The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-6-2025
ISPRS, EARSeL & DGPF Joint Istanbul Workshop “Topographic Mapping from Space” dedicated to Dr. Karsten Jacobsen’s 80th Birthday
29–31 January 2025, Istanbul, Türkiye

An Assessment of the Use of Visual Odometry in Localization

İrem Yakar [1], Ramazan Alper Kucak [1], Serdar Bilgi [1]

[1] ITU, Department of Geomatics Engineering, 80626 Maslak Istanbul, Türkiye - (yakari, kucak15, bilgi) @itu.edu.tr

Keywords: Localization, LiDAR, Visual Odometry, iPad Pro.

Abstract

Localization is the process of determining the path, position and orientation of a robot in an environment. The data from different sensors such as LiDAR, inertial measurement units (IMU) or cameras are used in the localization process. This task is of fundamental importance to enable robots to navigate autonomously and perform tasks effectively. The robot localization can be performed utilizing different techniques either using hardware or software designs. Visual localization algorithms can be shown as one of the localization techniques that enable robots to determine their position and orientation. In this aspect, visual odometry is one of the mostly used methods. It is a technique that enables robots to determine their positions and movements by analysing sequential images. It tracks features in consecutive images to determine the movement of the camera between those frames which enables the determination of the robot’s position in the environment. The process commonly involves detecting and matching key points or features in the images, such as corners or edges, and then using algorithms to calculate the camera's motion. Visual odometry is useful in environments where Global Navigation Satellite Systems (GNSS) or other external positioning systems cannot operate. In this study, the use of visual odometry is assessed in comparison with the iPad Pro LiDAR and steel tape results to determine the distances between each image-taking point. The iPad and steel tape results were taken as the ground truth and the root mean square values were determined by comparing the algorithm and their results.

1. Introduction

Localization can be defined as identifying a robot’s position concerning its environment. Localization is an important aspect of any autonomous robot since understanding its position is a necessary foundation for planning and executing future actions. In a usual robot localization framework, a map of the robot’s surroundings is available and the robot has sensors to observe the environment and track its movement. Then, the localization task involves determining the robot's position and orientation utilizing the information obtained using these sensors. Robot localization methods must also be able to handle noisy data, providing both an estimate of the robot's position and a measure of the uncertainty linked to that estimate (Huang and Dissanayake, 1999).

In general sense, several technologies can be used for accurate localization, and the choice of method depends largely to the environment.
Global Navigation Satellite Systems (GNSS) are the most commonly used methods for localization at outdoor, where satellite connections are readily available. GNSS refers to a group of geo-referenced high orbit satellite systems, including the GPS (Global Positioning System) operated by the United States, GLONASS (Russia), Galileo (European Union), and
BeiDou (China), which use signals to broadcast position (longitude, latitude, and altitude) and time information to receivers on Earth (Moradbeikie et al., 2021). GNSS technology is essential for producing maps (gathering the survey data), navigation for land-air-sea transportation vehicles, all engineering applications requiring location data, military purposes and location-based services.

GNSS technology has its limitations, especially in indoor areas or densely constructed urban areas where satellite signals may be blocked or diminished. GNSS can be only used in outdoor environment because satellite signals cannot be received by the GNSS antennas inside the closed areas/buildings. For this reason, it is required different system providing indoor navigation at these types of places (Faria et al., 2010). In these situations, disparate techniques for localization is necessary. The relative location information can be obtained through calculating the signal strength of multiple access points with the techniques such as Wi-Fi triangulation and Bluetooth beacons (Bilgi et al., 2017). Wi-Fi localization systems are dependent upon Wi-Fi coverage and are not adjusted for recalculating location Ahmetovic et al., 2017). These methods work by triangulating signals from devices such as Wi-Fi routers or Bluetooth and calculating distances based on their relative signal strength. While, methods may not be as accurate as GNSS, they are applicable for indoor applications where exterior barriers prevent GNSS signal performance from being optimal (Tiku and Pasricha, 2023).

In this study, the performance of the visual odometry was examined by comparing the localization results with the iPad Pro LiDAR and steel tape measurements. The distances between the points where photographs were taken were compared with the algorithm and the iPad Pro & steel tape results, and Root Mean Square Error (RMSE) values were calculated.

2. Methodology

2.1 Visual Odometry

Visual odometry is a method to determine an object’s motion (its rotation and orientation) in an environment utilizing a sequence of images. It can be classified as a special case of Structure from Motion (SFM) is where both the 3D structure of the environment and pose information are obtained using unordered or in-order images (Scaramuzza and Fraundorfer, 2011).

In recent years, different types of visual odometry methods have been presented. These methods can be classified into two categories: monocular (Campbell et al., 2005) and stereocamera (Mahon et al., 2008). Then, they can be divided into feature matching (Talukder et al., 2003), feature tracking (Dornhege and Kleiner, 2006), and optical flow techniques (Zhang et al., 2009). Nister et al. (2006) initially introduced visual odometry, since it

is similar to the concept of wheel odometry. They presented newly developed methods for determining camera poses utilizing visual input either monocular or stereo. The study focused on the outlier rejection using Random Sample Consensus (RANSAC) where the false feature matches are rejected (Fischler and Bolles, 1981). Nister et al. (2006) were the first to apply feature tracking in all frames rather than limiting feature matching to sequential images. This method helps to eliminate the feature drift usually associated with cross-correlation-based tracking methods (Scaramuzza and Fraundorfer, 2011). They also introduced a RANSAC-based motion estimation method that relies on 3D-to-2D reprojection error rather than the Euclidean distance error between 3D points. It was demonstrated that using 3D-to-2D reprojection errors produces more accurate estimates compared to 3D-to-3D errors (Henry et al., 2012).

The robotic space mission on Mars (Cheng et al., 2005) in 2003 is an example of the usage of visual odometry that aimed to examine the surface and the geology of the planet using two rovers. Scaramuzza and Siegwart (2008) performed another study utilizing visual odometry in an outdoor environment. They used a monocular omni-directional camera and performed a fusion approach using two different methods: In the first method; the Scale Invariant Feature Transform (SIFT) feature extraction was used and the RANSAC was also used for the outlier rejection (Lowe, 2004). In the other method, an appearance-based method proposed by Comport et al. (2007) was utilized for the pose estimation.

Visual odometry progressively determines a vehicle's movement by examining sequential camera images and calculating the relative pose between perspectives using 2D bearing vectors extracted from detected features. At time k, the visual odometry algorithm processes two consecutive images, I_k and I_{k-1} , as input and generates an incremental motion estimate relative to the local camera reference frame. This motion is expressed as $\delta *_{k,k-1} \in \mathrm{R}^3$ (Ouerghi et. al., 2018):

$$\delta *_{k,k-1} = (\Delta s *_k, \Delta\theta_k) \quad (1)$$

Visual odometry techniques can be categorized depending on the imagery they use during the monocular or stereoscopic process- and the processing methods are either direct (image/appearance-based) or feature-based. The techniques may utilize a combination of feature tracking, feature matching, or optical flow (Scaramuza and Fraundorfer, 2011; Yousif et al., 2015). Most of the visual odometry systems use images from a pair of cameras that are mounted on a robot since the majority of them produce 3D navigation information from a set of images. The triangulation of image features is used to calculate the robot's velocity and displacement easily depending on the distance between cameras and their capture frame rate (Hartley, 1997). On the other hand, monocular visual odometry is more challenging, and it has only recently gained attention. Monocular visual odometry estimates motion and reconstructs the environment using at least three consecutive 2D images and their associated bearing data. A parallel tracking and mapping (PTAM) algorithm (Klein and Murray, 2007), is utilized in many monocular applications. The algorithm was initially developed for augmented reality (AR), and its speed and robustness, relying solely on existing map features, have made it a popular choice among visual odometry researchers (Lim and Braunl, 2020). The types of visual odometry can be seen in Figure 1.

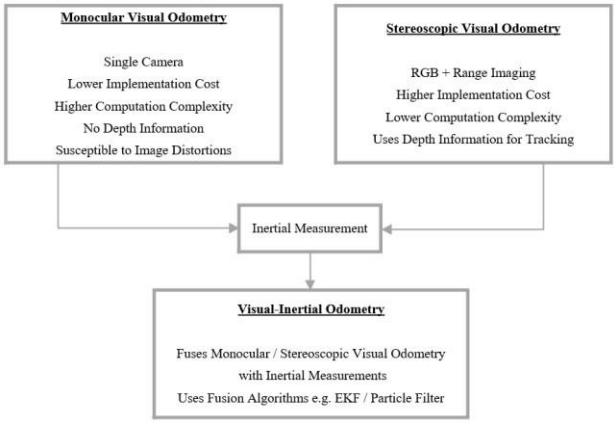

Figure 1. Types of visual odometry (Lim and Braunl, 2020).

2.2 Light Detection and Ranging (LiDAR)

Light detection and ranging (LiDAR) technology has been widely used in different types of applications in recent years. LiDAR is a measurement system that quickly produces large amounts of 3D point cloud data. The features of LiDAR systems have advanced significantly, leading to a configuration with notably reduced weight, size, cost, and power (SWaP) requirements (Liner, 2015). The classification of LiDAR instruments differs based on the applications. According to, the classification based on the measurement platform, there are two types of LiDAR: aerial and terrestrial. Generally, time of flight (TOF) measurements are used to capture spatial information which is a core function of all LiDAR instruments. LiDAR systems that collect spatial data are available in three types: one-dimensional (1D), two-dimensional (2D), and threedimensional (3D), with 2D and 3D spatial data acquisition facilitated by optical deflection systems. Spatial information is of fundamental importance for generating a precise 3D map of the environment. However, it alone is insufficient for applications that involve object detection. The second type of LiDAR instrument is designed to measure spectral data, like the laser return intensity (LRI), to offer supplementary information. Additionally, certain applications require the capture of temporal information alongside spatial and spectral data. The aim can be achieved with the use of the repeated LiDAR technique, which involves collecting temporal data from a target environment over a specific period (Robin and Jacky, 2014).

LiDAR architecture is referred to as "the art of LiDAR instrumentation, which involves both hardware and software" (Xinzhao, 2012 & 2016). A fully operational LiDAR system includes four key subsystems: laser rangefinder, beam deflection, power management and master controller units, as illustrated in Figure 2. Each of these subsystems is critical, as a failure in any one of them can fail the functioning of the system.
Without the beam deflection subsystem, the LiDAR can still operate as a 1D system, commonly referred to as a laser rangefinder (LRF) (Raj et al., 2020). The basics of LiDAR systems can be seen in Figure 2.

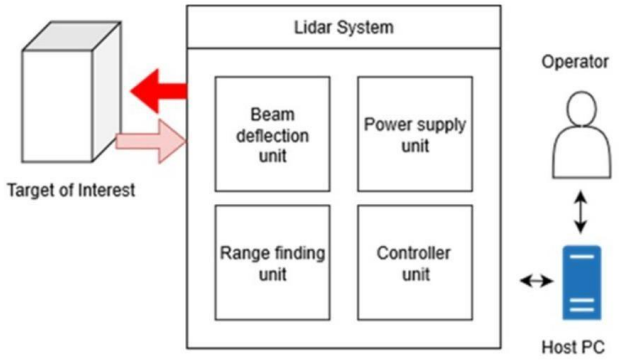

Figure 2. LiDAR system (Raj et al., 2020).

2.2.1 iPad Pro LiDAR: The IPad Pro and iPhone 12 Pro with built-in LiDAR were released in 2020 which can be considered as a great innovation in tablet&smartphone market. The LiDAR integrated into these devices primarily focuses on Augmented Reality (AR) applications, and its performance in this area has been tested in different studies (Spreafico et al., 2021). The IPad Pro and iPhone 12 Pro took the attention of the researchers in both indoor and outdoor studies. They use sensors developed by Sony for the LiDAR Scanners. The IPhone and IPad LiDAR uses a “time of flight” technology that measures the time it takes for light pulses (approximately 940 nm near-infrared range) (URL 1). They are particularly valuable when speed, portability, and cost-efficiency are critical. Their technical capabilities, cost-effectiveness, and user-friendly design make them an attractive alternative to traditional techniques, such as Terrestrial Laser Scanning (TLS) and Photogrammetric cameras, which are commonly utilized in different applications (Chiabrando et al., 2011).

The internal structure of the iPad Pro LiDAR can be seen in Figure 3.

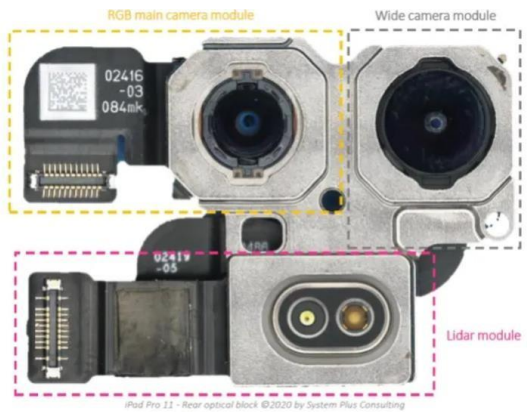

Figure 3. iPad Pro LiDAR (Yoshida, 2020).

3. Application

The study focuses on obtaining camera poses from consecutive images (Figure 4).

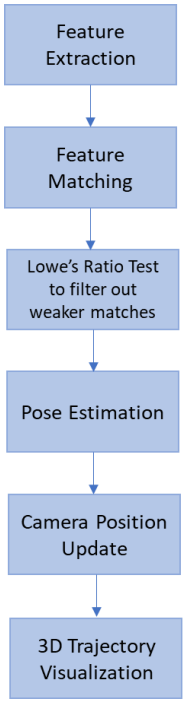

Figure 4. The workflow of visual odometry.

The ORB (Oriented FAST and Rotated BRIEF) algorithm is used for feature detection. The features are then matched between successive image frames using a brute-force matcher (BFMatcher) with Hamming distance. Lowe's ratio test is utilized to improve feature-matching accuracy. Therefore, only matches where the distance of the nearest neighbour is significantly less than that of the second nearest neighbour are kept (threshold value is 0.75). The detected key points and matched features between images can be seen in figures 5 and 6.

Figure 5. Detected key points.

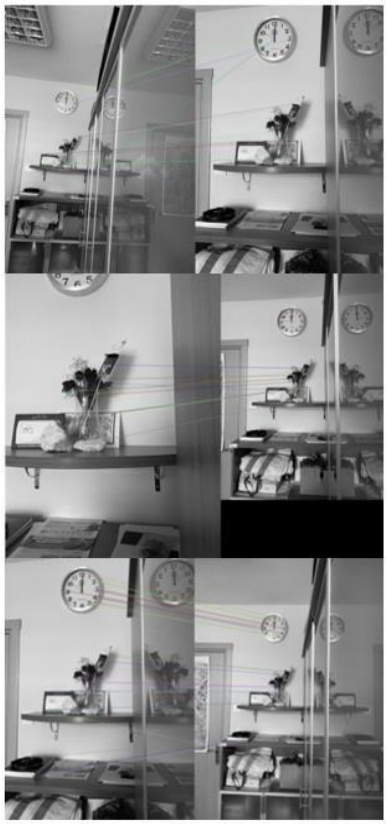

Figure 6. The matched features between images.

RANSAC is used to estimate the fundamental matrix utilizing these filtered matches. The essential matrix is computed using a predefined intrinsic camera matrix and the fundamental matrix. The relative rotation and translation of the camera between two frames are then estimated utilizing the essential matrix. While each image is being processed, the cumulative rotation and translation are being tracked, assuming the first image as the origin. The trajectory of the camera is obtained by updating the positions iteratively. Then a 3D plot is obtained to display the camera's trajectory. The localization result can be seen in figure 7.

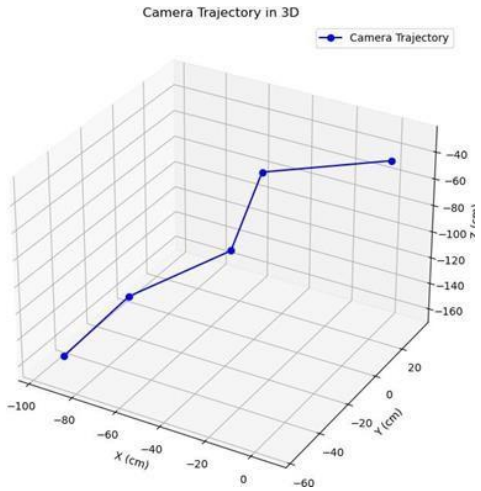

Figure 7. Localization with visual odometry.

The same area was scanned using iPad Pro’s LiDAR and modelled in the SiteScape software (Figure 8). The distances were also measured on the scan data and then compared with the results from the visual odometry.

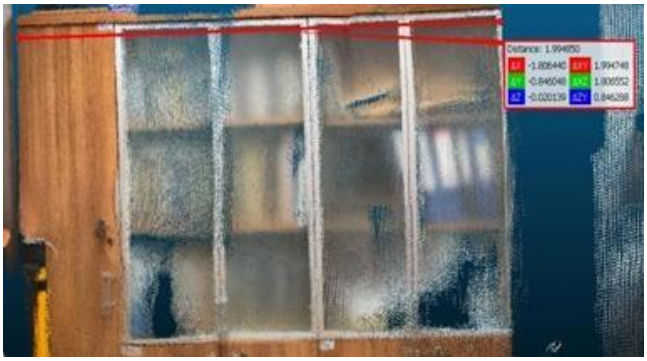

Figure 8. The point cloud data of iPad Pro in SiteScape Software.

The distances between the points where the photographs were captured were also measured with steel tape, on the ground. The same distances were also calculated with iPad Pro LiDAR point cloud. Consequently, a comparison between the distances obtained by visual odometry and Apple iPad Pro and steel tape measurements were acquired. The distances between each point where the photographs were captured can be seen in Table 1.

Image Pair	Distance (Visual Odometry) (cm)	Distance (Steel Tape) (cm)	Distance (iPad Pro) (cm)
1 to 2	50.0	40.5	39.5
2 to 3	50.0	40.5	40.1
3 to 4	50.0	40.5	39.4
4 to 5	50.0	40.5	41.0

Table 1. The distances between each point.

The RMSE values were calculated by comparing the visual odometry results with the distance obtained by iPad Pro and steel tape. The comparison with steel tape gives a RMSE of 9.50 cm while the comparison with iPad Pro gives a RMSE of 10.02 cm.

4. Conclusion

In this study, the performance of visual odometry was evaluated by comparing the distances between each image capturing point with the iPad Pro and steel tape measurements. The RMSE values are 9.50 cm in comparison with steel tape measurements and 10.02 cm in comparison with iPad Pro LiDAR measurements respectively. Since, the localization process is easy and fast with visual odometry. However, the accuracy needs should be taken into consideration based on the application. For instance, in environments requiring high accuracy, such as robotic surgery or high-stakes construction, visual odometry may need to be augmented with more reliable measurement techniques. The findings show the importance of selecting the appropriate measurement method based on the specific requirements and accuracy needs of the task that is carried out.

References

Ahmetovic, D., Murata, M., Gleason, C., Brady, E., Takagi, H., Kitani, K., Asakawa, C., 2017. Achieving Practical and Accurate Indoor Navigation for People with Visual Impairments. *14th Web for All Conference on The Future of Accessible Work*, No. 31.

Bilgi, S., Ozturk, O., Gulnerman, A., 2017, Navigation System for Blind, Hearing and Visually Impaired People in ITU Ayazaga

Campus. *IEEE, International Conference on Computing Networking and Informatics (ICCNI)*, 1-5.

Campbell, J., Sukthankar, R., Nourbakhsh, I., Pahwa, A., 2005. A robust visual odometry and precipice detection system using consumergrade monocular vision. *IEEE International Conference on Robotics and Automation (ICRA2005)*, 3421–3427.

Cheng, Y., Maimone, M., Matthies, L., 2005. Visual odometry on the Mars exploration rovers. *IEEE International Conference on Systems, Man and Cybernetics*, 1, 903–910.

Chiabrando, F., Piatti, D., Rinaudo, F., 2011. New sensors for cultural heritage metric survey: the ToF cameras. *Geoinformatics*, FCE CTU, 6, 307-313.

Comport, A., Malis, E., Rives, P., 2007. Accurate quadrifocal tracking for robust 3D visual odometry. *IEEE International Conference on Robotics and Automation*, 40–45.

Dornhege, C., Kleiner, A., 2006. Visual odometry for tracked vehicles. *IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR)*.

Faria, J, Lopes, S., Fernandes, H., Martins, P. Barroso, J., 2010. Electronic white cane for blind people navigation assistance. *In: World Automation Congress (WAC)*, 1-7.

Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395. doi.org/10.1145/358669.358692.

Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D., 2012. RGB-D mapping: using kinect-style depth cameras for dense 3D modeling of indoor environments. *The International Journal of Robotics Research.*, 31(5), 647–663.

Klein, G., Murray, D., 2007. Parallel tracking and mapping for small AR workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, 225-234.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

Hartley, R.I., Sturm, P., 1997. Triangulation. *Computer Vision and Image Understanding*, 68(2), 146–157.

Huang, S., Dissanayake, G., 1999. Robot localization: An introduction. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1-10.

Lim, K.L., Bräunl, T., 2020. A review of visual odometry methods and its applications for autonomous driving. *arXiv preprint arXiv:2009.09193*.

Liner, J., 2015. SWaP: The RF solution that can mean the difference between flying high and being grounded; Analog Devices: Norwood, MA, USA.

Mahon, I., Williams, S., Pizarro, O., Johnson-Roberson, M., 2008. Efficient view-based slam using visual loop closures. *IEEE Transactions on Robotics*, 24(5), 1002–1014.

Moradbeikie, A., Keshavarz, A., Rostami, H., Paiva, S., Lopes, S.I., 2021. GNSS-free outdoor localization techniques for resource-constrained IoT architectures: A literature review. *Applied Sciences*, 11(22), 10793. doi.org/10.3390/app112210793.

Nister, D., Stewenius, H., 2006. Scalable recognition with a vocabulary tree. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, 2161–2168. doi.org/10.1109/CVPR.2006.264.

Ouerghi, S., Boutteau, R., Savatier, X., Tlili, F., 2018. Visual odometry and place recognition fusion for vehicle position tracking in urban environments. *Sensors*, 18(4), 939.

Raj, T., Hanim Hashim, F., Baseri Huddin, A., Ibrahim, M.F., Hussain, A., 2020. A survey on LiDAR scanning mechanisms. *Electronics*, 9(5), 741.

Robin, G.J., Jacky, C., 2014. Making a difference: Examples of the use of repeat LiDAR datasets to guide river management decisions following extreme floods. *In Proceedings of the 7th Australian Stream Management Conference*, Townsville, Australia, 232–239.

Talukder, A., Goldberg, S., Matthies, L., Ansar, A., 2003. Realtime detection of moving objects in a dynamic scene from moving robotic vehicles. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), Proceedings*, 2, 1308–1313.

Tiku, S., Pasricha, S., 2023. An overview of indoor localization techniques. *Machine Learning for Indoor Localization and Navigation*, 3-25.

Scaramuzza, D., Siegwart, R., 2008. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Transactions on Robotics*, 24(5), 1015–1026. doi.org/10.1109/TRO.2008.2004490.

Scaramuzza, D., Fraundorfer, F., 2011. Visual odometry: part Ithe first 30 years and fundamentals. *IEEE Robotics & Automation Magazine*, 18(4), 80–92.

Spreafico, A., Chiabrando, F., Teppati Losè, L., Giulio Tonolo, F., 2021. The ipad pro built-in lidar sensor: 3d rapid mapping tests and quality assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 63-69.

Xinzhao, C. Lecture 07, Fundamentals of Lidar remote sensing. 2012. Available online: http://superlidar.colorado.edu/Classes/Lidar2012/LidarLecture07_Architecture.pdf (accessed on 1 December 2024).

Xinzhao, C. Lecture 08, Fundamentals of Lidar remote sensing. 2016. Available online: http://superlidar.colorado.edu/Classes/Lidar2016/Lidar2016_Lecture08_Architecture.pdf (accessed on 1 December 2024).

Xinzhao, C. Lecture 36, Lidar Architecture and Lidar Design. 2016. Available online: http://superlidar.colorado.edu/Classes/Lidar2016/Lidar2016_Lecture36_LidarDesignArchitecture.pdf (accessed on 1 December 2024).

Xinzhao, C. Lecture 41, Lidar Architecture and Lidar Design. 2012. Available online: http://superlidar.colorado.edu/Classes/Lidar2012/LidarLecture41_LidarDesign1.pdf (accessed on 1 December 2024).

Yousif, K., Bab-Hadiashar, A., Hoseinnezhad, R., 2015. An overview to visual odometry and visual SLAM: Applications to mobile robotics. *Intelligent Industrial Systems*, 1(4), 289-311.

Yoshida, J., 2020. Breaking Down iPad Pro 11’s LiDAR Scanner. https://www.eetimes.com/breaking-down-ipad-pro11s-lidar-scanner/2/ (accessed on 4 December 2024).

Zhang, T., Liu, X., Kühnlenz, K., Buss, M., 2009. Visual odometry for the autonomous city explorer. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS’09)*, 3513–3518. IEEE Press, Piscataway. doi.org/10.1109/IROS.2009.5354675.

URL 1. https://www.techinsights.com/blog/sony-d-tof-sensorfound-apples-new-lidar-camera. Accessed on December 10Th, 2024.