Journal of AI and Data Mining
Vol 8, No 1, 2020, 1-12.
DOI: 10.22044/JADM.2019.7209.1855

Camera Arrangement in Visual 3D Systems using Iso-disparity Model to Enhance Depth Estimation Accuracy

M. Karami[1], A. Mousavinia[2*] and M. Ehsanian[1]

1. Faculty of Electrical Engineering, K. N. Toosi University of Technology, Shariati Ave., Tehran, Iran.
2. Faculty of Computer Engineering, K. N. Toosi University of Technology, Shariati Ave., Tehran, Iran.

Received 26 June 2018; Revised 28 December 2018; Accepted 08 February 2019
**Corresponding author: moosavie@kntu.ac.ir (A. Mousavinia).*

Abstract
In this paper we address the problem of automatic arrangement of cameras in a 3D system in order to enhance the performance of depth acquisition procedure. Lacking ground truth or a priori information, a measure of uncertainty is required to assess the quality of reconstruction. The mathematical model of iso-disparity surfaces provides an efficient way to estimate the depth estimation uncertainty which is believed to be related to the baseline length, focal length, panning angle, and pixel resolution in a stereo-vision system. Accordingly, we first present analytical relations for a fast estimation of the embedded uncertainty in depth acquisition and then these relations, along with the 3D sampling arrangement are employed to define a cost function. The optimal camera arrangement will be determined by minimizing the cost function with respect to the system parameters and the required constraints. Finally, the proposed algorithm is implemented on some 3D models. The simulation results demonstrate a significant improvement (up to 35%) in depth uncertainty in the achieved depth maps compared with the traditional rectified camera setup.

Keywords: *Computer Vision, Correspondence Field, Camera Arrangement, Depth Estimation, Iso-disparity, Uncertainty.*

1. Introduction

Nowadays, a 3D reconstruction system is an inseparable part of most ongoing computer vision, modeling and robotic systems.

A 3D system design typically involves a range of performance trade-offs. Depending on the application, different system parameters and camera setups may be required. In such systems the depth calculation is always associated with a certain level of uncertainty. This uncertainty mainly arises from two sources, correspondence matching error and camera arrangement [1]. Most research works have focused on the former case, where recovering the best possible reconstruction from a given input data is of primary concern while the latter case has not been considered as it deserves, despite its importance. In many cases, there is the possibility of steering the acquisition process. In vision metrology, for example, where often images of objects are captured by a camera on a robotic arm, the goal is to measure the 3D coordinates of a scene or object as accurately as possible using visual information. Given constraints on physical placement of the cameras, selecting a good set of parameters to achieve well-conditioned triangulation, is an optimization problem, regarding camera arrangement.

Changing the camera arrangement, can dramatically change the mapping between the disparity space and the 3D space. As a consequence, the spread uncertainty will not remain constant across the scene. Thus arranging cameras in an appropriate manner in a 3D system can strongly affect the system performance.

In practice, the automatic optimization of camera configuration is inherently a quite difficult problem because of the very multi-modal and non-convex behavior of the cost function. Therefore, the process of setting up cameras is usually preferred to be done manually.

In this paper, we will address the automatic camera arrangement problem to increase the 3D reconstruction quality using the iso-disparity

relations [2]. The mathematical model of the iso-disparity surfaces can be used to analyze the space sampling behavior of pairs of cameras in 3D space in order to reliably control the system parameters with respect to the required constraints and target properties. We exploit this capability to derive an analytical relation to calculate depth uncertainty as well as variation in disparity value anywhere in field of view (FoV) for a general camera setup. A combination of uncertainty and disparity variation will be used to define a cost function. By minimizing the cost function with respect to the system parameters, the appropriate camera arrangement will be extracted.In this work, we only consider the geometric aspects of the problem, and do not account for the availability of texture or object occlusion, which are, of course, issues in a real system.In the sequel, Section 2 presents some related works that address the camera arrangement problem. A brief introduction to iso-disparity surfaces and other related concepts are presented in Section 3. Proposed algorithm for camera arrangement will be described in Section 4. The algorithm performance is evaluated by applying it to some 3D test models in Section 5. Finally, Section 6 is conclusion.

2. Related works

In many applications such as SLAM (simultaneous localization and mapping), robotics, and vision metrology, the success of algorithms depends strongly upon the input images, which makes viewpoint planning important. Accordingly, selecting the appropriate camera parameters to achieve a desired accuracy has been the subject of a group of research works in the field of computer vision field.

The early works on camera arrangement mostly tried to change the baseline length to manage uncertainty values [3, 4]. This idea arises from the well-known fact that the depth estimation accuracy is inversely proportional to the baseline length. However, in practice this procedure is restricted to implementation and performance. Increasing the baseline introduces two detrimental effects. First, it increases the stereo-minimum range, limiting the closest objects that can be seen simultaneously by two cameras. Secondly, it makes the matching process harder due to introducing the occlusion.

To overcome these detrimental effects, it is suggested to consider the other camera parameters such as focal length and orientation. Changing these parameters alters the topology of captured 3D sampling space. Modeling and estimating the behavior of the sampling space with respect to camera setup variations has been addressed in some of the recent research works. In [5], a scene has been captured by a number of cameras, and the goal is to create a uniform space sampling with uniformly distribution of the corresponding point along the scene.

To tackle the camera arrangement problem, Safaei et al. [5], have introduced the concept of correspondence field (CF) as a set of points associated with intersection of rays on an epipolar plane, which shows the location of all 3D points on the epipolar plane that can be sampled by the 3D system. They used CF to set the system parameters to achieve the desired sample density in a scene. More sample density at a region of space means less uncertainty in 3D reconstruction.

Before [5], the spatial topology of samples in the FoV of cameras had been studied by Pollefeys and Sinha [10]. They had demonstrated that the samples in the CF with the same disparity were observed as a family of iso-disparity conics on an epipolar plane. They discussed the effect of these conics on the performance of a general stereo-system to provide the most optimal configuration for an active stereo-head with the capability of panning and zooming. According to [10], as the iso-disparity contours more closely follow the expected object surface, the object surface will fit within a smaller disparity range and less disparities have to be searched. Consequently, the quality and efficiency of the matching results will be improved.

The arrangement of samples with the same disparity is called 2-surfaces as in [1], which indeed is the other name for iso-disparity surfaces. The authors in [1] have tried to maximize the sum of 2-surface density and alignment between CF and the directions of object surfaces in the scene. To implement this, the 2-surface gradient field is derived. An objective function is defined accordingly to optimize the arrangement for depth estimation.

Both [10] and [1] verify that camera setup in a 3D system has a significant impact on the quality of 3D reconstruction. They investigated this case using similar concepts as iso-disparity surfaces or 2-surfaces that is the representative of 3D sample arrangement.

In both [5] and [10], a 2D profile of iso-disparity layers on an epipolar plane has been considered. In [2], the exact general relations of 3D form of iso-disparity layers with respect to camera parameters have been extracted in a compact matrix form. It has been demonstrated that the iso-disparity layers appear in the form of a set of cylindrical quadric surfaces. The distance between the iso-disparity layers specifies the maximum achievable depth resolution in a 3D system. Increasing the distance

between these layers leads to an increase in the uncertainty of depth estimation, and consequently, a reduction in the localization accuracy. According to this, in [2], these distances have been calculated using iso-disparity relations to estimate the amount of embedded uncertainty of estimated depth at each 3D point.

Authors in [6] have constructed an active stereo-vision system comprising two identical sensor cameras mounted on a specific frame allowing different pan configurations and baseline variations. They proposed to actively vary the parameters of a stereo-vision system, i.e. distance between the two cameras and vision angles to improve the output quality. However, they did not introduce any effective measure on setting up the cameras to achieve this goal.

In [7], an optimization of a camera placement is presented to improve the localization accuracy. This goal is achieved by calculation of the localization at one specific point. With the assumption of Gaussian distribution for pixel quantization error, the authors have tried to estimate the localization error in the 3D space. The best position means where it maximizes the expected value of the accuracy. In this case the orientation is defined such that the observed point is mapped onto the center of the image.

Similarly, [8] works on placement of stereo-cameras in space to reduce the reconstruction error and improve the spatial resolution of the final 3D output. Authors have proposed an optimization framework using an error-based objective function under some given constraints. They generated an initial solution to the optimization problem using Genetic Algorithms, and then refined it using the Gradient Descent algorithm. The error-based objective function is defined based on minimizing the stereo-localization error, obtained from the stereo-localization geometry and maximizing the pixel resolution for each one of the stereo-cameras.

In [9], another approach has been reported, and the visibility is in the center of attention. The relation between the reconstruction quality and the camera arrangement has been analyzed, and a new camera positioning strategy based on the properties of the synthetic object surface has been proposed. They segmented the object into different parts and modeled them as simple shapes and considered their visibilities separately. They divided the object surface into two groups: 1- regions whose curvatures are modeled as sinusoidal functions, and 2- regions that can be modeled as flat planes. A binary tree decision algorithm with the candidate viewing direction is used to select the best candidates to achieve the best visibility.

Since the iso-disparity model suitably unlocks the potential of sample alteration in the entire or some selected sub-region in a 3D scene, this paper exploits it in order to setup an algorithm for selecting an appropriate camera arrangement to reduce disparity map errors as well as propose a quite fast and accurate uncertainty estimation algorithm in a general camera configuration.

3. Iso-disparity layers

The ideas introduced in [5] and [10] as CF and iso-disparity to model sampling space of 3D systems are very similar. This idea has been matured in [2] by extracting exact relations of iso-disparity layers with respect to the system parameters.

In [2], it has been demonstrated that, in their general 3D form, the iso-disparity layers are a set of quadric surfaces. Depending on the camera configuration, these surfaces will be elliptic, hyperbolic or parabolic cylinders while the axis of these cylinders are always parallel to the y axis. The rectified setup is a degenerated case where the quadric surfaces turn to fronto-parallel planes. In [2], for the first time, the exact relation of these surfaces with respect to the intrinsic and extrinsic camera parameters are extracted as the following equations:

$$\boldsymbol{Q}=\begin{bmatrix} a_3 & 0 & a_4 & a_1 \\ 0 & 0 & 0 & 0 \\ a_4 & 0 & a_5 & a_2 \\ a_1 & 0 & a_2 & a_0 \end{bmatrix} \tag{1}$$

where the coefficients a_0 to a_5 are defined as what follow.

$$a_0 = c_1c_2\,(tan(\alpha_1)-tan(\alpha_2) \\ -\rho\, tan(\alpha_1)tan(\alpha_2)) \tag{2}$$

$$a_1 = (c_1+c_2)(tan(\alpha_2)-tan(\alpha_1) \\ +\rho\, tan(\alpha_1)tan(\alpha_2)) \tag{3}$$

$$a_2 = (c_1-c_2)(1+tan(\alpha_1)tan(\alpha_2)) \\ +\rho\,(c_1\, tan(\alpha_1)+c_2\, tan(\alpha_2)) \tag{4}$$

$$a_3 = tan(\alpha_1)-tan(\alpha_2)-\rho\, tan(\alpha_1)tan(\alpha_2) \tag{5}$$

$$a_4 = -\rho\,(tan(\alpha_1)+tan(\alpha_2)) \tag{6}$$

$$a_5 = tan(\alpha_1)-tan(\alpha_2)-\rho \tag{7}$$

In these relations, $\rho=\frac{rd}{f}$, where r is the pixel width; f is focal length; α_1 and α_2 are the panning angles of the first and second cameras respectively; the camera centers are located, respectively, at $\boldsymbol{C}_1=(c_1,0,0)^T$ and $\boldsymbol{C}_2=(c_2,0,0)^T$;

and d is the disparity in non-rectified camera setup, which is derived during extracting iso-disparity relations as:

$$d = \frac{f}{r}\left(\frac{x - c_2 - z\,tan(\alpha_2)}{z + (x - c_2)tan(\alpha_2)} - \frac{x - c_1 - z\,tan(\alpha_1)}{z + (x - c_1)tan(\alpha_1)}\right) \quad (8)$$

The detailed prove of these relations has been given in [2].

Figure 1 depicts a typical converged camera setup with the associated iso-disparity surfaces and their corresponding iso-disparity conics on the xz-epipolar plane.

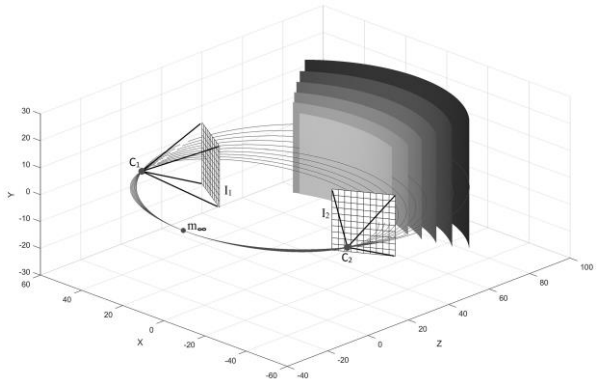


Figure 1. Iso-disparity surfaces for a typical converged camera setup [2].

Because of the independency of iso-disparity surfaces with respect to the y coordinate, it is possible to shrink the computational dimension from $3D$ to $2D$.

In this context, it is appropriate to take a 2D profile of iso-disparity layers into account as an intersection of iso-disparity surfaces with mid-epipolar plane including the principal axes of cameras. Therefore, the iso-disparity layers turn to a set of conics on the xz-epipolar plane with the following general equation:

$$f(z,x) = a_5 z^2 + a_4 zx + a_3 x^2 + a_2 z + a_1 x + a_0 = 0 \quad (9)$$

The coefficients a_0 to a_5 are as defined in relations 2 to 7.

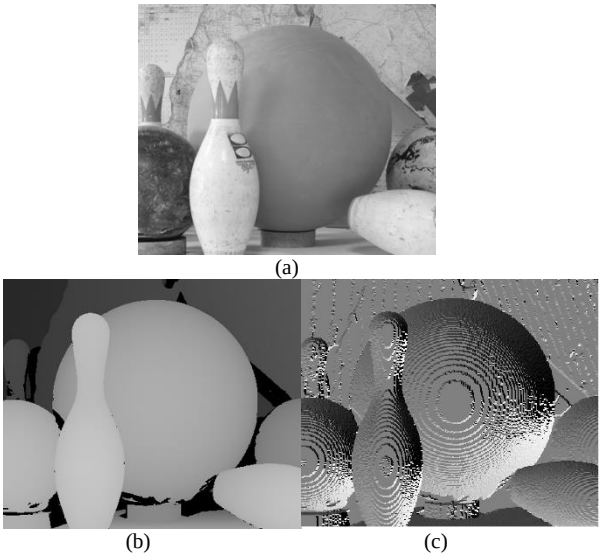
4. Camera arrangement

As the scene visibility can be changed dramatically with viewpoint, most modern multi-view algorithms look for an appropriate method to specify the best views to achieve the best 3D reconstruction accuracy and completeness. In most cases, the camera arrangement is done manually based on the operator's experiences. To automate this task one can define an optimization problem under given constraints. Unfortunately, different camera configurations may result in similar reconstruction errors. This arises from the very multi-modal and non-convex nature of cost function defined in this problem. Consequently, the selection of an appropriate criterion to determine the parameter values, is the vital key in any automatic camera arrangement algorithm.

4.1. 3D sampling

Varying the cameras' parameters and their relative pose with respect to each other will change the sampling behavior of 3D system, and consequently, the form of iso-disparity layers. In [1] and [10] it has been proved that the more closely the iso-disparity layers follow the object surface, there is less variation in disparity, and consequently, a significant reduction in depth-levelling errors. In contrast, when the iso-disparity layers are perpendicular to the normal of the object's surface, there will be a significant depth variations and high uncertainty values.

These effects are depicted in Figure 2.



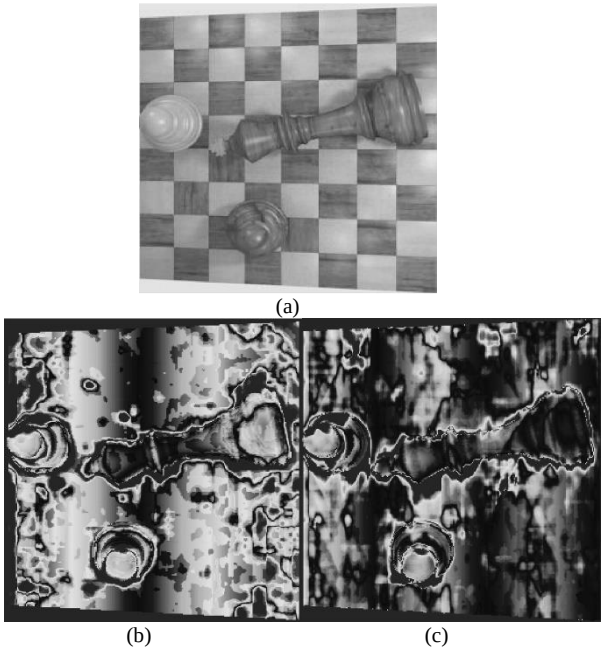
(a) (b) (c)

Figure 2. (a) Captured scene by a rectified camera setup [11] (b) Associated disparity map (c) Intersection of iso-disparity layers with object surface.

Figure 2(a) shows the scene that is captured by a rectified camera setup and its associated disparity map in Figure 2(b) [11]. In a rectified setup, the iso-disparity layers are fronto-parallel planes. The intersection of these planes with object surfaces is a set of curves that are depicted in Figure 2(c). It can be observed that the larger the difference between the normal vectors, the less is the distance between the curves. In such regions, a small error in point-matching easily causes to shift from one depth layer to the next layer, and hence, increasing error in calculating disparity map. Where the normal vector of object surface is not compliant with the iso-disparity layer normal vector, there is an occlusion between the object surface depth levels as well. It means the less part of object

surface is visible which results incompleteness of the reconstructed model.

Figure 3 demonstrates the effects of changing the CF parameters on the disparity map. Figure 3(a) shows a chess scene, while Figures 3(b) and 3(c) show the associated disparity maps estimated in a parallel stereo-arrangement and non-rectified camera arrangement respectively.



(a)

(b) (c)

Figure 3. (a)A chess scene (b) Associated disparity map for rectified setup (c) Calculated disparity map for non-rectified setup [1].

Figure 3 emphasizes again the direct effect of selecting the proper camera arrangement on the quality of the reconstruction result.

Ideally, the best arrangement occurs when an iso-disparity layer follows completely the object surface. Of course in practice, it is unlikely to happen that all object surface points lie on a single iso-disparity layer. However, it is possible to look for an arrangement in which the surface points are confined within a minimum number of disparity layers. Based on these facts, a promising approach for camera arrangement is to sample 3D points of object in the scene in such a way that disparity variations become as small as possible and 3D points fall in the smallest disparity layer set. Thereupon, it is likely to minimize the error of depth estimation and the area of occlusion.

To achieve this goal, Equation (8) establishes a relation between 3D points on the object surface and their disparity value.

Let d be a set of all disparity values associated with the scene 3D points. The objective is to select the parameter vector $\boldsymbol{p}$ to minimize the variance of d.

Accordingly, a cost function is defined as follows:

$$\min_{\boldsymbol{p}} \mathrm{V}(\boldsymbol{p}) = \frac{1}{N}\sum_{i=1}^{N}\left(d_i(z_i, x_i, \boldsymbol{p}) - \mu\right)^2 \qquad (10)$$

$$subject\,to \quad LB \le \boldsymbol{p} \le UB$$

Where $d_i(z_i, x_i, \boldsymbol{p})$ is the disparity value associated with the ith 3D point that is calculated as:

$$d_i(z_i, x_i, \boldsymbol{p}) = \frac{f}{r}\left(\frac{x_i - c_2 - z_i\, tan(\alpha_2)}{z_i + (x_i - c_2)tan(\alpha_2)} - \frac{x_i - c_1 - z_i\, tan(\alpha_1)}{z_i + (x_i - c_1)tan(\alpha_1)}\right) \qquad (11)$$

And

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i(z_i, x_i, \boldsymbol{p}) \qquad (12)$$

while LB and UB are given the lower and upper bounds of parameter vector $\boldsymbol{p} = (f, c_1, c_2, \alpha_1, \alpha_2, r)^T$, and N is the total number of triangulated points in 3D space.

4.2. Estimation of uncertainty

The accuracy of reconstruction is a major concern in all 3D systems. As mentioned in Section 3, theoretically, given two images, captured from different vantage points, and information on intrinsic and extrinsic parameters of cameras, it is possible to determine the exact location of each point in 3D space. However, in practice, the depth spatial quantization uncertainty, caused by a discrete sensor, results in uncertainty in the localization process that grows quadratically proportional to the distance from the camera baseline. This puts restriction on accuracy of the system at each point of the scene.

When the ground truth is not available, the uncertainty of estimation is the only gauge for quality assessment of a reconstruction.

It has been mentioned in [1] that the embedded uncertainty in localization of points in 3D arises from two major sources: correspondence matching error and camera arrangement. Usually, the correspondence matching algorithm is fixed but camera position and orientation is adjustable with some degree of freedom.

Traditionally, the accuracy of 3D systems is estimated by measuring the error of depth, $\dot{o}_z$, as:

$$\dot{o}_z = \frac{\partial \mathrm{T}}{\partial d}\dot{o}_d \qquad (13)$$

where T is the triangulation function and $\dot{o}_d$ is the error of disparity [1]. Since d is estimated by a

correspondence matching algorithm, the error $\grave{o}_d$ stems from the matching process. On the other hand $\frac{\partial \mathrm{T}}{\partial d}$ is related to the camera configuration on the scene and its parameters.

In non-rectified setups, because of the lack of consensus about definition of disparity, there is no straightforward relation for depth uncertainty. In [12] and [13], the authors have suggested the intersection volume of cones or pyramids corresponding to each pixel from two images as a measure of uncertainty. In [2] and [1], the distance between the iso-disparity layers has been proposed as a measure of this uncertainty in depth estimation. In all the above-mentioned works, geometrical methods are exploited to estimate the amount of uncertainty.

However, using relation (8) in [2] it is possible to extract a close form, analytical relation for fast and straightforward estimation of depth uncertainty.

As demonstrated in [2], in a general case, each point $\boldsymbol{M}=(x,y,z)^T$ in 3D space, is assigned to a discrete iso-disparity layer in CF whose disparity value can be calculated as:

$$\mathrm{T}^{-1}=d(z,x,\boldsymbol{p})=\frac{f}{r}\left(\frac{x-c_2-z\tan(\alpha_2)}{z-(x-c_2)\tan(\alpha_2)}-\frac{x-c_1-z\tan(\alpha_1)}{z-(x-c_1)\tan(\alpha_1)}\right) \quad (14)$$

Iso-disparity layers are a set of quadratic equations and as a function are not invertible in their whole domain. However, in a stereo-system, it is possible to define a partial inverse on the FoV as the area of interest.

One way to calculate the uncertainty according to relation (13) is to calculate the inverse of d and then differentiating it, which results in a rather complex expression. However, a more convenient way is to follow the rule of differentiating inverse of a function.

Let $f(x)$ be an invertible function, with $f^{-1}(x)$ as its inverse. If $f(x)$ is differentiable on an interval I, then $f^{-1}(x)$ is also differentiable if $f'(x)\neq 0$ for any $x\in I$ as:

$$\left(f^{-1}\right)'(x)=\frac{1}{f'\left(f^{-1}(x)\right)} \quad (15)$$

This relation also holds for multivariate functions where the Jacobian of the inverse function is simply the reciprocal of the Jacobian of the function [14]. Accordingly, it is easy to show that the magnitude of localization uncertainty at point $\boldsymbol{M}=(x,y,z)^T$ in a general camera setup can be calculated as:

$$\mathrm{U}(x,y,z)=\left(\frac{\partial d(z,x)}{\partial z},\frac{\partial d(z,x)}{\partial x}\right)^{-1} \quad (16)$$

where

$$\frac{\partial d(z,x)}{\partial z}=\frac{f}{r}\left(\frac{(x-c_2)\left(1+\tan^2(\alpha_2)\right)}{\left(z+(x-c_2)\tan(\alpha_2)\right)^2}-\frac{(x-c_1)\left(1+\tan^2(\alpha_1)\right)}{\left(z+(x-c_1)\tan(\alpha_1)\right)^2}\right) \quad (17)$$

and

$$\frac{\partial d(z,x)}{\partial x}=\frac{f}{r}\left(\frac{z\left(1+\tan^2(\alpha_2)\right)}{\left(z+(x-c_2)\tan(\alpha_2)\right)^2}-\frac{z\left(1+\tan^2(\alpha_1)\right)}{\left(z+(x-c_1)\tan(\alpha_1)\right)^2}\right) \quad (18)$$

It is clear that $\frac{\partial d(z,x)}{\partial x}=0$. Finally the uncertainty at point $\boldsymbol{M}$ can be calculated as:

$$\mathrm{U}(x,y,z)=\frac{1}{\sqrt{\left(\frac{\partial d(z,x)}{\partial z}\right)^2+\left(\frac{\partial d(z,x)}{\partial x}\right)^2}} \quad (19)$$

This determines the maximum achievable accuracy of depth estimation for a given camera arrangement. Figure 4 shows the uncertainty value for three typical configurations in FoV. The interesting point is that in non-rectified setup, the expected uncertainty at the sides of FoV is significantly less than that of the middle of FoV.

In a similar way it is possible to calculate the sensitivity of a 3D system with respect to different parameters using the iso-disparity relations.

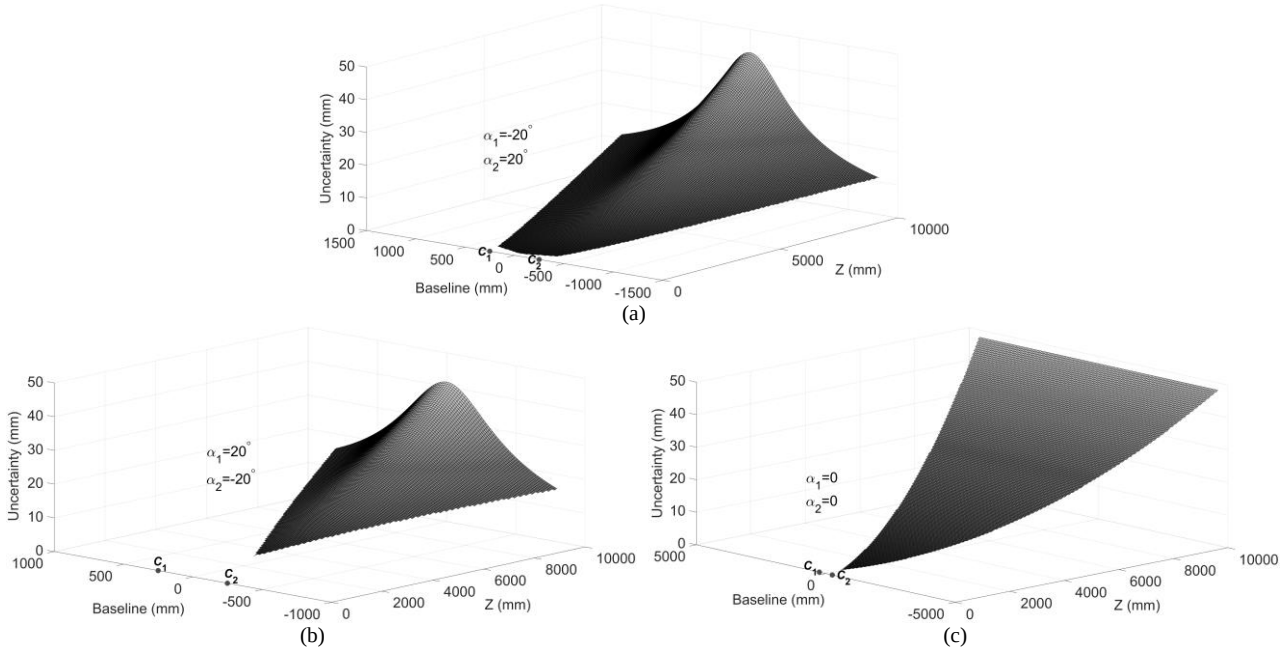


Figure 4. The uncertainty diagram for a typical (a) converged camera setup with $\alpha_1 = -20$ and $\alpha_2 = 20$ (b) diverged camera setup with $\alpha_1 = 20$ and $\alpha_2 = -20$ (c) rectified camera setup.

4.3. Optimization algorithm

Here we consider the problem of finding the position and orientation of cameras between two predetermined locations such that the reconstruction accuracy of the observed geometry is maximized while the best alignment occurs between the iso-disparity layers and the object surface in the scene. To do this we further assume the followings:

- The algorithm is stated with initially rectified cameras.
- An initial estimate of the camera parameters is available.
- We only consider the geometric aspects of the problem and do not account for availability of texture or other issues related to the matching algorithm.

This is a nonlinear, multi-objective, multivariate, constrained optimization problem. Sequential Quadratic Programming (SQP) is a successful iterative method for the constrained nonlinear optimization problems of the form:

$$\begin{aligned} &\min && f(x) \\ &\textit{over} && x \in \mathrm{R}^n \\ &\textit{subject to} && h(x) = 0, g(x) \le 0. \end{aligned} \tag{20}$$

where $f: \mathrm{R}^n \to \mathrm{R}$ is the objective function, the functions $h: \mathrm{R}^n \to \mathrm{R}^m$ and $g: \mathrm{R}^n \to \mathrm{R}^m$ describe the equality and inequality constraints.

The set of points that satisfy the equality and inequality constraints are referred to as feasible points.

The underlying ideas for the SQP method is to model the problem at a given approximate solution, x^k, by a quadratic programming sub-problem, and then to use the solution to this sub-problem to construct a better approximation, x^{k+1}

Nonlinear optimization problems can have multiple local solutions; the global solution is the local solution corresponding to the least value of f. Similar to all iterative methods, SQP is sensitive to the starting point, and only guarantees to find a local solution so we will repeat the optimization algorithm multiple times with different random start points, and the best solution will be selected.

4.4. Cost function

Minimizing the uncertainty and the disparity variations simultaneously, in most cases, are conflicting objectives that must be combined in a single cost function to make an efficient and fast optimization algorithm possible; otherwise, a multi-objective optimization algorithm must be utilized to solve the problem, which is usually slow.

Because these values are from two different kind of quantities, it is required to normalize these values to achieve a meaningful combination. For this purpose, their value at the rectified setup is selected as the normalization factor. In this way, the cost function is defined as:

$$C(\boldsymbol{p}) = \frac{1}{N}\left(\frac{\sum_{i=1}^{N} \mathrm{U}(\boldsymbol{p}, \boldsymbol{M}_i)}{\mathrm{U}_{rect}} + \zeta \frac{\sum_{i=1}^{N} \mathrm{V}(\boldsymbol{p}, \boldsymbol{M}_i)}{\mathrm{V}_{rect}} \right) \tag{21}$$

where $\zeta > 0$ is a constant factor to highlight the roll of each factor at the envisaged application. In this relation, U_{rect} and V_{rect} are the uncertainty and disparity variation in the initial rectified setup respectively, and calculated as:

$$U_{rect} = \frac{1}{N}\sum_{i=1}^{N} U\left(\boldsymbol{p}_{rect}, \boldsymbol{M}_i\right) \quad (22)$$

$$V_{rect} = \frac{1}{N}\sum_{i=1}^{N} V\left(\boldsymbol{p}_{rect}, \boldsymbol{M}_i\right) \quad (23)$$

Here N is the number of triangulated points taken into account.

4.5. Algorithm outline

All steps are outlined in algorithm 1.

Algorithm 1 Camera arrangement algorithm

Require: Images from rectified cameras.
Ensure: $LB \le \boldsymbol{p} \le UB$.

Search for reliable match points on image pair.
Triangulate match points.
Calculate U_{rect} and V_{rect}.
Set maximum number of iteration (It_{max}).
Set $t = 0$.
while $t \le It_{max}$
  Set the initial parameters as $\boldsymbol{p}^{(0)}$.
  Run SQP algorithm.
  if $C^{(t+1)}(\boldsymbol{p}) < C^{(t)}(\boldsymbol{p})$
    $\boldsymbol{p}_{best} = \boldsymbol{p}^{(t+1)}$.
  end if
  $t = t + 1$.
end while
return Camera arrangement parameters ($\boldsymbol{p}_{best}$).

5. Experimental results

In this section, the proposed algorithm is implemented in Matlab environment to testify the ability of the proposed algorithm in a two-camera 3D system. The algorithm has been applied to several 3D models and the results of optimized setup have been compared with the initial rectified camera setup.

The "Sift" function in the "Vlfeat" toolbox [15] is used in order to extract feature points of image pair in the triangulation step. In order to obtain the best possible answer in the optimization step, the SQP algorithm is applied multiple times with different random start points and specified constraints to acquire a near global extremum of the cost function.

To obtain the distance between the estimated point cloud and the reference point cloud (ground truth), first two point clouds were aligned using the iterative closest point (ICP) algorithm. The alignment parameters consist of a translation, rotation, and uniform scale. The quality of these alignments was inspected manually to insure a correct alignment between two point clouds, and if necessary an initial manual transformation was applied to acquire the best alignment result.

Then the Euclidean distance between each point on the estimated point cloud and the nearest point on the reference point cloud is calculated. Given these distances, it is possible to compute the summary statistics useful in comparing the accuracy of the reconstruction algorithms.

5.1. Dataset

Most of the available datasets are not suitable for evaluating our camera arrangement algorithms because these datasets usually provide a few fixed views, while we need a free-moving camera to capture arbitrary views based on the optimized parameters. Thus we selected some 3D models from the well-known Stanford 3D scanning repository [16] and a large geometric model archive of Georgia Institute of Technology [17].

Importing these models in the 3DSmax software makes us capable of rendering appropriate views with the desired specifications.

In the experiments, the first camera center is restricted to $20 \le x \le 500mm$ on the x axis and $-500 \le x \le -20mm$ for the second camera. The cameras are able to pan between ± 30 degrees with a focal length of $35mm$ and an image resolution of 4096×2731 pixels.

The algorithm is implemented on a laptop with Intel core i7 6700 CPU using the Matlab software.

5.2. Evaluation on 3D models

Table 1 shows the results of reconstruction after the optimal set up of cameras and compares it with the initial rectified setup. In this table the amount of uncertainty is considered as a gauge of reconstruction accuracy. This table shows a significant improvement in reconstruction after optimizing camera arrangement. Traditionally, in a rectified setup the uncertainty of depth estimation can be measured as $\partial_z \approx \frac{z^2}{fb}$ [18]. It is clear that the accuracy is inversely proportional to the baseline length. Of course, in non-rectified setup this relation is still held but in more complex form [1], meaning that in order to decrease the uncertainty, the baseline must be increased proportionally.

However, results of table 1 show that this can be achieved without significant increase or even with reduction in baseline in a non-rectified system. This is important especially when there is a limited space or motion restriction for the cameras.

Table 1. Comparison between rectified and optimized camera setups in changing of baseline and improvement in depth estimation uncertainty for five 3D models. The minus sign in the baseline change column means reduction in the baseline.

Model	Baseline			Optimized angle (Degree)		Uncertainty improvement (%)
	Rectified (mm)	Optimized (mm)	Change (%)	α_1	α_2	
Bunny	500	480.18	-13.96	13.18	-14.92	25.01
Blade	500	470.00	-6.00	-21.27	23.21	35.15
Buddha	500	476.06	-4.79	16.50	-19.54	25.47
Horse	500	481.81	-3.64	-17.53	11.37	31.48
Dragon	500	496.90	-0.62	15.72	14.77	25.63

In the baseline change column, the minus sign means a decrease in the baseline length, and in the uncertainty column the percentage of reduction of uncertainty is presented.

Figure 5 shows the optimized camera arrangement and the associated parameters for each model calculated by the proposed algorithm. In this figure, the calculated point cloud for each model, the coordinate of camera centers (C_1 and C_2) on the baseline length and the camera panning angles (α_1 and α_2) are displayed for the optimized setup. To clarify how the form of iso-disparity layers change regarding to the object surface, the shapes of iso-disparity layers after optimization are also depicted in this figure for each setup in the form of conic sections on the mid-epipolar plane. The accuracy improvement and baseline reduction with respect to the rectified setup are shown for each model in percentage as well.

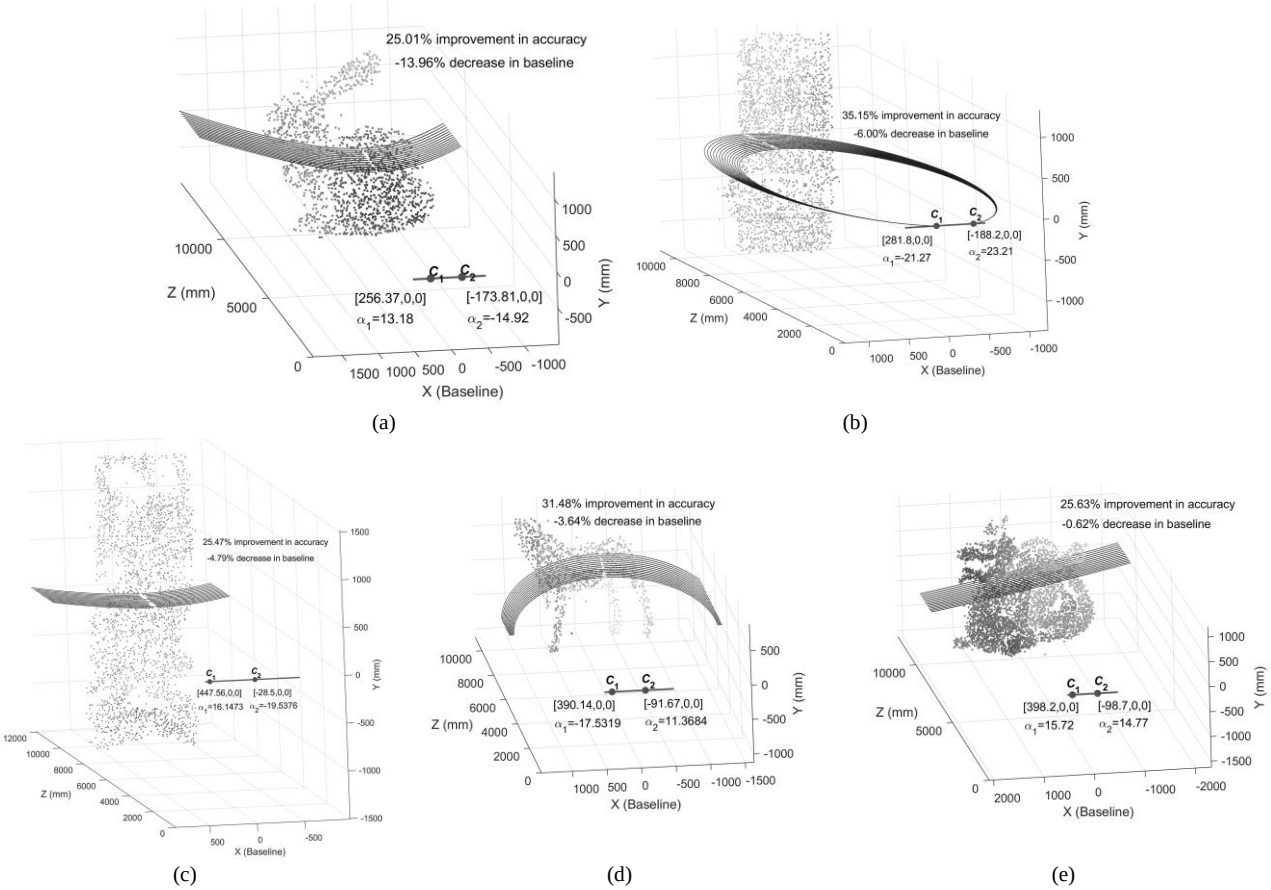


Figure 5. Optimized camera parameters and the associated iso-disparity layers for (a) the Stanford Bunny (b) the Turbine Blade (c) the happy Buddha, (d) The Horse and (e) the dragon.

Figure 6 shows a rough reconstruction of models without any pre- or post-processing smoothing techniques, just using triangulated points, before and after the optimization process. In this figure,

the first row shows the original models. The second and third rows are the reconstructed results for the rectified and optimized camera setups, respectively. The improvement in quality of surface reconstruction is clear but for a further clarification, a part of the reconstructed outputs is magnified in the fourth and fifth rows for rectified and optimized setups. Because of the decrease in depth estimation uncertainty, due to the optimum resampling of the data after the optimization process, a smoother surface with less geometric noise has been achieved.

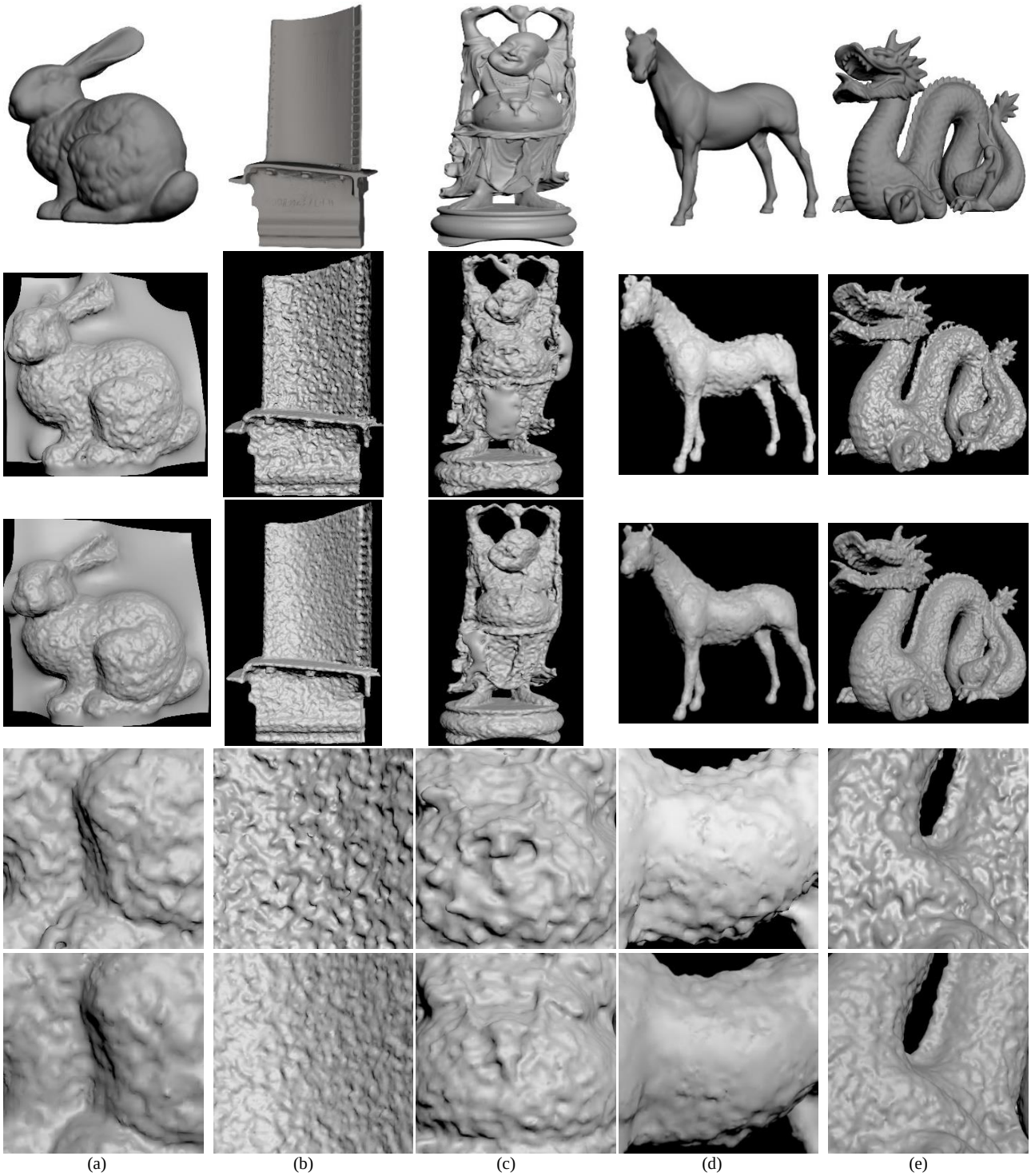


Figure 6. Columns from left to right (a) the Stanford Bunny [16] (b) the Turbine Blade [17] (c) the happy Buddha [16] (d) the Horse [17] and (e) The dragon [16]. The first row is the original models. The second row is the reconstruction for rectified setup. The third row shows reconstruction for optimized setup. The fourth and fifth rows are a magnified part of rows two and three, respectively, to clarify improvement in surface reconstruction for given models after camera arrangement optimization.

5.3 Time complexity analysis

In this paper we have used SIFT as a well-known feature-based algorithm to extract the correspondence points. The time spent on the matching process depends on image size. Here, in the experiments, the CPU time spent on the optimization algorithm was almost negligible compared to that on the stereo-matching. On a laptop with an Intel core i7 6700HQ CPU with a single core enabled, it takes about 12 s, on average, by stereo-matching whereas the optimization process is finished within about 40ms for 1000 points.

For the prohibitive time complexity of the matching algorithm, using the optimization algorithm is completely a reasonable choice in comparison with the methods in which multiple image pairs are processed. Moreover, such methods introduce additional problems to the system like inconsistency in depth values or occluding boundaries.

On the other hand, the computing cost will grow in matching algorithms correspondingly with the number and size of the images. For example as expressed in [4], increasing resolution by a factor of 2 would require $2^6 = 64$ times more computational efforts. This is while, it is possible to fix the execution time of the proposed optimization algorithm by utilizing a down-sampled version of point cloud to the extent that the original form of the model is specified to speed up the algorithm significantly and make it time-independent, which is another superiority of the proposed algorithm over using multiple image pairs to reduce the depth estimation uncertainty.

5.4 Comparison with [1]

The proposed algorithm is compared with one of the recent works on camera arrangements in which a similar idea is used to optimize cameras.

In [1] the iso-disparity surface concept (or 2-surfaces as it is called in [1]) is used to optimize the arrangement of cameras in a 3D system. The idea behind this paper is established based on the iso-disparity gradient field that includes both density and direction of iso-disparity surfaces. The objective function is defined as the sum of the magnitude of the inner product between this gradient and the object surface normal over a region of interest.

Since the scene surface is initially unknown, an expectation maximization (EM) approach is utilized to estimate the surface and adjusts the setup parameters to maximize the objective function.

1- In [1], implementation starts with a conventional parallel setting.
2- Estimation of object surface given current arrangement, consisting of four stages.
   a. The first stage is depth estimation.
   b. The second stage is noise reduction to remove outliers.
   c. In the third stage, the point cloud will be decomposed into a number of clusters using a clustering technique to identify point groups with a sufficient density and an appreciable separation.
   d. The estimation of object surface normal at each point is done by a PCA analysis on a local point patch.
3- A trust region reflective optimization method is utilized to solve constrained non-linear optimization problem.

By comparison, steps (1) and (2.a) are common between [1] and the proposed algorithm. In step (3), the proposed method uses a SQP algorithm instead of a trust region reflective optimization in order to optimize the objective function. Both of these algorithms belong to the same category of optimization methods with a similar computational complexity.

However, the differences appear in stages (2.b), (2.c), and (2.d). Because the estimation is done locally in [1], it is sensitive to outliers and a noise reduction is required in the (2.b) stage; but in the proposed algorithm the alignment is performed globally so if a sufficient number of samples is selected, the outlier's impact will not be considerable.

Stages (2.c) and (2.d) are related to the object surface normal estimation. These stages are not required in the proposed algorithm. Aside from the computational effort to cluster samples in stage (2.c), according to the complexity analysis in [1], the estimation of object surface normal by PCA algorithm is a function of the number of estimating points N and the number of neighbors B Its complexity is in the order of $\mathrm{O}\left(B^2 N\right)$, where N is equal to the pixel number. By comparison, in the proposed algorithm, these processes are not required so the camera arrangement optimization step will run much faster than the work in [1], while the final results are meaningfully still better.

6. Conclusion

This paper addresses the problem of optimizing the arrangement of two cameras to improve the accuracy of a 3D reconstruction. The iso-disparity concept was exploited to establish a relation between the objects in the scene and the camera arrangement parameters. Based on these relations an optimization algorithm was defined to minimize

the disparity variation and the uncertainty in depth estimation. During this process a novel uncertainty estimation was proposed for the first time in this paper. The proposed algorithm was implemented on 3D models to verify the efficacy of the algorithm. The implementation results demonstrate up to 35% improvement in depth estimation uncertainty within a reasonable time, while the execution time of optimization step of the proposed algorithm is almost fix for different image sizes.

References

[1] Fu, S., Safaei, F. & Li, W. (2017), Optimization of camera arrangement using correspondence field to improve depth estimation, IEEE Transactions on Image Processing, vol. 26, no. 6, pp. 3038-3050.

[2] Karami, M., Mousavinia, A. & 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.

[3] Nakabo, Y., Mukai, T., Hattori, Y., Takeuchi, Y. & Ohnishi, N. (2005), Variable baseline stereo tracking vision system using high-speed linear slider, in: Robotics and Automation, ICRA 2005. Proceedings of the IEEE International Conference on, 2005, pp. 1567-1572.

[4] Gallup, D., Frahm, J.-M., Mordohai, P. & Pollefeys, M. (2008), Variable baseline/resolution stereo, in: Computer Vision and Pattern Recognition, CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[5] Safaei, F., Mokhtarian, P., Shidanshidi, H., Li, W., Namazi-Rad, M. & Mousavinia, A. (2013), Scene-adaptive configuration of two cameras using the correspondence field function, IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 2013, pp. 1-6.

[6] E. Dumont, E., Constantin, C., Esse, A., Gréaux, A. & Techer, F., (2015), Optimized Stereoscopic 3-D Object Reconstruction, International Journal of Information and Electronics Engineering, vol. 5, no. 1, 2015, pp. 35-39.

[7] Szalóki, D., Koszó, N., Csorba, K. & Tevesz, G., (2013), Optimizing camera placement for localization accuracy, IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI), 2013, pp. 207-212.

[8] Malik, R. & Bajcsy, P., (2008), Automated Placement of Multiple Stereo Cameras, the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras - OMNIVIS, Marseille, France. 2008.

[9] Qian, N. & Lo, C. Y., (2015), Optimizing camera positions for multi-view 3D reconstruction, International Conference on 3D Imaging (IC3D), 2015, december, pp. 1-8.

[10] Pollefeys, M. & Sinha, S. (2004), Proceedings, Iso-disparity Surfaces for General Stereo Configurations, Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. pp. 509-520.

[11] Hirschmuller, H. & Scharstein, D. (2007), Evaluation of cost functions for stereo matching, IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-8.

[12] Fooladgar, F., Samavi, S., Soroushmehr, S. & Shirani, S. (2013), Geometrical analysis of localization error in stereo vision systems, Sensors Journal, IEEE vol. 13, no. 11, pp. 4236-4246.

[13] Wu, J. J., Sharma, R. & Huang, T. S. (1998), Analysis of uncertainty bounds due to quantization for three-dimensional position estimation using multiple cameras, Optical Engineering, Vol. 37, no. 1, pp. 280-292.

[14] Kaplan, W. (2002), Advanced Calculus, Addison-Wesley higher mathematics, Addison-Wesley, 2002.

[15] Vedaldi, A. & Fulkerson, B. (2008), VLFeat: An open and portable library of computer vision algorithms, Available: http://www.vlfeat.org/.

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

[17] Large Geometric Model Archive, Georgia Institute of Technology website, Available: https://www.cc.gatech.edu/projects/large_models/

[18] Cyganek, B. & Siebert, J. P. (2009), An Introduction to 3D Computer Vision Techniques and Algorithms. John Wiley & Sons, 2009.

نشریه هوش مصنوعی و داده کاوی

بهینه‌سازی چیدمان دوربین‌ها در سیستم‌های سه‌بعدی با استفاده از مدل ایزو-دیسپریتی برای بهبود دقت در تخمین عمق

مظاهر کرمی[1]، امیر موسوی نیا[2,*] و مهدی احسانیان[1]

[1] دانشکده برق، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.

[2] دانشکده کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.

ارسال ۲۰۱۸/۰۶/۲۶؛ بازنگری ۲۰۱۸/۱۲/۲۸؛ پذیرش ۲۰۱۹/۰۲/۰۸

چکیده:

در مقاله حاضر، مسئله چیدمان خودکار دوربین‌ها در یک سیستم سه‌بعدی به منظور بهبود در عملکرد فرآیند استخراج عمق مورد بررسی قرار گرفته است. در زمان نبود مدل مرجع یا هرگونه اطلاعات پیشین، برای قضاوت در مورد عملکرد و دقت یک سیستم ۳بعدی، تنها مرجع تصمیم‌گیری محاسبه عدم‌قطعیت است. مدل ریاضی سطوح ایزو-دیسپریتی راه حلی موثر را برای تخمین عدم‌قطعیت مرتبط با پارامترهای داخلی و خارجی دوربین در محاسبه عمق در اختیار می‌گذارد. بر این اساس، در ابتدا به کمک روابط ایزو-دیسپریتی، روشی تحلیلی برای تخمین عدم‌قطعیت موجود در محاسبه عمق معرفی می‌کنیم و سپس به کمک روابط به دست آمده، یک تابع هزینه به نحوی تعریف می‌شود که میدان تناظری بیش‌ترین انطباق را با سطح شئ داشته و در عین حال کم‌ترین میزان عدم‌قطعیت در تخمین عمق حاصل شود. چیدمان بهینه دوربین‌ها با کمینه کردن این تابع هزینه نسبت به پارامترهای سیستم و با در نظرگرفتن قیود حاکم به دست می‌آید. در نهایت الگوریتم پیشنهادی بر روی برخی از مدل‌های ۳بعدی پیاده‌سازی می‌شود. نتایج پیاده‌سازی بهبودی قابل توجه (تا ۳۵٪) در بازسازی نهایی را درمقایسه با سیستم‌های یکسوشده شناخته شده نشان می‌دهد.

کلمات کلیدی: بینایی ماشین، تخمین عمق، چیدمان دوربین‌ها، ایزو-دیسپریتی، عدم‌قطعیت، میدان تناظری.