ENS 492 – Graduation Project

(Implementation)

Final Report

Reality Capture from Multiple Video Cameras

Group Members:

İpek Nur

Dolu

Özge

Karasu

Serkan

Kütük

Supervisor: Özgür Erçetin

Date: 7 January 2024

Sabancı Üniversitesi

Table of Contents

1. EXECUTIVE SUMMARY

This project addresses the knowledge gap in Intelligent Transportation Systems (ITS) through the development of a strategic approach. The primary objective is to leverage multiple video streams and advanced photogrammetry techniques to create precise 3D models of observed vehicles. By improving vehicle tracking and recognition in ITS, the project aims to propel advancements in traffic management, emergency response systems, and road safety.

In addition to overcoming challenges posed by moving cameras, shifting lighting, and processing resource requirements during the development process, the project places a strong emphasis on optimizing the computational efficiency of the proposed approach. This involves implementing cutting-edge algorithms to enhance real-time processing, ensuring practical applicability in dynamic traffic environments.

While the comprehensive evaluations of vehicle behavior and traffic patterns were not conducted due to project scope constraints, the focus remains on creating accurate 3D models that offer valuable insights. These insights are intended to play a pivotal role in informing future transportation system designs, with a specific focus on maximizing operational effectiveness and user-friendliness.

Beyond the creation of accurate 3D models, the system's viability and efficacy undergo thorough validation and assessment according to predetermined standards. This meticulous evaluation process ensures the potential of the developed approach to significantly enhance ITS capabilities, contributing to the overall advancement of intelligent transportation technologies.

In conclusion, this project not only addresses current challenges in ITS through precise 3D modeling but also emphasizes the optimization of computational efficiency. By providing valuable insights, rigorous testing, and adherence to practical limitations, the proposed method holds the promise of making substantial contributions to the field of Intelligent Transportation Systems.

2. PROBLEM STATEMENT

The original problem centered around the knowledge gap in Intelligent Transportation Systems (ITS), particularly as it related to the accuracy and effectiveness of 3D modeling for the tracking and identification of vehicles. Developing a real-time strategy to extract accurate information from multiple video streams while taking into account variables like moving cameras, changing lighting, and processing resource requirements was the challenge.

The motivation for addressing this problem stemmed from the critical impact it could have on advancing ITS. Road safety, emergency response systems, and traffic management all benefit greatly from improved vehicle tracking and recognition. Additional driving force behind the motivation was the possibility of offering insightful observations about vehicle behavior and traffic patterns. To close the knowledge gap and create more efficient and user-friendly transportation systems was the goal.

The main objective was to develop a practical plan for producing accurate 3D models of observed vehicles through the use of sophisticated video processing techniques. By doing this, the project hopes to provide accurate data for emergency response and traffic management, as well as solutions for real-world ITS challenges. The aim to develop a workable and applicable solution was reflected in the focus on building a robust model that could handle dynamic environmental factors, like moving cameras and changing lighting. The project's ultimate goals were to advance the field by supplying a technique for precise 3D modeling, resolving existing problems, and furnishing insightful information about traffic patterns and vehicle behavior.

Research in the field that has already been done in the context of this study offers insightful information that advances the project. Although there is not a direct parallel study that perfectly matches the inquiry, the approaches and strategies used in previous research have been helpful. By adding these established techniques, the approach becomes more robust, which emphasizes how important it is to draw on earlier research for methodological foundations and study-relevant insights.

One of the research studies that contributed to the project was conducted by Ranftl et al. on depth estimation. In their paper, “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer” (Ranftl et al., 2020), the authors acknowledged the challenges in depth estimation across various scenarios. They investigated methods to train robust monocular depth estimation models capable of performing effectively in diverse environments. In their research, Ranftl et al. specifically focused on creating a diverse dataset using existing 3D movies. This approach aimed to develop a model capable of performing well in different environments. While their research provides a model that functions across various settings, it is crucial to note that the study was conducted with a focus on monocular depth estimation. While their work inspired and assisted this project, it introduced some challenges during adaptation. For instance, the model performed admirably when objects were close to the camera, but it exhibited limitations in points where objects were far away. This aspect is a noteworthy consideration in the application of their findings to the project.

Another research that had aided this project was Ranftl et al.’s research on “Vision Transformers for Dense Prediction.” In this paper, Ranftl et al. discussed the ability of "dense vision transformers," a novel architectural approach that replaces convolutional networks in dense prediction tasks, to produce predictions that are more globally coherent and finely tuned than those of fully convolutional networks. (Ranftl et al., 2021)

Further research was conducted in this area by Kerbl et al. in a paper titled, “3D Gaussian Splatting for Real-Time Radiance Field Rendering.” In this paper, the authors discuss advancements in 3D scene representation, particularly focusing on the challenges and trade-offs associated with existing methods such as meshes, points, and Neural Radiance Field (NeRF) techniques. They introduce a novel approach that combines the strengths of different representations, presenting a 3D Gaussian representation that allows for optimization with state-of-the-art visual quality and competitive training times. This new approach utilizes a tile-based splatting solution for real-time rendering, achieving high-quality results on various datasets. The authors emphasize the benefits of their method, which leverages anisotropic 3D Gaussians for radiance field representation, an optimization method for Gaussian properties, and a fast, differentiable rendering approach for GPUs, enabling real-time rendering with high quality for novel-view synthesis. The contributions include the introduction of anisotropic 3D Gaussians, an optimization method for their properties, and a GPU-based rendering approach. The experimental results demonstrate the effectiveness of their method, showcasing comparable or superior quality to previous implicit radiance field approaches while achieving real-time rendering capabilities for novel-view synthesis. (Kerbl et al., 2023)

2.1. Objectives/Tasks

In this section, the objectives and tasks of the project will be listed and corresponding intended results of each of them will be mentioned.

1. Capturing photos: Capturing sufficient number of photos was one of the compelling tasks as there were not any sources on the internet of the same scene with multiple angles, thus, the project members had to capture photos on their own.
2. Pre-processing Procedures: The objective here was to carry out image correction and feature detection to prepare the images for 3D reconstruction.
    a. Feature Detection and Matching: In this step of pre-processing procedures, the objective was to perform a satisfactory feature detection. Many algorithms such as SIFT were considered for this step however they did not perform as well as the FAST algorithm, as a result the FAST algorithm was used.
    b. Depth Estimation: Dept estimation proved to be a challenge, the project members first tried to calculate the depths of the object using various formulas however, the result did not satisfy the project's requirements. Thus, MIDAS was used.
3. 3D Reconstruction: The objective of this task was to utilize a photogrammetry library such as OpenCV to create a 3D model of the scene. OpenCV was used as planned as it is the most used and widespread library in this area of work.

2.2. Realistic Constraints

The project has multiple realistic constraints such as cost, environmental constraints, and privacy constraints.

Photogrammetry and 3D modeling are expensive and challenging areas. While there are open-source software programs and libraries available, running the photos is quite costly. During the project, a rather small dataset, a 30-second video, was used to achieve results faster. Despite this, running the project required a significant amount of run time and machines with high-performance processing units. Provided that this project is used with larger datasets, the cost would automatically be higher.

In the beginning of this project, it was anticipated that environmental constraints would affect the project results as the performance could change depending on the quality of images that are captured in different weather conditions. However, environmental constraints did not affect the project as predicted. The pre-trained models used in this project were already trained in various environments, thus, the project did not get affected.

One of the primary constraints for the project revolves around privacy concerns. To ensure image quality, project members had to capture videos on campus. Notably, this process does not raise privacy issues, as the identity plates of cars are not captured, and any images where cars or people can be identified are neither used nor shared.

This project applies multiple engineering and scientific standards. Machine learning can be used for photogrammetry or reality capture, and it is important to adhere to engineering and scientific standards to ensure accuracy and efficiency in the project. One relevant standard is the H.264 compression standard, which is commonly used for video compression and is supported by most web browsers and video players. H.264/AVC uses rate distortion optimization (RDO) technique to get the best coding result in terms of maximizing coding quality and minimizing resulting data bits. (Wu, et al., 2005)

Another engineering standard is ASTM E57, which was created by ASTM International (previously known as the American Society for Testing and Materials). The standard, whose full name is "Standard Specification for 3D Imaging Data Exchange," specifies the content and format criteria for transferring 3D imaging data between various hardware and software systems. (ASTM, 2006)

JPEG and PNG are two examples of image compression standards that are essential to the effective storage and transmission of digital images. Joint Photographic Experts Group, or JPEG, is a widely accepted standard for reducing digital photographs while preserving acceptable image quality. The JPEG method is an excellent option for applications with limited storage space since it can achieve high compression ratios. (International Organization for Standardization, 1994) PNG, on the other hand, is an image compression standard that is also widely used and is renowned for its lossless compression technique. As a result, PNG is a great option for applications that call for high-fidelity image replication since it can compress image data without sacrificing quality. The project can make sure that huge amounts of image data are saved effectively without sacrificing image quality by using these compression standards. (International Organization for Standardization, 2004)

Another standard that will be followed in this project is ISA/IEC 62443 which is a series of standards developed by the International Society of Automation (ISA) that provides a comprehensive framework for cybersecurity in industrial automation and control systems (IACS). This standard is designed to help organizations establish a proactive and systematic approach to protecting their IACS from cyber threats. It includes guidelines for risk assessment, security policies, network architecture, access control, and incident response. The ISA/IEC 62443 standard provides a common language and framework for cybersecurity across different industries and sectors, ensuring a consistent level of protection and resilience in IACS. (International Society of Automation, & International Electrotechnical Commission, 2013)

3. METHODOLOGY

In this section methods and techniques used in reaching the results will be described in detail. In the initial phase of video visualization, the project aimed to extract specific information from a full image context using the YOLOv3 model, optimized for car detection. Key files—coco.names, yolov3.cfg, and yolov3.weights—were crucial. YOLOv3 was trained to recognize various objects, including cars, using coco.names. Yolov3.cfg provided essential information about the model's architecture, and yolov3.weights included pre-trained weights. Due to the significant size of YOLOv3 binaries, YOLOv5 was chosen for efficiency on the Mac platform, focusing on improved vehicle detection accuracy by combining the model with essential files.

With YOLOv5, it streamlined the process by reducing the need to include files in the workspace. Already trained models are stored in an online repository called PyTorch Hub. To directly access the model, the API facilitated by the "torch.hub.load" script was used. Since the pre-trained model exhibited the ability to detect objects outside the scope of the requirements, a refinement process was necessary during the integration of the off-the-shelf model. Labels suitable for the target such as "car", "truck", "bus" were selected from the model's label set. A filtering process was then required to save only those tags for later use. The results of the object detection model, applied by using YOLOv5 can be seen in Figure 1.

After completing the detection part with implementing YOLOv5, it is needed to find the intrinsic and extrinsic parameters of the camera in order to reflect depth and real-world data to the 3D model. (Nielsen, 2020) Intrinsic parameters, also known as internal parameters, are parameters specific to the camera itself, such as focal length and lens distortion. These include the focal length of the view pattern and the keypoint, which refers to the center of the image plane. Fixes lens distortion and image distortion problems. Extrinsic parameters are the parameters used to define the transformation between the camera and the outside world. (Zhang, 2014) These parameters may change with changes in the position or change of the camera. These include the translation vector, which represents the displacement of the optical center of the camera, and the matrix, which does not change the displacement of the camera. The extrinsic matrix combines these contacts for their presence and representation in the camera's world coordinate system.

Starting from intrinsic parameters, it has been observed that common examples of determining intrinsic parameters include the use of objects with known width and length corresponding to the camera angle. The most commonly used technique to calibrate the camera was the Chessboard. (OpenCV, 2014) Along with OpenCV, pre-existing software solutions were used; Among these, the "camera calibration" implementation in MATLAB was assumed to be the most reliable. Using this application made it easier to determine the actual parameters of the camera, as shown in Figure 2.

In the next stage of camera calibration, it became necessary to determine external parameters. Considering that the internal parameters obtained with the chessboard technique are associated with the internal features of the camera, it was thought that these determined parameters should be integrated into the project. This process provided comprehensive calibration that included both internal and external parameters, facilitating an accurate representation of the camera's spatial relationship with the external environment within the scope of the project requirements. (Mathworks, n.d.)

Another method used in the project is the Disparity Map which is a visual representation that shows differences in pixel displacement between corresponding points in stereo images. In the context of computer vision and stereo vision systems, a stereo pair consists of two images taken from slightly different perspectives, simulating the way the human eye perceives depth. Disparity Map is obtained by calculating disparity values for each pair of pixels in stereo images. The disparity represents the apparent shift between the pixel coordinates of the same point in the left and right images. (Radke, 2009) This shift is a direct result of the depth of the scene and the separation between camera perspectives. The depth estimation method began to be used instead of the camera calibration method, as it was more suitable for the purposes of the project. However, due to the difficulties explained in detail in the "Problems Encountered" section, a solution was sought by researching an alternative methodology.

The MIDAS method, a machine learning model that estimates depth from the image, began to be used. First, this model began to split videos into separate frames. Then, similar to the procedure implemented with YOLOv5, the integration of MIDAS's "DPT_Large" model, characterized by its high accuracy despite its relatively slower inference speed, was facilitated in code via PyTorch. (Ranftl et al., 2022) Taking into account the subtle effectiveness of MIDAS models at specific dimensions, the segmented frames were resized without compromising their focal point. In particular, the assignment of pixel values within these frames bounded by a range between 0 and 255 resulted in a type of "depth estimate”. (Ranftl et al., 2022) Given the MIDAS model's natural tendency to assign high values to nearby entities, it became clear that the matrix values should be inverted; where 255 represented closeness and 0 represented distance. A comprehensive analysis revealed that MIDAS demonstrated significant accuracy in estimating depth at close range, however its performance dropped significantly at greater distances. The results obtained from MIDAS can be seen in Figure 3.

A reference point in the form of a nearby object in the camera's field (specifically on the pavement) was used strategically to reduce observed disparities in value allocation at nearby points across successive frames. Assuming that the reference point was specified as X in the current frame, an X * Y calculation was made, giving the estimated true depth value. The determined Y factor was then systematically applied to all pixels within the frame. This process of normalizing the values and bringing them closer to real-world depth values involved creating a new Y factor for the depth matrix with each successive frame. Ultimately, the matrix designed for the video was transferred and applied to the primary module for further analysis and implementation.

After the depth matrix was successfully integrated into the primary module, an evaluation was made by subjecting the integrated system to visualizations. Then the previously developed YOLOv5 module was merged with the core module. Initially, object detection within the frame was performed. The YOLO model provided the center point of the defined object along with length and width values corresponding to the dimensions of the bounding box. Simultaneously, points predetermined through the FAST algorithm between consecutive frames in the core module were kept in a variable called 'X' for use in visualization. After this initial detection, only points falling within the bounding box were retained, creating an effective filtering process. Essentially, only points related to detected objects were preserved for subsequent visualization.

Building upon this refined detection and filtering process, the subsequent visualization stage yielded the intermediate 3D model depicted in the accompanying image. The point cloud data, although of low resolution, marks a critical advancement in the ENS492 Graduation Project, presenting a partially reconstructed vehicle from a single perspective. Despite some scattered data points, this visualization is instrumental in the ongoing development and calibration of object detection and depth estimation algorithms, setting a foundation for precise vehicle modeling in Intelligent Transportation Systems. The result can be seen in Figure 4.

The objective was to enhance the performance of the module, which initially operated on a singular perspective, by incorporating an additional point of view into the visualization process.

To achieve this, corresponding points in images sourced from videos of both angles were aligned directly. However, this methodology failed to produce accurate results. For this reason, scaling, translation, and rotation are performed on the 3D coordinate plane.

First, one of the two-point clouds obtained from the module made for a single angle was rotated 180 degrees on the y-axis. Each point of that single-point cloud was rotated by using the rotation matrix given in Figure 5.

After the rotation process, minimum and maximum points on the x, y, and z axes were detected, a total of 12 in both point clouds, as can be seen in Figure 6 and Figure 7.

Subsequent to the identification of the specified points, three-dimensional bounding boxes, encompassing the relevant regions, were constructed within both point clouds. This is illustrated in Figure 8. The constructed 3D bounding boxes served as a foundational element in the scaling process.

After the scaling process was completed, translation was applied and both point clouds were brought to the same point on the plane. The results obtained can be seen in Figure 9, Figure 10.

4. RESULTS & DISCUSSION

The objective of the project was to advance the development of Intelligent Transportation Systems (ITS) through the generation of precise three-dimensional (3D) models of vehicles utilizing video streams and photogrammetry techniques. Notable enhancements were achieved in object detection through the implementation of YOLOv5 and camera calibration. Nevertheless, challenges arose during the process of cloning binary folders and executing camera calibration procedures. Following that, in order to overcome these challenges, changes were made to the methodology, such as switching from YOLOv3 to YOLOv5 and using depth estimate in place of camera calibration. The project required modifications to its methodology and techniques due to a number of unforeseen obstacles and requirements revisions. These included issues with video quality, operating system compatibility, and the need for alternate approaches for camera calibration and depth estimation. The project is behind schedule due to the encountered problems. Despite significant progress, including the resolving of various technical challenges and the integration of multiple technologies, the project is still ongoing. This research contributes significantly to the field of Intelligent Transportation Systems (ITS) by offering a more effective method for creating 3D models of cars, which may enhance traffic management and road safety. The use of advanced object detection and depth estimation techniques represents a valuable contribution to the field.

In conclusion, the project has advanced significantly toward its objectives, but it has also faced several challenges and is currently behind schedule. The contributions made so far are significant in advancing the capabilities of ITS.

5. IMPACT

There will be major scientific, technological, and socioeconomic effects from this research. By combining advanced object identification and depth estimation technologies—specifically, YOLOv5 and MIDAS—it improves the subject of Intelligent Transportation Systems (ITS) scientifically. This method greatly increases the effectiveness and safety of traffic control systems by making it possible to create 3D models of vehicles with greater precision using video data. From a socioeconomic perspective, this project aims to increase the efficiency of the transportation system by reducing accidents and traffic congestion. The project's unique approach to ITS, which combines state-of-the-art computer vision and photogrammetry techniques, opens new opportunities for commercial applications in traffic monitoring and highway safety improvements.

In addition, the project has no legal obstacles to further development and potential commercialization in the industry, making it a suitable candidate for entrepreneurs in the field of advanced transportation systems.

6. ETHICAL ISSUES

The project uses video sources to create three-dimensional (3D) models of cars for intelligent transportation systems. Two potential ethical problems that need to be carefully considered have been identified by this project. Firstly, if this project or the proposed solution is implemented in real life in the future, either for traffic management or vehicle safety, privacy issues may arise, as video footage is gathered in public places to develop the 3D models. Even though no personal information is intended to be recorded or processed, it is possible that people will be identified from the video footage gathered. Second, people might be concerned about the possibility of abusing the 3D models made using the project's methodology. In addition to traffic analysis, these models can be used for tracking or surveillance.

7. PROJECT MANAGEMENT

The initial project plan provided a detailed plan to create a new approach to Intelligent Transportation Systems (ITS). Photogrammetry methods and the use of multiple video streams were key components to create precise 3D models of the cars. Moving cameras, changing lighting, and handling resource requirements were among other issues the plan focused on solving.

Verifying the feasibility and effectiveness of the proposed method required rigorous testing and adherence to predefined standards.

The team initially used YOLOv3 for object detection and traditional camera calibration methods. However, challenges such as poor video quality and operating system incompatibility led to significant changes. The final plan included the use of YOLOv5 for improved object detection and the MIDAS method for depth estimation, addressing compatibility and efficiency issues.

The team learnt during the project how crucial it is to be flexible in project management, how to overcome unforeseen obstacles, and how valuable it is to use different approaches to accomplish tasks. The experience made clear how important it is to continually assess project plans and make necessary adjustments to achieve success in changing settings.

The final project plan includes using YOLOv5 for improved object detection and the MIDAS method for depth estimation. This plan aims to create more accurate 3D vehicle models for Intelligent Transportation Systems, addressing initial challenges around video quality and operating system compatibility. The team adapted their approach to overcome these challenges, demonstrating the importance of flexibility and continuous evaluation in project management.

8. CONCLUSION AND FUTURE WORK

The successful application of YOLOv5 for enhanced object identification and the usage of MIDAS for depth estimation in 3D vehicle modeling are the two most significant outcomes of the project. The Intelligent Transportation Systems benefit from these developments. The main issues with limitations were compatibility with various operating systems and video quality. The next obvious steps would be to improve these methods, make 3D models more accurate, and maybe expand the application to traffic control in real time. Subsequent research by students may concentrate on refining these techniques in a range of scenarios and incorporating the technology into larger-scale ITS frameworks for thorough traffic analysis and enhanced safety.

9. APPENDIX

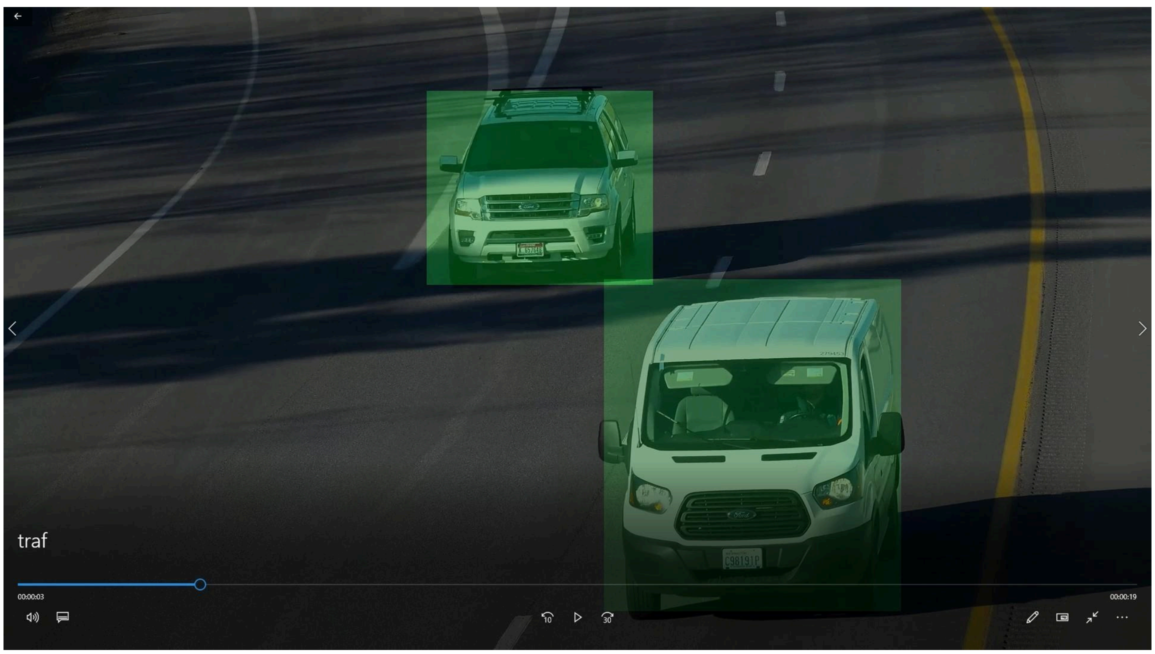

Figure 1: Results of YOLOv5, object detection model.

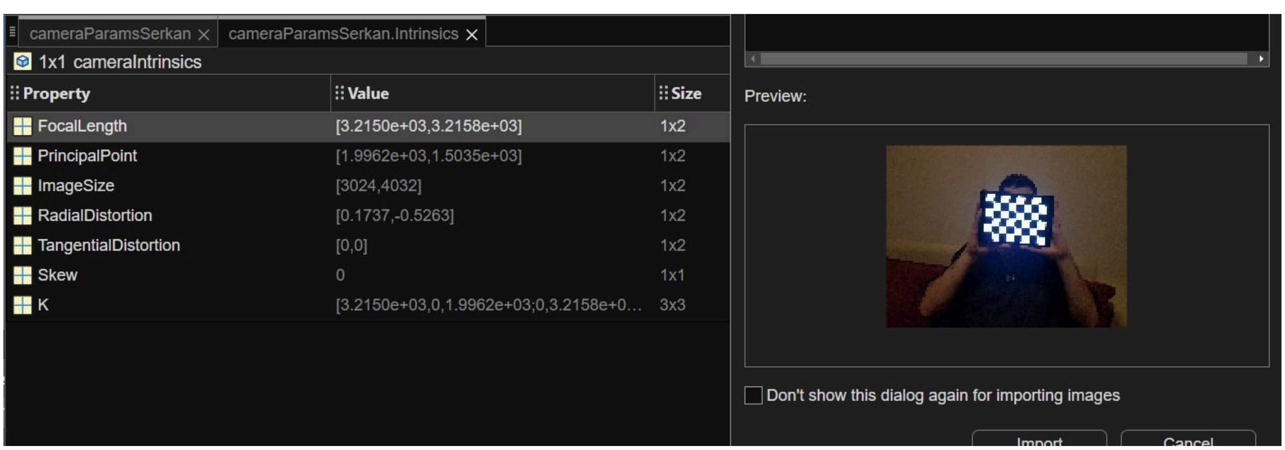

Figure 2: Using the chessboard technique.

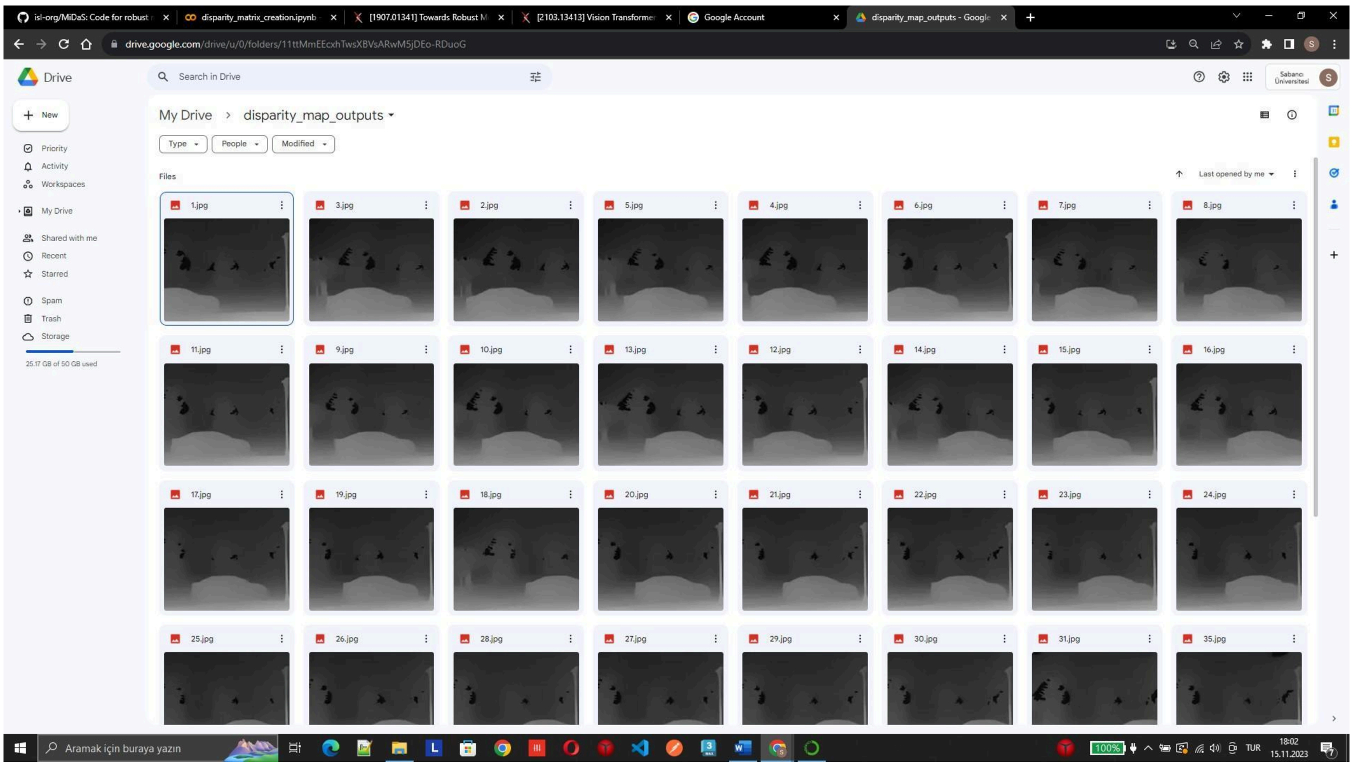

Figure 3: MIDAS results.

Figure 4: Obtained Point Cloud

$$R = \begin{bmatrix} \cos(\pi) & 0 & \sin(\pi) \\ 0 & 1 & 0 \\ -\sin(\pi) & 0 & \cos(\pi) \end{bmatrix}$$

Figure 5: Applied Rotation Matrix 1

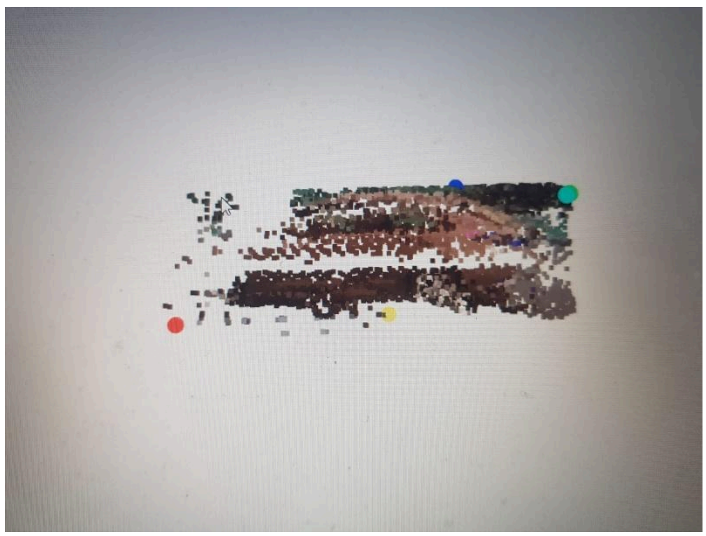

Figure 6: Visualization of extremes (1)

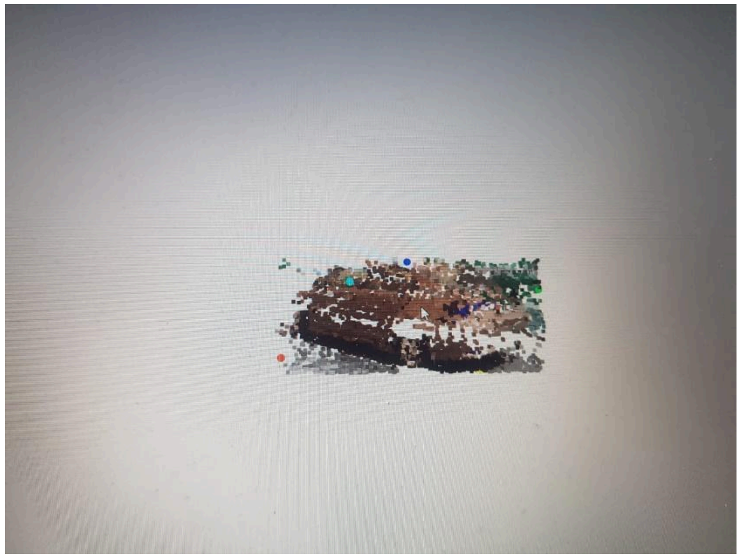

Figure 7: Visualization of extremes (2)

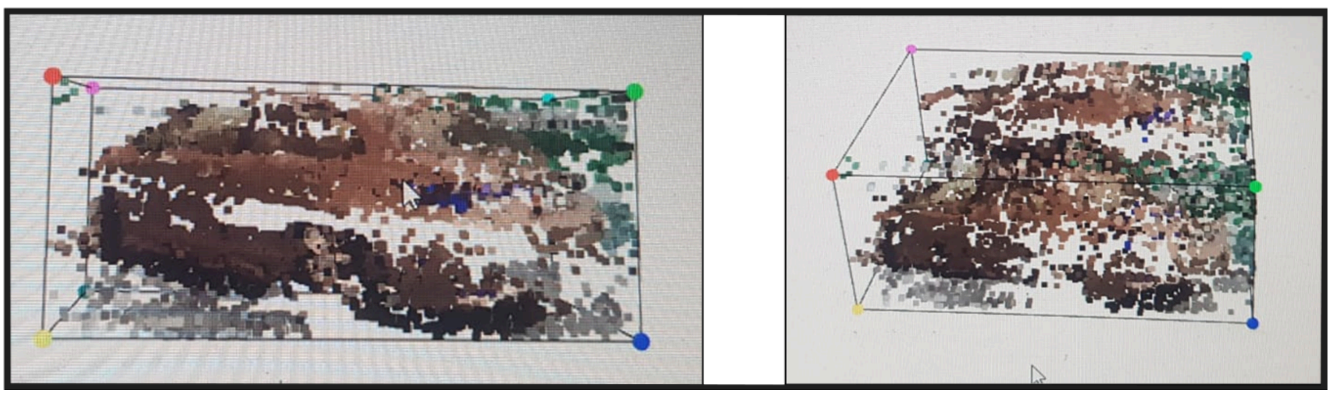

Figure 8: Visual. of 3D bounding boxes

Figure 9: Right Side of Point Cloud

Figure 10: Left Side of Point Cloud

10. REFERENCES

ASTM International. (n.d.). Committee E57 on 3D Imaging Systems. Retrieved April 26, 2023, from https://www.astm.org/get-involved/technical-committees/committee-e57 Ge, Z. (2023). United States Patent No. 11605220.

D. Wu et al., "Fast intermode decision in H.264/AVC video coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 7, pp. 953-958, July 2005, doi: 10.1109/TCSVT.2005.848304.

International Organization for Standardization. (1994). Digital compression and coding of continuous-tone still images: Requirements and guidelines, ISO/IEC 10918-1:1994. https://www.iso.org/standard/18902.html

International Organization for Standardization. (2004). Digital compression and coding of continuous-tone still images: Requirements and guidelines, ISO/IEC 15948:2004. https://www.iso.org/standard/29581.html

International Society of Automation, & International Electrotechnical Commission. (2013) Security for industrial automation and control systems Part 3-3: System security requirements and security levels, ANSI/ISA-62443-3-3 (99.03.03)-2013. From https://www.isa.org/products/ansi-isa-62443-3-3-99-03-03-2013-security-forindu Jiang, X., Zhang, Y., & Chen, Y. (2021). Deep learning-based 3D reconstruction and tracking in dynamic environments. Journal of Advanced Transportation, 2021, 1-17.

Kopf, J., Rong, X., & Huang, J.-B. (2021). Robust Consistent Video Depth Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Lambert, R.B. and Messer, K.J. (2023) “Video surveillance and object recognition.” U.S. Patent No. 20220020255. Washington, DC: U.S. Patent and Trademark Office.

Lou, J.-G. et al. (2005) 3-D Model-Based Vehicle Tracking. Available at: https://www.researchgate.net/publication/3328022_3-D_ModelBased_Vehicle_Tracking

. Lu, L., Liao, C., Zhang, W., & Yu, S. (2018). Multi-camera tracking system for crowd monitoring. Journal of Visual Communication and Image Representation, 56, 1-9.

MathWorks. (n.d.) Camera Calibration. MATLAB Documentation. Retrieved Month Day, Year, from https://www.mathworks.com/help/vision/ug/camera-calibration.html

Messer, K. J., & Lambert, R. B. (2023). United States Patent No. 11568723.

National Highway Traffic Safety Administration. (2017). 2016 fatal motor vehicle crashes: Overview. U.S. Department of Transportation. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456

Neupane, B., Horanont, T., & Aryal, J. (2022). Real-Time Vehicle Classification and Tracking Using a Transfer Learning-Improved Deep Learning Network. Sensors, 22(10), 3813. https://doi.org/10.3390/s22103813

Nielsen, N. H. (February, 2020). ComputerVision: Computer Vision Project. GitHub. https://github.com/niconielsen32/ComputerVision

OpenCV. (2023). About OpenCV. Retrieved from https://opencv.org/about/

Radke, R. J. (2009). Multi-View Geometry for Camera Networks. In H. Aghajan & A. Cavallaro (Eds.), Multi-Camera Networks (pp. 3-27). Academic Press. ISBN 9780123746337. https://doi.org/10.1016/B978-0-12-374633-7.00003-3

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2022). Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(3).

Viola, P., Jones, M.J. Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154 (2004).

https://doi.org/10.1023/B:VISI.0000013087.49260.fb

Szeliski, R., Hsiao, E., Sinha, S., Ramnath, K., Zitnick, C., & Baker, S.

(2015). OBJECT IDENTIFICATION USING 3-D CURVE MATCHING

U.S. Patent No.

9690977. Washington, DC: U.S. Patent and Trademark Office.

Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media.

T.C. Ulaştırma Ve Altyapı Bakanlığı (2020). Available at: https://www.uab.gov.tr/uploads/pages/bakanlik-yayinlari/ulusal-akilli-ulas-imsistemleri-strateji-b elgesi-ve-2020-2023-eylem-plani-eng.pdf.

Zhang, Z. (2014). Camera Parameters (Intrinsic, Extrinsic). In: Ikeuchi, K. (eds) Computer Vision. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-31439-6_152

Zhang, C., Zhang, J., & Chen, C. (2019). Real-time 3D reconstruction for augmented reality applications using multiple cameras and GPUs. Sensors, 19(14), 3152