

Article

# Target Image Mask Correction Based on Skeleton Divergence

Yaming Wang, Zhengheng Xu, Wenqing Huang \* , Yonghua Han and Mingfeng Jiang

Pattern Recognition and Computer Vision Lab, Zhejiang Sci-Tech University, Hangzhou 310000, China; yamingwang2000@163.com (Y.W.); 201720601011@mails.zstu.edu.cn (Z.X.); hanyonghuahan@163.com (Y.H.); peterjiang0517@163.com (M.J.)

\* Correspondence: patternrecog@163.com

Received: 25 October 2019; Accepted: 23 November 2019; Published: 25 November 2019



**Abstract:** Traditional approaches to modeling and processing discrete pixels are mainly based on image features or model optimization. These methods often result in excessive shrinkage or expansion of the restored pixel region, inhibiting accurate recovery of the target pixel region shape. This paper proposes a simultaneous source and mask-images optimization model based on skeleton divergence that overcomes these problems. In the proposed model, first, the edge of the entire discrete pixel region is extracted through bilateral filtering. Then, edge information and Delaunay triangulation are used to optimize the entire discrete pixel region. The skeleton is optimized with the skeleton as the local optimization center and the source and mask images are simultaneously optimized through edge guidance. The technique for order of preference by similarity to ideal solution (TOPSIS) and point-cloud regularization verification are subsequently employed to provide the optimal merging strategy and reduce cumulative error. In the regularization verification stage, the model is iteratively simplified via incremental and hierarchical clustering, so that point-cloud sampling is concentrated in the high-curvature region. The results of experiments conducted using the moving-target region in the RGB-depth (RGB-D) data (Technical University of Munich, Germany) indicate that the proposed algorithm is more accurate and suitable for image processing than existing high-performance algorithms.

**Keywords:** skeleton divergence; edge merging; TOPSIS; simultaneous optimization binary; mask correction

## 1. Introduction

Moving-target mask correction is a process by which a complete target object binary mask is recovered using scattered target pixels. Target pixel mask correction based on foreground motion is a key step in both point-cloud motion stitching and graphics stitching. It is also widely used in 3D target reconstruction, image segmentation, and video frame-rate improvement. Researchers have proposed various algorithms for image binary mask correction, most of which focuses on a single strategy [1–14]. In general, these algorithms fall into two categories: image feature-based methods and model-based optimization methods.

Image feature-based methods primarily use the feature information contained in the image to drive mask correction [1]. The splicing of multiple images is achieved using constant local features, which overcomes the influence of the input image's own signal-to-noise ratio and external illumination on the results. However, in the case of mask repair for continuous video frames showing the motion of participants in a scene, there is a lack of precision. Levin et al. [2] presented in detail the image-stitching cost function set in the gradient domain to evaluate the image-stitching quality. Compared with the traditional method, their approach overcomes the influence of luminosity inconsistency, but does not

overcome the influence of the external environment. Zhang and Liu [3] proposed a local splicing method for processing parallax. However, their proposed method lacks robustness in environments with small parallax. Zomet et al. [4] proposed a method that visually measures the quality of stitching using the similarity between the stitched image, the input image, and the visibility of the seam. However, their proposed method is still not robust in scenes with environmental change. Mills and Dudek [5] proposed the use of dynamic elements to stitch and compose images. However, their approach lacks precision in the case of multi-target scenes. Zheng et al. [6] proposed an image-stitching method based on a projected uniform plane. However, their proposed method is computationally intensive, and is thus not suitable for practical scene applications. Fang et al. [7] presented in detail a color-mixing method to eliminate color matching inconsistencies in pre-aligned images. However, the method presented does not overcome the influence of external illumination on experimental results. Consequently, it is computationally expensive, which also affects its practical application.

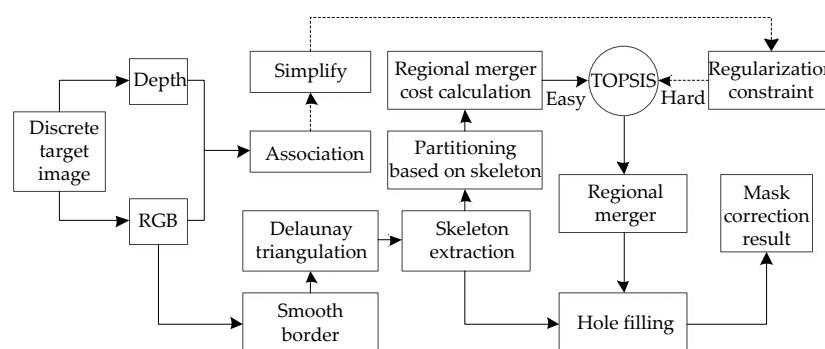
Model-based optimization methods primarily model the mask correction problem and add constraint items to achieve mask correction by minimizing the difference between the mask correction and the real scene. Hu et al. [8] presented a fusion score splicing and feature-matching method to solve the problem of 3D point-cloud stitching quality. However, their algorithm easily falls into local minima in certain complicated environments and does not provide the best stitching effect. Chen et al. [9] used the point-cloud density feature to optimize both stitching accuracy and scene modeling accuracy, but the density model is sensitive to motion interference from dynamic participants in a scene, affecting the stitching accuracy. Jovanovski and Li [10] demonstrated the extraction of depth information from a specific image group and achieved higher precision stitching using depth and image information. However, the scene that is selected by the system is a specific scene, and so the system lacks generalization. Hu et al. [11] proposed reconstruction of 3D point clouds using a continuous frame image and combining matching constraints and image feature scores to achieve high-precision scene stitching. However, owing to the number of point clouds, the calculation demands of the system are high, leading to low system efficiency. Wang et al. [12] discussed in detail the use of a transformation matrix obtained by key point matching to improve the stitching precision of a scene, although the selection of key points was affected by dynamic participants in the scene, leading to low robustness in practical applications. Wang et al. [13] used the dotted-line feature of a 2D image to perform closed-loop detection and achieve high-precision scene stitching reconstruction through probability conversion of the octree structure. However, the point-line feature constraint of the system was complicated, not robust, and invalid in terms of the original scene. Koller [14] proposed a method for image filtering 3D objects using 2D manifolds, although the calculation requirements of their system were high, and system efficiency was low.

In this study, a target image mask correction method based on skeleton divergence is proposed. The skeleton and discrete pixel region edge constraints are used to avoid excessive shrinkage and expansion during pixel region recovery, which ensures high precision of discrete pixel shaping. The edge-guided strategy is used to merge the local pixel regions, and the sparse point-cloud regularization constraints and TOPSIS (technique for order of preference by similarity to ideal solution) [15] decision-making are used to reduce the error accumulation caused by the combination cost error and improve the robustness of the system. The main contributions of this paper are as follows. Firstly, the idea of skeleton divergence is applied to mask correction, and the idea based on the skeleton is a good way to reduce the problem of excessive contraction and excessive expansion. Secondly, based on the work of predecessors, a complete high-precision mask correction system has been presented. We compare our algorithm with some existing algorithms. Experimental results prove our algorithm can achieve higher mask correction accuracy and efficiency.

The remainder of this paper is organized as follows. Section 2 introduces each part of the system. Section 3 outlines the experiment conducted and analyzes the results obtained. Section 4 presents our concluding remarks.

## 2. Materials and Methods

We propose a simultaneous source and mask-images optimization model to model and process discrete pixels, which is based on skeleton divergence, as shown in Figure 1. Firstly, the boundary curve of the discrete pixel region is represented by scale-invariant intrinsic variables, and then the boundary of the discrete pixel region is smoothed by the angle bilateral filtering based on the directional corner. After a new set of directional corner parameters are obtained and utilized, the new vertex coordinate equation of the smoothed boundary curve is established to reconstruct the smoothed boundary curve. Secondly, based on the Delaunay triangulation [16], the boundary discrete curve is utilized to extract the skeleton of the global discrete region, and partition discrete pixel regions using a single connected region surrounded by skeletons and boundaries. Then, taking the skeleton as the divergent center of the internal pixel combination, the discrete pixel edge similarity of the local region is used to guide the internal discrete pixel merging. Thirdly, in order to reduce the error accumulation and improve the region merge accuracy, the technique for order of preference by similarity to ideal solution (TOPSIS) [15] is employed to make a strategic decision in the screening stage of the candidate merge strategies. According to the value of the minimum merge cost of each candidate merge strategy, it is judged whether it is "hard" or "easy" to choose the optimal from many candidate merge strategies. When it is easy to make a choice, the calculated three single-channel statistical costs are used. The three values of single-channel statistical cost and the minimum merge cost are considered as evaluation indicators as well as strategies to sort and then get the optimal strategy. When it is hard to make a choice, the point cloud regularization constraint results corresponding to the candidate merger strategies are added into the original evaluation indicator set for an optimal merge strategy. Fourth, considering the huge computational complexity of dealing with dense point clouds, the acquisition of the matching degree of the point cloud regularization constraint is based on the sparse point cloud. In the same way as the method in [17], with the help of incremental and hierarchical clustering, and iterative simplification, these point clouds are simplified in advance so that they are concentrated in the high-curvature region, a surface deformation model is established [18], and then the system accuracy is guaranteed while reducing the amount of calculation in the process. Finally, with the regional constraints of the global skeleton and boundary, combined with the morphological features and structural similarity, the hole filling of the merged discrete pixel regions is performed to improve the accuracy of the binary mask correction and efficiency.



**Figure 1.** The proposed mask correction algorithm flow. TOPSIS, technique for order of preference by similarity to ideal solution.

### 2.1. Mask Optimization Based on Skeleton Divergence

Boundary denoising and smoothing operations are performed on the original and mask images. After extracting the skeletons of the source and mask images, the images are optimized using the image constraints near the image skeleton. This study employs the angle-based filtering method in the process of boundary smoothing, using the scale-invariant intrinsic variables of the edge pixel connection as the filtering independent variable [19]. The intrinsic amount of expansion is the reference

quantity that represents the spatial movement and expansion of the image edge vector unit, in which the directional corner can reflect the trend and degree of curvature of the global edge vector and can reflect the smoothness of the local plane image. The angle of the edge vectors is subject to bilateral filtering, not only because of the application scenario and method of [20], but also because bilateral filtering considers the spatial distance difference between edge vectors. The influence of the difference in the edge vector angle is determined, and the set of intrinsic amounts of expansion, including the new directional corner parameter, may be obtained, as shown in Figure 2. The scale-invariant intrinsic variables of the discrete curve are expressed as

$$\Omega = \{(r_i, \theta_i)\}_{i=1}^{n-1}, \quad (1)$$

and

$$\begin{cases} r_i = l_i / l_{i-1} = \|\overline{v_{i+1}v_i}\| / \|\overline{v_{i-1}v_i}\| \\ \theta_i \in (-\pi, \pi] \end{cases}, \quad (2)$$

where  $r_i$  is the ratio of the modulus lengths of the two edge vectors  $\overline{v_{i+1}v_i}$ ,  $\overline{v_{i-1}v_i}$  which has the common vertex  $v_i$ , and  $\theta_i$  is the directed rotation angle of the two adjacent edge vectors. The counterclockwise direction is positive. Take the vertex  $v_i$  as an example and the discrete representation of the angle bilateral filtering is as follows [18].

$$\theta_i^* = \begin{cases} \frac{\theta_1 + c(v_2, v_1)s(\theta_2, \theta_1)\theta_2}{1 + c(v_2, v_1)s(\theta_2, \theta_1)}, i = 1 \\ \frac{c(v_{i-1}, v_i)s(\theta_{i-1}, \theta_i)\theta_{i-1} + \theta_i + c(v_{i+1}, v_i)s(\theta_{i+1}, \theta_i)\theta_{i+1}}{c(v_{i-1}, v_i)s(\theta_{i-1}, \theta_i) + 1 + c(v_{i+1}, v_i)s(\theta_{i+1}, \theta_i)}, i = 2, 3, \dots, n-2 \\ \frac{c(v_{n-2}, v_{n-1})s(\theta_{n-2}, \theta_{n-1})\theta_{n-2} + \theta_{n-1}}{1 + c(v_{n-2}, v_{n-1})s(\theta_{n-2}, \theta_{n-1})}, i = n-1 \end{cases}, \quad (3)$$

where  $\theta_i^*$  is the new directional corner at the vertex  $v_i$  of the curve after bilateral filtering. Let  $c(v_j, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d(v_j, v_i)}{\sigma_d}\right)^2\right)$  and  $d(v_j, v_i)$  be the Euclidean distance between the two vertices  $v_j, v_i$ ;  $s(\theta_j, \theta_i) = \exp\left(-\frac{1}{2}\left(\frac{\delta(\theta_j, \theta_i)}{\sigma_r}\right)^2\right)$  and  $\delta(\theta_j, \theta_i) = |\theta_j - \theta_i|$  is the difference between the directed corners at the two vertices  $v_j, v_i$ . The size of the weight factor  $\sigma_d, \sigma_r$  depends on the specific situation. The larger the  $\sigma_d$ , the greater the effect of the distance size. The larger the  $\sigma_r$ , the greater the effect of the angle difference. This study uses the same weight factor settings as [20]:  $\sigma_d = 4.5, \sigma_r = 2.5$ .

With this new set of scale-invariant intrinsic variables, a set of vertex-coordinate linear equations is then established to recover the smooth image edge vector. The process of solving the vertex-coordinate linear equations may provide no solution. Suppose the curve after smoothing is  $C^*$ , its vertex is  $v_i^*(x_i^*, y_i^*)$ , and its scale-invariant intrinsic variables are expressed as

$$\Omega^* = \{(r_i^*, \theta_i^*)\}_{i=1}^{n-1}. \quad (4)$$

The relationship between the adjacent sides  $\overline{v_{i+1}^*v_i^*}, \overline{v_i^*v_{i-1}^*}$  of the common vertex  $v_i^*$  is

$$\overline{v_{i+1}^*v_i^*} = r_i^* R(\theta_i^*) \overline{v_i^*v_{i-1}^*}. \quad (5)$$

Among them,

$$R(\theta_i^*) = \begin{pmatrix} \cos \theta_i^* & -\sin \theta_i^* \\ \sin \theta_i^* & \cos \theta_i^* \end{pmatrix}. \quad (6)$$

Then Equation (5) can be rewritten as

$$\begin{pmatrix} x_{i+1}^* - x_i^* \\ y_{i+1}^* - y_i^* \end{pmatrix} = r_i^* \begin{pmatrix} \cos \theta_i^* & -\sin \theta_i^* \\ \sin \theta_i^* & \cos \theta_i^* \end{pmatrix} \begin{pmatrix} x_i^* - x_{i-1}^* \\ y_i^* - y_{i-1}^* \end{pmatrix}. \quad (7)$$

Equation (7) is a set of two linear equations for the vertex coordinates  $x_{i-1}^*, y_{i-1}^*, x_i^*, y_i^*, x_{i+1}^*, y_{i+1}^*$ , according to the vertex subscripts  $i = 1, 2, \dots, n-1$  in sequence, a linear system of equations can be obtained:

$$\Upsilon Z = 0, \quad (8)$$

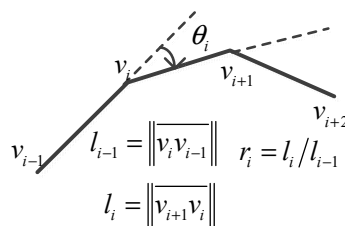
where  $\Upsilon$  is a matrix of order  $(2n-2) \times (2n+2)$ ,  $Z$  is a vector consisting of all vertex coordinates of curve  $C^*$ .  $Z = (x_0^*, y_0^*, x_1^*, y_1^*, \dots, x_{n-1}^*, y_{n-1}^*, x_n^*, y_n^*)^T$ , as can be seen from the literature [20], the matrix  $\Upsilon$  is a row full-rank matrix, its rank is  $2n-1$ . In this study, the coordinates of two vertices  $v_x, v_y$  are set in advance as the benchmark for mask correction, that is, the coordinates of the two vertices of  $v_x^*, v_y^*$  are the same as  $v_x, v_y$ , respectively. Substituting the coordinates of the two points into the linear Equation (5), a new system of linear equations can be obtained:

$$\bar{\Upsilon} Z = b_0, \quad (9)$$

where  $\bar{\Upsilon}$  is a sparse matrix of  $(2n+2) \times (2n+2)$  orders, and  $b_0$  is a  $(2n+2)$  dimensional column vector. Since  $\bar{\Upsilon}$  is a full-rank square matrix, the solution of the linear Equation (6) is unique, and the vertex coordinate set of the curve  $C^*$  after smoothing is obtained. Using the smoothed edge vector set, the image skeleton is then extracted simultaneously from the source and mask images, and skeleton processing proceeds using Delaunay triangulation [16]. The skeleton is then used as the center for local mask correction and localized area merging with edge guidance, for simultaneous optimization of source and mask images. Similar to the method in [20], in this study, the skeleton extraction is divided into three steps, which are briefly described as follows.

1. Delaunay triangulation is performed on the polygons enclosed by the obtained boundary vertices, and the Voronoi diagram is obtained at the same time.
2. The watershed algorithm is used to detect the raised points on the boundary.
3. With the obtained dual map and raised point information, part of the triangular piece is deleted according to the principle of homotopy equivalence, and the final approximate skeleton is obtained.

As the pixel grayscale gradient and edge guidance information near the skeleton exhibit similar gradients, the effect of edge guidance is optimized while reducing the required search area. At the same time, the extraction of the skeleton also ensures that the shapes of the source and mask images are not subject to global structural distortion. Regularization constraint verification of the sparse point cloud was also conducted, which reduced error accumulation in the mask correction process without requiring a significant amount of calculation.



**Figure 2.** The scale-invariant intrinsic variables representation of discrete curves.

## 2.2. Discrete Pixel Edge Similarity-Guided Merging

Assuming that each single-channel amplitude or intensity format image of an RGB image, subject to a fully developed speckle pattern, can be represented by the following multiplicative noise model [21].

$$I(x, y) = \varepsilon(x, y) \xi(x, y), n \in \mathbb{Z}^2, \quad (10)$$

where  $I(x, y)$  represents the single-channel information modeling of the image,  $n = (x, y)$  representing the pixel coordinates,  $\varepsilon(x, y)$  is the reflection coefficient of the real scene, and  $\xi(x, y)$  is the multiplicative coherence coefficient.

We treat the image separately under each channel. When the edge vector and skeleton of the source and mask images are extracted, the pixels in the edge and neighborhood of the skeleton are preferentially optimized, and discrete pixel regions are merged into a larger pixel region to achieve high efficiency and precision. In terms of discrete pixel arrangement optimization, this study introduces a common boundary length penalty region-merging technique for regional merging of motion-compensated fragmented regions, to achieve discrete pixel mask correction. First, the region similarity function is constructed using the first-order statistic of the edge information of the discrete pixel region. The region adjacency graph (RAG) and nearest neighbor graph (NNG) are then used to represent the regional relationship before and after the region-merging process. Finally, the NNG of the region is used to rapidly search for the local adjacent region with the least cost, merge the regions, update the RAG and NNG, and iteratively continue the process until the minimum merge cost is less than a preset merge cost error threshold. The preset merge cost error threshold is the same as the setting in [21], and is set to 1.2. The minimum merge cost of discrete RGB pixels is the sum of three single-channel minimized merge costs. Take the red channel image as an example, the minimized merge cost function for discrete pixel merging is as follows [21].

$$E(I_{P_1^r}(x, y), I_{P_2^r}(x, y)) = E_S(I_{P_1^r}(x, y), I_{P_2^r}(x, y)) + \eta \Gamma_S(P_1^r, P_2^r), \quad (11)$$

where  $P_1, P_2$  represent discrete RGB pixel regions.  $P_1^r, P_2^r, P_1^g, P_2^g, P_1^b, P_2^b$  represent their respective discrete single-channel pixel regions.  $E(I_{P_1^r}(x, y), I_{P_2^r}(x, y))$  is the single-channel merge cost function and  $E_S(I_{P_1^r}(x, y), I_{P_2^r}(x, y))$  is the single-channel statistical similarity cost, which describes the statistical similarity between the combined result and the image data.  $\Gamma_S(P_1^r, P_2^r)$  is the single-channel skeleton and boundary deformation cost, which is the deformation cost of the combined results on the skeleton and the boundary. As discrete pixel regions merge, the statistical cost increases, and the deformation costs of the skeleton and boundary gradually decreases.  $\eta$  is the normalization parameter, empirically set to 1.8. Starting from an initial state, the designed region-merging strategy is used to combine the adjacent sub-regions with the strongest edge similarity, until the optimization termination condition is satisfied and the final result is output. The single-channel statistical similarity cost is characterized by the statistical similarity measure (SSM).

$$E_S(I_{P_1^r}(x, y), I_{P_2^r}(x, y)) = \frac{\sqrt{L} \left( 1 - \min \left\{ \frac{\bar{m}_1}{\bar{m}_2}, \frac{\bar{m}_2}{\bar{m}_1} \right\} \right)}{\sqrt{\frac{1}{\#P_1^r} \left( \min \left\{ \frac{\bar{m}_1}{\bar{m}_2}, 1 \right\} \right)^2 + \frac{1}{\#P_2^r} \left( \min \left\{ \frac{\bar{m}_2}{\bar{m}_1}, 1 \right\} \right)^2}}, \quad (12)$$

among it,

$$\begin{aligned} \bar{m}_1 &= \frac{1}{\#P_1^r} \sum_{P_1^r} I_{P_1^r}(x, y) \\ \bar{m}_2 &= \frac{1}{\#P_2^r} \sum_{P_2^r} I_{P_2^r}(x, y) \end{aligned} \quad (13)$$

where  $L$  is the apparent number of discrete single-channel images,  $\bar{m}_1, \bar{m}_2$  represent the sample means of regions  $P_1^r, P_2^r$ , respectively, and  $\#P_1^r, \#P_2^r$  represent the number of pixels in regions  $P_1^r, P_2^r$ , respectively. The deformation cost of the skeleton and boundary is measured using the Hausdorff distance [22]. The Hausdorff distance is a measure describing the degree of similarity between two sets of points. Considering that the deformation cost of the skeleton and boundary is tiny in many cases, and that the Hausdorff method divides the skeleton and the boundary into small segments, the degree of similarity is fine enough. At the same time, the processing of this study is only pixel operation, and no large



external noise, so the Hausdorff distance was chosen. For the two connected discrete single-channel pixel regions  $P_1^r, P_2^r$  to be merged, the vertex sets of the skeleton and boundary information before and after the region combination are respectively represented as:  $\Phi = \{l_1, l_2, \dots, l_n\}$ ,  $\Psi = \{k_1, k_2, \dots, k_n\}$ , the distance is defined as:

$$\Gamma_S(P_1^r, P_2^r) = H(\Phi, \Psi) = \max(h(\Phi, \Psi), h(\Psi, \Phi)), \quad (14)$$

among them,

$$\begin{aligned} h(\Phi, \Psi) &= \max(l \in \Phi) \min(k \in \Psi) |l - k| \\ h(\Psi, \Phi) &= \max(k \in \Psi) \min(l \in \Phi) |k - l| \end{aligned} \quad (15)$$

where  $|\cdot|$  is the Euclidean distance between the set of points  $\Phi$  and  $\Psi$ .

In order to avoid error accumulation in the process of merging discrete pixel regions guided by continuous edges, this study uses the TOPSIS method [15], which approximates the ideal solution. The TOPSIS method orders ideal merge strategies, providing additional point-cloud regularization constraint results. This method provides the optimal merge strategy while reducing the impact of cumulative error. At the same time, the decision process is used to determine the verification of the point-cloud regularization constraint, which reduces the amount of calculation required and improves system efficiency.

### 2.3. Decision-Based Sparse RGB-Depth Regularization Verification

Decisions based on TOPSIS [15] are used in this study. When it is difficult to find the optimal splicing strategy, or the minimum merge cost of the splicing strategy is always greater than the preset merge cost error threshold, decisions are made using the sparse point-cloud regularization constraint. Because of the large amount of depth image point-cloud information data in the system, the global regularization constraint inevitably leads to high computational cost, and redundant data limit the effectiveness of the system. Considering the need to centrally sample the high curvature region while simplifying the point cloud, in order to reduce the computational complexity and improve the accuracy of the detail mask correction, this study refers to the method and application scenario of [17]. Cloud information is refined and pre-processed with the help of incremental and hierarchical clustering, iterative simplification concentrates the point-cloud sampling in the high-curvature region. At the same time, because the adopted skeleton-extraction process causes the point cloud to be partitioned by the local structure, the scope of the search is further reduced, and the calculation expenditure is reduced. A surface deformation model may then be established [18]. When the edge-guided discrete pixel regions are merged, there is an area deformation error related to the discrete pixel region, and the optimal stitching strategy is screened using the size of the edge stitching error.

The TOPSIS [15] process is summarized as follows.

Suppose that the set of strategy to be determined is  $D = \{d_1, d_2, \dots, d_n\}$ . If the  $m$  attributes of the measure  $d_i$  are  $a_{i1}, a_{i2}, \dots, a_{im}$ , then the decision matrix can be set to  $A = (a_{ij})_{n \times m}$ , and thus the normalized matrix  $B = (b_{ij})_{n \times m}$  can be obtained, among it,

$$b_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^n a_{ij}^2}}, i = 1, 2, \dots, n; j = 1, 2, \dots, m. \quad (16)$$

The weighted canonical matrix is  $C = (c_{ij})_{n \times m}$ . The manually set weight vector is  $W = [w_1, w_2, \dots, w_m]$ , then,

$$c_{ij} = w_j \cdot b_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, m. \quad (17)$$

Determine positive ideal solution  $C^*$  and negative ideal solution  $C_0$ ,

$$c_j^* = \begin{cases} \max c_{ij} \\ \min c_{ij} \end{cases}, c_j^0 = \begin{cases} \max c_{ij} \\ \min c_{ij} \end{cases}. \quad (18)$$

Calculate the distance from each solution to the positive and negative ideal solutions, as follows.

$$s_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, i = 1, 2, \dots, m$$

$$s_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, i = 1, 2, \dots, m. \quad (19)$$

Calculate the comprehensive evaluation value and select the strategy with the highest value of  $f_i^*$  as the optimal merge strategy.

$$f_i^* = \frac{s_i^0}{s_i^0 + s_i^*}, i = 1, 2, \dots, m. \quad (20)$$

In this study, when it is easy to screen out the optimal strategy, four evaluation indicators are selected, namely the minimum combined cost error and the overall minimum cost error of the three channels, i.e.,  $m = 4$ . When it is difficult to select the optimal strategy, five evaluation indicators are selected, namely the minimum combined cost error, the overall minimum cost error of the three channels, and the matching ratio of the combined result of each combining strategy under the minimum regularization constraint, i.e.,  $m = 5$ . The value of the weight vector is derived empirically, when  $m = 4$ ,  $W = [0.4, 0.2, 0.2, 0.2]$ . Moreover, when  $m = 5$ ,  $W = [0.2, 0.2, 0.2, 0.2, 0.2]$ .

#### 2.4. Hole Filling

An image that is merged using discrete pixel regions is affected by insufficient surface texture and noise. Image regions are prone to voids after mask correction. Hole filling presents great advantages in terms of post-image restoration accuracy and avoidance of accumulated errors [23]. The employed void-filling method is mainly based on structural similarity and the image hole filling of morphological features. The image is first subjected to regional morphological expansion processing, then connectivity analysis is conducted. Subsequently, the regularization constraints are combined to recover the cavity information. In addition, skeleton and boundary information increases the overall structural constraints for hole filling, improving the accuracy of hole filling.

### 3. Results and Discussion

In order to verify the efficacy of the proposed algorithm, RGB-depth (RGB-D) data (Technical University of Munich, Germany) were used to manually mark data set sequences, the RGB-depth (RGB-D) data (Technical University of Munich, Germany) is a video sequence containing depth information and RGB information for background motion recorded in an office environment using a handheld kinect. We manually labeled the moving target and background pixels of each frame in the image sequence. The computational equipment used in this study were as follows. CPU: Xeon (Xeon) E5 2.40 GHz six core, GPU: GTX1060 6G, memory: 1866 MHz 32 GB. The algorithms were programmed in C++ and then called on the MATLAB 2017b platform under the Windows 10 operating system.

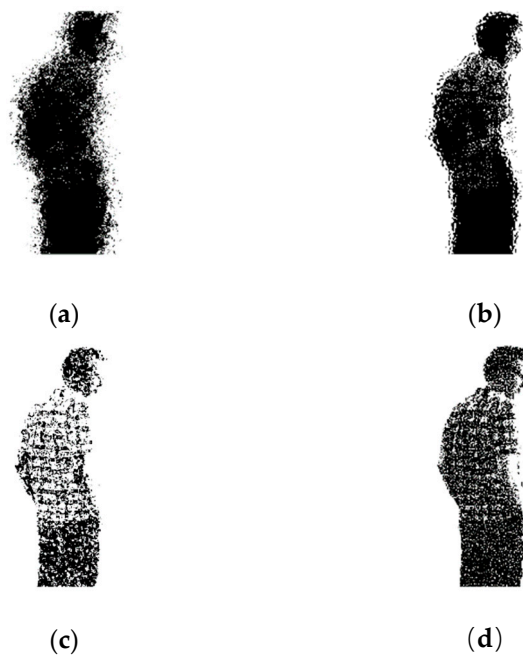
In the traditional motion compensation algorithm, due to the influence of texture loss and computation error accumulation in the target pixel region, the target region is excessively discrete and distorted. In order to achieve the correction of the target pixel region after the traditional motion compensation, the source images in this paper are discrete pixel regions generated after the same motion compensation [24] for moving targets, refer to [24], the square length of the selected sub-block is 7. Suppose that the internal pixel distribution of the moving target  $O_1$  at time  $t$  is  $S_1(t)$ , and the moving target  $O_1$  is restored by motion compensation for the internal pixel distribution  $S_1(t + 1)$  of



the moving target  $O_1$  at time  $t + 1$ . The pixel distribution  $S'_1(t)$ ,  $S''_1(t)$ , at the moment is the source image, and  $S'_1(t)$  is mask corrected to obtain  $S''_1(t)$ , and  $S_1(t)$  and  $S''_1(t)$  are compared to evaluate the effect of mask correction, in Section 3.1, Section 3.2, and Section 3.4, in order to give the reader an intuitive demonstration of the effect of mask correction on shape contours and demonstrate the positive effects of the structure of each part of the paper, this paper used a self-contrast approach. In Sections 3.3 and 3.5, comparing  $S_1(t)$  and  $S''_1(t)$ , the mask correction effect is demonstrated by a large number of quantitative experiments. The number of image pairs used in this study is 1357, which is approximately 1.57 G, and the image resolution is  $640 \times 480$ , PNG format. The size of the images shown in the experimental results in this article is 380 dpi  $\times$  382 dpi.

### 3.1. Single Target Mask Correction Effect

As shown in Figure 3, Comparing a and b, the contour of the human body of b is clearer, because the shape and structure of the binary mask based on the edge similarity of the skeleton region are well corrected. Comparing c and b, the plaid pattern on the man's shirt and the outline of the human body are clearer in c, and many edge pixel masks in b are corrected, which fully reflects TOPSIS and points. The cloud regularization constraint plays a positive role in the elimination of error accumulation in b. Comparing c and d, it can be seen that the outline, clothing, and facial details of the man in d are more complete, and the hole filling technique improves the accuracy of the binary mask correction.

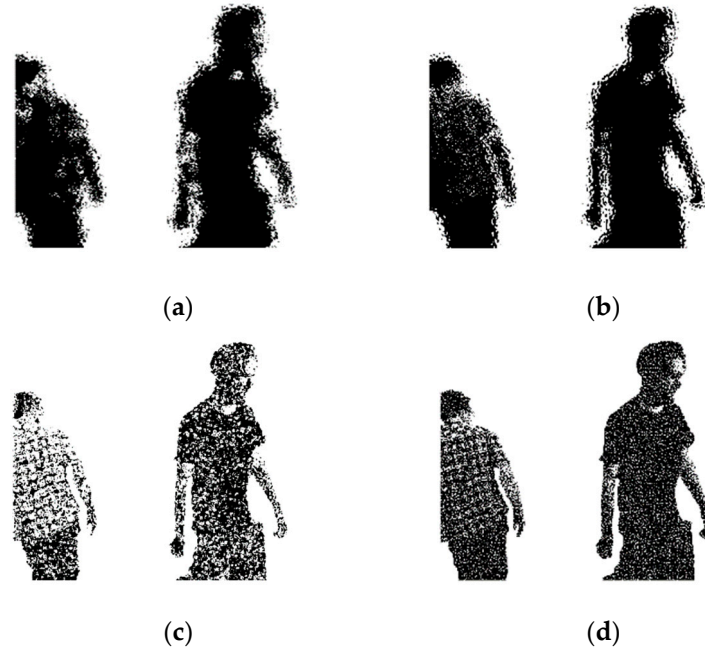


**Figure 3.** Single-objective experimental rendering of the proposed algorithm. (a) Binary mask effect of the source image, (b) mask correction effect using only skeleton-based edge guidance, (c) the result of adding TOPSIS and point cloud regularization constraints to the (b) system, and (d) binary masking effect of the algorithm.

### 3.2. Multi-Target Mask Correction Effect

As shown in the multi-objective mask correction described in Figure 4, comparing Figure 3a with Figure 4a, the number of targets in Figure 4a is increased. Similar to the results in Figure 3, with the improvement of the system structure, the algorithm gradually shows better correction of the binary mask, and it is not affected by the increase of the target. Obviously, the increase of the target means that the difficulty of skeleton extraction increases. The effect of the mixture of discrete pixel binary masks on each target on the correction accuracy of the mask will also be enhanced. However, the skeleton extraction in this paper is based on the edge contour information to extract the skeleton. When the

edge contour information is extracted, the two targets will be taken as two. The single connected regions extract the skeleton separately, avoiding the mutual influence between the targets. At the same time, using the edge similarity guidance of the skeleton divergence, the adjacent sub-regions are merged according to the edge similarity, and the skeleton-centered divergence and the adjacent sub-regions are merged. Merging makes the edge features of the pixel regions to be merged more concentrated, which greatly reduces the influence of other target pixel binary masks that are mixed in.



**Figure 4.** Experimental effect of the proposed algorithm in the multi-objective case. (a) Binary mask effect on source image, (b) mask correction effect using only skeleton-based edge guidance, (c) the result of adding TOPSIS and point cloud regularization constraints to the (b) system, and (d) binary value of the algorithm mask effect.

### 3.3. Multiple Indicator Comparison of Sequence Experimental Results for the Proposed Algorithm and Other Algorithms

The proposed algorithm is compared with four other algorithms: GOP [25], CUR [26], IFMCaS [8], and AIF [27]. These are all relatively advanced algorithms in the field of mask optimization. GOP utilizes a point-to-point iterative nearest-point algorithm framework; CUR uses the circumferential shape feature of the local surface to obtain accurate matching points from the initial matching points; IFMCaS is a fusion score splicing and feature matching method to solve the problem of image stitching quality; AIF utilizes the spatial angle invariant feature to stitch the image. The adopted evaluation criteria include precision (Pr), recall (Re), specificity (Sp), false positive rate (FPR), false negative rate (FNR), percentage of wrong classification (PWC), and F-measure (F1), which are defined as follows [28].

$$\begin{aligned} \text{Pr} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \text{Re} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{Sp} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}}, \text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \\ \text{PWC} &= \frac{100 \times (\text{FN} + \text{FP})}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}, \text{F1} = \frac{2 \times \text{Pr} \times \text{Re}}{\text{Pr} + \text{Re}} \end{aligned} \quad (21)$$

For these criteria, the true positive (TP) is the number of pixels correctly detected as foreground, the false positive (FP) is the number of pixels erroneously detected as foreground, the true negative (TN) is the number of pixels correctly detected as the background, and the false negative (FN) is the number of pixels erroneously detected as background.

As shown in Equation (21), the larger the values of Pr (indicating that a method is not sensitive to noise), Re, Sp, and F1 (indicating that a lower amount of correct information is missing in the

target), the better the experimental result. The smaller the values of FPR, FNR, and PWC (the effects of background subtraction and illumination), the better the experimental result.

Table 1 shows the results of experiments using all the data sets used in this study. As shown in Table 1, The proposed algorithm outperforms the other algorithms for all the experimental criteria. The larger values of Pr, Re, Sp, and F1 indicate that the algorithm is better at identifying the target pixel, mainly because this paper is based on the skeleton divergence algorithm. The overall strategy of mask correction is established by using skeleton partitioning, which leads to the edge of the area to be merged. The features are more concentrated, and the target pixels are easier to identify. The algorithm is compared with other algorithms. The CUR and AIF algorithms are all for global point correction. The edge feature information is scattered and susceptible to image noise. GOP utilizes the point-surface correlation feature, and IFMCaS uses the feature matching integrity score to improve the partial mask correction accuracy. However, the error accumulation cannot be avoided in the partial detail mask correction, and the target pixel is not easily recognized. The PWC value is better than other algorithms, which shows that the proposed algorithm is robust to the interference of holes and noise in the binary mask. Because this paper uses the skeleton partition to establish the overall strategy of mask correction, the pixel features to be merged are more concentrated, and the influence of interference is reduced. At the same time, because this paper uses the hole filling technology, the cavity filling technology utilizes the morphology. Feature and structural similarity features reduce the effects of image noise.

**Table 1.** Comparison of image processing results for various algorithms.

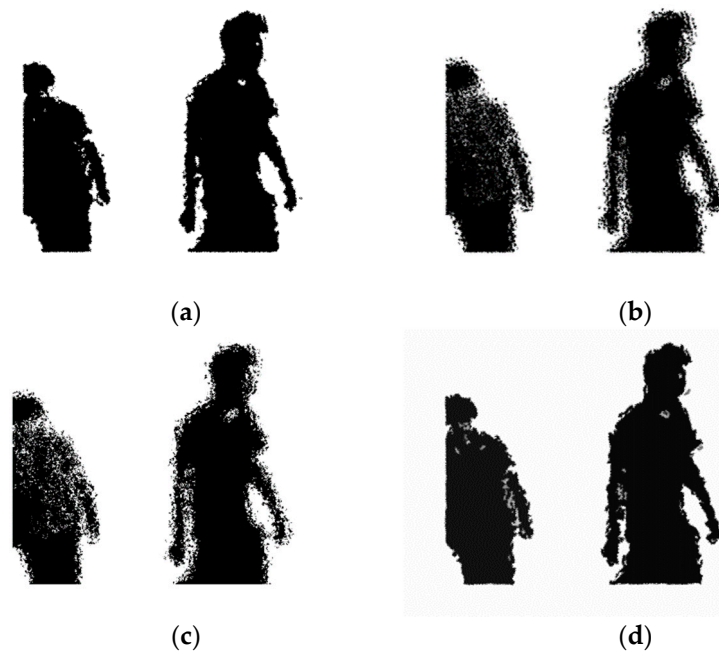
Method	Pr	Re	Sp	FPR	FNR	PWC	F1
Prop	0.703	0.896	0.987	0.013	0.104	0.611	0.755
GOP [25]	0.692	0.795	0.983	0.017	0.205	1.128	0.567
CUR [26]	0.688	0.676	0.980	0.020	0.324	0.882	0.631
IFMCaS [8]	0.575	0.777	0.972	0.028	0.223	1.217	0.526
AIF [27]	0.674	0.538	0.979	0.021	0.462	0.783	0.421

### 3.4. Single Frame Comparison of Experimental Results for the Proposed Algorithm and Other Algorithms

As shown in Figure 5, the mask correction results for the four comparison algorithms are compared with the experimental results in Figure 4d. Compared with CUR and AIF, GOP and IFMCaS maintain a slightly better contour information of the human body, but they are far behind the algorithm in the correction of binary mask details, for example, the mask correction effect on the arm contour and the clothing grid pattern. None of the algorithms in this paper are clear; CUR and AIF lag far behind the proposed algorithm in the correction of edge contours, and the adhesion in the arm part is increased and not easy to separate, because the correction algorithm of this paper is based on skeleton divergence, using skeleton partitioning. The overall strategy of mask correction is established, and the CUR and AIF algorithms are all for the correction of global points. It is easy for the points in different regions to accumulate errors, which make it difficult to accurately classify; GOP utilizes point-surface correlation characteristics, and IFMCaS utilization characteristics. Although the matching integrity score improves the partial mask correction accuracy, it cannot reduce the error accumulation in the detail correction, resulting in poor mask correction effect.

At the same time, Table 2 shows the time cost for outputting these experimental results of the five algorithms. It should be noted that the time cost of the algorithm proposed in this article is the time it takes to output the result of Figure 4d, and the time cost of the others is the time it takes to output the experimental results of Figure 5. These five time cost values are all obtained under the same preconditions. The time cost is the time it takes for the computer to run the entire mask correction process, starting with a source image input and ending with the corresponding mask correction image output. It is timed and displayed by the running program inside the computer.

As shown in Table 2, the calculation cost of the proposed algorithm is lower than that of similar algorithms, mainly because the algorithm uses point cloud regularization constraints, but the merging strategy of this study adopts the strategy of RGB information as the auxiliary depth information. Moreover, in the early stage of the point cloud regularization constraint, the point cloud is refined and pre-processed, and the iterative simplification is carried out by means of increment and hierarchical structure clustering, so that the point cloud sampling is concentrated in the high curvature region, and the mask correction accuracy is ensured. At the same time, the system efficiency is improved.



**Figure 5.** Mask correction of the 82nd frame in f3/w\_h. (a) GOP, (b) CUR, (c) IFMCaS, and (d) AIF.

**Table 2.** Time cost in different mask correction algorithms.

Mask Correction Algorithms	Computing Time Cost (s)
GOP	0.532
CUR	0.618
IFMCaS	0.471
AIF	0.672
Prop	0.301

### 3.5. Effect of the Algorithm on Different Sequences

As shown in Table 3, the dynamics of the scenes in sequences f3/w\_h and f3/w\_r are slightly higher than those of sequences f2/d\_p and f3/s\_hz. These four sequences are all video sequences in the TUM RGBD data set. The scenes are all office scenes, in which there are two people in the f3/w\_h and f3/w\_r sequences; and one person in the f2/d\_p and f3/s\_h sequences. The complexity of the motion of dynamic participants in a scene affects motion compensation, which renders mask correction more difficult. However, the results provided in Table 3, which represent a variety of scene complexities, indicate that the proposed algorithm is robust to dynamic scenes. The information fusion decision method significantly reduces the influence of uncertain factors on pixel-attribute judgment decisions; thus, the robustness of mask correction is improved. The results for f2/d\_p are slightly better than those of f3/s\_h mainly because the people in the f2/d\_p scene wore clothes that are easier to distinguish from the rest of the scene.

**Table 3.** Average signal-to-noise ratio of moving target pixels.

Sequence	GOP	CUR	IFMCaS	AIF	Prop
f2/d_p	28.31	27.32	28.34	27.44	29.63
f3/s_h	27.97	26.81	27.82	28.32	28.72
f3/w_h	27.41	28.57	28.12	27.01	28.71
f3/w_r	26.53	26.40	27.62	26.94	28.64

#### 4. Conclusions

This study was conducted to model and process discrete pixels in complex images. We proposed a binary mask correction model based on skeleton divergence of the source and mask images. Discrete pixel region edges were extracted with bilateral filtering, edge smoothing was performed, and the skeleton of the global discrete pixel region was extracted. The skeleton was then used as the center of local mask correction to optimize edge gradients and achieve region merging. The TOPSIS method [15] was used to determine the ideal merge strategy, and additional point-cloud regularization constraint verification was adopted. The mask-correction effect of the proposed algorithm in both single-objective and multi-target cases was tested experimentally and compared with four other high-quality processing algorithms. The effectiveness of the proposed algorithm was verified using various discrete source images of ranging complexity, with our results indicating that it outperformed all the other algorithms. One shortcoming of this approach is that a fixed-value method was adopted in the setting of the scatter length optimization step of the skeleton model. This resulted in a lack of flexibility and should be the subject of further research. Nonetheless, this highly robust algorithm presents a promising option for use in image-processing applications.

**Author Contributions:** Conceptualization, Z.X. and W.H.; Methodology, Z.X. and Y.W.; Software, Z.X.; Validation, Y.H.; Formal analysis, Z.X. and W.H.; Investigation, Z.X. and Y.W.; Writing—original draft preparation, Z.X.; Writing—review and editing, Z.X. and W.H.; Funding acquisition, Y.W., Y.H. and M.J.

**Funding:** This research was supported by the Natural Science Foundation of Zhejiang Province (LZ20F020003, LY17F020034, LSZ19F010001), the National Natural Science Foundation of China (61272311, 61672466), and the 521 Project of Zhejiang Sci-Tech University.

**Acknowledgments:** Thanks for the test data set provided by researchers at the Technical University of Munich, Germany.

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

#### References

1. Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73. [\[CrossRef\]](#)
2. Levin, A.; Zomet, A.; Peleg, S.; Weiss, Y. Seamless image stitching in the gradient domain. In Proceedings of the Computer Vision—Proceedings of ECCV 2004, Prague, Czech Republic, 11–14 May 2004; pp. 377–389.
3. Zhang, F.; Liu, F. Parallax-tolerant image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3262–3269.
4. Zomet, A.; Levin, A.; Peleg, S.; Weiss, Y. Seamless image stitching by minimizing false edges. *IEEE Trans. Image Process.* **2006**, *15*, 969–977. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Mills, A.; Dudek, G. Image stitching with dynamic elements. *Image Vis. Comput.* **2009**, *27*, 1593–1602. [\[CrossRef\]](#)
6. Zheng, J.; Wang, Y.; Wang, H.; Li, B.; Hu, H.M. A Novel Projective-Consistent Plane based Image Stitching Method. *IEEE Trans. Multimed.* **2019**, *21*, 2561–2575. [\[CrossRef\]](#)
7. Fang, F.; Wang, T.; Fang, Y.; Zhang, G. Color Blending for Seamless Image Stitching. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1115–1119. [\[CrossRef\]](#)
8. Hu, F.; Li, Y.; Feng, Y. Continuous Point Cloud Stitch based on Image Feature Matching Constraint and Score. *IEEE Trans. Intell. Veh.* **2019**, *4*, 363–374. [\[CrossRef\]](#)



9. Chen, M.; Tang, Y.; Zou, X.; Huang, K.; Li, L.; He, Y. High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm. *Opt. Lasers Eng.* **2019**, *122*, 170–183. [[CrossRef](#)]
10. Jovanovski, B.L.; Li, J. Image-Stitching for Dimensioning. U.S. Patent Application 16/140,953, 24 January 2019.
11. Hu, F.; Bai, L.; Li, Y.; Chen, H.; Li, C.; Su, X. Environmental Reconstruction for Autonomous Vehicle Based on Image Feature Matching Constraint and Score. In Proceedings of the PRICAI 2018: Trends in Artificial Intelligence, Nanjing, China, 28–31 August 2018; pp. 140–148.
12. Ding, C.; Liu, H.; Li, H. Stitching of depth and color images from multiple RGB-D sensors for extended field of view. *Int. J. Adv. Robot. Syst.* **2019**, *16*. [[CrossRef](#)]
13. Wang, P.; Liu, J.; Zha, F.; Liu, H.; Sun, L.; Li, M. A fast 3D map building method for indoor robots based on point-line features extraction. In Proceedings of the 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Hefei and Tai'an, China, 27–31 August 2017; pp. 576–581.
14. Koller, D.O. Image filtering on 3D objects using 2D manifolds. U.S. Patent 6,756,990, 29 June 2004.
15. Nanayakkara, C.; Yeoh, W.; Lee, A.; Moayedikia, A. Deciding discipline, course and university through TOPSIS. *Stud. High. Educ.* **2019**, 1–16. [[CrossRef](#)]
16. Feng, L.; Alliez, P.; Busé, L.; Delingette, H.; Desbrun, M. Curved optimal delaunay triangulation. *ACM Trans. Graph. (Tog)* **2018**, *37*, 61. [[CrossRef](#)]
17. Robert, L.F.; Gros, N.; Noutary, Y.; Malleus, L.; Fisichella, T.; Lingrand, D.; Precioso, F.; Samoun, L. Keypoint-Based Point-Pair-Feature for Scalable Automatic Global Registration of Large RGB-D Scans. U.S. Patent 10,217,277, 26 February 2019.
18. Yu, T.; Zheng, Z.; Guo, K.; Zhao, J.; Dai, Q.; Li, H.; Pons-Moll, G.; Liu, Y. Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7287–7296.
19. Murthy, P.A.; Defenu, N.; Bayha, L.; Holten, M.; Preiss, P.M.; Enss, T.; Jochim, S. Quantum scale anomaly and spatial coherence in a 2D Fermi superfluid. *Science* **2019**, *365*, 268–272. [[CrossRef](#)] [[PubMed](#)]
20. Zhang, D.M. Research on Shape Optimization Algorithm of Discrete Curved Surfaces. Ph.D. Thesis, Zhejiang University, Hangzhou, China, 2010. (In Chinese).
21. Shui, P.L.; Zhang, Z.J. Fast SAR image segmentation via merging cost with relative common boundary length penalty. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6434–6448. [[CrossRef](#)]
22. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 850–963. [[CrossRef](#)]
23. Jung, S.; Song, S.; Chang, M.; Park, S. Range image registration based on 2D synthetic images. *Comput. Aided Des.* **2018**, *94*, 16–27. [[CrossRef](#)]
24. Zhang, J.; Su, Q.; Liu, P.; Zhu, Q.; Zhang, K. An improved VSLAM algorithm based on adaptive feature map. *Acta Autom. Sin.* **2019**, *45*, 553–565. (In Chinese) [[CrossRef](#)]
25. Mitra, N.J.; Gelfand, N.; Pottmann, H.; Guibas, L. Registration of point cloud data from a geometric optimization perspective. In Proceedings of the Eurographics/ACM SIGGRAPH symposium on Geometry Processing, Nice, France, 8–10 July 2004; pp. 22–31.
26. He, B.; Lin, Z.; Li, Y.F. An automatic registration algorithm for the scattered point clouds based on the curvature feature. *Opt. Laser Technol.* **2013**, *46*, 53–60. [[CrossRef](#)]
27. Jiang, J.; Cheng, J.; Chen, X. Registration for 3-D point cloud using angular-invariant feature. *Neurocomputing* **2009**, *72*, 3839–3844. [[CrossRef](#)]
28. Goyette, N.; Jodoin, P.M.; Porikli, F.; Konrad, J.; Ishwar, P. A novel video dataset for change detection benchmarking. *IEEE Trans. Image Process.* **2014**, *23*, 4663–4679. [[CrossRef](#)] [[PubMed](#)]

