1. Łukasz MACIURA[1], 2. Dariusz WÓJCIK[1,2], 3. Michał MAJ[1,2], 4. Dariusz MAJEREK[1], 5. Bartłomiej Kiczek[1]

Netrix S.A., R&D Centre, Lublin, Poland (1), University of Economics and Innovation, Lublin, Poland (2)
ORCID: 1. 0000-0001-8657-3472; 2. 0000-0002-4200-3432; 3. 0000-0002-7604-8559; 4. 0000-0002-0035-7187; 5. 0000-0003-2834-4000

doi:10.15199/48.2023.02.32

# A stereo vision system on Jetson device using deep learning

***Abstract***. *This paper presents a stereo vision system created on a Jetson device with a GPU. Two cameras attached to the cover are connected to the device. Before using the system, the calibration procedure should be done. The stereo vision system uses two deep learning algorithms (one for disparity map extraction and the second for panoptic segmentation) with custom preprocessing and postprocessing stages. Results from both algorithms were used to calculate point clouds for every registered object.*

***Streszczenie.*** *W artykule przedstawiono system stereowizyjny stworzony na urządzeniu Jetson z GPU. Do urządzenia podłączone są dwie kamery przymocowane do obudowy. Przed użyciem systemu należy przeprowadzić procedurę kalibracji. System stereowizyjny wykorzystuje dwa algorytmy głębokiego uczenia (jeden do ekstrakcji mapy rozbieżności, a drugi do segmentacji panoptycznej) z niestandardowymi etapami wstępnego i końcowego przetwarzania. Wyniki z obu algorytmów posłużyły do obliczenia chmur punktów dla każdego zarejestrowanego obiektu (**System widzenia stereo na urządzeniu Jetson wykorzystujący głębokie uczenie**).*

**Keywords:** stereo vision, Jetson, deep learning.
**Słowa kluczowe:** stereowizja, Jetson, głębokie uczenie.

## Introduction

Classical stereo vision systems use calibrated dual cameras to obtain 3D information. After the calibration process [1, 2] necessary for further image rectification and distortion movement, the disparity map could be generated using stereo-matching algorithms. There are multiple stereo matching algorithms available which could be divided into two groups: classical algorithms (i.e. Semi Global Box Matching algorithm [4] belongs to the Box Matching category [3]) and deep learning solutions like PSMNet (Pyramid Stereo Matching Network) algorithm [5].

Many optimization methods are used for analysis or image reconstruction [6-15]. The deep learning PSMNU algorithm [16] for disparity map extraction was used in the created system. The Panoptic FPN deep learning algorithm [17] from the detectron2 library [18] was used for object detection and segmentation. The segmentation results of the Panoptic FPN algorithm need correction and improper objects removement which could be realized by custom postprocessing stage using additional information from the disparity map. Finally, the point clouds are extracted for each object in the scene using information from panoptic segmentation together with a disparity map of the scene.

## Calibration of cameras

Proper disparity map extraction requires calibrated cameras. Precisely, the calibration of cameras in the stereo vision system allows for further removal of the radial distortion [1] from images and stereo rectification [1,2] in order to produce a stereo pair where epipolar lines in both images are colinear and parallel to scanned horizontal lines in images (canonical stereovision system).

The calibration requires a set of unique stereo pairs with special checkerboards placed in different positions and angles. In order to perform proper calibration, about 100 stereo pairs with scenes containing special chessboards in different positions are necessary. After collecting images, the pair of chessboard points sets are obtained from both images in stereo. The most popular algorithm for chessboard point extraction is based on Hessian [19]. Before the extraction, the sharpening filtration is performed using the convolution images with a $3x3$ sharpening Laplace mask.

After the extraction of chessboard corner points, these points are analyzed. Sometimes the chessboard points extraction algorithm detects these points in both images in stereo pairs in reverse order. In the beginning, the directions of both point sets are checked, and if the direction is different for corresponding points in the stereo pair (as in Fig. 1), then the points set in the second image are reverted. One of four possible directions is recognized by the difference between the first and last point extracted by the algorithm implemented in the OpenCV library [20] (the algorithm detects points row by row).

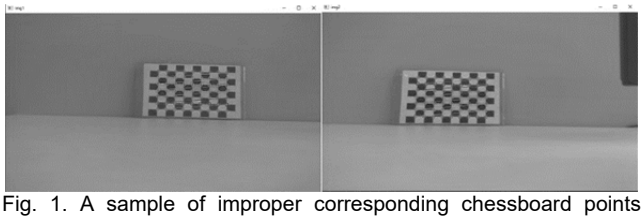

Fig. 1. A sample of improper corresponding chessboard points extracted

If the order of corresponding points is correct, then the outstanding distance of points from rows and column lines designated from all chessboard points for each point is designated for each stereo pair separately. The stereo pair is rejected if the maximal outstanding distance is bigger than the threshold (as in the sample presented in figure 2).

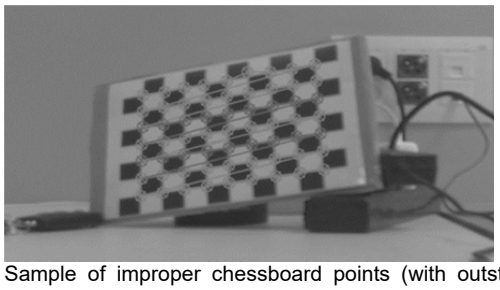

Fig. 2. Sample of improper chessboard points (with outstanding one point in the first row)

After the set of proper corresponding chessboard points and lists are extracted, the proper calibration stage can be performed. The stereo–calibration algorithm based on fundamental matrix calculation [1,2] is implemented in the OpenCV library [20]. After the calibration stage, the special maps for stereo image transformation are prepared.

## Stereo pair rectification and preprocessing

After rectifying the stereo–pair based on maps obtained during the calibration stage, the stereo pair has to be preprocessed because the rectification produces images

with the wrong area (black background). Therefore, the thresholding operation is performed at the beginning to designate the proper image area in stereo pair images (Fig. 3).

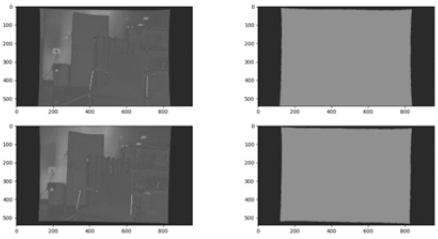

Fig. 3. Rectified stereo pair (on the left) and thresholded rectified stereo pair (on the right)

After thresholding, the common proper area based on both images in the stereo pair is computed by AND logical operation performed on thresholded rectified images from the stereo pair (Fig. 4).

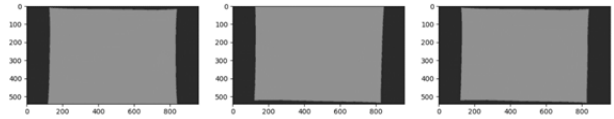

Fig. 4. Designation of common proper area (on the right) based on the left (on the left) and right (in the center) camera image

After computing a common proper area like the right image in Fig. 4, the border lines could be designated in the following way:

a) The central row and column from the common proper area image are extracted, and the amount of nonzero elements is calculated,

b) The left index is designated by analysis of subsequent columns. Searching of the left index is stopped when the analyzed column has at least 90% of nonzero elements amount of central column,

c) The right index is designated similarly but by decreasing the index starting from the right column,

d) The top index is designated by analysis of subsequent rows and comparing the number of nonzero values with the number of nonzero values of the central row (threshold was also set to 90%).

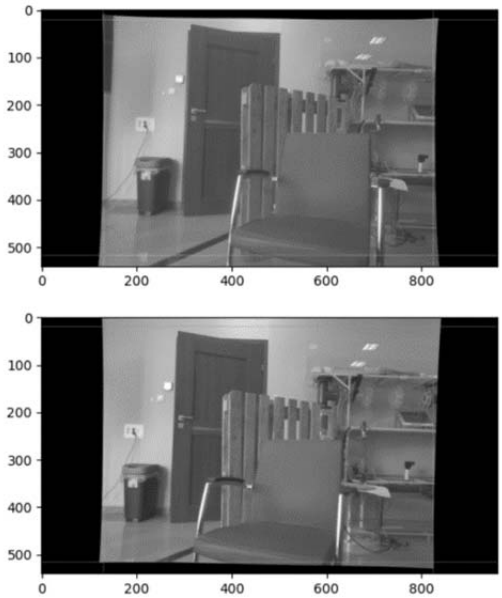

Fig. 5. Using designed parameters for cutting proper image area in the left and right image in the stereopair

e) The bottom index is designated similarly to the top but by decreasing the index starting from the bottom row. Finally, designated indices are used for rectangle proper image area designation (as we can see in Fig. 5), and these parts of images are cut out as input images for further processing by deep learning algorithms.

**Disparity map generation**

The disparity in stereo vision [2] is the difference in the horizontal coordinates from stereo pair images:

$$\Delta_x = x_2 - x_1$$

where $x_2$ is the horizontal position of a point on the right image from stereo pair while $x_1$ is the horizontal position of a point on the left image from stereo pair.

There are many disparity map calculation algorithms based on rectified stereo pairs of images. Classical algorithms such as Box Matching [3] and Semi Global Box Matching [4] use image processing techniques to match points in pairs by scanning images along corresponding epipolar lines. Nowadays, significantly better performance is achieved by deep learning algorithms based on features from convolutional neural networks, as in PSMNU [16], which is based on the PSMNet model [5].

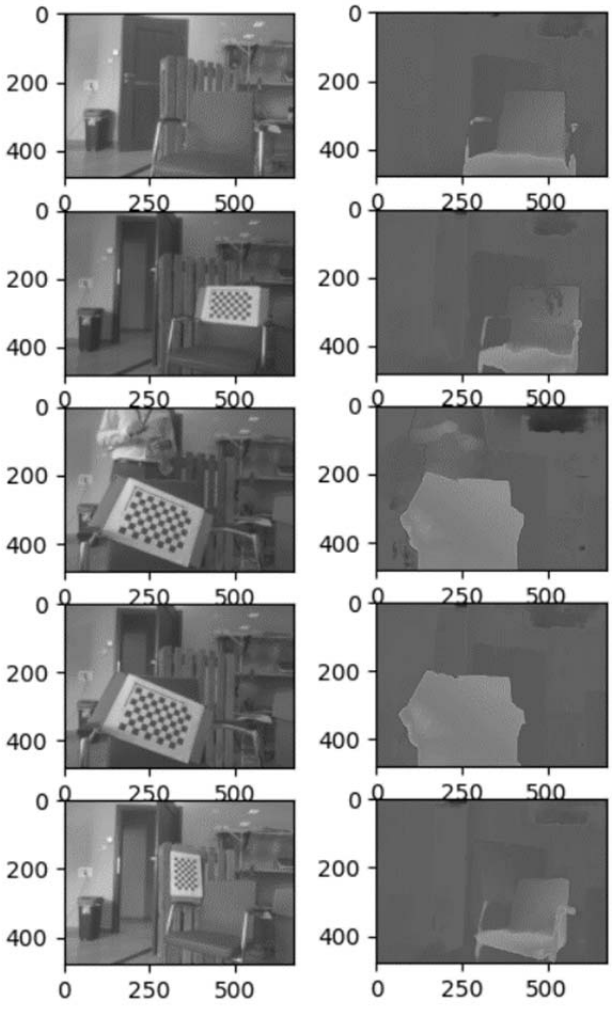

Fig. 6. Left images from stereo pair (on the left) and the results of disparity maps using PSMNU algorithm (on the right)

The review of different solutions that could be used on Jetson device led us to choose a PSMNU model with modified parameters. The maximal disparity value was set to 192 (the bigger value needs more GPU memory which exceeds the limits of a portable device).

The preprocessed left images from stereo pair (left images) and the sample disparity maps using the PSMNU algorithm (right images) are presented in figure 6.

**Objects detection and extraction**

To extract objects from the scene, the Panoptic Feature Pyramid Network algorithm [17] implemented in the detectron2 library [18] (from Facebook research) was used. The panoptic FPN deep learning solution allows for accurate object extraction and segmentation with significantly better results than the Mask R-CNN algorithm [21], most known for this purpose. The used model segments all objects in the scene, including the background. Sample results of panoptic segmentation are presented in Fig. 7.

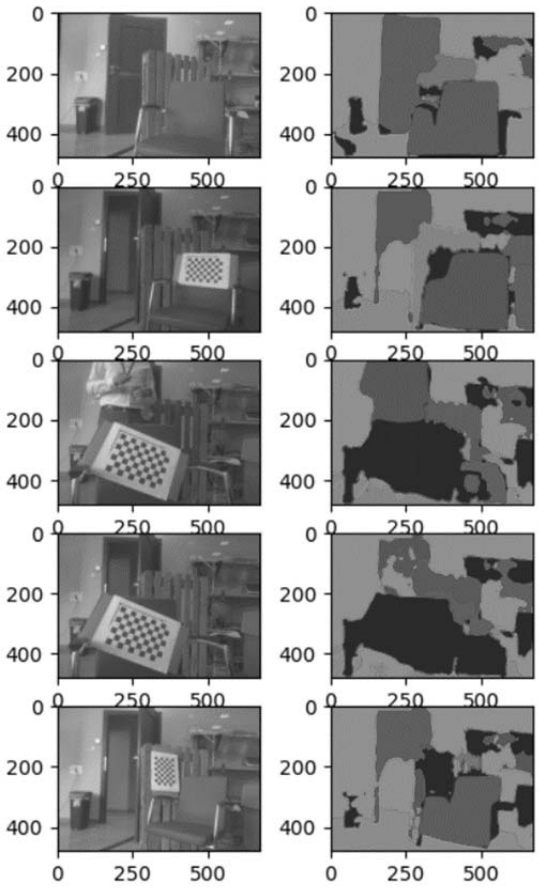

Fig. 7. Sample Panoptic FPN segmentation results (on the right)

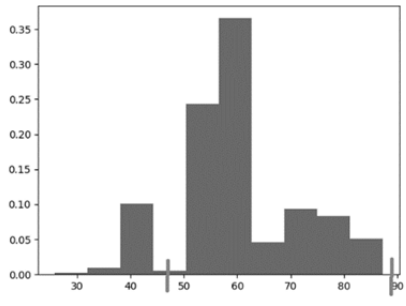

Fig. 8. Sample histogram thresholding result from step 1b

**Postprocessing of objects segmentation**

The designed postprocessing contains the following stages:

1) Wrong fragments of objects movement not belonging to them on the basis of disparity map histogram:
   a. Extraction of disparity map for each object
   b. Calculation of disparity histograms and designation of thresholds automatically for proper disparity values for a given object on the basis of histograms (Fig. 8). The histogram indices (left one and right one) starts in the place of histogram peak and are changing (increasing and decreasing respectively) while are bigger than given threshold obtained experimentally. After stopping changing, the indices indicate the range of disparity values for a given object.
   c. Wrong parts of objects removement on the basis of previously designated thresholds
2) Recalculation of disparity histograms for each object and parallel processing of those histograms:
   a. Disparity map extraction for each object after processing in the previous step
   b. Calculation of disparity histograms for each object
   c. Verification in each histograms ranges which object gave higher value and assignment of object index into this histogram range
   d. Creation of a new image with objects indices and saving indices of objects in places related to disparity values in ranges assigned to given objects designated in the previous step
   e. During the last steps, the situation that the given object could be erased from segmentation results (its points will be assigned to other objects with bigger histogram values in the particular ranges) may occur – then the numbers of indices will be reorganized the following numbers 1..n will respond to really existing objects where n is the amount of the current object.

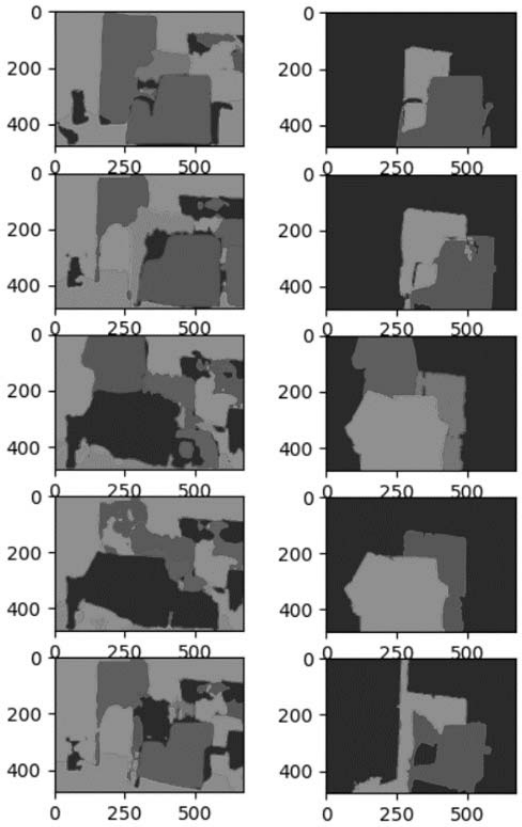

Fig. 9. Samples of final post processing of panoptic segmentations

1) Extraction of binary image for each object and its parts (sic.) indexation. When an object consists of many separate sub-objects, only greater sub–objects are saved, and all other sub-objects of the given object will be removed from the result of segmentation.

2) Background objects movement – these objects which have contact with image borders will be removed. After the removement the object's indices are reorganized in a similar way as in step 2e.
3) Removement of improper objects on the basis of the Blair – Bliss coefficient (wrong objects often have very complicated shapes and higher Blair – Bliss coefficient value. All objects with a coefficient greater than the specified threshold will be removed after removement object indices are reorganized in a similar way as in step 2e.

The sample results of the panoptic segmentation postprocessing algorithm are presented in figure 9 (right column) with the original panoptic segmentation (on the left column) for comparison.

**Points clouds computation**

Based on disparity, the depth [2] could be computed using the following equation:

$$Z = \frac{bf}{\Delta_z}$$

where $b$ is base distance between cameras, $f$ is the focal length and $\Delta x$ is disparity. After depth calculation, the remaining 3D coordinates [2] could be computed:

$$X = \frac{uZ}{f} \quad Y = \frac{vZ}{f}$$

where *(u, v)* pair is the 2D position of a point in one of the images from a stereo pair and *(X, Y, Z)* is the position of the 3D point.

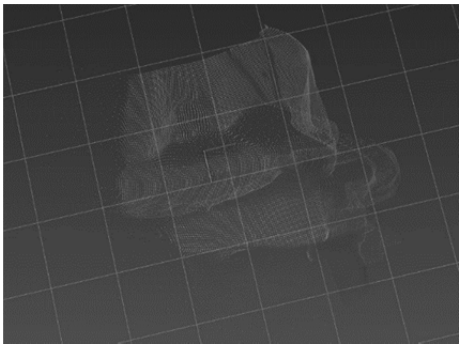

Fig. 10. Sample of generated points cloud on the basis of the segmented human object and its disparities

Finally, separate disparity maps are extracted for each segmented object, and 3D point clouds are computed based on object disparity maps. The sample point cloud for one object is presented in figure 10.

**Conclusion**

The paper presents a novel stereo vision system with dual cameras implemented in a portable Jetson device with GPU using deep learning algorithms.

The created system enables calibration of cameras, rectification (using available deep learning algorithms with custom preprocessing and postprocessing stages to obtain disparity map), extraction of objects in the scene and finally, computation of a 3D point cloud for every object.

In the system, two ready deep learning algorithms were used, but the scientific value of this paper relies on joining algorithms into one system and using custom preprocessing and postprocessing algorithms in order to improve the results of the used algorithms.

The system was implemented using Python language and among others, PyTorch and detectron2 libraries.

***Authors**: dr inż. Łukasz Maciura, Netrix S.A, R&R Centre, Lublin, Poland, E-mail: lukasz.maciura@netrix.com.pl; dr inż. Dariusz Wójcik, Netrix S.A., University of Economics and Innovation, Netrix S.A., R&D Centre, Lublin, Poland,, E-mail: dariusz.wojcik@netrix.com.pl; mgr inż. Michał Maj, University of Economics and Innovation, Netrix S.A., R&D Centre, Lublin, Poland, E-mail: michal.maj@netrix.com.pl; dr Dariusz Majerek, Netrix S.A., R&D Centre, Lublin, Poland, dariusz.majerek@netrix.com.pl; mgr Bartłomiej Kiczek, Netrix S.A., R&D Centre, Lublin, Poland, E-mail: bartlomiej.kiczek@netrix.com.pl;*

REFERENCES

[1] Hartley R., Zisserman A., Multiple View Geometry in computer vision, Cambridge University Press, 2003
[2] Cyganek B. Siebert J., An Introduction to 3D Computer Vision Techniques and Algorithms, John Wiley and Sons, 2009
[3] Zhang M., Experimental Implementation of Stereo Matching Algorithms in Halide, MIT, 2016
[4] H. Hirchmuller, Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Trans. Pattern Analysis and Machine Intelligence, 2008
[5] Chang, J. Ren and Chen, Y. Sheng, „Pyramid Stereo Matching Network", Procedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018
[6] Rymarczyk T., Kłosowski G., Hoła A., Sikora J., Tchórzewski P., Skowron Ł., Optimising the Use of Machine Learning Algorithms in Electrical Tomography of Building Walls: Pixel Oriented Ensemble Approach, Measurement, 188 (2022), 110581.
[7] Koulountzios P., Rymarczyk T., Soleimani M., Ultrasonic Time-of-Flight Computed Tomography for Investigation of Batch Crystallisation Processes, Sensors, 21 (2021), No. 2, 639.
[8] Kłosowski G., Rymarczyk T., Niderla K., Rzemieniak M., Dmowski A., Maj M., Comparison of Machine Learning Methods for Image Reconstruction Using the LSTM Classifier in Industrial Electrical Tomography, Energies 2021, 14 (2021), No. 21, 7269.
[9] Rymarczyk T., Król K. Kozłowski E., Wołowiec T., Cholewa-Wiktor M., Bednarczuk P., Application of Electrical Tomography Imaging Using Machine Learning Methods for the Monitoring of Flood Embankments Leaks, Energies, 14 (2021), No. 23, 8081.
[10] Majerek D., Rymarczyk T., Wójcik D., Kozłowski E., Rzemieniak M., Gudowski J., Gauda K., Machine Learning and Deterministic Approach to the Reflective Ultrasound Tomography, Energies, 14 (2021), No. 22, 7549.
[11] Kłosowski G., Rymarczyk T., Kania K., Świć A., Cieplak T., Maintenance of industrial reactors supported by deep learning driven ultrasound tomography, Eksploatacja i Niezawodnosc – Maintenance and Reliability; 22 (2020), No 1, 138–147.
[12] Gnaś, D., Adamkiewicz, P., Indoor localization system using UWB, Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12 (2022), No. 1, 15-19.
[13] Styła, M., Adamkiewicz, P., Optimisation of commercial building management processes using user behaviour analysis systems supported by computational intelligence and RTI, Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12 (2022), No 1, 28-35.
[14] Korzeniewska, E., Krawczyk, A., Mróz, J., Wyszyńska, E., Zawiślak, R., Applications of smart textiles in post-stroke rehabilitation, Sensors (Switzerland), 20 (2020), No. 8, 2370.
[15] Sekulska-Nalewajko, J., Gocławski, J., Korzeniewska, E., A method for the assessment of textile pilling tendency using optical coherence tomography, Sensors (Switzerland), 20 (2020), No. 13, 1–19, 3687.
[16] Hu Y., Zhen W., Scherer S., Deep – Learning Assisted High – Resolution Binocular Stereo Depth Reconstruction, presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA), 2019.
[17] Kirillov A., Girshick R., He K., Dollar P., Panoptic Feature Pyramid Networks, arXiv.org > cs > arXiv:1901.02446v2, 2019
[18] Y. Wu, et al., Detectron2, https://github.com/facebookresearch/detectron2, 2019
[19] Yu Liu, Shuping Liu, Yang Cao, Zengfu Wang, „A practical algorithm for automatic chessboard corner detection", ICIP 2014
[20] The OpenCV Reference Manual, http://opencv.org
[21] He K., Gkioxari G., Dollr P., Girshick R., Mask R-CNN arXiv:1703.06870