

Article

# Automatic Multiview Alignment of RGB-D Range Maps of Upper Limb Anatomy

Luca Di Angelo <sup>1</sup>, Paolo Di Stefano <sup>1</sup>, Emanuele Guardiani <sup>1,\*</sup>, Paolo Neri <sup>2</sup>, Alessandro Paoli <sup>2</sup>  
and Armando Viviano Razionale <sup>2</sup>

<sup>1</sup> Department of Industrial and Information Engineering and Economics, University of L'Aquila, 67100 L'Aquila, Italy; luca.diangelo@univaq.it (L.D.A.); paolo.distefano@univaq.it (P.D.S.)

<sup>2</sup> Department of Civil and Industrial Engineering, University of Pisa, 56122 Pisa, Italy; paolo.neri@unipi.it (P.N.); alessandro.paoli@unipi.it (A.P.); armando.viviano.razionale@unipi.it (A.V.R.)

\* Correspondence: emanuele.guardiani@univaq.it

**Abstract:** Digital representations of anatomical parts are crucial for various biomedical applications. This paper presents an automatic alignment procedure for creating accurate 3D models of upper limb anatomy using a low-cost handheld 3D scanner. The goal is to overcome the challenges associated with forearm 3D scanning, such as needing multiple views, stability requirements, and optical undercuts. While bulky and expensive multi-camera systems have been used in previous research, this study explores the feasibility of using multiple consumer RGB-D sensors for scanning human anatomies. The proposed scanner comprises three Intel<sup>®</sup> RealSense<sup>™</sup> D415 depth cameras assembled on a lightweight circular jig, enabling simultaneous acquisition from three viewpoints. To achieve automatic alignment, the paper introduces a procedure that extracts common key points between acquisitions deriving from different scanner poses. Relevant hand key points are detected using a neural network, which works on the RGB images captured by the depth cameras. A set of forearm key points is meanwhile identified by processing the acquired data through a specifically developed algorithm that seeks the forearm's skeleton line. The alignment process involves automatic, rough 3D alignment and fine registration using an iterative-closest-point (ICP) algorithm expressly developed for this application. The proposed method was tested on forearm scans and compared the results obtained by a manual coarse alignment followed by an ICP algorithm for fine registration using commercial software. Deviations below 5 mm, with a mean value of 1.5 mm, were found. The obtained results are critically discussed and compared with the available implementations of published methods. The results demonstrate significant improvements to the state of the art and the potential of the proposed approach to accelerate the acquisition process and automatically register point clouds from different scanner poses without the intervention of skilled operators. This study contributes to developing effective upper limb rehabilitation frameworks and personalized biomedical applications by addressing these critical challenges.

**Keywords:** depth cameras; 3D optical scanning; upper limb anatomy; automatic point cloud alignment; neural network



**Citation:** Di Angelo, L.; Di Stefano, P.; Guardiani, E.; Neri, P.; Paoli, A.; Razionale, A.V. Automatic Multiview Alignment of RGB-D Range Maps of Upper Limb Anatomy. *Sensors* **2023**, *23*, 7841. <https://doi.org/10.3390/s23187841>

Academic Editor: Andrea Facchinetti

Received: 28 July 2023

Revised: 4 September 2023

Accepted: 10 September 2023

Published: 12 September 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

Digital representations of anatomical structures play a crucial role in various applications within the biomedical field. For example, an accurate 3D model depicting the anatomy of the upper limb can be utilized to design customized devices for personalized rehabilitation or human-machine interfaces (HMIs) [1]. Enhancing upper limb functionality holds immense importance for patients afflicted with sports injuries, strokes, or dystonia disorders, as the upper limb is extensively involved in daily activities. Novel technologies like virtual, augmented, and mixed reality offer promising solutions to overcome conventional treatment limitations [2]. Frameworks for upper limb rehabilitation can be defined

by integrating virtual reality gaming and tailored controllers, enabling the recreation of therapeutic movements. Patients can enter a virtual gaming environment tailored to their rehabilitation needs and engage with the game and virtual environment using a specifically configured controller [3].

The initial and crucial step in generating a digital representation of an anatomical part involves 3D scanning [4]. In this regard, optical techniques prove to be particularly suitable as they cause minimal invasiveness to the patient [5]. However, 3D scanning of the forearm is a complex process due to the requirements of multiple scans from various angles and ensuring the arm remains stable to prevent motion artifacts caused by involuntary movements. Furthermore, the presence of optical undercuts complicates the acquisition of finger geometry. Numerous approaches are available for capturing human body parts. Photogrammetry is the underlying principle for 3D body scanners, which can simultaneously capture all sides of the human body. These systems employ multiple cameras arranged in a cylindrical architecture to capture the target anatomy from different viewpoints [6–8]. Nonetheless, these systems can be cumbersome and expensive, depending on the cameras used.

A practical alternative is given by handheld 3D scanners that can be manually moved around the patient [9,10]. This approach requires longer acquisition times due to the scanner's need to cover the entire target anatomy [4]. A drawback of this method is the possibility of unintentional patient movement during the scanning process, resulting in artifacts in the scanning outcome. Furthermore, high-level handheld scanners, which offer substantial reductions in acquisition time compared with low-level handheld scanners, tend to be expensive [4,11]. For this reason, in the last few years, many studies have been conducted to assess the feasibility of using consumer RGB-Depth (RGB-D) sensors (e.g., Kinect by Microsoft, Structure by Occipital, RealSense by Intel) to scan human anatomies [12,13]. RGB-D cameras are compact, portable, and low-cost [4]. In this regard, some architectures have been developed to scan human arms, considering the simultaneous use of multiple RGB-D sensors rigidly assembled in circular jigs [14,15]. These solutions have the advantage of offering affordable architectures that can ensure the necessary dimensional accuracy.

However, rapidly obtaining the arm's complete anatomy requires multiple sensors to capture acquisitions from various viewpoints, effectively resolving optical undercut issues. Consequently, this leads to the development of stationary, bulky, and non-portable systems [14]. In contrast, lightweight configurations can be achieved by integrating a limited number of sensors, although this approach necessitates multiple scanner placements around the patient's arm to capture the complete anatomy [15]. This technique is more flexible and adaptable to different patients' needs but requires accurate registration of the acquired multiple views into a common reference frame, which is conventionally composed of two alignment steps [16,17]: a local coarse alignment and a global fine registration. The first step is based on manually identifying common landmarks on different scans. While widely used, the traditional manual procedure has several limitations that negatively impact performance. The manual selection of corresponding points is time-consuming and subjective, relying on the user's visual inspection and interpretation of the acquired data. This can lead to different results depending on the operator, making the procedure unreliable and non-repeatable. Furthermore, the absence of markers on the patient's arm complicates the process: the operator must rely solely on visual inspection to identify corresponding points, leading to errors and inconsistencies. The global fine registration, instead, is based on applying iterative closest point (ICP) techniques on overlapping regions of adjacent scans. Aligning 3D human body scans captured from different viewpoints presents challenges, especially when there is low overlap and a lack of relevant topology, as is often the case with the human forearm, which typically exhibits a quasi-cylindrical shape.

To overcome these limitations, this paper proposes an automatic alignment procedure for human forearm scans from a low-cost handheld scanner. The scanner consists of three Intel® RealSense™ D415 depth cameras rigidly assembled on a lightweight circular jig, enabling the simultaneous acquisition of point clouds from three different viewpoints.

Extrinsic calibration is performed to determine the positioning and orientation of the three depth sensors, aligning the three point clouds into a common reference frame for each scanner pose.

The primary contribution of this research lies in providing an automated procedure to align scans obtained from different scanner poses, where extrinsic calibration cannot be used due to the operator moving the scanner around the target anatomy. Regarding registration methods, feature-based approaches are commonly employed to identify meaningful and robust geometrical descriptors, which serve as correspondences to estimate transformations between distinct point clouds [18]. These features can be artificially attached to the patient's skin as markers or manually identified in the acquired point clouds during post-processing. However, both approaches are ineffective in clinical scenarios. Attaching markers to the patient's skin is generally not feasible, and relying on manual intervention during post-processing requires skilled operators, which undermines the goal of automation.

Consequently, it is necessary to automatically extract local and/or global shape descriptors by analyzing the geometric information in the point clouds [19,20]. It is important to note that the accuracy of extracting these shape descriptors improves with high-density point clouds due to their reliance on discrete differential geometry. Nevertheless, the resolution of the depth cameras used in the 3D scanner must provide more information for the abovementioned methods to be utilized. As a result, an automatic procedure has been developed in the present work to identify common key points between scans obtained from different scanner placements. RGB images from the depth cameras are processed using a neural network to detect a dataset of relevant key points on the hand, which are then projected onto the point cloud to identify corresponding 3D points. A smaller dataset of relevant key points on the forearm is also included in the analysis. This additional dataset is generated by directly processing the point cloud with a specially developed algorithm that searches for matches along the forearm's skeleton line. These two datasets, obtained from different scanner positions, are then used to achieve an automatic rough 3D alignment of the point clouds in a common reference frame. Finally, an iterative-closest-point (ICP) algorithm, supported by a specially designed function to be minimized, refines the alignment through fine registration.

The method was tested on the alignment of upper limb anatomies with multiple views for five subjects acquired with a low-cost handheld 3D scanner. The results are critically discussed by using as a reference a gold-standard manual alignment pipeline with commercial software and compared with the available implementations of published methods. The study aims to address two significant issues that arise when obtaining a 360° acquisition of an upper limb: speeding up the acquisition process and automatically registering point clouds derived from different scanner poses.

## 2. Related Works

Registering partially overlapped point clouds presents challenges due to limited and blurred information caused by occlusion, missing points, outliers, and measurement noise. According to [21], the existing methods in the related literature can be categorized into two main groups:

- Optimization-based methods;
- Learning-based methods.

### 2.1. Optimization-Based Registration Methods

These methods, which are typically applied to a pair of point clouds, generally involve two main steps:

- Estimating correspondences between the two point clouds to achieve a coarse-registration solution;
- Estimating the rigid body transformation by solving a least-squares problem to obtain the final solution.

A coarse estimation can be performed by extracting local 3D descriptors. These descriptors typically calculate statistics related to the principal component analysis (PCA) or other discrete geometric properties, such as surface normal, curvatures, second-order moment invariants, etc. A comprehensive review of 3D point cloud descriptors is provided by Han et al. [22]. However, these descriptors often fail to provide a complete and unambiguous representation of local shape geometry. Moreover, uncertainties arise due to factors like point cloud quality, resolution, surface noise, and outliers, further impacting the accuracy of these geometric properties. These challenges become even more pronounced when dealing with partially overlapping point cloud registration. To address these issues, the published methods for registering partially overlapping point clouds often rely on randomized alignment or overlap ratio estimation [23]. Two widely used randomized alignment-based techniques are four-point congruent sets (4PCS) [24] and Super4PCS [25]. These methods exploit the invariance of the ratio between lines formed by four coplanar points, eliminating the need for complex geometric calculations. The robustness of these algorithms increases with larger overlap areas compared with the average density of the point cloud, making them more noise-resilient.

The iterative closest point (ICP) [26] algorithm is commonly used for fine registration; however, it is prone to local optima, is unable to differentiate between inlier and outlier correspondences, and performs global registration that considers non-overlapping areas. Therefore, specific methods have been developed to enhance registration performance for partial overlap problems. The trimmed ICP method, for example, introduced overlap-ratio estimation [27,28] to exclude outliers and increase the algorithm's robustness when analyzing point clouds with a relatively low overlap ratio (below 50%). Chetverikov et al. [27] proposed the trimmed ICP (TrICP) algorithm, which exploits the least-trimmed-squares approach to minimize a certain number of smaller values and sort the square errors. This method estimates the optimal transformation at each iteration step, proving high robustness for incomplete and noisy data with overlap ranges from 0.4 to 1. To reduce false correspondences and enhance robustness against outliers, Du et al. [29] introduced a new objective function that considers the overlapping percentage between the two point sets. Dong et al. [30] proposed the LieTrICP algorithm, which, after determining the correspondences, estimates the transformation by minimizing the trimmed squared distances of the point pairs using the Lie group parametrization. Incorporating the Lie group parametrization allows the algorithm to handle data with anisotropic scale transformation and low overlap rates [31]. Additionally, Wang et al. [28] proposed the parallel-trimming iterative-closest-point (PTrICP) method for fine point cloud registration. PTrICP incorporates parallel overlap rate estimation during the iterative registration process to improve the algorithm's robustness. However, it is worth noting that all of these registration algorithms require a good initial estimation for satisfactory results [32].

## 2.2. Learning-Based Registration Methods

Learning-based registration methods use learning of the latent representations of the point clouds for their registration.

The first published method applying deep learning to point cloud registration is PointNetLK [33]. It uses a learnable global representation for point clouds based on PointNet [34]. This method's main contribution is modifying the Lucas and Kanade (LK) algorithm [35] for optimization. Specifically, it avoids the need for convolutions on the PointNet representation. Sarode et al. [36] proposed a similar approach by replacing the LK algorithm with a multi-layer perceptron to improve noise robustness. One limitation of these methods is that they do not consider the adverse effects caused by non-overlapping regions. To overcome this limitation, Zhou et al. [37] proposed SCANet, which consists of two main modules: a spatial self-attention aggregation (SSA) module and a channel cross-attention regression (CCR) module. The SSA module efficiently extracts different-level features from point clouds to capture local and global information. The CCR module was innovatively used for pose estimation by enhancing relevant information and suppressing redundant

information. The results of SCANet show a significant improvement in registration accuracy compared with state-of-the-art methods while reducing the number of weights. Xu et al. [38] proposed OMNet, another global-feature-based registration network robust to noise and different partial overlaps. OMNet uses learning masks to reject non-overlapping regions and allows iterative mask prediction and transformation estimation modification. However, estimating the masks without interaction with feature information is challenging, as pointed out in [39]. To address this challenge, Zhang et al. [39] proposed a two-stage partially overlapping point cloud registration method, which can deal with point clouds having inhomogeneous density. It uses a network with overlapping-region-prediction and pose-estimation modules. The first module extracts features from the two point clouds using the DGCNN algorithm [40] to predict possible overlapping regions. The second module combines PointNet with a self-attention mechanism to capture global information and achieve robust transformation regression. When tested on different datasets, this method proved to have limited accuracy compared with other methods, leading the authors to conclude that its results should be used as the initial pose for ICP.

In recent years, several research papers have introduced graph-based learning methods to process point clouds [40–47]. Unlike point-based methods, these approaches do not directly handle individual points as input. Instead, they construct local structures using information obtained from 3D shapers, treating them as graphs. Shen et al. proposed KCNet [42], which extracts local structural features using kernel correlation layers. Wang et al. [40] introduced EdgeConv, a novel operation for learning from point clouds, aiming to capture local geometric features while maintaining permutation invariance. Their reported results demonstrated that the model could learn to semantically group points by dynamically updating a relationship graph across layers. Zhang et al. [41] introduced LDGCNN, which draws inspiration from PointNet [34] and DGCNN [40], to construct a directed graph for point cloud data. The constructed graph extracts features, which are then utilized for point cloud classification and segmentation. The results indicate that LDGCNN reduces the network's model size by eliminating the transformation network and optimizing the network architecture by connecting hierarchical features from various dynamic graphs. However, the main drawback of the graph-based methods mentioned above is the standard definition of convolution, which overlooks the differences between neighboring points. To address this issue, Wang et al. introduced a new graph attention convolution method called GAC in their study [44]. GAC utilizes learnable kernel shapes that can adapt dynamically to the object's structure for point cloud semantic segmentation. Similarly, in another work, Wang et al. [32] presented STORM, a structure-based overlapping approach that employs differentiable sampling. This method predicts overlaps and generates partial correspondences for partial point cloud registration. The authors used densely connected EdgeConv layers [40] along with two transformers, demonstrating favorable outcomes, particularly in cases where the overlap ratio was low.

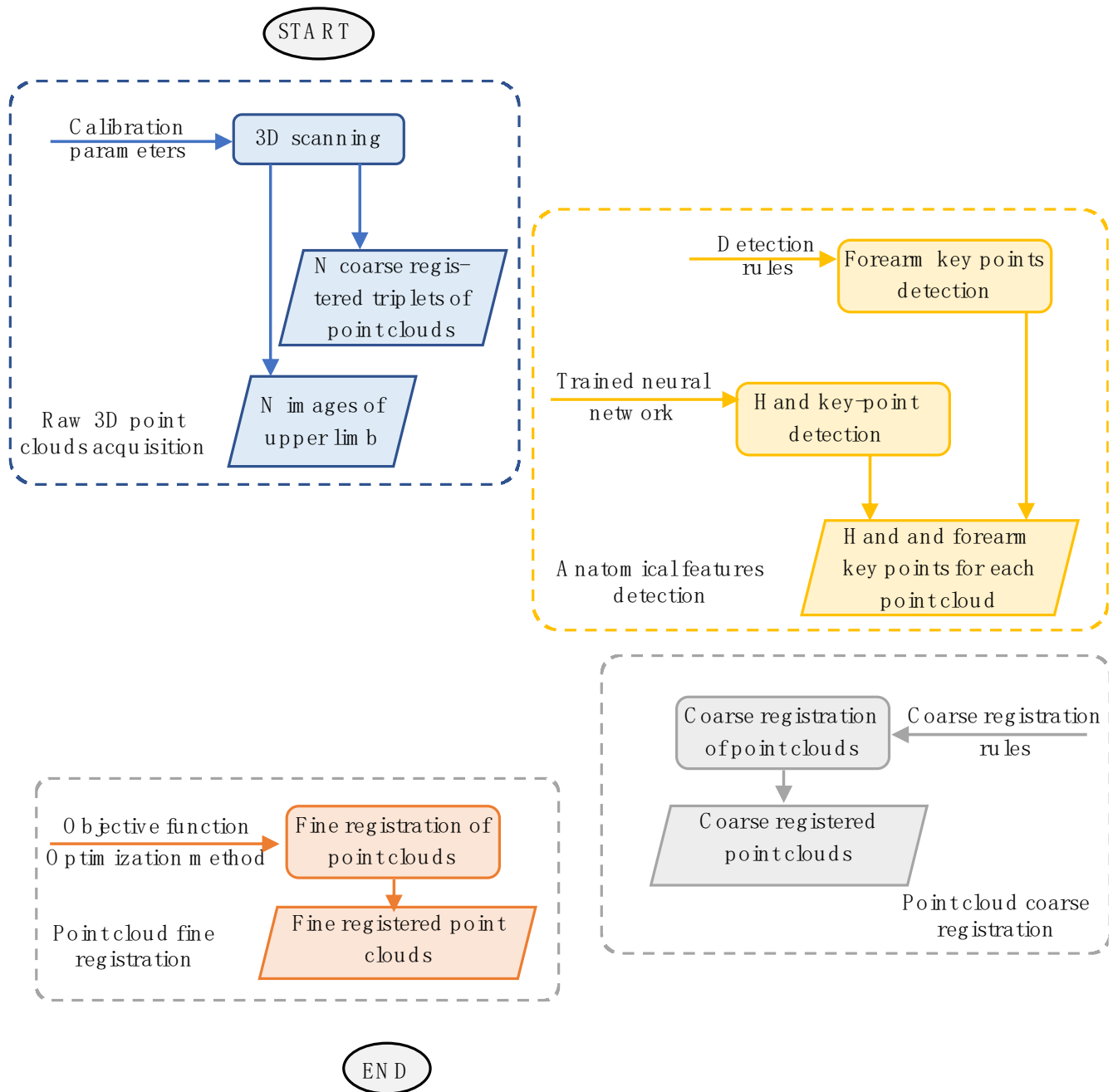
### 3. The Proposed Methodology

State-of-the-art analysis indicates that the morphology of the upper limb with automorphisms, combined with the small overlapping areas and high level of noise obtainable from different scans, makes applying existing optimization-based registration methods difficult. The same reasons, to which must be added the need for many training samples considering individual variations, make using learning-based methods unfeasible. To address these challenges, this research paper introduces a novel optimization-based method that focuses on registering anatomical features. The proposed approach, outlined in Figure 1, involves the following key steps:

- Acquisition of raw 3D point clouds;
- Detection of anatomical features:
  - Hand key-point detection;
  - Forearm key-point detection;
- Coarse registration of the point clouds;



- Fine registration of the point clouds.



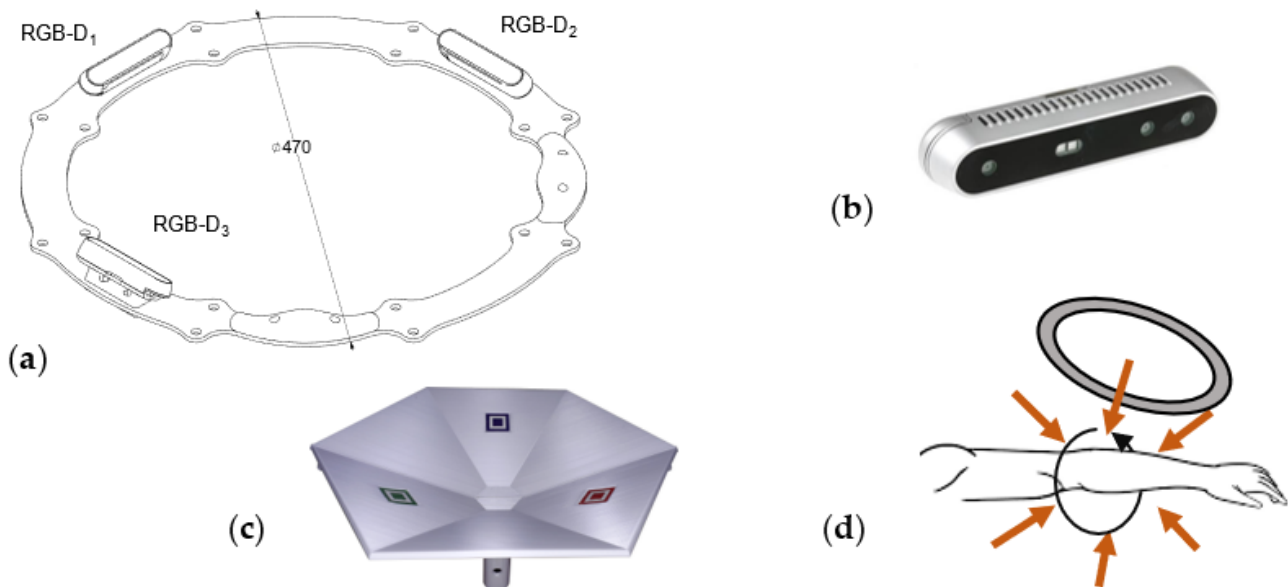
**Figure 1.** Flow-chart of the proposed method.

The framework developed in the present work integrates a neural network to identify relevant key points on the patient's hand with an automatic recognition process for forearm key points. Finally, a global registration by an ICP algorithm based on a specifically designed function to be minimized is carried out to refine the alignment results.

### 3.1. Acquisition of Raw 3D Point Clouds

The 3D scanner consists of three Intel® RealSense™ D415 depth cameras arranged at 90° angles on a circular frame (Figure 2a). The D415 sensor (Figure 2b) is a compact device (99 mm × 20 mm × 23 mm) equipped with an RGB camera (resolution up to 1920 × 1080 pixels, frame rate 30 fps), an infrared projector, and two IR cameras (resolution up to 1280 × 720 pixels, frame rate 90 fps). Unlike structured-light approaches that

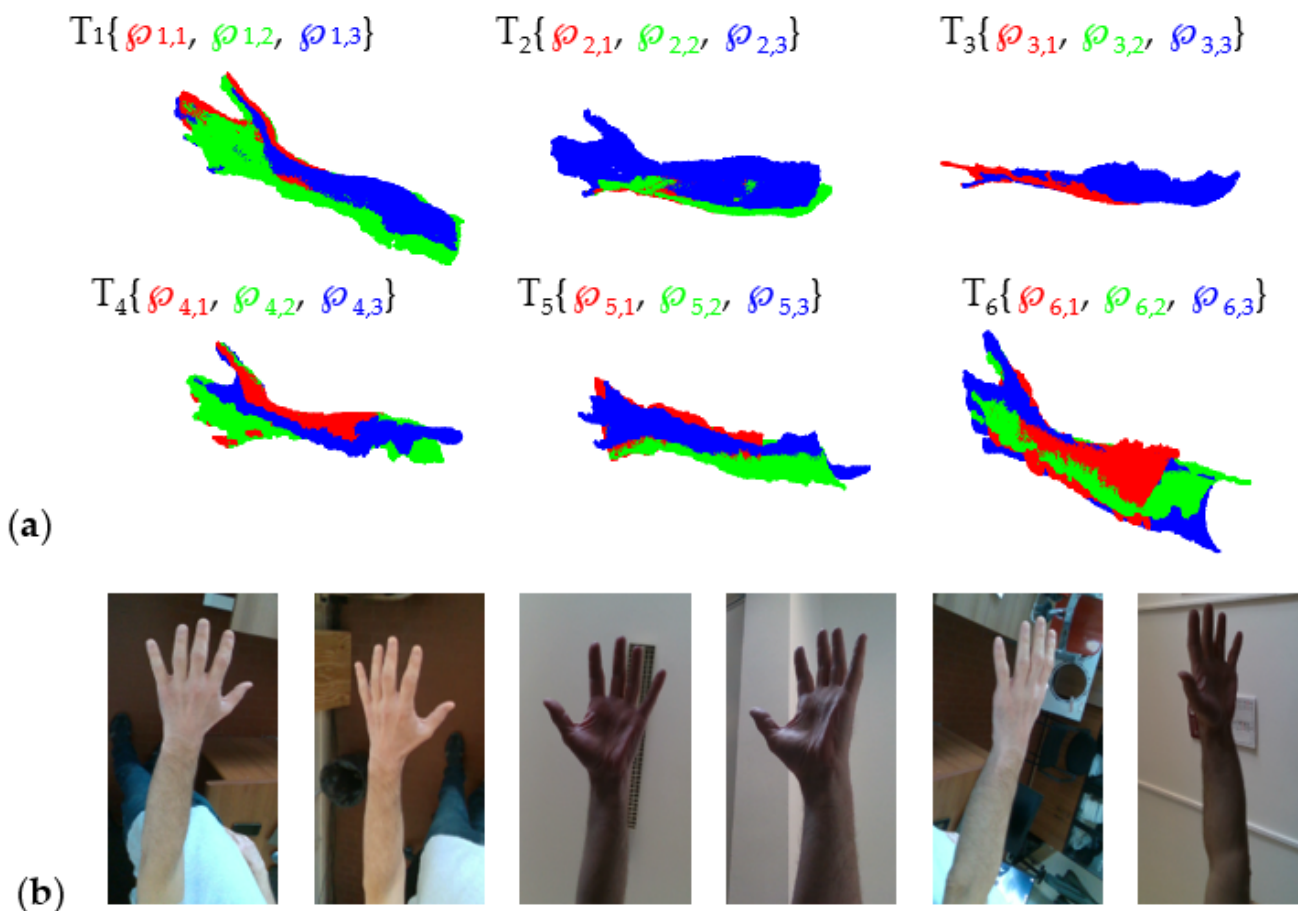
use sequential projection of fringe patterns (such as the SR300 Intel® RealSense™ depth camera), the infrared projector of the D415 sensor projects a single static pattern, thus avoiding interference issues among multiple sensors. The D415 camera has wide acquisition depth sensing, ranging from about 160 mm to 10,000 mm. However, as demonstrated in [48], the maximum accuracy is obtained for close distances (from 150 mm to 500 mm), with corresponding 3D reconstruction errors below 1 mm. In the present work, three depth cameras were arranged to acquire the target shape in the range 250 mm–450 mm in room ambient lighting conditions. The scanner generates three separate point clouds, automatically registered into a common reference frame through an extrinsic calibration process after the assembly stage. The extrinsic calibration, which solves the placement parameters between the depth cameras, uses a 3D calibration specimen and a coarse-to-fine scheme [49]. The calibration specimen consists of six differently oriented planar surfaces and three colored markers affixed on the external surfaces (Figure 2c). The scanner captures the specimen, thus providing three point clouds, one for each sensor. The marker's centers on the captured RGB images are detected and reprojected on the corresponding point clouds using the sensor's factory intrinsic calibration data for each sensor acquisition. The markers' 3D coordinates are then used to obtain a coarse alignment of the point clouds, which is then refined by minimizing the distance between the point clouds through an iterative-closest-point (ICP) algorithm. The resulting roto-translation matrices, which transform each point cloud from the sensor reference frame to the common reference frame, are saved in a calibration file. This file aligns the three point clouds corresponding to a specific scanner pose during the 3D scanning of the patient's arm. For more comprehensive information regarding the scanner's architecture and the extrinsic calibration process, please refer to [50]. The scanner captures several consecutive frames of the limb to obtain a 360° scan of the patient's arm. A user, such as a therapist, can hold the scanner with one or two hands and move it around the patient's arm in a circular path (see Figure 2d).



**Figure 2.** 3D scanner architecture: (a) arrangement of the three depth cameras on the circular frame; (b) picture of a single sensor; (c) calibration specimen; (d) use of the scanner for the arm's acquisition.

The results of the acquisition process are (Figure 3):

- $N$  triplets  $\{T_1 (\varphi_{1,1}, \varphi_{1,2}, \varphi_{1,3}), T_2 (\varphi_{2,1}, \varphi_{2,2}, \varphi_{2,3}), \dots, T_N (\varphi_{N,1}, \varphi_{N,2}, \varphi_{N,3})\}$  of coarsely registered point clouds (Figure 3a);
- $3 \times N$  images of the upper limb anatomy (Figure 3b).



**Figure 3.** The results of the raw-data acquisition process: (a)  $N$  triplets of coarse registered point clouds; (b) some of the  $3 \times N$  images of upper limb anatomy from different orientations.

The adopted depth-camera-based scanner makes the realignment of point clouds more challenging due to the low density of the point clouds and the high level of noise that characterizes the raw data.

### 3.2. Anatomical Features Detection

One of the most innovative aspects of the proposed method is achieving the initial alignment of  $N$ -point clouds through anatomical feature recognition. This operation is highly complex because it involves extracting high-level information from low-level information, such as the coordinates of the points, which are affected by noise and sparse distribution. To address this challenge, the proposed framework combines a neural network for identifying significant key points in 2D images of the patient's hand with an automatic recognition process for key points on the forearm in the 3D point cloud.

#### 3.2.1. Hand Key-Point Detection by a Neural Network

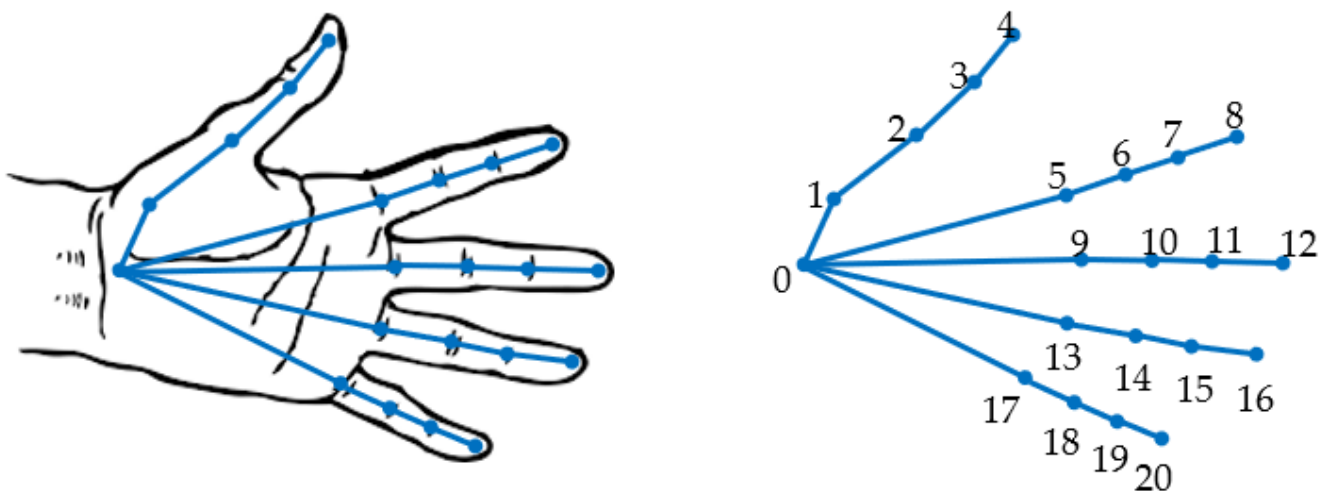
Detecting hand key points using a neural network (NN) is a computer vision task that involves identifying specific landmarks on fingers, palms, and wrists. The use of NNs in studying hand motion has gained popularity due to their capacity to handle vast amounts of data and learn from it [51]. In the present study, the effectiveness of an NN in identifying landmarks on the patient's hand is significantly enhanced by providing three RGB images for each scanner pose, thereby increasing the likelihood of recognizing relevant landmarks in at least one of the images. Additionally, various pre-trained detection models of hand key points are available as open-source software libraries, such as the OpenPose model [52] and the MediaPipe Hand Landmarker model [53]. These models can be fine-tuned using new data to enhance accuracy for specific tasks.



The MediaPipe Hand Landmarker model (v. 0.8.8.1), developed by the Google Research team, is used in this work. It is a component of the MediaPipe framework, which offers a flexible and scalable solution for real-time computer vision and machine learning tasks like hand tracking, facial recognition, and object detection. The MediaPipe Hand Landmarker model is specifically designed for real-time hand tracking and hand gesture recognition. It combines a deep neural network and computer vision algorithms to accurately detect and track hands, even in challenging environments like low light or cluttered backgrounds.

The neural network used in the MediaPipe Hand Landmarker model consists of multiple stages, including a hand segmentation network, a hand pose estimation network, and a hand landmark localization network. It outputs a set of 2D hand key points representing the locations of fingers, palms, and wrists. The MediaPipe Hand Landmarker model is implemented in Python, and its open-source code simplifies its integration into the overall scanning workflow, coded in Matlab.

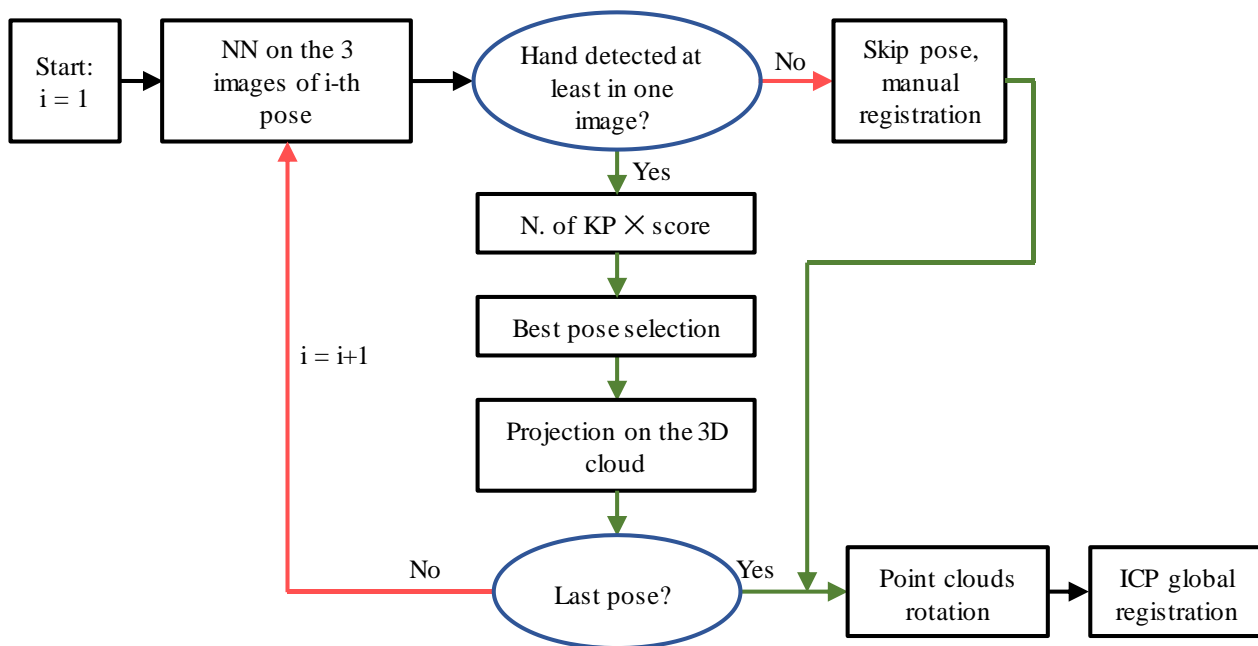
When an RGB image of a hand is provided as input, the neural network returns a list of identified key points along with a recognition score. Under optimal conditions, the algorithm can identify 21 hand features, including four key points on each finger and one key point on the wrist. The output is always sorted in the same sequence of key points, ensuring consistent numbering, as shown in Figure 4. In cases where one or more key points are not visible in a particular image, the network assigns a “not-a-number” value to those undetected key points. As a result, the output always consists of 21 points in the same sequence, eliminating any confusion regarding landmark numbering. This data structure is crucial for comparing the same key points detected from different viewpoints, as it ensures consistent indexing and point correspondence.



**Figure 4.** Hand key-point numbering in the NN.

Each acquisition frame includes data from three separate Intel® RealSense™ D415 sensors. The calibration of the scanner ensures that the three point clouds are aligned. Consequently, the corresponding RGB image for each point cloud can be analyzed to identify the hand key points. In theory, this process would provide three sets of hand key points that could be utilized for registering the clouds in subsequent frames. This redundancy in the data significantly increases the success rate. However, despite the hand’s orientation to the cameras and finger pose, some sensors may fail to detect the hand key points due to occlusion. Using a single sensor may result in a lower number of identified key points or complete failure. Therefore, an automated algorithm must assess each frame and select the sensor with the highest detection rate among the three options. It is important to note that cloud registration requires a minimum of three points; thus, detecting all 21 key points is unnecessary. However, more key points lead to better performance by improving registration constraints. When a hand pose does not allow the detection of at least three

key points in any of the three images, the automatic coarse alignment fails, and an operator must intervene to align the relative scan to the previous ones. A metric was established to determine the quality of key-point detection by multiplying the identification score (provided by the NN algorithm for each RGB image) by the number of detected key points. This criterion allows the selection of the key-point group with the most landmarks and the best identification among the three sensors. This criterion allows selection of the key-point group with the most landmarks and the best identification among the three sensors. Once the best 2D key-point set is chosen, its coordinates can be projected onto the 3D point cloud using the D415 factory calibration. This results in a set of 3D key points that can be stored alongside the point cloud for registration purposes. An overview of the automatic selection method is depicted in Figure 5. It should be noted that if none of the sensors provide reliable key-point identification for a specific pose, manual registration will be necessary.



**Figure 5.** Automatic selection of successful key points.

This process generates a set of 3D key points for each point cloud. However, due to the results obtained from the neural network (NN), the key points associated with each frame may consist of different landmarks. This poses a challenge, because cloud alignment requires two sets of corresponding 3D points. Typically, the NN may fail to identify a few random key points on the hand, which limits the usability of the entire set for rough registration. To overcome this issue, the intersection of the key-point lists from the two point clouds to be registered is considered. A list of corresponding indexes is stored alongside the point cloud. Only the key points with the selected indexes are used to determine the alignment roto-translation during the registration step.

### 3.2.2. Forearm Key-Point Detection

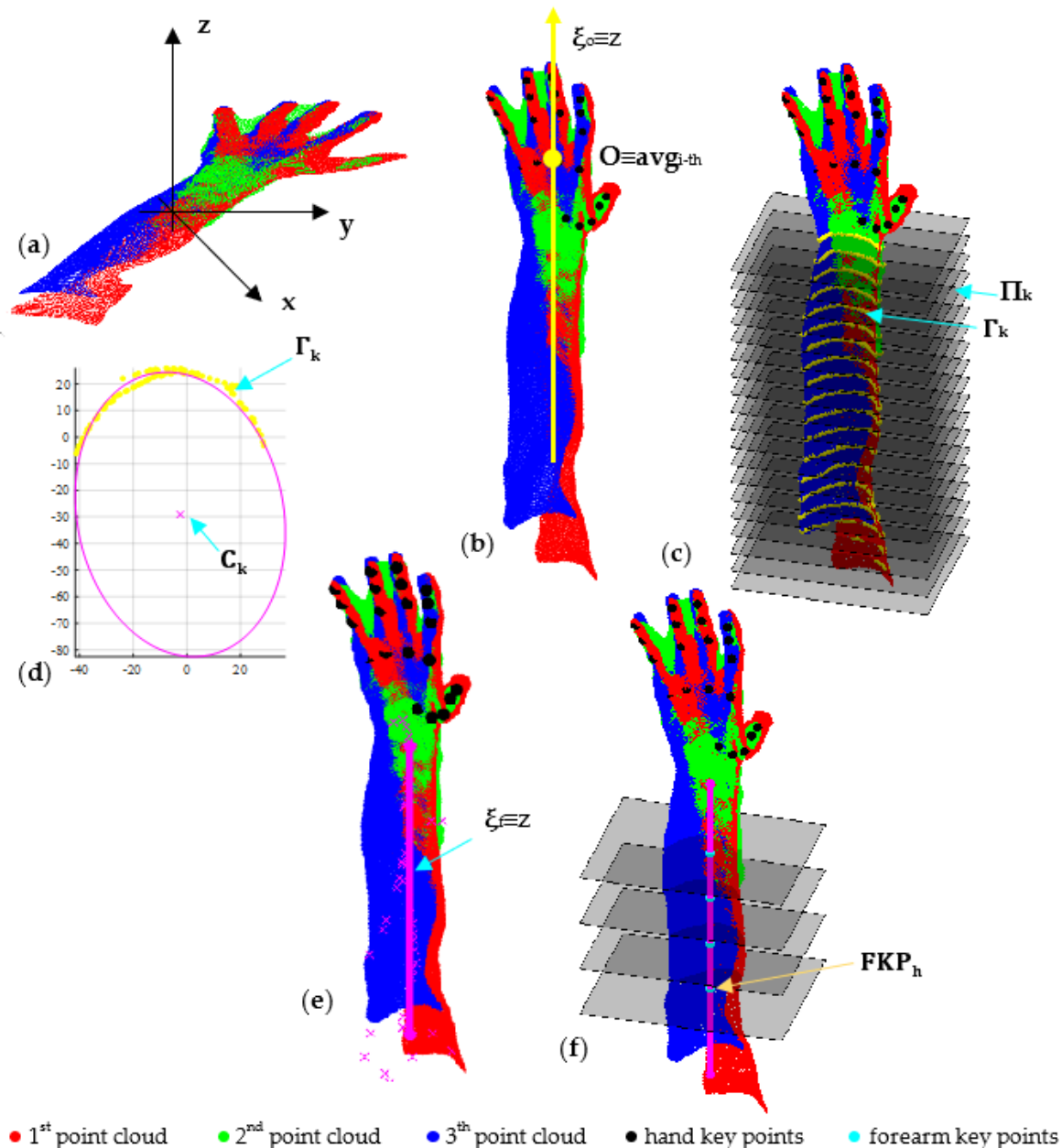
To detect the forearm key points, each triplet captured by the three depth cameras and registered using the method proposed in Section 3.1 undergoes the following steps:

- Initial evaluation of the approximate skeleton line of the forearm;
- Final evaluation of the approximate skeleton line of the forearm;
- Detection of forearm key points.

Let  $T_i\{\varphi_{i,1}, \varphi_{i,2}, \varphi_{i,3}\}$  represent the  $i$ -th triplet corresponding to the  $i$ -th scanner pose (Figure 6a). The direction of the first-attempt skeleton line of the forearm ( $\xi_0$ ) is calculated using the point cloud principal component of inertia: the longitudinal axis  $\xi_0$  is associated with a lower moment of inertia of  $T_i$ . In Figure 6b,  $T_i$  is transformed so that  $\xi_0$  aligns with

the  $z$ -axis of the global reference system and the origin ( $O$ ) with the barycenter ( $\text{avg}_{i\text{-th}}$ ) of the hand key points detected for  $T_i$ . Starting from  $\xi_o$ , the final direction  $\xi_f$  is obtained by following these steps:

- Introducing a set of planes ( $\Pi$ ) perpendicular to the  $z$ -axis with a defined pitch and intersecting  $T_i$  from a fixed height below the barycenter (Figure 6c);
- Clustering the points ( $\Gamma_k$ ) obtained for each plane intersection  $\Pi_k$  (Figure 6c);
- Approximating each  $\Gamma_k$  with an ellipse whose center is  $C_k$  (Figure 6d);
- Approximating the centers  $C_k$  with a line by the RANSAC algorithm, thus obtaining  $\xi_f$  (Figure 6e).

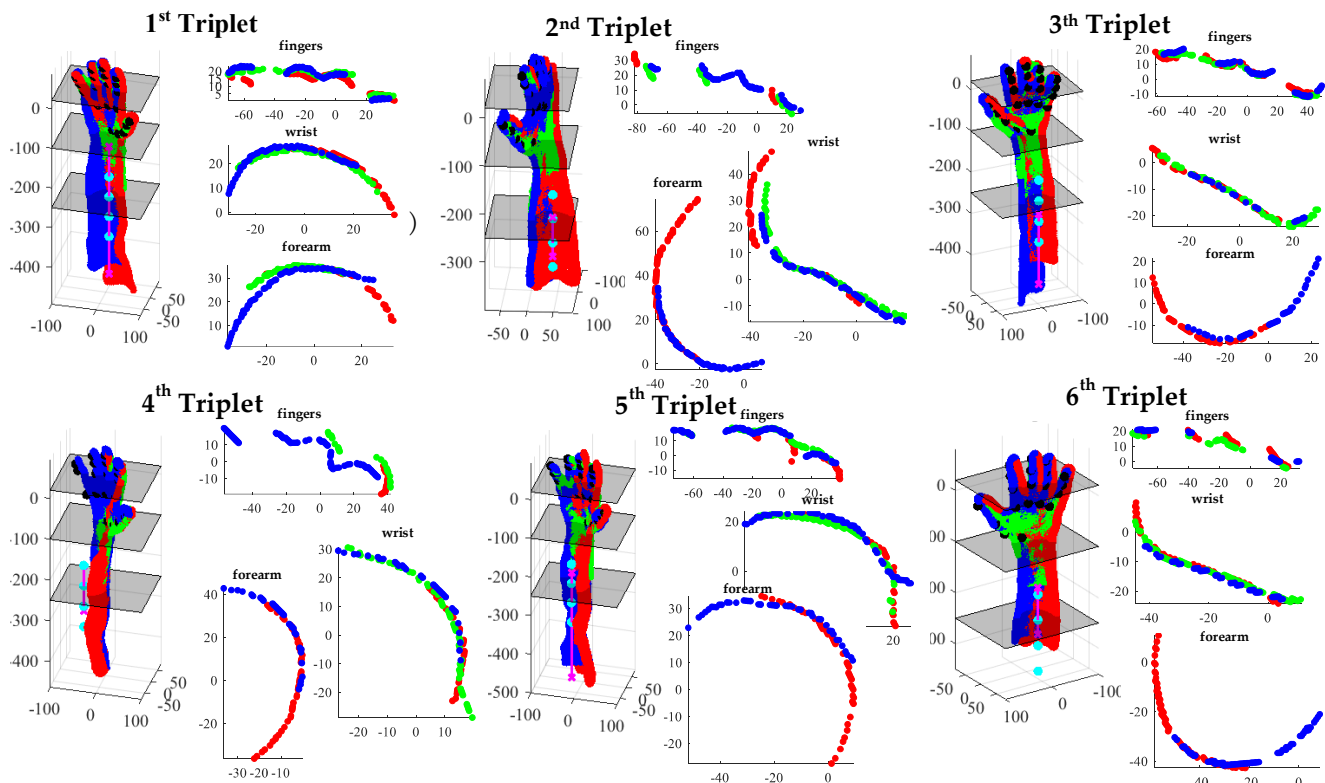


**Figure 6.** Key steps of the automatic process for forearm key-point detection: (a) the  $i$ -th triplet  $T_i$  corresponding to the  $i$ -th scanner pose; (b) alignment of  $T_i$  so that  $\xi_o$  coincides with the  $z$ -axis of the global reference system; (c) intersecting points  $\Gamma_k$  between  $T_i$  and the plane  $\Pi_k$  perpendicular to  $T_i$ ; (d) ellipse approximating  $\Gamma_k$  with a center at  $C_k$ ; (e) line approximating the center  $C_k$ ; (f)  $\text{FKP}_h$  identification.

The forearm key points (FKP<sub>h</sub>) are finally identified as the intersection points between  $\xi_f$  and the set of planes perpendicular to  $\xi_f$ , which start from a fixed height below the barycenter (avg<sub>i-th</sub>) (Figure 6f).

### 3.3. Coarse Registration of Point Cloud

Figure 7 shows the outcome of registration performed after the acquisition stage and recognition of key points for the hand and forearm in the six triplets illustrated in Figure 3. Additionally, the same figure presents cross-sections at various distances to the barycenter of the identified hand markers:

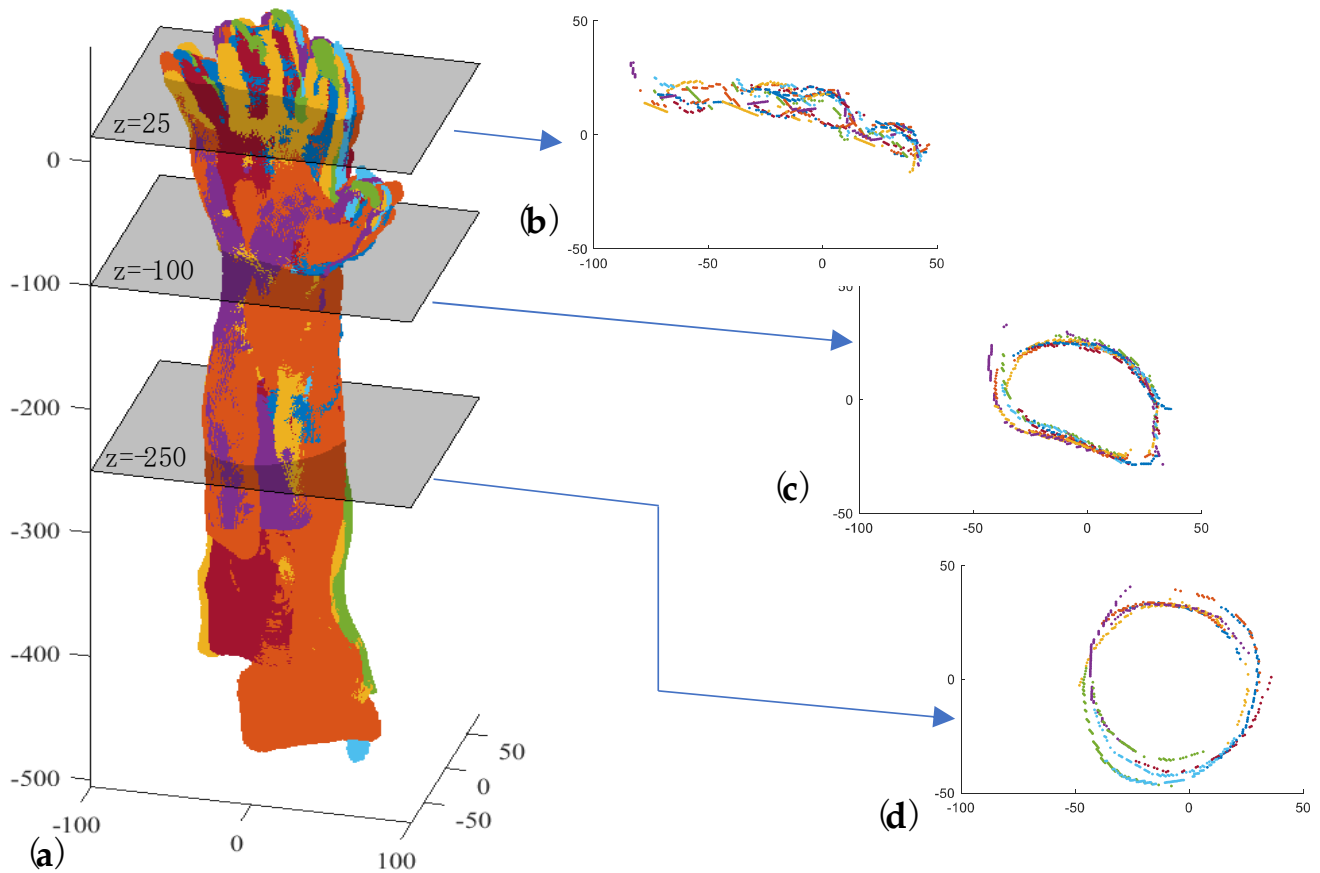


**Figure 7.** The results of the automatic registration performed after the acquisition stage and key–point detection for the six triplets reported in Figure 3.

- $z = 25$  (fingers);
- $z = -100$  (wrist);
- $z = -250$  (forearm).

The registration quality is influenced by extrinsic calibration as described in Section 3.1.

As a first step, the coarse-registration algorithm selects the triplet with the highest number of recognized hand key points as the fixed reference ( $T_f$ ) cloud. Subsequently, all other triplets are registered to this fixed triplet using their corresponding key points on the hand and forearm. To achieve this, the roto-translation matrix is obtained by solving the weighted absolute orientation problem using Horn’s quaternion-based method [54]. Figure 8 shows the result of the coarse registration of the six triplets reported in Figure 3.



**Figure 8.** (a) The results of the coarse registration of the six triplets reported in Figure 3. Each color represents a sub-pointcloud coming from a depth sensor. (b) cross-section at  $z = 25$  mm; (c) cross-section at  $z = 100$  mm; (d) cross-section at  $z = 250$  mm.

### 3.4. Fine Registration of Point Clouds

The final step of the proposed framework involves the simultaneous fine registration of all the  $T_i$  point clouds. For this purpose, an original automated procedure, summarized in the following, has been coded into a Python application.

One of the original elements of this approach is the method implemented for the fine registration of two point clouds ( $\varphi_{fixed}$  and  $\varphi_{moving}$ ), which finds the rigid rotation  $[R^{opt}]$  and translation  $[t^{opt}]$  such that:

$$[R^{opt}, t^{opt}] = \underset{R,t}{\operatorname{argmin}} \left\| \frac{1}{f_1 f_2} \right\| \quad (1)$$

The  $f_1$  term in the objective function evaluates the number of points in the clouds whose distance is lower than a specified threshold value,  $toll$ . It is defined as:

$$f_1 = \frac{|D|}{\min(M, N)} \quad (2)$$

where:

$M$  = number of points of  $\varphi_{fixed}$

$N$  = number of points of  $\varphi_{moving}$

$D$  is a  $M \times N$  matrix, defined as follows:

$$D_{M \times N} = \begin{cases} d_{ij} = \text{distance}(P_i \in \varphi_{fixed}, j \in \varphi_{aligned}) & \text{if distance} < toll \\ d_{ij} = NULL & \text{otherwise} \end{cases} \quad (3)$$



whose  $d_{ij}$  is a no-null element representing the Euclidian distance between the  $i$ -th point of  $\varphi_{\text{fixed}}$  and the  $j$ -th point of  $\varphi_{\text{moving}}$  if it is lower than the *toll* threshold value.

The term  $f_2$  refines the angular mismatch between the two point clouds and is defined as:

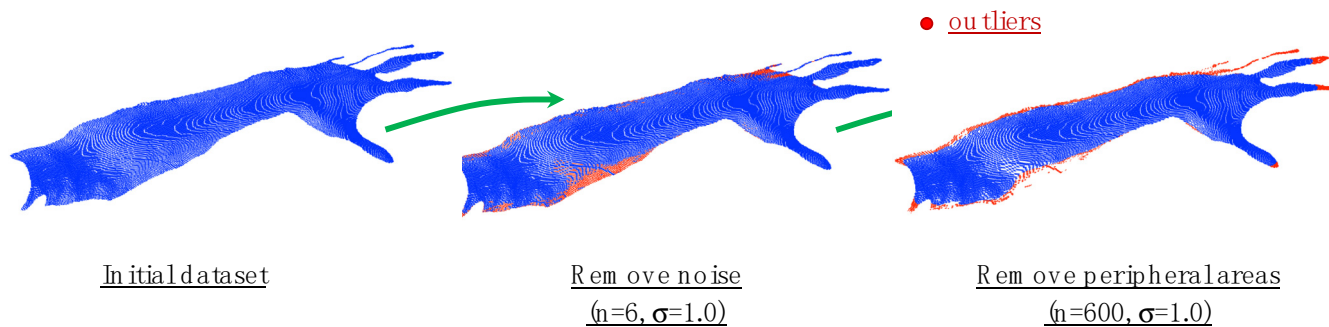
$$f_2 = \frac{\sum_{\substack{0 \leq i \leq M \\ 0 \leq j \leq N}} v_{ij}}{|D|} \tag{4}$$

where  $v_{ij}$  are the elements of the  $V_{M \times N}$  matrix, determined as:

$$V_{M \times N} = \begin{cases} v_{ij} = \hat{n}(P_i) \bullet \hat{n}(Q_j) & \text{if } d_{ij} < \textit{toll} \\ v_{ij} = \textit{NULL} & \textit{otherwise} \end{cases} \tag{5}$$

with  $\hat{n}(P)$  being the exterior normal estimated for the point  $P$ .

The constrained nonlinear optimization problem resulting from the proposed fine registration of two point clouds is solved using a Python implementation of the sequential least-squares programming method [55]. This optimization method was selected for its efficiency in solving constrained nonlinear optimization problems subject to the lower and upper bounds values of angle rotation and distance translation. During the preliminary experimental analysis, it was noticed that there was a significant impact of outliers resulting from the depth camera acquisitions on the results of the fine-registration process. In this specific application, two kinds of outliers can be identified in the scans: a first one of random nature (noise) and a second one of systematic origin, identifiable in the peripheral areas of the hand and the arm. Both of these outliers were removed from the original datasets by applying an SOR filter implemented in Open3D (see Figure 9 for details).



**Figure 9.** The results of the fine registration of the six triplets reported in Figure 3.

The second innovative aspect of the global fine registration involves a strategy to realign all the point clouds  $\{\varphi_1, \varphi_2, \dots, \varphi_{np}\}$  using the optimization method described above for two point clouds. This strategy is summarized in Algorithm 1. Figure 10 shows the results of the fine registration of the six triplets reported in Figure 3, highlighting a significant improvement to the coarse registration shown in Figure 8.

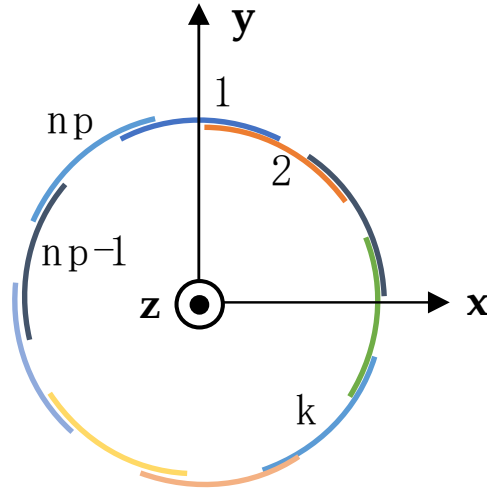
---

**Algorithm 1.** Developed algorithm for the fine global registration of  $np$  point clouds.

---

**Require:** Point clouds  $\{\varphi_1^0, \varphi_2^0, \dots, \varphi_{np}^0\}, N_{iter\_ex}$

Labeling of point clouds from 1 to  $np$  in a clockwise direction:



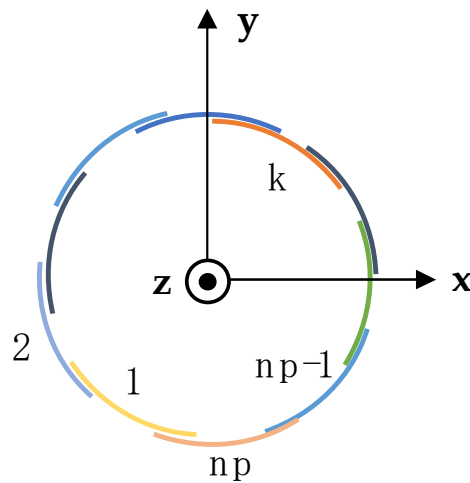
for  $h = 1$  to  $N_{iter\_ex}$  do

    Select randomly the index of the fixed point cloud:

    Index\_fixed\_point\_cloud = random(1, np)

    Reassign the labels to point clouds such that:

- Index\_fixed\_point\_cloud  $\leftarrow$  1
- Labels from 2 to  $np$  in a clockwise direction



for  $k = 2$  to  $np$  do

$[R^k, t^k] = \text{fine\_registration}(\varphi_{\text{fixed}}, \varphi_k^{h-1})$

    Update the  $\varphi_k$  alignment:

$\varphi_k^h = R^k \bullet \varphi_k^{h-1} + t^k$

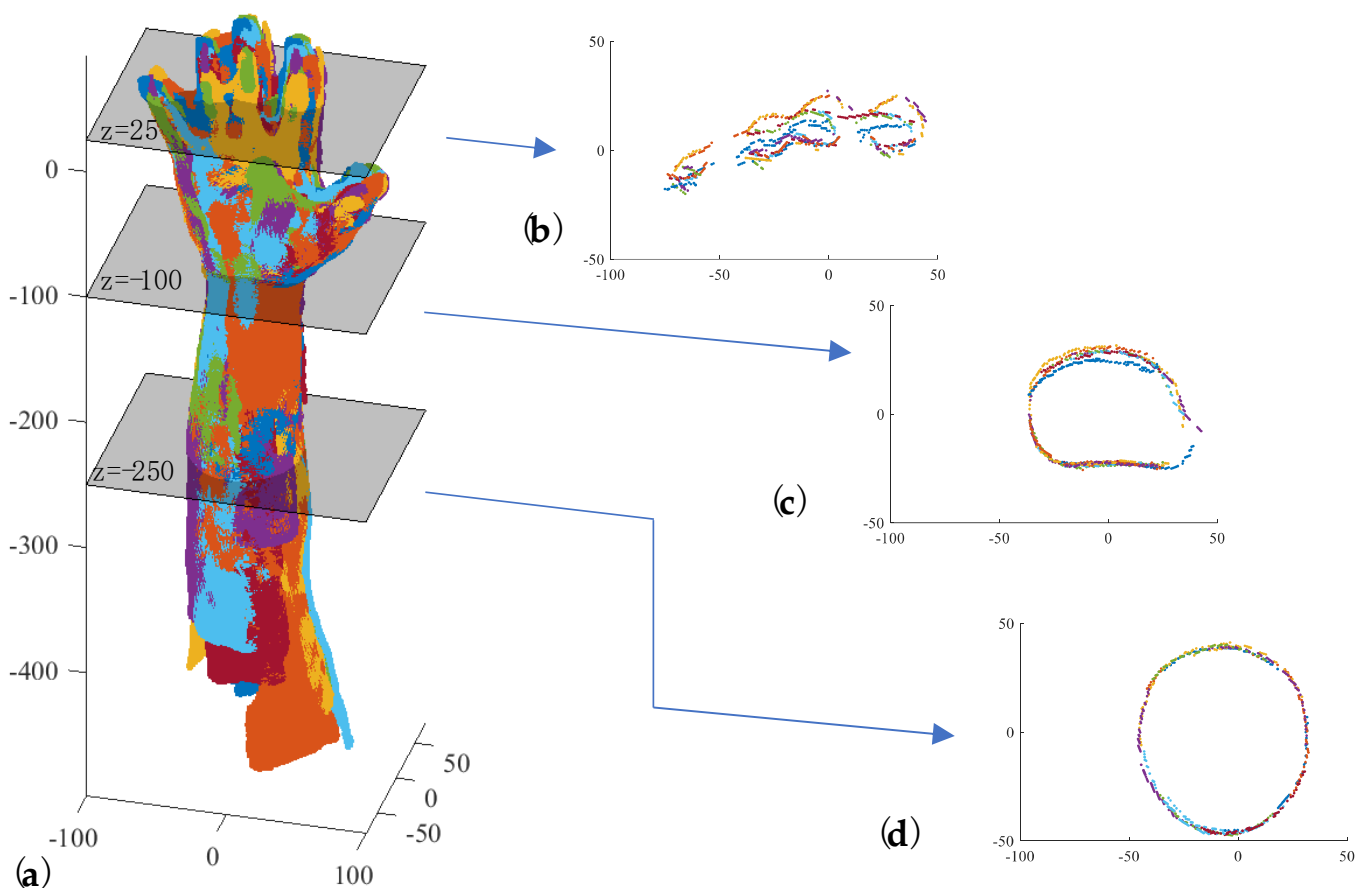
    Update the  $\varphi_{\text{fixed}}$ :

$\varphi_{\text{fixed}} \leftarrow \varphi_{\text{fixed}} \cup \varphi_k^h$

end for

end for

---



**Figure 10.** (a) The results of the fine registration of the six triplets reported in Figure 3. Each color represents a sub-pointcloud coming from a depth sensor. (b) cross-section at  $z = 25$  mm; (c) cross-section at  $z = 100$  mm; (d) cross-section at  $z = 250$  mm.

#### 4. Results and Discussion

The computer-based methodology proposed in this study was implemented using custom software coded in MATLAB 2022. This decision allowed for the definition of a unified framework for automatically realigning point clouds by incorporating functions written in Python. Table 1 presents the parameter values used to implement various steps within the proposed procedure, as applied in the case studies.

**Table 1.** List of parameters and corresponding values for each step of the proposed methodology.

Step	Parameter	Values
Hand key-point detection	Detection score threshold	0.5
Forearm key-point detection	$H_{FKP}$ : heights below the barycenter to define forearm key points	−200 mm:−50 mm:−350 mm
Coarse registration of point clouds	$w_H$ : Horn’s quaternion-based method weights [54]	1
Fine registration of point clouds	SLSQP Voxel grid	0.75 mm
	$Toll$ : distance threshold value (Equations (3) and (5))	1.5 mm
	SLSQP Stopping criteria	$\left\  \frac{1}{f_1 f_2}^{iter} - \frac{1}{f_1 f_2}^{iter-1} \right\  < 10^{-4}$
	$N_{iter\_ex}$ (Algorithm 1)	9

This study evaluates the proposed methodology's potential for realigning morphologically complex shapes, specifically focusing on human upper limbs. The arms of healthy subjects (five volunteers from the Department of Civil and Industrial Engineering at the University of Pisa) were used to ensure that the quality of the overall acquisition was not influenced by pathologies such as strokes or dystonia disorders.

#### 4.1. Assessment of the Performance of the Proposed Methodology

Firstly, to quantitatively assess the performance of the proposed methodology, the obtained final registration results were compared with those obtained by a gold-standard approach: manual coarse alignment of the point clouds by an operator followed by an ICP algorithm for fine registration (Algorithm 2) using commercial software (i.e., Geomagic Studio 2014<sup>®</sup> v. 1.0 (from 3D Systems, Rock Hill, SC, USA)). The resulting point clouds were transformed into a common reference system to enable a fair comparison between the automatic and manual methods. In this reference system, the origins correspond to the respective centers of gravity of the point clouds. At the same time, the z- and x-axis directions are associated with the lower and intermediate inertia values, respectively.

---

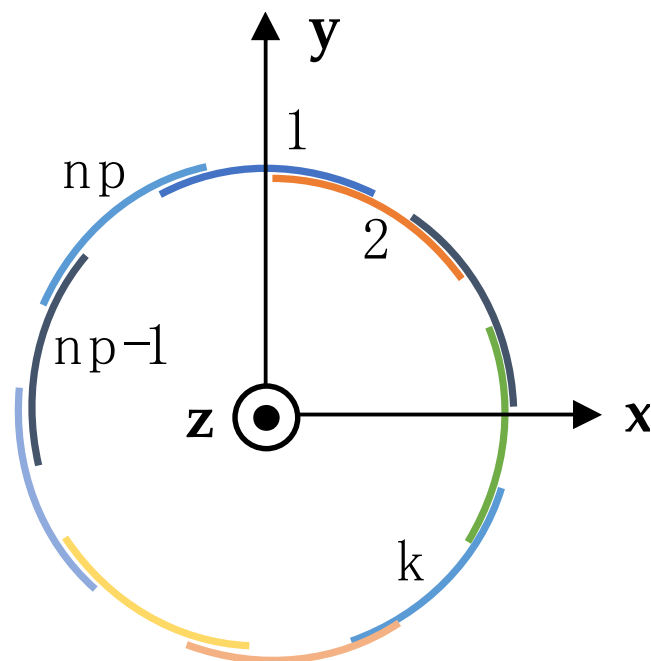
**Algorithm 2.** Algorithm for the fine global registration of np point clouds using the standard semi-automatic alignment pipeline.

---

**Require:** Point clouds  $\{\varphi_1^0, \varphi_2^0, \dots, \varphi_{np}^0\}$

Labeling of point clouds from 1 to np in a clockwise direction

---



for  $h = 1$  to  $np-1$  do

    Fix  $\varphi_h^0$  and leave  $\varphi_{h+1}^0$  floating.

    Manual selection of at least three corresponding points (i.e., fingertips, elbow, wrist bones, natural skin marks) on common areas of  $\varphi_h^0$  and  $\varphi_{h+1}^0$

    Rough alignment between the two adjacent point clouds using the 3-2-1 pairwise registration process.

    Fine registration between the two adjacent point clouds by ICP algorithm.

end for

Global registration by ICP algorithm

---

Figure 11 illustrates, as an example, the results obtained for two cases. Each subplot in the figure displays a map of distances between the reference (clouds realigned with the manual method) and the alignment achieved with the proposed method. Additionally, a histogram is provided for each case, showing the frequencies and cumulative frequencies of the obtained discrepancies. The cumulative frequencies are specifically marked for distances of 5 mm and 10 mm. The two cases were selected to show the worst-case scenario (case#1) and the best-case scenario (case #2) obtained during the trials. As can be seen, even for case #1, 95% of the points have a distance lower than 5 mm, while, for case #2, the percentage rises to 98%.

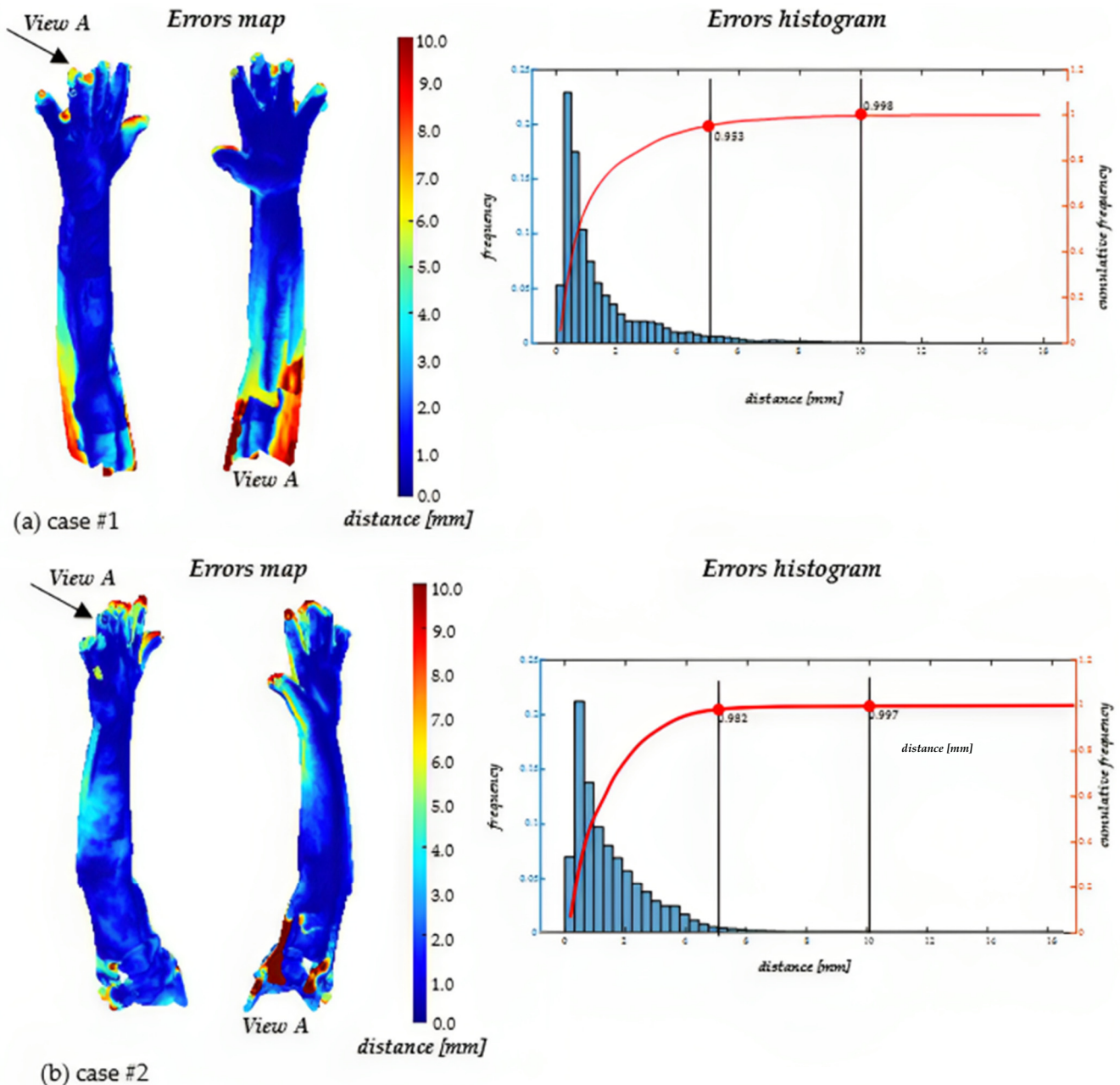


Figure 11. Examples of the results obtained.



Table 2 reports the results in terms of the means and standard deviation values of the same parameters obtained by analyzing all the cases. Most deviations were below values considered acceptable by the medical personnel involved in this research activity. In particular, an average of 96% of the points were found to have a distance of less than 5 mm compared with the model obtained with the conventional pipeline, while the average distance between the two clouds was less than 2 mm. The distinction between hand and arm was made to verify whether differences appear between almost uniform surfaces (like those of the arm) and surfaces characterized by a high degree of optical undercuts (for example, the fingers). The finger region poses the most significant challenges during the acquisition process, primarily due to self-undercuts, rapid slope variations of the surface, and the tendency for involuntary movements of the fingers. Even if the acquisition by the three RGB-D cameras is almost simultaneous for each scanner pose, the reconstruction of the overall limb anatomy requires the scanner to be moved around the arm. For this reason, any non-rigid limb movement during the acquisition process (which can last up to 50 s) effectively modifies the geometry to be reconstructed, resulting in data that cannot be perfectly realigned regardless of the methodology adopted.

**Table 2.** Mean distances and standard deviation values from the gold standard.

Parameter		Hand	Arm	Total
% points dist < 3 mm	mean	0.884	0.859	0.870
	std	0.088	0.073	0.014
% points dist < 6 mm	mean	0.979	0.975	0.976
	std	0.021	0.018	0.010
Dist, mm	mean	1.429	1.634	1.522
	std	0.572	0.363	0.026

Nevertheless, the data in Table 2 confirm that the two procedures (i.e., conventional manual registration and fully automated registration) provide almost indistinguishable results, since a relative point-by-point distance of less than 2 mm is considered negligible for soft anatomical tissues. Additionally, it is worth noting that the manual registration procedure requires an experienced operator with a point-cloud-processing background (thus, not medical staff) and about 20 min for a full 360° scan. On the other hand, the proposed prototype implementation of a fully automated procedure requires up to 30 min without human intervention. Specifically, the two procedures are differentiated by the time distribution of their two main steps. The gold-standard approach is time-consuming at the coarse alignment step since the operator must visually identify and select corresponding points on adjacent point clouds. The subsequent refinement using ICP takes less time due to optimized routines within commercial software due to the skilled and trained operator's ability to visually identify corresponding landmarks. This visual identification results in an initial alignment of the point clouds, serving as a well-suited initialization for the subsequent optimization process. On the other hand, the developed approach is much faster during coarse initial alignment since identifying corresponding landmarks is automatic (neural network + forearm key-point detection) but is much slower during the convergence of the fine registration. The substantial advantage of a fully automated process, devoid of any operator intervention with specialized training, outweighs any potential increase in the time needed to achieve the outcome, making its effective use in clinical centers possible. Finally, it is essential to acknowledge that the developed procedure could benefit from an improved software implementation.

#### 4.2. Improvements to the State of the Art

The performance of the proposed method was compared with algorithms that had been previously published in the literature and for which implementations were available. This choice was made to ensure a fair comparison among various techniques, as inaccuracies in implementing procedures developed by other authors could compromise the integrity of the comparison. Even minor errors might lead to cascading effects, resulting in significantly divergent outcomes, making meaningful algorithmic comparisons challenging. The overall accuracy of the procedure hinges on the precision of individual phases, including scanning, coarse registration, and fine registration. Given that the proposed method's objectives involve generating accurate 3D models of upper limb anatomy through a low-cost handheld 3D scanner, this section of the study specifically compares the proposed method's accuracy with the current state of the art for both coarse and fine registration. In particular:

- The proposed algorithm for coarse registration is compared with the following:
  - ✓ 4PCS [24];
  - ✓ Super4PCS [25];
  - ✓ PCA [56];
  - ✓ FPFH [57].
- The proposed algorithm for fine registration is compared with the following:
  - ✓ ICP point-to-plane [58];
  - ✓ ICP trimmed (Tukeyloss) point-to-plane [59].

The comparison was carried out using the same five test cases and the corresponding gold standards presented in the previous section. Additionally, for assessing the outcomes produced by the compared fine-registration algorithms (specifically, ICP point-to-plane and ICP trimmed point-to-plane), the proposed coarse-registration method provides the initial solution.

Table 3 compares the mean distances and the standard deviation values of the analyzed coarse-registration methods with the corresponding gold standard. The parameters listed for each of the five methods are those for which the best results were obtained. Values for 4PCS [24] and Super4PCS [25] are not reported as the methods do not provide acceptable solutions. It is also evident that the proposed procedure, based on the realignment of specific anatomical features of the upper limb, allows for much more accurate initial alignment of 3D point clouds than other analyzed state-of-the-art methods.

**Table 3.** Comparison of mean distances and standard deviation values from coarse-registration methods and the gold standard.

Method	Parameters	Distances [mm]					
		Total		Hand		Arm	
		Mean	Std	Mean	Std	Mean	Std
Proposed method	See Table 1	2.440	0.742	2.871	1.011	2.175	0.725
4PCS [24]	$d = 5, n = 200, t = 1000$	--	--	--	--	--	--
Super4PCS [25]	$d = 5, n = 200, t = 1000$	--	--	--	--	--	--
PCA	$n = 3$	9.616	0.446	6.233	1.162	11.638	0.160
FPFH	$n = 3$	5.137	1.012	4.154	0.911	5.722	1.031

Table 4 compares the mean distances and the standard deviation values of the analyzed fine-registration methods with the corresponding gold standard. Again, in this case, the parameters listed for each of the three methods are those for which the best results were obtained. By implementing the original function to be minimized, the proposed method achieves a notably higher level of accuracy in aligning point clouds compared with the methods examined in the current state of the art. Moreover, the low standard deviation

values associated with the proposed method denote the promising robustness of the results obtainable by aligning the upper limb point clouds.

**Table 4.** Comparison of mean distances and standard deviation values from fine-registration methods and the gold standard.

Method	Parameters	Distances [mm]					
		Total		Hand		Arm	
		Mean	Std	Mean	Std	Mean	Std
Proposed method	See Table 1	1.429	0.572	1.634	0.363	1.522	0.026
ICP point-to-plane	$d = 0.75$ , $knn = 30$	4.425	3.733	4.619	4.027	4.315	3.575
ICP trimmed (Tukeyloss) point-to-plane	$d = 0.75$ , $knn = 30$ , $\sigma = 1.5$	2.964	1.079	3.095	1.274	2.891	0.972

## 5. Conclusions

This research paper proposes a novel approach for automatically aligning and registering point clouds obtained by scanning upper limb anatomy with a depth-camera-based scanner. The methodology integrates information extracted from both 2D and 3D data. Specifically, a pre-trained neural network (NN) was exploited to automatically detect hand key points by processing RGB images captured by the depth cameras. Additionally, an automated algorithm was developed to identify the forearm's skeleton line directly from the point cloud data. The combination of hand key points and forearm key points allows for better constraint of the coarse registration of point clouds obtained from different scanner poses. The automatic recognition of forearm features further enhances the methodology's robustness, enabling it to handle a wide range of hand poses and even point clouds with limited overlap and noise.

Additionally, the coarse alignment step has proven effective in providing a reliable starting alignment for the subsequent fine-registration process, which is obtained through an innovative registration algorithm (inspired by ICP). To evaluate the proposed methodology, its results were analyzed using, as the reference, a gold-standard point cloud re-alignment pipeline conducted using a hybrid manual and automatic procedure within commercial software. The assessment demonstrated the effectiveness of the proposed approach in achieving accurate registration results. Most deviations were found to be below acceptable values, with approximately 96% of the points having a distance of less than 5 mm compared with the model obtained with the conventional pipeline (with an average distance value lower than 2 mm). This makes the proposed method a feasible alternative to manual registration, which also requires the intervention of a skilled operator, introducing some subjectivity in identifying common anatomical landmarks. Finally, the improvements over the state of the art were verified by comparing the proposed method with existing published methods for which implementations were accessible.

The main aim of the proposed methodology is to contribute to developing an effective and personalized upper limb rehabilitation framework. In this context, assembling a compact hand-held scanner and an automated procedure for registering partially overlapped and noisy point clouds is of utmost importance to allow the procedure to be completed in clinics without the intervention of reverse engineering experts. Although the approach has demonstrated encouraging outcomes, it is essential to conduct additional validation and testing using more comprehensive and varied datasets. This is necessary to thoroughly evaluate its performance in various scenarios, especially in clinical environments involving actual patients exhibiting potential movement disorders. Subsequent clinical validation will encompass the three-dimensional reconstruction of upper limb anatomy for individuals experiencing diminished mobility in their upper limbs. This includes cases related to patients who have undergone sports injuries or suffered strokes. The main limitation of the proposed approach is given by the 3D scanner, which requires multiple placements around

the arm to complete the data acquisition process. This could be an issue in the case of patients not being able to maintain the arm in the same configuration for the time required to complete the scanning process (as in the case of fingers or wrist movements). However, the choice of the proposed scanner and scanning strategy was oriented towards defining a tradeoff between cost and accurate results. Future design efforts should be towards the definition of alternatives to mitigate or remove this limitation.

**Author Contributions:** Conceptualization, L.D.A., P.D.S., E.G., P.N., A.P. and A.V.R.; methodology, L.D.A., E.G., P.N. and A.P.; software, E.G. and P.N.; validation, E.G. and P.N.; formal analysis, L.D.A., E.G., P.N. and A.P.; investigation, L.D.A., E.G., P.N. and A.P.; resources, P.D.S. and A.V.R.; data curation, L.D.A., E.G., P.N. and A.P.; writing—original draft preparation, L.D.A., E.G., P.N. and A.P.; writing—review and editing, P.D.S. and A.V.R.; supervision, L.D.A., P.D.S. and A.V.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, E.G., upon reasonable request.

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

## References

- Perry, J.C.; Brower, J.R.; Carne, R.H.R.; Bogert, M.A. 3D Scanning of the Forearm for Orthosis and HMI Applications. *Front. Robot. Ai* **2021**, *8*, 576783. [[CrossRef](#)]
- Leong, S.C.; Tang, Y.M.; Toh, F.M.; Fong, K.N.K. Examining the effectiveness of virtual, augmented, and mixed reality (VAMR) therapy for upper limb recovery and activities of daily living in stroke patients: A systematic review and meta-analysis. *J. Neuroeng. Rehabil.* **2022**, *19*, 93. [[CrossRef](#)]
- Urquhart, L.; Petrakis, K.; Hansen, J.P.; Wodehouse, A.; Mariani, M.E.; Lauer-Schmaltz, M.W.; Loudon, B. Prototyping Approaches for Rehabilitation Devices: From Product Embodiment to Data Management. *Comput. Aided Des. Appl.* **2023**, *20*, 145–157. [[CrossRef](#)]
- Bartol, K.; Bojanic, D.; Petkovic, T.; Pribanic, T. A Review of Body Measurement Using 3D Scanning. *IEEE Access* **2021**, *9*, 67281–67301. [[CrossRef](#)]
- Haleem, A.; Javaid, M. 3D scanning applications in medical field: A literature-based review. *Clin. Epidemiol. Glob.* **2019**, *7*, 199–210. [[CrossRef](#)]
- Grazioso, S.; Selvaggio, M.; Di Gironimo, G. Design and development of a novel body scanning system for healthcare applications. *Int. J. Interact. Des. M* **2018**, *12*, 611–620. [[CrossRef](#)]
- Zeraatkar, M.; Khalili, K. A Fast and Low-Cost Human Body 3D Scanner Using 100 Cameras. *J. Imaging* **2020**, *6*, 21. [[CrossRef](#)] [[PubMed](#)]
- Straub, J.; Kerlin, S. A Very Low-Cost 3D Scanning System for Whole-Body Imaging. In *Smart Biomedical and Physiological Sensor Technology XII*; SPIE: Bellingham, WA, USA, 2015; Volume 9487, pp. 122–131.
- Kersten, T.P.; Lindstaedt, M.; Starosta, D. Comparative Geometrical Accuracy Investigations of Hand-Held 3d Scanning Systems—AN Update. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *422*, 487–494. [[CrossRef](#)]
- Kalinowski, P.; Hindmarch, J.; Luhmann, T. Accuracy Investigations of Hand-Held Scanning Systems Using Different Dumbbell Artefacts. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *43*, 401–407. [[CrossRef](#)]
- Dessery, Y.; Pallari, J. Measurements agreement between low-cost and high-level handheld 3D scanners to scan the knee for designing a 3D printed knee brace. *PLoS ONE* **2018**, *13*, e0190585.
- Redaelli, D.F.; Barsanti, S.G.; Biffi, E.; Storm, F.A.; Colombo, G. Comparison of geometrical accuracy of active devices for 3D orthopaedic reconstructions. *Int. J. Adv. Manuf. Tech.* **2021**, *114*, 319–342. [[CrossRef](#)]
- Vitali, A.; Togni, G.; Regazzoni, D.; Rizzi, C.; Molinero, G. A virtual environment to evaluate the arm volume for lymphedema affected patients. *Comput. Meth Prog. Bio* **2021**, *198*, 105795. [[CrossRef](#)] [[PubMed](#)]
- Buonamici, F.; Carfagni, M.; Puggelli, L.; Servi, M.; Volpe, Y. A Fast and Reliable Optical 3D Scanning System for Human Arm. In *Advances on Mechanics, Design Engineering and Manufacturing III*; Roucoules, L., Paredes, M., Eynard, B., Morer Camo, P., Rizzi, C., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 268–273.
- Neri, P.; Barone, S.; Paoli, A.; Razonale, A.V.; Tamburrino, F. A Depth-Camera Based System for the Real-Time Scanning of Upper Limb Anatomy. In *Proceedings of the International Conference on Design, Simulation, Manufacturing: The Innovation Exchange*, Rome, Italy, 9–10 September 2021; Springer: Cham, Switzerland, 2022; pp. 245–255.
- Bernardini, F.; Rushmeier, H. The 3D model acquisition pipeline. *Comput. Graph. Forum* **2002**, *21*, 149–172. [[CrossRef](#)]

17. Bonarrigo, F.; Signoroni, A. Global registration of large collections of range images with an improved Optimization-on-a-Manifold approach. *Image Vis. Comput.* **2014**, *32*, 437–451. [[CrossRef](#)]
18. Li, L.H.; Wang, R.W.; Zhang, X.P. A Tutorial Review on Point Cloud Registrations: Principle, Classification, Comparison, and Technology Challenges. *Math. Probl. Eng.* **2021**, *2021*, 9953910. [[CrossRef](#)]
19. Pickup, D.; Sun, X.; Rosin, P.L.; Martin, R.R.; Cheng, Z.; Lian, Z.; Aono, M.; Ben Hamza, A.; Bronstein, A.; Bronstein, M.; et al. Shape Retrieval of Non-rigid 3D Human Models. *Int. J. Comput. Vis.* **2016**, *120*, 169–193. [[CrossRef](#)]
20. Berretti, S.; Daoudi, M.; Turaga, P.; Basu, A. Representation, Analysis, and Recognition of 3D Humans: A Survey. *Acm T Multim Comput.* **2018**, *14*, 1–36. [[CrossRef](#)]
21. Yan, Y.J.; An, J.Y.; Zhao, J.; Shen, F.R. Hybrid optimization with unconstrained variables on partial point cloud registration. *Pattern Recogn.* **2023**, *136*, 109267. [[CrossRef](#)]
22. Han, X.-F.; Feng, Z.-A.; Sun, S.-J.; Xiao, G.-Q. 3D point cloud descriptors: State-of-the-art. *Artif. Intell. Rev.* **2023**, *56*, 12033–12083. [[CrossRef](#)]
23. Cheng, L.; Chen, S.; Liu, X.Q.; Xu, H.; Wu, Y.; Li, M.C.; Chen, Y.M. Registration of Laser Scanning Point Clouds: A Review. *Sensors* **2018**, *18*, 1641. [[CrossRef](#)]
24. Aiger, D.; Mitra, N.J.; Cohen-Or, D. 4-points congruent sets for robust pairwise surface registration. *ACM Trans. Graph.* **2008**, *27*, 1–10. [[CrossRef](#)]
25. Mellado, N.; Aiger, D.; Mitra, N.J. SUPER 4PCS Fast Global Pointcloud Registration via Smart Indexing. *Comput. Graph. Forum* **2014**, *33*, 205–215. [[CrossRef](#)]
26. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *Ieee T Pattern Anal.* **1992**, *14*, 239–256. [[CrossRef](#)]
27. Chetverikov, D.; Svirko, D.; Stepanov, D.; Krsek, P. The Trimmed Iterative Closest Point Algorithm. In Proceedings of the 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002; pp. 545–548.
28. Wang, X.; Zhu, X.H.; Ying, S.H.; Shen, C.M. An Accelerated and Robust Partial Registration Algorithm for Point Clouds. *IEEE Access* **2020**, *8*, 156504–156518. [[CrossRef](#)]
29. Du, S.Y.; Zhu, J.H.; Zheng, N.N.; Liu, Y.H.; Li, C. Robust iterative closest point algorithm for registration of point sets with outliers. *Opt. Eng.* **2011**, *50*, 087001. [[CrossRef](#)]
30. Dong, J.M.; Peng, Y.X.; Ying, S.H.; Hu, Z.Y. LieTrICP: An improvement of trimmed iterative closest point algorithm. *Neurocomputing* **2014**, *140*, 67–76. [[CrossRef](#)]
31. Dong, Z.; Liang, F.X.; Yang, B.S.; Xu, Y.S.; Zang, Y.F.; Li, J.P.; Wang, Y.; Dai, W.X.; Fan, H.C.; Hyyppa, J.; et al. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. *ISPRS J. Photogramm.* **2020**, *163*, 327–342. [[CrossRef](#)]
32. Wang, Y.J.; Yan, C.G.; Feng, Y.T.; Du, S.Y.; Dai, Q.H.; Gao, Y. STORM: Structure-Based Overlap Matching for Partial Point Cloud Registration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 1135–1149. [[CrossRef](#)]
33. Aoki, Y.; Goforth, H.; Srivatsan, R.A.; Lucey, S. PointNetLK: Robust & Efficient Point Cloud Registration using PointNet. In Proceedings of the 2019 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr 2019), Long Beach, CA, USA, 16–20 June 2019; pp. 7156–7165.
34. Qi, C.R.; Su, H.; Mo, K.C.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85.
35. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence—Volume 2, Vancouver, BC, Canada, 24–28 August 1981; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1981; pp. 674–679.
36. Sarode, V.C.; Li, X.; Goforth, H.; Aoki, Y.; Rangaprasad, A.S.; Lucey, S.; Choset, H. PCRNNet: Point Cloud Registration Network using PointNet Encoding. *arXiv* **2019**, arXiv:1908.07906.
37. Zhou, R.Q.; Li, X.X.; Jiang, W.S. SCANet: A Spatial and Channel Attention based Network for Partial-to-Partial Point Cloud Registration. *Pattern Recogn. Lett.* **2021**, *151*, 120–126. [[CrossRef](#)]
38. Xu, H.; Liu, S.C.; Wang, G.F.; Liu, G.H.; Zeng, B. OMNet: Learning Overlapping Mask for Partial-to-Partial Point Cloud Registration. In Proceedings of the 2021 IEEE/Cvf International Conference on Computer Vision (Iccv 2021), Virtual, 11–17 October 2021; pp. 3112–3121.
39. Zhang, W.H.; Zhang, Y.; Li, J.L. A Two-Stage Correspondence-Free Algorithm for Partially Overlapping Point Cloud Registration. *Sensors* **2022**, *22*, 4697. [[CrossRef](#)] [[PubMed](#)]
40. Wang, Y.; Sun, Y.B.; Liu, Z.W.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. [[CrossRef](#)]
41. Zhang, K.G.; Hao, M.; Wang, J.; Chen, X.X.; Leng, Y.Q.; de Silva, C.W.; Fu, C.L. Linked Dynamic Graph CNN: Learning through Point Cloud by Linking Hierarchical Features. In Proceedings of the 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Shanghai, China, 26–28 November 2021.
42. Shen, Y.R.; Feng, C.; Yang, Y.; Tian, D. Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 21–26 July 2018; pp. 4548–4557.



43. Dominguez, M.; Dhamdhere, R.; Petkar, A.; Jain, S.; Sah, S.; Ptucha, R. General-Purpose Deep Point Cloud Feature Extractor. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (Wacv 2018), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1972–1981.
44. Wang, L.; Huang, Y.C.; Hou, Y.L.; Zhang, S.M.; Shan, J. Graph Attention Convolution for Point Cloud Semantic Segmentation. In Proceedings of the 2019 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr 2019), Long Beach, CA, USA, 16–20 June 2019; pp. 10288–10297.
45. Simonovsky, M.; Komodakis, N. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 29–38.
46. Xu, Q.G.; Sun, X.D.; Wu, C.Y.; Wang, P.Q.; Neumann, U. Grid-GCN for Fast and Scalable Point Cloud Learning. In Proceedings of the 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5660–5669.
47. Shi, W.J.; Rajkumar, R. Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1708–1716.
48. Carfagni, M.; Furferi, R.; Governì, L.; Santarelli, C.; Servi, M.; Uccheddu, F.; Volpe, Y. Metrological and Critical Characterization of the Intel D415 Stereo Depth Camera. *Sensors* **2019**, *19*, 489. [[CrossRef](#)] [[PubMed](#)]
49. Chen, C.; Yang, B.S.; Song, S.; Tian, M.; Li, J.P.; Dai, W.X.; Fang, L.N. Calibrate Multiple Consumer RGB-D Cameras for Low-Cost and Efficient 3D Indoor Mapping. *Remote Sens.* **2018**, *10*, 328. [[CrossRef](#)]
50. Neri, P.; Paoli, A.; Aruanno, B.; Barone, S.; Tamburrino, F.; Razonale, A.V. 3D scanning of Upper Limb anatomy by a depth-camera-based system. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2023**, 1–12. [[CrossRef](#)]
51. Liu, Y.; Jiang, J.; Sun, J.H. Hand Pose Estimation from RGB Images Based on Deep Learning: A Survey. In Proceedings of the 2021 IEEE 7th International Conference on Virtual Reality (Icvr 2021), Foshan, China, 20–22 May 2021; pp. 82–89.
52. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *Ieee Trans. Pattern Anal.* **2021**, *43*, 172–186. [[CrossRef](#)]
53. MediaPipe Hand Landmarker. Available online: [https://developers.google.com/mediapipe/solutions/vision/hand\\_landmarker](https://developers.google.com/mediapipe/solutions/vision/hand_landmarker) (accessed on 9 September 2023).
54. Horn, B.K.P.; Hilden, H.M.; Negahdaripour, S. Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Am. A* **1988**, *5*, 1127–1135. [[CrossRef](#)]
55. Kraft, D. *A Software Package for Sequential Quadratic Programming*; Tech. Rep. DFVLR-FB 88-28; DLR German Aerospace Center, Institute for Flight Mechanics: Koln, Germany; p. 1988.
56. Tipping, M.E.; Bishop, C.M. Mixtures of Probabilistic Principal Component Analyzers. *Neural Comput.* **1999**, *11*, 443–482. [[CrossRef](#)]
57. Rusu, R.B.; Blodow, N.; Beetz, M. Fast Point Feature Histograms (FPFH) for 3D Registration. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 3212–3217.
58. Chen, Y.; Medioni, G. Object modelling by registration of multiple range images. *Image Vis. Comput.* **1992**, *10*, 145–155. [[CrossRef](#)]
59. Babin, P.; Giguère, P.; Pomerleau, F. Analysis of Robust Functions for Registration Algorithms. *arXiv* **2018**, arXiv:cs.RO/1810.01474.

**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.