autonomous landing, etc. However, the existing literature only briefly describe various methods, lacking a systematic classification and summary of application scenarios [2–4], and there are few studies on autonomous landing of UAVs in complex scenes. In order to further promote the research in the field of UAV autonomous landing, combine the existing algorithms to prepare for the future research work on autonomous landing in complex scenes, it is necessary to sort out and analyze the existing results.

After studying the existing achievements, this paper innovatively classifies the autonomous landing of UAVs based on vision into static scenes, dynamic scenes and complex scenes. According to the different detection targets, static scenes are divided into cooperative target based and natural scenario based. According to the carrier of the moving platform, the dynamic scene is divided into vehicle-based and ship-based. Complex scenes include the selection of safe landing areas for UAVs, vision-based multi-sensor fusion, etc. Figure 1 provides autonomous landing classification of this paper. Finally, we summarize the problems to be solved of the existing achievements in the field, provide solutions and discuss the future development direction.

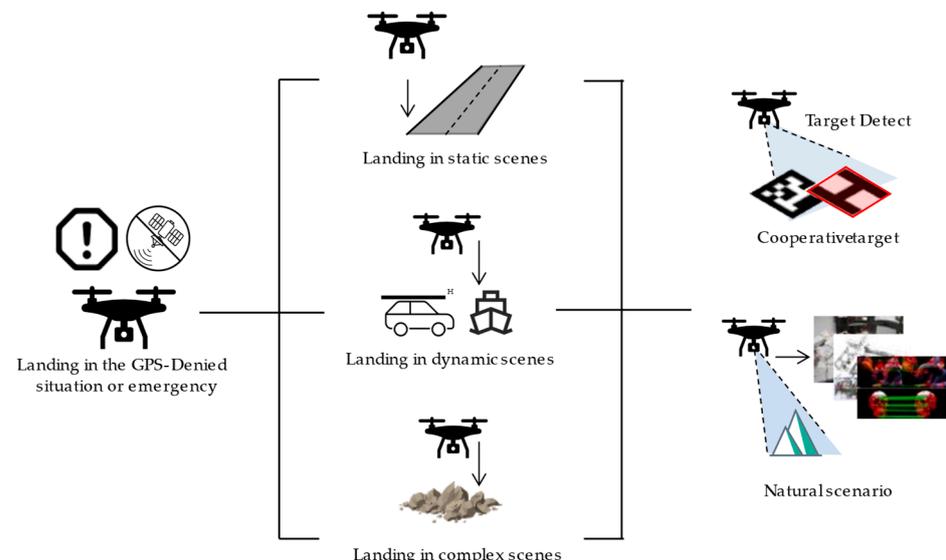


Figure 1. UAV autonomous landing classification.

The rest of the paper is organized as follows. In Section 2, we discuss UAV autonomous landings in static scenes. UAV autonomous landings in complex scenes are expounded in Section 3. In Section 4, UAVs landing in dynamic scenes are provided correspondingly. Section 5 involves the problems to be solved, the future development and gives workable solutions.

2. Autonomous Landing of UAVs in Static Scenes

As the short detectable distance of visual inspection, the built-in map is usually used firstly in engineering for autonomous landing of the UAV. Then, the UAV is guided to the landing area through the global positioning system or inertial navigation system. At last, vision is used for guidance. Vision-guided UAV landing is mainly divided into three stages: target detection, flight guidance and autonomous landing:

- (1) The visual task load and other sensors on the UAV are called to collect and resolve the information from the environment around the landing area, capture artificial landmarks, and perform feature point calculation, feature extraction, feature matching and other operations on the image to achieve landmark tracking and detection. The relative pose information of the UAV and the landmark is continuously returned to the head controller.
- (2) Flight control system corrects and guides the aircraft's position and attitude according to the relative position and missed target amount returned by the vision module in order to make it converge to the target with a relatively appropriate direction and speed.
- (3) Flight control system completes the real-time correction of the landing attitude at high speed, and successfully landed at the target location at last.

In the research field of autonomous landing of UAVs, landing on a static target site or a standard runway and apron is the basis for studying autonomous landing in dynamic scenarios. The autonomous landing of UAVs in static scenes could be divided into two different types: autonomous landing based on cooperative target and autonomous landing based on natural scenario. The concept of scene-based landing is that UAVs can identify environmental characteristics to land in the absence of artificial cooperative targets.

2.1. Cooperative Targets Based Autonomous Landing

The most important step for autonomous landing based on cooperative targets is the detection and recognition of the artificially designed marker. It can accurately identify and solve the current flight attitude of the UAV by extracting the features, so as to realize the guidance of the autonomous landing of the UAV. However, in some special scenarios, cooperation targets cannot be laid manually, especially for scenarios like UAV post-disaster rescue, which puts forward higher requirements for the autonomy of UAVs, demanding UAVs to analyze the surrounding environment autonomously and make correct feedback results at the same time.

2.1.1. Classical Feature-Based Solutions

In feature-based methods, an artificial marker is a kind of identification mark that is designed artificially according to geometric pattern or the principle of certain geometric laws. How to design accurate and efficient identification patterns and applying appropriate feature detection algorithms could be an important way to improve the autonomous landing capability of UAVs. The current mainstream identification marks can be divided into the following categories: "T", "H", circular, rectangular, and combination marks. Some cooperative target images are shown in Figure 2 [5–10].



Figure 2. Partial landing signs.

(1) T shape

In 2006, Tsai et al. used Canny detection [11], Hough transform and Hu invariant moment to detect “T” shape marks, and the pose of the aircraft is estimated based on parallel lines. The root mean square errors of the attitude angles of this method are 4.8° , 4.2° , and 4.6° , respectively. However, the parallel line pose estimation causes great errors when the aircraft experiences larger pitch and attitude changes, and image noise has a significant impact on the accuracy of the algorithm as well.

Aiming at solving the problem that the imaging quality of visual images has a great impact on the recognition accuracy, Xu proposed to use infrared images for target detection, and used the temperature difference between the target and the surrounding environment to eliminate the impact from imaging quality on detection to a certain extent [12]. Using adaptive threshold image segmentation and based on the characteristics that the temperature of the cooperative target is significantly higher than the environment and occupies a relatively small proportion in the field of view, they proposed to use the maximum peak position of the histogram and obtain the threshold by the Otsu method to generate a binary image. For edge processing, they used the Sobel operator, and then obtained the edge information through the chain code and segment table method. At last, the affine invariant moment was used to solve the pose. Experimental data showed that the detection time of the proposed method is 17.2 ms, and the accuracy is 97.2%.

(2) H shape

The H shape is almost the earliest cooperative target in the field of autonomous landing of the UAV. The University of Southern California has taken a lead in studying the autonomous landing system of UAVs. They used the AVATAR unmanned helicopter and realized positioning of the UAV by identifying the target “H” [6]. The proposed system used the Hu invariant moment and combined differential GPS to calculate the attitude and achieved the autonomous landing with a pose error of 4.2 cm and an attitude error of 7° , which is the earliest autonomous landing with fast speed and strong robustness.

Yang used a white H shape cooperative target to perform median filtering to remove noise in signal transmission, then perform image segmentation, and extract the segmented images through a fixed threshold [13]. The simulation experiments showed that a Hu invariant moment does not resolve the position and attitude very well, especially when the landmark is changed in rotation. Therefore, they used Zernike Moments, of which the most important characteristic from its definition, while Hu moments have to obtain such invariance by complex calculation [14].

Invariant moment is an excellent method for solving pose and attitude, but the problem is that the value of a Hu invariant moment of the same cooperative target at different sizes and angles will differ by more than 20% [15]. Meanwhile, many tasks require the UAV and cooperative target to match really well in directions. The orientation is kept corresponding, and the Hu moment invariance effect is not ideal in the case of this specified orientation. Therefore, Zeng designed an H-shaped logo with a triangle, indicating the direction to determine the landing direction [16]. In the detection algorithm, the team first processed and smoothed the image by filtering, Gaussian blurring, etc., and then segmented the processed image through a depth-first search algorithm and compared the searched area with a threshold. According to the truth, features in the H logo cannot be described in a unified way, although it provides a variety of information for the UAV, and the Hu invariant moment is not ideal for this situation, the team proposed an image registration algorithm, which matches the image with the template images. The orientation is calculated with methods such as Hough transform, line segment detection, and Helen’s formula. After experimental verification, the average detection success rate of this method is 97.42%, and the average image detection time is less than 60 ms, which indicates the advantages of strong robustness and high precision. However, the image processing algorithm runs on the ground station, and the image data link wastes a large amount of time. The UAV also has poor anti-signal interference ability when it lands autonomously.

(3) Round, rectangular shapes

Shakernia et al. in 1999 designed a cooperative target consisting of six white rectangles and a black rectangle [17]. They assumed that all feature points are on the same plane and proposed a new geometric estimation scheme. Finally, they discussed a nonlinear controller based on differential flatness. The axial positioning accuracy is 5 cm, and the attitude angle error is within 5° . The simulation shows that the method has low computational cost and is easy to implement. It can achieve stable landing even when the noise level of the image is large. However, the disadvantage is that the feature points have limitations, and the algorithm has low environmental adaptability and robustness.

In 2005, Zheng et al. designed a double-circle cooperation target using eight common tangent points generated by the double circle to generate 21 feature points with perspective projection invariance. They established world coordinates on the cooperative target at the same time. The position information of the coordinate point is calculated based on the segmentation of the ellipse to realize the resolving of the pose. The method has strong anti-interference and anti-noise ability. Simulation experiments show that when the noise deviation reaches 1.5 pixels at a distance of 10 m from the camera, the single-axis error is less than 6 cm, and the single-axis attitude error is less than 0.7° . Moreover, the method has high processing speed. For 768×576 images, its feature extraction and labeling time is less than 9 ms. The disadvantage is that the complexity of the experimental environment is obviously insufficient, which cannot show the robustness of the method under the interference of complex environments and complex backgrounds.

In 2009, Sven designed a landing sign composed of concentric circles with different inner and outer diameter ratios [18]. The cooperative target has adaptability at different heights. By adding more rings based on the ratio of the radius between the rings, the algorithm can achieve detection at higher heights. The team does not rely on any other sensors, but only installs the optical flow sensor to resolve the pose. The experimental results show that the UAV can hover stably for 5 min directly above the cooperative target. The landing error is 3.8 cm, and the maximum error is 23 cm. Considering the error accumulation of the optical flow sensor and the error caused by long-term hovering, this result basically meets the requirements of autonomous landing.

In 2016, Benini et al. designed a cooperative target consisting of two concentric circles and two six circles used for camera pose estimation [19]. After image pre-processing, algorithms, such as noise reduction, the detection algorithm, will be divided into two stages: searching for the main structure of the marker and detecting the inner ellipse. After detection, several groups of ellipses consistent with the characteristics of the cooperative target will be obtained. Noise caused by camera vibration and camera occlusion will decline as the Kalman filtering is processing. The experimental results showed that the minimum frame rate is 30 fps when the size of the image is 640×480 . Regardless of the complexity of the image, the time required for GPU computation is 3 ms, while the error is less than 8% of the diameter of the cooperative target.

In 2018, Yuan designed a system called robust and quick response landing pattern (RQRLP) based on the hierarchical vision detection [20]. The system is divided into three stages: namely "Approaching", "Adjustment" and "Touchdown". During the three stages, the different detection methods of the cooperative target contours were adopted respectively. The markers were extracted and used during the approaching stage in the RQRLP. During the landing stage, optical flow sensors achieved robust pose estimation by tracking the previous position. Meanwhile, a federated filter based on the extended Kalman filter is costumed to integrate these vision solutions. The method is suitable for the landing height of 0–20 m, the error of which is less than 0.0639 m in the range of 20 m. The attitude angle error of the method is less than 0.0818° . Compared with other methods, it has higher recognition distance, stronger robustness, and more computing power which can deal with the image with 1920×1080 resolution.

(4) Combination shape

Combination logos usually refer to the arrangement and combination of standard geometric figures, such as “H”, “T”, circles and rectangles, and also include two-dimensional codes and barcode types of cooperative logos that can carry information. In 2011, the University of Michigan developed an open-source UAV autonomous landing vision system, AprilTag, based on a two-dimensional code combined graphic cooperative target [21]. It is similar in concept to the ARTag system, but has improved performance compared to both ARTag and ARToolkit in the field of autonomous UAV landing [22]. The system is divided into two parts: detector and encoder. The detector utilizes a gradient-based clustering method while applying low-pass filtering to reduce noise to a usable level, and finally uses quaternary detection to determine the location of cooperative targets. The false detection rate of this method is less than 0.3%. Additionally, as the project is open source, it is favored by many researchers. For example, Zhou has conducted research on autonomous landing based on AprilTag [23–25].

In 2016, the University of Michigan team carried out iterative optimization of AprilTag, and proposed AprilTag2. AprilTag2 optimized the tag detector and used adaptive thresholds to reduce random noise interference. It improved computational efficiency, and adopted continuous boundary segmentation to provide more accurate boundaries information, and used edge refinement to improve the accuracy of pose estimation at last [26]. These optimizations not only improve efficiency and robustness, but also reduce false positives.

In 2013, Yang et al. studied the landing of UAVs under complex backgrounds [27]. They designed a combined logo with H placed in a ring. Using the method of projective geometry, the five degrees of freedom of a UAV's pose were calculated by identifying ellipses. Another degree of freedom is calculated using H. This method can achieve accurate identification in the presence of various edge information and can distinguish it from a single ring and H logo. The landing error is 2.4 cm and 8.6 cm in the two axes, respectively. The attitude angle error is 6° .

In 2017, Nguyen et al. designed a three-circle nesting composition [28], which is divided into eight parts at the same time. The image processing speed of this method is about 40 ms. At a height of 10 m, the errors of the three axes are 7.6 cm, 1.4 cm, and 9.5 cm, respectively, and the attitude angle errors are 1.8° , 1.15° , and 1.09° , respectively.

In 2018, Zhang et al. proposed a UAV autonomous landing system based on hierarchical identification method and designed a cooperative identification based on multi-layer nested two-dimensional coding [29]. Through the identification of nested QR codes, the system maintained a good positioning effect at the multi-scale level. At the same time, the experimental results showed that the system has good robustness and anti-interference ability of environmental factors, such as noise and temperature, while its requirement for computing capability is low. However, it only relies on the stratum identification for positioning, lacking other environmental information.

2.1.2. Machine Learning-Based Solutions

A feature-based solution needs different feature extraction methods for different cooperative targets, and it is difficult to accurately extract the deep features of cooperative target images. Therefore, traditional feature-based methods meet a theoretical choke point in environmental adaptability and algorithm robustness. Since machine learning-based methods are able to learn high-level semantic features from a large amount of data, they have been widely studied and applied to the field of object detection. Machine learning-based target detection methods can be further divided into two categories: classifier-based methods and deep learning-based methods.

(1) Classifier-based methods

The classifier-based identification detection is similar to the target detection tasks, which is a method combining sliding window and machine learning. For the image in each window, the features are extracted as the input of the classifier, the training result is

obtained afterwards as the final learning model. The commonly used classifiers include Support Vector Machine, AdaBoost, K-Nearest Neighbor, etc. [30–32]. This method has high accuracy under certain circumstances, but it still relies on manual feature extraction. Different feature extraction methods need to be designed for different identifications in different application scenarios, so the method also lacks robustness.

In 2012, based on the cooperative target composed of six circles, Li et al. extracted affine invariant moments as input features, and used an SVM classifier for detection [33]. For undistorted and distorted images, the classification accuracy was 98.25% and 92.1%, the recognition time of a single image was 7×10^{-3} ms. Compared with the traditional geometric invariant moment and BP neural network, it has extraordinary real-time performance, and the required computing power is relatively low. However, the portability of it is not strong, and environmental factors cause great interference to it.

In 2014, Verbandt used the SVM classifier to detect the cooperative target and resolved the pose through the Hough transform. The experimental data showed that the landing error was within 5 cm, and it had a robust performance under different ambient lighting and sharpness.

(2) Deep learning-based methods

The basic idea of the identification detection algorithm based on deep learning is to optimize the target detection algorithm from the perspective of detection speed, detection accuracy and model lightweight, and integrate dehazing, deblurring and other algorithms to adapt to complex scenes in practical applications, achieving high precision and real-time identification detection in the end to meet the original requirements. Compared with traditional feature-based detection algorithms, deep learning-based detection algorithms have good generalization, less dependence on specific logos, high robustness, and broad application prospects [34,35].

In 2017, Chen et al. used faster regional neural networks (Faster R-CNN) to detect cooperative targets and realize autonomous landing of the UAV [36]. After detecting the cooperative target, the least squares ellipse fitting and Shi-Tomasi corner were adopted for pose calculation. The R-CNN method proved that it is not only capable of extracting features from the color and texture level of the target, but also extracting high-level features from multiple color channels. Even if there is noise interference, the accuracy of recognition can be guaranteed. Experiments showed that the accuracy of the SVM and BP network is greatly reduced in the case of image distortion. Compared with the YOLO algorithm, which is also a convolutional neural network, Faster R-CNN has higher accuracy because of RPN. In the case of image distortion or no distortion, the recognition accuracy can reach 99.2% and 97.8% respectively, and the average detection time is 0.081 s, which meets the requirements of real-time and accuracy and has high robustness. Within 5 m height, the error of x, y, z axis was not more than 1.5 cm in the position estimation, and the error of the direction estimation was within three degrees.

In 2018, Nguyen et al. replaced the original adaptive template matching with LightDenseYOLO and used Profile Checker to further improve the accuracy rate at the same time [37]. LightDenseYOLO is divided into two parts, one is the feature extraction network, the other is the label detection module YOLOv2. The experimental results showed that the new model inherits the good feature extraction effect of the LightDense network and has both the high real-time performance and robustness of YOLO. It runs at a speed of 50 ms on a desktop computer and about 20 fps on a Snapdragon 835 processing platform. In the process, the time required for contour detection is only 10 ms. If traditional methods, such as Hough transform, are used, the detection time will be longer.

In 2020, Noi Quang Truong et al. proposed a two-phase framework of deblurring and object detection, by adopting a slimmed version of the deblur generative adversarial network model called SlimDeblurGAN and a you only look once version 2 (YOLOv2) detector, respectively [38]. It considered the performance of a combination of motion deblurring and marker detection for autonomous UAV landings. The processing speed of the SlimDeblurGAN algorithm is 54.6 fps, and the total processing speed reaches 20.3 fps.

Using this method can improve the robustness of the system, but the disadvantage is that the calculation complexity is high.

2.2. Natural Scenario Based Autonomous Landing

The autonomous landing of UAVs based on the natural scenario is evidently different from that of the cooperative target. Specifically, the method based on the cooperative target can perform accurate positioning and navigation according to the markers that are highly differentiated from the background. The algorithm design can be custom-made to match traditional features. However, the markers must be set up in advance, otherwise it will not be recognized. In some scenarios with high requirements for UAV autonomy, such as rescue missions, UAVs often arrive at the scene before humans and perform preliminary task processing. In this situation, the method of placing cooperative markers in advance do not satisfy the mission requirements. Therefore, scene-based autonomous landing has been developed. The existing technologies in this field include autonomous landing based on scene matching [39], which is autonomous landing based on SLAM 3D scene modeling

2.2.1. Scene Matching-Based Solutions

A scene matching navigation system has the characteristics of simple structure, passive type, high positioning accuracy, etc. It uses image sensors to obtain regional images near the flight or target area and match them with the stored reference images to obtain aircraft position data. Different from other active navigation systems, scene matching navigation can be combined with an inertial navigation system to form a highly autonomous and high-precision navigation system as an auxiliary navigation method. It can make the UAV achieve precise autonomous landing in specific scenarios.

In 2008, Gianpaolo Conte et al. researched on the problem of scene matching correctness in a UAV autonomous landing system and developed a method showing how to detect incorrect image registration [40,41]. Its vision system uses image registration and visual odometers to work stably in unstructured environments. At the same time, a multi-sensor fusion vision-aided architecture is used, the core of which is the Kalman filter, inertial and position sensors. The image registration technique developed here is based on the Sobel edge detector, which is robust to illumination changes as well as geographic feature changes. In the experiment, the vehicle flies at a speed of 3 m/s at an altitude of 60 m, and the final landing error was about 3 m. The system can stably guide the autonomous landing of the UAV.

In 2008, Andrew et al. built a map reference database of the runway on the UAV [42]. After the UAV landed on the expected area roughly, the scenario information was matched to the information collected from the processor on the UAV to provide distance parameters and attitude angle parameters between the UAV and the runway. The advantage is that it does not depend on the model of the UAV, the parameters of the camera or the style of the landing runway and does not require the support of cooperative targets. The weakness lies in a large angle deviation when a UAV approaches the runway, which is easy to cause harm to flight safety.

In 2010, Cesetti et al. researched on a UAV safe landing system based on natural landmarks [43]. The operator can define the target area or navigation point path from high-resolution satellite or aerial images and use the airborne scene feature matching algorithm to control the system to land autonomously. Moreover, using the optical flow method to construct a sparse terrain map, two detection methods were proposed for safe landing: one to use SIFT features to estimate optical flow, and a simple classifier with binary thresholds is used to determine whether the surface is flat and has conditions for landing; the other to observe whether the target features change linearly during the vertical landing process. The proposed method did not require cooperative targets, and experimental results showed that it was robust to illumination changes and occlusions.

In 2012, Northwestern Polytechnical University realized the autonomous landing of UAVs based on the key frame method of natural landmarks. The researchers used

image feature extraction technology to autonomously extract key frames containing natural landmarks in real-time image sequences, and innovatively gave up the feature matching between real-time images and benchmark images. Through a detection mode based on the inter-frame image matching technology, they used the dynamic key frame images calculated between real-time images to accomplish “relative” scene matching, which overcomes the difficulty of tracking and matching image features. At the same time, a dynamic key frame management mechanism [44] is proposed on the basis of this method [45]. This method can greatly reduce the cumulative error caused by long-term video mosaic without increasing the computational complexity, and it is robust to image blur and noise interference.

In 2015, Wang et al. proposed a fast scene registration algorithm based on FREAK features to solve the problem that the scene matching processing of traditional local invariant features took too much time. The FAST-Difference method is used to extract feature points, and the FREAK descriptor is used to calculate the feature vector. Finally, the RANSAC algorithm is used to eliminate the mismatched points, and the least squares method is used to calculate the spatial geometric transformation parameters between the two images. The experimental results showed that compared with the classic SIFT and SURF algorithm, the detection speed is greatly improved with 38 ms under the image size of 235×472 . The detection accuracy is comprehensively improved, and the robustness is enhanced at the same time.

2.2.2. Near-Field 3D Reconstruction-Based Solutions

Importantly, 3D reconstruction means that the UAV platform relies on its own sensors and poses information to obtain the surrounding 3D environment data. Due to the particularity of the UAV platform and the autonomous landing task, the algorithm that realizes the positioning function in the autonomous landing system can be narrowly classified as the SLAM algorithm. In the existing research, the UAV is generally navigated to the vicinity of the target through its own inertial navigation or global positioning system, and then guided to land by the SLAM algorithm. Since most of the SLAM algorithms used in autonomous landing tasks are used for near-field 3D reconstruction and positioning, they are defined as near-field 3D reconstruction.

Visual SLAM mainly includes two parts: front end and back end. The front end is responsible for grabbing sensors and pose data and estimating the state of the UAV. The back end optimizes the data generated by the front end, and performs loop closure detection [46]. At the same time, with the continuous rise of artificial intelligence and deep learning, artificial intelligence technology is constantly being combined with SLAM technology, which has achieved remarkable results.

The multi-sensor information fusion research carried out by the GRASP Laboratory of the University of Pennsylvania realized the accurate indoor and outdoor environment positioning and modeling of UAVs [47]; ETH Zurich studies indoor precise positioning algorithms for multi-rotor UAVs [48]; the Vision Laboratory of Munich University of Technology is engaged in research on V-SLAM algorithms and 3D environment reconstruction for multi-rotor UAVs [49].

Shen Shaoxie researched the fusion strategy of aircraft V-SLAM and IMU, and proposed the VINS-Mono algorithm and the MVDepthNet algorithm for monocular depth estimation [50,51], which can overcome the inherent characteristics and disadvantages of monocular V-SLAM. At present, it has been widely used on multi-rotor UAV platforms. Wang et al. of the National University of Singapore proposed a complete set of UAV navigation systems based on visual optical flow and laser SLAM [52]. The main idea is to combine IMU with tachymeter data to robustly estimate the speed and position of the UAV.

In 2021, Cui proposed a precise landing method based on binocular SLAM [53]. The team simplified the traditional SLAM framework, which contains front end sensor information fusion and back end nonlinear optimization. In the improved SLAM, the ORB algorithm was used to extract a fixed number of features, and the GMS strategy based on the assumption of motion smoothness was used to evaluate the detection results. Based on

the results of 100 actual scene experiments, the maximum error of this method is 1 m, while the error of the traditional SLAM method is 4 m, which obviously improves the accuracy.

In 2018, Yang studied UAVs landing in unknown areas based on monocular SLAM in emergency situations [54]. The system proposes a new map representation method that combines 3D features with intermediate processes to remove noise and build grid maps with different heights, combines 3D point clouds with grid maps and computes each height of the grid, which can be used to build a visible grid map. At the same time, in order to improve the speed and accuracy of UAV landing area recognition, in accordance with image segmentation based on the mean shift principle, the grid map is smoothed and divided. At last, the height information and obstacle information of the ground are obtained. Then, the shortest path algorithm is used for planning, and the UAV is navigated to the landing destination. The effectiveness and robustness of the method have been demonstrated in multiple real experimental scenarios. The disadvantage is that image processing and SLAM are conducted at the ground station, which may have large signal interference and delay.

In 2020, Li used truncated sign function to model the landing area in real-time, generate low noise depth images, and automatically analyze the terrain for landing. The depth camera achieves noise reduction and meets the requirements of autonomous landing.

In order to facilitate the research and comparison of the detection methods adopted by various research institutions, the literature related to static scenes has been sorted out and summarized in Table 1.

Table 1. Comparison of different autonomous landing solutions in static scenes.

Target Type	Method	Precision	Height	Image Resolution	Processing Speed	Speed
T	[11] 1 Canny detection 2 Hough transform 3 Hu invariant moment	Pose 4.8°	2–10 m	737 × 575	25 Hz	N/A
	[12] 1 Adaptive threshold selection method 2 Infrared images 3 Sobel method 4 Affine moments	Success rate 97.2%	N/A	N/A	58 Hz	N/A
H	[6] 1 Hu invariant moment, 2 Combines differential GPS	Position 4.2 cm Pose 7° Landing < 40 cm	10 m	640 × 480	10 Hz	0.3–0.6 m/s
	[13] 1 Images extract, 2 Zernike Moments	Position X 4.21 cm Position Y 1.21 cm Pose 0.56°	6–20 m	640 × 480	>20 Hz	N/A
	[16] 1 Image registration, 2 Image segment, 3 Depth-first search 4 Adaptive threshold selection method	Success rate 97.42%	N/A	640 × 480	>16 Hz	/N/A
Round Rectangular	[17] 1 Differential fitness 2 New geometric estimation scheme	Position 5 cm Pose 5°	N/A	N/A	N/A	N/A
	[18] 1 Optical flow sensor 2 Fixed threshold 3 Segmentation 4 Contour detection	Position 3.8 cm	0.7 m	640 × 480	70–100 Hz	0.9–1.3 m/s
	[19] 1 Kalman filtering 2 Function ‘solvePnP/Ransac’	Position error less than 8% of the diameter of the cooperative target.	2.5 m	640 × 480	30 Hz	N/A
	[20] 1 Robust and quick response landing pattern 2 Optical flow sensors 3 Extended Kalman filter	Position 6.4 cm pose 0.08°	20 m	1920 × 1080	7 Hz	N/A

Table 1. Cont.

Target Type	Method	Precision	Height	Image Resolution	Processing Speed	Speed
Combination	[21] 1 QR code digital coding system 2 Gradient-based clustering method 3 Quad detection	(<40 m) Landing position <0.5 m Success rate > 97%	50 m	400 × 400	30 Hz	N/A
	[26] 1 Coding scheme 2 Tag boundary segmentation method 3 Image gradients 4 Adaptive thresholding	Position < 0.2 m Pose < 0.5°	0–20 m	640 × 480	45 Hz	N/A
	[27] 1 Projective geometry 2 Adaptive thresholding 3 Ellipse fitting 4 Image moments method	Position (2.4, 8.6) cm Pose 6°	10 m	640 × 480	60 Hz	N/A
	[23] 1 AprilTags 2 HOG 3 NCC	Position < 1%	4 m	N/A	N/A	0.3 m/s
	[28] 1 Profile-checker algorithm 2 Template matching 3 Kalman filtering	Position (7.6, 1.4, 9.5) cm pose (1.8° 1.15° 1.09°)	10 m	1280 × 720	40 Hz	N/A
	[29] 1 Canny 2 Adaptive thresholding 3 Levenberg–Marquardt (LM)	Position < 10 cm	3–10 m	640 × 480	30 Hz	N/A

3. Autonomous Landing of UAVs in Dynamic Scenes

In the process of UAVs performing complex tasks, the target is generally moving. At the same time, the realization of the technology of landing on moving targets also lays the foundation for the autonomous landing of UAVs in complex scenarios. Compared with static scenes, the complexity of dynamic scenes is much higher, which requires increasing levels of the navigation system and control system of UAVs. Meanwhile, the mission objectives of the two types of landing are almost the same, which means the autonomous landing methods mentioned above in static scenes are also widely used in dynamic scenes.

Summarizing the previous research results, we found that the autonomous landing of UAVs on moving platforms can be roughly divided into the following directions: external auxiliary equipment guidance, vision-based navigation, multi-sensor data fusion, and emerging learning-based control method, which has also been developed rapidly. Due to the complexity of the moving platform, the current target detection and recognition algorithms cannot reliably identify and track the characteristics of the landing platform itself. Therefore, most of the existing research adds cooperative targets and achieves autonomous navigation by identifying the cooperative targets with distinctive features. Based on the different mission scenarios, the achievements will be divided into vehicle-mounted platforms and ship-borne platforms for research.

3.1. Autonomous Landing on Vehicle-Based Platform

In 2013, Cheng et al. made use of the LED luminous circular markers on the vehicle platform to guide the UAV to land [55]. In this system, the UAV sent the collected image information to the ground station first, then the ground station instructions to the UAV after image processing and pose calculation. Finally, the UAV controlled the position through closed-loop PID to land. Experiments show that in 15 autonomous landing experiments, the success rate is 88.24% when the landing platform moves at a speed of 1.2 m/s. The error of the landing position is 5.79 cm and 3.44 cm in the x-axis and y-axis, respectively. The team has successfully realized the functions of autonomous take-off, tracking and landing, proving robustness in dealing with complex environments. The disadvantage of this system is that the wireless image transmission is sometimes disturbed, resulting in

misjudgment caused by image noise. The wireless transmission has a certain delay, which means it is not conducive to high-speed autonomous landing.

In 2017, Baca et al. proposed a fast and robust visual localization method in the International Robot Competition for UAV autonomous landing on the vehicle [9]. The adaptive thresholds were used for image segmentation and detection and the detected positions of the car were filtered using an unscented Kalman filter-based technique with an assumed car-like model of the vehicle. The model predictive control was used to track the vehicle and estimate its future trajectory. With the platform moving at 15 km/h the system landed within 25.1 s and won the first place in the competition. The method was tested in a variety of environments, such as grass, snow, and concrete surfaces. When it was tested at a wind speed of 10 km/s, experimental results indicated the robustness.

To enhance the accuracy of UAV autonomous landing on a vehicle, Davide et al. estimated relative position using circle detection and cross detection based on the cooperative target designed by themselves in 2017. The Kalman filter algorithm was used for data filtering, and the PnP algorithm and inertial measurement were used to assist in the calculation of the UAV position. With these methods, the stable landing of the moving platform at a speed of 1.5 km/h was realized.

Relying solely on vision-guided autonomous landing is limited by algorithms and computing power; therefore, some studies focus on vision-led multi-sensor fusion technology to guide the autonomous landing of UAVs.

In 2016, Chen et al. designed a tracking and landing of moving platforms algorithm based on vision and laser ranging [56]. Cooperative targets were nested with rectangles and circles of different colors, then detected by color threshold method. In the cooperative objective-based landing process, the most difficult phase is the accurate altitude measurement of the UAV when it is over the horizon or distance is too close. In this system, the monocular vision system was used for detection and pose calculation, and the lidar ranging sensor is used to measure the relative height of the UAV. The UAV successfully landed on a platform with a moving speed of 1 m/s, and it could still accurately detect and estimate the pose even when the cooperative target was largely occluded. The disadvantage is that when the moving platform suddenly accelerates or decelerates, the UAV may have a large overshoot, so it can only be applied to the slow speed of the moving platform.

In 2017, Araar et al. combined landing markers of different sizes to achieve accurate identification and detection of cooperative targets at different heights [57]. To remove image noise, they proposed to fuse extended Kalman filter and extended H_∞ to improve the efficiency and accuracy of detection. They fused inertial measurements and pose estimation from vision to improve the sampling rate and adaptability in the case of short-time interference. Verified by simulation and real scenarios, the UAV was capable of landing on a moving platform of 50 km/h with the error of 8 cm.

In 2018, Yang et al. proposed a relative position estimation method based on binocular vision [58]. Meanwhile, based on the improved Yolov3 algorithm, the real-time detection and tracking of the target was realized. Compared with the results of the traditional algorithm, the accuracy and real-time performance were improved, and the robustness and stability of the system in the complex surrounding environment also took a giant leap.

In 2019, Alejandro Rodriguez-Ramos et al. designed a visual landing system based on a deep learning algorithm [59]. The system is based on deep learning algorithm DDPG, which represents a success in the application of neural networks in reinforcement learning. The algorithm was a policy-based deep reinforcement learning algorithm designed for continuous state space and action space. The team conducted experiments on artificial noise, and the experiments showed that the maximum landing time is 17.67 s, and the maximum axial error is about 6 cm, maintaining good accuracy, which can prove the anti-interference and robustness of the system positioning.

In 2020, Cai et al. researched on the adaptive moving target landing of UAVs based on the AprilTag cooperative target system. The team proposed a monocular vision algorithm and depth estimation strategy based on deep learning. In the case of medium and long

distances, UAVs used YOLOv3-Tiny to identify and judge small targets and track their three-dimensional positions [60]. An algorithm integrating YOLOv3-Tiny and TLD was proposed for tracking, which solves the problem of real-time and long-term tracking accuracy decline of classical algorithms through CPU and GPU hybrid processing. When approaching the UAV, the AprilTag cooperative target was used to locate and land, and the UAV's flight trajectory was output by adopting the measure of Kalman filtering and altitude estimation.

3.2. Autonomous Landing on Ship-Based Platforms

The movement of the vehicle platform can be abstracted into two-dimensional movement in some way. However, on the more complex and changeable sea, the ship is affected by the environment with the three-dimensional movement characteristics. Therefore, it requires the UAV to resolve the six degrees of freedom relative position information of the ship and the UAV itself in real-time, and that requires extremely high speed and accuracy of the UAV system. Furthermore, the water mist and weather phenomena that occur on the sea surface have a great impact on the imaging quality of the visual sensor as well, which greatly increases the difficulty of autonomous landing on carrier-based platforms.

In 2003, Qiu conducted autonomous landing of unmanned helicopters based on the binocular stereo vision system, which initially met the practical requirements. Since the vision system was easily disturbed in the maritime environment, it cannot complete tasks around the clock. In 2009, Xu installed a heating module on the "T" logo, which can overcome the influence of water mist to a certain extent [12]. Using infrared cameras to achieve autonomous landing at night, the recognition speed of the UAV can reach 17.2 ms, and an accuracy rate of 97.2%. Aiming at the influence of the environment and waves during the landing process, in 2014, Sanchez-Lopez proposed to use Hu invariant moment features and decision tree discriminant units, including multi-layer perceptrons, and other geometric properties to detect international common landing signs [61]. The experiment showed that the method can well recognize the target at different heights and under certain occlusion, and it was not sensitive to illumination.

In 2015, Portuguese researchers proposed a method based on airborne cameras to realize the landing of UAV on ships at sea during the movement process [62]. The UAV detected the guidance signs on the ship through an on-board RGB camera. The relative position, velocity and attitude of the UAV and the visual guidance signs on the ship were estimated by Kalman filter and efficient perspective-n-point method. By combining the information, the flight trajectory of the UAV during the landing process was generated at last. The deficiency lies in the misjudgment caused by reflections on the water and other situations.

In 2016, the British company Roke Manor developed AutoLand Technology in partnership with the Defense Science and Technology Laboratory. The system can autonomously identify obstacles on board for autonomous landing without GPS system guidance. At the same time, the company tested it in different scenarios and sea conditions, and the system performed well.

In 2017, Polvara proposed an autonomous landing scheme for the case where the landing scene is the deck of a ship [63]. The scheme used the cooperative target to obtain the relative pose of the UAV as a six-degree-of-freedom, which combined extended Kalman filtering techniques, inertial navigation data and used the position of the target platform at one moment to estimate the current position. It can obtain a stable and accurate cooperative target, tracking results and detection results. However, the scheme was not tested in real environment, so its availability is unknown. In the following year, the team designed a vision-based system for autonomous landing of a quadrotor UAV on a disturbed deck, and simulated experiments were carried out in complex environments, such as adverse weather. The system performed well and was robust in the simulated environment.

In 2019, the Huazhong University of Science and Technology team realized the autonomous take-off and landing of the UAV on the autonomous unmanned boat HUSTER-68

during the navigation. The cooperation target adopted a multi-circle nesting method, which can ensure that the UAV can identify targets accurately and efficiently at various heights.

In 2020, Li et al. proposed a new feature extraction structure for the problems of small field of view and large image scale changes in unmanned helicopters [64]. The algorithm used the mutual correction of SSD algorithm and KCF algorithm and took the advantages of deep learning algorithm into account. With the advantages of high precision and strong real-time filtering algorithms together, the proposed method has a detection success rate of 91.1% and an average processing speed of 9 ms, which meets the mission requirements of UAV autonomous landing.

At present, in the field of autonomous landing in dynamic scenes, researchers mainly focus on overcoming the three-dimensional motion characteristics of the moving platform and the cooperative target recognition caused by the environment. They have made significant progress in detection accuracy and real-time performance. However, most of the research is aimed at the cooperative target rather than the platform structure, and it is difficult for the current vision algorithm to efficiently solve the relative pose without relying on the cooperative target. Extending the algorithm to the detection of the whole platform will effectively improve the detection distance and effectively enhance the adaptability and robustness of the vision algorithm.

We summarize the various research results of autonomous landing of UAVs in dynamic scenes, as shown in Table 2.

Table 2. Comparison of different autonomous landing solutions in dynamic scenes.

Type	Method	Precision	Hight	Image Resolution	Processing Speed	Platform Speed
Vehicle-based	[55] 1 Hough transform 2 Adaptive thresholding 3 Erosion and dilation 4 Visual-Inertial Data Fusion	Position (5.79, 3.44) cm Successful rate 88.24%	1.2 m	640 × 480	25 Hz	<1.2 m/s
	[9] 1 Model predictive controller 2 Nonlinear feedback controller 3 Linear Kalman filter	Landing in 25 s Position error < 10 cm	N/A	752 × 480	30 Hz	4.2 m/s
	[65] 1 Visual-inertial odometry 2 Extended Kalman filter	N/A	3 m	752 × 480	80 Hz	<1.2 m/s
	[56] 1 LiDAR scanning 2 PnP 3 Adaptive thresholding	Landing in 60 s Have larger overshoot when backward	3 m	N/A	N/A	1 m/s
	[57] 1 Extended Kalman filter 2 Extended H_{∞} 3 AprilTag Landing Pad 4 PnP 5 Visual-Inertial Data Fusion	Position < 13 cm	2 m	N/A	N/A	1.8 m/s
	[58] 1 State estimation algorithm 2 Nonlinear controllers 3 Convolutional neural network 4 Velocity observer 5 Nonlinear controller	Position < (10, 10) cm	1.5 m–8 m	512 × 512	N/A	1.5 m/s
	[59] 1 Gazebo-based reinforcement learning framework 2 Deep deterministic policy gradients	Landing in 17.67 s Position error < 6 cm The action space does not include altitude	N/A	N/A	20 Hz	1.2 m/s
Ship-based	[12] 1 Afine Invariants moment 2 Infrared radiation images 3 Otsu method and iterative method 4 Adaptive thresholding	Success rate 97.2%	N/A	N/A	58 Hz	N/A
	[61] 1 Kalman filter 2 feature matching 3 Image threshold 4 Artificial neural networks 5 Hu moments	Position error (4.33, 1.42) cm	1.5 m	640 × 480	20 Hz	N/A
	[62] 1 Kalman filter 2 Efficient perspective-n-point (EPnP)	N/A	N/A	N/A	N/A	N/A
	[63] 1 Extended Kalman filter 2 Visual-Inertial Data Fusion	Landing in 40 s	3 m	320 × 240	N/A	N/A

4. Autonomous Landing in Complex Scenes

The static scene and the dynamic scene can be collectively concluded as the simple scene, which is characterized by a clear autonomous landing task objective, and each sensor works well in the experiment under ideal conditions. Although dynamic scenes are more complex, they still rely too much on the method of setting up manual calibration of the exact location of the landing, which cannot meet the task goal of autonomous landing on complex surfaces, such as post-earthquake disaster areas, complex mountains, and hills. These target areas often require UAVs to arrive before people and detect or release loads in the area. Moreover, due to the limited passage of people, it is impossible to accurately place cooperative targets to guide the landing. The complex scene is defined as the scene that is mentioned above, with multiple tasks or complex landing environments. Autonomous landing in a complex scene refers to the use of limited sensors to achieve autonomous landing in a designated non-flat or moving with unpredictable speed area to complete the target task. It can be seen that the concept of the landing area is different from a landing target point. It requires the UAV to have the ability to autonomously identify and select a landing area, and it also puts forward high demands for the fusion of UAV sensor data and the accuracy together with adaptability of the algorithm. At present, there is relatively little research on autonomous landing of UAVs in complex scenarios. This chapter will briefly illustrate the existing achievements.

Garcia-Pardo pointed out that the key problem of visual aided UAV landing in an unknown area is how to find a safe area suitable for landing [66]. The safe area needs to satisfy two conditions: one is to meet the size of the UAV landing area, the other is no obstacles in the landing area. In addition, an important assumption is that the edges of obstacles in the image have higher contrast than that of flat areas. Based on this assumption, a fixed threshold was set for the aerial landing area image for segmentation in the experiment. The part with the contrast above the threshold was regarded as an obstacle area, or it was seen as a flat area. Then, a circular area suitable for the landing of the UAV was selected from the area lower than the set threshold. Finally, the feasibility of the scheme was verified by the UAV flight experiment. The disadvantage is that it has not been tested in a complex textured environment, and the landing center point is easy to shift. In addition, the influence of ground physical properties on landing is not considered, and the identification of landing relies on the setting of thresholds, which is not robust.

Fitzgerald and Mejias adopted a UAV emergency landing area selection method based on a monocular camera combined with a digital elevation model [67,68]. The steps of autonomous landing of UAV were primary selection of landing area, identification of landing candidate area, flatness analysis using DEM, and comprehensive decision-making. Although this method performs well in the designed scene, it only relies on the Canny operator for edge extraction, which means the algorithm has limitations. In the flat estimation stage, it only relies on DEM calculation, lacking robustness.

Multi-sensor fusion technology adds a variety of data information compared with visual algorithm, such as depth, pose, etc. Moreover, the stability and accuracy of the technology are stronger. Among them, the combination of vision algorithm and depth camera is used mostly. The three-dimensional information of the environment is constructed through depth information, so as to accurately search and determine the area that can be landed. The combination of vision algorithm and lidar also constructs a complete dense three-dimensional map that is applied to complete the landing.

Scherer proposed a method to select the unknown landing zone using lidar combined with monocular camera. The method is mainly divided into two steps: rough assessment of the landing zone and fine assessment of the landing zone. First, the entire landing area was divided into several units according to the size of the landing area required by the UAV. Then, the height, mean, and variance were obtained from the point cloud data measured by lidar in each unit. After that, the appropriate landing candidate area was selected by setting the threshold. Then, it was assumed to perform surface fitting on the cloud point map of the candidate area to obtain a more accurate terrain and finally complete the landing.

The method has been tested on a variety of terrains, and it has good performance and robustness. However, the cloud point data obtained by lidar are prone to interference, resulting in noise, which affects the accuracy of the cloud point map.

Fan et al. of the Chinese Academy of Sciences completed the three-dimensional reconstruction through vision algorithm and used statistical algorithms to find the safe landing area from the reconstructed point cloud image. The method was based on 3D scene recognition with high accuracy and reliability. It can be seen that the method based on 3D scene reconstruction will definitely become the mainstream of research in this field. However, its disadvantage is that its large amount of calculation is not conducive to the real-time understanding of the scene by the UAV. Therefore, Huang proposed a fast point cloud segmentation and flat area recognition method based on the geometric features of the point cloud. This method filters the point cloud, and then uses the improved RANSAC to fit the point cloud surface to screen out the flat area suitable for landing. After experimental verification, the terrain relief error identified by this method is less than 0.125 m, which satisfies the requirements of the autonomous landing requirements of the aircraft.

In addition, Yang studied the unstructured emergency autonomous landing of UAVs based on SLAM in the case of missing GPS signals in 2018 [54]. In this system, monocular vision SLAM was used to construct cloud point maps and locate the UAV. Then, they used a new map based on 3D features and mid-pass filter to denoise and build grids at different heights, and finally divided the grid to calculate a safe landing area. The experimental data showed that the landing starts at a height of 20 m, and it takes 52 s. The experimental verification has been carried out in various complex scenes, which proved the robustness of the system. At the same time, the sparse point cloud can be divided into different heights to meet the needs of autonomous landing.

Li built a system that can land on complex and rough surfaces. Using a binocular RGB-D depth camera as a depth sensor and a truncated signed distance function to conduct real-time 3D modeling of the landing area, they successfully achieved the low noise image which meets the needs of landing. At the same time, in order to meet the landing planning under the condition of limited computing power, the team designed a landing search algorithm consisting of coarse screening, layering, scoring, mapping, search and decision-making. The simulation and real flight test were conducted with satisfactory results. After calculating the landing point, the actual landing error is 7.4 cm at most. The disadvantage is that they did not operate more experiments under different environmental conditions to verify the robustness of the system.

5. Summary and Suggestion

Vision-based navigation has attracted wide attention due to its advantages of rich scene information, strong anti-interference ability, high accuracy, and low cost. Vision-based autonomous landing for UAV uses the real-time target and environmental information as the basic data, which are then processed by onboard computer, and provide the position and attitude for the decision-making and control system, so as to guide the UAV to land autonomously in static scenes, dynamic scenes and complex unknown environments. Therefore, vision-based autonomous landing has become one of the research hotspots for decades, and it has been widely utilized in military and commercial fields. However, there are still several key problems to be solved in vision-based autonomous landing technology:

Marker design and detection in complex scenes. In complex landing environments, the target maker may be disturbed by other objects, or it would be hard to detect the marker. Therefore, how to design the target landing marker to be easy to find, and the robust target detection algorithms adaptable to the complex environment will be a future research field. Color feature is an alternative solution used as a cooperative target marker. It is more recognizable than shape feature, especially in complex landing scenes where many disturbance objects exist, and it would be hard to find the landing marker without color features.

The ability of anti-interference environment perception. In complex tasks, such as rescue, battlefield environment perception, explosive removal, etc., which may be performed together with autonomous landing task, visual perception faces the challenge of smoke and fire, and the UAV needs to locate explosion source, a person, or damaged equipment with huge appearance changes in color, texture, and shape. Autonomous landing in post-earthquake ruins, in dense forest rescue or other challenging environments is also necessary. In these situations, it would be a great challenge to identify the landing area and guide the UAV to land. Anti-interference environment perception is an important but difficult ability for the UAV autonomous landing. Fortunately, the further improvement of artificial intelligence technology and machine learning technology provides a feasible solution for the functional demand. The improvement of the image processing capability of the hardware makes it possible for real-time and robust computation of the artificial intelligence related perception algorithms.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Gautam, A.; Sujit, P.B.; Saripalli, S. A survey of autonomous landing techniques for UAVs. In Proceedings of the 2014 International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA, 27–30 May 2014; pp. 1210–1218. [\[CrossRef\]](#)
2. Kong, W.; Zhou, D.; Zhang, D.; Zhang, J. Vision-based autonomous landing system for unmanned aerial vehicle: A survey. In Proceedings of the 2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI), Beijing, China, 28–29 September 2014; pp. 1–8. [\[CrossRef\]](#)
3. Yang, Z.; Li, C. Review on vision-based pose estimation of UAV based on landmark. In Proceedings of the 2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST), Shenzhen, China, 14–16 April 2017; pp. 453–457. [\[CrossRef\]](#)
4. Chen, P.; Zhou, Y. The Review of Target Tracking for UAV. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 1800–1805. [\[CrossRef\]](#)
5. Sharp, C.S.; Shakernia, O.; Sastry, S.S. A vision system for landing an unmanned aerial vehicle. In Proceedings of the 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164), Seoul, Korea, 21–26 May 2001; Volume 1722, pp. 1720–1727. [\[CrossRef\]](#)
6. Saripalli, S.; Montgomery, J.F.; Sukhatme, G.S. Vision-based autonomous landing of an unmanned aerial vehicle. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), Washington, DC, USA, 11–15 May 2002; pp. 2799–2804. [\[CrossRef\]](#)
7. Saripalli, S.; Montgomery, J.F.; Sukhatme, G.S. Visually guided landing of an unmanned aerial vehicle. *IEEE Trans. Robot. Autom.* **2003**, *19*, 371–380. [\[CrossRef\]](#)
8. Jung, Y.; Lee, D.; Bang, H. Close-range vision navigation and guidance for rotary UAV autonomous landing. In Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden, 24–28 August 2015; pp. 342–347. [\[CrossRef\]](#)
9. Baca, T.; Stepan, P.; Spurny, V.; Hert, D.; Penicka, R.; Saska, M.; Thomas, J.; Loianno, G.; Kumar, V. Autonomous landing on a moving vehicle with an unmanned aerial vehicle. *J. Field Robot.* **2019**, *36*, 874–891. [\[CrossRef\]](#)
10. Verbandt, M.; Theys, B.; De Schutter, J. Robust marker-tracking system for vision-based autonomous landing of VTOL UAVs. In Proceedings of the International Micro Air Vehicle Conference and Competition, Delft, Netherlands, 12–15 August 2014; pp. 84–91. [\[CrossRef\]](#)
11. Tsai, A.C.; Gibbens, P.W.; Stone, R.H. Terminal phase vision-based target recognition and 3D pose estimation for a tail-sitter, vertical takeoff and landing unmanned air vehicle. In Proceedings of the Pacific-Rim Symposium on Image and Video Technology, Hsinchu, Taiwan, 10–13 December 2006; pp. 672–681. [\[CrossRef\]](#)

12. Xu, G.; Zhang, Y.; Ji, S.; Cheng, Y.; Tian, Y. Research on computer vision-based for UAV autonomous landing on a ship. *Pattern Recognit. Lett.* **2009**, *30*, 600–605. [[CrossRef](#)]
13. Yang, F.; Shi, H.; Wang, H. A vision-based algorithm for landing unmanned aerial vehicles. In Proceedings of the 2008 International Conference on Computer Science and Software Engineering, Wuhan, China, 12–14 December 2008; pp. 993–996. [[CrossRef](#)]
14. Hu, M.-K. Visual pattern recognition by moment invariants. *IEEE Trans. Inf. Theory* **1962**, *8*, 179–187. [[CrossRef](#)]
15. Yol, A.; Delabarre, B.; Dame, A.; Dartois, J.É.; Marchand, E. Vision-based absolute localization for unmanned aerial vehicles. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 3429–3434. [[CrossRef](#)]
16. Zeng, F.; Shi, H.; Wang, H. The object recognition and adaptive threshold selection in the vision system for landing an unmanned aerial vehicle. In Proceedings of the 2009 International Conference on Information and Automation, Zhuhai, Macau, 22–24 June 2009; pp. 117–122. [[CrossRef](#)]
17. Shakernia, O.; Ma, Y.; Koo, T.J.; Hespanha, J.; Sastry, S.S. Vision guided landing of an unmanned air vehicle. In Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304), Phoenix, AZ, USA, 7–10 December 1999; pp. 4143–4148. [[CrossRef](#)]
18. Lange, S.; Sunderhauf, N.; Protzel, P. A vision based onboard approach for landing and position control of an autonomous multirotor UAV in GPS-denied environments. In Proceedings of the 2009 International Conference on Advanced Robotics, Munich, Germany, 22–26 June 2009; pp. 1–6.
19. Benini, A.; Rutherford, M.J.; Valavanis, K.P. Real-time, GPU-based pose estimation of a UAV for autonomous takeoff and landing. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 3463–3470. [[CrossRef](#)]
20. Yuan, H.; Xiao, C.; Xiu, S.; Zhan, W.; Ye, Z.; Zhang, F.; Zhou, C.; Wen, Y.; Li, Q. A Hierarchical Vision-Based UAV Localization for an Open Landing. *Electronics* **2018**, *7*, 68. [[CrossRef](#)]
21. Olson, E. AprilTag: A robust and flexible visual fiducial system. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3400–3407. [[CrossRef](#)]
22. Fiala, M. Comparing ARTag and ARToolkit Plus fiducial marker systems. In Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications, Ottawa, ON, Canada, 1 October 2005; p. 6. [[CrossRef](#)]
23. Li, Z.; Chen, Y.; Lu, H.; Wu, H.; Cheng, L. UAV autonomous landing technology based on AprilTags vision positioning algorithm. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 8148–8153. [[CrossRef](#)]
24. Wang, Z.; She, H.; Si, W. Autonomous landing of multi-rotors UAV with monocular gimbaled camera on moving vehicle. In Proceedings of the 2017 13th IEEE International Conference on Control & Automation (ICCA), Ohrid, Macedonia, 3–6 July 2017; pp. 408–412. [[CrossRef](#)]
25. Wu, H.; Cai, Z.; Wang, Y. Vision-based auxiliary navigation method using augmented reality for unmanned aerial vehicles. In Proceedings of the IEEE 10th International Conference on Industrial Informatics, Beijing, China, 25–27 July 2012; pp. 520–525. [[CrossRef](#)]
26. Wang, J.; Olson, E. AprilTag 2: Efficient and robust fiducial detection. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 4193–4198. [[CrossRef](#)]
27. Yang, S.; Scherer, S.A.; Zell, A. An Onboard Monocular Vision System for Autonomous Takeoff, Hovering and Landing of a Micro Aerial Vehicle. *J. Intell. Robot. Syst.* **2012**, *69*, 499–515. [[CrossRef](#)]
28. Nguyen, P.H.; Kim, K.W.; Lee, Y.W.; Park, K.R. Remote Marker-Based Tracking for UAV Landing Using Visible-Light Camera Sensor. *Sensors* **2017**, *17*, 1987. [[CrossRef](#)]
29. Xiu, S.; Wen, Y.; Xiao, C.; Yuan, H.; Zhan, W. Design and Simulation on Autonomous Landing of a Quad Tilt Rotor. *J. Syst. Simul.* **2020**, *32*, 1676. [[CrossRef](#)]
30. Kotsiantis, S.B. Supervised Machine Learning: A Review of Classification Techniques. In Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, Amsterdam, Netherlands, 10 June 2007; pp. 3–24.
31. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Suthaharan, S., Ed.; Springer: Boston, MA, USA, 2016; pp. 207–235.
32. Kramer, O. K-Nearest Neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Kramer, O., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23.
33. Li, Y.; Wang, Y.; Luo, H.; Chen, Y.; Jiang, Y. Landmark recognition for UAV Autonomous landing based on vision. *Appl. Res. Comput.* **2012**, *29*, 2780–2783. [[CrossRef](#)]
34. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [[CrossRef](#)]
35. Lee, T.; McKeever, S.; Courtney, J. Flying Free: A Research Overview of Deep Learning in Drone Navigation Autonomy. *Drones* **2021**, *5*, 52. [[CrossRef](#)]
36. Chen, J.; Miao, X.; Jiang, H.; Chen, J.; Liu, X. Identification of autonomous landing sign for unmanned aerial vehicle based on faster regions with convolutional neural network. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 2109–2114. [[CrossRef](#)]

37. Nguyen, P.H.; Arsalan, M.; Koo, J.H.; Naqvi, R.A.; Truong, N.Q.; Park, K.R. LightDenseYOLO: A Fast and Accurate Marker Tracker for Autonomous UAV Landing by Visible Light Camera Sensor on Drone. *Sensors* **2018**, *18*, 1703. [[CrossRef](#)]
38. Truong, N.Q.; Lee, Y.W.; Owais, M.; Nguyen, D.T.; Batchuluun, G.; Pham, T.D.; Park, K.R. SlimDeblurGAN-Based Motion Deblurring and Marker Detection for Autonomous Drone Landing. *Sensors* **2020**, *20*, 3918. [[CrossRef](#)] [[PubMed](#)]
39. Yan, J.; Yin, X.-C.; Lin, W.; Deng, C.; Zha, H.; Yang, X. A Short Survey of Recent Advances in Graph Matching. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, New York, NY, USA, 15–19 October 2016; pp. 167–174. [[CrossRef](#)]
40. Conte, G.; Doherty, P. An Integrated UAV Navigation System Based on Aerial Image Matching. In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; pp. 1–10. [[CrossRef](#)]
41. Conte, G.; Doherty, P. Vision-Based Unmanned Aerial Vehicle Navigation Using Geo-Referenced Information. *EURASIP J. Adv. Signal Process.* **2009**, *2009*, 387308. [[CrossRef](#)]
42. Miller, A.; Shah, M.; Harper, D. Landing a UAV on a runway using image registration. In Proceedings of the 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA, USA, 19–23 May 2008; pp. 182–187. [[CrossRef](#)]
43. Cesetti, A.; Frontoni, E.; Mancini, A.; Zingaretti, P.; Longhi, S. A Vision-Based Guidance System for UAV Navigation and Safe Landing using Natural Landmarks. *J. Intell. Robot. Syst.* **2009**, *57*, 233. [[CrossRef](#)]
44. Zhao, L.; Qi, W.; Li, S.Z.; Yang, S.-Q.; Zhang, H. Key-frame extraction and shot retrieval using nearest feature line (NFL). In Proceedings of the 2000 ACM Workshops on Multimedia, Los Angeles, CA, USA, 30 October–3 November 2000; pp. 217–220. [[CrossRef](#)]
45. Li, Y.; Pan, Q.; Zhao, C. Natural-Landmark Scene Matching Vision Navigation based on Dynamic Key-frame. *Phys. Procedia* **2012**, *24*, 1701–1706. [[CrossRef](#)]
46. Saputra, M.R.U.; Markham, A.; Trigoni, N. Visual SLAM and Structure from Motion in Dynamic Environments: A Survey. *ACM Comput. Surv.* **2018**, *51*, 37. [[CrossRef](#)]
47. Shen, S.; Mulgaonkar, Y.; Michael, N.; Kumar, V. Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 4974–4981. [[CrossRef](#)]
48. Cadena, C.; Carlone, L.; Carrillo, H.; Latif, Y.; Scaramuzza, D.; Neira, J.; Reid, I.; Leonard, J.J. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Trans. Robot.* **2016**, *32*, 1309–1332. [[CrossRef](#)]
49. Engel, J.; Sturm, J.; Cremers, D. Camera-based navigation of a low-cost quadcopter. In Proceedings of the 2012 IEEE/RISJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 2815–2821. [[CrossRef](#)]
50. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [[CrossRef](#)]
51. Wang, K.; Shen, S. MVDepthNet: Real-Time Multiview Depth Estimation Neural Network. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 248–257. [[CrossRef](#)]
52. Wang, F.; Cui, J.-Q.; Chen, B.-M.; Lee, T.H. A Comprehensive UAV Indoor Navigation System Based on Vision Optical Flow and Laser FastSLAM. *Acta Autom. Sin.* **2013**, *39*, 1889–1899. [[CrossRef](#)]
53. Cui, T.; Guo, C.; Liu, Y.; Tian, Z. Precise Landing Control of UAV Based on Binocular Visual SLAM. In Proceedings of the 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS), Wuhan, China, 14–16 May 2021; pp. 312–317. [[CrossRef](#)]
54. Yang, T.; Li, P.; Zhang, H.; Li, J.; Li, Z. Monocular Vision SLAM-Based UAV Autonomous Landing in Emergencies and Unknown Environments. *Electronics* **2018**, *7*, 73. [[CrossRef](#)]
55. Cheng, H.; Chen, Y.; Li, X.; Wong Wing, S. Autonomous takeoff, tracking and landing of a UAV on a moving UGV using onboard monocular vision. In Proceedings of the 32nd Chinese Control Conference, Xi'an, China, 26–28 July 2013; pp. 5895–5901.
56. Chen, X.; Phang, S.K.; Shan, M.; Chen, B.M. System integration of a vision-guided UAV for autonomous landing on moving platform. In Proceedings of the 2016 12th IEEE International Conference on Control and Automation (ICCA), Kathmandu, Nepal, 1–3 June 2016; pp. 761–766. [[CrossRef](#)]
57. Araar, O.; Aouf, N.; Vitanov, I. Vision based autonomous landing of multirotor UAV on moving platform. *J. Intell. Robot. Syst.* **2017**, *85*, 369–384. [[CrossRef](#)]
58. Yang, T.; Ren, Q.; Zhang, F.; Xie, B.; Ren, H.; Li, J.; Zhang, Y. Hybrid Camera Array-Based UAV Auto-Landing on Moving UGV in GPS-Denied Environment. *Remote Sens.* **2018**, *10*, 1829. [[CrossRef](#)]
59. Rodriguez-Ramos, A.; Sampedro, C.; Bavle, H.; de la Puente, P.; Campoy, P. A Deep Reinforcement Learning Strategy for UAV Autonomous Landing on a Moving Platform. *J. Intell. Robot. Syst.* **2018**, *93*, 351–366. [[CrossRef](#)]
60. Nepal, U.; Eslamiati, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22*, 464. [[CrossRef](#)]
61. Sanchez-Lopez, J.L.; Pestana, J.; Saripalli, S.; Campoy, P. An Approach Toward Visual Autonomous Ship Board Landing of a VTOL UAV. *J. Intell. Robot. Syst.* **2013**, *74*, 113–127. [[CrossRef](#)]
62. Morais, F.; Ramalho, T.; Sinogas, P.; Marques, M.M.; Santos, N.P.; Lobo, V. Trajectory and guidance mode for autonomously landing an UAV on a naval platform using a vision approach. In Proceedings of the OCEANS 2015, Genova, Italy, 18–21 May 2015; pp. 1–7. [[CrossRef](#)]

63. Polvara, R.; Sharma, S.; Wan, J.; Manning, A.; Sutton, R. Towards autonomous landing on a moving vessel through fiducial markers. In Proceedings of the 2017 European Conference on Mobile Robots (ECMR), Paris, France, 6–8 September 2017; pp. 1–6. [[CrossRef](#)]
64. Li, J.; Wang, X.; Cui, H.; Ma, Z. Research on Detection Technology of Autonomous Landing Based on Airborne Vision. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *440*, 042093. [[CrossRef](#)]
65. Falanga, D.; Zanchettin, A.; Simovic, A.; Delmerico, J.; Scaramuzza, D. Vision-based autonomous quadrotor landing on a moving platform. In Proceedings of the 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), Shanghai, China, 11–13 October 2017; pp. 200–207. [[CrossRef](#)]
66. Garcia-Pardo, P.J.; Sukhatme, G.S.; Montgomery, J.F. Towards vision-based safe landing for an autonomous helicopter. *Robot. Auton. Syst.* **2002**, *38*, 19–29. [[CrossRef](#)]
67. Fitzgerald, D.; Walker, R.; Campbell, D. A Vision Based Forced Landing Site Selection System for an Autonomous UAV. In Proceedings of the 2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 5–8 December 2005; pp. 397–402. [[CrossRef](#)]
68. Mejias, L.; Fitzgerald, D.L.; Eng, P.C.; Xi, L. Forced landing technologies for unmanned aerial vehicles: Towards safer operations. *Aer. Veh.* **2009**, *1*, 415–442.