


Article

Salient Preprocessing: Robotic ICP Pose Estimation Based on SIFT Features

Lihe Hu¹, Yi Zhang^{2,*}, Yang Wang¹, Gengyu Ge¹ and Wei Wang¹ 

¹ School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

² Advanced Manufacturing and Automatization Engineering Laboratory, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

* Correspondence: zhangyi@cqupt.edu.cn; Tel.: +86-023-6248-0054

Abstract: The pose estimation can be effectively solved according to the feature point matching relationship in RGB-D. However, the extraction and matching process based on the whole image's feature point is very computationally intensive and lacks robustness, which is the bottleneck of the traditional ICP algorithm. This paper proposes representing the whole image's feature points by the salient objects' robustness SIFT feature points through the salient preprocessing, and further solving the pose estimation. The steps are as follows: (1) salient preprocessing; (2) salient object's SIFT feature extraction and matching; (3) RANSAC removes mismatching salient feature points; (4) ICP pose estimation. This paper proposes salient preprocessing aided by RANSAC processing based on the SIFT feature for pose estimation for the first time, which is a coarse-to-fine method. The experimental results show that our salient preprocessing algorithm can coarsely reduce the feature points' extractable range and interfere. Furthermore, the results are processed by RANSAC good optimization, reducing the calculation amount in the feature points' extraction process and improving the matching quality of the point pairs. Finally, the calculation amount of solving R, t based on all the matching feature points is reduced and provides a new idea for related research.

Keywords: salient preprocessing; salient object's SIFT feature; matching quality; calculation amount



Citation: Hu, L.; Zhang, Y.; Wang, Y.; Ge, G.; Wang, W. Salient Preprocessing: Robotic ICP Pose Estimation Based on SIFT Features. *Machines* **2023**, *11*, 157. <https://doi.org/10.3390/machines11020157>

Academic Editors: Praneel Chand and Antonios Gasteratos

Received: 13 November 2022

Revised: 18 January 2023

Accepted: 20 January 2023

Published: 23 January 2023



Copyright: © 2023 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

SLAM (robot simultaneous localization and mapping) can realize the autonomous exploration of unknown environments. Visual odometry is the critical technology of SLAM, which uses the relationship between matching feature points in the adjacent frame images to estimate the camera's rotation R and translation t for calculating the camera's pose change [1]. Among many visual cameras, the vision sensor RGB-D camera can provide depth information for the positioning and perception of the mobile robot's environment. It can be directly used to estimate the camera pose after matching the feature points in RGB-D images. Moreover, using the ICP (iterative closest point) to solve the pose estimation between two matching 3D feature point sets is a very reliable method to solve the RGB-D visual odometry problem [2–4].

ICP (iterative closest point) uses the correspondence relationship between the two-point clouds and solves the R, t by solving a least squares problem. It is the most effective and reliable 3D-3D pose estimation method [5–7], which is suitable for RGB-D SLAM and lidar SLAM. However, the lidar data features are not rich enough to know the matching relationship between two data point sets. A better matching relationship can be obtained according to the feature points in the visual RGB-D. The entire pose solution problem becomes easier than the ICP based on lidar [8]. The ICP can better complete the rigid body registration of the point cloud based on RGB-D. The related work is mature and is widely used because of the advantages of high precision and fast operation speed [9–11]. However, in the actual registration process, the effect will not be perfectly achieved. The effect is

not ideal for containing outliers, noise points, missing point clouds, edge disturbance, etc. Furthermore, the improper selection of the initial position will also lead to registration failure. Here, we carry out research in this paper from the perspective of improving the pose estimation quality by improving feature point matching.

ICP is calculated based on the matching point pairs in the images. In the process of image feature extraction, description, and matching, the amount of calculation is tremendous, and the speed is slow, which is the bottleneck of the traditional ICP algorithm [12]. The traditional ICP algorithm calculates the corresponding relationship based on all matching feature points in the images. The extraction and matching based on this large number of feature points is time-consuming. In subsequent studies, related researchers proposed various feature selection methods [13]. The primary purpose of these methods is to reduce the feature point extraction and matching number, that is, how to select the fewest features with high quality to represent all the original image's feature information. It mainly includes the following research aspects: the filtering (sampling) of the point cloud [14], the matching method of the corresponding point cloud [4], the weighting of the matching point pair [3,15], and the screening of the matching point pair [16,17]. Related research mainly focuses on the feature selecting and matching way of the scene. Here, we introduce the salient attention mechanism by using the extraction and matching of salient object's features to represent the whole image's features for pose estimation. It is theoretically feasible and full of application potential to reduce the whole image's feature point extraction and matching to save time. Among the many kinds of features, we choose to perform pose estimation based on SIFT (scale-Invariant feature transform) features because the SIFT feature remains invariant to rotation, scale scaling, and brightness changes, and is a very stable local feature [18,19].

RANSAC (random sample consensus) means consistent random sampling. The RANSAC is now widely used in image registration and splicing [20]. It means taking random samples from matching samples to find consistent sample points and removing incorrect matching feature point pairs. The RANSAC algorithm is based on a set of sample data sets containing abnormal data to calculate the data's task model parameters and obtain effective sample data [21]. One common problem is feature points incorrectly matching when using existing algorithms without RANSAC. These incorrectly matching points greatly impact the matching effect and pose estimation. Therefore, we propose using salient preprocessing methods to coarsely eliminate the incorrectly matching feature points and using the RANSAC algorithm to further finely eliminate the incorrectly matching points. Our previous exploration experiments show that the image salient preprocessing in this paper still has the possibility of noise influence of feature points incorrectly matching, so we further improve the matching feature point quality based on the RANSAC processing.

The salient object is the most critical object in the scene, which is the response of modern computer vision tasks to imitate human beings to the objective nature. Objectively speaking, it is the response weight of human eyes to specific regions in the frequency domain and chroma space (HSV, LAB). After years of evolution, it is natural attention and unconditioned reflection of feature sparsity [22–24]. Relevant studies show that feature point extraction and matching are time-consuming in ICP pose estimation based on feature points. Increasing 3D point sets to be matched will severely reduce the ICP algorithm's calculation efficiency [25,26]. Here, we propose a new feature selection method, which introduces the salient attention mechanism. We use the most critical objects' feature points to represent the whole scene. The pose estimation is only based on the critical objects' matching feature points, which reduces the influence of background noise and improves the matching quality because the larger the scene, the greater the probability of error matching, and without any background feature involved in the feature matching. Theoretically, the background feature points can be effectively eliminated by salient processing. It is feasible to reduce the number of matching feature points, and thus reduce the calculation of pose estimation based on feature points.

This is the highlight of this article:

- (1) Our method proposes reducing the SIFT feature points' extraction and matching based on salient preprocessing, which saves the calculation of solving R, t based on all the matching feature points and with less processing time.
- (2) The interference from the background's feature points can be eliminated, and the matching quality of feature points can be improved after salient preprocessing.
- (3) Our algorithm uses a coarse-to-fine method to eliminate the wrong matching feature points, simultaneously improving the quality of the matching point pairs and reducing the number of point sets.
- (4) We propose a salient object's feature point selection method to improve the real-time performance of ICP pose estimation while achieving good robustness. Moreover, we analyze the main influencing factors that affect its related performance.

2. Related Work

The standard ICP algorithm has a high calculation amount for feature point extraction and matching, which is sensitive to initial transformation and easily falls into the local optimum. Since ICP was proposed, many improved ICP algorithms have been proposed. We will list some influential research cases here and list the problems that our work is committed to solving:

Point-to-Plane ICP. Zeng Y et al. [2] proposed a new weight method for the ICP algorithm of point-to-plane error metric to improve the accuracy and robustness of the ICP algorithm. Li J et al. [3] presented a new symmetric point-to-plane distance metric whose functional zero-set is a set of locally second-order surfaces. Then, they introduced a robust adaptive loss to construct their robust symmetric metric. Compared with directly calculating the point-to-point distance used in the cost function of the original ICP algorithm, the point-to-plane is the distance from the original vertex to the face where the target vertex is located. Furthermore, it should be noted that the optimization of point-to-plane is a nonlinear problem, and calculation is slow and generally uses linearization approximation. The point-to-plane will converge faster than the point-to-point, but the speed is relatively slow due to the optimization being a nonlinear problem. So, in this article, we still use the most classic and simplest point-to-point algorithm as the basis for further salient preprocessing for better real-time performance.

Plane-to-Plane ICP. Wang J et al. [4] proposed an acceptance-rejection sampling-based two-step point filter to exclude the points that rarely benefit the lidar odometry performance, reducing the distribution approximation errors when the GICP works as a plane-to-plane iterative closest point (ICP). Pavan N L et al. [5] presented a plane-based matching algorithm to find plane-to-plane correspondences using a new parametrization based on complex numbers, which avoids the ambiguity in the calculation of the rotation angle formed between normal vectors of adjacent planes. The plane-to-plane considers the point cloud's local structure and calculates the face-to-face distance, similar to the point-to-plane that considers the target point cloud's local structure [27,28]. The above-related literature put forward a new idea of point cloud matching. The improvement of the point cloud matching method can reduce the logarithm of feature matching. Further, we propose reducing the feature matching consideration from the source of point cloud data through salient preprocessing, which can be used to enrich the research on correlation matching algorithms.

Generalized ICP (GICP). Min Z et al. [6] first formally formulated the generalized PS registration problem probabilistically. Especially, positional and orientational information was incorporated into the registration. Makovetskii A et al. [7] proposed an exact closed-form solution for orthogonal registration of point clouds based on the generalized point-to-point ICP algorithm. Moreover, they used points and normal vectors to align 3D point clouds, while the common point-to-point approach uses only the coordinates of points. The generalized ICP (GICP) comprehensively considers point-to-point, point-to-plane, and plane-to-plane. In comparison, the strategy, accuracy, and robustness have been improved. The related literature proposed a new point cloud matching idea, and the excessive de-

tails and consideration of data can be reduced from the point cloud data source through salient preprocessing.

Normal iterative closest point (NICP). Serafin J et al. [9] presented a novel online method to align point clouds recursively. The algorithm relies on a least-squares formulation of the alignment problem that minimizes an error metric depending on these surface characteristics. Jia S et al. [10] presented a novel method called color support normal iterative closest point (color NICP) to align point clouds recursively. Their algorithm takes advantage of not only the 3D structure but also the texture information of the color image. NICP considers the normal vector and local curvature and further utilizes the local structure information of the point cloud. The experimental results in their paper are better than those of GICP. The above-related literature puts forward a new idea of point cloud matching. However, we found that the improved ICP algorithm above does not introduce a salient attention mechanism to imitate the human perspective to process tasks. Here, we propose further reducing the data consideration from the source of point cloud data through salient preprocessing, which can be used to enrich the work of relevant algorithms in point cloud matching.

The above research demonstrates that the improved ICP method of point cloud matching can reduce feature matching and improve pose estimation robustness. Relevant research shows that the extraction and matching of feature points occupy a considerable proportion of the entire process of ICP pose estimation [5,9,15]. With the increase in the number of spatial matching points, the efficiency of the ICP algorithm solution will decrease seriously [25,26]. So, in addition to the above-improved point cloud matching method, we also focus on choosing fewer feature points to represent all the original point set's feature information. The existing related research work revolves around the following work: (1) uniform sub-sampling [11]; (2) random sampling [29]; (3) feature-based sampling [14]; (4) normal-space sampling [12]; and (5) curvature sampling [13].

As represented by the above, various methods are used to reduce the extraction and matching of feature points, and various sampling and matching methods are proposed. However, there is rarely a report on introducing the salient attention mechanisms in related research fields. The salient object is the most critical in an image. It is more representative of the whole image, which is used in reducing the whole image's feature point extraction and matching to save time, and is theoretically feasible and full of application potential. Recent influential articles, such as Wan T et al. [30], offer three strategies to increase the robustness of the iterative closest point (ICP) algorithm involving the salient object detection (SOD) method, and their experimental scheme. These are interesting, but complex. Yao Run Zhao et al. [31] proposed a joint objective to align both salient color points and background points based on the color-supported generalized ICP. Furthermore, they fully leveraged geometric and texture information, but their method lacks consideration from the perspective of reducing matching features. In this paper, we propose a simple coarse-to-fine approach to study the help of salient preprocessing and RANSAC processing in reducing feature point matching. Our feature point selection method is based on the feature points on the salient object for sampling and using RANSAC for optimizing. Our method can effectively reduce the noise interference from incorrect environmental matching point pairs and reduce the feature points' extraction and matching, thereby improving the real-time performance of ICP pose estimation.

3. Methodology

In this section, we provide flowchart Figure 1 of the algorithm proposed in this paper with the main experimental operation explanations. Furthermore, we describe the main algorithm theories involved in the proposed method, including three main components: 3.1. Salient preprocessing; 3.2. RANSAC removes incorrectly matching salient SIFT feature points; 3.3. Pose estimation based on matching salient feature points.

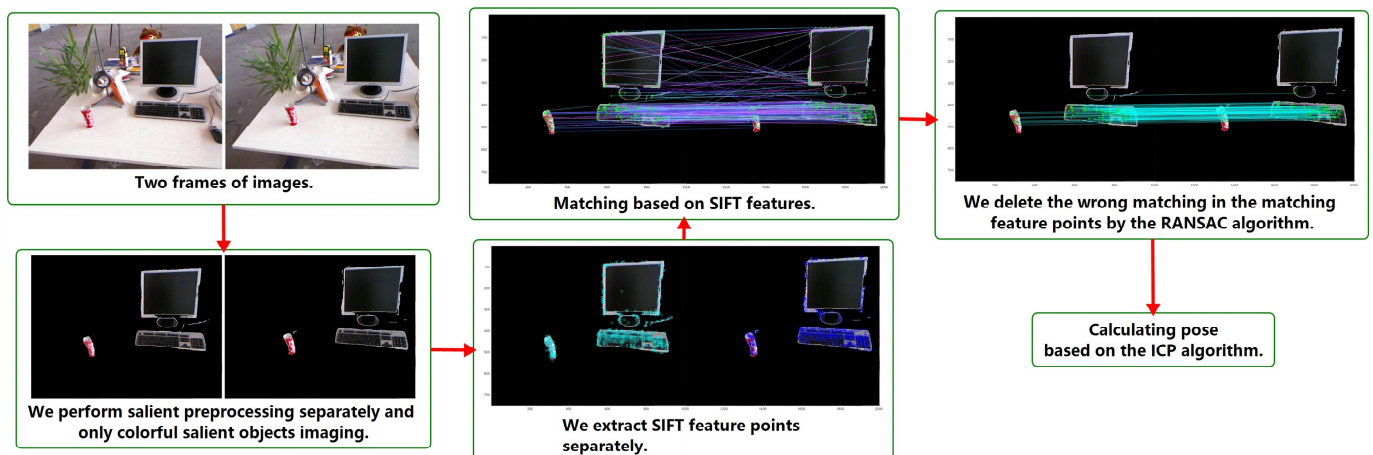


Figure 1. The flowchart of the method proposed in this paper contains explanations. Our proposed method includes salient preprocessing, feature extraction, feature point matching, RANSAC processing, and ICP pose estimation.

3.1. Salient Preprocessing

Salient object detection can be defined as detecting the areas in the image with the most obvious and prominent contrast with the background [23,24], which is often used as a preprocessing stage for advanced vision tasks, such as video SOD [32], visual tracking [33], SLAM [34], etc. We propose to process images using a salient object detection neural network architecture [35] as the preprocessing stage of pose estimation based on the RGB-D camera. We use salient preprocessing to eliminate the extraction and matching of background feature points. Furthermore, we use salient objects' feature points to replace the entire image's feature points to improve the real-time performance of ICP pose estimation. The salient preprocessing neural network architecture [35] introduces a dynamic feature integration strategy to choose favored features dynamically in an end-to-end learning manner. It runs in real time, while producing good detection results. This dynamic strategy can largely ease the process of architecture construction and promote the backbone to adjust its parameters for solving multiple problems adaptively.

3.2. RANSAC Removes Incorrectly Matching Salient SIFT Feature Points

SIFT (scale invariant feature transform) is the most classic local feature extraction method [36]. The SIFT feature remains invariant to rotation, scale scaling, and brightness changes and is a very stable local feature [18,19]. So, in this article, we perform ICP pose estimation based on SIFT features.

There are still incorrectly matching pairs of feature points in our proposed algorithm after salient preprocessing. The incorrectly matching points belong to outliers in statistics, which are interference noises. These mismatches will have a great impact on the subsequent camera pose estimation. Too many mismatches will make the camera pose estimation more and more outrageous. Therefore, we need to improve the matching quality through RANSAC further.

RANSAC (random sample consensus) adopts an iterative method to estimate the parameters of the task model from a group of observed data that contains outliers. The RANSAC algorithm assumes that the data contain correct and abnormal data (or noise). The correct data are labeled as inliers, and the abnormal data are labeled as outliers [20,21]. RANSAC also assumes that given a set of correct data, the task model parameters that conform to these data can be calculated. The core idea of the algorithm is randomness and hypothesis. The randomness is to randomly select sampling data according to the occurrence probability of correct data. The randomness method can approximate the correct results according to the law of large numbers. The hypothesis is to assume that the selected sampling data are all correct and then use these correct data to calculate other points by the

task model and give a score to the result. Finally, the highest-scoring task model was treated as the model for the entire data set [37,38]. Relevant studies have fully demonstrated that RANSAC can effectively improve matching point pairs' quality.

3.3. Pose Estimation Based on Matching Salient Feature Points

ICP (iterative closure point) has the advantages of simplicity and low computational complexity compared with other algorithms, and it has become the most popular rigid point cloud registration method [14,29]. The ICP algorithm assigns correspondence iteratively based on the nearest distance criterion and obtains the rigid transformed least squares for the two-point clouds. That is a process of continuously determining the corresponding relationship and continuing to iterate until the minimum value of least squares is reached [3,39]. Currently, many point cloud matching algorithms are based on the ICP improvement, which is simple and does not need to segment the point cloud. When the initial matching is good, the accuracy and convergence are good. When the initial matching is bad, the accuracy and convergence are bad. The features of lidar data are not rich enough, so it is not easy to know the matching relationship between the two-point sets. A better matching relationship can be obtained in visual RGB-D according to the matching feature points, and the whole pose estimation problem has become simpler [8,40,41]. This paper proposes to improve the real-time performance of ICP pose estimation based on an RGB-D camera by reducing the image feature point extraction and matching range.

The solution of ICP is to solve \mathbf{R} and \mathbf{t} to minimize the following formula:

$$E(\mathbf{R}, \mathbf{t}) = \frac{1}{P_s} \sum_{i=1}^{|P_s|} \|P_t^i - (\mathbf{R} * P_s^i + \mathbf{t})\|^2, \quad (1)$$

where P_s^i and P_t^i are the corresponding points in the origin point cloud P_s (source) and the target point cloud P_t (target), respectively.

The idea of the ICP algorithm is that if we know the correspondence between the points of the two-point clouds, we can solve the rotation transformation \mathbf{R} and translation transformation \mathbf{t} in the pose transformation by solving the least squares problem.

4. Experiments

In this section, we conduct the experiments of our proposed method. The method contains the following content: (1) perform salient preprocessing experiments on images; (2) the experiments prove that salient preprocessing can effectively reduce the extraction and matching of feature points; (3) according to the improvement of pose estimation after salient preprocessing, we analyze the influencing factors and conduct the verification analysis; (4) we conduct comparative experiments and analyses. The experiments have verified that the salient preprocessing algorithm can effectively reduce the calculation amount in the feature point extraction and matching process. Our method's relative pose error and real-time performance are reasonable, providing a new research idea for processing feature points in pose estimation.

The data set used in this paper is shown in Table 1:

Table 1. The test set data information details.

Category	Data Type	Number	Sampling Premise
TUM [42]	RGB-D image	600 pairs	Salient imaging is clear, and the Raw RGB image pairs are the same scene.

Our experiments under the condition of the GPU are NVIDIA GeForce GTX 1650, the processor is AMD Ryzen 5 4600H with Radeon Graphics, the memory is 8GB, the main frequency of the processor is 3GHz, and the highest turbo frequency is 4 GHz.

Here, we use the subset with the ground truth trajectory in TUM. There are 600 image pairs in total with the true pose values and clear salient imaging selected as the test data set for the proposed algorithm (the subset includes: sequence 'freiburg1_plant', sequence

'freiburg1_teddy', sequence 'freiburg2_coke', sequence 'freiburg2_flowerbouquet', sequence 'freiburg2_flowerbouquet_brownbackground', sequence 'freiburg2_metallic_sphere', sequence 'freiburg3_cabinet', sequence 'freiburg3_large_cabinet', sequence 'freiburg3_teddy', Sequence 'freiburg2_desk', sequence 'freiburg1_xyz', sequence 'freiburg1_rpy', sequence 'freiburg2_xyz', sequence 'freiburg2_rpy', sequence 'freiburg1_360', sequence 'freiburg1_desk', etc.).

4.1. Salient Preprocessing Experiment

Figure 2 is a schematic diagram of the salient preprocessing method in this paper.

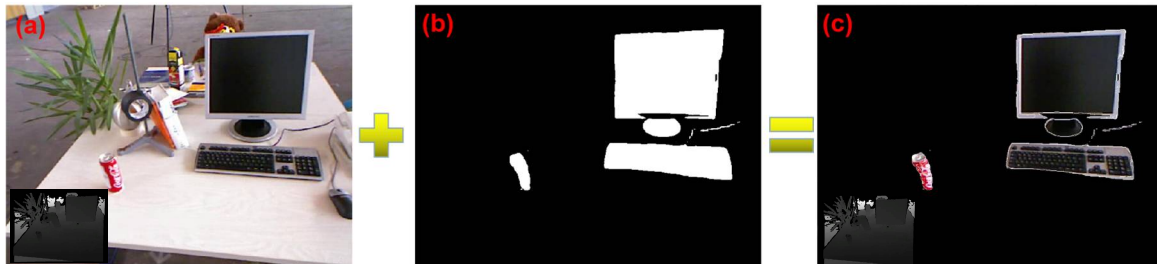


Figure 2. We use a salient grayscale image (b) to mask image (a) to get colorful salient object imaging (c).

The following are the steps of salient preprocessing:

- (1) We process the original image Figure 2a to obtain a salient image Figure 2b with a gray value imaging.
- (2) We set the alpha channel of the salient object part where the pixel value is greater than 200 to transparent in the gray value image. That is, its alpha value is set to 0 at the place where the pixel value is greater than 200.
- (3) Under the same coordinate system, we use Figure 2b with the transparent channel to mask Figure 2a for obtaining Figure 2c with colorful salient object imaging.

Our salient preprocessing method makes the background image masked with black pixels. Compared with Figure 2b that salient objects in gray value imaging, the salient preprocessing step help Figure 2c get salient objects with richer texture features, which are more conducive to SIFT feature points' extraction and matching.

4.2. The Advantages of Salient Preprocessing

Here, we compare the two images before and after salient preprocessing. Figure 3(a) has more extracted feature points than Figure 3(a)#. After our salient preprocessing processes in the original image, the feature points in the background do not exist. However, the feature points in the new salient image imaging are dense at the edge of the contour. In this case, the total number of feature point differences in the image before and after salient preprocessing is still obvious. The total number of extracted feature points after salient preprocessing is less than the total number of feature points before salient preprocessing by more than 10%. That shows that our proposed salient preprocessing method can reduce the number of extracted feature points.

Figure 3(b),(b)# are based on the SIFT feature for feature matching comparison. The matching quality of the feature points in Figure 3(b)# is better after the salient preprocessing, and there is no background feature at this time without the noise point interference. Image similarity (FLANN method) improves from 41.23% in Figure 3(b) to 49.77% in Figure 3(b)#.

Figure 3(c),(c)# are the feature point matching maps before and after the salient preprocessing with RANSAC removing the incorrectly matching pairs. Our salient preprocessing without RANSAC processing can improve the matching feature quality with image similarity improving. However, with RANSAC further processing, this advantage is not obvious because the RANSAC and salient preprocessing algorithm can both remove the noise of feature points in the image and improve the matching quality of feature points. RANSAC can mask the advantages of coarsely salient preprocessing after RANSAC finely processing, such as the matching quality of Figure 3(c),(c)# are both well after RANSAC. The effect

of improving feature point matching after salient and RANSAC processing is not very obvious. However, from the whole process of our proposed algorithm, over 95% of image pairs that the incorrectly matching feature point pairs are reduced in the process from salient preprocessing to RANSAC processing after we observed and calculated the statistics for all of the images.

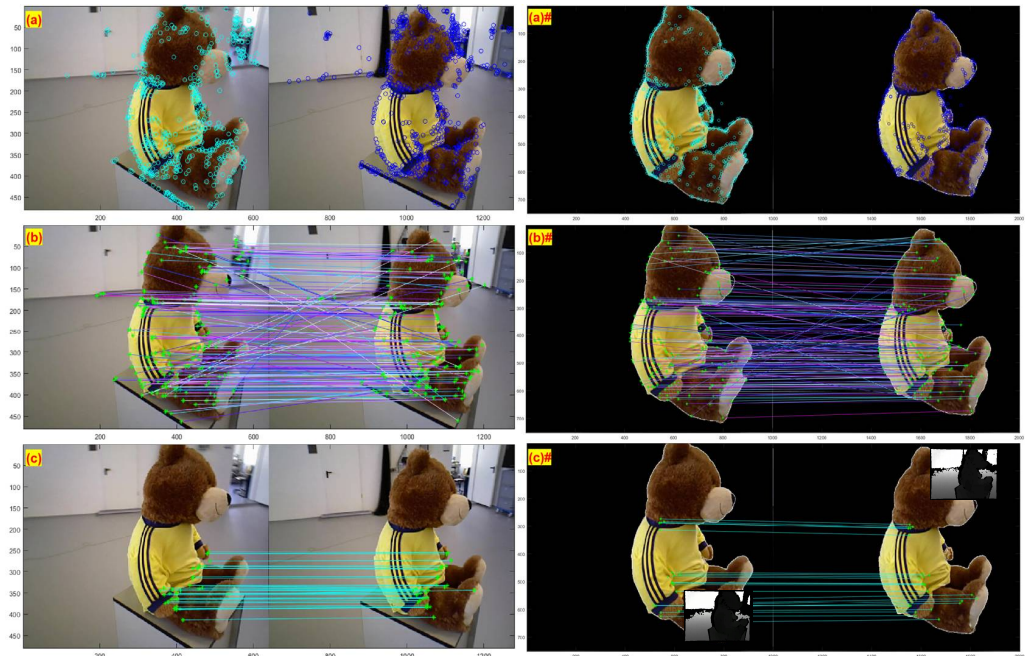


Figure 3. Comparative experiments before and after salient preprocessing. Including feature point extraction, feature point matching, and feature point matching after RANSAC processing. The feature points in the RGB map and the feature points in the Depth map have the same coordinate relationship in the same plane coordinate system, and it is easy to obtain the spatial positions of these feature points.

It can be seen from Table 2 that before and after salient preprocessing, whether or not RANSAC processing is performed, salient preprocessing can reduce the number of feature extractions to save time and improve feature matching quality. It can be seen from the comparison that under the same conditions, RANSAC processing can further improve the quality of feature matching. Figure 3(b)# shows the matching feature point pairs of the extracted feature points from Figure 3(a)# after salient preprocessing. The matching feature points after RANSAC processing are shown in Figure 3(c)#, and Figure 3(c)# has fewer feature point pairs available for pose estimation compared with Figure 3(c). The feature point extraction and matching of the whole process of our method are less than the corresponding horizontal position image without salient preprocessing, which saves the extraction and matching time of feature points.

Table 2. Information details of Figure 3.

Category	Extracted Key Points	Feature Point Matching Pairs	Number of Matching Pairs after RANSAC Processing
Without salient preprocessing	Left IMG 611, Right IMG 666	202	30
Salient preprocessing	Left IMG 627, Right IMG 535	172	12

4.3. Main Influencing Factors

Our experiments found that the quality of feature point matching can be improved after salient preprocessing (the pose estimation becomes better after salient preprocessing

(without RANSAC) takes a proportion of 92.5%). However, the matching feature points are wrong in some cases, and the incorrectly matching feature points cannot be effectively removed by RANSAC after salient preprocessing, and thus cannot effectively be used for pose estimation. After experimental analysis, we infer that the relative pose error is large because the feature points cannot be matched effectively and are affected by the following influencing factors, including (1) large image frame differences; (2) object contour and texture information; and (3) salient imaging quality differences.

The improved pose estimation after the salient preprocessing and RANSAC processing takes a proportion of 85.6%, which is not apparent compared with the improved pose estimation only based on salient preprocessing and which takes a proportion of 92.5%. That is because RANSAC will finely eliminate the wrong matching point pairs in all the matching point pairs. The advantages of salient preprocessing with RANSAC are not apparent, no matter whether we use coarsely salient preprocessing. The interference of incorrectly matching noise points will be finely removed in the later RANSAC stage. Our proposed salient preprocessing plus RANSAC is a coarse-to-fine way to eliminate incorrectly matched pairs and has fewer incorrectly matched pairs compared with using salient preprocessing only or the parallel comparison test without salient preprocessing, as shown in Table 7. However, our proposed method also obtains fewer feature pairs making it more susceptible to receiving the interference of incorrectly matching feature points than the original image using RANSAC processing. That is why the improvement of pose estimation after salient preprocessing plus RANSAC processing is less obvious than that only based on salient preprocessing in Table 3.

Table 3. The performance of the proposed method in improving pose estimation.

Category	Unable to Estimate the Pose by Our Method (Relative Pose Error $t > 0.1$ m/s, $R > 2$ deg/s)	The Pose Estimation Becomes Better after Salient Preprocessing (without RANSAC)	The Pose Estimation Becomes Better after Salient Preprocessing (with RANSAC)	Improved Real-Time Performance
Our method	7.9%	92.5%	85.6%	96.5%

The performance of the loss mainly comes from the influence of the poor matching quality of the feature points. It should be noted in this part of the experiment that the evaluation of pose estimation is based on the relative pose error for comparison.

About 7.9% cannot effectively estimate the pose in the experimental data set by our method, which comes from the influence of feature point extraction and matching quality. From the improved real-time performance in Table 3, we can see that as long as our salient preprocessing plus RANSAC processing is carried out, then our method can effectively improve the real-time performance of pose estimation, except for some influencing conditions (such as in Section 4.3.2, the feature points become more after salient preprocessing). Our algorithm's real-time performance improved 96.5% of the image pairs after the salient preprocessing using RANSAC processing compared with the original image using RANSAC processing, mainly owing to salient preprocessing reducing the feature point extraction and matching.

The verification tests of the three main influencing factors of the matching feature points quality that influence pose estimation are as follows:

In our analysis experiments below, in Figure 4(a),(a)#, Figure 5(a),(a)# and Figure 6(a),(a)# are the extracted feature points before and after the salient preprocessing. Figure 4(b),(b)#, Figure 5(b),(b)# and Figure 6(b),(b)# are the matching SIFT feature points before and after the salient preprocessing without RANSAC processing. Figure 4(c),(c)#, Figure 5(c),(c)# and Figure 6(c),(c)# are the matching SIFT feature points before and after the salient preprocessing with RANSAC processing.

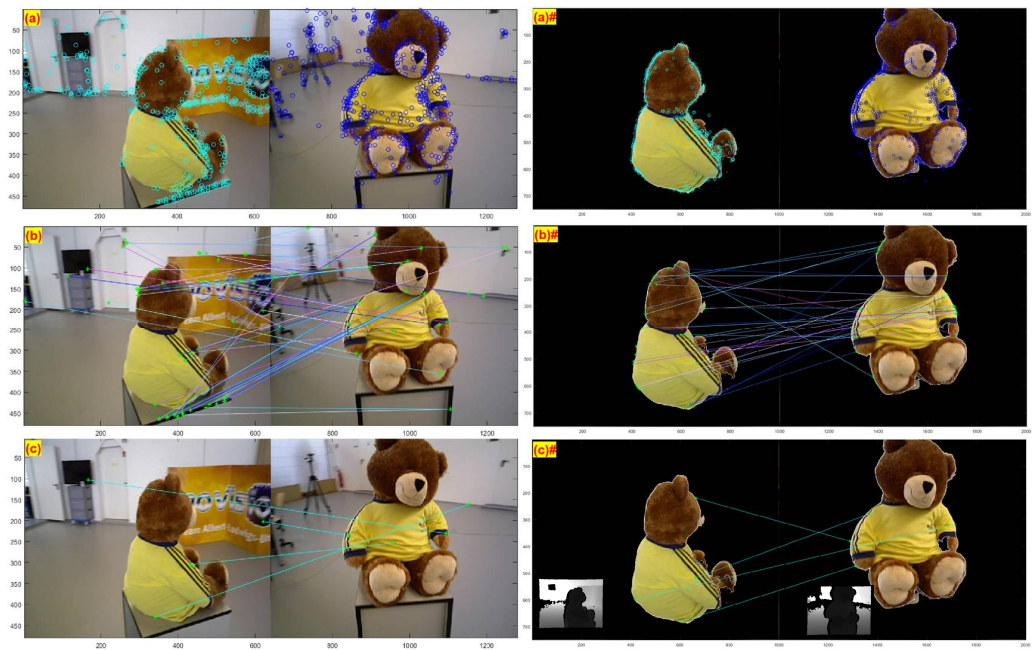


Figure 4. Feature matching cannot be performed effectively even after the salient preprocessing when the difference between frames is large. The feature points in the RGB map and the feature points in the Depth map have the same coordinate relationship in the same plane coordinate system, and it is easy to obtain the spatial positions of these feature points.

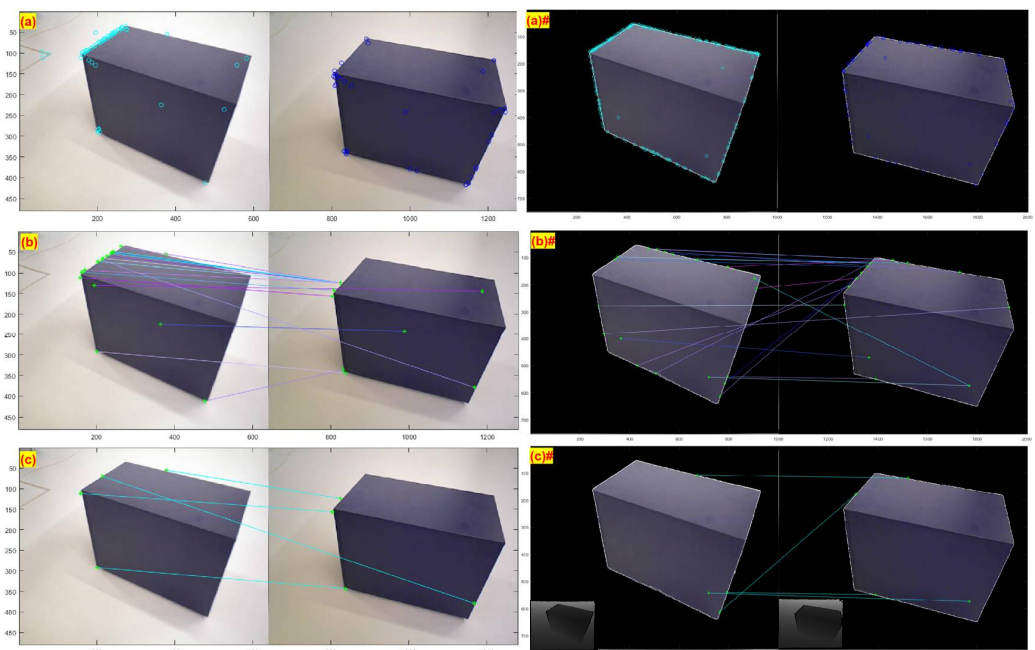


Figure 5. Feature matching cannot be performed effectively even after the salient preprocessing when the contour features are symmetrical and lack texture information. The feature points in the RGB map and the feature points in the Depth map have the same coordinate relationship in the same plane coordinate system, and it is easy to obtain the spatial positions of these feature points.

4.3.1. The Large Difference between Image Frames

Figures 3 and 4 are from the same subset of TUM, but the feature points' extraction and matching quality are very different. The reason is that the difference between image frames (the interval time of two image frames) in Figure 4 is too large to calculate pose

estimation effectively. In contrast, the difference between image frames in Figure 3 is much smaller and can effectively estimate the pose.

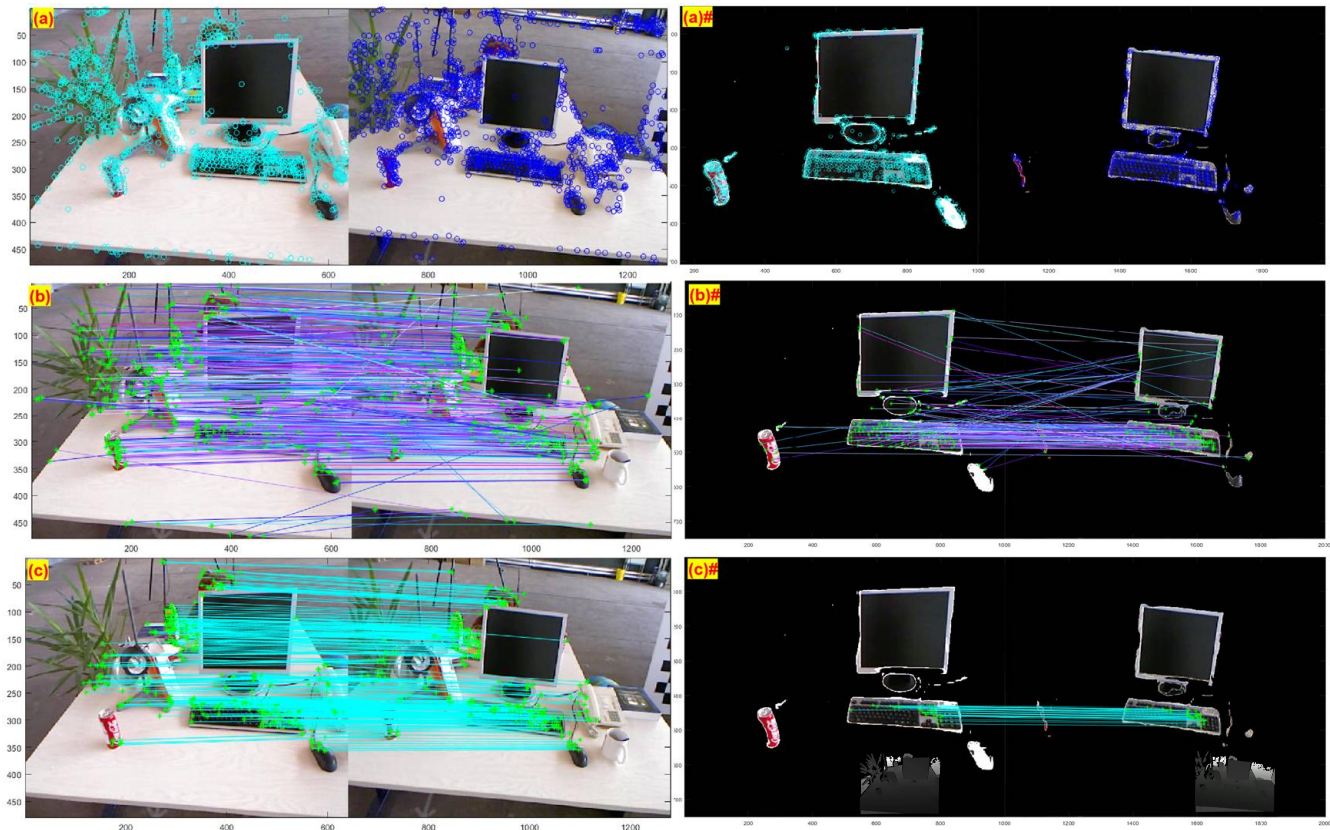


Figure 6. Feature matching numbers will be reduced after the salient preprocessing when obvious salient imaging differences exist. The feature points in the RGB map and the feature points in the Depth map have the same coordinate relationship in the same plane coordinate system, and it is easy to obtain the spatial positions of these feature points.

When there is a large time interval between the two image frames, there will be a large change in the imaging angle. As shown in Figure 4(b),(c) and Figure 4(b)#,(c)#, whether or not salient and RANSAC processing is performed, the matching quality of feature points is poor, and the pose estimation cannot be performed effectively. Figure 4(c)# is the last step of our experiment, and its final result has relatively better feature point extraction and matching than the parallel comparison experiment shown in Figure 4(c) because without any background interference points. All operations, including the result in Figure 4(c)#, the previous feature extraction in Figure 4(a)#, and the feature matching before RANSAC processing in Figure 4(b)# as a whole show that we can improve the real-time performance by reducing the time of feature points’ extraction and matching compared with the comparison experiments on the left in Figure 4 (Table 4).

Table 4. Information details of Figure 4.

Category	Extracted Key Points	Feature Point Matching Pairs	Number of Matching Pairs after RANSAC Processing
Without salient preprocessing	Left IMG 549, Right IMG 626	42	4
Salient preprocessing	Left IMG 350, Right IMG 615	30	4

4.3.2. Object Contour and Texture Information

The lacking contour features and texture information will affect the extraction and matching of SIFT features.

When the contour features of the objects in the two images are symmetrical, and the texture information is single, the feature point matching cannot be effectively carried out regardless of whether the images are after salient preprocessing, as shown in Figure 5(b),(b)# and Figure 5(c),(c)#. The pose estimation is invalid at this time. In our experiment, the whole feature point extraction and matching are relatively reduced after salient preprocessing, and the percentage of real-time improvement accounts for 96.5%, as shown in Table 3. Where the error situation is, as shown in Figure 5(a), the feature points become more after salient preprocessing, as shown in Figure 5(a)#. The matching logarithm of feature points after salient preprocessing is also more (Figure 5(b)# compared with Figure 5(b)). We infer that it is due to the salient object having a symmetrical contour and lacking texture information. In this situation, the imaging boundary of the salient object will change white width due to the influence of the imaging quality of the previous salient preprocessing, as shown in Figure 5(a)#. White stripes with varying widths are generally considered texture parts, and their characteristics are extracted as SIFT features. That is why the number of feature points on the contour boundary increases after salient preprocessing (Table 5). The final results of Figure 5(c),(c)# after RANSAC processing show that their feature point matching is poor, no matter whether it is after salient preprocessing. That is because the object lacks texture information at the symmetrical shape. The features in the image can easily be mismatched with the symmetrical part of another image, making the feature point matching ineffective for estimating pose.

Table 5. Information details of Figure 5.

Category	Extracted Key Points	Feature Point Matching Pairs	Number of Matching Pairs after RANSAC Processing
Without salient preprocessing	Left IMG 70, Right IMG 40	17	4
Salient preprocessing	Left IMG 197, Right IMG 79	20	4

4.3.3. Salient Imaging Differences

The salient preprocessing process of the robot camera at different positions is easily affected by the shooting angle in some situations, which makes the imaging of the salient object have certain differences. Such as, some smaller salient objects have smaller areas of salient imaging easily affected by salient preprocessing from different shooting angles, and this makes the salient small object can be clearly imaged at one shooting angle but may not fully be imaged at another. Moreover, the extracted features on this small salient object cannot be matched in different image frames, and this reduces the number of matching feature point pairs. If the salient object is large, the imaging area is relatively large, and the effect of this influencing factor is much smaller because of the more extracted features available for matching.

When the main objects in the scene have certain shapes and textures, our proposed salient preprocessing method can greatly reduce the feature point extraction and matching, as shown in Table 6, thereby saving time. However, our experiments show that when there is a salient imaging difference between the two frame images, the results in extractable feature differences affect the pose estimation. As shown in Figure 6(b)#, the salient imaging of the Cola and mouse in the two images are quite different. The matching feature point pairs in Figure 6(c)# are far less than in Figure 6(c) after the RANSAC processing. Too few matching point pairs will affect the pose estimation quality, and the fewer matching feature points are more susceptible to receiving the feature point noise interference. That is why the pose estimation quality after salient preprocessing is not as good as the pose estimation based on the original image (both after being processed by RANSAC). The

matching feature point pairs in the original image after RANSAC processing in Figure 6(c) with more matching feature points and have a stronger anti-interference for feature point noise than the image after salient preprocessing and RANSAC processing in Figure 6(c)#.

4.4. Comparison Test Analysis

In this part, we use parallel contrast experiments to illustrate the performance of the proposed algorithm in feature extraction, feature point matching, time-consuming of pose estimation, and so on.

Table 6. Information details of Figure 6.

Category	Extracted Key Points	Feature Point Matching Pairs	Number of Matching Pairs after RANSAC Processing
Before salient preprocessing	Left IMG 1629, Right IMG 1602	591	321
Salient preprocessing	Left IMG 885, Right IMG 592	164	29

In this section, we will illustrate the effectiveness of the coarse-to-fine way in our algorithm in Table 7. After salient preprocessing, the matching feature pairs in the images are further improved by RANSAC processing to eliminate the wrong matching feature points and improve the matching quality.

Table 7. The salient preprocessing can preliminarily reduce the incorrectly matching feature point pairs, and then the RANSAC processing can further eliminate the incorrectly matching feature point pairs.

Category	The Incorrectly Matching Logarithm of Figure 3 (before RANSAC Processing)	The Incorrectly Matching Logarithm of Figure 3 (after RANSAC Processing)	600 Pairs of Images
Without salient preprocessing	172	The incorrect matching pairs difference is not apparent before and after salient preprocessing.	Over 95% of image pairs reduce the incorrectly matching feature point pairs containing the process from salient preprocessing to RANSAC processing.
Salient preprocessing	160		

Salient preprocessing can greatly reduce the number of wrong matching logarithms before RANSAC processing, which is convenient for improving the matching quality of feature points in the scene in a real-time way. The difference in the incorrectly matching logarithm between the image after salient preprocessing (with RANSAC) and the original image (with RANSAC) is not obvious in most pairs. That is because RANSAC can finely remove the mismatching logarithm, which weakens the advantages of salient processing. However, we found that the average error matching logarithm after the salient preprocessing (with RANSAC) is slightly lower than that in the original image (with RANSAC). That is because incorrectly matching logarithms being obviously reduced after the salient preprocessing (with RANSAC) affects the average value of incorrectly matching logarithms.

Due to the lack of the ground truth of matching feature point pairs, we manually count the incorrectly matching feature point pairs after enlarging the picture. It can be seen from Table 7 that there are fewer incorrectly matching feature point pairs after salient preprocessing than those without salient preprocessing (before RANSAC processing). Furthermore, we take Figure 3 as an example to calculate the incorrectly matching feature point pairs, in which neither the salient preprocessed nor the non-salient processed images have obviously incorrectly matching feature point pairs after RANSAC processing and are not detected by us. Nevertheless, we take Figure 6 as an example. There are no obvious wrong matching pairs in Figure 6(c)# after the salient preprocessing using RANSAC processing compared with some incorrect matching feature pairs detected in the original image using RANSAC processing. Similar blurry contrast observations were made in 600 image pairs with an accuracy of over 95% in reducing the incorrectly matching feature point pairs containing the process of salient preprocessing and RANSAC processing. The less than 5% part is mainly the case because there is no apparent difference in the number of

incorrectly matching feature point pairs that the naked eye identifies, and the salient object has a symmetrical contour and lacks texture information after salient preprocessing. The experiments demonstrate that the proposed method can obtain fewer incorrectly matching feature points after salient preprocessing. Further, it can reduce the incorrectly matching feature points in a more detailed manner based on RANSAC. It is worth noting that the total number of matching pairs obtained after salient preprocessing and RANSAC processing in Figure 3(c) is 12 pairs, which is less than the 30 matching pairs in Figure 3(c) without salient preprocessing. This means our method is used for solving pose estimation and can save time based on fewer matching feature point pairs.

The following compares the feature extraction, feature matching, and time-consuming pose solution of the SIFT+RANSAC before and after salient preprocessing.

In our experiment, the average number of feature point extraction and matching decreased after salient preprocessing. The average number of extracted feature points in Table 8 is based on the total number of feature points extracted in all images in the data set divided by the number of images. The average matching logarithm of feature points before and after RANSAC processing in Table 8 is the sum of the matching logarithm of all feature points in the 600 image pairs divided by the number of image pairs. The average calculation time in this paper is the average value of the time spent in the pose estimation based on ICP, including SIFT feature point extraction and matching, RANSAC processing, and pose estimation.

Table 8. Comparison of feature point extraction and matching before and after salient preprocessing.

Category	The Average Number of Extracted Feature Points (One Image).	Average Matching Feature Point Pairs (One Pair of Images).	Match Logarithms after RANSAC Processing (One Pair of Images).	Average Pose Estimation Time (One Pair of Images).
Without salient preprocessing	522.8	181.1	69.4	1.92s
Salient preprocessing	351.5	123.7	43.3	1.49 s
Percentage of saved time				22.40%

The average calculation time in this part is the average value of the time spent in the pose estimation according to the matching feature point pairs based on the ICP analyzed in Section 3.3.

The percentage of saved time after salient preprocessing in Table 8 is the ratio of the average time difference spent on the pose solution before and after salient preprocessing to the average time spent on the pose solution before salient preprocessing. The formula is as follows:

$$Time_{saved} = \frac{Time_{without\ processing} - Time_{processing}}{Time_{without\ processing}}, \quad (2)$$

where $Time_{saved}$ means the percentage of saved time after salient preprocessing, $Time_{without\ processing}$ means average time spent on the pose solution before salient preprocessing, and $Time_{processing}$ means average time spent on the pose solution after salient preprocessing.

Through comparative experiments, it can be seen that masking the image background through salient preprocessing can effectively reduce the feature point extraction and matching range. The matching logarithm of feature points after RANSAC processing is also reduced. Our method reduces the time of solving the pose by 22.40% on average because the salient preprocessing algorithm proposed in this paper can reduce the number of SIFT feature point extraction and matching. In theory, our method can also be realized by other types of feature points.

In addition to comparing the pose estimation based on the SIFT + RANSAC method before and after salient preprocessing, we also compare the pose estimation algorithms based on ORB+RANSAC and SURF + RANSAC, respectively, as shown in Table 9.

Table 9. The performance comparison of the proposed method and related methods.

Method	The Average Number of Extracted Feature Points (One Image).	Average Matching Feature Point Pairs (One Pair of Images).	Match Logarithms after RANSAC Processing (One Pair of Images).	Average Pose Estimation Time (One Pair of Images)	Relative Pose Error (RPE)
Ous	351.5	123.7	43.3	1.49 s	0.057 m/s, 1.512 deg/s
SIFT+RANSAC [19,40,41]	522.8	181.1	69.4	1.92s	0.054 m/s, 1.467 deg/s
ORB+RANSAC [40,41,43]	494	177	59	0.36s	0.076 m/s, 1.791 deg/s
SURF+RANSAC [40,41,44,45]	353	159	62	0.84 s	0.062 m/s, 1.619 deg/s

In the experiment in the table, we use a pair of images as the statistics unit to calculate the data. Furthermore, we use the 600 pairs of images described above as the data set to calculate the average matching feature point pairs, pose error, etc., and take the average value. It should be noted that some image pairs with poor matching effects in the data set will affect the average relative pose error performance. All the experiments in this paper use GPU to accelerate the feature matching and ICP pose estimation to shorten the running time. Each kind of comparative experiment here is based on the content of several articles because we cannot rely only on one article to realize the operation of the comparative experiment. The parameter settings of feature extraction and RANSAC processing in the parallel comparison experiment are consistent. The relative pose error contains translation and rotation errors and can be assisted using the EVO tool [46].

The calculation method and details of the relative pose error (RPE) are involved in Table 9.

The TUM data set contains the ground truth trajectory of the camera (including timestamp and actual pose). Then, under the same timestamp, we calculate the relative pose of two image frames before and after the salient preprocessing. This paper uses the relative pose error to calculate the difference between the pose changes within the identical two timestamps. After aligning timestamps, both the actual pose and the estimated pose can be obtained at the same time interval. We are, according to Equation (3), to obtain the relative pose error. The relative pose error is calculated according to the inputted actual pose value and the estimated pose.

The relative pose error $E_{R,i}$ of frame i is defined as follows:

$$E_{R,i} = \left(Q_i^{-1}Q_{i+\Delta}\right)^{-1} \left(P_i^{-1}P_{i+\Delta}\right), \quad (3)$$

where Q_i is the actual value of the pose of the image, P_i is the estimated value of the pose, and Δ represents the interval time. Then, the actual value of the image pose after Δ time is $Q_{i+\Delta}$ and the estimated value $P_{i+\Delta}$. Our experiment calculates the pose based on 600 pairs of images in the data set. The actual pose and the estimated pose of the previous frame are the same ($Q_i = P_i$), and after the Δ time interval, the true value of the pose $Q_{i+\Delta}$ and the estimated value of the pose $P_{i+\Delta}$ are not the same. The Δ of the 600 image pairs in the used data sets is the interval time of two image frames with clear salient imaging, which is much larger than the Δ in the two adjacent frames of the standard data set. That is why this article's relative pose error data are universally larger no matter what feature is used.

In this paper, ORB+RANSAC has the largest relative pose error, but the best real-time performance. We analyze that this is because ORB feature extraction is fast and in real time. However, the disadvantage is that ORB features do not have scale and rotation invariance and are sensitive to noise.

In our experiments, the SURF has certain rotation and scale transformation robustness, and its real-time performance is better than SIFT. Although SIFT takes longer, the feature extraction and matching quality are much higher than SURF. Our proposed method is based on SIFT because the pose estimation using SIFT features can extract some local features of the image. Moreover, the SIFT feature has the advantage of maintaining invariance to rotation, scale scaling, brightness changes, and maintaining a certain degree of stability to angle changes and noise. Using the SIFT feature can more effectively identify an object

in different scenes and be more accurate for pose estimation. It can be seen from Table 9 that SIFT+RANSAC has better relative translation error and rotation error performance than ORB+RANSAC and SURF+RANSAC, no matter whether after salient preprocessing. The SURF feature has less robustness and better real-time performance than SIFT, which makes the pose estimation based on SURF+RANSAC have better real-time performance and a larger relative pose error than SIFT+RANSAC. The SURF feature has better feature robustness and worse real-time performance than the ORB feature. So, the pose estimation based on SURF+RANSAC in Table 9 takes more time to calculate the pose, but has a smaller relative error than the ORB-RANSAC-based pose estimation.

The SIFT-RANSAC with salient preprocessing has fewer relative translation and rotation errors than ORB-RANSAC and SURF-RANSAC. That is because the SIFT feature is the most time-consuming but robust feature. After the salient preprocessing, the number of extraction and matching logarithms of feature points is greatly reduced compared with the SIFT+RANSAC without salient preprocessing, and the time-consuming result is reduced. However, there is no advantage in the time-consuming amount compared to ORB-RANSAC and SURF-RANSAC because there is no salient preprocessing step in the other related comparison experiments. Nevertheless, our proposed method is based on salient preprocessing to reduce feature points, so the amount of time consumed in salient preprocessing needs to be considered.

In our comparative experiments, the method proposed in this paper can effectively improve real-time performance based on SIFT features. However, translation and rotation errors are larger than the SIFT+RANSAC without salient preprocessing. It can be seen from Table 7 that there are fewer incorrectly matching feature point pairs after salient preprocessing than in the comparative experiment without salient preprocessing. Most image pairs have improved their pose estimation after our salient preprocessing. However, where a small part of individual pose estimation values is low, this makes the overall average RPE values lower than those without salient preprocessing. The reasons that the RPE becomes larger after salient preprocessing are analyzed as follows: (1) our proposed method can reduce the number of incorrectly matching logarithms. However, the total number of matching logarithms is also reduced, and the pose estimation is more susceptible to receiving interference from incorrectly matching point pairs; and (2) the experiments show that the relative pose error in some cases becomes larger because they receive interference from the main influencing factors, such as those shown in Figure 5, where the edge of the salient objects' imaging contour has relatively rich texture information after salient preprocessing when the salient objects have symmetrical outlines and lack texture information. The texture parts have more feature points that can be extracted and more mismatching feature pairs. The mismatching feature points on the salient object's symmetrical contour will affect the pose estimation and worsen the average value of the relative pose error.

We further use the YCB-Video data set to verify our above comparative analysis (in Table 10). The objects in the YCB-Video scene have high-quality 3D models and good visibility in depth.

Table 10. Comparative tests on different data sets.

Method	Average Pose Estimation Time (One Pair of Images, TUM)	Relative Pose Error (RPE) (TUM [42])	Average Pose Estimation Time (One Pair of Images, YCB-Video)	Relative Pose Error (RPE) (YCB-Video [47])
Ous	1.49 s	0.057 m/s, 1.512 deg/s	1.32 s	0.033 m/s 1.023 deg/s
SIFT+RANSAC [19,40,41]	1.92 s	0.054 m/s, 1.467 deg/s	1.97 s	0.032 m/s 0.987 deg/s
ORB+RANSAC [40,41,43]	0.36 s	0.076 m/s, 1.791 deg/s	0.24 s	0.062 m/s 1.545 deg/s
SURF+RANSAC [40,41,44,45]	0.84 s	0.062 m/s, 1.619 deg/s	0.67 s	0.041 m/s 1.203 deg/s

The image size of YCB-Video and TUM are both 640 * 480, and the experimental conditions are consistent (YCB-Video is only temporarily used in Table 10).

The test results on the YCB-Video data set further verify our analysis conclusion of the comparative experiment of pose estimation based on the different features above. The lack of background features after salient processing will decrease the correspondence between frames, which will harm the robustness. That is why our proposed method is consistently slightly worse than SIFT+RANSAC in relative pose estimation, but there is no doubt that our proposed method has better real-time performance than other methods while ensuring good pose estimation. Our method based on YCB-Video has less relative pose error than TUM. We infer this is because the objects in YCB-Video have high-quality 3D models and good visibility in depth, which is conducive to the extraction of high-quality spatial feature points for pose estimation. The shape and color of these high-quality objects are different from the background and are located in the center of the field of view, which is easier to be detected as salient objects and is more conducive to reliable feature extraction on salient objects. The performance of various methods without salient preprocessing based on YCB-Video is generally better than that of the TUM data set. This is because the objects in YCB-Video have high-quality 3D models and good visibility in depth. This is conducive to extracting high-quality spatial feature points on these objects for pose estimation.

This paper proposes to reduce the computational complexity of pose estimation and improve the feature extraction and matching quality in a coarse-to-fine way for the first time. It verifies the feasibility of the proposed scheme through comparative experiments. The method proposed in this paper has good relative translation error and rotation error and performs well in terms of the amount of time consumed on the whole, which reduces the relative pose estimation error and is more robust compared with the pose estimation based on ORB and SURF features, respectively, and has better real-time performance than the traditional pose estimation based on SIFT features. Although our proposed algorithm can save computation regarding feature extraction and feature point matching, it is based on the premise of clear salient imaging. Furthermore, the experiments show that the salient preprocessing's imaging quality and real-time performance will impact the proposed method. With the improvement of salient detection neural network architecture, the practicability of our proposed method for pose estimation based on salient preprocessing will also be further improved.

5. Conclusions

The extraction and matching of feature points for the traditional ICP pose estimation process are time-consuming and lack robustness. In order to balance these two challenges, we first propose a coarse-to-fine method. After salient preprocessing, the matching SIFT feature pairs in the images are further improved by RANSAC processing to eliminate the wrong matching feature points and improve the matching quality. The influencing factors that affect the pose estimation quality after salient preprocessing are analyzed experimentally. The proposed algorithm is influenced by the difference between image frames, salient objects' contour plus texture, and salient preprocessing's detection imaging quality. We analyze the three situations and infer that the relative pose error becomes larger because the feature points cannot be matched effectively. Our method is more suitable for static and texture-rich asymmetric object scenes by salient preprocessing. This paper verifies the advantages of the proposed algorithm in the time-consuming amount by comparing the performance differences based on the SIFT feature before and after salient preprocessing. This paper also compares our method with the mainstream algorithm based on ORB or SURF. The experimental results show that our algorithm processing based on the same kind of feature can effectively reduce the feature points' extraction and matching and reduce the pose estimation time. Moreover, our proposed algorithm has improved pose estimation compared with the mainstream algorithm. In this paper, we propose a new feature point selection method, which uses salient objects' feature points to replace the entire image's feature points to improve the real-time performance and provide a reference for real-time pose estimation research.

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

Funding: This research is supported in part by the Research Project of China Disabled Persons Federation on assistive technology under Grant 2022CDPFAT-01, in part by the Science and Technology Planning Project of Chongqing Changshou District under Grant cskj2022014, in part by the National Nature Science Foundation of China under Grant 51775076, 51905065, and in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant KJ1704072 and KJZD-K201903801.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Mo, J.; Islam, J.; Sattar, J. Fast Direct Stereo Visual SLAM. *IEEE Robot. Autom. Lett.* **2022**, *7*, 778–785. [\[CrossRef\]](#)
- Zeng, Y.; Jiang, Y. Weight Algorithm Based Depth Camera Point-to-Plane ICP Algorithm. In Proceedings of the 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, CYBER 2021, Jiaying, China, 27–31 July 2021.
- Li, J.; Hu, Q.; Zhang, Y.; Ai, M. Robust symmetric iterative closest point. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 219–231. [\[CrossRef\]](#)
- Wang, J.; Xu, M.; Foroughi, F.; Dai, D.; Chen, Z. FasterGICP: Acceptance-Rejection Sampling Based 3D Lidar Odometry. *IEEE Robot. Autom. Lett.* **2022**, *7*, 255–262. [\[CrossRef\]](#)
- Pavan, N.L.; dos Santos, D.R.; Khoshelham, K. Global Registration of Terrestrial Laser Scanner Point Clouds Using Plane-to-Plane Correspondences. *Remote Sens.* **2020**, *12*, 1127. [\[CrossRef\]](#)
- Min, Z.; Wang, J.; Pan, J.; Meng, M.Q.-H. Generalized 3-D Point Set Registration with Hybrid Mixture Models for Computer-Assisted Orthopedic Surgery: From Isotropic to Anisotropic Positional Error. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1679–1691. [\[CrossRef\]](#)
- Makovetskii, A.; Voronin, S.; Kober, V.; Voronin, A. A regularized point cloud registration approach for orthogonal transformations. *J. Glob. Optim.* **2022**, *83*, 497–519. [\[CrossRef\]](#)
- Gao, X.; Zhang, T.; Liu, Y.; Yan, Q. *14 Lectures on Visual SLAM: From Theory to Practice*; Publishing House of Electronics Industry: Beijing, China, 2017.
- Serafin, J.; Grisetti, G. NICP: Dense Normal Based Point Cloud Registration. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Hamburg, Germany, 28 September–2 October 2015; Volume 2015.
- Jia, S.; Ding, M.; Zhang, G.; Li, X. Improved Normal Iterative Closest Point Algorithm with Multi-Information. In Proceedings of the 2016 IEEE International Conference on Information and Automation, IEEE ICIA 2016, Ningbo, China, 1–3 August 2016.
- Pomerleau, F.; Colas, F.; Siegwart, R.; Magnenat, S. Comparing ICP variants on real-world data sets. *Auton. Robots* **2013**, *34*, 133–148. [\[CrossRef\]](#)
- Rusinkiewicz, S.; Levoy, M. Efficient variants of the ICP algorithm. In Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling, Quebec City, QC, Canada, 28 May–1 June 2001. [\[CrossRef\]](#)
- Rodolà, E.; Albarelli, A.; Cremers, D.; Torsello, A. A simple and effective relevance-based point sampling for 3D shapes. *Pattern Recognit. Lett.* **2015**, *59*, 41–47. [\[CrossRef\]](#)
- Servos, J.; Waslander, S.L. Multi-Channel Generalized-ICP: A robust framework for multi-channel scan registration. *Robot. Auton. Syst.* **2017**, *87*, 247–257. [\[CrossRef\]](#)
- Zhu, Z.; Xiang, W.; Huo, J.; Yang, M.; Zhang, G.; Wei, L. Non-Cooperative Target Pose Estimation based on Improved Iterative Closest Point Algorithm. *J. Syst. Eng. Electron.* **2022**, *33*, 1–10. [\[CrossRef\]](#)
- Yue, X.; Liu, Z.; Zhu, J.; Gao, X.; Yang, B.; Tian, Y. Coarse-fine point cloud registration based on local point-pair features and the iterative closest point algorithm. *Appl. Intell.* **2022**, 12569–12583. [\[CrossRef\]](#)
- Yu, J.; Yu, C.; Lin, C.; Wei, F. Improved Iterative Closest Point (ICP) Point Cloud Registration Algorithm based on Matching Point Pair Quadratic Filtering. In Proceedings of the 2021 International Conference on Computer, Internet of Things and Control Engineering, CITCE, Guangzhou, China, 12–14 November 2021; pp. 1–5. [\[CrossRef\]](#)
- Ran, Y.; Xu, X. Point cloud registration method based on SIFT and geometry feature. *Optik* **2020**, *203*, 163902. [\[CrossRef\]](#)
- Hossein-Nejad, Z.; Nasri, M. An adaptive image registration method based on SIFT features and RANSAC transform. *Comput. Electr. Eng.* **2017**, *62*, 524–537. [\[CrossRef\]](#)
- Stanković, I.; Brajović, M.; Lerga, J.; Daković, M.; Stanković, L. Image denoising using RANSAC and compressive sensing. *Multimed. Tools Appl.* **2022**, *81*, 44311–44333. [\[CrossRef\]](#)

21. Li, J.; Hu, Q.; Ai, M. Point Cloud Registration Based on One-Point RANSAC and Scale-Annealing Biweight Estimation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9716–9729. [[CrossRef](#)]
22. Borji, A.; Cheng, M.-M.; Hou, Q.; Jiang, H.; Li, J. Salient object detection: A survey. *Comput. Vis. Media* **2019**, *5*, 117–150. [[CrossRef](#)]
23. Zhuge, M.; Fan, D.P.; Liu, N.; Zhang, D.; Xu, D.; Shao, L. Salient object detection via integrity learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *1*. [[CrossRef](#)]
24. Hu, L.; Zhang, Y.; Wang, Y.; Jiang, Q.; Ge, G.; Wang, W. A simple information fusion method provides the obstacle with saliency labeling as a landmark in robotic mapping. *Alex. Eng. J.* **2022**, *61*, 12061–12074. [[CrossRef](#)]
25. Wang, R.; Su, C.; Yu, H.; Wang, S. Six-dimensional Target Pose Estimation for Robot Autonomous Manipulation: Methodology and Verification. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *1*. [[CrossRef](#)]
26. Du, T.; Shi, S.; Zeng, Y.; Yang, J.; Guo, L. An Integrated INS/Lidar Odometry/Polarized Camera Pose Estimation via Factor Graph Optimization for Sparse Environment. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [[CrossRef](#)]
27. Anderson, J.D.; Raettig, R.M.; Larson, J.; Nykl, S.L.; Taylor, C.N.; Wischgoll, T. Delaunay walk for fast nearest neighbor: Accelerating correspondence matching for ICP. *Mach. Vis. Appl.* **2022**, *33*, 31. [[CrossRef](#)]
28. Reyes-Aviles, F.; Fleck, P.; Schmalstieg, D.; Arth, C. Compact World Anchors: Registration Using Parametric Primitives as Scene Description. *IEEE Trans. Vis. Comput. Graph.* **2022**, *1–13*. [[CrossRef](#)] [[PubMed](#)]
29. Wu, P.; Li, W.; Yan, M. 3D scene reconstruction based on improved ICP algorithm. *Microprocess. Microsystems* **2020**, *75*. [[CrossRef](#)]
30. Wan, T.; Du, S.; Cui, W.; Yao, R.; Ge, Y.; Li, C.; Gao, Y.; Zheng, N. RGB-D Point Cloud Registration Based on Salient Object Detection. *IEEE Trans. Neural Networks Learn. Syst.* **2022**, *33*, 21621947. [[CrossRef](#)] [[PubMed](#)]
31. Yao, R.; Du, S.; Wan, T.; Cui, W. A robust registration algorithm based on salient object detection. *Multimed. Tools Appl.* **2022**, *81*, 34387–34400. [[CrossRef](#)]
32. Wang, W.; Shen, J.; Xie, J.; Cheng, M.-M.; Ling, H.; Borji, A. Revisiting Video Saliency Prediction in the Deep Learning Era. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 220–237. [[CrossRef](#)] [[PubMed](#)]
33. Hong, S.; You, T.; Kwak, S.; Han, B. Online Tracking by Learning Discriminative Saliency Map with Convolutional Neural Network. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015; Volume 1.
34. Li, J.; Pei, L.; Zou, D.; Xia, S.; Wu, Q.; Li, T.; Sun, Z.; Yu, W. Attention-SLAM: A Visual Monocular SLAM Learning From Human Gaze. *IEEE Sens. J.* **2021**, *21*, 6408–6420. [[CrossRef](#)]
35. Liu, J.-J.; Hou, Q.; Cheng, M.-M. Dynamic Feature Integration for Simultaneous Detection of Salient Object, Edge, and Skeleton. *IEEE Trans. Image Process.* **2020**, *29*, 8652–8667. [[CrossRef](#)]
36. Bansal, M.; Kumar, M. 2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimedia Tools Appl.* **2021**, *80*, 18839–18857. [[CrossRef](#)]
37. Yang, J.; Huang, Z.; Quan, S.; Zhang, Q.; Zhang, Y.; Cao, Z. Toward Efficient and Robust Metrics for RANSAC Hypotheses and 3D Rigid Registration. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 893–906. [[CrossRef](#)]
38. Xu, X.; Zhang, L.; Yang, J.; Liu, C.; Xiong, Y.; Luo, M.; Tan, Z.; Liu, B. LiDAR-camera calibration method based on ranging statistical characteristics and improved RANSAC algorithm. *Robot. Auton. Syst.* **2021**, *141*, 103776. [[CrossRef](#)]
39. Maken, F.A.; Ramos, F.; Ott, L. Stein ICP for Uncertainty Estimation in Point Cloud Matching. *IEEE Robot. Autom. Lett.* **2021**, *7*, 1063–1070. [[CrossRef](#)]
40. Mur-Artal, R.; Tardos, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [[CrossRef](#)]
41. Whelan, T.; Johannsson, H.; Kaess, M.; Leonard, J.J.; McDonald, J. Robust Real-Time Visual Odometry for Dense RGB-D Mapping. In Proceedings of the IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013.
42. Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; Cremers, D. A Benchmark for the Evaluation of RGB-D SLAM Systems. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012.
43. Zhang, H.; Zheng, G.; Fu, H. Research on image feature point matching based on ORB and RANSAC algorithm. *J. Phys. Conf. Ser. IOP Publ.* **2020**, *1651*, 012187. [[CrossRef](#)]
44. Zhang, J.; Yin, X.; Luan, J.; Liu, T. An improved vehicle panoramic image generation algorithm. *Multimedia Tools Appl.* **2019**, *78*, 27663–27682. [[CrossRef](#)]
45. Abu Bakar, S.; Jiang, X.; Gui, X.; Li, G.; Li, Z. Image Stitching for Chest Digital Radiography Using the SIFT and SURF Feature Extraction by RANSAC Algorithm. *J. Phys. Conf. Ser.* **2020**, *1624*. [[CrossRef](#)]
46. Michael Grupp. Python package for the evaluation of odometry and SLAM. Available online: <https://github.com/MichaelGrupp/evo> (accessed on 1 October 2022).
47. Xiang, Y.; Schmidt, T.; Narayanan, V.; Fox, D. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. In Proceedings of the Robotics: Science and System XIV, Pittsburgh, PA, USA, 26–30 June 2018. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.