



## Research paper

# Camera Arrangement using Geometric Optimization to Minimize Localization Error in Stereo-vision Systems

H. K. Ardakani<sup>1\*</sup>, A. Mousavinia<sup>1</sup> and Farzad Safaei<sup>2</sup>

1. Electrical and Computer Engineering Faculty, K.N.Toosi University of Technology, Tehran, Iran.

2. Faculty of Engineering and Information Sciences, University of Wollongong, New South Wales, Australia.

## Article Info

### Article History:

Received 03 July 2020

Revised 06 April 2021

Accepted 14 May 2021

DOI:10.22044/jadm.2021.9855.2117

### Keywords:

Computer Vision, Camera Arrangement, Correspondence Field, Geometric Optimization.

\*Corresponding author:  
h.k.ardakani@gmail.com(H. K. Ardakani).

## Abstract

Stereo-machine vision can be used as a space sampling technique, and the cameras parameters and configuration can effectively change the number of samples in each volume of space called the space sampling density (SSD). Using the concept of voxels, in this work, we presents a method to optimize the geometric configuration of the cameras in order to maximize SSD, which means minimizing the voxel volume and reducing the uncertainty in localizing an object in a 3D space. Each pixel's field of view (FOV) is considered as a skew pyramid. The uncertainty region will be created from the intersection of two pyramids associated with any of the cameras. Then the mathematical equation of the uncertainty region is developed based on the correspondence field as a criterion for the localization error including the depth error as well as the X and Y axes error. This field is completely dependent on the internal and external parameters of the cameras. Given the mathematical equation of the localization error, the camera's configuration optimization is addressed in a stereo-vision system. Finally, the validity of the proposed method is examined by the simulation and empirical results. These results show that the localization error is significantly decreased in the optimized camera configuration.

## 1. Introduction

In the area of machine vision, one of the most important issues is finding the 3D position estimation of objects [1]. A method for the semi-dense monocular simultaneous localization and mapping (SLAM) has been presented by Zhou, Y. in [2]. In this method, a probabilistic depth map model built on the Bayesian estimation is combined with the main framework of the state-of-the-art direct method LSD-SLAM. In most research works, the 3D location of an object has been determined by solving the geometrical equations according to the camera configurations after finding the corresponding pixels in each image pairs [3]. Furthermore, we require the camera parameters that are obtained through the camera calibration [4]. The 3D location of the object may be obtained from the intersection of two corresponding rays crossing the center of

cameras and the center of the corresponding pixels. Thus a more accurate estimation of this intersection results in a more accurate 3D object localization. Several methods have been proposed in order to solve this problem [5-11]. Some of these methods have studied the geometry of light and rays [6-7].

The method presented in [6] moves the probabilistic corresponding pixels to force the rays to intersect each other at the desired object surface. In [9], the mid-point of the common perpendicular of these two rays is considered as the intersection point. In order to improve this method, some optimization techniques have been proposed in [8, 10, 11]. Rafael Weilharter in [8] has proposed an end-to-end deep learning architecture for a 3D reconstruction from the high-resolution images. While many approaches

focus on improving the reconstruction quality alone, he primarily focused on decreasing the memory requirements in order to exploit the abundant information provided by the modern high-resolution cameras. The 3D object reconstruction from depth image streams using the Kinect-style depth cameras has been extensively studied in [10]. In [11], Zheng-Ning Liu et al. have introduced a novel computation- and memory-efficient cascaded 3D convolutional network architecture, which learns to reconstruct the implicit surface representations as well as the corresponding color information from the noisy and imperfect RGB-D maps. This proposed 3D neural network performs reconstruction in a progressive and coarse-to-fine manner, achieving an unprecedented output resolution and fidelity.

Some works have considered the pixel quantization in a CCD as an error source. In [1], the intersection of pyramids originated from two pixels has been used in order to estimate the localization error. They have used the approximate volume of the intersection region as the localization error. In [12], a model has been proposed in order to analyze the quantization error in CCD of cameras using a geometrical method. In this research work, it was assumed that each pixel had a circular shape, and FOV of each pixel was a conic. Therefore, in order to find the object's localization error using two cameras, the intersection of two corresponding cones from the cameras was used. Since finding the intersection of two cones, specially two skew cones, is very complex, three methods have been proposed in order to simplify this problem. In the first proposed method, all the points mapped to a pixel have been considered. In the second method, intersection of a ray and a cone has been used instead of the intersection of two cones. In the third method, using the Lagrange method, the minimum and maximum points of the intersection region of two cones have been calculated in all three dimensions.

Many other research works have considered the cameras arrangement in the multi-view and stereo-vision systems. Since the distance between the cameras (baseline length) in the stereo-vision system is often constant, the error increases when the object is further away from the camera. In [13], a stereo-vision system has been proposed with a variable camera distance, which is constructed using a fast slide bar. In this system, the distance between the cameras is varied by

sliding them along this bar. In addition, a method to control the distance between the cameras (baseline) has been proposed based on the object distance. In [14], a system with multiple cameras and different baselines and resolutions has been proposed in order to keep the localization error constant. Malik and Bajcsy [15] have discussed the configuration of the cameras in a stereo-vision system. They positioned the cameras to improve the image resolution and to reduce the localization error. When two cameras are close to each other, the depth perception error increases compared to when the distance between the two cameras is large. On the other hand, the depth resolution increases as long as the distance between the cameras and the object decreases. Therefore, in this study, the authors have used the genetic algorithm and gradient descent methods in order to find proper locations for the cameras with respect to each other and the object.

The statistical behavior of the 3D localization error has been studied in [16, 17]. In [16], the parameters of a stereo-vision system have been studied in order to minimize the 3D reconstruction error. They have obtained the error variance based on the stereo-vision system parameters, and classified the model error under two categories. In the first category, the quantization error and the worst-case analysis state have been considered in order to obtain the upper and lower limits of error. In the second category, the Gaussian error has been studied in order to analyze the localization error. In [17], Wenhardt has proposed 2D and 3D models to find an analytical model for the localization error and the stereo vision-system parameters. In the 2D model, the distance between the cameras and the optimum focal length is obtained by minimizing the error, while in the 3D model, a Monte Carlo simulation is applied.

The concept of Correspondence Filed (CF) has been proposed in [18], which can represent this problem in a mathematical model. CF describes the spatial topology of the intersecting rays of the cameras, arranged in a number of layers or surfaces with same disparity values, referred to as the iso-disparity layers [18]. This field is completely dependent on the cameras' configuration and position and their view angle toward the scene as well as the external and internal parameters of the cameras. Therefore, the intersecting region of the cameras' FOV depends on the CF parameters. In reference [19], a mathematical framework has been developed in

order to investigate the geometrical variations of these layers with respect to the camera intrinsic and extrinsic parameters. In [20, 21], the optimization of camera arrangement to reduce the error of depth estimation has been addressed by the application of CF. This research work has employed the iterative calculations, and all the CF parameters have not been assumed as variable.

In this study, the 3D Euclidean distance error has been used as the error measure, and is simply named as the localization error from now on. In this method, each pixel is considered as a square, and therefore its FOV will be in the form of a skew pyramid. The uncertainty region will be created from the intersection of the two pyramids originated from each center of cameras. The absolute value of the volume of this region is proportional to the localization error. The CF theory is utilized in order to obtain the required mathematic for this calculation. The mathematical equations obtained are then used to optimize the parameters of the correspondence field in a stereo-vision system including the configuration and position of the cameras and their view angles toward the scene and also the camera parameters. The rest of this paper is organized as what follows. In Section 2, the camera model and the correspondence field are discussed. In Sections 3 and 4, the proposed mathematical model and optimization are studied. In Section 5, the simulation results are presented. In the final section, the conclusions are presented.

## 2. Camera Model and Correspondence Field

Typically, a pin hole camera model is utilized for camera modeling [18-21]. Equations (1-3) state the mathematical relationship between the spatial coordinates  $(X', Y', Z')$  and the image coordinates  $(x, y)$  in the pin hole camera model.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (1)$$

$$R = R_X \times R_Y \times R_Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) \\ 0 & \sin(\theta_x) & \cos(\theta_x) \end{bmatrix} \times \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) \\ 0 & 1 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix} \times \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) & 0 \\ \sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$T = \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} \quad (3)$$

where  $f$  is the camera focal length,  $R$  is the rotation matrix, and  $T$  is the translation vector with respect to space coordinate system. We will consider that the  $f$  values are equal in the  $x$  and  $y$  directions and that there is no lens distortion to simplify the equations.

The topology of intersection of the camera rays produce CF, which depends on the configuration, location, and orientation of the cameras [18]. As it is obvious in Figure 1, the intersection of the light rays creates a number of layers in the correspondence field associated with each disparity, which shows the computable depth layers in the scene. These layers are close to each other in the vicinity of the cameras, and they become farther when getting away from them. The non-uniform distribution of these layers causes the variability of accuracy in the estimation of depth at different distances. However, as each ray is actually a pyramid, and not a line in space, the real intersection of two rays is an irregular hexahedron, as shown in Figure 2. A voxel is used in order to define a regular grid in a three dimensional space, and as the volume in Figure 2 is irregular, we named it as 2 Pyramid Voxel or 2PV.

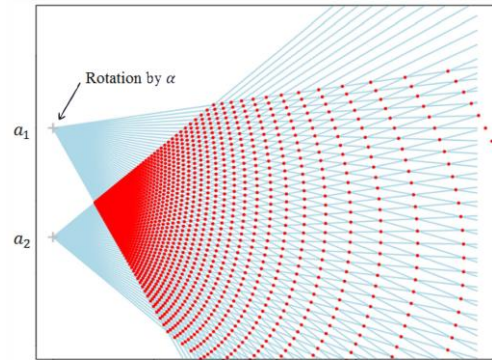


Figure 1. Intersection of light rays appears as layers in the correspondence field [18].

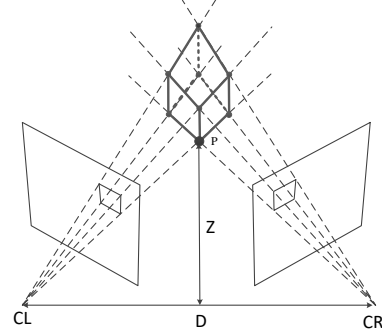


Figure 2. 2PV as element of space sampling.

### 3. Calculation of Uncertainty 3D Region: 2 Pyramid Voxel (2PV) Volume

In this section, the simple mathematic used to calculate the volume of a 2PV is presented. In the first step, the coordinates of vertices are calculated based on the correspondence field parameters, and then the volume of irregular hexahedron that is the same as the 2PV volume is calculated.

#### 3.1. Obtaining Vertices' Spatial Coordinates of 2PV

Consider two rays from cameras that are on the same epipolar plane. Supposing that there is no tilt and that the pixels of the left and right cameras are  $\mu \times \mu$ , 2PV will be created, as in Figure 2, where CL and CR are the centers of the left and right cameras and D is the baseline length. 2PV has 8 vertices, and we wish to calculate their coordinate in 3D. Consider the point P, which is the intersection of the line passing CL and one of the corners of the pixel in the left image and the similar line from the right camera.

In order to find the intersection of these two lines, consider Figure 3, which is the top view of Figure 2 in more details.

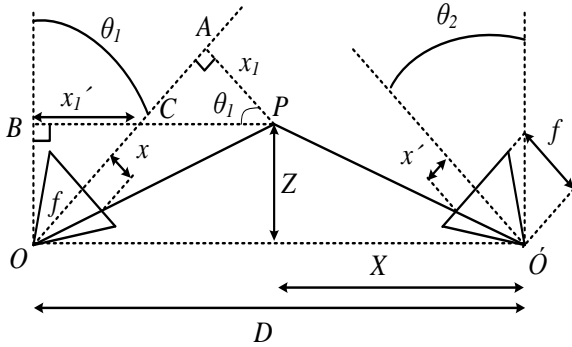


Figure 3. Top view of Figure 2.

In the ACP triangle it can be written that:

$$x_1 = (D - X - x_1') \cos(\theta_1) \quad (4)$$

By re-writing Equation (4), we have:

$$x_1 = (D - X) \cos(\theta_1) - Z \sin(\theta_1) \quad (5)$$

In the OBC and ACP triangles it can be written that:

$$OA = (D - X) \sin(\theta_1) + Z \cos(\theta_1) \quad (6)$$

By Equations (5) and (6) it can be written that:

$$\frac{x}{(D - X) \cos(\theta_1) - Z \sin(\theta_1)} = \frac{f}{(D - X) \sin(\theta_1) + Z \cos(\theta_1)} \quad (7)$$

Similarly, for the camera on the right side, we have:

$$\frac{x'}{X \cos(\theta_2) - Z \sin(\theta_2)} = \frac{f}{X \sin(\theta_2) + Z \cos(\theta_2)} \quad (8)$$

Using Equation (7), we obtain the following equation:

$$D - X = Z \frac{\cos(\theta_1) + \frac{f}{x} \sin(\theta_1)}{\frac{f}{x} \cos(\theta_1) - \sin(\theta_1)} \quad (9)$$

From Equation (8) it can be written that:

$$X = Z \frac{\cos(\theta_2) + \frac{f}{x'} \sin(\theta_2)}{\frac{f}{x'} \cos(\theta_2) - \sin(\theta_2)} \quad (10)$$

By dividing Equation (10) by Equation (9), we have:

$$\frac{X}{D - X} = K_1 \quad (11)$$

$$K_1 = \frac{x' \cos(\theta_2) + f \sin(\theta_2)}{f \cos(\theta_2) - x' \sin(\theta_2)} \times \frac{f \cos(\theta_1) - x \sin(\theta_1)}{x \cos(\theta_1) + f \sin(\theta_1)} \quad (12)$$

where  $K_1$  only depends on the parameters of the correspondence field. Now, using Equation (11), we may obtain X as follows:

$$X = \frac{D \times K_1}{1 + K_1} \quad (13)$$

Now, using Equation (10), Z may be calculated as:

$$Z = X \times K_2 \quad (14)$$

$$K_2 = \frac{f \cos(\theta_2) - x' \sin(\theta_2)}{x' \cos(\theta_2) + f \sin(\theta_2)} \quad (15)$$

Figure 4 may also be drawn along Y. The following equation is correct in the OAP triangle:

$$Y = \frac{y'}{f} (Y \sin(\theta_2) + Z \cos(\theta_2)) \quad (16)$$

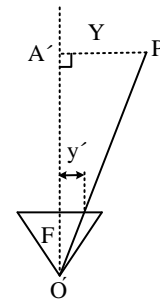


Figure 4. Drawn Figure 3 along Y.

Now, the coordinates of the intersection of two

lines from the center of two cameras is available based on the parameters of the correspondence field. This point is one of the 8 vertices of 2PV. Similarly one can calculate the coordinates of the other vertices.

### 3.2. Calculating Volume of a 2PV Given Its Vertex Point Coordinates

We may calculate the volume of the regular hexahedron of Figure 5 giving it the vertex points, as follows [22]:

Equation (17) may be written as:

$$V = \left| \det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \right| \quad (18)$$

where  $a_i$ ,  $b_i$ , and  $c_i$  are the components of the vectors  $\vec{ae}$ ,  $\vec{ab}$ , and  $\vec{ad}$ , respectively. Moreover, we may obtain each vector by subtracting its end point and starting point.

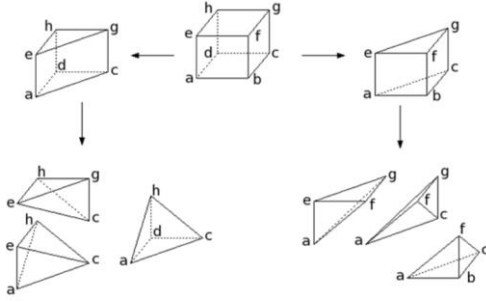


Figure 5. A regular hexahedron divided to 6 tetrahedrons.

In order to obtain the volume of a 2PV, we may divide it into 6 tetrahedrons (similar to Figure 5). Each one of these tetrahedrons may be considered as one-sixth of a regular hexahedron [22]. Therefore, we may calculate the volume of a regular hexahedron as follows:

$$V = \sum_{i=1}^6 V_i \quad (19)$$

where  $V_i$  is the volume of tetrahedron  $i$ . Due to the fact that each tetrahedron may be considered as one-sixth of a regular hexahedron, the volume of each tetrahedron may be calculated as follows:

$$V_i = \frac{1}{6} \left| \det \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix} \right| = \frac{1}{6} \left| \det \begin{bmatrix} a_{1i} & a_{2i} & a_{3i} \\ b_{1i} & b_{2i} & b_{3i} \\ c_{1i} & c_{2i} & c_{3i} \end{bmatrix} \right| \quad (20)$$

where  $a_{ji}$ ,  $b_{ji}$ ,  $c_{ji}$  for  $j = 1$  to 3 are the components of the vectors  $a_i$ ,  $b_i$ , and  $c_i$ , respectively. In this way, the volume of each

irregular hexahedron may be calculated accurately.

According to Equation (17), it should be noticed that if the order of vectors in this equation is correct, the number inside the  $| |$  is always positive. Numerically we may easily calculate the volume using Equation (17) without considering the order. However, parametrically, it is easier to notice the order, so that we will not require the absolute value operator.

$$V = |\vec{ae} \cdot (\vec{ab} \times \vec{ad})| = |\vec{ab} \cdot (\vec{ae} \times \vec{ad})| = |\vec{ad} \cdot (\vec{ab} \times \vec{ae})| \quad (17)$$

### 3.3. Calculating 2PV based on Parameters of Correspondence Field

Using Equations (13, 14), and (16) with different  $x$ ,  $x'$ , and  $y'$ , we may obtain the coordinates of the vertices of the non-regular hexahedron, which is the 2PV.

In Figure 6, the coordinate system of the left side and right side cameras are shown. The center of the coordinate system is the center of the camera ( $i$ ,  $j$ , and  $k$  are the intended pixel numbers).

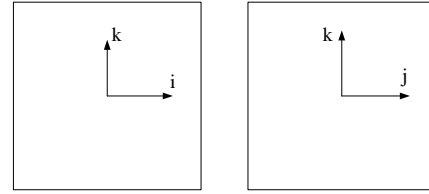


Figure 6. Coordinate system for the left side and right side cameras.

If we consider each pixel as a square with dimension  $\mu \times \mu$ , then we may calculate  $x$ ,  $x'$ , and  $y'$  in order to find the vertex points of 2PV (pixels  $i$  and  $k$  in the left-side camera and pixels  $j$  and  $k$  in the right-side camera), as shown in Table 1.

Table 1. Vertex points of 2PV.

Pixel Vertices	2PV vertices			
	$P_1$	$P_2$	$P_3$	$P_4$
$x$	$i \times \mu$	$i \times \mu$	$(i+1) \times \mu$	$(i+1) \times \mu$
$x'$	$-j \times \mu$	$-(j+1) \times \mu$	$-(j+1) \times \mu$	$-j \times \mu$
$y'$	$k \times \mu$	$k \times \mu$	$k \times \mu$	$k \times \mu$
Pixel Vertices	2PV vertices			
	$P_5$	$P_6$	$P_7$	$P_8$
$x$	$i \times \mu$	$i \times \mu$	$(i+1) \times \mu$	$(i+1) \times \mu$
$x'$	$-j \times \mu$	$-(j+1) \times \mu$	$-(j+1) \times \mu$	$-j \times \mu$
$y'$	$(k+1) \times \mu$	$(k+1) \times \mu$	$(k+1) \times \mu$	$(k+1) \times \mu$

Now, the coordinates of the vertex points of 2PV ( $P_i$ ) may be calculated using Equations (13, 14), and (16).

$$P_i^T = [X_i \ Y_i \ Z_i] \quad (21)$$

Using Equations (19) and (20), one can easily calculate the volume of a 2PV.

$$V_1 = \frac{1}{6} \det \begin{bmatrix} P_6^T - P_2^T \\ P_1^T - P_2^T \\ P_4^T - P_2^T \end{bmatrix} \quad (22)$$

while  $P_i$ s are the 3D coordinates of the 2PV vertices.

Results:

$$V_1 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_1 h_2^2}{h_3 h_4^2 h_5} \quad (23)$$

in which:

$$h_1 = f \cos(\theta_2) + (j+1)\mu \sin(\theta_2) \quad (24)$$

$$h_2 = f \cos(\theta_1) - (i+1)\mu \sin(\theta_1) \quad (25)$$

$$h_3 = (f^2 + i(j+1)\mu^2) \sin(\theta_1 + \theta_2) + f\mu(i-j-1) \cos(\theta_1 + \theta_2) \quad (26)$$

$$h_4 = (f^2 + (i+1)(j+1)\mu^2) \sin(\theta_1 + \theta_2) + f\mu(i-j) \cos(\theta_1 + \theta_2) \quad (27)$$

$$h_5 = (f^2 + (i+1)j\mu^2) \sin(\theta_1 + \theta_2) + f\mu(i-j+1) \cos(\theta_1 + \theta_2) \quad (28)$$

Similarly:

$$V_2 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_6(h_7^2 - h_8)}{h_4 h_5 h_9^2} \quad (29)$$

$$V_3 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_6 h_2^2}{h_4 h_5^2 h_9} \quad (30)$$

$$V_4 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_1 h_{10}^2}{h_3^2 h_4 h_9} \quad (31)$$

$$V_5 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_1 h_{10}^2}{h_3 h_4 h_9^2} \quad (32)$$

$$V_6 = \frac{D^3 f^2 \mu^3}{6} \cdot \frac{h_6(h_7^2 - h_8)}{h_3 h_4 h_5 h_9} \quad (33)$$

in which:

$$h_6 = f \cos(\theta_2) + j\mu \sin(\theta_2) \quad (34)$$

$$h_7 = f \cos(\theta_1) - (i+0.5)\mu \sin(\theta_1) \quad (35)$$

$$h_8 = \frac{\mu^2 \sin(\theta_1)^2}{4} \quad (36)$$

$$h_9 = (f^2 + ij\mu^2) \sin(\theta_1 + \theta_2) + f\mu(i-j) \cos(\theta_1 + \theta_2) \quad (37)$$

$$h_{10} = f \cos(\theta_1) - i\mu \sin(\theta_1) \quad (38)$$

Substituting Equation (23, 29-33) in (19), 2PV can be calculated.

#### 4. Optimization of 2PV Volume

Looking more at the  $h_i$  equations show that we can neglect the 1 and 0.5 terms compared to the

values of  $i$  and  $j$ , which are the pixel numbers, and take a value between 0 and the maximum resolution. This assumption reduces Equation (19) into a more compact one of Equation (39):

$$\tilde{V} = D^3 f^2 \mu^3 \frac{h_6 h_{10}^2}{h_9^4} \quad (39)$$

Solving Equation (9) and (10) for  $i$  and  $j$ , we have:

$$i = \frac{f[(D-X) \cos(\theta_1) - Z \sin(\theta_1)]}{\mu[Z \cos(\theta_1) + (D-X) \sin(\theta_1)]} \quad (40)$$

$$j = \frac{f[X \cos(\theta_2) - Z \sin(\theta_2)]}{\mu[Z \cos(\theta_2) + X \sin(\theta_2)]} \quad (41)$$

Now, substituting Equations (40) and (41) in Equation (39):

$$\tilde{V} = \frac{\mu^3}{DZf^3} [Z \cos(\theta_2) + X \sin(\theta_2)]^3 [Z \cos(\theta_1) + (D-X) \sin(\theta_1)]^2 \quad (42)$$

The derivative of  $\tilde{V}$  with respect to  $\theta_1$  is shown in Equation (43), and can be used to find the extremum value of equation  $\tilde{V}$ .

$$\frac{\partial \tilde{V}}{\partial \theta_1} = [Z \cos(\theta_1) + (D-X) \sin(\theta_1)] [(D-X) \cos(\theta_1) - Z \sin(\theta_1)] \quad (43)$$

Equating the right side of Equation (43) to zero yields two groups of answers as:

$$\tan(\theta_1) = \frac{D-X}{Z} \quad (44)$$

$$\tan(\theta_1) = \frac{-Z}{D-X} \quad (45)$$

The right side of Equation (44) is positive, so the answer for  $\theta_1$  is either in the first or the third quarter. When  $\theta_1$  is in first quarter, the object is on the camera FOV, and the answer is accepted, while for  $\theta_1$  on the 3th quarter, the object is located behind the camera, and the answer is rejected.

Considering Equation (45), one can see that the right side of the equation is negative, which means that  $\theta_1$  should be at the second or the fourth quarter. This means that the object is behind the camera or is not on its FOV, which is not our answer.

The same process goes for the  $\theta_2$  value, and we can use Equation (46) in order to find an acceptable answer for  $\theta_2$ .

$$\tan(\theta_2) = \frac{X}{Z} \quad (46)$$

Solving Equation (44) for the sine and cosine values of  $\theta_1$ , we have:

$$\sin(\theta_1) = \frac{D - X}{\sqrt{Z^2 + (D - X)^2}} \quad (47)$$

$$\cos(\theta_1) = \frac{Z}{\sqrt{Z^2 + (D - X)^2}} \quad (48)$$

The same would be for  $\theta_2$  as:

$$\sin(\theta_2) = \frac{X}{\sqrt{Z^2 + X^2}} \quad (49)$$

$$\cos(\theta_2) = \frac{Z}{\sqrt{Z^2 + X^2}} \quad (50)$$

Using the trigonometric identities, we obtain:

$$\sin(\theta_1 + \theta_2) = \frac{ZD}{\sqrt{Z^2 + X^2}\sqrt{Z^2 + (D - X)^2}} \quad (51)$$

$$\cos(\theta_1 + \theta_2) = \frac{Z^2 - X(D - X)}{\sqrt{Z^2 + X^2}\sqrt{Z^2 + (D - X)^2}} \quad (52)$$

Now, substituting Equations (47-52) in the Equation (29-33), we can re-write Equation (19) as:

$$\tilde{V}_\theta = \frac{D\mu^3(X^2 + Z^2)^{\frac{3}{2}}(Z^2 + (D - X)^2)}{6fF} \quad (53)$$

$F$  is itself a function of  $Z$ ,  $D$ , and  $X$  (Appendix shows the details of the  $F$  function). Using  $F$  will simplify our next equations. In order to find the optimum distance between the cameras ( $D$ ), which minimizes the  $\tilde{V}_\theta$  value, we can obtain the derivative of  $\tilde{V}_\theta$  with respect to  $D$  as:

$$\frac{\partial \tilde{V}_\theta}{\partial D} = \left[ \frac{(Z^2 + (D - X)^2 + 2D(D - X))F}{F^2} - \frac{D(Z^2 + (D - X)^2)F'}{F^2} \right] \quad (54)$$

Setting the right side of Equation (54) to zero and solving for  $D$  gives us the value of  $D$ , in which  $\tilde{V}_\theta$  would have an extremum. However, it is very difficult to find an analytic solution for  $D$ . Suppose that  $\tilde{D}$  is such an answer. Now, Equation (55) can guarantee that the right side of (54) is zero.

$$\frac{F'}{F} = \frac{1}{\tilde{D}} + \frac{2(\tilde{D} - X)}{Z^2 + (\tilde{D} - X)^2} \quad (55)$$

Equation (55) is a differential equation with a solution in the form of Equation (56), in which  $K$  is an unknown constant.

$$F = K\tilde{D}(Z^2 + (\tilde{D} - X)^2) \quad (56)$$

Re-arranging Equation (53) with the value of  $F$  as in Equation (56) and  $\tilde{D}$ , we obtain:

$$\tilde{V}_{\theta D} = \frac{\tilde{D}\mu^3(X^2 + Z^2)^{\frac{3}{2}}(Z^2 + (\tilde{D} - X)^2)}{6fK\tilde{D}(Z^2 + (\tilde{D} - X)^2)} \quad (57)$$

Considering the parameter  $X$ , we can do the same for Equation (57) and arrive at Equation (58).

$$\frac{\partial \tilde{V}_{\theta D}}{\partial X} = 0 \Rightarrow (X^2 + Z^2) \left( Z^2 + (\tilde{D} - X)^2 \right) = 0 \quad (58)$$

The left side of Equation (58) is always a positive number, so we can only minimize this value. Again we compute the derivative of Equation (58) with respect to  $X$  and equate the result to zero:

$$2X \left( Z^2 + (\tilde{D} - X)^2 \right) - 2(\tilde{D} - X)(X^2 + Z^2) = 0 \quad (59)$$

Now we get:

$$(\tilde{D} - 2X)(X(\tilde{D} - X) - Z^2) = 0 \quad (60)$$

Solving Equation (60) for  $X$ , we reach:

$$\tilde{X} = \frac{\tilde{D}}{2} \pm \sqrt{\frac{\tilde{D}^2}{4} - Z^2}, \frac{\tilde{D}}{2} \quad (61)$$

Considering Equation (61), we can conclude that:

1. If  $\tilde{D} \leq 2Z$ , then there is one real answer,  $\tilde{X} = \tilde{D}/2$  corresponding to minimum of  $\tilde{V}_{\theta D}$ .
2. If  $\tilde{D} > 2Z$ , then there are three distinct answers with  $\tilde{D}/2$  pointing to a local maximum for  $\tilde{V}_{\theta D}$  and two other answers with respect to the two local minima of  $\tilde{V}_{\theta D}$ .

The result of  $\tilde{X}$  fitting into the second term of Equation (60) would apply to Equation (62) too.

$$\frac{(\tilde{D} - X)}{Z} = \frac{Z}{X} \quad (62)$$

The left side of Equation (62) is equal to the right side of (44), and its right side is equal to the inverse of the right side of Equation (46). Thus we can say:

$$\tan(\theta_1) = \cotan(\theta_2) \quad (63)$$

The solution for Equation (63) is:

$$\theta_1 + \theta_2 = \frac{\pi}{2} + k\pi \quad k = 0, \pm 1, \pm 2, \dots \quad (64)$$

In other words, the axes of the two cameras have to be perpendicular to each other in the optimized condition. Regarding the results of Equation (57), it can be said that  $\tilde{D}$  at least has to be equal to  $2Z$ .



In such a condition, regarding Equations (42) and (43),  $\theta_1 = \theta_2 = \pi/4$ . Furthermore, if  $\tilde{D} > 2Z$ , the perpendicularity of the cameras' axes is enough, and there will be two optimized states that are the answers of Equation (61). However, in this condition, the distance of the object from two cameras is not the same and there is no geometric symmetry. In addition, according to Equation (57), the optimal camera focal length ( $f$ ) will be maximal. This maximum value should be determined in a way that the required region is within FOV of the two cameras. In this way, if the distances of this spatial region to the cameras differ, then the cameras' focal lengths have to be different as well. A larger focal length implies a larger lens and more cost. Therefore, it can be said that the optimum condition is  $\tilde{D} = 2Z$ .

$\mu$  is the last parameter of the correspondence field to be discussed. According to Equation (57), the volume of each 2PV is proportional to the third power of  $\mu$ , which is related to the CCD technology used.

## 5. Simulation and Experimental Results

In this section, the validity of the proposed method is examined by the simulation and empirical results.

### 5.1. Simulation

In order to validate our analytical results, which determine the condition of camera posing to have the minimum voxel volume, we developed a simulation setup. As the voxel size is directly related to the localization error, we used an array of  $11 \times 11 \times 11$  dots located inside a 1 cubic meter space as our test bed, as shown in Figure 7 (with  $\tilde{X} = 0$ ). The cube center is at the coordinate  $(0, \tilde{X}, 5)$ . After setting the camera parameters and  $\tilde{X}$  value, using Equations (13, 14), and (16), the position of each point in the array is calculated, and the mean RMS error with respect to the actual value is obtained.

Starting with  $D = 2Z$ , the three values of 40, 45, and 50 degrees are chosen for  $\theta_2$  and  $\theta_1$  is changed from 30 to 60 degrees in 0.03 degree steps. The mean RMS error value for each step is calculated, and the total results are shown in Figure 8. Equations (44) and (46) predict 39.95, 45, and 49.25 degrees as the optimum value for  $\theta_1$  according to the  $\theta_2$  values. Looking at Figure 8, we can see that this result is validated through the simulation too. In addition  $\theta_1 = \theta_2 = 45$  degrees corresponds to the best result.

For  $D > 2Z$ ,  $D = 11\text{ m}$ , the simulation is repeated for the  $\theta_2$  values of 32.69, 40, 47.73, 50, and

57.31 degrees. According to Equation (64), for  $D = 11\text{ m}$ , the optimum values of  $\theta_2$  and  $\theta_1$  are 32.69 and 57.31 degrees.

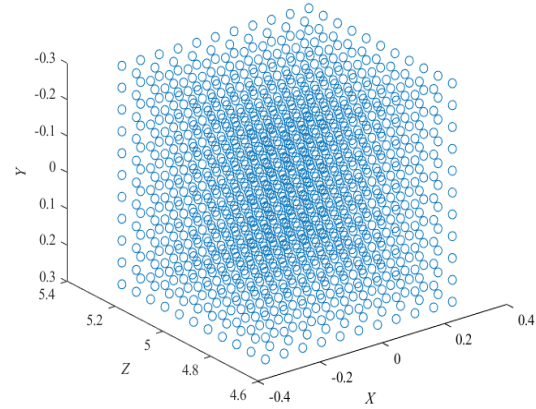


Figure 7. An array of  $11 \times 11 \times 11$  dots as our test bed for  $\tilde{X} = 0$ .

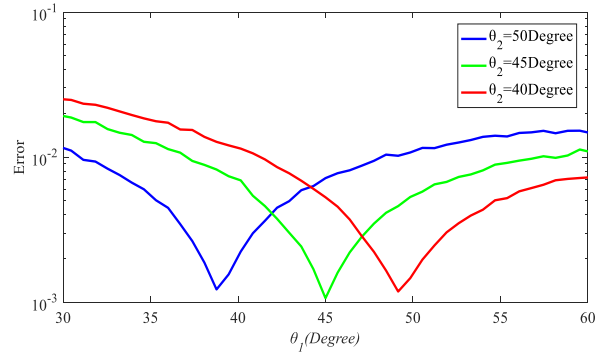


Figure 8. Mean RMS error versus variation of  $\theta_1$  for  $D = 2Z$  and different values of  $\theta_2$ .

When  $\tilde{X} = D/2$ ,  $\theta_2$  and  $\theta_1$  would be 47.73 degrees as Equations (44) and (46) predict.  $\theta_1$  in changed again from 30 to 60 degree in 0.03 degree steps, and the mean RMS error is calculated for each step, Figure 9 shows the results. As it can be seen, at 32.69 and 57.31 degree the minimum error is less than the other cases.

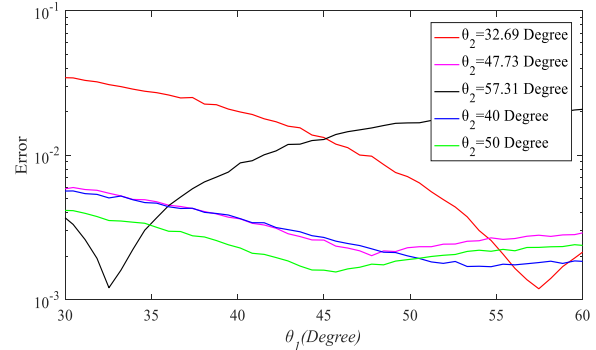


Figure 9. Mean RMS error versus variation of  $\theta_1$  for  $D > 2Z$  and different values of  $\theta_2$ .

Figure 10 shows the mean RMS error, while  $X$  is changed from 0 to 14 m in step of 0.14 m, and  $\theta_2$  and  $\theta_1$  chosen as Equations (44) and (46). Three



different cases for  $D < 2Z$  ( $D = 7\text{ m}$ ),  $D = 2Z$  ( $D = 10\text{ m}$ ),  $D > 2Z$  ( $D = 15\text{ m}$ ) are reported. It is clear that for  $D < 2Z$ , the optimized value is  $D/2 = 3.5\text{ m}$ . When  $D = 2Z$ , we have  $D/2 = 5\text{ m}$  again as the optimum value.

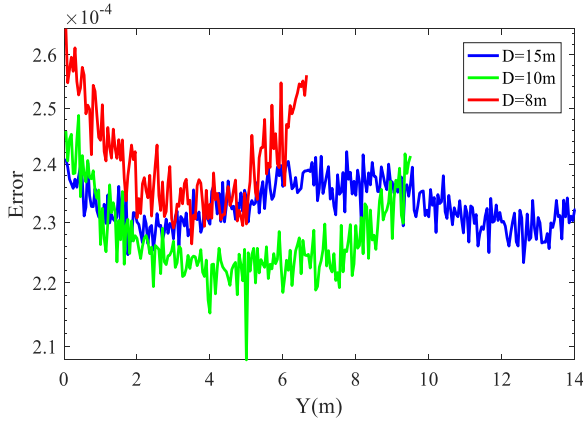


Figure 10. Mean RMS error versus  $X$  variation for different values of  $D$ .

While  $D > 2Z$ , we have two minimum values at 1.91 and 13.09 m and a local maximum at  $D/2 = 7.5\text{ m}$ . Again, the simulation results are as what Equation (61) is predicting.  $f$  and  $\mu$  are evident and require no further investigation.

## 5.2. Experimental Results

In this section, the validity of the proposed method is experimentally assessed. For this purpose, the Stanford Scanning Models are applied [23]. These models provide stereo-images per any desired configuration of the camera and object settings. Four statues of Bunny, Buddha, Dragon, and Asian dragon in the Stanford datasets are used for the test, as in Figure 11.

Five different camera setups are used for comparison. In the first setup, the cameras are arranged in a conventional rectified configuration. Equations (44) and (46) are used in order to choose the optimized value for  $\theta_1$  and  $\theta_2$  used as setup 2. In setup 3, first, the optimum value for  $D$  is chosen, and then the optimum values for  $\theta_1$  and  $\theta_2$  are calculated using Equations (44) and (46). The setups 4 and 5 are the same as the setups 2 and 3, except that the focal length has been optimized in a way that the object will fill the entire FOV. Table 2 shows the setup configurations in details.

Table 2. Five different camera setups details.

Camera arrangement	$\theta_1$	$\theta_2$	$D(\text{m})$	$X(\text{m})$	$f$
Setup 1	0	0	0.4	0.2	5300
Setup 2	2.86	2.86	0.4	0.2	5300
Setup 3	45	45	8	4	5300
Setup 4	2.86	2.86	0.4	0.2	5800
Setup 5	45	45	8	4	7500



Figure 11. Examples of the applied datasets [23].

Figures 12-a, 12-b, and 12-c show the experimental results for the setups 1, 2, and 3, respectively. Stereo-images of Bunny for setup 2 are shown in Figure 13.

Table 3 shows the mean absolute error for the three configurations, in summary. Observing the results, it is clear that the configuration of setup 3 has reduced the localization error about one order of magnitude compared to the rectified settings of setup 1.

Table 3. Mean absolute error for different camera arrangements.

Setup 3	Setup 2	Setup 1
3.33e-5	3.27e-4	3.27e-4

The second experiment is very similar to the first one, except that the focal length is increased to the maximum value, so the object will fill the entire FOV, and includes the setups 1, 4 and 5. Figures 14-a, 14-b, and 14-c show the experimental results for setups 1, 4, and 5, respectively. Table 4 shows the mean absolute error for the second experiment, in summary. In comparison to the rectified case of setup 1, we have almost a 96.7% reduction in the localization error in setup 5.

Figure 15 shows the results of repeating the experiment on Buddha, Dragon, and Asian dragon. Again, Figures 15-a, 15-b, and 15-c show the experimental results for setups 1, 4, and 5, respectively. The columns 3 to 5 in Table 4 show these results. A reduction of 30 to 40 times in the localization error can be observed.

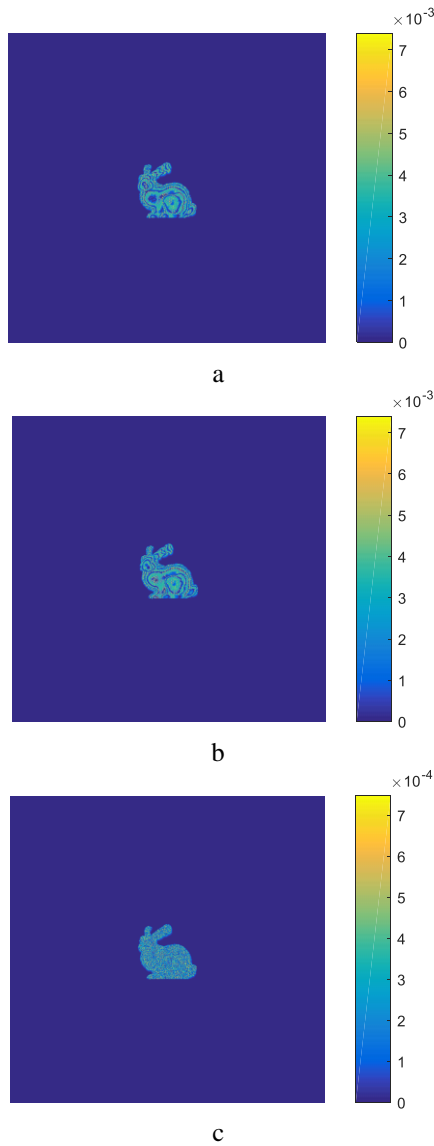


Figure 12. Effect of different CF parameters' optimization on the localization error. a) Initial parallel cameras, b)  $\theta_1$  and  $\theta_2$  optimization, c) optimization of  $\theta_1$ ,  $\theta_2$ , and  $D$ .

Table 4. Comparison between mean absolute errors for different camera arrangements.

Camera arrangement	Error reduction	Bunny	Buddha	Dragon	Asian Dragon
Setup 1	Reference Error	3.31e-4	1.60e-4	1.84e-4	1.09e-4
Setup 4	Error	1.37e-4	6.71e-5	4.37e-5	5.08e-5
	Reduction	58.6%	58.1%	76.2%	53.4%
Setup 5	Error	7.53e-6	3.42e-6	1.99e-6	3.58e-6
	Reduction	97.7%	97.8%	98.9%	96.7%

In order to evaluate the behaviour of our method, the localization error around the obtained optimum points is investigated through three experiments. In the first experiment,  $D = 2Z$  and  $X$  is chosen as the optimum value according to Equation (61), while a set of different values around the optimum points are assigned to  $\theta_1$  and  $\theta_2$ .

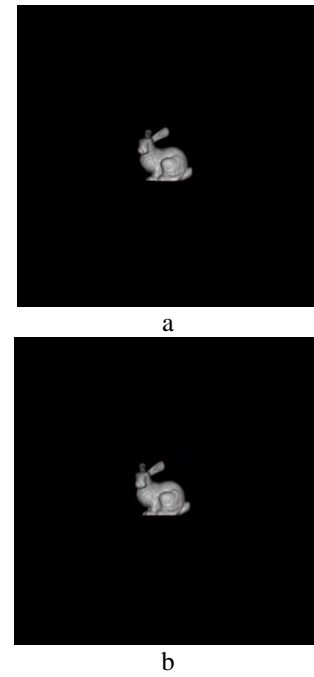


Figure 13. Stereo-images of Bunny for setup 2. a) Right camera, b) Left camera.

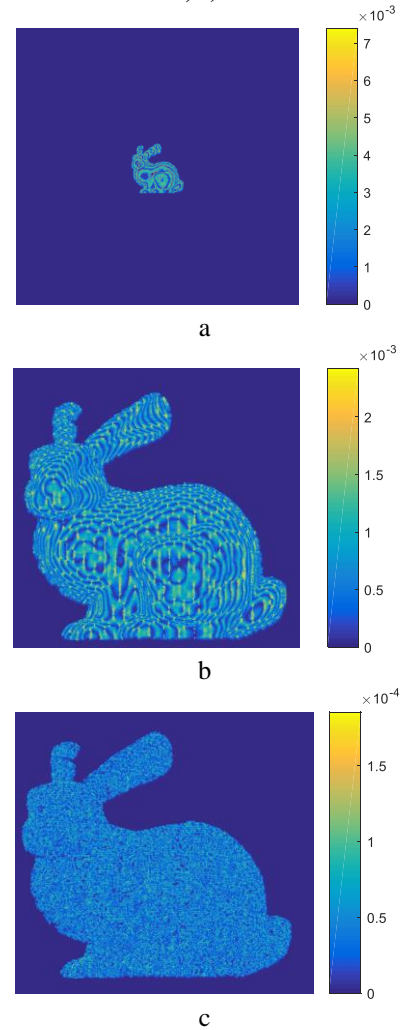


Figure 14. Effect of different CF parameters' optimization on the localization error. a) Initial parallel cameras, b)  $\theta_1$ ,  $\theta_2$ , and  $f$  optimization, c) optimization of  $\theta_1$ ,  $\theta_2$ ,  $D$ , and  $f$ .

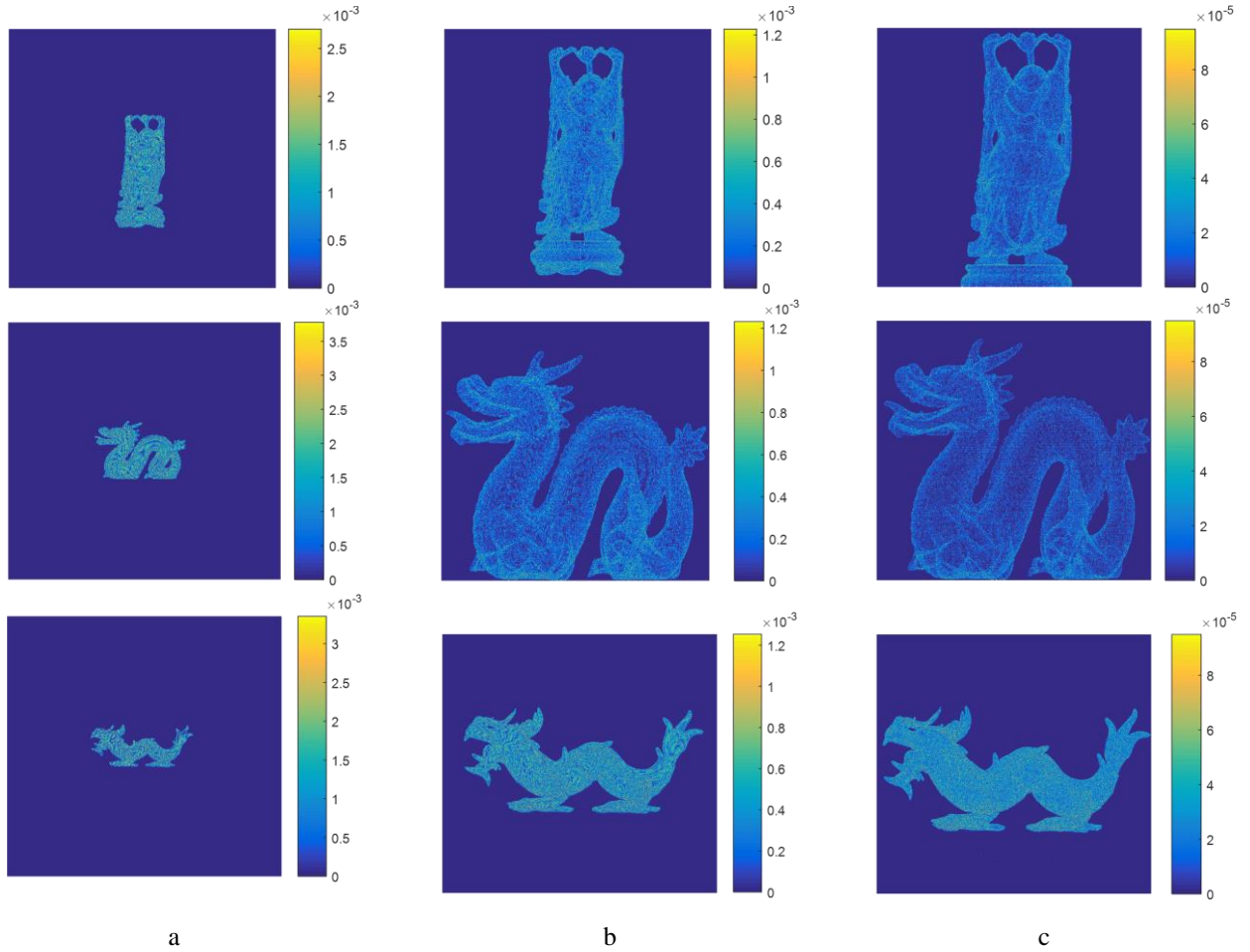


Figure 15. Effect of different CF parameters' optimization on the localization error for Buddha, Dragon, and Asian Dragon. a) Initial parallel cameras, b)  $\theta_1$ ,  $\theta_2$ , and  $f$  optimization, c) optimization of  $\theta_1$ ,  $\theta_2$ ,  $D$ , and  $f$ .

Table 6 shows the mean absolute localization error for the Bunny dataset at some combinations for  $\theta_1$  and  $\theta_2$  around their optimum value (from Table 5 when  $D = 2Z = 8\text{ m}$ ).

Table 5. Optimized  $\theta_1$  value per  $\theta_2$  values for  $D = 2Z$ .

$\theta_2$ (Degrees)	$\theta_1$ (Degrees)
42.0	47.7
43.0	46.9
44.0	46.0
45.0	45.0
46.0	44.0
47.0	42.8
48.0	41.6

Table 6. Mean absolute error ( $\times 10^{-4}$ ) versus variation of  $\theta_1$  for  $D = 2Z$  and different values of  $\theta_2$ .

$\theta_2$ (Degrees)	$\theta_1$ (Degrees)						
	41.6	42.8	44.0	45.0	46.0	46.9	47.7
42.0	1.7996	1.6750	1.5552	1.4551	1.3546	1.2731	<b>1.2058</b>
43.0	1.6392	1.5240	1.4108	1.3227	1.2340	<b>1.1652</b>	1.5422
44.0	1.5370	1.4255	1.3221	1.2374	<b>1.1634</b>	1.6738	2.1081
45.0	1.4411	1.3371	1.2394	<b>1.1629</b>	1.6976	2.1244	2.6292
46.0	1.3432	1.2462	<b>1.1649</b>	1.6938	2.3607	2.8886	3.1615
47.0	1.2528	<b>1.1668</b>	1.7288	2.3711	2.9520	3.4738	3.6885
48.0	<b>1.1647</b>	1.7058	2.4343	2.9849	3.5487	4.0668	4.7697

The bold value in Table 6 shows the minimum in each row, and it is clear that they occur at the

predicted  $\theta_1$  value. The global minimum is at  $\theta_1 = \theta_2 = 45$  degrees.

The second experiment is essentially as the first one, except that we have  $D > 2Z$ . Here, Table 8 shows the results for this case, and the three bold numbers enclosed by squares are the three answers of Equation (61), while Table 7 shows the corresponding optimum pairs of  $\theta_1$  and  $\theta_2$  for  $D > 2Z$ .

Table 7. Optimized  $\theta_1$  value per  $\theta_2$  values for  $D > 2z$ .

$\theta_2$ (Degrees)	$\theta_1$ (Degrees)
24.00	61
31.37	58.63
40.00	54.67
48.30	48.3
54.67	40
58.63	31.37
61.00	24

In the third experiment,  $\theta_1$  and  $\theta_2$  are chosen as the optimum value according to Equations (44) and (46), while a set of different values around the optimized points are assigned to  $D$  and  $X$ .

Table 9 shows the forecasted values of optimum  $X$  for some different  $D$  values. For each selected  $D$  value from Table 9, a set of  $X$  values around its

corresponding optimum value is selected.

**Table 8. Mean absolute error ( $\times 10^{-4}$ ) versus variation of  $\theta_1$  for  $D > 2Z$  and different values of  $\theta_2$ .**

$\theta_2$ (Degrees)	$\theta_1$ (Degrees)								
	24	31.37	40	48.3	54.67	58.63	61		
24.00	4.8788	4.8272	4.2207	3.2059	2.1660	1.5348	<b>1.2087</b>		
31.37	4.3134	4.2333	3.5902	2.5741	1.6720	<b>1.1775</b>	2.9528		
40.00	3.6626	3.4294	2.7298	1.8463	<b>1.1832</b>	3.8316	5.9459		
48.30	2.8517	2.4764	1.8426	<b>1.1947</b>	5.0453	8.5946	9.3228		
54.67	2.0921	1.6848	<b>1.1875</b>	6.5011	10.057	11.274	11.815		
58.63	1.5138	<b>1.1730</b>	5.1793	9.4405	12.820	13.603	13.643		
61.00	<b>1.2040</b>	4.2688	7.9951	10.959	1.3200	13.434	13.995		

The mean absolute localization error for all of these combinations are calculated and shown in Table 10. Again, the minima on each row are bolded. As Equation (61) tells for  $D \leq 2z = 8m$  we have one answer, and for  $D > 2Z$ , there are two answers.

**Table 9. Optimized  $X$  value per  $D$  values.**

$D(m)$	$X(m)$	
7.0	3.5	-
7.5	3.75	-
8.0	4.0	-
8.5	2.82	5.68
9.0	2.44	6.56

**Table 10. Mean absolute error ( $\times 10^{-4}$ ) versus  $X$  variation for different values of  $D$ .**

$D$	$X$								
	2.00	2.44	2.82	3.50	3.75	4.00	5.68	6.56	7.00
7.0	1.218	1.209	1.201	<b>1.183</b>	1.183	1.185	1.218	1.252	1.291
7.5	1.220	1.203	1.196	1.179	<b>1.177</b>	1.178	1.195	1.215	1.236
8.0	1.203	1.197	1.190	1.173	1.165	<b>1.163</b>	1.163	1.178	1.188
8.5	1.178	1.175	<b>1.167</b>	1.170	1.172	1.174	<b>1.166</b>	1.172	1.178
9.0	1.175	<b>1.167</b>	1.170	1.172	1.173	1.174	1.172	<b>1.168</b>	1.170

## 6. Conclusions

Spatial positioning of the objects is an important issue in the machine vision application. Different parameters can influence the positioning accuracy, one of which is the cameras' configuration and their viewing angle toward the scene. This paper presented a method for the geometric optimization of cameras in order to minimize the localization error. For this purpose, FOV of each pixel was assumed as a skew pyramid, and the uncertainty region was created from the intersection of these two pyramids. Then the mathematical equation of the uncertainty volume was calculated based on the correspondence field parameters as a criterion for the localization error. Then, given the mathematical relation of the localization error, the geometric optimization of camera configuration in the stereo-vision system was addressed. This meant that the cameras' position and their viewing angles and also camera parameters were optimized, which was not possible till now, and was determined according to the experience. Finally, the validity of the results obtained was

evaluated by simulation and the empirical results. The empirical results indicated that in the optimized state, a 40-fold reduction in the localization error was possible. It must be noted that the optimal configuration requires a larger focal length, implying higher costs and computational load.

## References

- [1] Wu, J., Sharma, R., and Huang, T. (1998). Analysis of uncertainty bounds due to quantization for three-dimensional position estimation using multiple cameras. *Optical Engineering journal*, vol. 37, no. 1, pp. 280–292.
- [2] Zhou, Y., Yan, F., and Zhou, Z. (2019). Handling pure camera rotation in semi-dense monocular SLAM. *Visual Computer*, vol. 35, no. 1, pp. 123–132.
- [3] Brown, M., Z., Burschka, D., and Hager, G. (2003). Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008.
- [4] Ardakani, H., K., Mousavinia, A., and Safaei, F. (2020). Four points: one-pass geometrical camera calibration algorithm. *Visual Computer*, Vol. 36, pp. 413–424.
- [5] Aliakbarpour, H. and Dias, J. (2012). Three-dimensional reconstruction based on multiple virtual planes by using fusion-based camera network. *IET Computer Vision*, vol. 6, no. 4, pp. 355 – 369.
- [6] Mandun, Z., Lichao, Q., Guodong, C., and Ming, Y. (2009). A triangulation method in 3D reconstruction from image sequences. *Second International Conference on Intelligent Networks and Intelligent Systems*, Tianjin, China.
- [7] Kanatani, K., Sugaya, Y., and Niitsuma, H. (2008). Triangulation from two views revisited: Hartley-Sturm vs. Optimal correction. *Proceedings of the British Machine Vision Conference*, Leeds, United Kingdom.
- [8] Weilharter, R. and Fraundorfer, F. (2021). HighRes-MVSNet: A Fast Multi-View Stereo Network for Dense 3D Reconstruction from High-Resolution Images. *IEEE Access*, vol. 9, pp. 11306–11315.
- [9] Hartley, R. & Sturm, P. (1997). Triangulation. *Computer Vision Image Understanding*, Vol. 68, No. 2, pp. 146–157.
- [10] Zhang, C. (2019). CuFusion2: Accurate and Denoised Volumetric 3D Object Reconstruction Using Depth Cameras. *IEEE Access*, vol. 7, pp. 49882–49893.
- [11] Liu, Z. -N., Cao, Y., Kuang, Z., Kobbelt, L., and Hu, S. (2021). High-Quality Textured 3D Shape Reconstruction with Cascaded Fully Convolutional Networks. *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 1, pp. 83–97.
- [12] Fooladgar, F., Samavi, S., Soroushmehr, S., and Shirani, S. (2013). Geometrical Analysis of

Localization Error in Stereo Vision Systems. *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4236–4246.

[13] Nakabo, Y., Mukai, T., Hattori, Y., et al. (2005). Variable baseline stereo tracking vision system using high-speed linear slider. *IEEE International Conference on Robotics and Automation*, Barcelona, Spain.

[14] Gallup, D., Frahm, J., Mordohai, P., and Pollefeys, M. (2008). Variable baseline/resolution stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA.

[15] Maliky, R. and Bajcsy, P. (2008). Automated placement of multiple stereo cameras. The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Marseille, France.

[16] Zhang, T. and Boulton, T. (2011). Realistic stereo error models and finite optimal stereo baselines. *IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, HI, USA.

[17] Wenhardt, S., Denzler, J., and Niemann, H. (2007). On minimizing errors in 3D reconstruction for stereo camera systems. *Pattern Recognition and Image Analysis*, vol. 17, no. 2, pp. 337–348.

[18] Safaei, F., Mokhtarian, P., Shidanshidi, H. et al. (2013). Scene-adaptive Configuration of Multiple Cameras using the Correspondence Field Function. *IEEE International Conference on Multimedia and Expo (ICME)*, San Jose, CA, USA.

[19] Karami, M., Mousavinia, A., and Ehsanian, M. (2017). A General Solution for Iso-Disparity Layers and Correspondence Field Model for Stereo Systems. *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3744–3753.

[20] Fu, S., Safaei, F., Li, W. (2017). Optimization of Camera Arrangement Using Correspondence Field to Improve Depth Estimation. *IEEE Trans. on Image Processing*, vol. 26, no. 6, pp. 3038–3050.

[21] Karami, M., Mousavinia, A., and Ehsanian, M. (2020). Camera Arrangement in Visual 3D Systems using Iso-disparity Model to Enhance Depth Estimation Accuracy. *Journal of AI and Data Mining*, vol. 8, no. 1, pp. 1-12.

[22] Jarrousse, O. (2014). Modified Mass-spring System for Physically-based Deformation Modeling. KIT Scientific Publishing.

[23] The Stanford 3D repository website (2014), Available: <http://graphics.stanford.edu/data/3Dscanrep/>.

## 7. Appendix

### 7.1. Details of Function F

Parameters  $m_1$  through  $m_{10}$  are defined as:

$$m_1 = X^2 + Z^2 \quad (65)$$

$$m_2 = D^2 Z(Z^2 - D(D - X)) \quad (66)$$

$$m_3 = -fD(3DX(D^2 + X^2) + m_1(4Z^2 - 6D^2)) \quad (67)$$

$$m_4 = -Zf^2(4m_1(m_1 - DX) + D^2(2D^2 + 7DX - 9X^2 - 14Z^2)) \quad (68)$$

$$m_5 = 3Df^3Z^2(4m_1 - 2D^2 - 3DX) \quad (69)$$

$$m_6 = 6D^2f^4Z^2 \quad (70)$$

$$m_7 = -DZ(m_1 - DX) \quad (71)$$

$$m_8 = f(2m_1(m_1 - 2DX) + D^2(2X^2 - Z^2)) \quad (72)$$

$$m_9 = 3Df^2Z(m_1 - DX) \quad (73)$$

$$m_{10} = D^2f^3Z^2 \quad (74)$$

Now,  $n_1$  and  $n_2$  can be calculated as:

$$n_1 = m_2\mu^4 + m_3\mu^3 + m_4\mu^2 + m_5\mu + m_6 \quad (75)$$

$$n_2 = m_7\mu^3 + m_8\mu^2 + m_9\mu + m_{10} \quad (76)$$

The desired  $F$  function can be summarized as:

$$F = \frac{n_2^2}{n_1} \quad (77)$$

## بهینه‌سازی هندسی چیدمان دوربین‌ها در سیستم دو دوربین جهت حداقل کردن خطای موقعیت‌یابی

حسین کمالی اردکانی<sup>۱\*</sup>، امیر موسوی‌نیا<sup>۱</sup> و فرزاد صفایی<sup>۲</sup><sup>۱</sup>دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.<sup>۲</sup>دانشکده مهندسی و علوم اطلاعات، دانشگاه وولونگونگ، وولونگونگ، استرالیا.

ارسال ۲۰۲۰/۰۷/۰۳؛ بازنگری ۲۰۲۱/۰۴/۰۶؛ پذیرش ۲۰۲۱/۰۵/۱۴

## چکیده:

سیستم بینایی ماشین دو دوربین می‌تواند به عنوان یک تکنیک نمونه برداری از فضا، مورد استفاده قرار گیرد و پارامترها و چیدمان دوربین‌ها می‌توانند به طور مؤثر تعداد نمونه‌ها فضایی (SSD) را تغییر دهند. با استفاده از مفهوم Voxels، این مقاله روشی را برای بهینه‌سازی چیدمان هندسی دوربین‌ها برای به حداقل رساندن SSD ارائه می‌دهد که به معنای به حداقل رساندن حجم Voxel و کاهش عدم قطعیت در موقعیت‌یابی یک شی در فضای سه بعدی است. میدان دید هر پیکسل به صورت یک هرم مورب در نظر گرفته می‌شود و در نتیجه منطقه عدم قطعیت از محل تلاقی دو هرم متناظر از دو دوربین به دست می‌آید. سپس معادله ریاضی منطقه عدم قطعیت بر اساس پارامترهای میدان تناظری به عنوان معیاری برای خطای موقعیت‌یابی از جمله خطای عمق و همچنین خطای محورهای X و Y محاسبه می‌شود. این میدان کاملاً به پارامترهای داخلی و خارجی دوربین بستگی دارد. با توجه به معادله ریاضی خطای موقعیت‌یابی به دست آمده، بهینه‌سازی چیدمان دوربین‌ها در یک سیستم دو دوربین مورد بررسی قرار می‌گیرد. سرانجام، اعتبار روش پیشنهادی با شبیه‌سازی و نتایج تجربی بررسی می‌شود. این نتایج نشان می‌دهد که خطای موقعیت‌یابی به طور قابل توجهی در چیدمان بهینه دوربین‌ها کاهش می‌یابد.

کلمات کلیدی: بینایی ماشین، چیدمان دوربین‌ها، میدان تناظری، بهینه‌سازی هندسی.