

# The impact of environmental factors on depth mapping from stereovision camera

*Wpływ czynników środowiskowych na mapowanie głębi z kamery stereowizyjnej*

**Abstract.** This article presents an experimental study on the impact of lighting and camera lens contamination on depth map computation in stereovision systems. Utilizing the Intel RealSense D435i camera, images were captured under various lighting conditions and with varying levels of lens dirtiness. The results demonstrate that lighting significantly affects accuracy of disparity computation and hence depth map accuracy, with maximum values of average errors exceeding 30 cm. The study underscores the importance of considering these factors in stereovision applications.

**Streszczenie.** Artykuł przedstawia eksperymentalne badanie wpływu oświetlenia i zanieczyszczenia soczewek kamer na dokładności wyznaczania głębi w systemach stereowizyjnych. Z wykorzystaniem kamery Intel RealSense D435i zarejestrowano obrazy w różnych warunkach oświetlenia oraz przy różnym stopniu zanieczyszczenia soczewek. Wyniki pokazują, że oświetlenie znacząco wpływa na dokładność obliczania dysparycji, a w konsekwencji, skutkuje błędami w wyznaczaniu mapy głębi z maksymalnymi wartościami błędów średnich większymi niż 30 cm. Badanie podkreśla znaczenie uwzględnienia tych czynników w zastosowaniach stereowizyj.

**Keywords:** stereovision, lighting conditions, camera lens contamination, depth map accuracy, disparity computation

**Słowa kluczowe:** stereowizja, warunki oświetlenia, zanieczyszczenie soczewek, dokładność wyznaczenia mapy głębi, obliczanie dysparycji

## Introduction

Stereovision systems represent a crucial tool in the field of three-dimensional scene reconstruction. Their application spans a wide range of domains, from medicine to industry and robotics [1, 2, 3, 4]. Stereovision enables the extraction of depth and spatial context from two-dimensional images. However, the effectiveness of these systems can be significantly compromised depending on lighting conditions and occurring disturbances, such as lens dirt or lighting non-uniformities.

Stereovision systems find applications in devices placed in controlled environments, such as body scanners in cabins, where environmental conditions are regulated and constant [2, 3], as well as in the systems employed in robots and industrial applications [4], aids for the visually impaired [5, 6], etc., where the environmental conditions may undergo dynamic changes. In the context of systems for the visually impaired, improving the effectiveness of 3D scene reconstruction can have a significant impact on users' quality of life, providing them with a more precise and comprehensive perception of their surroundings.

In this work we report on a developed stereovision-based Electronic Travel Aid (ETA) intended to help visually impaired individuals in navigating their surroundings [6]. We conducted tests with visually impaired participants and we identified notable fluctuations in the accuracy of 3D scene reconstruction across varying lighting conditions and camera disturbances. While we recognized the influence of lighting on reconstruction effectiveness, we deliberately pursued an investigation into the precise extent to which different factors impact scene reconstruction quality. This observation emphasizes the critical need for comprehensive research and adjustment of stereovision systems to address a wide range of environmental conditions.

## Stereovision

Stereovision is a passive technique for measuring the depth of a three-dimensional scene using at least two images of the same scene captured from different positions in space [7, 8]. Among its main advantages are its operation in natural lighting conditions (passive method), the possibility of using low-cost digital cameras, and the existence of effective algorithms for determining the coordinates of points in the scene mapped in stereoscopic images. The main

disadvantages include the high computational cost associated with scene reconstruction, the need for camera calibration and synchronization, and the inability to reconstruct areas of the scene occluded by objects and regions of uniform brightness (non-textured).

Calibration is a crucial step in stereo vision systems, as it ensures that the cameras are accurately aligned to capture the scene. This involves adjusting the internal and external parameters of each camera, including focal length, principal point, distortion coefficients, and the relative pose of the cameras with respect to each other. The canonical camera arrangement, where the cameras are positioned at a fixed distance (base) and their optical axes are parallel, is a fundamental concept in stereo vision systems. Once calibrated, the images from both cameras can be rectified to a common coordinate system, which is essential for accurate depth estimation and feature tracking.

Point matching is the fundamental step in reconstructing 3D scene geometry. It involves finding points in stereoscopic images that correspond to the same point in the scene. Information about this relationship is referred to as a depth map or a disparity map. If the disparity is determined for all points in the scene (where possible), we call it a dense disparity map. If the disparity is determined only for a certain group of points, such as characteristic points, we refer to it as a sparse disparity map [7, 8] (Fig. 1).

The operation of algorithms in this group involves assigning to each point in one of the images, called the reference image, a corresponding point in the other image, which represents the same point in the scene. Effective and easily implementable methods involve block matching. For each pixel in the reference image, an  $N \times N$  block (where  $N$  is an odd number) of pixels surrounding it is selected, and then a search is conducted for the corresponding block in the second image based on a chosen similarity measure.

Block-based methods are extensively utilized in stereovision applications due to their simplicity and effectiveness. A comprehensive review of prominent block-based disparity estimation methods reveals several key approaches. The Sum of Absolute Differences (SAD) algorithm minimizes absolute differences between pixel intensities (as depth map estimation typically relies on monochromatic cameras), while Sum of Squared Differences (SSD) computes disparities by minimizing squared intensity

differences [8], albeit with sensitivity to variations and noise. Adaptive window sizes and hierarchical approaches dynamically adjust window sizes and employ multi-resolution strategies for improved accuracy and efficiency [8, 9, 10]. Notably, these methods play crucial roles in diverse stereovision applications, with ongoing research aimed at further enhancing their performance and versatility.

Another problem is the verification of the correctness of depth map determination. One method involves analysing the matching function values for different disparity values. When a certain minimum is reached, it is considered that the disparity value has been correctly determined. Another method involves searching for matches first in one image and then in the other. Obtaining the same disparity value in both cases leads to considering the result as correct [7, 8].

The mentioned drawbacks of stereovision reconstruction can be alleviated by incorporating an IR illuminator that projects a pattern onto the three-dimensional scene. This pattern facilitates depth searching in scenarios with insufficient lighting and in areas lacking clear structure. It should be clearly emphasized that in the case of a very bright scene, the pattern projected by the IR illuminator may not be visible when recorded by stereovision camera. Such a solution has indeed been utilized in stereovision cameras like the Intel RealSense series. The depth calculation mechanism employed by the Intel® RealSense™ D400 series depth camera relies on stereo vision principles [11]. This entails the utilization of both a left and right imager, along with an optional infrared projector. The infrared projector emits a non-visible static IR pattern, enhancing depth accuracy, particularly in low-texture environments. As the left and right imagers capture the scene, the data is then transmitted to the depth imaging vision processor (Intel® RealSense™ Vision Processor D4). Here, depth values for each pixel are computed by correlating corresponding points between the left and right images and analysing the displacement between them. However, detailed implementation specifics are not publicly available [11].

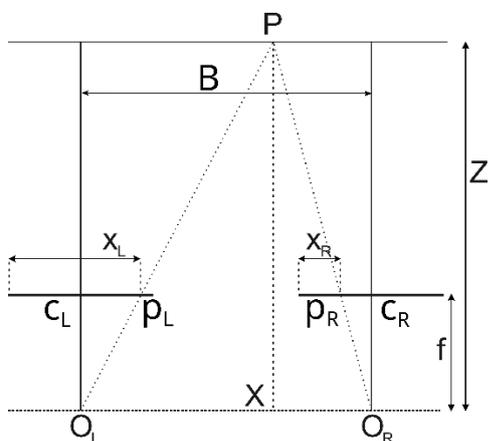


Fig. 1. Canonical camera arrangement in stereovision systems. Points  $O_L$  and  $O_R$  are camera foci,  $f$  - focal length of the camera, distance  $B$  between them is the base of a stereoscopic system. Let point  $P$  in three-dimensional scene have coordinates  $(X, Y, Z)$ . Images of this point in the left and right cameras are respectively points  $p_L(x_L, y)$  and  $p_R(x_R, y)$ . Disparity is the difference  $x_L - x_R$

### The experimental setup and results

The Intel RealSense D435i stereovision camera was securely mounted on a tripod to ensure the camera remained stationary. The rig is illustrated in Figure 2. The camera was connected to an Android tablet. For image acquisition purposes, an application was developed to save the visual image, left and right stereovision images, and depth map (resolution 640x480). Default reconstruction parameter

values proposed in Realsense SDK were used. Additionally, a digital luxmeter was placed in the scene to measure the lighting intensity. The luxmeter was placed 30 cm from the cameras in the plane of the camera sensors. It measures the luminous flux incident on a surface per unit area at the location of the luxmeter. In good lighting conditions, the readings of the luxmeter are strongly dependent on its position in the scene, whereas in poor lighting conditions the readings are not dependent on its position. The same observation applies to the stereovision cameras, as the left and right cameras can receive very different illumination patterns in strongly illuminated scenes.



Fig. 2. Mounting the stereo camera on a tripod to prevent position changes during tests

To conduct the analysis, depth maps (which are derived from the disparity map) were considered. Since no reference depth map was available, it was decided not to consider the difference between a reference map and one determined under different lighting conditions. Instead, each depth map was compared pairwise with all other depth maps to demonstrate how maps of the same scene may differ during different lighting conditions. The camera was kept stationary throughout the process, and the measurements were taken throughout the day, resulting in changes in lighting conditions as natural daylight changes. The indicator of the difference between maps is the average distance difference value of points  $\varepsilon$  (in the  $z$ -coordinate) from the origin of the coordinate system, determined for points that were correctly identified according to the algorithm (i.e., points for which the depth map was determined in both compared images).

$$(1) \quad \varepsilon = \frac{1}{N} \sum_{i=1}^N |z_1 - z_2|$$

where  $N$  is the number of points for which the depth map was correctly determined in both compared images,  $z_1$  is the depth of the point in one of the compared maps, and  $z_2$  is the depth of the same point in the other map.

The results of  $\varepsilon$  values are presented in Figure 3. It can be observed that the maximum error reaches 323mm. Figure 4 illustrates the number of pixels for which only one depth map was correctly determined in the pair-wise comparison of all depth maps, showing cases where a pixel had a valid depth measurement in one image but not the other. Thus, we examine the impact of lighting on the certainty of disparity determination, understood as a binary map (whether a value was determined or not). In both sets of results, an anomaly is evident at an illumination level of 2410 Lux. Stereovision images for this case are shown in Figure 5.

The anomaly was caused by direct sunlight illuminating the camera and overexposing a section of the building's wall.

For lighting conditions below 1000 Lux, the influence of structured lighting on reconstruction effectiveness can be observed. Figure 6 shows images captured by stereovision reference camera under various lighting conditions, where white points in the scene are caused by the pattern projected by the IR illuminator.

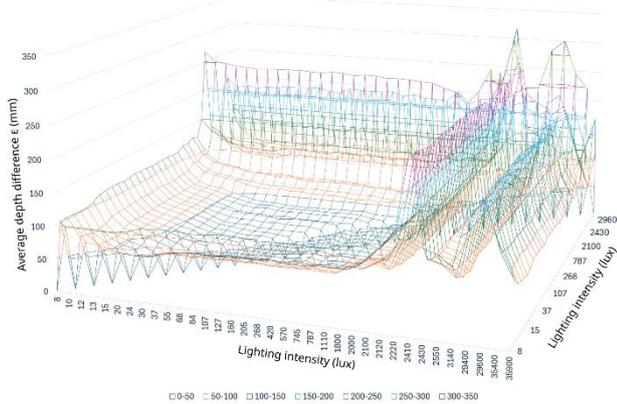


Fig. 3. Average depth difference for different illumination levels. The value of parameter  $\varepsilon$  (z-axis) [in mm] for depth maps determined from images recorded under different lighting conditions [in lux – axes x,y]

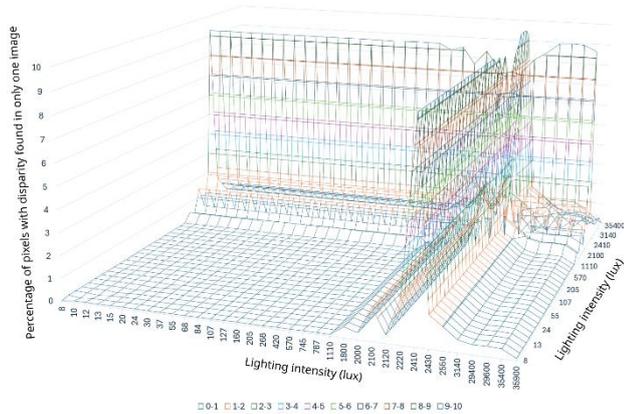


Fig. 4. Analysing lighting influence on disparity determination: Percentage of pixels with disparity unfound in the compared image. Percentage of pixels for which the disparity was determined in only one image (z-axis), for depth maps determined from images recorded under different lighting conditions [in lux – axes x,y]

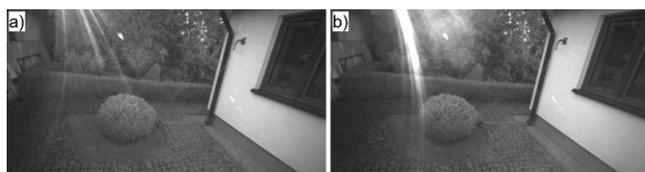


Fig. 5. Left and right images recorded by the stereo camera under illumination  $L = 2410$  lux

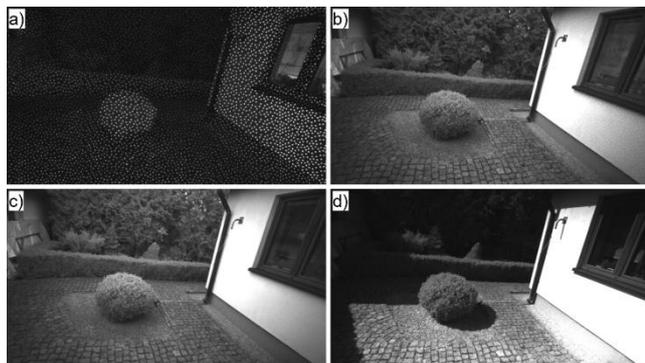


Fig. 6. The impact of lighting on images recorded by the stereo camera: a) 8 lux, b) 23 lux, c) 457 lux, d) 29000 lux

The next test involved intentionally dirtying the lenses of one and two stereovision cameras and comparing the obtained data with the image acquired from a clean camera. The contamination consisted of placing drops of clean water simulating rainfall. This contamination is marked in the table as WILC (Water-Induced Lens Contamination), where WILC1, WILC2, and WILC3 represent three different cases of water droplet placement on the cameras in the stereovision system. The resulting "blurring" effect for each case is shown in Figure 7b-d. The results are presented in Tables 1 and 2 and in Figures 7 and 8.

Table 1. Values of parameter  $\varepsilon$  [in mm] for depth maps obtained from images captured under various levels of water-induced lens contamination (marked in the table as WILC).

	Clean	WILC 1	WILC 2	WILC 3
Clean	0	76	127	3727
WILC 1	76	0	146	4431
WILC 2	127	146	0	6688
WILC 3	3727	4431	6688	0

Table 2. Percentage of pixels excluded due to insufficient disparity calculation in one image for depth maps obtained from images captured under various levels of water-induced lens contamination.

	Clean	WILC 1	WILC 2	WILC 3
Clean	0,0	6,3	15,4	80,9
WILC 1	6,3	0,0	19,9	82,8
WILC 2	15,4	19,9	0,0	86,0
WILC 3	80,9	82,8	86,0	0,0

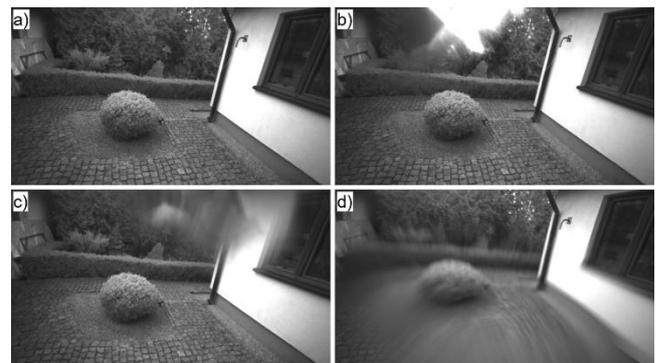


Fig. 7. Images of one of the cameras of a stereo camera system recorded under the same lighting conditions, but with different levels of dirtiness

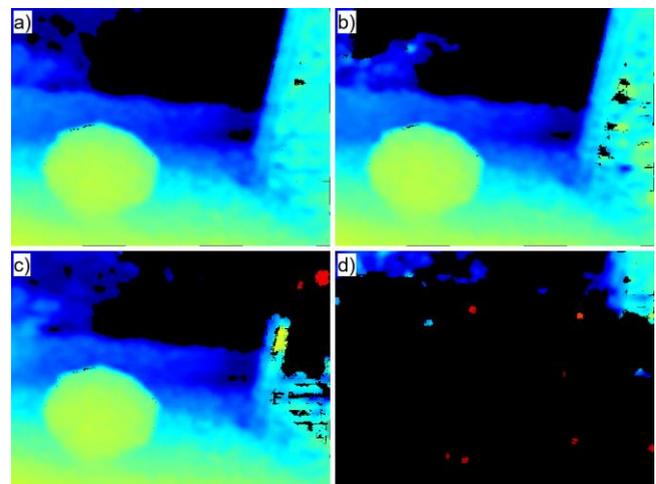


Fig. 8. Depth images of a stereo camera system recorded under the same lighting conditions, but with different levels of dirtiness

The calculations presented in Figures 3 and 4 were repeated after filtering out depth map values greater than 3 meters (the range of information presentation for a visually impaired person) (see Fig. 9 and 10). Pixels with depth values exceeding  $th=3$  meters were considered undetectable and treated the same as those for which the depth map was not computed, despite the fact that depth estimation algorithms were able to determine these values and their built-in verification methods confirmed them as valid. The maximum error  $\varepsilon$  for the same depth maps decreased from 323 mm to 74 mm. While the maximum percentage of pixels where disparity was determined in only one image decreased, it is worth noting that for images recorded under lighting conditions without anomalies, the values increased significantly. This is because, under favorable lighting conditions, the number of pixels for which disparity could be determined in only one image increased. This phenomenon results from the high instability of depth map calculations—many points can be identified where the distance value was  $>3$  m in one image and  $<3$  m in the other.

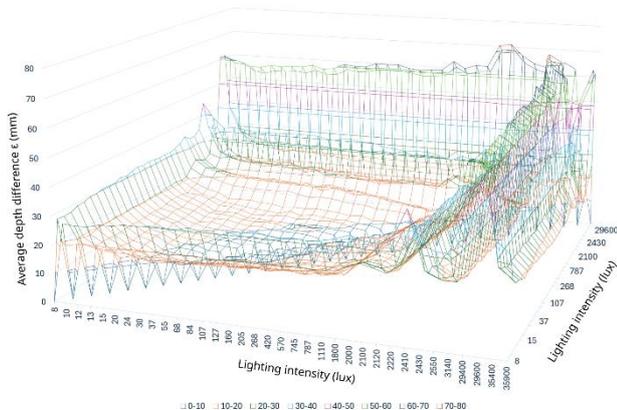


Fig. 9. Average depth difference for different illumination levels. The value of parameter  $\varepsilon$  (z-axis) [in mm] for depth maps determined from images recorded under different lighting conditions [in lux – axes x,y], with depth values limited to a maximum of 3 meters.

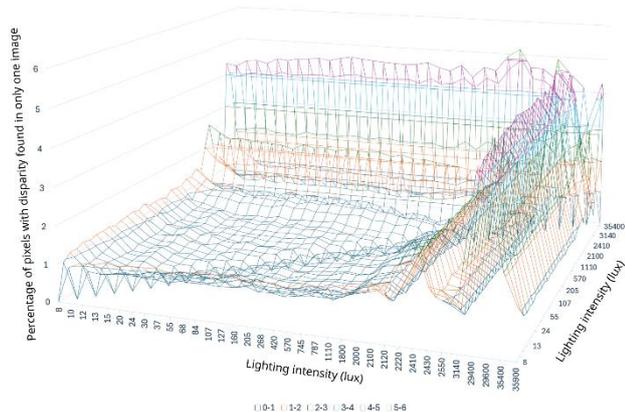


Fig. 10. Analysing lighting influence on disparity determination: Percentage of pixels with disparity unfound in the compared image. Percentage of pixels for which the disparity was determined in only one image (z-axis), for depth maps determined from images recorded under different lighting conditions [in lux – axes x,y], with depth values limited to a maximum of 3 meters.

Table 3 shows the value of parameter  $\varepsilon_M$  which represents maximum value of the average depth difference  $\varepsilon$ , the minimum (*min*) and maximum (*max*) number of pixels (%) for which disparity was determined, for different values of the analyzed range  $th$ . It can be observed that as the range increases, the maximum value of the Average Depth Difference determined for each image pair also increases. It

is should be noted that the values shown in Table 3 apply to the analyzed scene, and they will be different for a different scene.

Table 3. The value of parameter  $\varepsilon_M$  (mm), and the min/max pixel count(%) with determined disparity for different values of the range  $th$  (m).

$th$ [m]	3	3,5	4	4,5	5	5,5	6	6,5	7
$\varepsilon_M$ [mm]	74	125	167	201	205	210	226	243	235
<i>min</i> [%]	12	23	37	44	54	61	67	75	85
<i>max</i> [%]	16	31	47	53	64	78	78	86	97

These results also suggest that the performance of algorithms based on segmentation or clustering in the disparity domain may vary significantly, even when images are recorded under conditions considered favorable for stereo-based reconstruction. Comparing the provided cases, it should be noted that to increase the reliability of the results, it is essential to appropriately select the maximum allowable range for 3D scene reconstruction. The farther a visible point is in the image, the smaller the disparity becomes, and consequently, the accuracy of depth estimation for that point decreases. The range of reconstruction depends on the camera resolution, the internal parameters of the stereo camera, and the baseline of the stereo system.

The quality of disparity maps is significantly influenced by lighting conditions, particularly when the image is overexposed. In such cases, bright uniform areas appear in very bright objects, making it impossible to determine their depth values. Furthermore, the IR illuminator's image is not visible in registered images when recorded on sunny day. Surprisingly, excellent results are achieved with minimal lighting using an IR illuminator. Despite being conducted on a single scene, this analysis demonstrates the significant impact of both the scene and lighting conditions on the outcomes. Lighting conditions have a profound effect on outdoor testing results for visually impaired individuals, highlighting the need to conduct tests in various lighting conditions and adapt algorithm parameters dynamically based on lighting conditions (and, in essence, the number of correctly identified depth values). Another issue is an accidental lens contamination, which can be easily achieved when adjusting camera settings by hand. In this case, a calibration procedure must be written to verify the quality of disparity maps and alert users to potential contamination. The results demonstrate that, despite its numerous benefits, stereovision does not offer a perfect solution, as environmental factors can significantly alter the performance of tested algorithms when operating in controlled laboratory settings.

### Impact of lighting and lens contamination on assistive navigation system for visually impaired individuals

The previously presented influence of lighting on the quality of disparity maps led to the necessity of considering these factors in three-dimensional scene reconstruction algorithms. The algorithm proposed in [6], based on detecting the ground plane and objects not satisfying its equation, was completed by adding the possibility of modifying key parameters in the task of finding obstacles for a blind person. Additionally, the ability to re-estimate the minimum distance between a point in the scene and the ground plane, which causes a given point to be treated as an obstacle, was added. This parameter could be adjusted within the range of 5 cm to 35 cm with a step size of 5 cm. In cases where scenes are recorded at night or during the

evening, due to the existence of structural lighting, this parameter can be reduced. On the other hand, during a sunny day, this parameter needs to be increased, otherwise, false objects appear in the scene.

The problem of camera contamination was solved by analysing the number of found pixels in the depth map during system initialization. A blind person, before using the system, must stand in an obstruction-free area with a visible ground plane and initialize the verification procedure. This procedure determines the angle of the camera's inclination relative to the ground plane and the distance from the ground plane. These information are then used to find obstacles, in case a significant part of the ground plane is invisible in the scene. A significant lack of found depth points for the lower part of the reference image (where the ground plane should be visible) causes the generation of a message requesting verification of the inclination angle and cleanliness of the camera's protective objective lens.

## Summary

The study analyses the impact of lighting and lens contamination on the accuracy of depth maps obtained using stereo vision principles. The Intel RealSense D435i stereo vision camera was used to capture images of a scene under various lighting conditions. All values were determined for the analyzed scene and specific lighting conditions. The impact of direct camera illumination and problems with depth map estimation depend on the orientation of the camera relative to the incoming sunlight. Lenses of one or both cameras were intentionally contaminated by simulating rainfall, and the results were compared with those obtained from a clean camera.

The results indicate that the maximum average error ( $\epsilon$ ) in depth values is 323 mm for the test scene, occurring when the image is overexposed. When the depth values were

restricted to a maximum of 3 meters, the error decreased significantly to 74 mm. Lighting conditions significantly affect the accuracy of disparity maps, particularly in bright uniform areas. Additionally, it was observed that using an infrared illuminator can improve the accuracy of depth maps in low-texture environments. Despite lighting conditions considered favorable for stereo vision, the disparity values for some pixels were inconsistent across images. The study emphasizes the importance of considering lighting conditions and lens contamination when developing stereo vision algorithms for visually impaired individuals. It is suggested that algorithm parameters should be dynamically adjusted based on lighting conditions and that a calibration procedure is necessary to verify the quality of disparity maps and alert users to potential contamination.

*This work has been partly supported by the European Union and the European Regional Development Fund, as part of the Smart Growth Operational Programme. This paper presents the results of industrial research and development work carried out under the project POIR.01.01.01-00-0091/18-00, which was carried out as part of the 'Fast Track' competition organized by the National Centre for Research and Development (Poland). The aim of the project was to develop an electronic system of non-visual presentation of the environment to support navigation in unfamiliar spaces (ETA - Electronic Travel Aid) for blind and visually impaired people, and its implementation will be the result.*

**Authors:** Dr. Eng. Piotr Skulimowski, Prof. Paweł Strumillo, Lodz University of Technology, Institute of Electronics, al. Politechniki 10, 93-590 Lodz, Poland, e-mail: piotr.skulimowski@p.lodz.pl, pawel.strumillo@p.lodz.pl; Szymon Trygar, Tyflos Sp. z o.o, ul. Drzewieckiego 9/4 39-300, Mielec, e-mail: tyflos@tyflos.pl.

## REFERENCES

- [1] Gutta V., Fallavollita P., Baddou N., Lemaire E. D., "Development of a Smart Hallway for Marker-Less Human Foot Tracking and Stride Analysis," in *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1-12, 2021, Art no. 2100412, doi: 10.1109/JTEHM.2021.3069353.
- [2] Feng X., Zhang X., Shi X., Li L., "AIRS: Autonomous Intraoperative Robotic Suturing Based on Surgeon-Like Operation and Path Quantification in Keratoplasty," in *IEEE Transactions on Industrial Electronics*, vol. 71, no. 9, pp. 11115-11124, Sept. 2024, doi: 10.1109/TIE.2023.3342277
- [3] Yao M., Xu B., "A Dense Stereovision System for 3D Body Imaging," in *IEEE Access*, vol. 7, pp. 170907-170918, 2019, doi: 10.1109/ACCESS.2019.2955915,
- [4] Ramon-Soria P., Gomez-Tamm A. E., Garcia-Rubiales F. J., Arrue B. C., Ollero A., "Autonomous landing on pipes using soft gripper for inspection and maintenance in outdoor environments," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019, pp. 5832-5839, doi: 10.1109/IROS40897.2019.8967850.
- [5] Caraiman S., Morar A., Owczarek M., Burlacu A., Rzeszotarski D., Botezatu N., Herghelegiu P., Moldoveanu F., Strumillo P., Moldoveanu, A. "Computer Vision for the Visually Impaired: the Sound of Vision System," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, 2017, pp. 1480-1489, doi: 10.1109/ICCVW.2017.175.
- [6] Skulimowski P., Strumillo P., Trygar Sz., Trygar W., Haptic Display of Depth Images in an Electronic Travel Aid for the Blind: Technical Indoor Trials, Pages 443-453, Lecture Notes in Networks and Systems (LNNS, volume 746), The Latest Developments and Challenges in Biomedical Engineering, Proceedings of the 23rd Polish Conference on Biocybernetics and Biomedical Engineering, Lodz, Poland, September 27–29, 2023
- [7] Cyganek B., Komputerowe przetwarzanie obrazów trójwymiarowych. Akademicka Oficyna Wydawnicza EXIT, 2002.
- [8] Brown M.Z., Burschka D., Hager G.D., Advances in Computational Stereo. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25(8), pp. 993–1008, 2003
- [9] Scharstein, D., Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision 47, 7–42 (2002). <https://doi.org/10.1023/A:1014573219977>
- [10] Hirschmuller H., "Stereo Processing by Semiglobal Matching and Mutual Information," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328-341, Feb. 2008, doi: 10.1109/TPAMI.2007.1166
- [11] Intel® RealSense™ Product Family D400 Series – Datasheet, rev. 018, 2024, <https://dev.intelrealsense.com/docs/intel-realsense-d400-series-product-family-datasheet>, available 2024.06.23