SLAM methods for Augmented Reality systems for flight simulators*

Onyeka J. Nwobodo[1[0000−0002−2706−6830]], Kamil Wereszczyński[1[0000−0003−1686−472X]], and Krzysztof Cyran[1[0000−0003−1789−4939]]

Department of Computer Graphics, Vision and Digital Systems,Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland
{onyeka.nwobodo, kamil.wereszczynski, krzysztof.cyran}@polsl.pl phone:

Abstract. In this paper, we present the review and practical evaluation of the flight simulators of Simultaneous Localization and Mapping methods. We present a review of recent research and development in the SLAM application in a wide range of domains, like autonomous driving, robotics and augmented reality (AR). Then we focus on the methods selected from the perspective of their usefulness in the AR systems for training and servicing the flight simulators. The localization and mapping in such an environment are much more complex than in others since the flight simulator is relatively small and close area. Our previous experiments showed that the built-in SLAM system in HoloLens is insufficient for such areas and has to be enhanced with additional elements, like QR codes. Therefore, the presented study on other methods can improve the localization and mapping of AR systems in flight simulators.

Keywords: SLAM, flight simulators, Deep Learning, LiDAR, Augmented Reality

Introduction

Simultaneous localization and mapping(SLAM) have attracted much interest recently, particularly in intelligent systems and robotics. There are problems with implementing SLAM due to its complexity. This problem has existed for over 30 years, especially when finding a solution[1]. However, various approximations have come close to resolving this challenging algorithmic problem following decades of mathematical and computational effort[2]. Solving the SLAM problem will enable a wide range of potential applications for autonomous robots[3]. A robot that can navigate its environment without human intervention is said

* The authors would like to acknowledge that this paper has been written based on the results achieved within the WrightBroS project. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No 822483. Disclaimer: The paper reflects only the author's view, and the Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

to be autonomous. For a robot to successfully navigate, it must have a thorough understanding of its surroundings and be able to track its position in that environment consistently and accurately. Scientists employ various methods to enhance the autonomy and self-discovery of robot navigation, leading to the development of augmented reality(AR) systems. The state-of-the-art SLAM technology uses multiple sensors, such as LIDAR, cameras, and IMU, to create highly accurate and detailed maps.

Simultaneous localization and mapping(SLAM) is a method robotics and computer vision professionals use to locate a robot while simultaneously creating a map of its surroundings. SLAM aims to estimate the agent's location and the environment's structure using sensor data such as lidar, cameras, GNSS receivers and antennas, an Inertial Measurement Unit(IMU), and odometry. It is known as "visual SLAM" when the SLAM algorithm is based on camera sensors, and "LiDAR SLAM" is based on laser scanners[4].
Microsoft HoloLens is a head-mounted device that projects augmented reality(AR) into the users' field of view. It is a direct-based feature with built-in visualization and no camera trajectory module. The HoloLens is an optical see-through gadget with numerous sensors, including an accelerometer, magnetometer, and gyroscope. Numerous good SLAM techniques have greatly aided the development of SLAM technology, including MonoSLAM[5], DTAM[6], LSD SLAM[7], RATSLAM[8], KinetFusion[9], RGB-D SLAM[10]. DeepVo[11].Various SLAM methods target different objectives; Microsoft HoloLens uses a variant of the SLAM technology to enable its augmented reality capabilities. Therefore, in this work, we discuss different SLAM methods for Microsoft HoloLens and flight simulators, furthermore compare these methods and the deep learning SLAM methodology in a flight simulator environment.

The SLAM problem has numerous current solutions, which can be categorized as filter-based and global optimization approaches. The filter-based approach is the classical approach that recursively performs prediction and update steps. They are typically thought of as a maximum posterior(MAP) method where the robot's previous distribution is estimated using data from sensors like the IMU. The likelihood distribution is constructed by combining the IMU data with those made by a camera or LiDAR[12]. The IMU sensor measurements utilize the prediction step of a standard SLAM filtering technique to forecast the vehicle's motion(odometry)[13]. At the same time, the estimated camera posture and measured image attributes are employed as a likelihood distribution to update the predictions in the update step[12]. This includes all the Kalman filter families (EKF, UKF and SEIF) and particle filters. The global optimization method relies on keeping a few keyframes in the environment and estimates the motion through bundle adjustment(BA). Kummer et al.[14] estimate the robot trajectory like the SLAM graph by processing all sensor measurements. They typically use least-squares adjustment methods and optimization, considered more accurate than the filtering process[15]. This is currently a popular approach for vision-based SLAM such as ORB-SLAM. Following this, other researchers have attempted to increase the effectiveness of SLAM systems by utilizing new fea-

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

ture extraction and matching algorithms, including orientated FAST and rotated BRIEF(ORB)[16] and Speeded-Up Robust Features[17].

SLAM has various applications in decision-making processes, including autonomous driving (on land, air, sea, and underwater), robotics, and augmented reality.
SLAM-based Augmented Reality(AR) is a technology that mixes real-world imagery and computer-generated graphics to improve user experience[18]. SLAM-based AR uses cameras and sensors to map the surroundings in real-time and precisely put virtual items in the real world. This enhances the stability and precision of virtual objects in the real environment and makes for a more immersive AR experience. To provide the appearance that the virtual and real worlds are completely integrated, displaying AR items in the proper place and adhering to the user's perspective demands a solution to numerous static and dynamic registration issues[19]. Recently, scientists have employed SLAM's accuracy and real-time performance for virtual registration in AR[20]. Regarding direct approaches[7], a camera-based method, LSD-SLAM, enables the construction of large-scale, semi-dense maps that do not require adjusting bundles of features. ORB SLAM is another approach to visual SLAM proposed by[21] that uses a feature-based approach to simultaneously estimate the camera pose and build a map of the environment. To partially resolve the scale ambiguity and provide motion cues without the use of visual features, Visual Inertial SLAM (VISLAM) can combine VSLAM with Inertial Measurement Unit(IMU) sensors(accelerometer, gyroscope, and magnetometer)[22]. The HoloLens device's ability to localize has been demonstrated by a recent study on the technology's position tracking[23]. According to [24], with the help of Microsoft's newly released spatial mapping capability, the HoloLens can map scenes in its immediate surroundings. Environmental objects such as walls, floors, and obstacles can be found using spatial mapping.

SLAM on Microsoft HoloLens is used to build 3D maps of an environment and track the device's position within it in real-time. It uses inbuilt cameras and sensors to detect and track distinctive environmental features and estimate the device's pose. The HoloLens device combines RGB cameras, depth sensors, and IMUs(Inertial Measurement Units) that can be used for SLAM. The device can use visual SLAM, visual-inertial SLAM, or RGB-D SLAM to estimate the camera pose and build an environment map. Using SLAM, the HoloLens device can provide a more accurate and stable AR experience, as it can track the user's movements and the position of objects in the environment. More recent work at Microsoft has concentrated on large-scale scene reconstruction utilizing voxel hashing[25] and RGB-D camera re-localization[26] for recovering from tracking failures. In the area of SLAM methods in flight simulators, Skurowski et al.[27] uses the QR codes for marking the crucial points in the simulator. We made this study to avoid including the additional and non-natural elements of the flight simulator cockpit.

The following is this paper's main contributions:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

1. The analysis is completed by running/discussing five selected states of the SLAM methods, which have been chosen to represent the diversity of the existing SLAM methods on Microsoft HoloLens and flight simulator,
2. The accuracy of the existing SLAM methods on Microsoft HoloLens is compared with the presented SLAM and deep learning methods in a flight simulator environment.

In numerous studies, SLAM techniques are reviewed and contrasted. Depending on the robot's goals, they provide information on the dependability, real-time estimation, and accurate depiction of the surroundings, as well as the location and orientation of the robot. As reviewed and compared by the following authors[28–32]. It is discovered, nonetheless, that none of the articles compares SLAM techniques using Microsoft HoloLens in a flight simulator environment. The primary **motivation** supporting this research is to evaluate the Microsoft HoloLens SLAM techniques and compare them to the chosen SLAM and deep learning techniques in flight simulator scenarios. The outcome will make selecting the SLAM methods most appropriate to HoloLens in a flight simulator environment easier for accurate, reliable, and long-term motion estimation to render smooth and stable holograms.

1 Materials and Methods

RGB-D SLAM: It is a system that creates a real-time 3D model of the surroundings by fusing RGB and D(depth) information. The additional depth information aids in overcoming some of the difficulties associated with using standard RGB-only SLAM techniques, such as coping with repeated or texture-less situations. The system uses an RGB-D camera, which records information about each pixel's color and depth, and algorithms to predict the camera's position and create a 3D environment model as the camera moves[33]. Modern RGB-D SLAM systems align point features whose spatial coordinates are determined by corresponding sensor depth data using the Iterative Closest Point (ICP) technique. However, because visual features typically lie at the edges of actual objects, noise frequently contaminates the depth measurements of features. Utilizing the benefits of the RGB-D camera's depth image is one technique to deal with this problem. For instance, planes can be inferred from the depth information, and the locations on these planes have less noise than the corners. Instead of using point characteristics as primitives, researchers have proposed using planes. Towards that end, Gao and Zhang[34] provide a technique that extracts and chooses reliable depth values or planer point features. Dai et al.[35] use Delaunay triangulation, in which changes in the triangle edges in adjacent frames are compared to assess the correlation of feature points and differentiate between dynamic from static map points. To create a realistic and immersive experience in a flight simulator environment, the RGB-D SLAM approach employing depth data from an RGB-D camera for HoloLens can be a solid choice.

Advantages:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

1. Improved spatial awareness: The HoloLens cameras' RGB and depth data can create a more realistic depiction of the surroundings, giving users enhanced depth perception and spatial awareness.
2. Real-time mapping: The HoloLens' ability to update its map of the surroundings in real-time thanks to RGB-D SLAM makes it perfect for usage in dynamic simulation scenarios like flight simulators.
3. Increased interaction: The HoloLens can offer customers a more interactive and immersive experience in the flight simulator environment with an accurate representation of the environment.

Limitations:

1. High computational requirements: High computing power is needed to produce an accurate map in real-time using the RGB-D SLAM algorithm. This presents a challenge for processing-constrained devices like the HoloLens.
2. Sensitivity to lighting conditions: RGB-D SLAM methods depend on precise depth data, which is susceptible to variations in lighting. The maps that the HoloLens produces may need to be more accurate.
3. Occlusions: It can be challenging for the HoloLens to effectively acquire depth information when environmental objects overlap in certain circumstances. The maps produced by the RGB-D SLAM algorithm may need to be revised.

ORB-SLAM(oriented Fast Rotation Brief SLAM): One of the latest monocular vision-based SLAM techniques with an open-source implementation is ORB-SLAM[36] using a real-time SLAM library called ORB-SLAM, monocular, stereo, and RGB-D cameras can calculate their camera trajectories and sparse 3D reconstructions. It uses ORB Rotated BRIEF(Binary Robust Independent Elementary Features) and Oriented FAST(Accelerated Segment Test) feature detectors developed by[16]. According to Mur-Artal et al.[21], this technique estimates position and maps from an image sequence in real-time. The primary ORB-SLAM process creates an environmental map comprising keyframes and map points. Each keyframe stores a list of 2D features and their location in ORB-SLAM coordinates. Tracking, local mapping, and loop closure are the three parallel threads that comprise the ORB-SLAM process.1)Using motion-only BA to reduce re-projection error and camera localization with each frame by finding matching features on the local map. 2)The local mapping to maintain control over, enhance local mapping and perform local BA. 3)Loop closure uses position graph optimization to find large loops and correct accumulated drift. After optimizing the position graph, this thread triggers a fourth thread to perform a full BA to determine the best solution for structure and motion. To address the issue of the low number of feature point extraction and the simple keyframe loss, Cai et al.[37] proposed an enhanced visual SLAM based on affine transformation for ORB feature extraction. However, the environments used in these methods are frequently static, which is insufficient to meet the demands of the complex task in a dynamic environment. Various approaches address the dynamic problem[38, 39].
Advantages:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

1. Real-time performance: In a flight simulator scenario where quick judgments are necessary, ORB-SLAM can give real-time mapping and localization of the environment, which is essential.
2. Robustness: ORB-SLAM excels in a flight simulator environment because it is highly resilient to changes in illumination conditions and can endure abrupt motions and shocks.
3. Open-source: Open-source platform ORB SLAM is easily customized and adaptable to fulfill unique requirements.

Limitations

1. Dependence on keypoints: ORB-SLAM relies on locating and following the keypoints in the image, which things like occlusions and changes in lighting can impact. The effectiveness of ORB-SLAM may be hampered if keypoints are not precisely identified or tracked.
2. Limited map-building capabilities: Real-time tracking is the purpose of ORB-SLAM. In comparison to other SLAM algorithms that place a higher priority on creating a comprehensive map. As a result, its map-building capabilities may need to be revised.
3. High computational requirements: In real-time applications, especially in a resource-restricted setting like a HoloLens, ORB SLAM takes a lot of processing resources, which can be difficult.

LSD-SLAM(Large scale direct monocular SLAM): It is a direct monocular SLAM method that tracks and maps objects directly using picture intensities rather than key points. Direct image alignment is used to follow the camera, and semi-dense depth maps created by filtering through several pixelwise stereo comparisons are used to estimate geometry[7]. Then, using a Sim (3) pose-graph of keyframes creates large-scale maps with scale-drift corrections, including loop closures. LSD-SLAM is a real-time executable on CPUs and even on contemporary smartphones. Traditional monocular SLAM techniques have limitations in accuracy, scalability, and performance in vast and dynamic situations. LSD-SLAM was created to address these issues. There are two primary phases to this method: 1)Keyframe selection: Keyframes are chosen based on the visualization information of the incoming image changes, and 2)Direct Image Alignment: The best transformation(i.e. rotation and translation) that aligns the two images is found by directly comparing the current image to the keyframes in the map. Forster et al.[40] suggested SVO(semi-direct visual odometry), a visual odometry(VO) without loop closure detection and re-localization, as a method for measuring motion. SVO follows the features from accelerated segment test(FAST) feature points and surrounding pixels by minimizing the photometric error to determine the camera's motion. Bergmann et al.[41] proposed an online photometric calibration that dynamically estimates the photometric parameters by solving the least squares equation of the feature tracker and modifies the exposure situation of the input sequence to improve the performance of direct visual odometry. It marks a significant advancement in direct formulation placement and mapping precision. Peixin Liu et al.[42] improved the visual slam technique based

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

on the sparse direct method, in which input sequences' silhouette and response functions were optimized based on the camera's photometric configuration. Optimization equations were developed by tracking the input sequence's Shi-Tomasi corners with sparse direct visual odometry (VO) pixel tracking. Additionally, the joint optimization equation was solved using the Levenberg-Marquardt(L-M) method, and the photometric calibration parameters in the VO were updated to realize real-time dynamic compensation of the input sequences' exposure. This reduced the impact of light variations on SLAM's accuracy and robustness. A Shi-Tomasi corner-filtered strategy further reduced the computational complexity of the suggested technique, and the loop closure detection was realized using the orientated FAST and rotated BRIEF(ORB) features.
Advantages:

1. Real-time performance: LSD-SLAM is appropriate for usage in flight simulators because it is made to work in real-time.
2. High accuracy: Regarding motion tracking, LSD-LAM is a dependable source because it can precisely follow the simulator's movement.
3. Low computational power requirement: LSD-SLAM is appropriate for low-end computer systems because it does not require much computational resources.

Limitations:

1. Difficulty handling dynamic environments: LSD-SLAM can struggle with dynamic environments, where objects rapidly move or change in real-time.
2. Sensitivity to lighting conditions: LSD-SLAM relies heavily on visual information, making it susceptible to lighting conditions and camera noise.
3. Need for good initialization: LSD-SLAM requires an accurate initial pose of the simulator, which can be challenging to achieve in specific flight scenarios.
4. Limited robustness: LSD-SLAM can be affected by insufficient or inaccurate data, leading to incorrect results or lost track.

VINS(Virtual Inertial SLAM): Virtual inertial SLAM is a hybrid technique that employs data from visual and inertial sensors, such as gyroscopes and accelerometers, to map the surroundings and estimate the camera's pose. The technique tracks the camera's movements and calculates its position in the surroundings using the visual elements seen in photos. The inertial measurements help increase the posture estimate's precision by supplying information about recent motion. The map is updated when new features are added, and old features are removed as the camera moves through the surroundings. Several benefits over conventional monocular or stereo-visual SLAM techniques are provided by utilizing visual and inertial information in VINS algorithms. For instance, VINS algorithms can accurately estimate the camera's motion even without visual input, which is frequently the case in chaotic or obscured surroundings. Additionally, VINS algorithms are suited for use in applications that call for long-term mapping or tracking because they may offer drift-free estimations of the camera's position and orientation over time. According to their method[43], inertial

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

measurements should be tightly integrated into a keyframe-based visual SLAM structure; in the visual SLAM system, both the re-projection error terms and the cost function IMU are optimized. The technique to integrate visual information and IMU can be classified into loosely coupled[44] and tightly coupled[43] techniques. This can be achieved by evaluating whether the state vector includes the visual information. The vision-based SLAM and the inertia-based modules are often executed separately as part of the loosely connected techniques. Additionally, the measures are estimated using the combination of the results. The loosely connected approach still suffers from the monocular SLAM drift issue. In a tightly coupled technique, the depth information from the monocular SLAM can be calculated, and the IMU deviation can be fixed by adding visual information to the state vector. This approach is more reliable and accurate than the traditional single-vision-sensor-based SLAM. Yin et al.[45] present a stereo visual-inertial SLAM system, using a technique to detect dynamic features that loosely coupled the stereo scene flow with an inertial measurement unit(IMU) and tightly coupled the dynamic and static characteristics with the IMU measurements for nonlinear optimization.
The Microsoft HoloLens SLAM algorithm uses inertial measurements from the HoloLens' IMU and high-quality visual information from the virtual environment to produce more precise and stable estimates of the device's position and orientation in a flight simulator environment than using just one of these sources alone.

Advantages:

1. High precision: Visual and inertial sensors fused can accurately estimate location and orientation.
2. Robustness: The system can operate effectively in low-light conditions or without lighting. It can still deliver accurate position estimates when a temporary loss of visual data occurs.
3. Performance in real-time: The system's ability to perform real-time estimation, allowing for smooth and responsive navigation in the virtual environment, suits it for robotics, augmented reality, and autonomous vehicle applications.
4. Autonomy: The device is equipped with Visual Inertial SLAM, which permits independent operation without external tracking systems.

Limitations:

1. Sensitivity to starting conditions: The system's accuracy is strongly influenced by the initial conditions, which might impact the system's operation if not properly calibrated.
2. Cost of computation: The system is challenging to implement on low-power devices due to the high computational resource requirements.
3. Drift over time: The system's mistakes may build up, causing drift in the estimated position and orientation.

Deep learning SLAM methods

Deep learning has shown increasingly clear advantages in image processing, especially in data fitting, feature extraction, and spatial transformation, which

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

has produced exciting results. The traditional manual algorithm compared with studies on recognition[46], image segmentation[47], object detection[48] and image classification[49] all of which showed the algorithm to be significantly better. Slam's deep learning implementation can circumvent the limitations imposed on visual odometry and scene recognition by manually creating features. Additionally, it helps build high-level semantics and develop the agent's knowledge base, improving the agent's perceptual and cognitive abilities. The model has undergone data-driven training to make it more consistent with people's interactions and settings. The evolution of visual SLAM from geometry-based methods to deep learning methods occurs. Recently, visual SLAM issues, including visual odometry[50] and loop closure[51], have been addressed using both supervised deep learning and unsupervised approaches. Due to these recent developments, deep learning techniques have a significant potential to address the complex problems of visual SLAM by incorporating adaptive and learning capabilities. Using an unsupervised end-to-end learning framework, Geng et al.[52] propose an ORB-SLAM method that better estimates camera motion and monocular depth. Liet al.[53] employed SegNet, a well-known semantic segmentation network, to segment the images and further separate dynamic objects. Zhang et al.[54] use YOLO[55] to recognize objects in the environment, and to increase the system's accuracy, they used a semantic map to filter dynamic feature points. The scientists' current work suggests fusing SLAM and deep learning to speed up the various visual SLAM system elements. Tateno et al.[56] describes a technique for accurate and dense monocular reconstruction where CNN-predicted dense depth maps from a deep neural network combine depth measurements with direct monocular SLAM. Li et al.[57], using UnDeepVO perform unsupervised training on stereo images, then utilize monocular images to estimate pose and create maps. Vijayanarasimhan et al.[58] present SfM-Net, a neural network that trains the image generated from the geometry to extract 3D structure, ego-motion, segmentation, object rotations, and translations in videos.

Advantages:

1. Fusion of SLAM with Deep Learning: One method integrates deep learning-based techniques with conventional SLAM algorithms. For instance, a deep neural network that enhances the accuracy of a traditional SLAM algorithm's output or a deep neural network that refines the output of a typical SLAM method.
2. Increased accuracy: Deep learning can potentially enhance the accuracy of maps produced by conventional SLAM techniques. To create more precise 3D maps, a deep neural network can be trained to estimate the depth of objects in a given environment.
3. Increasing Robustness: Deep learning can also improve the robustness of conventional SLAM systems. For instance, a deep neural network can be taught to recognize and rectify faults in the SLAM output, improving its ability to handle difficult situations like low light levels or rapid motions.

Limitations:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

1. Data collection and annotation: Deep learning algorithms need a lot of labeled data to be trained, which might be challenging to gather and label in a flight simulator environment.
2. High computational requirements: Deep learning techniques demand a lot of computer power, which can be difficult for a device like the Hololens, which has a relatively low processing capacity.
3. Model size: Deep learning models can get rather large, making it difficult to store and deploy them on a system like Hololens.
4. Deep learning algorithms can be expensive to train, making them less desirable for commercial flight simulator applications.

2 Results and discussion

We used HoloLens 2 to run SLAM techniques in the flight simulator, and we got access to real-time live-streaming video of the simulator cockpits from every camera. Utilizing the research mode (RM) depicted in Figure 1, data is collected from the depth sensors, IMU, and gray-scale cameras. We used the Python library to connect to the HoloLens 2 server, send commands and configuration data, receive and decode data, and analyze the data for usage with other libraries.

The main goal of utilizing HoloLens to run the SLAM methods on the cockpit was to determine how well each method could be customized to find patterns and structures in the cockpit or map the flight simulator cockpit for precise and reliable pose estimation. Our test results are displayed below. In our result, as compared in Table 1, the SLAM methods discussed in our methods section, we inferred with our experimental result, as shown in figure1 the performance of SLAM methods in a flight simulator environment. RGB-D SLAM is considered moderate compared with other SLAM techniques because flight simulators often have a controlled and structured environment, so there are fewer variables and obstacles to consider when detecting features in the environment. ORB and LSD SLAM are visual-based techniques that use features extracted from the images to perform SLAM. ORB SLAM is considered poor because it is not robust to lighting changes or occlusion; it makes it difficult for the SLAM algorithm to distinguish between different objects in the environment, while LSD SLAM is poor in terms of accuracy in a large-scale environment. VINS fuses information from

Methods	Map quality	Localization Accuracy	Computational Cost	Robustness	Efficiency	In flight simulator
RGB-D	Moderate	Good	High	Poor	Fair	Moderate
ORB	Low	Good	Moderate	Good	Good	Poor
LSD	High	Excellent	Low	Fair	Fair	Poor
VINS	High	Excellent	High	Excellent	Good	Good
DL	Good	Excellent	Very high	Good	Low	Good

Table 1. Comparative table of SLAM methods and flight simulator environment

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

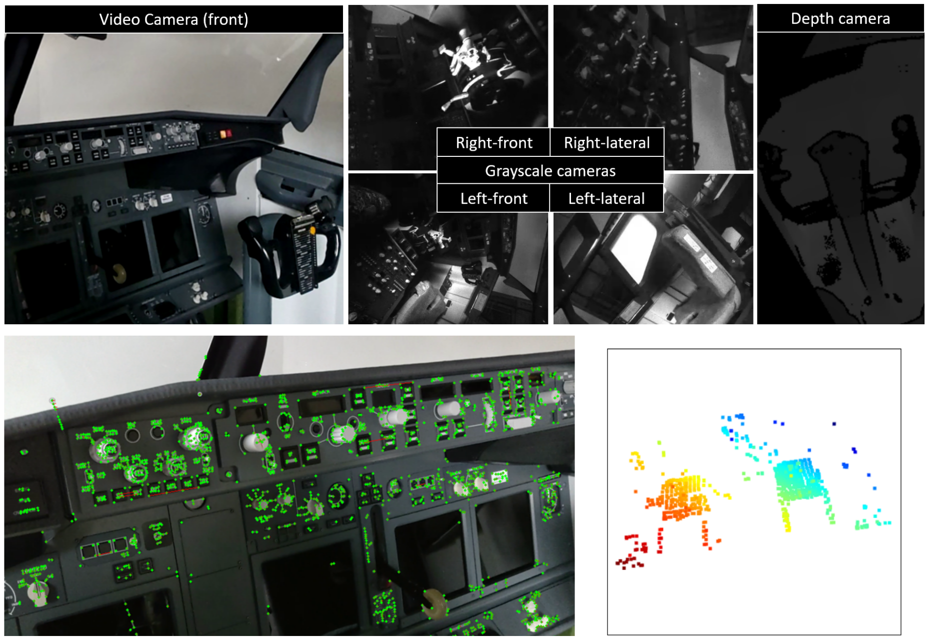

Fig. 1. Top picture shows the simultaneous capture from all the Hololens 2 sensors. The left picture shows the feature detection, while the right one - the 3D Map of the cockpit.

visual and inertial sensors; therefore, it is considered good because it provides accurate results without additional sensors. Deep Learning (DL) is considered good because it can learn to extract useful features from the sensor data and adapt to a different environment.

3 Conclusion

This paper discussed and analyzed different SLAM methods in a flight simulator environment. We obtained a high-resolution live stream video of a flight simulator cockpit with a high dynamic range using HoloLens 2. We perform feature detection on the video stream, and the SLAM algorithm then extracts features from the cockpit structures. These futures can be used to estimate the robot's position and orientations relative to the environment's features, track the robot's movement, and build a map of the environment in real-time. Comparing different SLAM techniques is challenging because their performance depends on the complexity of the surrounding environment and the type of sensors being employed. Nevertheless, this work still suggests selecting a suitable SLAM technique for HoloLens-based flight simulator scenarios. Future research should concentrate on enhancing the effectiveness of Visual Inertial SLAM and upgrading it based

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

on the findings of this publication. We will demonstrate how the proposed techniques may be utilized to develop more accurate and reliable algorithms for feature detection, tracking, and aligning virtual objects in the real world while using augmented reality goggles. This will help to reduce the incidence of tracking failures. We will present the development of better sensor fusion techniques by integrating with other sensors and fusing data from multiple sensors; the system can better understand its environment and its position. This model can aid in improving the SLAM system's accuracy and help it better forecast aircraft behavior under various flight conditions. In augmented reality, one of the significant challenges is the ability to operate in a dynamic and unstructured environment. We will create an algorithm capable of a wider range of environments in handling challenging conditions, such as low light, moving objects and dynamic scenes. We will also investigate how to make algorithms better by using machine learning to track the location and orientation of an AR device in the actual world. The program can recognize features in real-world surroundings and follow their movement to estimate the position and orientation of the device by training machine learning on an extensive data collection of images. Additionally, we will apply SLAM methods to reduce the latency caused by occlusions, where real-world objects obscure virtual objects. SLAM techniques can help to ensure that virtual items are perfectly aligned with the actual world, even when the user moves and the environment changes, by updating the device's position and orientation continually.

References

1. C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
2. Y. Chen, "Algorithms for simultaneous localization and mapping," *February*, vol. 3, pp. 1–15, 2013.
3. G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.
4. Y. Nava, *Visual-LiDAR SLAM with loop closure*. PhD thesis, KTH Royal Institute of Technology, 2018.
5. T. Sun, Y. Liu, Y. Wang, and Z. Xiao, "An improved monocular visual-inertial navigation system," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11728–11739, 2020.
6. R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *2011 international conference on computer vision*, pp. 2320–2327, IEEE, 2011.
7. J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pp. 834–849, Springer, 2014.
8. R. Berkvens, D. Vandermeulen, C. Vercauteren, H. Peremans, and M. Weyn, "Feasibility of geomagnetic localization and geomagnetic ratslam," *International journal on advances in systems and measurements*, vol. 7, no. 1-2, pp. 44–56, 2014.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

9. R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE international symposium on mixed and augmented reality*, pp. 127–136, Ieee, 2011.
10. X. Meng, W. Gao, and Z. Hu, "Dense rgb-d slam with multiple cameras," *Sensors*, vol. 18, no. 7, p. 2118, 2018.
11. S. Wang, R. Clark, H. Wen, and N. Trigoni, "Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," in *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 2043–2050, IEEE, 2017.
12. S. A. Mohamed, M.-H. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen, and J. Plosila, "A survey on odometry for autonomous navigation systems," *IEEE access*, vol. 7, pp. 97466–97486, 2019.
13. S. Karam, V. Lehtola, and G. Vosselman, "Integrating a low-cost mems imu into a laser-based slam for indoor mobile mapping," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 149–156, 2019.
14. R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g 2 o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation*, pp. 3607–3613, IEEE, 2011.
15. J.-E. Deschaud, "Imls-slam: Scan-to-model matching based on 3d data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2480–2485, IEEE, 2018.
16. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*, pp. 2564–2571, Ieee, 2011.
17. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
18. H.-C. Chi, T.-H. Tsai, and S.-Y. Chen, "Slam-based augmented reality system in interactive exhibition," in *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp. 258–262, IEEE, 2020.
19. R. T. Azuma, "The most important challenge facing augmented reality," *Presence*, vol. 25, no. 3, pp. 234–238, 2016.
20. Z. Zhang, M. Shu, Z. Wang, H. Wang, and X. Wang, "A registration method for augmented reality system based on visual slam," in *2019 International Conference on Electronic Engineering and Informatics (EEI)*, pp. 408–411, IEEE, 2019.
21. R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
22. H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao, "Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1974–1982, 2018.
23. J. Cyrus, D. Krcmarik, R. Moezzi, J. Koci, and M. Petru, "Hololens used for precise position tracking of the third party devices-autonomous vehicles," *Communications-Scientific letters of the University of Zilina*, vol. 21, no. 2, pp. 18–23, 2019.
24. M. A. Hoffman, "Microsoft hololens development edition," *Science*, vol. 353, no. 6302, pp. 876–876, 2016.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

25. M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger, “Real-time 3d reconstruction at scale using voxel hashing,” *ACM Transactions on Graphics (ToG)*, vol. 32, no. 6, pp. 1–11, 2013.
26. B. Glocker, J. Shotton, A. Criminisi, and S. Izadi, “Real-time rgb-d camera relocalization via randomized ferns for keyframe encoding,” *IEEE transactions on visualization and computer graphics*, vol. 21, no. 5, pp. 571–583, 2014.
27. P. Skurowski, K. Nurzyńska, M. Pawlyta, and K. A. Cyran, “Performance of QR Code Detectors near Nyquist Limits,” *Sensors*, vol. 22, p. 7230, Sept. 2022.
28. J. Cheng, L. Zhang, Q. Chen, X. Hu, and J. Cai, “A review of visual slam methods for autonomous driving vehicles,” *Engineering Applications of Artificial Intelligence*, vol. 114, p. 104992, 2022.
29. A. Juneja, L. Bhandari, H. Mohammadbagherpoor, A. Singh, and E. Grant, “A comparative study of slam algorithms for indoor navigation of autonomous wheelchairs,” in *2019 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pp. 261–266, IEEE, 2019.
30. Q. Zou, Q. Sun, L. Chen, B. Nie, and Q. Li, “A comparative analysis of lidar slam-based indoor navigation for autonomous vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6907–6921, 2021.
31. M. U. Khan, S. A. A. Zaidi, A. Ishtiaq, S. U. R. Bukhari, S. Samer, and A. Farman, “A comparative survey of lidar-slam and lidar based sensor technologies,” in *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, pp. 1–8, IEEE, 2021.
32. X. Zhou and R. Huang, “A state-of-the-art review on slam,” in *Intelligent Robotics and Applications: 15th International Conference, ICIRA 2022, Harbin, China, August 1–3, 2022, Proceedings, Part III*, pp. 240–251, Springer, 2022.
33. S. Klose, P. Heise, and A. Knoll, “Efficient compositional approaches for real-time robust direct visual odometry from rgb-d data,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1100–1106, IEEE, 2013.
34. X. Gao, R. Wang, N. Demmel, and D. Cremers, “Ldso: Direct sparse odometry with loop closure,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2198–2204, IEEE, 2018.
35. W. Dai, Y. Zhang, P. Li, Z. Fang, and S. Scherer, “Rgb-d slam in dynamic environments using point correlations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 373–389, 2020.
36. D. Kiss-Illés, C. Barrado, and E. Salamí, “Gps-slam: an augmentation of the orb-slam algorithm,” *Sensors*, vol. 19, no. 22, p. 4973, 2019.
37. L. Cai, Y. Ye, X. Gao, Z. Li, and C. Zhang, “An improved visual slam based on affine transformation for orb feature extraction,” *Optik*, vol. 227, p. 165421, 2021.
38. B. Bescos, J. M. Fácil, J. Civera, and J. Neira, “Dynaslam: Tracking, mapping, and inpainting in dynamic scenes,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.
39. J. Cheng, Y. Sun, and M. Q.-H. Meng, “Improving monocular visual slam in dynamic environments: An optical-flow-based approach,” *Advanced Robotics*, vol. 33, no. 12, pp. 576–589, 2019.
40. C. Forster, M. Pizzoli, and D. Scaramuzza, “Svo: Fast semi-direct monocular visual odometry,” in *2014 IEEE international conference on robotics and automation (ICRA)*, pp. 15–22, IEEE, 2014.
41. P. Bergmann, R. Wang, and D. Cremers, “Online photometric calibration of auto exposure video for realtime visual odometry and slam,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 627–634, 2017.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46

42. P. Liu, X. Yuan, C. Zhang, Y. Song, C. Liu, and Z. Li, "Real-time photometric calibrated monocular direct visual slam," *Sensors*, vol. 19, no. 16, p. 3604, 2019.
43. T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
44. S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time on-board visual-inertial state estimation and self-calibration of mavs in unknown environments," in *2012 IEEE international conference on robotics and automation*, pp. 957–964, IEEE, 2012.
45. H. Yin, S. Li, Y. Tao, J. Guo, and B. Huang, "Dynam-slam: An accurate, robust stereo visual-inertial slam method in dynamic environments," *IEEE Transactions on Robotics*, 2022.
46. Q. Cheng, S. Zhang, S. Bo, D. Chen, and H. Zhang, "Augmented reality dynamic image recognition technology based on deep learning algorithm," *IEEE Access*, vol. 8, pp. 137370–137384, 2020.
47. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
48. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
49. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
50. T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1851–1858, 2017.
51. X. Gao and T. Zhang, "Unsupervised learning to detect loops using deep neural networks for visual slam system," *Autonomous robots*, vol. 41, pp. 1–18, 2017.
52. M. Geng, S. Shang, B. Ding, H. Wang, and P. Zhang, "Unsupervised learning-based depth estimation-aided visual slam approach," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 543–570, 2020.
53. F. Li, W. Chen, W. Xu, L. Huang, D. Li, S. Cai, M. Yang, X. Xiong, Y. Liu, and W. Li, "A mobile robot visual slam system with enhanced semantics segmentation," *IEEE Access*, vol. 8, pp. 25442–25458, 2020.
54. L. Zhang, L. Wei, P. Shen, W. Wei, G. Zhu, and J. Song, "Semantic slam based on object detection and improved octomap," *IEEE Access*, vol. 6, pp. 75545–75559, 2018.
55. J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
56. K. Tateno, F. Tombari, I. Laina, and N. Navab, "Cnn-slam: Real-time dense monocular slam with learned depth prediction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6243–6252, 2017.
57. R. Li, S. Wang, Z. Long, and D. Gu, "Undeepvo: Monocular visual odometry through unsupervised deep learning," in *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 7286–7291, IEEE, 2018.
58. S. Vijayanarasimhan, S. Ricco, C. Schmid, R. Sukthankar, and K. Fragkiadaki, "Sfm-net: Learning of structure and motion from video," *arXiv preprint arXiv:1704.07804*, 2017.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_46