The Twenty-Seventh International Symposium on Artificial Life and Robotics 2022 (AROB 27th 2022),
The Seventh International Symposium on BioComplexity 2022 (ISBC 7th 2022),
The Fifth International Symposium on Swarm Behavior and Bio-Inspired Robotics 2022 (SWARM 5th 2022)
ONLINE, January 25-27, 2022

Stereo Vision-based Object Detection and Depth Estimation from 3D Reconstructed Scene for an Autonomous Multi Robotic Rescue Mission

Jackrit Suthakorn[1†*], Mayur Kishore[1], Songpol Ongwattanakul[1], Fumitoshi Matsuno[2], Mikhail Svinin[3], and Branesh M. Pillai[1]

[1]Center for Biomedical and Robotics Technology (BART LAB), Faculty of Engineering, Mahidol University, Thailand
(* Correspondence to: jackrit.sut@mahidol.ac.th)
[2]Department of Mechanical Engineering and Science, Kyoto University, Japan
[3]College of Information Science and Engineering, Ritsumeikan University, Japan

Abstract: The 3D reconstruction of objects and depth extraction is inevitable for autonomous surveillance and mobile robot navigation in Safety, Security, and Rescue Robotics (SSRR) missions, which can be executed with the help of high-end sensors and cameras which elevates the overall operation cost. When it comes to hazardous unknown environment missions the risk factor of using these delicate sensors are at high risk, in which the malfunction of any of these is indispensable. Considering these aspects this project entails creating a stereo vision system comprising 2 low-cost HD cameras for SSRR that can extract depth from 3D reconstructed scenes and analyze performance by cross verifying with distance estimated by heat map generation. The reconstructed 3D view is used to estimate the depth of the target of interest. Depth-based object extraction is performed using binary masking and thresholding. The results are compared and assessed against the original distance. On evaluating the obtained distance with the location of Unmanned ground vehicle (UGV) the coordinates of the target are established, thereby allocating UGV with the target coordinates. When compared with the practical experiments it was found that the evaluated distance had an accuracy of more than 90% with the real-time distance.

Keywords: Rescue robotics, Safety security rescue robotics (SSRR), Unmanned aerial vehicle (UAV), Unmanned ground vehicle (UGV), Depth estimation.

1. INTRODUCTION

The primary objective of Safety security rescue robotic missions is based on identifying and determining the location of the target. The heterogeneous robotic operations are usually held with the help of UAV, UGV and USV [1-3]. When compared to UGV and USV, UAV always possess an upper hand in the overall visibility factor of the scenario. As a result, in the majority of SSRR operations initial map generation and localization [4] is executed with the help of UAV. By employing various technologies like image fusion, visual slam, stereo imaging etc. with the help of high-end sensors and cameras clear 3D maps of the surroundings and coordinates of the objects can be estimated with high accuracy [5].

Image fusion (IF) is a rapidly developing field that combines images acquired from diverse sensors to create an informative picture that may be used to make decisions. Integrating diverse photos improves the analytical and perceptual picture quality. Appropriate image fusion can preserve crucial information by extracting all relevant data from photos without causing discrepancies in the final image. Simultaneous localization and mapping (SLAM) is the computing issue of creating or upgrading a representation of an unexplored area while maintaining information of the targets position at the same time [6, 7]. Visual SLAM is a technique in which camera is incorporated with SLAM is a beneficial way to enhance the reliability rate, especially in large-scale outside environments [8, 9].

Stereo imaging is a passive technology that matches elements observed in distinct photographs of the same area to reconstruct the environment's topology. The output of this computation will be the 3D point cloud where each 3D point corresponds to a pixel in one of the images [10].

Laser range finders, visual sensors such as stereo vision Real sense, Kinect RGB-D camera, and time of flight camera are more suitable than other types since they capture accurate information of the surroundings [11]. Besides great measurement accuracy and availability, they have several drawbacks, such as high cost and maintenance, installation, and high computational cost, which ultimately results in the overall mission cost. When it comes to low-end distance sensors such as ultrasonic sensors, infrared (IR) sensors the level of accuracy has to be sacrificed. In this research we adapted the technology of stereo vision imaging with the evaluation of images obtained from normal HD cameras for obtaining a significantly accurate and economically viable solution for constructing the 3D image [12] of the environment and to detect the location of target. For the replication of rescue scenario NIST standard based arenas were constructed. The overview of the research is shown in Fig.1.

2. METHODOLOGY

The primary goal of deploying robots in emergency rescue missions is to improve the security and efficiency of rescue responders working in dangerous or difficult situations. The National Institute of Standards and Technology (NIST) have created a standardized testing procedures [13] to explain how consistently the robot can navigate the specified sorts of environments in a reasonably significant fashion, giving rescue operators enough assurance to decide the robot's relevance [14]. The output information obtained in this test procedure

† Jackrit Suthakorn, is the presenter of this paper.

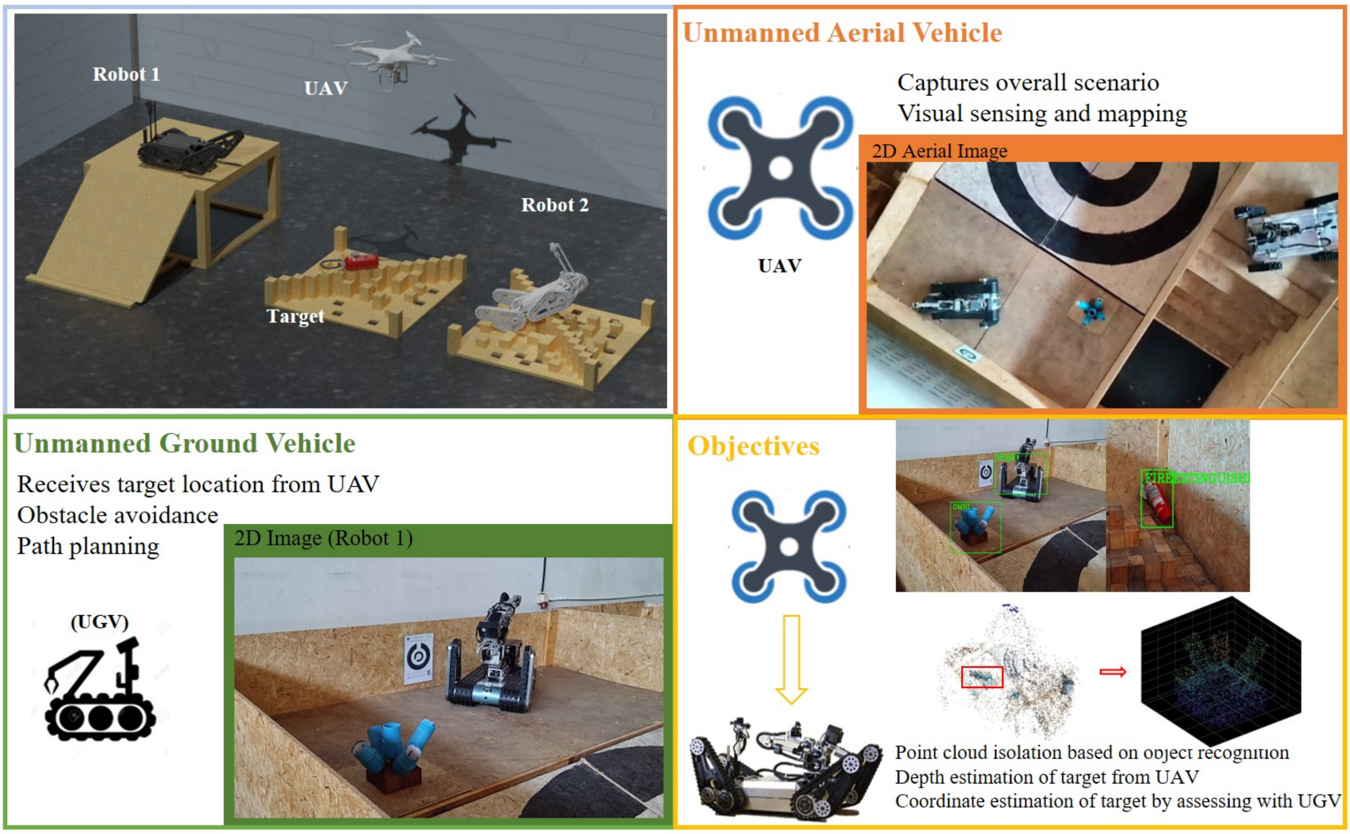

Fig. 1. Overview of stereo vision-based object detection and depth estimation

reflects the abilities of the rescue robot. The output factors include parameters like time taken to accomplish the task, distance travelled, obstacle avoidance, effectiveness in communication, ability of maneuvering in different terrains, etc.

The next important step is the calibration of cameras. Due to the use of two the case of two cameras the chances of shift, oscillation or shake are high. Therefore, the concept of camera calibration plays a vital role in transforming the 2D image to a real world 3D image. The main goal of calibration is to estimate the intrinsic and extrinsic parameters of the camera for the transformation calculations. Extrinsic parameter describes the location of cameras coordinate system within the world coordinate system. Intrinsic parameters are utilized for the mapping from 3D camera coordinates to 2D image coordinates, accurate camera calibration also helps to increase the accuracy of depth calculation. Stereo system consists of two cameras hence compared to the conventional calibration it is important to calibrate orientation of second camera to that of the first camera. Followed by the calibration, the robots are employed to the testing arena.

During these testing conditions the operators controlling the robots will be completely isolated from the testing field such that the visibility and sound factors are completely avoided. Ones the task commence several images of the events are captured by the UAV which will act as the input parameter. The images are then rectified to remove the distortions and horizontally align the image pairs. After rectification, both picture planes have the same y value. Then a disparity map is created by matching every pixel of the images

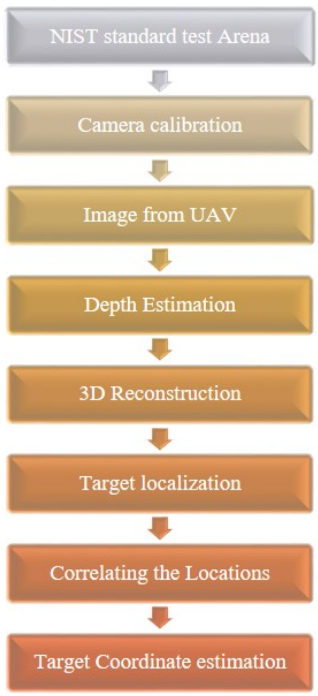

Fig. 2. Methodology for 3D reconstructed scene for autonomous multi robots

from both the cameras and compute the distance between pixel values. With the help of disparity map and the stereo parameters the 3D version of the image is reconstructed in the point cloud. Flow chart for the methodology section is

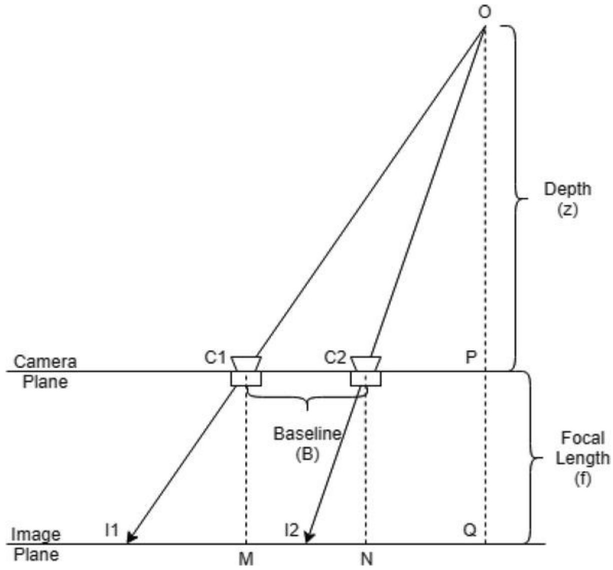

Fig. 3. Model for depth estimation [15]

shown in Fig.2.

The model of depth estimation is represented in Fig.3 depth of object from the stereo unit can be written as eq.1

$$Z = \frac{fb}{x_L - x_R} = \frac{fb}{d} \quad (1)$$

where,
Z- Depth of the object point from the stereo unit in cm
b-Baseline distance between the 2 camera units in cm
d- Pixel disparity
$(x_L - x_R)$- disparity between corresponding left and right image points

To measure the distance between two objects from the image, first select two pixel points from each object, by estimating the euclidean distance between these pixels the real- time distance can be calculated. For measuring the distance D_{pix} (eq.2) between the pixels of the image, consider two points a and b, of coordinates (x_1, y_1) and (x_2, y_2)

$$D_{Pix}(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

For calculating the real time distance R_D,

$$R_D = \frac{2.54}{96} P_R H_R D_{Pix} \quad (3)$$

$$H_R = \frac{Z}{D_c} \quad (4)$$

where,
R_D- real distance from UAV image
P_R- Pixel ratio
H_R- Ratio of UAV height to calibrated position
f- Effective focal length of the stereo unit in pixels
O- Object point in the world frame
C_1, C_2- Camera 1 and camera 2
I_1, I_2- Corresponding image from camera 1 and camera 2

By correlating the values including coordinates of ground robot and UAV along with the distance between ground robot with the target, the height of the UAV, coordinates of the target can be estimated. After projecting these coordinates and distances to a 2D plane, with the help of euclidean distance formula the coordinates of the target within the region of interest is determined.

3. RESULTS

The heterogeneous robots are deployed to the NIST standard arena for conducting the experiment. Among this, UAV possessing the highest visibility factor captures the images of the scenario. The selected pair of images for conducting the experiments is shown in Fig. 4

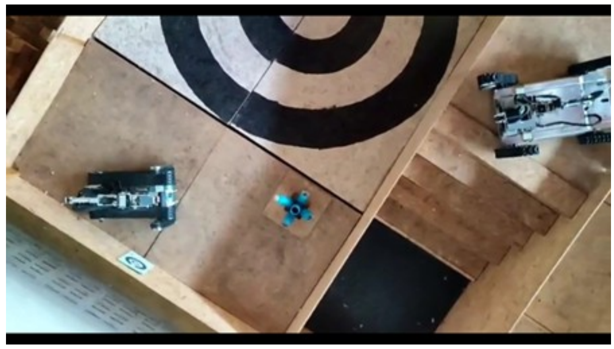

Fig. 4. Image for estimation

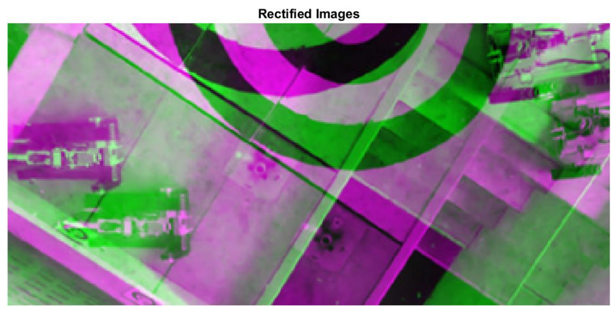

Fig. 5. Rectified image

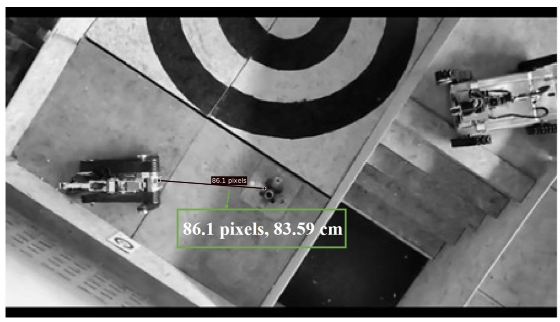

Fig. 6. Pixel distance estimation between two points

The height at which the UAV is being operated is figured out with the help of depth estimation as shown in eq.1 and it was found out to be 3415.95 cm. Fig.5 was obtained during the experimental procedure. The points for the distance calculations are selected from the image. Pixel distance value between the first robot and the target was estimated as 86.1pixels from the eq.2. The corresponding real-world distance was obtained from eq.3 and the value is 83.59 cm as shown in Fig.6. The accuracy of the distance was obtained

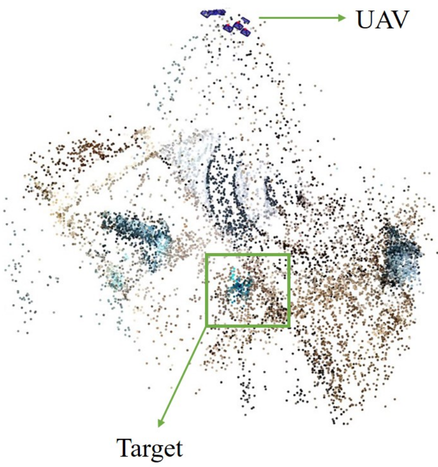

Fig. 7. Point cloud isolation of UAV image based on object recognition

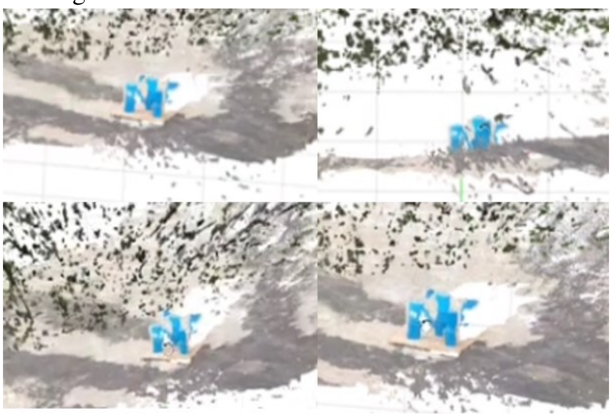

Fig. 8. 3D reconstructed image

by comparing it with the real-time measured value and was found to be 94.76%. The point cloud isolation of UAV image constructed from the UAV is as shown in Fig.7. The 3D point cloud image of the target is reconstructed and is shown in Fig.8. Thereby accomplishing successful replication of the proposed method from heterogeneous robots.

Fig. 9. Target image from ground robot

The point cloud estimation from the UGV was also obtained in order to validate the method by comparing the accuracy of results. The image of the target obtained from ground robot is shown in Fig.9. The disparity Map generated is shown in Fig.10 and the point cloud estimation of the image was constructed as shown in Fig.11 and an accurate replication of the target was achieved.

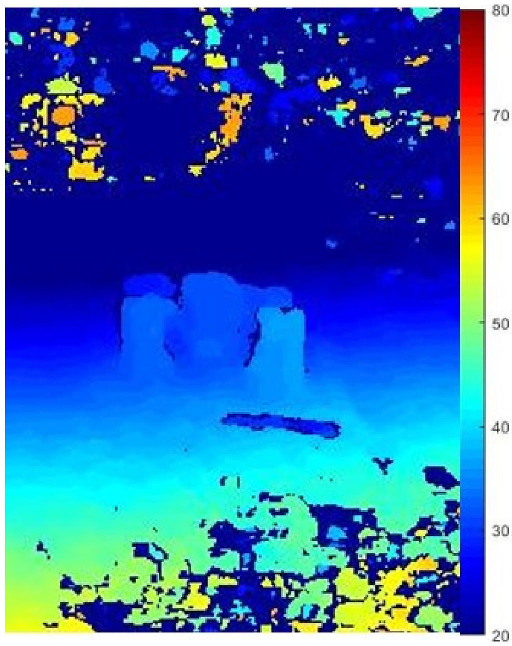

Fig. 10. Disparity Map

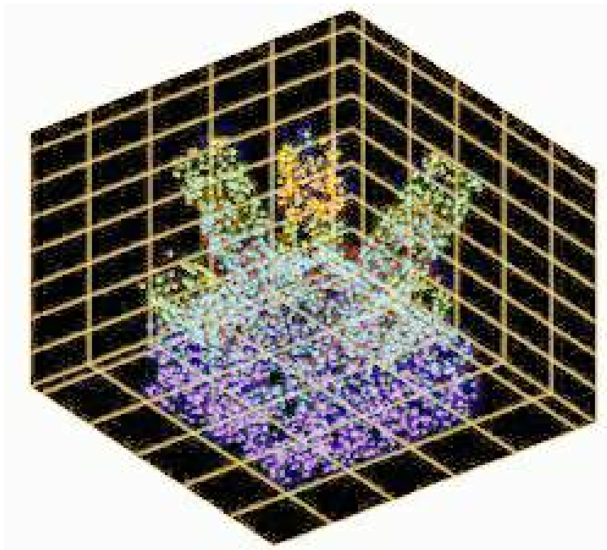

Fig. 11. Point cloud isolation of target based on object recognition

4. CONCLUSION

The domain of 3D image reconstruction and localization is very vast and still have tremendous number of research gaps especially for the reduction in computational cost, error elimination methods and localization of objects. The 3D reconstruction of a 2D image with the stereo vision technology have been executed, thereby allowing to give an outline of the physical properties of the target to the robots along with the coordinate determination of the target object. When compared with the real time dimensions it was found that the estimated value showed a maximum error of 10%. The experimental results illustrates that the proposed system can be induced in the real-time rescue missions incorporated with heterogeneous robots and can play a vital role in reducing the overall operation cost for SSRR missions.

5. ACKNOWLEDGEMENT

This research is partially supported by the e-ASIA Joint Research Program (P-19-50869) Grant through National Science and Technology Development Agency (NSTDA) and Mahidol University, Thailand.

REFERENCES

[1] Dimitrios Chatziparaschis, Michail G Lagoudakis, and Panagiotis Partsinevelos. Aerial and ground robot collaboration for autonomous mapping in search and rescue missions. *Drones*, 4(4):79, 2020.

[2] Samira Hayat, Evşen Yanmaz, Christian Bettstetter, and Timothy X Brown. Multi-objective drone path planning for search and rescue with quality-of-service requirements. *Autonomous Robots*, 44(7):1183–1198, 2020.

[3] Daniel S Drew. Multi-agent systems for search and rescue applications. *Current Robotics Reports*, pages 1–12, 2021.

[4] Md Shah Alam and Jared Oluoch. A survey of safe landing zone detection techniques for autonomous unmanned aerial vehicles (uavs). *Expert Systems with Applications*, page 115091, 2021.

[5] Adrián Bañuls, Anthony Mandow, Ricardo Vázquez-Martín, Jesús Morales, and Alfonso García-Cerezo. Object detection from thermal infrared and visible light cameras in search and rescue scenes. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 380–386. IEEE, 2020.

[6] Arda SEZEN and Çigdem TURHAN. A systematic mapping review of image annotation studies for obtaining information retrieval from images. *Bilişim Teknolojileri Dergisi*, 13(4):423–434.

[7] Hongling Wang, Chengjin Zhang, Yong Song, and Bao Pang. Master-followed multiple robots cooperation slam adapted to search and rescue environment. *International Journal of Control, Automation and Systems*, 16(6):2593–2608, 2018.

[8] Hongling Wang, Chengjin Zhang, Yong Song, Bao Pang, and Guangyuan Zhang. Three-dimensional reconstruction based on visual slam of mobile robot in search and rescue disaster scenarios. *Robotica*, 38(2):350–373, 2020.

[9] Ramil Safin, Roman Lavrenov, and Edgar Alonso Martínez-García. Evaluation of visual slam methods in usar applications using ros/gazebo simulation. In *Proceedings of 15th International Conference on Electromechanics and Robotics" Zavalishin's Readings"*, pages 371–382. Springer, 2021.

[10] Wenbang Deng, Kaihong Huang, Xieyuanli Chen, Zhiqian Zhou, Chenghao Shi, Ruibin Guo, and Hui Zhang. Rgb-d based semantic slam framework for rescue robot. In *2020 Chinese Automation Congress (CAC)*, pages 6023–6028. IEEE, 2020.

[11] Shangmeng He, Ziyuan Tong, Guobin Ma, Mingyun Fan, Ling Lingzhou, and Shoufeng Tang. Research on stereo vision matching algorithm for rescue robot. In *2017 International Conference on Robotics and Automation Sciences (ICRAS)*, pages 35–38. IEEE, 2017.

[12] Wei-Chung Chen, Shuuichi Ihara, and Makoto Hasegawa. Proposal of a rescue operation support system based on 3d reconstruction, gps, and digital pen. In *International Workshop on Advanced Imaging Technology (IWAIT) 2020*, volume 11515, page 115152I. International Society for Optics and Photonics, 2020.

[13] Zhuming Bi. Overview of testing platform for development of integrated robotic systems at nist. In *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 899–904. IEEE, 2019.

[14] Kaya Turgut and Burak Kaleci. A pointnet application for semantic classification of ramps in search and rescue arenas. *International Journal of Intelligent Systems and Applications in Engineering*, 7(3):159–165, 2019.

[15] Satyarth Praveen. Efficient depth estimation using sparse stereo-vision with other perception techniques. *Coding Theory*, page 111, 2020.

