International Journal on
Drones

Volume 2, Issue 2, 2026
July—December
DOI (Journal): 10.37591/IJD

https://journals.stmjournals.com/ijd

Research IJD

JATAYU: A Stereo Vision–Based UAV for Autonomous Navigation and Human Following in GPS-Denied Environments

Cecilia Dinesh[1,*], Anas Fodkar[2], Amey Shelar[3], Tushar Nandy[4], Nilamabri G Narkar[5]

Abstract

Traditional unmanned aerial vehicle (UAV) systems primarily rely on Global Positioning System (GPS)-based navigation and conventional computer vision techniques. However, these approaches face significant challenges in GPS-denied environments, such as dense urban areas, indoor spaces, and disaster-affected regions where GPS signals may be weak, unavailable, or unreliable. In addition, limitations in real-time processing, object detection accuracy, and environmental perception can reduce UAV effectiveness in dynamic and unpredictable situations. This paper presents JATAYU, a smart autonomous UAV system designed to operate reliably without GPS by integrating deep learning-based human detection with stereo vision for real-time depth estimation. The proposed system employs a dual-camera configuration to estimate the distance between the UAV and detected objects accurately. The perception unit continuously processes visual information and communicates with the flight control system, enabling the UAV to make informed navigation decisions and maintain reliable operation in complex environments. The proposed system achieves an average human detection accuracy of 92.3%, demonstrating its effectiveness in identifying people under challenging operating conditions. The system is specifically designed for high-risk applications where affordability, reliability, and real-time performance are essential. By combining deep learning, stereo vision, and autonomous flight control, JATAYU provides robust perception and navigation capabilities while reducing dependence on GPS infrastructure. Furthermore, the system is optimized for embedded hardware, making it suitable for deployment on resource-constrained UAV platforms. Its lightweight and cost-effective design makes JATAYU a versatile solution for applications such as surveillance, disaster management, search-and-rescue missions, and emergency response. Overall, the proposed UAV demonstrates reliable real-time human detection and navigation capabilities in GPS-denied environments.

Keywords: UAV, STEREO vision, deep learning, human detection, autonomous drone

***Author for Correspondence**
Cecilia Dinesh
E-mail: cecilia14062004@gmail.com

[1,2,3,4]Student, Department of Computer Engineering, Xavier Institute of Engineering, Mahim, Mumbai, India
[5]Faculty Guide, Department of Computer Engineering, Xavier Institute of Engineering, Mahim, Mumbai, India

Received Date: May 26, 2026
Accepted Date: August 04, 2026
Published Date: August 20, 2026

Citation: Cecilia Dinesh, Anas Fodkar, Amey Shelar, Tushar Nandy, Nilamabri G Narkar. JATAYU: A Stereo Vision–Based UAV for Autonomous Navigation and Human Following in GPS-Denied Environments. International Journal on Drones. 2026; 2(2): 30–38p.

INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have emerged as pivotal tools in modern aerial monitoring, infrastructure inspection and search-and-rescue operations due to their rapid deployment and inherent maneuverability. Beyond simple observation, contemporary applications increasingly demand that UAVs perform complex human-centric tasks such as real-time detection, persistent tracking and behavioral monitoring. The change demands a shift from remote-controlled aircraft to full-fledged autonomously controlled vehicles that have multi-functional capabilities within uncertain environments.

One crucial barrier in autonomous operation of drones lies in the extensive use of the GPS system. While GPS offers a standard for global localization it's reliability falters in "GPSdenied" environments such as dense urban "canyons," or regions affected by electronic interference and signal multipath effects. These vulnerabilities pose crucial safety risks and operational limitations, driving the urgent need for robust and perception based navigation alternatives that do not depend on external satellite signals.

Recent literature has proposed various frameworks to bridge this gap in autonomy. The interaction between humans and multi-UAV drone swarm collaboration has been studied for improving the interaction between human operators and UAV swarms [1]. Vision-based techniques, particularly Visual Simultaneous Localization and Mapping (V-SLAM), enable UAVs to understand the environment and calculate ego-motion with cameras mounted on the vehicle [2, 3]. The proliferation of these systems has been supported by benchmark datasets like KITTI, which provide the necessary ground truth for evaluating perception algorithms [4]. Additionally, visual odometry has improved environmental perception capabilities in aerial robotics [5], and radio frequency-based machine learning models have been applied to classify UAVs [6].Despite these advancements, high-performance solutions often necessitate expensive LiDAR sensors or high-end GPU hardware, which are frequently cost-prohibitive or too heavy for lightweight, low-cost platforms.

Stereo vision presents a compelling alternative for the 3D scene understanding. By calculating the disparity between the synchronized frames from dual cameras these systems can re-construct spatial data and estimates distances in real time[8, 9]. However commercial stereo sensors often remain expensive.

In this paper, we introduce JATAYU, an intelligent, multi-purpose UAV platform engineered for autonomous monitoring and human-following. JATAYU utilizes a custom made and cost effective stereo vision system consisting of synchronized Raspberry Pi modules.This system is designed such that it generates highly accurate depth data at much lower costs than the industrial-grade sensors. By integrating stereo depth estimation with deep learning-based human detection, the platform achieves reliable autonomous navigation and persistent monitoring even in GPS denied scenarios.

The primary contribution of this work is the design and implementation of a scalable, budget friendly UAV architecture that harmonizes stereo vision with intelligent perception providing a robust solution for autonomous operations in challenging real world environments.

METHODS

The JATAYU system follows a modular architecture which is designed to enable perception driven autonomous monitoring using a low cost stereo vision system. This system involves an architecture for real-time sensory inputs, onboard computation, and flight control within a data-processing pipeline. This system is composed of three major blocks including perception, onboard processing, and flight control.

Perception System

The perception subsystem uses a custom stereo camera consisting of two Raspberry Pi camera modules which is mounted on a rigid bracket. The stereo configuration enables realtime depth perception by computing disparity between synchronized image pairs. Compared to commercial stereo sensors, this design provides a low cost solution while maintaining reliable depth estimation.

To ensure accurate stereo matching, the cameras are aligned with a fixed baseline and synchronized using a clock-based triggering mechanism. The synchronization accuracy is approximately 1 ms, as shown in Figure 1.The stereo matching algorithm used is SGBm with WLS filter.

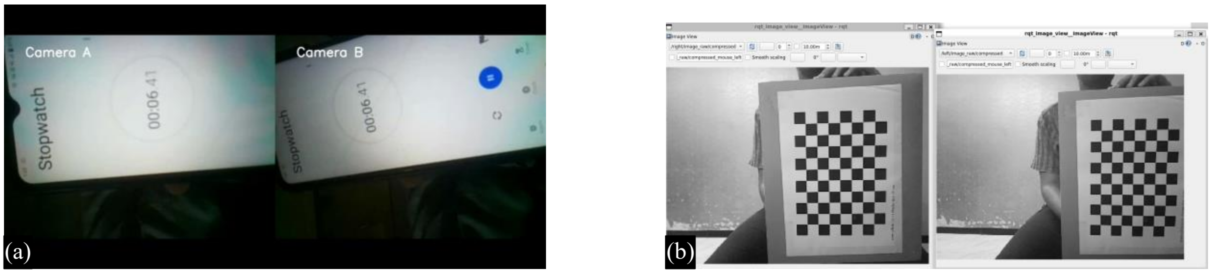


Figure 1. Stereo camera synchronization used for simultaneous image capture The control input is:(a) Clock-Based Synchronization (b) ROS Monitoring

Onboard Processing

Perception and decision-making tasks are executed entirely on a Raspberry Pi 5 onboard computing unit, with no reliance on external servers or cloud processing during flight operations.

The distance error is defined as:

$$e_d = Z - Z_{ref} \tag{1}$$

where Z represents the measured depth and Z_{ref} is the desired following distance. A proportional controller generates the forward velocity command.

$$v_x = k_d \cdot e_d \tag{2}$$

Software Framework

The system is built on the Robot Operating System (ROS) for inter-module communication. OpenCV is applied to stereo vision computations, while Nanodet is employed for object detection and MAVLink serves for communicating with the flight controller.

Hardware Implementation

The UAV platform incorporates the stereo vision system, onboard computing unit and flight controller in a compact and lightweight package for autonomous flight. During operation, stable sensor mounting and reliable flight performance is ensured by the hardware configuration. Summarized in Table 1 are the detailed specification of the hardware and systems.

Human Detection

Real-time human detection is performed using NanoDet-Plus (m-416 variant), a lightweight anchor-free single-stage object detector designed for deployment on resource-constrained embedded hardware. NanoDet-Plus was selected over heavier alternatives such as YOLOv5 or Faster R-CNN due to its significantly lower inference time on ARM-based processors, making it well-suited for the Raspberry Pi platform. The model was deployed using weights pretrained on the COCO dataset, which contains 80 object categories including the person class. On the standard COCO validation benchmark, NanoDet-Plus achieves a mean Average Precision (mAP) of 20.6 across all 80 categories.

Table 1. Hardware and system specifications.

Component	Specification
UAV Frame	F450 quadrotor frame
Flight Controller	Pixhawk (ArduPilot firmware)
Onboard Computer	Raspberry Pi 5
Camera	Raspberry Pi Camera Module v2 × 2
Stereo Baseline	62.9 mm
Communication	MAVLink over UART
Flight Time	Approximately 20 min

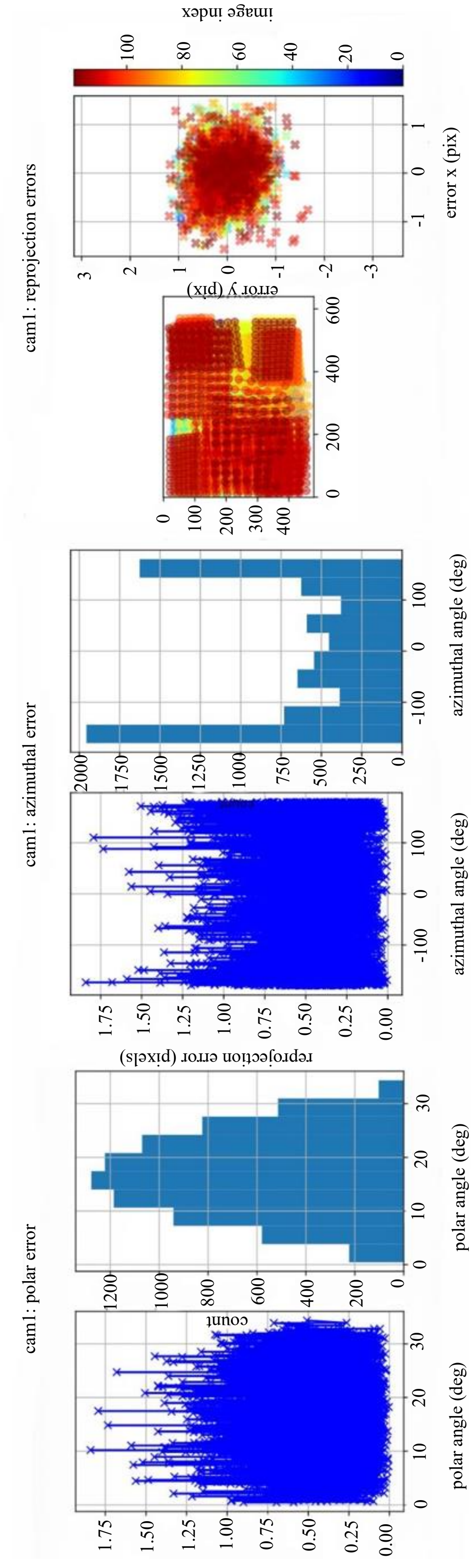


Figure 2. Camera calibration results using checkerboard pattern

Stereo Camera System

Before computing disparity, rectification of incoming stereo image streams is performed using the calibrated intrinsic and extrinsic parameters. The Semi-Global Block Matching (SGBM) algorithm is applied to generate disparity maps and the Weighted Least Squares (WLS) filtering is used to smoothen the disparity map and reduce artefacts at object edges. The WLS filter takes the disparity map and the confidence map from the left and applies the one to the other to get a smooth, dense disparity map with sharp disparities at the object boundaries. It is a two-phase pipeline, SGBM and WLS post-filtering, that dramatically increases the quality of depth map in particular for the textured outdoor surface that is present during testing.

Depth estimation is computed using:

$$z = \frac{fB}{d} \quad 3$$

Calibration and Validation

Stereo calibration was performed using a checkerboard pattern. The calibration parameters are summarized in Table 2 and the calibration results are shown in Figure 2.

Each camera follows the pinhole projection model:

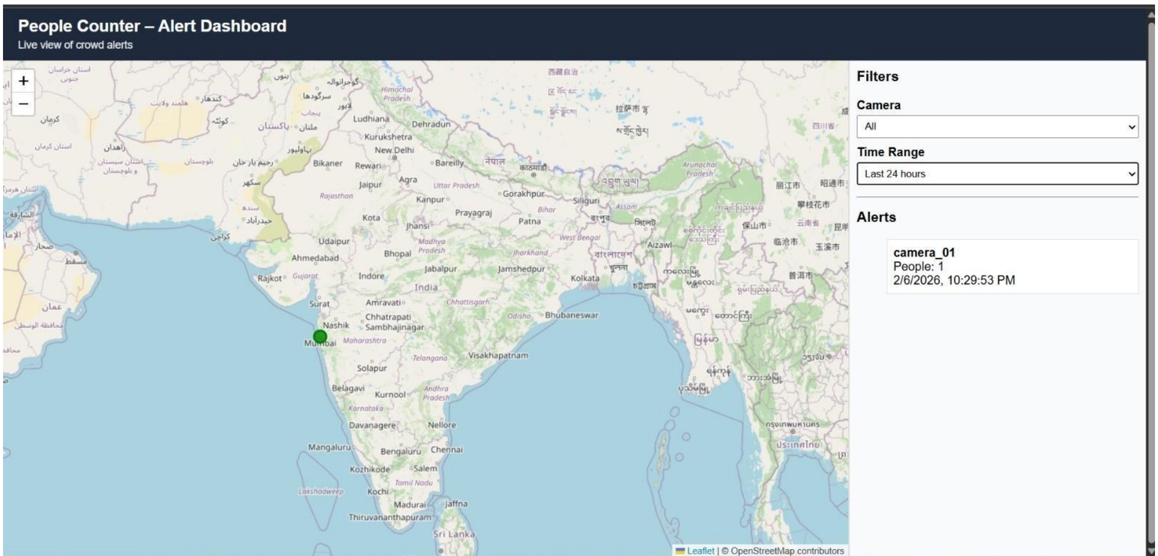
$$uv1 = \mathrm{KRt}XYZ1 \quad (4)$$

System Monitoring

A web-based dashboard is used for real-time telemetry monitoring and system tracking, as shown in Figure 3. It displays key flight parameters such as altitude, speed, and battery status, enabling efficient supervision and control during operation.

Table 2. Stereo camera calibration parameters.

Parameter	Value
Calibration Target	Checkerboard
Checkerboard Size	10×7
Square Size	0.02 m
Stereo Baseline	0.0629 m
Cam0 Focal Length	586.07 px
Cam1 Focal Length	590.58 px
Average Reprojection Error	≈0.37 px



(a)

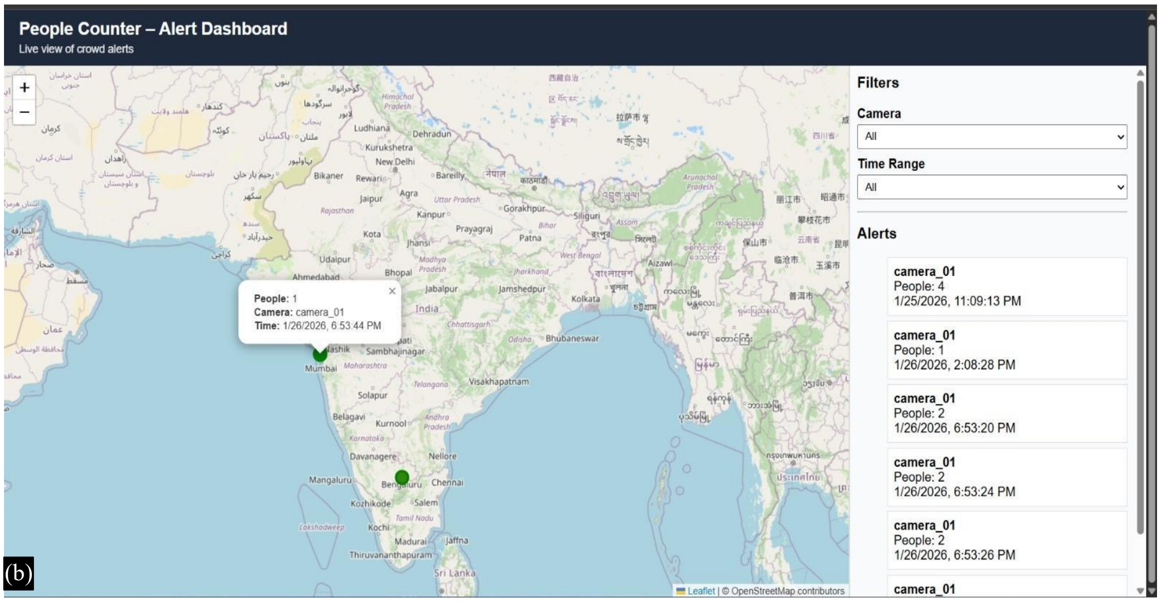


Figure 3. Web-based dashboard for UAV telemetry and real-time tracking (a) Interface (b) Tracking.

Table 3. Experimental setup.

Parameter	Value
Environment	Open-field outdoor environment under natural daylight
Detection Target	Human targets
Evaluation Metric	Detection accuracy, depth error
Test Frames	500
Depth Range Tested	0.64 m – 4.0 m
Detection Model	NanoDet-Plus (COCO pretrained)
Stereo Algorithm	SGBM + WLS filtering

Experimental Setup

The proposed UAV system was tested in an open field, outdoor environment, in natural daylight conditions to evaluate the UAV system in realistic operating conditions. The human target experiments were performed to evaluate the robustness of the detection, tracking, and depth estimation modules in the autonomous operation system, in which targets moving with different patterns and distances from the camera were used. The UAV platform was equipped with a stereo vision system, onboard processing and autonomous flight control to carry out real-time perception and following human flight in dynamic outdoor scenarios. The entire UAV platform for experiments is presented in Figure 4 and the details of the experimental setup and performance metrics are presented in Tables 2 and 3, respectively.

RESULTS AND DISCUSSION

Detection Performance

The AI perception module is based on the lightweight NanoDet object detection network [7] for real-time human detection and counting. The detection performance remained stable with different crowd density and motion patterns under the outdoor operating conditions.

It is crucial to note that the COCO mAP score is the general multi-class detection performance for all object classes in the COCO. The detection accuracy for peds of 92.3% reported in this study, on the other hand, is a measurement performed specifically for this class under controlled outdoor daylight environment that mimics the UAV operational environment. Thus, the two measurements cannot be compared directly. No further optimization and the pretrained model was tested in real world conditions.

Figure 4. JATAYU UAV platform used for experimental evaluation.

Table 4. Stereo vision performance evaluation.

Metric	Value
Depth Estimation Error	± 7 cm
Minimum Measurable Distance	0.64 m
Camera Frame Rate	15 FPS
FPS After Processing	7 FPS
Depth Map Density	87%
Processing Delay	21.64 ms

Depth and Following Performance

Stereo vision provides consistent depth estimation thus enabling smooth trajectory control and stable autonomous following behavior.

Monitoring Performance

The system generates alerts when predefined crowd thresholds are exceeded. Alerts include GPS coordinates, timestamps, drone identifiers and people counts, enabling real time monitoring through the networked platform.

Table 4 shows that the proposed system achieves an accuracy of depth estimation of around ±7 cm, and a processing speed of 7 frames per second (FPS), sufficient for real-time human-following applications.

The results demonstrate that low cost hardware with AI based perception and stereo vision enables reliable autonomous people following and real time monitoring with good scalability through cloud integration. However, its shortcomings are low computational capabilities on board, lighting effects, and network dependence. Future work involves edge AI acceleration, adaptive networking strategies and multi-drone coordination for large-scale monitoring.

CONCLUSION

This work presents JATAYU an AI based UAV system for human following and crowd monitoring. The system integrates deep learning based human detection with a low cost stereo vision setup to enable effective autonomous operation in real world environments.

This system is well equipped with features such as real-time monitoring and cloud-based data visualization. Therefore it can be used for surveillance and crowd control applications. Overall, the results demonstrate that reliable autonomous following drone and crowd monitoring can be achieved using cost effective hardware and AI driven perception.

Future work will focus on improving navigation using visual SLAM for GPS-denied environments, enhancing computational efficiency through edge AI deployment and enabling coordinated multi-UAV systems for large-scale monitoring tasks.

Acknowledgments
We thank the Department of Computer Engineering of Xavier Institute of Engineering for their support and for providing the infrastructure required to conduct our project.

Ethics And Informed Consent Statement
This study did not involve clinical trials or sensitive human data. All UAV experiments were conducted in compliance with local regulations in open environments, and no institutional ethics approval was required. All individuals within the UAV's field of view were informed about the study and provided verbal informed consent. No personally identifiable, facial, or biometric data were collected, and all data were used solely for research purposes.

Data Availability Statement
This study uses the publicly available COCO dataset for training and evaluation. The experimental UAV data, including stereo images, calibration parameters, and detection logs, are available from the corresponding author upon reasonable request.

Author Contributions Statement
- Cecilia Dinesh: Conceptualization, Software, Writing – Original Draft, Project Administration.
- Anas Fodkar: Methodology, Investigation, Software Integration, System Integration, Validation.
- Amey Shelar: Investigation, Data Collection, Visualization.
- Tushar Nandy: Data Curation, Formal Analysis, Software.
- Nilambari G. Narkar: Supervision, Writing – Review & Editing, Resources.

Conflict of Interest
The authors have indicated that they have no competing financial interests nor personal relationships which could have appeared to bias the work reported in this paper. This research has been done while pursuing the academic project at the Department of Computer Engineering, Xavier Institute of Engineering, Mumbai, India. No external funding, sponsorship or commercial support was provided for the design, development or evaluation of the JATAYU system. The authors do not have any financial interest in the hardware components, software frameworks or tools employed in this study.

REFERENCES
1. Barr H, Levy D, Rosenfeld A, Maksimov O, Kraus S. Advising agent for supporting human–multi-drone team collaboration. arXiv. 2025;arXiv:2502.17960. doi:10.48550/arXiv.2502.17960.
2. Mur-Artal R, Montiel JMM, Tardós JD. ORB-SLAM: A versatile and accurate monocular SLAM system. IEEE Transactions on Robotics. 2015;31(5):1147–1163. doi:10.1109/TRO.2015.2463671.
3. Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM. In: European Conference on Computer Vision (ECCV); 2014. p. 834–849.
4. Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2012. p. 3354–3361.
5. Wang J, Gao X, Shen S. Visual odometry and mapping for aerial robots. Robotics and Autonomous Systems. 2017;87:118–134. doi:10.1016/j.robot.2016.10.015.

6. Huynh-The T, Pham Q-V, Nguyen T-V, da Costa DB, Kim D-S. RF-UAVNet: High-performance convolutional network for RF-based drone surveillance systems. IEEE Access. 2022;10:1618–1630. doi:10.1109/ACCESS.2021.3139555.
7. Liu Z, Ding K, Xu Q, Song Y, Yuan X, Li Y. Scene images and text information-based object location of robot grasping. IET Cyber-Systems and Robotics. 2022. doi:10.1049/csy2.12049.
8. Achtelik M, Weiss S, Siegwart R. Stereo vision and UAV navigation in GPS-denied environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2011. p. 1–6.
9. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision. 2002;47(1–3):7–42. doi:10.1023/A:1014573219977.

