Augmented and Mixed Reality Scene Alignment Through 3D-to-3D Learning-Based Cross-Source Point Cloud Registration

Juan Sardi-Barzallo[1], Norbert Haala[2], Volker Coors[1]

[1] Centre for Geodesy and Geoinformatics, Stuttgart Technical University of Applied Sciences (HFT Stuttgart), Stuttgart, Germany. (juan.sardi-barzallo, volker.coors)@hft-stuttgart.de
[2]Institute for Photogrammetry and Geoinformatics, University of Stuttgart, Stuttgart, Germany. (norbert.haala@ifp.uni-stuttgart.de)

Keywords: Cross-Source Point Cloud Registration, Mixed Reality, Augmented Reality Camera localization, Scene alignment.

Abstract

Rapid advancements in reality capture technology and increasing accessibility to devices capable of generating point cloud data have led to a greater prevalence of applications requiring the interaction and integration of cross-source Point Cloud Data (PCD). Augmented and Mixed Reality (AR/MR) technologies are similarly pivotal for integrating digital and physical environments by overlaying Digital Twin (DT) models onto real-world contexts. While AR/MR devices can generate real-time 3D point clouds, integrating this data with high-precision sources, such as LiDAR, remains a research challenge—particularly regarding scene alignment and camera localization. Accurate positioning of AR/MR users within large-scale pre-scanned environments enables the deployment of digital content at predefined locations without prior field preparation. However, conventional vision-based methods are often vulnerable to environmental variations, complicating reliable localization. This work proposes an exclusively 3D-to-3D-based methodology for AR/MR scene alignment and camera localization, addressing cross-source registration in scenarios with significant size disparity. By combining a learning-based Voxel Representation and Hierarchical Correspondence Filtering (VRHCF) method with the Truncated Least Squares Estimation And SEmidefinite Relaxation (TEASER++) algorithm, our approach effectively manages asymmetric heterogeneous point cloud data. The results demonstrate promising gross registration performance, particularly in extensive indoor settings. Although computational constraints and outlier sensitivity in outdoor environments warrant further investigation, this study underscores the potential of advanced 3D-to-3D methodologies to enable seamless interaction between digital and physical worlds.

1. Introduction

1.1 Overview

The advent of the fourth industrial revolution marks a significant milestone in the digitization of production models. It depends on the development of Cyber-Physical Systems (CPS), such as Digital Twin (DT) systems (Zhou et al., 2015), which can facilitate real-time communication, interaction, and data exchange among humans, products, and devices during production processes (Tao et al., 2019; Zhou et al., 2015)). In this context, Extended Reality (XR) technologies are prominent interface tools that support visualization, interaction, and real-time contextual augmentation between physical assets and human cognitive functions (Kunz et al., 2022).

Nevertheless, these possibilities of natural interaction pose several challenges. Technology still struggles to achieve sufficient quality in reproducing immersive features (collision, occlusion), especially in scene alignment, which aligns virtual and physical objects so they appear spatially consistent and realistic, and is needed to recreate real-life-like immersive experiences. MR and AR applications depend on accurate scene alignment for correct operation and expected functionality ((Li et al., 2020)). Accurate camera pose estimation and scene alignment are both essential for enabling and improving fine-grained input and for selecting 3D digital content ((Kern et al., 2023)). They also help avoid coexistence failures between real-world and digital content ((Zhou et al., 2009)).

The alignment of AR and MR scenes can be understood as the accurate, consistent superposition of virtual 3D content (models or elements) at specific locations within the real environment, aiming to achieve placement accuracy, real-time interactivity, and 3D tracking ((El Barhoumi et al., 2022)). In general, scene alignment presents three significant challenges. The first is identifying the location where the device is initialized by gathering information about the immediate context, such as images or 3D data. The second is comparing the gathered information to a previous representation of the environment, such as a map. Finally, if the device is located within the expected area, its position and rotation are determined, which is similar to the visual place recognition problem that has been studied extensively in robotics ((Lowry et al., 2016; Herbers and König, 2019).

1.2 Mixed Reality Camera Location challenge

The methods used for addressing the camera location challenge can be classified into sensor-based, vision-based, and hybrid methods. (Kharroubi et al., 2020; Huang et al., 2023a). Sensor-based methods use external systems like Wi-Fi, Radio Frequency Identification (RFID), or Global Navigation Satellite System (GNSS) to determine camera pose, offering reliability in large outdoor areas but facing issues such as signal degradation (e.g., in the presence of urban canyons) and requiring significant setup, especially indoors (Haneefa et al., 2024; Herbers and König, 2019; Huang et al., 2023a). Vision-based methods rely on analyzing environmental images (physical markers or natural features) to extract distinctive, unique, highly invariant identifiers through the use of point detectors and descriptor algorithms to identify unique features for camera pose estimation (Lowe, 2004; Bay et al., 2006; Rublee et al., 2011; Long Quan and Zhongdan Lan, 1999; Xie and Zou, 2024), but they struggle with changing lighting, dynamic occlusions, and at initialization stage at complex and large scale outdoors environments (Marchand et al., 2016; Radanovic et al., 2023; El Barhoumi et al., 2022; Huang et al., 2023a).

On the other hand, Hybrid methods combine sensor and visual data to enhance robustness and accuracy, thereby reducing computational complexity but at the cost of increased implementation complexity and operational expenses (El Barhoumi et al., 2022; Messi et al., 2024; Huang et al., 2023a). Recently, 3D-based methods have shown promise as more efficient solutions for searching for 3D-to-3D correspondences to estimate camera pose, which depends on the availability of 3D coordinates of the observed points (Radanovic et al., 2023). These methods offer the advantage of being unaffected by environmental changes in light and color. Additionally, due to their similarity to the 'point cloud registration problem' (Haneefa et al., 2024), they can leverage methods developed for point cloud registration, particularly deep learning approaches. Although these approaches are still in the early stages of development (Guo et al., 2020), they have demonstrated improved performance over traditional handcrafted methods (Radanovic et al., 2023).

1.3 Point Cloud Registration

Point Cloud Registration (PCR) addresses the problematic of estimating the transformation between two point clouds (Huang et al., 2021), that is to say, to infer the translation and rotation parameters between a source point cloud – which will be moved from its original location – and target point cloud – which is in a fixed position - in order to achieve their alignment on a same coordinate system space. This aforementioned translation comprises a two stages, the coarse-alignment, which brings together the correspondent point clouds and the fine-alignment, which subsequently iterates on the data refining the alignment and achieving higher degree of accuracy. The relevance of PCR applications in fields such as computer vision, robotics, mapping, remote sensing is widely recognized (Huang et al., 2021; Xu et al., 2023; Zhang et al., 2025; Zhao et al., 2024b) jointly with more recent and rapid evolving fields such as Extended Reality (XR) (Ma et al., 2024; Osipov et al., 2023).

In the current scenario, where the trend of accessing high-quality point cloud information of the built environment is increasing (Haleem et al., 2022; Xu et al., 2025) and, where mixed reality devices are capable of reconstruct on real time the surrounding environment (Teruggi and Fassi, 2022), Rigid Pairwise PCR takes on relevance by providing the possibility of infer user localization and object pose estimation (Zhao et al., 2025). Nevertheless, to register point cloud data which have been captured with different sensors is a very challenging task due to density and scale difference, large amount of outliers, large percentage of rotation and low or partial overlapping conditions (**?**Huang et al., 2021).

The proliferation of point cloud data sources in several fields (e.g., urban mapping, environmental monitoring, and traffic modeling), coupled with technological advancements, has led to an increased prevalence of multi-sensor data (Teuscher et al., 2025). Despite of this, most of existing available PCR studies uses point cloud data from the same sensor, type and density (Huang et al., 2023b), with overlapping regions and similar size models. Therefore, it is imperative to study cross-source point cloud registration methods that better address the complexity of data from different sensors and mitigate the limited progress in this field in recent years (Zhao et al., 2024b).

Cross-Source Point Cloud Registration (CSPCR), also referred as Cross-Modality PCR (Ma et al., 2024), Multi-source PCR (Lee et al., 2024; Xu et al., 2025), Cross-Device PCR (Osipov et al., 2023) or Cross-Domain PCR (Poiesi and Boscaini, 2022), can be addressed from two perspectives optimization-based methods, learning-based methods (Huang et al., 2021). The first one pursues to find the optimal correspondence solution in front of large correspondence outliers (caused by density and location differences), using so called “hand-crafted” features on the original point clouds to find correspondences usually using methods such as Iterative Closest Point (ICP). The second one seeks to find robust point descriptors among the point cloud data in order to infer the transformation matrix by feature-based correspondences or direly by regression from the features (Huang et al., 2023b).

Within this context learning-based methods shows to largely outperform hand-crafted ones, however the its capabilities for generalization in scenarios with different point cloud data with which they were originally trained, it is still an open challenge (Poiesi and Boscaini, 2022). The some studies suggest that combination between optimization-based and learning-based methods could achieve high accuracy and efficiency (Huang et al., 2021; Osipov et al., 2023), nevertheless despite the big amount of deep-learning based proposed methods currently (Wei et al., 2024) there are few promising tested studies dealing with real-world point cloud data.

This research proposes the use of Cross-Source Point Cloud Registration Via Voxel Representation and Hierarchical Correspondence Filtering (VRHCF) (Zhao et al., 2024a), a promising Cross-Source PCR learning-based method, in combination with Truncated least squares Estimation And SEmidefinite Relaxation (TEASER++) (Yang et al., 2020) a robust registration algorithm, in order to address the large amount of outliers correspondences when dealing with real-world cross-source point cloud data registration at MR scenarios of large-scale difference and low overlap.

2. Related Studies

On the field of point cloud registration, numerous traditional and neural-based methods have been developed for rigid registration and for determining transformations between point clouds. As outlined by Lyu et al. (2024), these methods are either one-stage or two-stage. One-stage methods, such as PointNetLK (Aoki et al., 2019) and AlignNet-3D (GroB et al., 2019), directly compute transformation parameters without identifying point correspondences, streamlining the process but increasing computational complexity. Conversely, two-stage approaches first establish point correspondences and then compute transformations. Techniques like iterative optimization combine traditional methods such as Iterative Closest Point (ICP) (Besl and McKay, 1992) and RANSAC (Fischler and Bolles, 1981) with neural registration for initial alignments. Sequential methods, such as FCGF (Choy et al., 2019) and SpinNet (Ao et al., 2021), convert raw point clouds into feature descriptors, using intermediate modules to enhance accuracy before applying methods like RANSAC for final parameter estimation (Lyu et al., 2024).

A critical limitation of many existing methods is their reliance on training on standardized same-source point cloud datasets, which limits their effectiveness in real-world cross-source applications (Zhao et al., 2024b). D3Feat (Bai et al., 2020) is notable for its kernel point convolution approach and its adaptability to varied environments, and it has been successfully used for learning-based MR camera localization in indoor environments (Radanovic et al., 2023).

D3Feat distinguished itself through its simplicity, adaptability to environments of varying scale and density, and reliable extraction of geometric correspondences. The workflow utilized a pre-scanned environmental point cloud as a server-hosted reference and a real-time mesh from the MR device. Both clouds were processed for keypoint detection and feature description, enabling effective matching. The study applied subsequently, RANSAC and ICP algorithms to estimate translation and rotation parameters. Server-side processing, such as downsampling and expanding the receptive fields of descriptors, addresses limitations of MR devices. However, D3Feat faces scalability issues in large indoor or outdoor spaces and potential bottlenecks during intensive feature processing, especially with RANSAC, highlighting areas for further research.

Accordingly, Lee et al. (2024) present a similar concept, proposing a methodology for using Real-World Cross-Source point cloud data to perform PCR of several small-sized point clouds into a large-sized point cloud using a coarse-to-fine approach. This approach uses specific algorithms to extract exterior wall and building outline information for precise localization of multi-source point clouds produced by Unmanned Aerial Vehicle (UAV) photogrammetry, Mobile Mapping System (MMS), and Terrestrial Laser Scanning (TLS). This approach effectively manages large outlier quantities but still leaves gaps, particularly in scalability to indoor environments.

Within this context, the VRHCF method (Zhao et al., 2024a) addresses cross-source registration challenges by utilizing spherical voxel feature representation and Hierarchical Correspondence Filtering (HCF), which counteract density inconsistencies and feature distribution differences without the need for downsampling. Complementing this, the TEASER method (Yang et al., 2020) provides a robust solution for 3D point cloud registration under extreme outlier conditions (up to 99%), leveraging a Truncated Least Squares cost function and a graph-theoretic framework to disentangle scale, rotation, and translation estimation. The integration of VRHCF and TEASER++ promises a robust and efficient solution for cross-source point cloud registration, particularly for MR scenarios involving spatial mesh, terrestrial and mobile LiDAR, and photogrammetry data—addressing gaps identified in previous research.

3. Methodology

The proposed methodology infers the camera location or MR scene alignment using a method that relies solely on 3D data, unaffected by factors such as lighting or color changes, and without any external sensor information (Figure1). To this end, the study expects two point clouds: a large target point cloud generated using common surveying devices, such as LiDAR or Structure from Motion (SFM), and a small source point cloud generated by Time-of-Flight (ToF) cameras commonly found in MR devices. The target point cloud—corresponding to the environment in which an MR scene is to be rendered—is captured in advance and stored in a database for retrieval. On the other hand, the MR device captures the source point cloud in real time while the user views their immediate surroundings. Due to the high processing requirements and the limitations of mobile MR devices in this regard, the captured data is sent

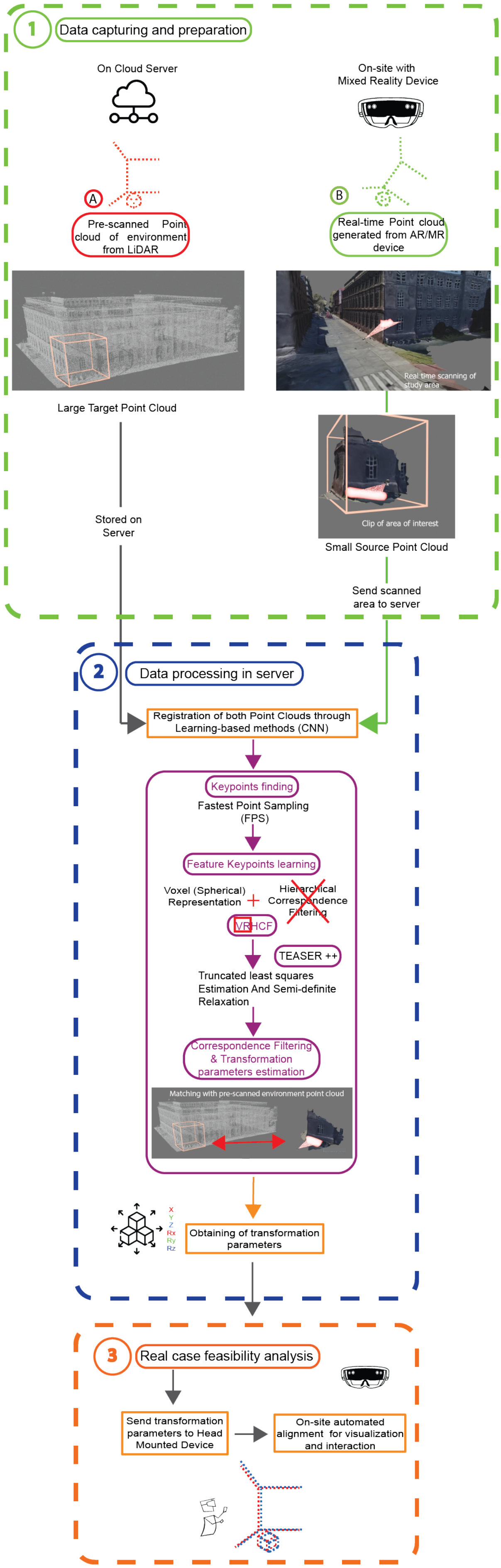

Figure 1. Overview of workflow diagram of the study. The presented project focuses in the first and second stages.

to a server for performing the point cloud registration. This study is inspired by the work of Radanovic et al. (2023) but differs by proposing an approach to the problem, using a different learning-based method for CS point cloud registration, and attempting to locate the subject within a considerably larger area. The study faces two major challenges regarding the registration of point clouds: the point clouds are generated by different sensors, and there is a significant disparity in the size of the point clouds to be registered—that is, the smaller point cloud (source point cloud) will be completely embedded within the larger point cloud (target point cloud). As a result, the scenario presented faces the well-known challenges of cross-source point cloud registration, including scale differences, significant rotation, and—particularly in this study—partial overlap, density differences, and large outliers (Huang et al., 2023a).

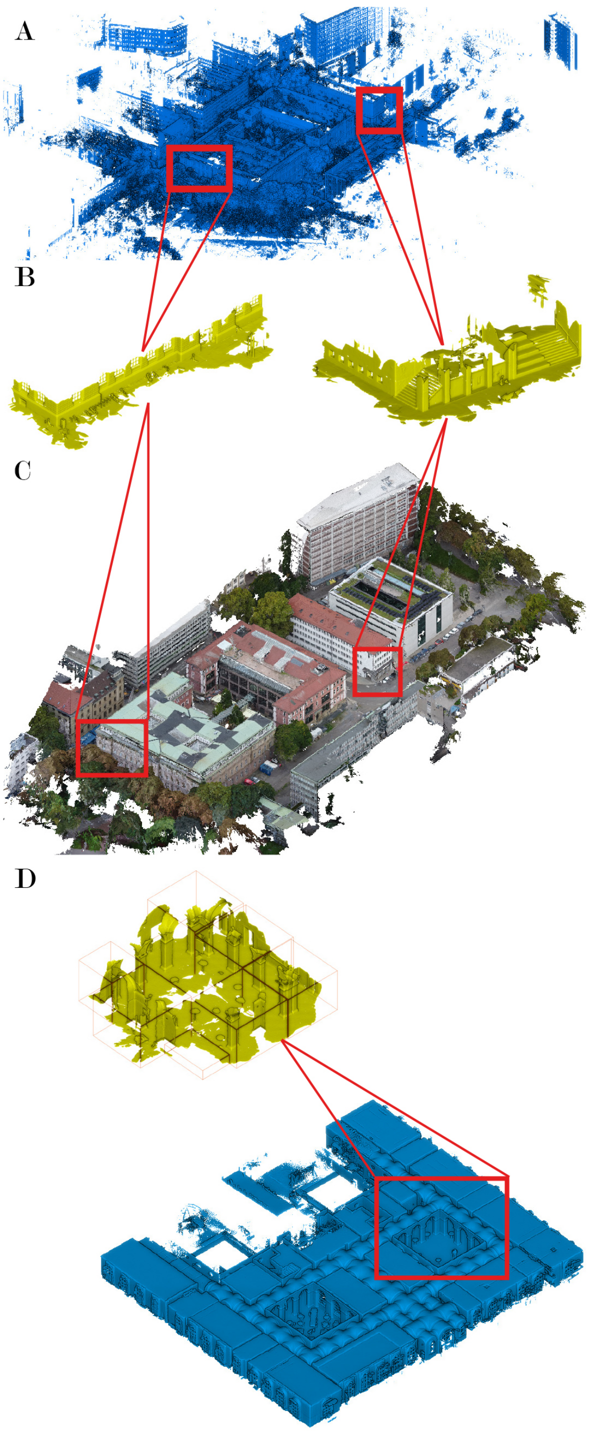

Figure 2. Datasets of the study area with their respective expected registration locations (in red). A. LiDAR outdoors dataset (in blue). B& C Outdoors MR Source data set (In yellow). D. MR outdoor dataset divided in Tiles, pointing out its registration location.

To address these challenges, this study analyzes the VRHCF framework Zhao et al. (2024a), which leverages spherical voxel-based feature representations to handle the varying densities and distributions of point clouds. This normalizes the density of the analyzed point clouds to achieve uniform distribution, then selects a fixed number of points as anchors to define the keypoints. Each generated keypoint selects a local patch in its vicinity with a fixed number of points (2048), which are subsequently processed into high-dimensional descriptors (64-dimensional) using the proposed SphereNet neural network. This provides a robust description of each point cloud generated by different sensors, which must be matched to continue the registration process.

The matching method proposed in the second part of VRHCF uses Hierarchical Correspondence Filtering (HCF) to address the high number of existing outliers in the matching. This method identifies the global distribution of voxels in the source point cloud that can be roughly matched to those in the target, and discards entire sections of the point cloud that are poorly matched. It then performs fine-level matching: once the general regions have been matched, it focuses on individual point correspondences using local geometric consensus.

However, this method relies heavily on spatial density, so a small area of true voxels (in a small source point cloud, as in the case described) can be statistically overwhelmed by the massive number of incorrectly matched voxels in a target point cloud that is several times larger. This can result in the first step of HCF—where general regions of voxels that can be matched are sought—ignoring actual matches, thereby failing to achieve the desired registration. For this reason, this study proposes not using the HCF part of the method and instead seeks to match keypoints and feature keypoints using the Spherical Voxel Representation with the Truncated Least Squares Estimation and Semidefinite Relaxation (TEASER++) algorithm (Yang et al., 2020). This certifiable, robust mathematical method is not based on local voxel density but instead uses a Maximum Clique graph formulation to generate a massive graph in which each spatially compatible match is connected by an edge. This edge, in turn, generates a perfectly connected subgraph or Maximum Clique, which, through the use of graph solvers, filters out 99% of existing outliers.

Once server-side processing is complete and the proposed point clouds are coarse registered, a traditional handcrafted method for fine registration, such as ICP (Wang and Zhao, 2017), may be used for fine refinement. The resulting transformation matrix is transmitted back to the MR device, where it can be applied to integrate MR technologies into the architectural design process (Sardi-Barzallo and Coors, 2025) and to visualize digital twin models within a physical environment at full scale, among other use cases. This work focuses on the feasibility analysis of combining the proposed methods for cross-source point registration using large, asymmetric point cloud data (see Figure 1, part two). The client-server communication architecture, as well as the development and testing of the MR application in real-world scenarios (see Figure 1, part three), are beyond the scope of this study.

3.1 Study area

This work was conducted in indoor and outdoor environments at the Stuttgart Technical University of Applied Sciences, in Stuttgart, Germany. The study area covered approximately 49,905.06 m² of exterior space and 3,257 m² of interior space and analyzed as independent datasets. Outdoor spaces enclose the public areas surrounding the university, including green areas with vegetation, public roads with vehicular traffic, and the surrounding built environment as a common urban space. The Indoor area consisted of the ground floor of the university's administrative building, a historical building with a symmetrical space distribution, featuring high vaulted ceilings and wide corridors. Figure 2 part D.

3.2 Sensors and Data Preparation

The survey was conducted using different devices for both interior and exterior environments. The interior of the building was surveyed with a LiDAR Mobile Mapping System (MMS) from NavVis, the VLX2, equipped with two 16-layer LiDAR sensors to capture 3D measurements in motion and four 20 MP cameras on top of the device, set to a minimum grid resolution of 11 mm. The resulting indoor point cloud had 84,327,063 points, and no further downsampling was applied for the posterior registration. The exterior areas were surveyed using a Terrestrial Laser Scanner (TLS) from Z+F, the model Imager 5010 with a phase-shift LiDAR sensor, and a minimum resolution grid of 31.5mm at 50 meters, with an estimated average time per scan station of 1m 38s. The resulting point cloud consisted of 127,286,539 points, which were later post-processed.

Data were collected using the Time-of-Flight (ToF) camera integrated into the Microsoft HoloLens 2. The resulting mesh was sampled into point clouds, yielding an average of 1.5 million points for indoor environments and 900,000 for outdoor environments. Surveys were conducted in randomly selected representative areas of the building's interior and exterior. For the interior, an area of approximately 14 m × 16 m (224 m2) was surveyed, exceeding the device's default spatial mapping cropping distance of 3 to 5 meters. This approach was employed to generate a comprehensive dataset from which smaller tiles could be extracted. Figure3

The area was divided into four tiles of similar size, with approximate dimensions of 8 m x 8 m, representing the typical area a user can survey in a single MR session. For the MR exterior datasets, the surveyed area was preserved at its original size (approximately 186 m2 for tile 1 and 202 m2 for tile 2) due to the greater asymmetry in the target point cloud in outdoor contexts.

On the other hand,to reduce noise in the analyzed outdoor target data, vegetation, mobile elements noise, and reflections resulting from data acquisition were removed through filtering and manual segmentation. Also, all point cloud data outside the ToF Camera's height range, which is 4 to 5 meters above the ground, was removed to avoid unnecessary processing, resulting in a point cloud of 43,763,481 points.

Before starting registration processing, the target point clouds were aligned to a common origin in the local reference system. On the other hand, all the preprocessed source point cloud data

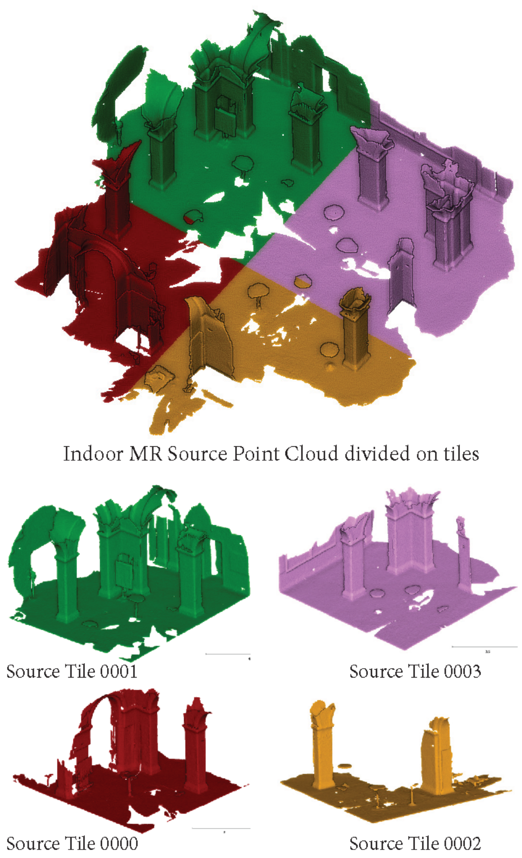

Figure 3. Indoor MR Source Point Cloud and its tiles

were sent to the origin of the Coordinate system with a random rotation to simulate the stage of running the MR application after a field survey.

3.3 Implementation

The project followed a logic in which the data from the MR device, obtained in real time in the field, was always treated as the source point cloud, while the previously scanned environment served as the target point cloud, which accommodated the significantly smaller point cloud. In all cases, the MR point cloud (source cloud) was entirely embedded within the target point cloud (LiDAR) due to the asymmetric size difference, resulting in a very low overlap of approximately 2-3% of the target point cloud's total covered area.

As the proposed workflow relies on detecting and matching keypoints and feature keypoints to establish sufficient similarities for reliable point cloud registration, it was necessary to extract different numbers of keypoints and feature keypoints for the source and target point clouds, depending on each point cloud's size. Based on empirical optimization and recognizing that processing time is an important factor for on-site use of MR applications, several tests were performed, extracting different numbers of Keypoints and feature keypoints. For each tile from the source point clouds, 1000, 3000, 5000, 10000, and 15000 keypoitns and features were extracted and processed for registration against 15000, 20000, 25000, 30000, 45000, 70000, 90000, and 270000 keypoints and feature keypoints from target point clouds. A total of 160 registration attempts were carried out for indoor data and 80 for outdoor areas.

The presented approach has the advantage that the most intensive processing task can be performed asynchronously, since the pre-scanned large target point cloud information can already be processed for keypoint and feature extraction and stored in the database, waiting for the on-site surveyed source point cloud to perform the registration. Despite this, it is necessary to consider that a larger number of keypoints and features will, in any case, affect overall processing time, as outlier-mismatch filtering will also increase it.

4. Results

This study evaluates learning-based methods for 3D-to-3D cross-source point cloud registration, with a focus on inferring transformation parameters for MR scene alignment and camera localization. The existing workflow, VRHCF, was assessed for its generalization capabilities using real-world cross-source point cloud data. We propose replacing keypoint and feature matching, along with outlier filtering, in HCF with the TEASER++ method, a certifiable, robust, graph-based mathematical approach, to address large-scale indoor and outdoor scenarios.

The quantitative analysis demonstrated that the proposed method accurately determines transformation parameters (translation and rotation) for MR devices in scenarios characterized by large size asymmetry between source and target point cloud data. The method achieved low translation and rotation errors of centimeter level and variate processing times. Figure 4

In indoor environments, 160 point cloud registration attempts were performed. Of these, 23.75% were successfully registered, meeting the thresholds of $\leq$ 0.07 m for translation error and $\leq$ 2.4° m for rotation error. An additional 8.75% achieved registration within a translation error range of 0.09–0.38 m and a rotation error of 11.02°. The lowest registration time was 1 minute and 52 seconds, yielding a translation error of 0.0694 m and a rotation error of 1.3191°.

Table 1 summarizes the optimal results, with translation error, rotation error, and processing time serving as the primary evaluation criteria. Not all analyzed source datasets yielded successful registrations within the specified thresholds. In the indoor scenario, tiles 0000 and 0002 failed to register, despite the quantity of extracted keypoints and features meeting the established thresholds. In the outdoor environment, seven successful registrations were achieved for source point cloud tile 1; the most accurate attempt exhibited a translation error of 0.067 m and a rotation error of 0.3067°, with a processing time of 2 minutes and 37 seconds. Conversely, outdoor source point cloud tile 2 failed to register.

5. Discussion

Cross-source point cloud registration as a method for deriving AR/MR camera localization and scene alignment shows promising potential but remains a nascent area of research due to unresolved challenges. The use of specialized learning-based methods to handle heterogeneous point cloud data, such as VRHCF and its proposed combination with TEASER++, offers encouraging preliminary insights into their application in real-world scenarios.

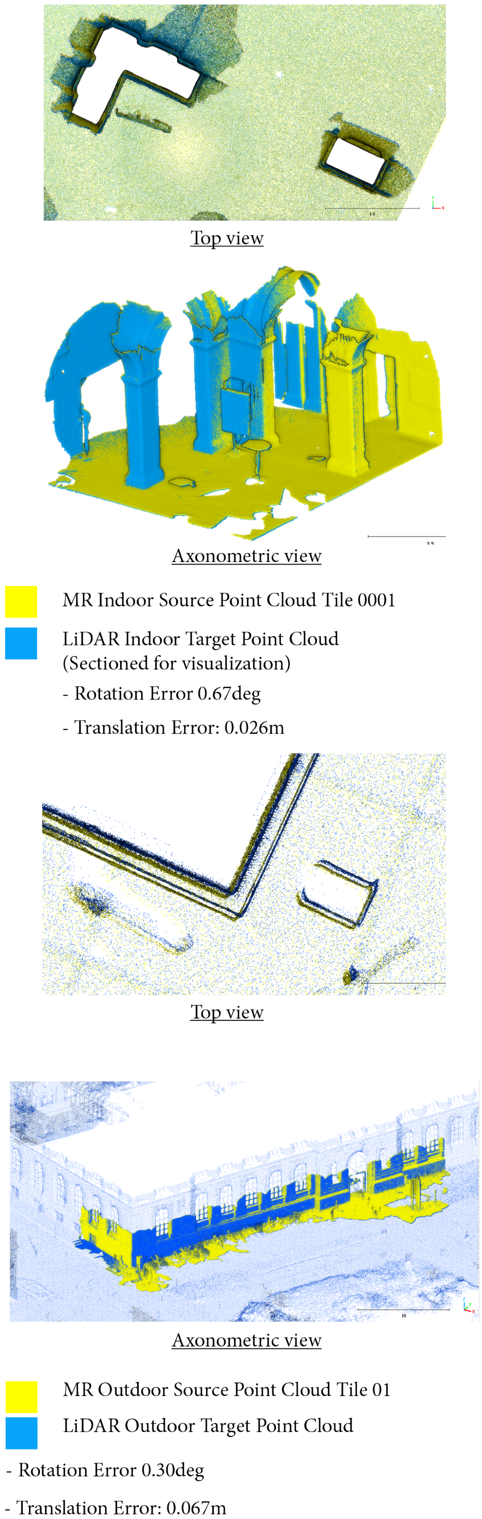

Figure 4. Cross-source Point cloud registration results. Independently from the starting location at the 3D shared space of the indoors and outdoors MR to LiDAR datasets, both achieved a successful coarse-registration. On top (indoors dataset) and down (first outdoors dataset) source (yellow) and target (blue) point cloud achieve successful coarse registration results.Further refinement could be applied for fine registration.

Table 1. Point cloud registration performance metrics categorized by error and processing time.

Source Point Cloud Tile	Source Keypoints & Features	Target Keypoints & Features	Rotation Error (degrees)	Translation Error (m)	Processing Time (min:sec)
Sorted by translation error (m)					
0003	**10000**	**270000**	**0.1453**	**0.0053**	**25:17**
0003	5000	270000	0.4685	0.0104	13:25
0003	10000	30000	0.6818	0.0114	11:25
0001	**1000**	**270000**	**0.4835**	**0.0196**	**02:46**
0001	3000	90000	0.4270	0.0212	04:30
Sorted by rotation error (degrees)					
0003	**10000**	**270000**	**0.1453**	0.0053	**25:17**
0003	15000	270000	0.2794	0.0158	35:53
0001	10000	90000	0.3208	0.0193	14:18
0003	10000	90000	0.3362	0.0327	14:38
0001	15000	270000	0.3851	0.0206	38:29
Sorted by overall processing time (min:sec)					
0003	**1000**	**20000**	**6.4049**	**0.2389**	**01:29**
0001	**1000**	**90000**	**1.3191**	**0.0694**	**01:52**
0003	1000	45000	0.8473	0.1045	01:38
0001	1000	270000	0.4835	0.0196	02:46
0001	3000	90000	0.4270	0.0212	04:30

However, further research is required to tackle the open challenges. While the proposed method requires successive approximations to determine the optimal number of keypoints and feature descriptors necessary for effective cross-source point cloud registration, it successfully inferred camera transformation parameters for indoor environments by registering MR source point clouds, which are approximately 98% smaller than the LiDAR target point cloud. State-of-the-art methods such as Radanovic et al. (2023) achieve registration in scenarios with an MR source point cloud approximately 88% smaller than the LiDAR target point cloud.

Despite the results suggesting an improvement in the asymmetry between the source and target point clouds for registration, the proposed method finds an obstacle at the required processing time. As shown in Table 1, the most accurate results for translation and rotation parameters require processing times of minutes to tens of minutes, which is a limitation, especially for on-site applications such as MR. Further research in this area should focus on feasibility analysis or strategy development to optimize processing times.

An important factor in the success or failure of 3D-to-3D-based cross-source point cloud registration is how descriptive the geometry captured by the source point cloud sensors is, and how it differentiates itself from its surroundings. Despite covering a similar area to the rest of the indoor dataset tiles, tiles 0000 and 0002 failed the CS point cloud registration due to their similarity to the surroundings. Description ambiguity, also known as perceptual aliasing, in point clouds can lead to false-positive matches. However, the remaining indoor tiles succeed in their registration attempts, with a mean success rate of 62.%. On the other hand, the low registration success rate in outdoor scenarios is attributed to the prevalence of data noise. Despite achieving some accurate results in registration, most attempts failed. Further experimentation is required to better assess the method’s use and feasibility in outdoor settings, such as for MR camera localization in urban scenarios.

Although preliminary results indicate potential for successful registration in cases with significant size discrepancies in point clouds using cross-source point cloud registration methods like the proposed VRHCF + TEASER++, the method may be counterproductive, as it continues to search for keypoints and feature descriptors based on local patches. This could complicate matters in the presence of large numbers of outliers or large volumes of data, limiting its scalability. A potential avenue for future research is to validate methods that learn larger-area descriptors to delimit smaller search keypoint and feature areas to support optimization at processing time.

6. Conclusions

In this study, we investigated the efficacy of cross-source point cloud registration as a key approach for AR/MR camera localization and scene alignment, particularly in practical, real-world settings. We used the Voxel Representation and Hierarchical Correspondence Filtering (VRHCF) method alongside the TEASER++ algorithm to address challenges posed by large, asymmetrically sized, heterogeneous point cloud data from various sensors. The results indicate that this combined approach could enhance 3D-to-3D MR scene alignment, especially within large indoor environments. The success rate indoors underscores the effectiveness of our methodology for overcoming challenges related to size discrepancies and outlier interference; however, it also highlights the need to optimize processing times. Although it has promising results, its use in large outdoor areas and urban environments requires accounting for the inherent noise in outdoor data. This research aims to support the advancement towards seamless AR/MR scene alignment through innovative cross-source point cloud registration methods. As the field evolves, these methods have the potential to transform AR/MR applications, improving accessibility, precision, and efficiency when integrating the digital and physical worlds.

Acknowledgment

The project on which this report is based was funded by the Federal Ministry of Research, Technology and Space (BMFTR) under grant number 03FHP206.The responsibility for the content of this publication lies with the authors.

References

Ao, S., Hu, Q., Yang, B., Markham, A., Guo, Y., 2021. SpinNet: Learning a General Surface Descriptor for 3D Point Cloud Registration. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Nashville, TN, USA, 11748–11757.

Aoki, Y., Goforth, H., Srivatsan, R. A., Lucey, S., 2019. PointNetLK: Robust & Efficient Point Cloud Registration using PointNet. arXiv:1903.05711 [cs].

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features. A. Leonardis, H. Bischof, A. Pinz (eds), *Computer Vision – ECCV 2006*, 3951, Springer Berlin Heidelberg, Berlin, Heidelberg, 404–417. Series Title: Lecture Notes in Computer Science.

Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256. http://ieeexplore.ieee.org/document/121791/.

Choy, C., Park, J., Koltun, V., 2019. Fully Convolutional Geometric Features. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Seoul, Korea (South), 8957–8965.

El Barhoumi, N., Hajji, R., Bouali, Z., Ben Brahim, Y., Kharroubi, A., 2022. Assessment of 3D Models Placement Methods in Augmented Reality. *Applied Sciences*, 12(20), 10620. https://www.mdpi.com/2076-3417/12/20/10620.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus. 24(6).

GroB, J., Osep, A., Leibe, B., 2019. AlignNet-3D: Fast Point Cloud Registration of Partially Observed Objects. *2019 International Conference on 3D Vision (3DV)*, IEEE, Québec City, QC, Canada, 623–632.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2020. Deep Learning for 3D Point Clouds: A Survey. arXiv:1912.12033 [cs].

Haleem, A., Javaid, M., Singh, R. P., Rab, S., Suman, R., Kumar, L., Khan, I. H., 2022. Exploring the potential of 3D scanning in Industry 4.0: An overview. *International Journal of Cognitive Computing in Engineering*, 3, 161–171. https://linkinghub.elsevier.com/retrieve/pii/S2666307422000171.

Haneefa, F. M., Shoufan, A., Damiani, E., 2024. The Essentials: A Comprehensive Survey to Get Started in Augmented Reality. *IEEE Access*, 12, 109012–109070. https://ieeexplore.ieee.org/document/10623629/.

Herbers, P., König, M., 2019. Indoor Localization for Augmented Reality Devices Using BIM, Point Clouds, and Template Matching. *Applied Sciences*, 9(20), 4260. https://www.mdpi.com/2076-3417/9/20/4260.

Huang, K., Wang, C., Shi, W., 2023a. Accurate and Robust Rotation-Invariant Estimation for High-Precision Outdoor AR Geo-Registration. *Remote Sensing*, 15(15), 3709. https://www.mdpi.com/2072-4292/15/15/3709.

Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K., 2021. PREDATOR: Registration of 3D Point Clouds with Low Overlap. arXiv:2011.13005 [cs].

Huang, X., Mei, G., Zhang, J., 2023b. Cross-source point cloud registration: Challenges, progress and prospects. *Neurocomputing*, 548, 126383. https://linkinghub.elsevier.com/retrieve/pii/S0925231223005064.

Kern, F., Tschanter, J., Latoschik, M. E., 2023. Virtual-to-Physical Surface Alignment and Refinement Techniques for Handwriting, Sketching, and Selection in XR. *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, Shanghai, China, 502–506.

Kharroubi, A., Billen, R., Poux, F., 2020. Marker-less mobile augmented reality application for massive 3d point clouds and semantics. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 255–261. https://isprs-archives.copernicus.org/articles/XLIII-B2-2020/255/2020/.

Kunz, A., Rosmann, S., Loria, E., Pirker, J., 2022. The Potential of Augmented Reality for Digital Twins: A Literature Review. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, IEEE, Christchurch, New Zealand, 389–398.

Lee, E., Kwon, Y., Kim, C., Choi, W., Sohn, H.-G., 2024. Multi-source point cloud registration for urban areas using a coarse-to-fine approach. *GIScience & Remote Sensing*, 61(1), 2341557. https://www.tandfonline.com/doi/full/10.1080/15481603.2024.2341557.

Li, J., Wang, C., Kang, X., Zhao, Q., 2020. Camera localization for augmented reality and indoor positioning: a vision-based 3D feature database approach. *International Journal of Digital Earth*, 13(6), 727–741. https://www.tandfonline.com/doi/full/10.1080/17538947.2018.1564379.

Long Quan, Zhongdan Lan, 1999. Linear N-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8), 774–780. http://ieeexplore.ieee.org/document/784291/.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. http://link.springer.com/10.1023/B:VISI.0000029664.99615.94.

Lowry, S., Sünderhauf, N., Newman, P., Leonard, J. J., Cox, D., Corke, P., Milford, M. J., 2016. Visual Place Recognition: A Survey. *IEEE Transactions on Robotics*, 32(1), 1–19. https://ieeexplore.ieee.org/document/7339473/.

Lyu, M., Yang, J., Qi, Z., Xu, R., Liu, J., 2024. Rigid pairwise 3D point cloud registration: A survey. *Pattern Recognition*, 151, 110408. https://linkinghub.elsevier.com/retrieve/pii/S0031320324001596.

Ma, N., Wang, M., Han, Y., Liu, Y.-J., 2024. FF-LOGO: Cross-Modality Point Cloud Registration with Feature Filtering and Local to Global Optimization. arXiv:2309.08966 [cs].

Marchand, E., Uchiyama, H., Spindler, F., 2016. Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, 22(12), 2633–2651. http://ieeexplore.ieee.org/document/7368948/.

Messi, L., Spegni, F., Vaccarini, M., Corneli, A., Binni, L., 2024. A First Evaluation of the Seamless Markerless Augmented Reality Registration System Supporting Facility Management.

Osipov, A., Ostanin, M., Klimchik, A., 2023. Comparison of Point Cloud Registration Algorithms for Mixed-Reality Cross-Device Global Localization. *Information*, 14(3), 149. https://www.mdpi.com/2078-2489/14/3/149.

Poiesi, F., Boscaini, D., 2022. Learning general and distinctive 3D local deep descriptors for point cloud registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://ieeexplore.ieee.org/document/9775606/.

Radanovic, M., Khoshelham, K., Fraser, C., 2023. Aligning the real and the virtual world: Mixed reality localisation using learning-based 3D–3D model registration. *Advanced Engineering Informatics*, 56, 101960. https://linkinghub.elsevier.com/retrieve/pii/S1474034623000885.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, IEEE, Barcelona, Spain, 2564–2571.

Sardi-Barzallo, J., Coors, V., 2025. From Mixed Reality to BIM: A Workflow for the Integration of Immersive Technologies at Architectural Design Process. *Proceedings of the 30th International Conference on 3D Web Technology*, ACM, Siena Italy, 1–9.

Tao, F., Qi, Q., Wang, L., Nee, A., 2019. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. *Engineering*, 5(4), 653–661. https://linkinghub.elsevier.com/retrieve/pii/S209580991830612X.

Teruggi, S., Fassi, F., 2022. Hololens 2 Spatial Mapping Capabilities in Vast Monumental Heritage Environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022, 489–496. https://isprs-archives.copernicus.org/articles/XLVI-2-W1-2022/489/2022/.

Teuscher, B., Walther, P., Wang, J., Werner, M., 2025. A Taxonomy of Point Cloud Search. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W6-2025, 185–192. https://isprs-annals.copernicus.org/articles/X-4-W6-2025/185/2025/.

Wang, F., Zhao, Z., 2017. A survey of iterative closest point algorithm. *2017 Chinese Automation Congress (CAC)*, IEEE, Jinan, 4395–4399.

Wei, Y., Yu, X., Song, R., Peng, H., Li, A., 2024. A Comprehensive Survey and Taxonomy on Point Cloud Registration Based on Deep Learning. *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, Jeju, South Korea, 8295–8303.

Xie, X., Zou, D., 2024. Depth-Based Efficient PnP: A Rapid and Accurate Method for Camera Pose Estimation. *IEEE Robotics and Automation Letters*, 9(11), 9287–9294. https://ieeexplore.ieee.org/document/10621448/.

Xu, J., Zhang, Y., Zou, Y., Liu, P. X., 2023. Point Cloud Registration With Zero Overlap Rate and Negative Overlap Rate. *IEEE Robotics and Automation Letters*, 8(10), 6643–6650. https://ieeexplore.ieee.org/document/10238787/.

Xu, M., Zhong, X., Zhong, R., 2025. A Multisource Heterogeneous Point Cloud Fine Registration Method for Large-Scale Outdoor Scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–23. https://ieeexplore.ieee.org/document/10965821/.

Yang, H., Shi, J., Carlone, L., 2020. TEASER: Fast and Certifiable Point Cloud Registration. arXiv:2001.07715 [cs].

Zhang, Y.-X., Gui, J., Yu, B., Cong, X., Gong, X., Tao, W., Tao, D., 2025. Deep Learning-Based Point Cloud Registration: A Comprehensive Survey and Taxonomy. arXiv:2404.13830 [cs].

Zhao, G., Ao, S., Zhang, Y., Xu, K., Guo, Y., 2025. Progressive Correspondence Regenerator for Robust 3D Registration. arXiv:2502.02163 [cs].

Zhao, G., Du, Z., Guo, Z., Ma, H., 2024a. VRHCF: Cross-Source Point Cloud Registration via Voxel Representation and Hierarchical Correspondence Filtering. arXiv:2403.10085 [cs].

Zhao, G., Guo, Z., Du, Z., Ma, H., 2024b. Cross-PCR: A Robust Cross-Source Point Cloud Registration Framework.

Zhou, K., Taigang Liu, Lifeng Zhou, 2015. Industry 4.0: Towards future industrial opportunities and challenges. *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, IEEE, Zhangjiajie, China, 2147–2152.

Zhou, Z., Karlekar, J., Hii, D., Schneider, M., Lu, W., Wittkopf, S., 2009. Robust Pose Estimation for Outdoor Mixed Reality with Sensor Fusion. C. Stephanidis (ed.), *Universal Access in Human-Computer Interaction. Applications and Services*, 5616, Springer Berlin Heidelberg, Berlin, Heidelberg, 281–289. Series Title: Lecture Notes in Computer Science.