Factors Influencing Accuracy of Estimating Position of Objects in a Multi-camera System

Krzysztof Klimaszewski[1], Tomasz Grajek[1], and Krzysztof Wegner[2]

[1]*Poznan University of Technology, Poznań, Poland,*
[2]*Mucha sp. z o.o., Poznań, Poland*

https://doi.org/10.26636/jtit.2024.2.1457

Abstract — Nowadays, research focusing on robotics, autonomous vehicles, and scene analysis shows a clear need for the ability to accurately reconstruct three-dimensional environments. One of the methods allowing to conduct such a reconstruction is to use a set of cameras and image processing techniques. This is a passive method. Despite being, in general, less accurate than its active counterparts, it offers significant advantages in numerous applications in which active systems cannot be deployed due to limited performance. This paper provides a theoretical analysis of the accuracy of estimating 3D positions of objects present at a given scene, based on images from a set of cameras. The analysis assumes a known geometrical configuration of the camera system. The important limiting factor in the considered scenario is the physical resolution of sensors – especially in the case of systems that are supposed to work in real time, with a high FPS rate, as the use of high-resolution cameras is difficult in such circumstances. In the paper, the influence of the geometric arrangement of the cameras is studied and important conclusions about the potential of three-camera configurations are drawn. The analysis performed and the formulas derived help predict the boundary accuracy values of any system using a digital camera. The results of an experiment that confirm the theoretical conclusions are presented as well.

Keywords — 3D imaging, measurement accuracy, optical imaging, stereovision

1. Introduction

Evaluation of the accuracy with which the position of a given object is estimated is a fundamental problem encountered when optimizing multi-camera systems. In many applications, estimation of the scene's depth and the position of specific objects is one of the most frequently performed tasks. The goal is to obtain as accurate an estimate as possible, with a certain set of limiting factors taken into consideration. Such factors include, most often, the cost and the physical size of the setup used. Additionally, the system needs to generate a manageable set of data that can be processed, extracted and transmitted. Real time data transmission and processing capabilities are of paramount importance in many applications.

The position and the size of an object are often measured based on an image acquired by video cameras. The majority of such systems rely on the parallax effect. Therefore, more than one camera and different angles of observation are required.

The simplest approach to such a system is a stereo pair – a set of two cameras having their optical axes aligned, so that the object of interest is visible in images acquired by both cameras. Images from a stereo pair can be used to estimate the distance of a set of points in the images (in some types of parcel sorting systems, this approach is used to determine the position of corners of the measured box). Then, the actual size of a 3D object is calculated [1], [2]. Usually, images from a stereo pair undergo a process of rectification, in which they are transformed to a form with the corresponding points from the pair of images situated along the same horizontal line. This significantly simplifies the matching process [3].

Accurate measurement of the dimensions of objects is important not only when sorting items, but also on production lines and conveyors. Therefore, different measurement methods are applied, with stereovision [4] enjoying a particularly strong position here. In the recent years, stereovision applications have been optimized by the use of neural networks [5], [6].

Other type of a multi-view system could be a crowd-based 3D object reconstruction, where based on images captured by random people, a 3D mesh or a 3D model of a photographed object is created [7].

Virtual/augmented reality applications have been gaining popularity in the recent years as well. In this approach, real-life objects are scanned and converted into a digital model [8]. The performance of such systems depends on the ability to precisely locate feature points in a given scene. Such problems are widely discussed in the literature, for example in [9]. Virtual/augmented reality solutions are used in a range of applications – from leisure to medical sectors. In the latter, they may be used to identify the location of surgical instruments or any other objects. In such applications, accuracy is utmost importance, as the displacements of feature points are small and any errors can influence the reconstructed orientation in a rather significant way. Therefore, the need for continuous optimization, even at the camera positioning stage, exists – as evidenced in [10], [11].

A visual robot trajectory planning system, where images from video cameras are used to control and plan the movement of

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License.
For more information, see https://creativecommons.org/licenses/by/4.0/

a robotic arm to avoid obstacles [12]–[15], is another area in which the solutions discussed in this paper may be applied. Stereo and multi-view systems may also be successfully used in the automotive industry, as the development of autonomous vehicles has been recently progressing at a very fast pace [16]. Here, the accuracy of locating objects within a scene is of utmost importance, since significant errors may result in injuries or deaths. Traffic management is another automotive application, with stereovision being commonly used for measurements [1], [2], [17].

Regardless of the type of the system, the need to optimize the accuracy of 3D positioning of objects always exists. Therefore, evaluation of a given camera system in terms of the level of accuracy it is capable of achieving is considered to be a prospective research area.

2. Related Work

The notion of accuracy has been considered from many different perspectives. In the 1980s, various authors considered the application of stereo pair-based robot navigation systems and investigated the problem of accurately estimating coordinates of objects within a given scene [18]. They noticed that the problem of accuracy cannot be simplified to triangulation and suggested a more complex model that has been relied upon in nearly all other projects since. In our paper, we analyze scenarios in which more cameras are used and attempt to analyze the impact of the spatial configuration of the cameras on the accuracy of the 3D position estimates.

In [19], the authors analyze depth resolvability of a system and show its dependency on different parameters of a stereo pair. A similar approach is presented in papers [20], [21], where the authors analyze the influence of the pixel size on the accuracy of estimating the distance to a given object. Their analysis is limited to a two-camera setup only. In [22], the authors note that the result of position estimation has the form of a three-dimensional figure located in space (a 3D cell). Such a figure is usually represented by a single point. The choice of this point does affect the accuracy of results. Again, only a stereo pair is considered in this paper.

The estimation error is also analyzed in papers [23] and [24], where a detailed analysis of several factors is presented for a stereo pair. Another in-depth analysis of the stereo pair is provided in [25]. In article [26], the authors investigate the method of finding an optimal arrangement of a stereo pair that would minimize the measurement errors for a given scenario with converging optical axes of the cameras. Their analysis is limited to a stereo pair as well.

A similar analysis for a stereo pair is found in [27]. This time, however, a stereo pair with parallel optical axes is considered. Another, more recent analysis of a stereo pair is given in papers [28], [29].

From the point of view of specific applications, the importance of studying the impact that the geometry of the system exerts on the results obtained in practice is shown, for instance, in [12], where stereo pair-based visual odometry is optimized by adjusting camera orientation. Theoretical calculations presented are augmented by experimental verification.

Many papers consider the size of a pixel as the limiting factor. However, when using more sophisticated algorithms, it is possible to obtain more accurate correspondences. Paper [30] provides an analysis of matching algorithms that give subpixel precision of disparity estimation. The analysis shows the possibility of obtaining accurate disparity results up to 1/10 or 1/20 of a pixel. With such an approach it is important to note the additional burden of the matching algorithm, which can not always be handled on a given platform, especially when requiring real-time results with high frame rates.

Moreover, it is very likely that due to the physical limitations of the system, i.e. dimensioning accuracy, repeatability, and accuracy of camera parameters, the overall accuracy will be impacted in a more prominent way. Therefore, a need to analyze the system and its properties exists.

A very important aspect of estimating the accuracy of a 3D scene was tackled in paper [31], where the authors showed the influence of a self-heating camera sensor and body on 3D scene estimation.

Systems comprising more than two cameras were also considered in the literature, although this subject is less popular. In [32], the authors propose the use of a third camera to improve the accuracy of estimating the location of objects. They do not, however, provide a study of the error-related factors. Recent analyses also consider a three-camera setup, as shown in [33]–[35]. The analysis is limited, however, and deals only with a specific linear camera arrangement. A more complex camera arrangement is considered in [36], where linear and rectangular matrix camera setups are considered. In our paper, we analyze a triangular camera arrangement using 3 cameras only.

Overall, review of the literature stresses the importance of accuracy. However, a limited number of analyses of various geometrical camera setups is observed. The above applies, in particular, to setups with more than two cameras.

In this paper, we present a closed-form 3D coordinate estimation solution that is based on multiple images, and provide information about the accuracy of such a measurement method, with various factors taken into consideration. Furthermore, we take into account the number and spatial configuration of the cameras required and describe their impact on the accuracy of estimating 3D positions in the context of multi-view processing.

It needs to be stressed that we consider, in this paper, only a theoretical model and we do not include the effects of noise or camera optics imperfections. We assume an ideal camera model that does not introduce distortions to the image. While this means that our results are not directly applicable to a real-life camera system, they are still useful in understanding the relationships between the number and position of the cameras and the boundaries of the 3D space containing the actual location of specific points.

2.1. Stereo Pair Approach

In a simple stereo-rectified image system, the process of calculating the distance to a given object is relatively easy, as it relies on trigonometric formulas [3], [37] and the relation of the so-called disparities d, i.e. relative shifts in the position of the object under consideration in both images. Distance z is:

$$z = f\frac{b}{d}\,, \tag{1}$$

where b is the system baseline (distance between two cameras observing an object), and f is the focal length of the cameras. It is assumed that both cameras have the same parameters. Dense depth estimation is one of the applications of the above-mentioned relationship [38]–[42].

This rudimentary system with two cameras has the advantage of simplicity and low cost, but more demanding applications require the use of a higher number of cameras. It is hard to use a two-camera system to generate a sufficient data set for refocusing or all-in-focus applications, see e.g. [43].

To address such problems, more sophisticated video systems were developed, composed of tens of cameras, positioned such that all of them observe the same object. The lightfield camera system may be one of such solutions. The cameras are arranged in a rectangular grid, pointing in the same direction and recording the same scene [44]. Another solution consists in using a single camera with a matrix of lenses in front of the sensor. Such sensors, however, are difficult to manufacture and offer limited performance in terms of the field of view and resolution [45].

Two-camera arrangements used for 3D coordinate estimation are described in the literature rather broadly. When more cameras are available, the 3D coordinate estimation process is usually performed by relying on pair-wise distance estimation for consecutive pairs of cameras. Then, the results from different pairs are aggregated, usually by means of measurement averaging [46], [47].

2.2. Close Form Solution of 3D Position Estimation from Multiple Images

Let us assume that an object – a point in 3D space located at position $M = [X\ Y\ Z]^T$ – is observed by a set of N cameras. The i-th camera is located at coordinates T_i and is oriented in a direction specified by a rotation matrix R_i (Fig. 1). All parameters of the i-th camera are specified by

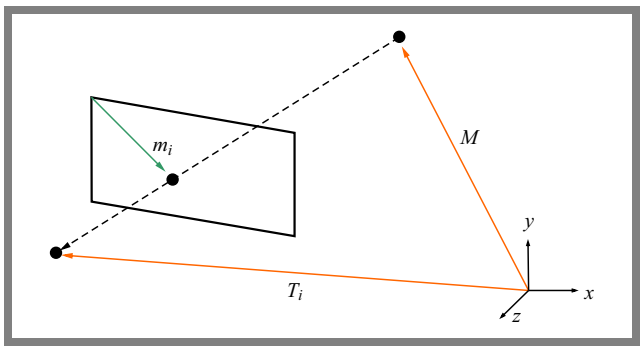

Fig. 1. Capturing the image of a scene using a single camera.

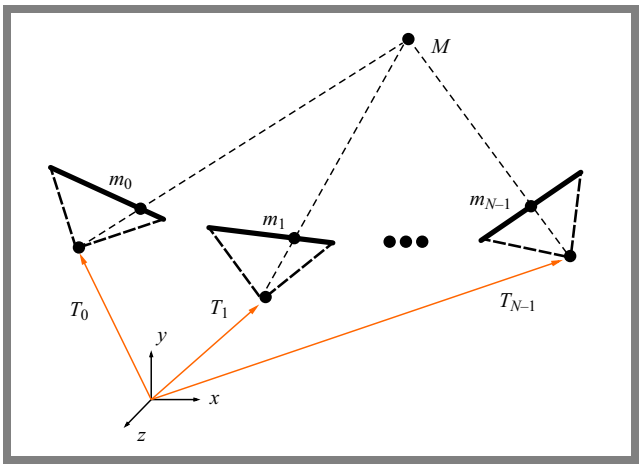

Fig. 2. Ideal multi-camera image acquisition setup.

K_i matrix. Cameras are not necessarily equal in that respect and may have different intrinsic parameters in the general case. However, in the most popular approach, all cameras in the system are equal and have similar parameters. Moreover, let us assume that all optical distortions have been removed. Point M is projected on the image plane of the i-th camera at position $m_i = [u_i\ v_i\ 1]^T$ expressed in homogenous image coordinates, where direction u spans from left to right and direction v spans from top to bottom. This situation is depicted in Fig. 2.

In such conditions, image acquisition of point M by i-th camera can be modeled as:

$$z_i\, m_i = K_i\, R_i(M - T_i)\,, \tag{2}$$

where z_i is the distance measured between the i-th camera at position T_i and the observed point M in the direction perpendicular to the image plane.

In ideal conditions, based on Eq. (2), one can calculate the position of point M based on the observed position m_i of that point, and distance z_i.

In real-world conditions, distance z_i is unknown, since it is lost in the acquisition process. Worse, in practical arrangements we do not know the accurate parameters of the i-th camera. Such parameters as the position of the camera, its orientation and other properties are frequently known only approximately [3]. Usually, camera parameters are determined in the course of a calibration procedure which is characterized by limited accuracy levels, with the error value being difficult to predict [48]. Moreover, due to the discrete nature of the image, we know the position m_i of the imaged point M up to a certain degree only, i.e. to the level of a pixel, half-pixel or quarter pixel. This scenario is presented in Fig. 3.

With all of the abovementioned reasons taken into consideration, we can only try to estimate the position of point M based on the image taken by the i-th camera. In such conditions, we can calculate an approximated position of point M, further denoted by M_i as:

$$M_i = z_i(K_i\, R_i)^{-1} m_i + T_i\,, \tag{3}$$

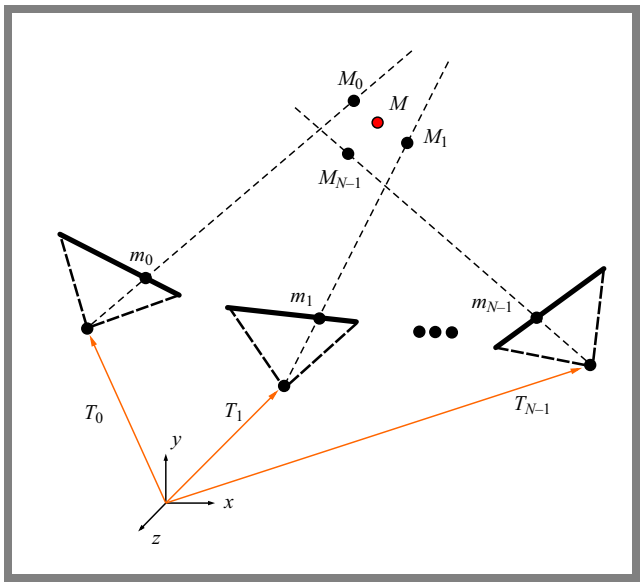

Fig. 3. Real-world multi-camera image acquisition setup.

Based on estimation from N cameras, a real position of point M in 3D space is then to be found by identifying the smallest distance between the true position of M and each estimate M_i.

Let us define a square norm of the distance between the true (unknown) position of M and the i-th camera based estimate M_i as $|S_i|^2$:

$$\begin{aligned} |S_i|^2 &= S_i^T\, S_i\,, \\ S_i = M - M_i &= M - z_i (K_i\, R_i)^{-1} m_i - T_i\,. \end{aligned} \tag{4}$$

The aim is to select such an estimate M' of the unknown position of M that it is closest to all estimates M_i, as this will minimize the error of the estimate of the actual M:

$$\arg\min_{M'} \sum_{i=0}^{N-1} |S_i|^2\,. \tag{5}$$

Substituting Eqs. (3) and (4) into Eq. (5), we get:

$$\begin{aligned} &\arg\min_{M'} \sum_{i=0}^{N-1} |S_i|^2 = \\ &\arg\min_{M'} \sum_{i=0}^{N-1} \left(M' - z_i (K_i\, R_i)^{-1} m_i - T_i\right)^T \\ &\cdot \left(M' - z_i (K_i\, R_i)^{-1} m_i - T_i\right)\,. \end{aligned} \tag{6}$$

For further simplification, let us assign the A_i parameter based on camera calibration and the image itself, such that:

$$A_i = (K_i\, R_i)^{-1}\, m_i\,. \tag{7}$$

Then, Eq. (6) simplifies to:

$$\begin{aligned} &\arg\min_{M'} \sum_{i=0}^{N-1} |S_i|^2 = \\ &\arg\min_{M'} \sum_{i=0}^{N-1} \left(M' - z_i A_i - T_i\right)^T \cdot \left(M' - z_i A_i - T_i\right)\,. \end{aligned} \tag{8}$$

In Eq. (8), there are $N+3$ unknown variables: all distances z_i and X, Y, Z coordinates of point M'.

Equation (8) can be minimized by differentiation with respect to all unknowns and finding a stationary point. Let us start with unknowns z_i:

$$\frac{\partial \sum_{k=0}^{N-1} |S_k|^2}{\partial z_i} = \frac{\partial |S_i|^2}{\partial z_i} = 2\, S_i^T \frac{\partial S_i}{\partial z_i} = 2\, S_i^T (-A_i) \tag{9}$$

and compare the resulting equations to 0:

$$2\, S_i^T (-A_i) = 0\,. \tag{10}$$

Next, we get z_i as:

$$\begin{aligned} 2\, S_i^T (-A_i) &= 0 \\ -2 \left(M' - z_i A_i - T_i\right)^T A_i &= 0 \\ \left(M' - z_i A_i - T_i\right)^T A_i &= 0 \\ M'^T A_i - z_i A_i^T A_i - T_i^T A_i &= 0 \\ M'^T A_i - T_i^T A_i &= z_i A_i^T A_i \\ z_i = \frac{M'^T A_i - T_i^T A_i}{A_i^T A_i}\,. \end{aligned} \tag{11}$$

From differentiation of Eq. (8) with respect to X, we get:

$$\begin{aligned} &\frac{\partial \sum_{i=0}^{N-1} |S_i|^2}{\partial X} = \sum_{i=0}^{N-1} 2 S_i^T \frac{\partial S_i}{\partial X} = \\ &\sum_{i=0}^{N-1} 2 S_i^T \frac{\partial M'}{\partial X} = \sum_{i=0}^{N-1} 2 S_i^T \cdot [\,1\ 0\ 0\,]^T\,. \end{aligned} \tag{12}$$

Similarly, from differentiation of Eq. (8) with respect to Y and Z, we get:

$$\begin{aligned} &\frac{\partial \sum_{i=0}^{N-1} |S_i|^2}{\partial Y} = \sum_{i=0}^{N-1} 2 S_i^T \frac{\partial S_i}{\partial Y} = \\ &\sum_{i=0}^{N-1} 2 S_i^T \frac{\partial M'}{\partial Y} = \sum_{i=0}^{N-1} 2 S_i^T \cdot [\,0\ 1\ 0\,]^T \\ &\frac{\partial \sum_{i=0}^{N-1} |S_i|^2}{\partial Z} = \sum_{i=0}^{N-1} 2 S_i^T \frac{\partial S_i}{\partial Z} = \\ &\sum_{i=0}^{N-1} 2 S_i^T \frac{\partial M'}{\partial Z} = \sum_{i=0}^{N-1} 2 S_i^T \cdot [\,0\ 0\ 1\,]^T\,. \end{aligned} \tag{13}$$

Equations (12) and (13) are equal to zero if and only if:

$$\sum_{i=0}^{N-1} S_i^T = 0 \,. \tag{14}$$

By expanding (14) and substituting Eq. (11), we get:

$$\begin{aligned}
\sum_{i=0}^{N-1} S_i^T &= 0 \\
\sum_{i=0}^{N-1} \left(M' - z_i A_i - T_i\right)^T &= 0 \\
\sum_{i=0}^{N-1} \left(M' - \frac{M'^T A_i - T_i^T A_i}{A_i^T A_i} A_i - T_i\right)^T &= 0 \\
\sum_{i=0}^{N-1} M' - \sum_{i=0}^{N-1} \frac{M'^T A_i - T_i^T A_i}{A_i^T A_i} A_i - \sum_{i=0}^{N-1} T_i &= 0 \\
NM' - \sum_{i=0}^{N-1} \frac{M'^T A_i - T_i^T A_i}{A_i^T A_i} A_i &= \sum_{i=0}^{N-1} T_i \\
NM' - \sum_{i=0}^{N-1} \frac{M'^T A_i}{A_i^T A_i} A_i - \sum_{i=0}^{N-1} \frac{-T_i^T A_i}{A_i^T A_i} A_i &= \sum_{i=0}^{N-1} T_i \\
NM' - \sum_{i=0}^{N-1} \frac{M'^T A_i}{A_i^T A_i} A_i + \sum_{i=0}^{N-1} \frac{T_i^T A_i}{A_i^T A_i} A_i &= \sum_{i=0}^{N-1} T_i \\
NM' - \sum_{i=0}^{N-1} \frac{M'^T A_i}{A_i^T A_i} A_i &= \sum_{i=0}^{N-1} T_i - \sum_{i=0}^{N-1} \frac{T_i^T A_i}{A_i^T A_i} A_i \,.
\end{aligned} \tag{15}$$

It can be proved that (see in Appendix):

$$\frac{M'^T A_i}{A_i^T A_i} A_i = \frac{A_i A_i^T}{A_i^T A_i} M' \,. \tag{16}$$

So, then:

$$\begin{aligned}
NM' - \sum_{i=0}^{N-1} \frac{M'^T A_i}{A_i^T A_i} A_i &= \sum_{i=0}^{N-1} T_i - \sum_{i=0}^{N-1} \frac{T_i^T A_i}{A_i^T A_i} A_i \\
NM' - \sum_{i=0}^{N-1} \frac{A_i A_i^T}{A_i^T A_i} M' &= \sum_{i=0}^{N-1} \left(T_i - \frac{T_i^T A_i}{A_i^T A_i} A_i\right) \\
\left(NI - \sum_{i=0}^{N-1} \frac{A_i A_i^T}{A_i^T A_i}\right) M' &= \sum_{i=0}^{N-1} \left(T_i - \frac{T_i^T A_i}{A_i^T A_i} A_i\right) \,.
\end{aligned} \tag{17}$$

Upon assigning a 3×3 matrix $\mathbf{A}$ and a 3-element vector $\mathbf{B}$, Eq. (17) can be rewritten as:

$$\mathbf{A} M' = \mathbf{B} \tag{18}$$

where:

$$\begin{aligned}
\mathbf{A} &= NI - \sum_{i=0}^{N-1} \frac{A_i A_i^T}{A_i^T A_i} \\
\mathbf{B} &= \sum_{i=0}^{N-1} \left(T_i - \frac{T_i^T A_i}{A_i^T A_i} A_i\right) \,.
\end{aligned} \tag{19}$$

The linear Eq. (18) can be easily solved to obtain the position of point M'.

3. Accuracy-Related Considerations

There are three main sources of inaccuracies affecting position estimation:

- accuracy of camera calibration – expressed by accuracy of variables T_i, K_i and R_i,
- accuracy of point detection in the images – the accuracy of m_i,
- the number of cameras used.

The accuracy of camera calibration depends on the features of the calibration software and the quality of calibration images used. It also depends on the quality of the calibration pattern. For contemporary high-resolution cameras and a typical grid or checkerboard-based calibration patterns, shape or coplanarity discrepancies of the calibration pattern (equaling fractions of a millimeter) will noticeably influence calibration accuracy. Unfortunately, it is difficult to improve calibration accuracy.

Similarly, the accuracy of the location of points in the images is limited as well. It may vary from several pixels in the case of neural network object detection, to ¼ of the pixel in the case of a simple Harris corner detector [49]. Some sophisticated methods can improve the accuracy and resolution of the location of corresponding points in images. Such accuracy will be constrained, however, in all cases, by the digital nature of the captured image and its rasterization [1], [2].

The last factor is the easiest to modify at the design stage. In theory, it is always possible to add more cameras to the analyzed system in order to improve its performance. This stems from the fact that noise introduced to the process of locating an object in a given scene, caused by inaccurate calibration, as well as by the localization of corresponding points in an image with limited raster and omnipresent image noise, can be averaged by the addition of more (still) inaccurate data from additional cameras.

This last factor is the one that has received the least attention in the literature. The most important question is whether it is always beneficial to add more cameras into a system and to what extent such an approach improves the accuracy of the position estimates?

In this paper, we address those questions by performing a theoretical analysis based on the formulas provided above. First of all, we present the phenomenon of position uncertainty caused by the limited granularity of data. We assume that granular data may result in a localization error that is determined by a few factors we attempt to analyze. We also assume that the correct localization of the object and the results of the localization estimation process form a certain bounded volume in space, later on referred to as uncertainty region. We evaluate what this uncertainty region of the space looks like.

The following model is used in further considerations. A small object is observed by N cameras. The question is how the

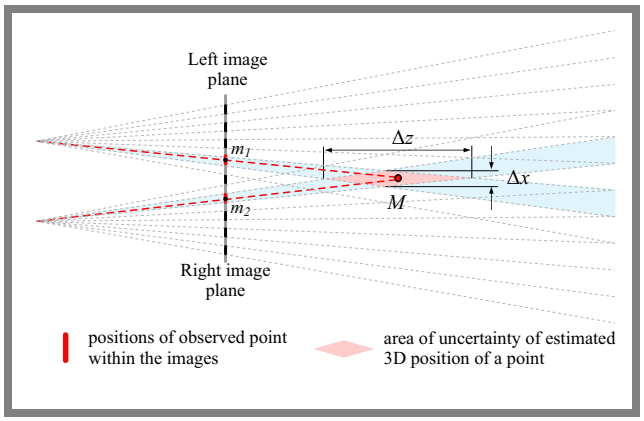

Fig. 4. Planar view of a real-world multi-camera image acquisition setup. x–z plane view is shown for simplicity. 3D shape of the area of uncertainty is shown in Fig. 12.

uncertainty of its location depends on the number of cameras used.

3.1. Stereo Pair

Figure 4 shows a simple stereoscopic system. The system is composed of two cameras. Light emitted from (or reflected off) a small object, i.e. point M, marked as a red dot, is travelling along the red paths to cameras 1 and 2. Those light rays results in cameras 1 and 2 observing point M as points m_1 and m_2, respectively. Without affecting generality, it is assumed that the cameras are perfectly aligned with their optical axes positioned parallel to each other. Thus, the optical axes are perpendicular to the baseline of the camera system. Such an arrangement provides perfectly rectified images of the scene and is commonly used in stereoscopic visual systems.

Because of the digital (rasterized) nature of the image and the limited granularity of image point representation, if object/point M is observed at some position m_1 e.g. $m_1 = [430 \;\; 650 \;\; 1]^T$ in the image from camera 1, it actually means that the object that emits (or reflects) that light was located inside the pixel cone defined by the focal length and the size of the sensor element (blue cones in Fig. 4). If the exact path of the light (red lines) was known, one could reconstruct the accurate position of point M. Unfortunately, in real world conditions we do not know the actual path of the light ray, but of a cone of rays (blue cones). Therefore, we can only reconstruct the area/volume in which point M should be located in order to be perceived as points of the image located at m_1 and m_2, respectively. This reconstruction volume of the possible true position of point M has a certain size which we can parameterize. Its dimensions may be calculated along all three axes: Δx, Δy, and Δz.

With the assumption of a perfectly aligned camera system, the translation and rotation matrices may be defined as follows:

$$\begin{aligned} T_1 &= [-b \;\; 0 \;\; 0]^T \\ T_2 &= [\, b \;\; 0 \;\; 0 \,]^T \,, \\ R_1 &= R_2 = I \end{aligned} \tag{20}$$

where a camera pair baseline distance is $2b$.

The intrinsic camera parameters are the same and equal to:

$$K_1 = K_2 = \begin{bmatrix} f & 0 & w/2 \\ 0 & f & h/2 \\ 0 & 0 & 1 \end{bmatrix} . \tag{21}$$

where f stands for focal length, while w and h are captured by the modeled cameras image resolution.

For the same point m after taking the images the location of projections of the point m onto the image plane of the cameras have the following coordinates:

$$\begin{aligned} m_1 &= [u_1 \;\; v_1 \;\; 1\,]^T \\ m_2 &= [u_2 \;\; v_2 \;\; 1\,]^T \end{aligned} . \tag{22}$$

Based on these assumptions, we can solve Eq. (18) for two cameras and obtain the position of the observed point M – see Eq. (23).

$$M' = \begin{bmatrix} \frac{b(4f^2(u_1-u_2)(u_1+u_2-w)+(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w)(-2(u_2v_1+u_1v_2)+h(u_1+u_2-w)+(v_1+v_2)w))}{4f^2((u_1-u_2)^2+(v_1-v_2)^2)+(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w)^2} \\ \frac{-b(4f^2(u_1-u_2)(h-v_1-v_2)+(h-2v_1)(h-2v_2)(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w))}{4f^2((u_1-u_2)^2+(v_1-v_2)^2)+(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w)^2} \\ \frac{2bf(4f^2(u_1-u_2)+(h-v_1-v_2)(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w))}{4f^2((u_1-u_2)^2+(v_1-v_2)^2)+(h(u_1-u_2)+2u_2v_1-2u_1v_2-v_1w+v_2w)^2} \end{bmatrix} . \tag{23}$$

$$\sigma M' = \begin{bmatrix} \sqrt{\left|\frac{\partial X'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial X'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial X'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial X'}{\partial v_2}\right|^2 \sigma v_2^2} \\ \sqrt{\left|\frac{\partial Y'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial Y'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial Y'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial Y'}{\partial v_2}\right|^2 \sigma v_2^2} \\ \sqrt{\left|\frac{\partial Z'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial Z'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial Z'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial Z'}{\partial v_2}\right|^2 \sigma v_2^2} \end{bmatrix} . \tag{24}$$

$$\sigma M' = \begin{bmatrix} \sqrt{\frac{\sigma u^2 Z^2(b^4Y^2+b^2(-2X^2Y^2+(Y^2+Z^2)^2)+X^2(X^2Y^2+(Y^2+Z^2)^2))}{2b^2f^2(Y^2+Z^2)^2}} \\ \sqrt{\frac{\sigma u^2 Z^2(b^2(Y^2+Z^2)^2+Y^2(X^2Y^2+(Y^2+Z^2)^2))}{2b^2f^2(Y^2+Z^2)^2}} \\ \sqrt{\frac{\sigma u^2 Z^4(X^2Y^2+(Y^2+Z^2)^2)}{2b^2f^2(Y^2+Z^2)^2}} \end{bmatrix} . \tag{25}$$

In order to estimate the uncertainty of the estimated location of point M depending on the uncertainty of the observed image positions m_1 and m_2, Eq. (23) could be expanded into a Taylor series around positions m_1 and m_2. We will assume here that only the uncertainty of the position is important, while other factors, i.e. intrinsic and extrinsic parameters of the camera: K_1 and K_2, T_1 and T_2 as well as R_1 and R_2, are known with perfect accuracy. The uncertainty of the position $\sigma M'$ expressed as standard deviation can be approximated via the propagation rule given by Eq. (24).

Note that $\sigma M'$ is actually a 3-component vector. After the necessary transformations and upon substituting Eq. (2) for obtaining u_1, v_1, u_2, v_2 from X, Y, Z, we get Eq. (25).

To study the accuracy of estimating the position of point m depending on its location in front of the camera system, a function of accuracy expressed as standard deviation along three main axes σX, σY, σZ could be drawn with respect to the exact Z coordinate of point M.

In the following calculations, we assume a perfect alignment of the camera system, the viewing angle of the cameras $fov = 60°$, image resolution of 1920×1080 pixels, and baseline $b = 0.5$ m. The only source of errors is the uncertainty of the exact location of corresponding points in the image, limited by image resolution. The accuracy of coordinates u and v of the positions of points m_1 and m_2, expressed as standard deviation σu and σv, is equal to 1:

$$\sigma u = \sigma v = 1 \ . \tag{26}$$

Therefore, for square pixels:

$$f = \frac{\frac{w}{2}}{\operatorname{tg} \frac{fov}{2}} \tag{27}$$

For a stereo pair, the minimum distance between the camera baseline and the object has a lower limit below which the object is not visible to both cameras. Although the limit has not been marked on all of the following figures, the value of the minimum distance for the considered case, when $Y = 0$, is:

$$\sqrt{3}\frac{b}{2} + \sqrt{3}X \ .$$

Accuracy estimation is presented in Fig. 5. In the conditions under consideration, accuracy of the estimated true position of point m positioned at the front of the two-camera system at a distance of 4 m is equal to 1 cm in the z direction and 0.2 cm in the x and y directions. At the distance of 7 m, accuracy drops to 4 cm. This means that if we measure the size of the object at 7 m, the error is up to 4 cm. This result is valid for a system with a 0.5 m baseline distance, meaning that the cameras observing the object were spaced 0.5 m apart.

In Fig. 6, the relationship between the accuracy expressed as standard deviation along three main axes σX, σY, σZ and camera baseline b is shown. It can be observed that by shortening the camera baseline, i.e. placing both cameras closer, accuracy is reduced and uncertainty, expressed as standard deviation, is increased. In general, this allows us

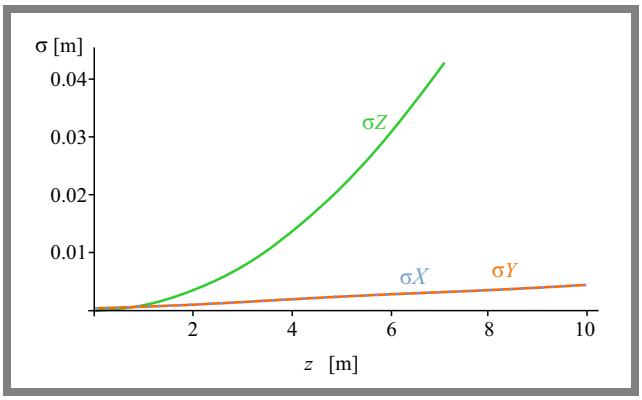

Fig. 5. Standard deviation of position estimation along z axis with respect to real position Z of point m ($X, Y = 0$, $b = 0.5$ m, $w = 1920$, $fov = 60°$)

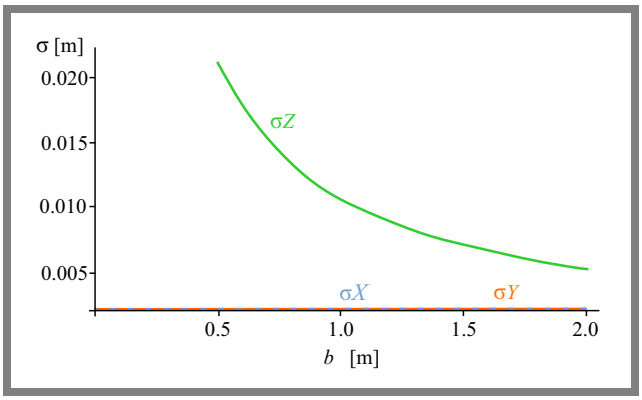

Fig. 6. Accuracy expressed as standard deviation σX, σY, σZ of position estimation along x, y and z axes with respect to camera baseline b ($X, Y = 0$, $Z = 5$ m, $w = 1920$, $fov = 60°$).

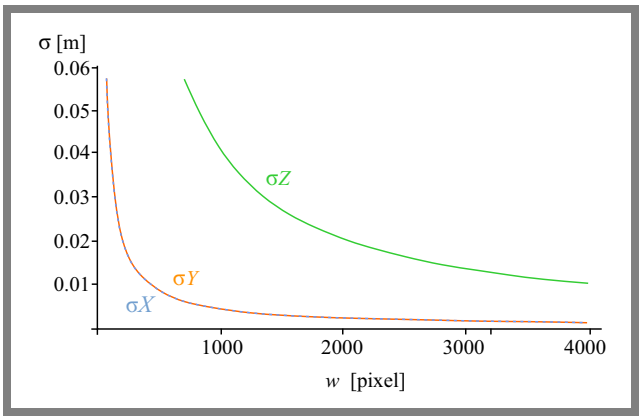

Fig. 7. Accuracy of position expressed as standard deviation σX, σY, σZ estimated along x, y axes and z axis, relative to camera resolution w ($X, Y = 0$, $Z = 5$ m, $b = 0.5$ m, $fov = 60°$).

to draw an expected conclusion – by doubling the camera distance (baseline) we can lower uncertainty by a factor of 2. Therefore, the universal relationship between distance uncertainty and baseline for a two-camera system is proved.

The accuracy of estimations is significantly impacted by the size of the pixel, i.e. resolution of the images taken. In Fig. 7, the dependency of accuracy – expressed as standard deviation – on the number of pixels along the image's horizontal axis (horizontal resolution) is illustrated. Inverse dependency is

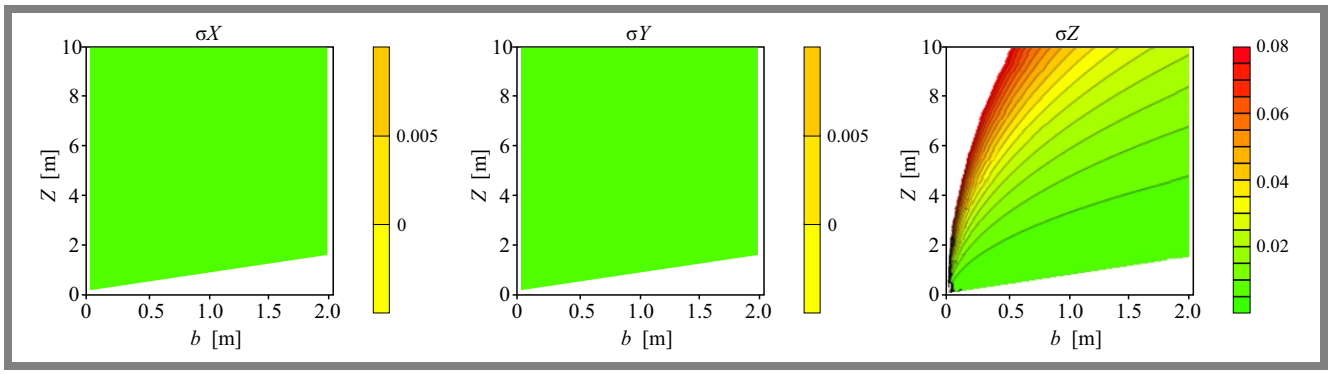

Fig. 8. Accuracy expressed as standard deviation σZ of position estimation along z axis with respect to position Z of point m and camera baseline b ($X = Y = 0$, $fov = 60°$, $w = 1920$). White areas at the bottom represent distances below the minimum limit.

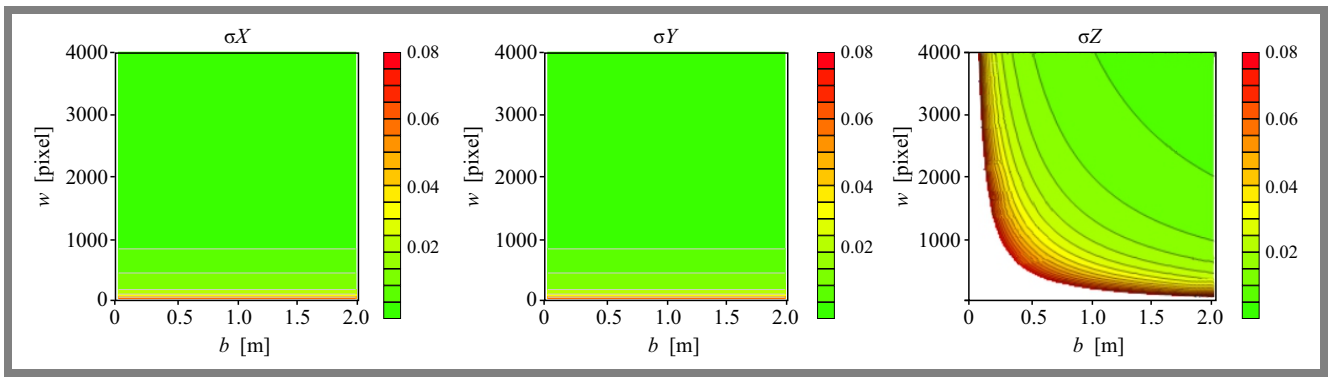

Fig. 9. Accuracy expressed as standard deviation σZ of position estimation along z axis with respect to baseline distance and image resolution w ($X = Y = 0$, $Z = 5$ m, $b = 0.5$ m, $fov = 60°$).

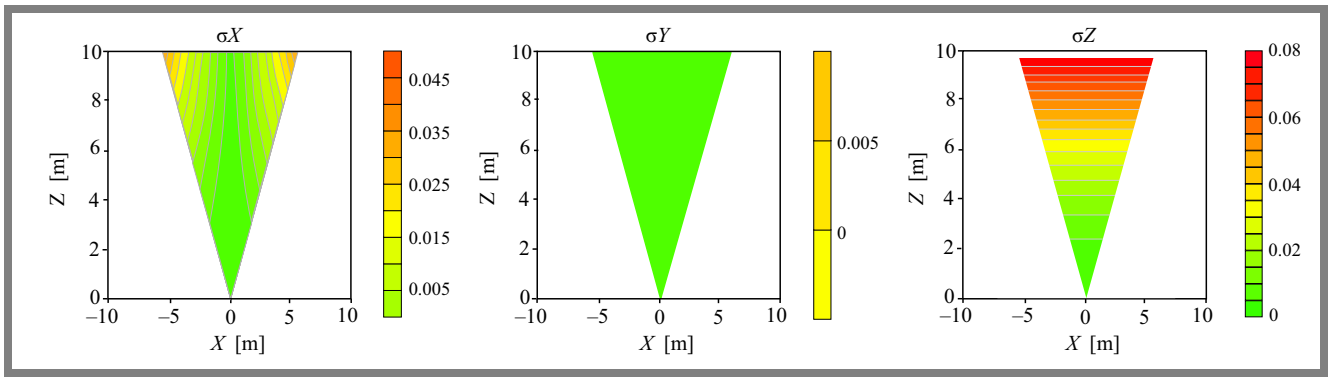

Fig. 10. Accuracy expressed as standard deviation σZ of position estimation along z axis with respect to position X, Z of point m ($b = 0.5$ m, $fov = 60°$, $w = 1920$). Accuracy is calculated only in the camera viewing cone.

clearly visible, although the actual values are less obvious. The range of values of the horizontal dimension of the image corresponds with the resolution of contemporary cameras, ranging from below Full HD (1920×1080 pixels) to 4K UHD (3840×2160 pixels).

Distance between the cameras is another factor influencing accuracy, as presented in Fig. 8.

Figure 9 shows that a doubled baseline distance and a doubled camera resolution have the same effect on position estimation accuracy σZ along the z axis. So, at locations where not much space is available to increase the baseline, the accuracy of measurements may be improved by increasing resolution. In all other cases, it is probably easier to increase the camera baseline. This remains true especially for distant objects, as long as they can be seen by both cameras after the camera baseline is increased. The fact that with a higher resolution of the camera, the requirements for calibration accuracy increase as well, is another issue that has more a practical meaning and that creates a set of completely new implications.

Figure 10 shows that position estimation accuracy σZ along the z axis is independent of positions X and Y directly in front of the camera system.

A similar analysis may be performed for the uncertainty of X and Y coordinates, and the results for the x axis are shown in Fig. 11. The lowest uncertainty can be obtained for the case when object m is positioned exactly in the middle between

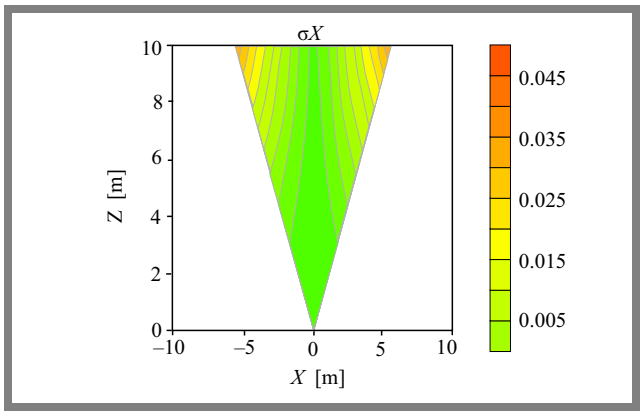

Fig. 11. Position estimation accuracy along x axis with respect to positions X and Z of point m for baseline $b = 0.5$ m, $Y = 0$, $fov = 60°$ and image resolution w = 1920. Accuracy is calculated only in the camera viewing cone.

the two cameras, along the line of symmetry. When the object moves farther away, the increase in uncertainty is negligible within the analyzed range of Z values, provided the object stays exactly in the middle between both cameras. As soon as the object moves away from the line of symmetry, the level of uncertainty rises at a rate that depends on the distance from the baseline. For example, in a system with cameras placed 50 cm apart, the object positioned 10 m from the camera baseline that is situated 5 m away from the axis of symmetry of the system has the X position uncertainty of approximately 4.5 cm. Hence, it is very important to keep the measured object in front of the camera system, i.e. close to the axis of symmetry of the camera rig.

For all the presented plots and examples, a high degree of accuracy of locating the corresponding points in the images (below 1 pixel) is considered. In practice, due to the inaccuracies of the algorithm used to localize the points, it can be higher than all standard deviation scales linearly with the accuracy.

It is also interesting to see the shape of the volume of the possible locations of point m in space. Figure 12 shows such a shape of the volume for the considered scenario, i.e. $fov = 60°$, $b = 0.5$ m, $w = 1920$. Note that for the purpose of improving legibility, the scale along the z axis is 20 times denser. The visualization does not take into account the blur of the light rays and simply assumes that perfect focusing conditions prevail and pixels of the square type are used. In reality, the shape will be bigger by at least a Minkowski product of the pixel shape and pixel light rays cone.

3.2. Three Camera Case

A three-camera system has left, center and right cameras, meaning that another device is added between the two cameras from the previous example, thus creating a system with 3 equally spaced cameras. Let us solve Eq. (18) for the case with three cameras ($N = 3$) with the assumption that the cameras are arranged along a straight line and are ideally aligned:

$$\begin{aligned} T_1 &= [-b \;\; 0 \;\; 0]^T \\ T_2 &= [\;\; 0 \;\; 0 \;\; 0]^T \\ T_3 &= [\;\; b \;\; 0 \;\; 0]^T \\ R_1 &= R_2 = R_3 = I \end{aligned} \quad , \tag{28}$$

$$K_1 = K_2 = K_3 \begin{bmatrix} f & 0 & w/2 \\ 0 & f & h/2 \\ 0 & 0 & 1 \end{bmatrix} , \tag{29}$$

$$\begin{aligned} m_1 &= [u_1 \;\; v_1 \;\; 1]^T \\ m_2 &= [u_2 \;\; v_2 \;\; 1]^T \\ m_3 &= [u_3 \;\; v_3 \;\; 1]^T \end{aligned} \quad . \tag{30}$$

where $2b$ is the camera baseline, f stands for focal length, w and h are the resolutions of the images captured with the use of the modeled cameras.

$$\sigma M' = \begin{bmatrix} \sqrt{\left|\frac{\partial X'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial X'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial X'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial X'}{\partial v_2}\right|^2 \sigma v_2^2 + \left|\frac{\partial X'}{\partial u_3}\right|^2 \sigma u_3^2 + \left|\frac{\partial X'}{\partial v_3}\right|^2 \sigma v_3^2} \\ \sqrt{\left|\frac{\partial Y'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial Y'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial Y'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial Y'}{\partial v_2}\right|^2 \sigma v_2^2 + \left|\frac{\partial Y'}{\partial u_3}\right|^2 \sigma u_3^2 + \left|\frac{\partial Y'}{\partial v_3}\right|^2 \sigma v_3^2} \\ \sqrt{\left|\frac{\partial Z'}{\partial u_1}\right|^2 \sigma u_1^2 + \left|\frac{\partial Z'}{\partial v_1}\right|^2 \sigma v_1^2 + \left|\frac{\partial Z'}{\partial u_2}\right|^2 \sigma u_2^2 + \left|\frac{\partial Z'}{\partial v_2}\right|^2 \sigma v_2^2 + \left|\frac{\partial Z'}{\partial u_3}\right|^2 \sigma u_3^2 + \left|\frac{\partial Z'}{\partial v_3}\right|^2 \sigma v_3^2} \end{bmatrix} . \tag{31}$$

$$\begin{aligned} \sigma X &= \left[\frac{1}{2b^2 f^2 (Y^2+Z^2)^2 (b^2+3(X^2+Y^2+Z^2))^2} \sigma u^2 Z^2 (9X^2(X^2+Y^2+Z^2)^2(X^2Y^2+(Y^2+Z^2)^2) + b^6(3X^2Y^2+2(Y^2+Z^2)^2) \right. \\ &\quad \left. + b^4(3X^4Y^2+4(Y^2+Z^2)^2(2Y^2+Z^2)+3X^2(5Y^4+4Y^2Z^2-Z^4)) + 3b^2(-5X^6Y^2+2(Y^2+Z^2)^4+3X^2(Y^2+Z^2)^2(Y^2+2Z^2)+X^4(-4Y^4+4Y^2Z^2+8Z^4))) \right]^{1/2} \\ \sigma Y &= \left[\frac{1}{6b^2 f^2 (Y^2+Z^2)^2 (b^2+3(X^2+Y^2+Z^2))^2} \sigma u^2 Z^2 (b^6(3Y^4+4Y^2Z^2+2Z^4) + 27Y^2(X^2+Y^2+Z^2)^2(X^2Y^2+(Y^2+Z^2)^2) \right. \\ &\quad \left. + 3b^4(7Y^6+16Y^4Z^2+13Y^2Z^4+4Z^6+X^2(3Y^4+8Y^2Z^2+4Z^4)) + 9b^2(X^4(Y^4+4Y^2Z^2+2Z^4)+(Y^2+Z^2)^2(5Y^4+6Y^2Z^2+2Z^4)+2X^2(3Y^6+10Y^4Z^2+9Y^2Z^4+2Z^6))) \right]^{1/2} \\ \sigma Z &= \left[\frac{\sigma u^2 Z^4 (b^6Y^2+3b^4(-X^2Y^2+3Y^4+4Y^2Z^2+Z^4)+27(X^2+Y^2+Z^2)^2(X^2Y^2+(Y^2+Z^2)^2)-9b^2(X^4Y^2-(Y^2+Z^2)^2(3Y^2+2Z^2)-2X^2(Y^4+4Y^2Z^2+3Z^4)))}{6b^2 f^2 (Y^2+Z^2)^2 (b^2+3(X^2+Y^2+Z^2))^2} \right]^{1/2} \end{aligned} \tag{32}$$

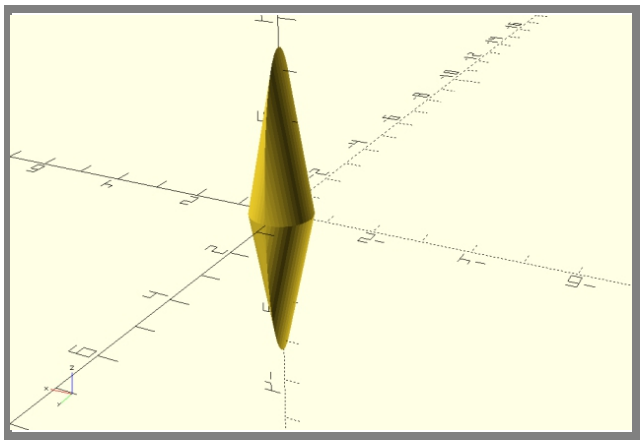

Fig. 12. Visualization of the uncertainty area of the position of point m based on the view from two cameras. z axis has been compressed 20× for better clarity. Cameras are placed below the image and the vertical axis is z.

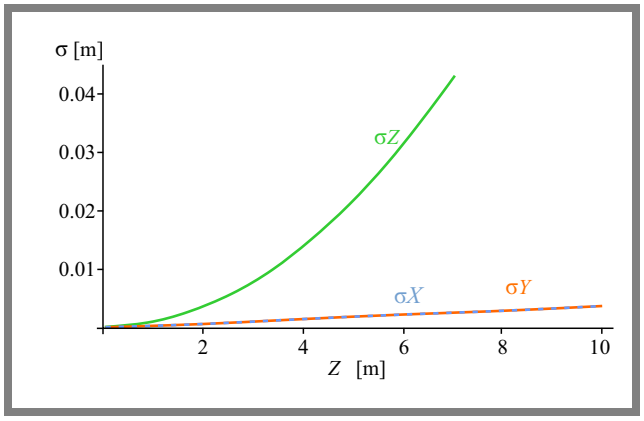

Fig. 13. Accuracy of position estimation along z axis, plotted as standard deviation σZ with varying position Z of point M.

In order to estimate uncertainty of the estimated location of point M depending on uncertainty of image positions m_1, m_2, m_3, the obtained formula for estimating position M' is expanded into a Taylor series and the uncertainty propagation rule is applied. As in the two-camera case described above, only uncertainty of the position expressed as standard deviation is considered and the remaining factors are known with infinite accuracy – see Eq. (31). Note that $\sigma M'$ is a 3-component vector.

After the transformations and by substituting Eq. (2) for obtaining u_1, v_1, u_2, v_2 from X, Y, Z, we get Eq. (32).

The derived formulas are helpful in analyzing the accuracy of estimating the position of point m depending on its location in front of the camera system.

Figure 13 illustrates a function of accuracy ΔZ with respect to the original position Z of point M, drawn for $fov = 60°$, image resolution 1920×1080 pixels, $b = 0.5$ m, and perfect alignment of the system. In Fig. 14, accuracy is expressed as standard deviation σZ with respect to camera baseline b. We have assumed that the accuracy of positions m_1, m_2 and m_3 in the captured images is $\sigma u = \sigma v = 1$ (within one pixel). Again, as for the two-camera case, doubling the distance between cameras lowers uncertainty 2 times (see Fig. 15).

As can be seen from Figs. 5 and 14, the resulting uncertainty in the Z direction remains unchanged. No improvement is therefore achieved by adding a third camera in the case of linear arrangement of the cameras – see Fig. 16.

The key to obtaining a better estimation is to arrange the cameras so that they are not situated along the same line. Here, we will analyze a triangular arrangement of 3 devices, where the cameras are situated at the vertices of an equilateral triangle, as shown in Fig. 17. The optical axes of all cameras are once again perpendicular to the plane defined by the triangle. It needs to be noted that in both arrangements the baseline between the most distant cameras is the same and equals $2b$.

As already considered above, the largest distance between the farthest cameras is an important factor limiting the accuracy of position estimation. Therefore, the biggest baseline between the cameras is fixed to be the same for both systems, as shown in Fig. 17.

Solving Eq. (18) for the case with three cameras ($N = 3$), with the assumption of a triangular camera arrangement, we get:

$$\begin{aligned} T_1 &= \begin{bmatrix} -b & -\frac{b\sqrt{3}}{2} \cdot \frac{1}{3} & 0 \end{bmatrix}^T \\ T_2 &= \begin{bmatrix} 0 & \frac{b\sqrt{3}}{2} \cdot \frac{2}{3} & 0 \end{bmatrix}^T \\ T_3 &= \begin{bmatrix} b & \frac{b\sqrt{3}}{2} \cdot \frac{1}{3} & 0 \end{bmatrix}^T \\ R_1 &= R_2 = R_3 = I \end{aligned} , \tag{33}$$

$$K_1 = K_2 = K_3 = \begin{bmatrix} f & 0 & \frac{w}{2} \\ 0 & f & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} , \tag{34}$$

$$\begin{aligned} m_1 &= [u_1 \;\; v_1 \;\; 1]^T \\ m_2 &= [u_2 \;\; v_2 \;\; 1]^T \\ m_3 &= [u_3 \;\; v_3 \;\; 1]^T \end{aligned} . \tag{35}$$

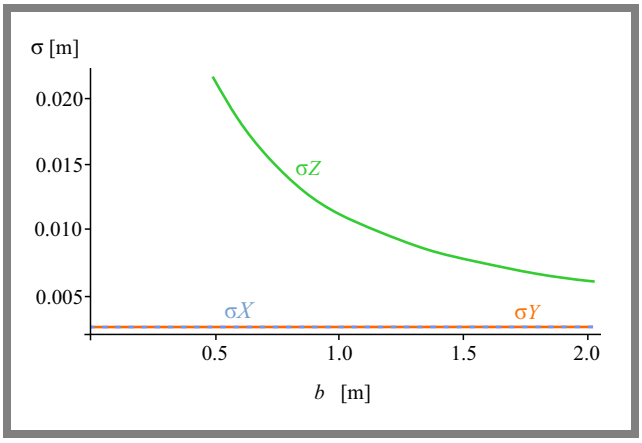

Fig. 14. Accuracy of position estimation along z axis shown as standard deviation σZ with respect to camera baseline b.

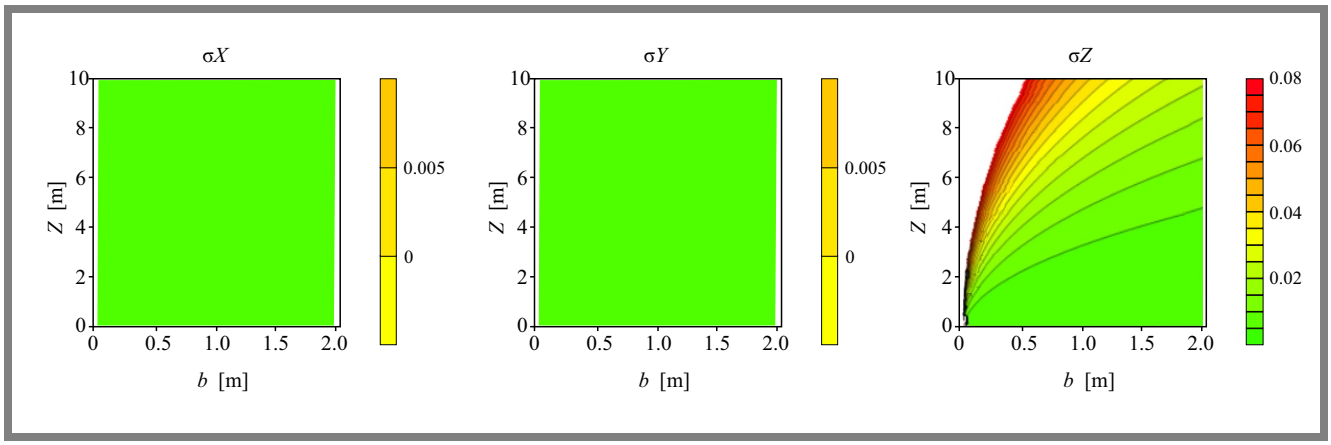

Fig. 15. Accuracy of position estimation along z axis shown as standard deviation σZ with respect to position Z of point m and camera baseline b.

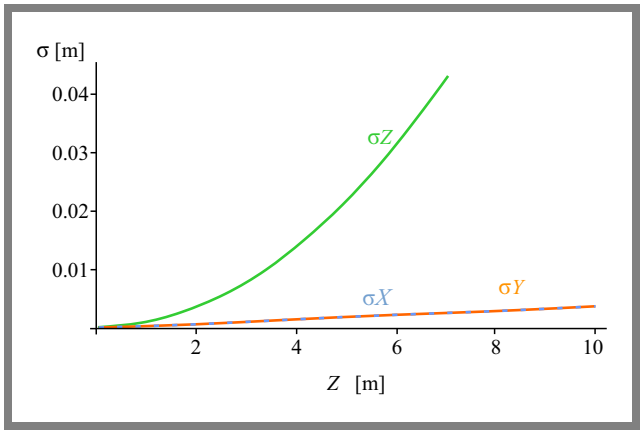

Fig. 16. Accuracy of position estimation along z axis plotted as standard deviation σZ with varying position Z of point M. The system is made up of 2 and 3 cameras with total baseline (distance between the farthest cameras) of $b = 0.5$ m. The curves are overlapping, so the middle camera in a 3-device setup does not improve the quality of estimation.

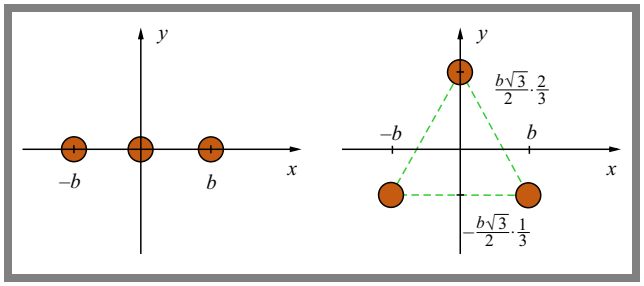

Fig. 17. Front view of the considered three camera systems.

The same procedure to assess uncertainty of the estimated location of point M is used depending on the uncertainty of the image positions m_1, m_2, m_3, as well as 1 pixel accuracy of image positions m_1, m_2, m_3 localization is assumed.

The result of the analysis is plotted in Fig. 18 as a function of accuracy, shown as standard deviation σZ with respect to the original position Z of point M for both linear and triangular systems, $fov = 60°$, image resolution 1920×1080 and $b = 0.5$ m.

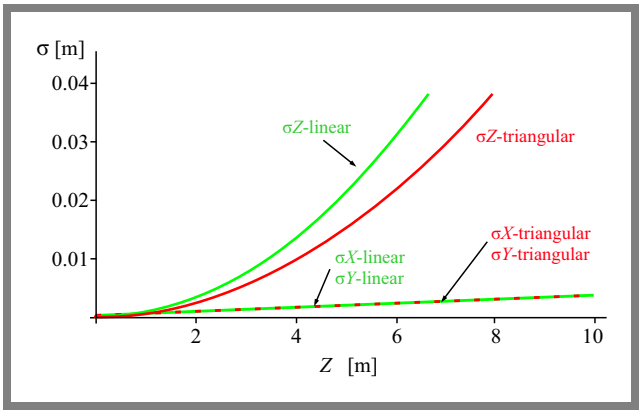

Fig. 18. Accuracy of position estimation in direction Z expressed as standard deviation σZ versus position Z of point m for system with $fov = 60°$, image resolution 1920×1080 and $b = 0.5$ m.

The plots presented in Figs. 18 and 19 confirm that the triangular camera arrangement is almost 30% better while using the same amount of space along the horizontal direction and the same number of cameras. There is no performance difference in the accuracy along x and y axes, at least when comparing a distance greater than twice the baseline $Z > 2b$. For smaller distances, the triangular arrangement performs

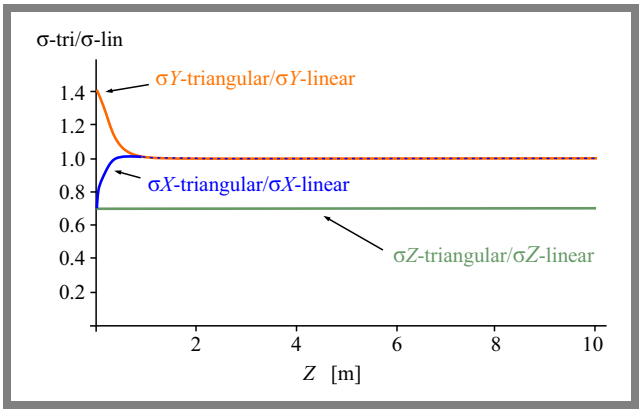

Fig. 19. Accuracy of position estimation along z axis for both setups with $fov = 60°$, image resolution 1920×1080 and $b = 0.5$ m. The distance between the two opposite cameras is equal to 1 m.

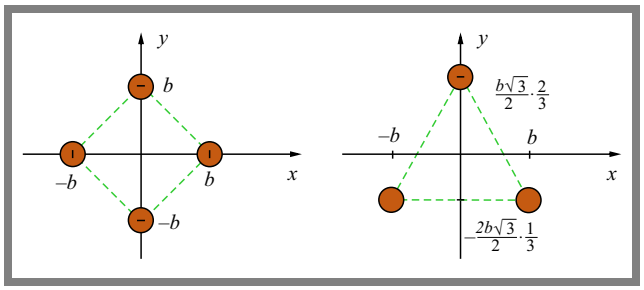

Fig. 20. Layout of a four-camera system – left, in comparison with the best three camera system – right.

differently than the linear arrangement, as evidenced in Fig. 19.

3.3. Four Camera System

Lastly, we have investigated whether the addition of a fourth camera would improve the measurement accuracy. In the setup under consideration, the cameras are positioned so that the longest distance between any two cameras is not greater than $2b$ and all inter-camera distances are maximized. To satisfy such requirements, the square camera arrangement is used, as depicted in Fig. 20.

Again, we have solved Eq. (18) with the considered assumptions:

$$\begin{aligned} T_1 &= [-b \;\; 0 \;\; 0]^T \\ T_2 &= [\;\; b \;\; 0 \;\; 0]^T \\ T_3 &= [\;0 \;\; -b \;\; 0]^T \\ T_4 &= [\;0 \;\;\; b \;\; 0]^T \\ R_1 &= R_2 = R_3 = R_4 = I \end{aligned} \quad . \tag{36}$$

$$K_1 = K_2 = K_3 = K_4 = \begin{bmatrix} f & 0 & \frac{w}{2} \\ 0 & f & \frac{h}{2} \\ 0 & 0 & 1 \end{bmatrix} , \tag{37}$$

$$\begin{aligned} m_1 &= [u_1 \;\; v_1 \;\; 1]^T \\ m_2 &= [u_2 \;\; v_2 \;\; 1]^T \\ m_3 &= [u_3 \;\; v_3 \;\; 1]^T \\ m_4 &= [u_4 \;\; v_4 \;\; 1]^T \end{aligned} , \tag{38}$$

where $2b$ is the largest camera baseline, f stands for focal length, w and h are captured image width and height, respectively.

To estimate uncertainty, as in the previous case, we assume that only uncertainty of the position expressed as standard deviation is important and the remaining coefficients are known with infinite accuracy.

Based on the derived equations, we can evaluate the accuracy expressed as standard deviation σZ versus position Z of point M for both triangular and square setups (Fig. 21). As one may see, the additional fourth camera does not improve the accuracy of the measurement (red and blue lines overlap).

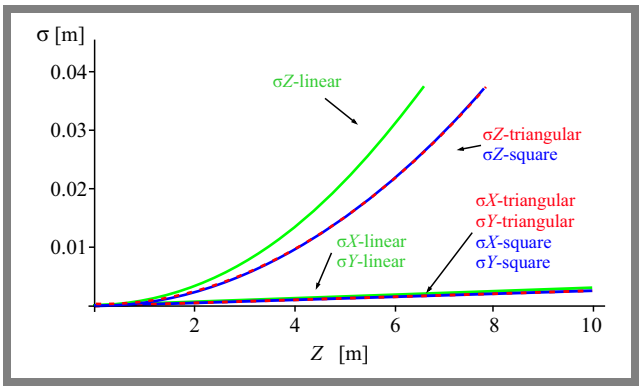

Fig. 21. Accuracy of position estimation along z axis, plotted as standard deviation σZ versus position Z of point m with $fov = 60°$, image resolution 1920×1080 and $b = 0.5$ m. The distance between cameras is as shown in Fig. 20.

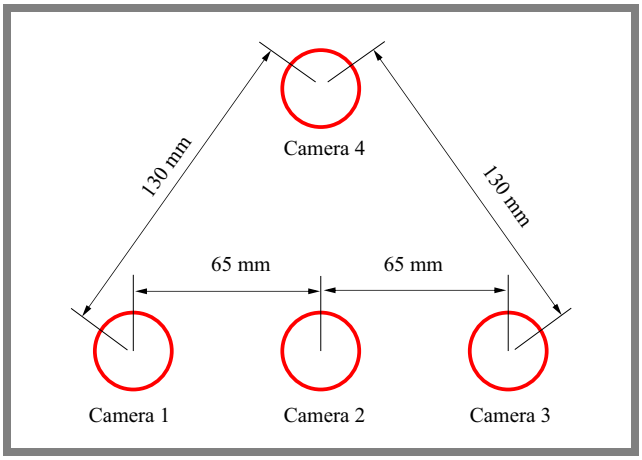

Fig. 22. Camera setup used in the experiment. The optical axes of the cameras are parallel and toward the reader.

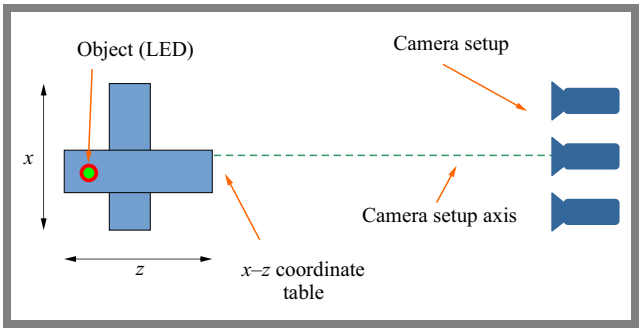

Fig. 23. Experimental setup details seen from above.

4. Experimental Verification

In order to verify the theoretical findings, we have measured by how much the light source can be moved before any changes will be observed in the images taken by the multi-camera acquisition system. As a source of light, we have used small (0603) red LED mounted on an x–z coordinate table, as such a setup allows to precisely control the position of the emitter. The camera rig and the LED have been placed 1 m from the four camera video acquisition system that is set up according to the schematic shown in Fig. 22. Three cameras, namely 1, 2 and 3, form a linear acquisition system, while cameras 1, 3

and 4 make up a triangular acquisition system. Both setups are similar to those considered previously in Section 3. The cameras used in the system were GoPro Hero 4 model with resolution of 1920×1080 pixels.

The light source was positioned at a number of different locations within a 0.2 mm grid pattern, along x and z axes. For each position of the LED, the scene was recorded by all four cameras (Fig. 23).

8×11=88 light source positions were recorded and 8×11×4 = 352 total images were processed. The 2D position of the LED on those images was estimated with the 1-pixel resolution, using both coordinates. The results were analyzed for two different triplets of cameras: the first being the linear arrangement made up of cameras 1, 2, and 3. The other set is in the triangular arrangement and is made up of cameras 1, 3, and 4.

For any triplet of the images taken by a given camera set and for different LED positions, the estimated 3D location of the LED stays the same if the estimated 2D position of the LED in all images for that camera triplet remains unchanged. In other words, for any true LED position in 3D space, as long as the 2D position in all the images does not change, the 3D position estimated from the images stays the same. Therefore, small moves of the LED in front of the camera setup will not immediately lead to a change in the captured images. There exist some ΔX and ΔZ by which the LED can be moved before the LED on the image moves by 1 pixel.

After estimating 2D positions of the LED using all cameras within the triplets, we cluster the images into groups in which the estimated 2D LED positions are the same, i.e. the estimated 3D location of the LED stays the same. Such clusters for the linear set are presented in Fig. 24 and for the triangular set in Fig. 25. Both figures represent the map of the positions of the test object on the x–z plane, being the plane of movement of the coordinate table – see Fig. 23. The dots with the same color mark the points on the x–z plane that would be projected to the same point in 3D space. As one may notice, in the case of the triangular camera arrangement the clusters are smaller than in the linear arrangement. This means that the triangular camera arrangement offers a higher degree of accuracy for 3D localization of points in 3D space.

For the linear arrangement, the average size is 3.667 mm^2, while for the triangular arrangement it equals 2.514 mm^2. The ratio of the areas equals 0.686 – a result that is very close to the value shown in Fig. 19 for the z axis. Since X and Y ratios are close to 1, the change in the area depends on Z estimation accuracy. For the 130 mm baseline, the scaled-up distance is 3.85 m and the $\frac{\Delta Z_{tri}}{\Delta Z_{lin}}$ is 0.7. The experimental result agrees with the theoretical estimation and confirms the outcome of our theoretical analysis for 3 camera arrangements.

5. Conclusions

In this paper, the problem of estimating the position of 3D objects (points) based on their 2D projections is presented and various factors impacting the accuracy of position estimation procedures used in multiview video systems, such as stereoscopic systems with a parallel optical axis, are analyzed. Based on the results of this research, the design of the system may be adjusted to achieve the required accuracy level, either by improving image resolution or increasing the camera baseline distance.

We have also proven that it is not always beneficial to add more cameras to existing systems in order to increase their accuracy. The decision depends on the geometry of the expanded system. For example, two- and three-camera systems placed along the same line show the same level of accuracy. A third camera that is not placed in line with the first two devices is the only solution that improves accuracy. Equilateral triangular camera arrangements outperform linear systems by 30%. Such an arrangement allows to obtain significantly better results for the case of ideal cameras, when compared to

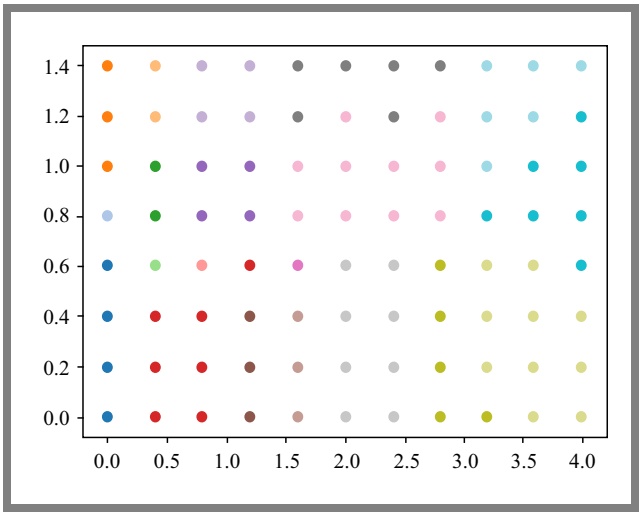

Fig. 24. Results for the linear arrangement (cameras 1, 3, and 4 as shown in Fig. 22). x (horizontal) and z (vertical) axes are scaled in millimeters. Positions marked with the same color are projected to the same point in 3D space after reconstruction.

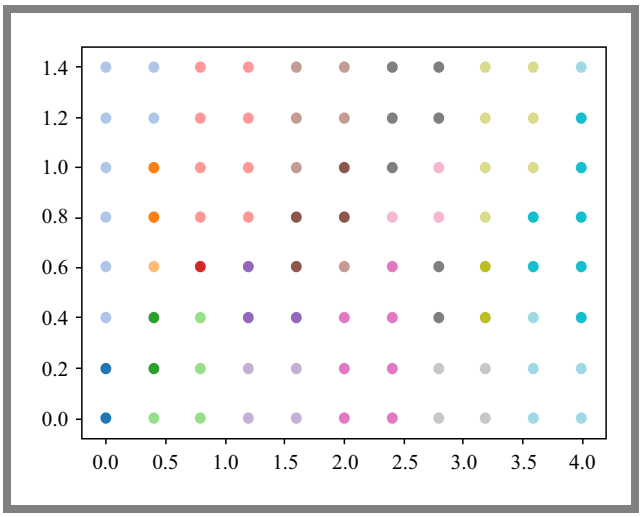

Fig. 25. Results for the triangular arrangement (cameras 1, 3, and 4 from Fig. 22). x (horizontal) and z (vertical) axes are scaled in millimeters. Positions marked with the same color are projected to the same point in 3D space after reconstruction.

a stereoscopic arrangement or three cameras in line, having the same resolution as in the triangular setup. In a real-life system with the same cameras, characterized by certain distortions and noise introduced to the image, such an arrangement also should outperform the linear setup, since any of the noise filtering algorithms can be applied, in both cases, with the same level of efficiency.

Adding a fourth camera while maintaining the same maximum distance between any two cameras in the set failed to increase the level of accuracy any further. Hence, the optimal solution for estimating 3D positions of an object consists in using a three-camera arrangement instead of the commonly used stereo pair, with the cameras positioned on the plan of an equilateral triangle. Such a setup offers, for the same baseline, a significant improvement in the accuracy of 3D position estimation when compared to the linear camera. If the number of cameras is higher than three, no improvement is achieved, since the average distance between the cameras decreases. Setups with 2 and 3 cameras are characterized by the highest average distance between cameras which is equal to the baseline.

The theoretical results were confirmed by an experiment with different camera arrangements.

Acknowledgments

The project has been financed by the National Centre for Research and Development under the LIDER Program (LIDER/34/0177/L-8/16/NCBR/2017). The work was supported by the Polish Ministry of Education and Science.

Appendix

Proof of formula (16) presented in Subsection 2.2.

$$\frac{M'^T A_i}{A_i^T A_i} A_i = \frac{A_i A_i^T}{A_i^T A_i} M' \tag{39}$$

$$\begin{aligned}
\frac{M'^T A_i}{A_i^T A_i} A_i &= \frac{1}{A_i^T A_i} \left(X A_{i,X} + Y A_{i,Y} + Z A_{i,Z}\right) A_i \\
&= \frac{1}{A_i^T A_i} \begin{bmatrix} (X A_{i,X} + Y A_{i,Y} + Z A_{i,Z}) A_{i,X} \\ (X A_{i,X} + Y A_{i,Y} + Z A_{i,Z}) A_{i,Y} \\ (X A_{i,X} + Y A_{i,Y} + Z A_{i,Z}) A_{i,Z} \end{bmatrix} \\
&= \frac{1}{A_i^T A_i} \begin{bmatrix} X A_{i,X} A_{i,X} + Y A_{i,Y} A_{i,X} + Z A_{i,Z} A_{i,X} \\ X A_{i,X} A_{i,Y} + Y A_{i,Y} A_{i,Y} + Z A_{i,Z} A_{i,Y} \\ X A_{i,X} A_{i,Z} + Y A_{i,Y} A_{i,Z} + Z A_{i,Z} A_{i,Z} \end{bmatrix} \\
&= \frac{1}{A_i^T A_i} \begin{bmatrix} A_{i,X} A_{i,X} A_{i,Y} A_{i,X} A_{i,Z} A_{i,X} \\ A_{i,X} A_{i,Y} A_{i,Y} A_{i,Y} A_{i,Z} A_{i,Y} \\ A_{i,X} A_{i,Z} A_{i,Y} A_{i,Z} A_{i,Z} A_{i,Z} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \\
&= \frac{1}{A_i^T A_i} A_i A_i^T M' = \frac{A_i A_i^T}{A_i^T A_i} M'
\end{aligned} \tag{40}$$

References

[1] T. Grajek, R. Ratajczak, M. Domański, and K. Wegner, "Limitations of Vehicle Length Estimation Using Stereoscopic Video Analysis", *20th International Conference on Systems, Signals and Image Processing IWSSIP*, Bucharest, Romania, pp. 27–30, 2013 (https:/doi.org/10.1109/IWSSIP.2013.6623441).

[2] R. Ratajczak, T. Grajek, K. Wegner, K. Klimaszewski, M. Kurc, and M. Domański, "Vehicle Dimensions Estimation Scheme Using AAM on Stereoscopic Video", *10th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS*, Krakow, Poland, pp. 478–482, 2013 (https:/doi.org/10.1109/AVSS.2013.6636686).

[3] R.I. Hartley, "Theory and Practice of Projective Rectification", *International Journal of Computer Vision*, vol. 35, pp. 115–127, 1999 (https:/doi.org/10.1023/A:1008115206617).

[4] L. Traxler, L. Ginner, S. Breuss, and B. Blaschitz, "Experimental Comparison of Optical Inline 3D Measurement and Inspection Systems", *IEEE Access*, vol. 9, pp. 53952–53963, 2021 (https:/doi.org/10.1109/ACCESS.2021.3070381).

[5] L. Wang *et al.*, "Parallax Attention for Unsupervised Stereo Correspondence Learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2108–2125, 2022 (https:/doi.org/10.1109/TPAMI.2020.3026899).

[6] S. Wang *et al.*, "Horizontal Attention Based Generation Module for Unsupervised Domain Adaptive Stereo Matching", *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6779–6786, 2023 (https:/doi.org/10.1109/LRA.2023.3313009).

[7] N. Snavely, S.M. Seitz, and R. Szeliski, "Photo Tourism: Exploring Photo Collections in 3D", *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 835–846, 2006 (https:/doi.org/10.1145/1141911.1141964).

[8] M. Yao and B. Xu, "A Dense Stereovision System for 3D Body Imaging", *IEEE Access*, vol. 7, pp. 170907–170918, 2019 (https:/doi.org/10.1109/ACCESS.2019.2955915).

[9] J. Odmins *et al.*, "Comparison of Passive and Active Fiducials for Optical Tracking", *Latvian Journal of Physics and Technical Sciences*, vol. 59, no. 5, pp. 46–57, 2022 (https:/doi.org/10.2478/lpts-2022-0040).

[10] C. Ye *et al.*, "Multi-viewpoint Optical Positioning Algorithm Based on Minimizing Reconstruction Error", *Measurement*, vol. 218, art. no. 113206, 2023 (https:/doi.org/10.1016/j.measurement.2023.113206).

[11] T. Zhang, J. Wang, S. Song, and M.Q.-H. Meng, "Wearable Surgical Optical Tracking System Based on Multi-Modular Sensor Fusion", *IEEE Transactions on Instrumentation and Measurement*, vol. 71, art no. 5006211, 2022 (https:/doi.org/10.1109/TIM.2022.3150828).

[12] V. Peretroukhin, J. Kelly, and T.D. Barfoot, "Optimizing Camera Perspective for Stereo Visual Odometry", *2014 Canadian Conference on Computer and Robot Vision*, Montreal, Canada, 2014 (https:/doi.org/10.1109/CRV.2014.9).

[13] H. M. Becerra and C. Sagues, *Visual Control of Wheeled Mobile Robots*, Springer, 118 p., 2014 (https:/doi.org/10.1007/978-3-319-05783-5).

[14] D. Murray and J.J. Little, "Using Real-Time Stereo Vision for Mobile Robot Navigation", *Autonomous Robots*, vol. 8, no. 2, pp. 161–171, 2000 (https:/doi.org/10.1023/A:1008987612352).

[15] V. Tadić *et al.*, "Application of the ZED Depth Sensor for Painting Robot Vision System Development", *IEEE Access*, vol. 9, pp. 117845–117859, 2021 (https:/doi.org/10.1109/ACCESS.2021.3105720).

[16] A. Geiger, P. Lenz, and R. Urtasun, "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite", *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2012 (https:/doi.org/10.1109/CVPR.2012.6248074).

[17] L. Yang *et al.*, "Vehicle Speed Measurement Based on Binocular Stereovision System", *IEEE Access*, vol. 7, pp. 106628–106641, 2019 (https:/doi.org/10.1109/ACCESS.2019.2932120).

[18] L. Matthies and S. Shafer, "Error Modeling in Stereo Navigation", *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987 (https:/doi.org/10.1109/JRA.1987.1087097).

[19] C. Chang and S. Chatterjee, "Quantization Error Analysis in Stereo Vision", *Conference Record of the Twenty-Sixth Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1037–1041, 1992 (https:/doi.org/10.1109/ACSSC.1992.269140).

[20] F. Fooladgar, S. Samavi, and S.M.R. Soroushmehr, "Geometrical Analysis of Altitude Estimation Error Caused by Pixel Quantization in Stereo Vision", *20th Iranian Conference on Electrical Engineering (ICEE2012)*, Tehran, Iran, 2012 (https:/doi.org/10.1109/IranianCEE.2012.6292444).

[21] F. Fooladgar, S. Samavi, S.M.R. Soroushmehr, and S. Shirani, "Geometrical Analysis of Localization Error in Stereo Vision Systems", *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4236–4246, 2013 (https:/doi.org/10.1109/JSEN.2013.2264480).

[22] C. Freundlich, M. Zavlanos, and P. Mordohai, "Exact Bias Correction and Covariance Estimation for Stereo Vision", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, pp. 3296–3304, 2015 (https:/doi.org/10.1109/CVPR.2015.7298950).

[23] L. Yang, B. Wang, R. Zhang, H. Zhou, and R. Wang, "Analysis on Location Accuracy for the Binocular Stereo Vision System", *IEEE Photonics Journal*, vol. 10, no. 1, 2017 (https:/doi.org/10.1109/JPHOT.2017.2784958).

[24] W. Sankowski, M. Włodarczyk, D. Kacperski, and K. Grabowski, "Estimation of Measurement Uncertainty in Stereo Vision System", *Image and Vision Computing*, vol. 61, pp. 70–81, 2017 (https:/doi.org/10.1016/j.imavis.2017.02.005).

[25] D. Jin, Y. Yang, "Sensitivity Analysis of the Error Factors in the Binocular Vision Measurement System", *Optical Engineering*, vol. 57, no. 10, pp. 104–109, 2018 (https:/doi.org/10.1117/1.OE.57.10.104109).

[26] X. Liu, W. Chen, H. Madhusudanan, L. Du, and Y. Sun, "Camera Orientation Optimization in Stereo Vision Systems for Low Measurement Error", *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 2, pp. 1178–1182, 2021 (https:/doi.org/10.1109/TMECH.2020.3019305).

[27] S. Gao, X. Chen, X. Wu, T. Zeng, and X. Xie, "Analysis of Ranging Error of Parallel Binocular Vision System", *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, Beijing, China, pp. 621–625, 2020 (https:/doi.org/10.1109/ICMA49215.2020.9233770).

[28] D. Liaw, S. Zhao, and Y. Hu, "A Study of Image Processing Based Object Depth Estimation", *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, Jeju, South Korea, pp. 949–954, 2019 (https:/doi.org/10.23919/ICCAS47443.2019.8971553).

[29] X. Guan, X. Li, J. Su, H. Pan, and L. Geng, "Integrative Evaluation of the Optimal Configuration for the Measurement of the Line Segments Using Stereo Vision", *Optik*, vol. 124, no. 11, pp. 1015–1018, 2013 (https:/doi.org/10.1016/J.IJLEO.2013.01.018).

[30] P. Pinggera, D. Pfeiffer, U. Franke, and R. Mester, "Know Your Limits: Accuracy of Long Range Stereoscopic Object Measurements in Practice", *Proceedings of 13th European Conference on Computer Vision*, Zurich, Switzerland, 2014 (https:/doi.org/0.1007/978-3-319-10605-2_7).

[31] L. Yu and G. Lubineau, "Modeling of Systematic Errors in Stereo-digital Image Correlation Due to Camera Self-heating", *Scientific Reports*, vol. 9, art. no. 6567, 2019 (https:/doi.org/10.1038/s41598-019-43019-7).

[32] U.R. Dhond and J.K. Aggarwal, "Binocular Versus Trinocular Stereo", *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 3, pp. 2045–2050, 1990 (https:/doi.org/10.1109/ROBOT.1990.126306).

[33] J. Wang *et al.*, "Stereo Matching Optimization with Multi-baseline Trinocular Camera Model", *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, London, Canada, 2020 (https:/doi.org/10.1109/CCECE47787.2020.9255786).

[34] S. Bi *et al.*, "High Precision Optical Tracking System Based on Near Infrared Trinocular Stereo Vision", *Sensors*, vol. 21, no. 7, 2021 (https:/doi.org/10.3390/s21072528).

[35] Y. Gao, H. Cui, X. Wang, and Z. Huang, "Novel Precision Vision Measurement Method Between Area-Array Imaging and Linear-Array Imaging Especially for Dynamic Objects", *IEEE Transactions on Instrumentation and Measurement*, vol. 71, art. no. 5022609, 2022 (https:/doi.org/10.1109/TIM.2022.3207826).

[36] W. Förstner and B.P. Wrobel, *Photogrammetric Computer Vision*, Springer Cham, 816 p., 2016 (https:/doi.org/10.1007/978-3-319-11550-4).

[37] B. Cyganek and J.P. Siebert, *An Introduction to 3D Computer Vision Techniques and Algorithms*, Hoboken: Wiley, 512 p., 2009 (https:/doi.org/10.1002/9780470699720).

[38] D. Mieloch and A. Grzelka, "Segmentation-based Method of Increasing the Depth Maps Temporal Consistency", *International Journal of Electronics and Telecommunications*, vol. 64, no. 3, pp. 293–298, 2018 (https:/doi.org/10.24425/123521).

[39] H. Xu and J. Zhang, "AANet: Adaptive Aggregation Network for Efficient Stereo Matching", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, pp. 1956–1965, 2020 (https:/doi.org/10.1109/CVPR42600.2020.00203).

[40] O. Stankiewicz and M. Domański, "Depth Map Estimation based on Maximum a Posteriori Probability", *IEIE Transactions on Smart Processing and Computing*, vol. 7, no. 1, pp. 49-61, 2018 (https:/doi.org/10.5573/IEIESPC.2018.7.1.049).

[41] D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz, and M. Domański, "Graph-based Multiview Depth Estimation Using Segmentation", *IEEE International Conference on Multimedia and Expo ICME*, Hong Kong, China, pp. 217–222, 2017 (https:/doi.org/10.1109/ICME.2017.8019532).

[42] R. Brandt, N. Strisciuglio, N. Petkov, and M. Wilkinson, "Efficient Binocular Stereo Correspondence Matching with 1-D Max-trees", *Pattern Recognition Letters*, vol. 135, pp. 402–408, 2020 (https:/doi.org/10.1016/j.patrec.2020.02.019).

[43] B. Busam, M. Hog, S. McDonagh, and G. Slabaugh, "SteReFo: Efficient Image Refocusing with Stereo Vision", *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, South Korea, pp. 3295–3304, 2019 (https:/doi.org/10.1109/ICCVW.2019.00411).

[44] B. Wilburn *et al.*, "High Performance Imaging Using Large Camera Arrays", *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 765–776, 2005 (https:/doi.org/10.1145/1073204.1073259).

[45] C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwass, "Automated Robust Metric Calibration Algorithm for Multifocus Plenoptic Cameras", *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 5, pp. 1197–1205, 2016 (https:/doi.org/10.1109/TIM.2015.2507412).

[46] O. Stankiewicz *et al.*, "A Free-viewpoint Television System for Horizontal Virtual Navigation", *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2182–2195, 2018 (https:/doi.org/10.1109/TMM.2018.2790162).

[47] N.S. Nair and M.S. Nair, "Scalable Multi-view Stereo Using CMA-ES and Distance Transform-based Depth Map Refinement", *13th International Conference on Machine Vision*, Rome, Italy, 2021 (https:/doi.org/10.1117/12.2587241).

[48] Z. Zhang, "A Flexible New Technique for Camera Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000 (https://doi.org/10.1109/34.888718).

[49] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", *Proceedings of the 4th Alvey Vision Conference*, UK, 1988, pp. 147–151, 1988 (https://doi.org/10.5244/C.2.23).

Krzysztof Klimaszewski, Ph.D.
Institute for Multimedia Telecommunications
https://orcid.org/0000-0001-5403-620X
E-mail: Krzysztof.Klimaszewski@put.poznan.pl
Poznan University of Technology, Poznań, Poland
http://www.multimedia.edu.pl

Tomasz Grajek, Ph.D.
Institute for Multimedia Telecommunications
https://orcid.org/0000-0002-7142-9416
E-mail: Tomasz.Grajek@put.poznan.pl
Poznan University of Technology, Poznań, Poland
http://www.multimedia.edu.pl

Krzysztof Wegner, M.Sc.
https://orcid.org/0000-0002-5671-0541
E-mail: kwegner@mucha.be
Mucha sp. z o.o., Poznań, Poland
https://www.mucha.be