*remote sensing*

MDPI

*Article*

# Relative Importance of Binocular Disparity and Motion Parallax for Depth Estimation: A Computer Vision Approach

**Mostafa Mansour [1,2], Pavel Davidson [1,*], Oleg Stepanov [2] and Robert Piché [1]**

[1] Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland

[2] Department of Information and Navigation Systems, ITMO University, 197101 St. Petersburg, Russia

* Correspondence: pavel.davidson@tuni.fi

Received: 4 July 2019; Accepted: 20 August 2019; Published: 23 August 2019


**Abstract:** Binocular disparity and motion parallax are the most important cues for depth estimation in human and computer vision. Here, we present an experimental study to evaluate the accuracy of these two cues in depth estimation to stationary objects in a static environment. Depth estimation via binocular disparity is most commonly implemented using stereo vision, which uses images from two or more cameras to triangulate and estimate distances. We use a commercial stereo camera mounted on a wheeled robot to create a depth map of the environment. The sequence of images obtained by one of these two cameras as well as the camera motion parameters serve as the input to our motion parallax-based depth estimation algorithm. The measured camera motion parameters include translational and angular velocities. Reference distance to the tracked features is provided by a LiDAR. Overall, our results show that at short distances stereo vision is more accurate, but at large distances the combination of parallax and camera motion provide better depth estimation. Therefore, by combining the two cues, one obtains depth estimation with greater range than is possible using either cue individually.

**Keywords:** binocular disparity; motion parallax; depth perception; proprioceptive sensors; unscented Kalman filter

## 1. Introduction

The human visual system relies on several different cues that provide depth information in static and dynamic environments: binocular disparity, motion parallax, kinetic depth effect, looming, perspective cues from linear image elements, occlusion, smooth shading, blur, etc. Information from multiple cues is combined to provide the viewer with a unified estimate of depth [1]. In this combination, the cues are weighted dynamically depending on the scene, observer motion, lighting conditions, etc.

Computer vision approaches that take into account combination of multiple cues can be implemented using semi-supervised deep neural networks [2]. In this approach the depth of each pixel in an image is directly predicted based on models that have been trained offline on large collections of ground truth depth data. Some approaches attempt to estimate depth from a single image using several monocular visual cues such as texture variations and gradients, defocus, color, haze, etc. [3]. However, such techniques require a significant amount of prior knowledge, since there is an intrinsic ambiguity between local image features and depth variations. Therefore, practical implementations usually incorporate monocular cues into a stereo system.

Bradshaw and Rogers [4] made important conclusion that the mechanisms that support the computation of depth from binocular disparity and motion parallax are independent, but, in general,

it is difficult to measure the depth accuracy of each cue separately and evaluate its importance for human visual perception. It has been proven in numerous experiments with human subjects that in short distances binocular disparity is the most important depth cue and for long distances motion parallax is the most important one [4–8]. However, this topic has been addressed mainly in noncomputer vision studies that provided only qualitative explanation of this phenomena and didn't provide algorithms that can be implemented in computer vision. The accuracy of stereo vision and motion parallax-based depth estimation have not been compared for the same viewing conditions.

Similar to biological visual systems machine vision can also use different visual cues for depth estimation. In robotic systems the depth is usually estimated via binocular disparity that is implemented using a so-called stereo rig: two identical cameras with parallel optical axes that are separated by a known distance [9–11]. The main difficulty in this approach is on the hardware side: the stereo rig has to be rigid, the cameras must be identical, synchronized, and have a wide viewing angle [10]. During fast vehicle motion no difference between capturing time is tolerated [12]. In this approach the cameras have to be carefully calibrated by computing the rotation and the translation between the cameras. This difficulty can be solved if a stereo camera comes in a single housing with preconfigured and precalibrated setup. However, those issues arise if separate cameras are used for wide-baseline stereo vision [13], for example, separate vehicle mounted cameras in cars and ships [14].

Another implementation problem is a point-by-point matching between the two images from the stereo setup to derive the disparity maps. This task is more difficult for stereo vision compared to monocular vision because the corresponding point is not necessarily in the same location in the two stereo images while in monocular vision the corresponding points between the successive images are almost in the same location provided that the frame-rate is high enough. There are many robust implementations for point by point stereo matching, but all of them are computationally complex and challenging for real-time stereo vision applications such as autonomous driving [14]. Real-time implementation of stereo vision with good resolution and high frame rate requires parallel computing using a dedicated GPU or FPGA [15].

Stereo vision can accurately recover the depth for near field objects, but the accuracy degrades with the distance and it becomes inaccurate for distant objects. In human vision, the binocular disparity is the most important depth cue when the distance is less than 5 m [16]. The accuracy of stereo cameras in depth estimation has been already studied in works by the authors of [17,18]. Ortiz et al. [18] used a ZED Stereo Labs camera with adjustable resolution. They concluded that the error in depth estimation of this stereo camera is directly related to the resolution of the camera. For the full HD resolution good accuracy is achieved for distances up to 8 m. For longer distances the depth estimation error increases exponentially and the camera does not compute reliable results for the depths beyond 15 m.

Different studies related to biological visual systems suggest that the combination of retinal image motion and vestibular signals related to head velocity can provide a rudimentary capacity for distance estimation [6,19]. However, if the observer is motionless and only the objects in the environment are moving, motion parallax provides only relative distances and no information about an object's absolute distance from the observer [20].

If the observer is moving its motion generates a baseline for range computations by triangulation. In human vision for distances larger than 10 m the motion parallax visual cue plays the most important role: depth sensing is then based on fusion of visual information (eyes) and egomotion perception (vestibular system). Regan et al. [21] described experiments in which pilots in the act of landing planes were deprived of the use of one eye. Their performance did not deteriorate significantly, so binocular cues cannot be important. Perceptual psychologists have extensively studied motion parallax and have shown that it is of paramount importance for the spatial orientation of car drivers [22] and for pilots landing an aircraft [23]. Many authors agree with Gibson [23,24] that the main depth cue in these cases is the so-called focus of expansion.

The accuracy of motion parallax for depth estimation has been also studied [25–28]. The performance of this approach is strongly affected by the mutual observer and feature point geometry, measurement accuracy of the observer's motion parameters and distance covered by the observer [29,30]. It was found that under favorable conditions the error in distance estimation does not exceed 1% of the distance to a feature point.

In this paper, we provide experimental evaluation of the depth estimation accuracy for two different approaches using the same data: binocular disparity implemented in a stereo camera and motion parallax implemented using images from one of these two cameras and the camera motion measurements. The goal of this paper is to provide quantitative comparison of binocular disparity and motion parallax in depth estimation in machine vision applications. We are also interested in determining whether the combination of binocular cues and motion parallax can significantly extend the range of traditional stereo cameras. The proposed approach can be applied to applications like autonomous driving, robotic navigation and augmented reality. By switching from stereo vision to motion parallax-based approach the range for depth estimation can be potentially extended to hundreds of meters.

The rest of the paper is organized as follows. Sections 2.1 and 2.2 describe our approaches for depth estimation using binocular disparity and motion parallax. Section 2.3 presents the hardware and experiment setup. The data processing steps are explained in Section 2.4. Section 3 present the field tests results. Finally, Section 4 summarizes advantages and disadvantages of stereo vision and motion parallax based approaches for depth estimation.

## 2. Methods

### *2.1. Depth Estimation Using Stereo Cameras*

A typical stereo camera is based on a frontal parallel configuration shown in Figure 1. The left and the right cameras have principal points at $(c_x^l, c_y^l)$ and $(c_x^r, c_y^r)$, respectively, and the two image planes are aligned horizontally with parallel optical axis. Therefore, for an image point in the left plane, there is a corresponding point located at the same row in the right plane. Both points have the same $y$-coordinate. The depth to a feature point ($P^c$) can be calculated as follows [31,32]

$$Z^c = \frac{fB}{x - x'} \tag{1}$$

where $Z^c = Z^{c_l} = Z^{c_r}$ is the depth to the feature point, $f$ is the focal length of the cameras, $B$ is the stereo camera baseline, and $(x, y)$ and $(x', y)$ are the left and the right image points, respectively. The difference between the x-coordinates of the image points is called a disparity. Equation (1) states that disparity is inversely proportional to the depth to a feature point. As a result, at long distances when the disparity is smaller than the pixel size, a stereo camera cannot retrieve the depth.

In practice, depth estimation using a stereo camera involves four steps [31] and errors in any of these steps lead to an error in the depth estimation:

1. Undistortion: In this first step, radial and tangential lens distortions are removed mathematically. The outputs of this step are undistorted images.
2. Rectification: This is the process of adjusting the angles and the distances between cameras to get a frontal parallel arrangement. This process is a crucial step in any stereo imaging algorithm. The outputs of this step are rectified and row-aligned images.
3. Correspondence: This is the process of finding common features in the left and right camera views. The output of this process is a disparity map, where the disparities are the differences in x-coordinates on the image plane.
4. Triangulation: In this process, the disparity map is converted to distances using the known geometric arrangement of the cameras. The output of this process is a depth map.

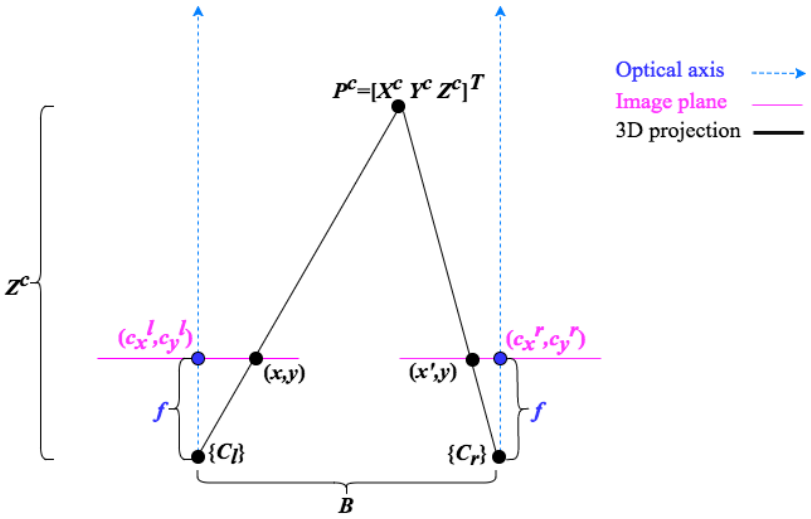


**Figure 1.** Frontal parallel stereo camera configuration for depth estimation.

The Accuracy of the Estimated Depth Using a Stereo Camera

According to Equation (1), the problem of depth estimation is ill-posed with respect to the disparity ($d = x - x'$). That is, a small error in disparity calculation will lead to a significant error in depth estimation. The influence of the errors in disparity on the computed depth can be derived from Equation (1) and is given by

$$|\delta Z^c| = \frac{(Z^c)^2}{fB}\delta d, \tag{2}$$

where $\delta Z^c$ and $\delta d$ are errors in the depth and computed disparity, respectively. Based on this equation we can make the following conclusions.

- The error of the estimated depth is proportional to the error of the computed disparity.
- The error of the estimated depth is proportional to the depth squared.
- The accuracy can be improved if the baseline is increased. However, in this case, the overlapping between cameras decreases, affecting the field of view at short distances.

### *2.2. Depth Estimation Using a Monocular Camera*

Movement of the object's projection on the image plane is specified by the observer's kinematic parameters (angular and translational velocities) by differential equations that relate the coordinates and velocity of a feature point's projection with the observer's translational and angular velocities [28]. These equations can be also applied to biological visual systems to determine the motion of the eye relative to the object from the velocity field of the changing retinal image [33]. These studies show that in the static environment the depth computed based on motion parallax is specified completely by the retinal velocity.

However, estimating distance to the feature points is often quite challenging due to very small motion between frames, especially when a feature point is located close to the focus of expansion. Successful approaches usually integrate proprioceptive sensors to estimate the observer's egomotion to produce a more robust sensing system than typical vision-only techniques [27,28]. The fusion of a bearing measurement provided by a camera with egomotion measurements leads to a nonlinear estimation problem, which can be solved by a nonlinear estimator.

Problem Formulation

Consider a feature point $P^c = [X^c, Y^c, Z^c]^T$ and its projection on the image plane $(x, y)$ shown in Figure 2. To define the problem of depth estimation to a feature point, the evolution of the state sequence is described using the process model [28]

$$\dot{X} = \mathbf{S}\left(X, \tilde{V}_z^c, \tilde{\omega}_y^c, q\right), \tag{3}$$

where $X = [x, y, \xi]^T$ is the state vector; $\xi = \frac{1}{Z^c}$ is the reciprocal of the depth; $\tilde{V}_z^c$ and $\tilde{\omega}_y^c$ are the measured camera linear and angular velocities, respectively; and $q = \left[\Delta\tilde{\omega}_y^c, \Delta\tilde{V}_y^c\right]$ is a vector of process noise whose components are modeled as independent zero-mean white noises with power spectral density (PSD) matrix.

$$Q_c = \begin{bmatrix} \sigma_\omega^2 & 0 \\ 0 & \sigma_V^2 \end{bmatrix}. \tag{4}$$

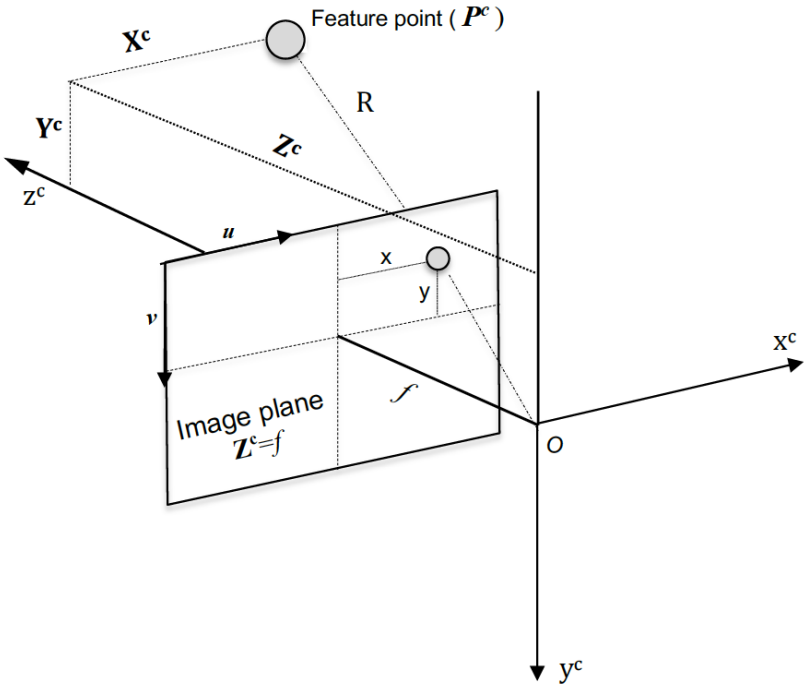


**Figure 2.** Projection of a feature point on the image plane. Reprinted from the work by the authors of [34].

For a wheeled robot moving on level terrain the process model (3) is given by the relation between the depth and the feature point projection on the image plane as follows [28],

$$\begin{aligned}
\dot{x} &= -\left(f + \frac{(x - c_x)^2}{f}\right)\left(\tilde{\omega}_y^c - \Delta\tilde{\omega}_y^c\right) + \left(\tilde{V}_z^c - \Delta\tilde{V}_z^c\right)\xi\,(x - c_x), \\
\dot{y} &= \left(\tilde{V}_z^c - \Delta\tilde{V}_z^c\right)\xi\,(y - c_y) - \frac{\left(\tilde{\omega}_y^c - \Delta\tilde{\omega}_y^c\right)(x - c_x)\,(y - c_y)}{f}, \\
\dot{\xi} &= \left(\tilde{V}_z^c - \Delta\tilde{V}_z^c\right)\xi^2 - \frac{\left(\tilde{\omega}_y^c - \Delta\tilde{\omega}_y^c\right)\xi\,(x - c_x)}{f}.
\end{aligned} \tag{5}$$

The objective is to recursively estimate the state vector in Equation (5) from the camera measurements, which can be described by

$$\mathbf{z}_k = HX_k + v_k, \tag{6}$$

where $v_k$ is a white noise measurement error vector with a covariance matrix $R$, $X_k$ is the state vector evaluated at time step $k$ ,and the measurement matrix is

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{7}$$

To solve the problem under investigation numerically, the model described by Equations (3)–(5) is converted to the difference equation

$$X_k = X_{k-1} + \int_{t_{k-1}}^{t_k} S\left(X_{k-1}, \tilde{V}^c_{z_{k-1}}, \tilde{\omega}^c_{y_{k-1}}\right) dt + G(X_{k-1})q_{k-1}, \tag{8}$$

where $q_{k-1}$ is the vector of discretized system noise that is considered to be zero-mean white Gaussian noise. The integral in Equation (8) can be estimated using numerical integration methods, e.g., the fourth order Runge–Kutta method. Equation (8) can be written as

$$X_k = \Psi\left(X_{k-1}, \tilde{V}^c_{z_{k-1}}, \tilde{\omega}^c_{y_{k-1}}, q_{k-1}\right). \tag{9}$$

The problem under investigation is a nonlinear problem because the process model is nonlinear. Hence, posterior distribution and the state vector can be estimated recursively using a Bayesian filter. In this paper, we used extended Kalman filter and unscented Kalman filter that gave similar results. The solution using extended Kalman filter was decribed in works by the authors of [28,34].

### *2.3. Hardware Description*

As a mobile platform the Robotnik TurtleBot2 unicycle robot was used (Figure 3). The robot was equipped with a Linux computer running Robot Operating System Kinetic (ROS; Kinetic), a calibrated IMU and an odometer. For our machine vision projects we installed additional equipment that included a Hokuyo UTM-30LX LiDAR range finder, Stereo Lab's ZED stereo camera, and NVIDIA Jetson TX2 GPU.

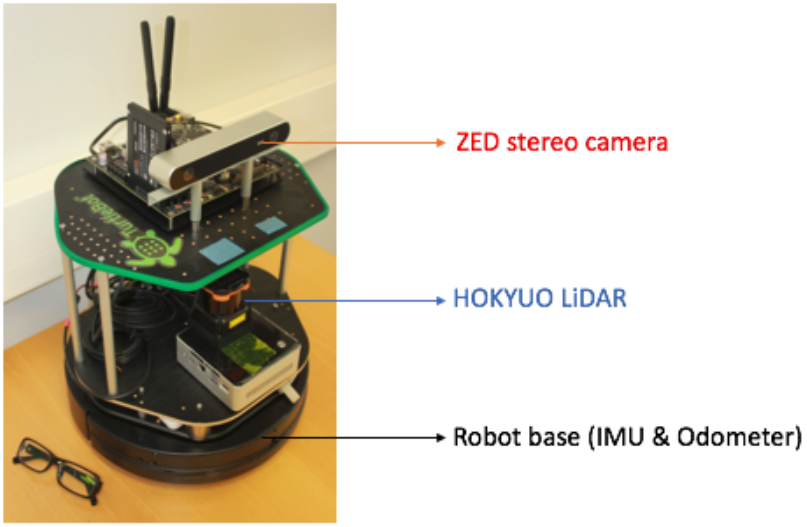


**Figure 3.** Robotnik TurtleBot 2 equipped with the devices in use.

#### 2.3.1. Hokuyo UTM-30LX

This is a 2D laser range finder (LiDAR) that has ~30 m range and distance measurement accuracy of ±30 mm for distances below 10 m and ±50 mm for distances above 10 m [35]. The scanning sector is 270° and angle resolution is 0.25°. Since this LiDAR provides high accuracy range measurements, it was used as ground truth in the field tests.

#### 2.3.2. NVIDIA Jetson TX2 Development Kit

Real-time operation of stereo cameras requires parallel computing using a dedicated GPU or FPGA. In our experiments with ZED stereo camera the real-time computations were performed by an Nvidia Jetson Tx2 GPU with 256 cores (1.3TFLOPs in total). These computational resources allow real-time computation of the depth map with Full HD resolution and 15 fps. The development kit included Jetson TX2 module with NVIDIA Pascal GPU with 8 GB of memory, ARM 128-bit CPUs (4 cores), 8 GB LPDDR4, 32 GB eMMC, and WiFi module. A Samsung 250 GB SSD was added for more drive space. CUDA Toolkit 9.0 was installed to take advantage of the parallel computing power of the GPU.

#### 2.3.3. ZED Stereo Camera

ZED stereo camera consists of two cameras separated by a 120 mm baseline. Each camera has a 4 megapixel sensor with resolution of 2 microns. The maximum range of this camera is 20 m and the maximum frame rate is 100 fps. The disparity and depth map are computed using ZED's software development kit (SDK). Since running the SDK is computationally expensive we used the parallel computing capabilities of the NVIDIA GPU to calculate the disparity and depth maps.

### *2.4. Data Processing*

The robot's base, the camera, and the LiDAR were connected to the main computer running ROS Kinetic. The following measurements were saved for postprocessing.

- Angular and linear velocities measured by the IMU and odometer.
- Distance values from the Hokuyo LiDAR.
- Rectified RGB images from the left ZED camera.
- The depth map computed with respect to the center of the left camera using ZED SDK and CUDA toolkit.

The robot is controlled by a computer over a wireless network. The computer drives the robot towards a feature point in a preplanned trajectory. The feature point is fixed on a vertical plane (wall). From the ranges to the plane measured by the LiDAR, depth to the feature point is extracted and transformed to the left camera coordinate frame to be used as a reference value.

A typical example of the experiment setup is shown in Figure 4. The stereo camera creates a depth map from which the distance to the feature point is derived. At the same time, the sequence of images from the left ZED camera together with IMU and odometer data are used as input to our algorithm for depth estimation. The collected data is time-stamped and stored for post-processing. The pipeline of the data processing includes four tasks: reference distance, feature point detection and tracking, stereo depth map, and motion parallax-based distance estimation (Figure 5).


**Figure 4.** An example of experiment setup.

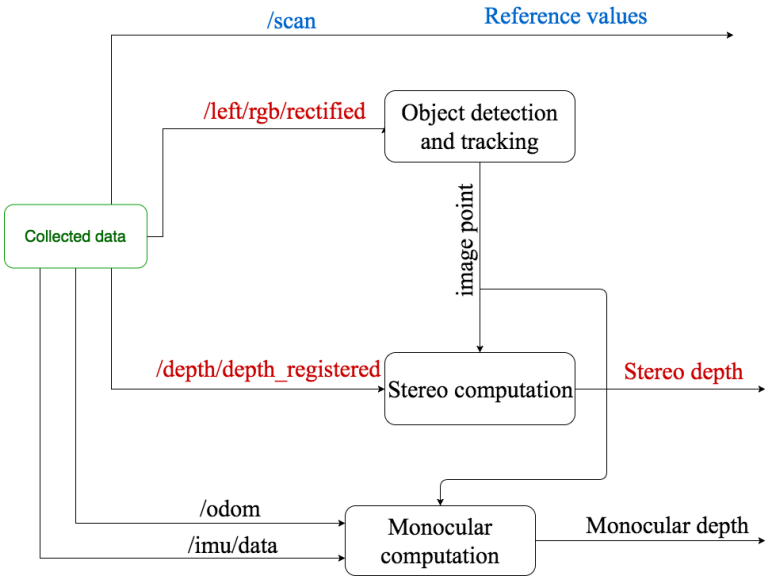


**Figure 5.** Data processing pipeline.

#### 2.4.1. Feature Point Ranging

The LiDAR publishes its time stamped messages on "/scan" ROS topic. Each message includes the ranges to the obstacles located in the horizontal plane of the LiDAR. From these ranges, the corresponding range to the feature point plane is extracted. This range is used as a reference value at every time step.

#### 2.4.2. Object Detection and Tracking

The left camera publishes its RGB rectified images on "/left/rgb/rectified" ROS topic. Instead of making an exhaustive search to find the corresponding points from frame to frame, an object tracker

is used first to produce a candidate region for the object of interest and then the feature matching algorithm will search for the corresponding image points inside this region.

#### 2.4.3. Stereo Computation

The output of the previous block is a corresponding image point. The depth to this point can be extracted from the depth map. ZED calculates depth map using the stereo camera geometric and intrinsic parameters and publishes this map on "/depth/depthregistered" ROS topic.

#### 2.4.4. Monocular Computation

In this block we use the algorithm presented by the authors of [28] to estimate the depth to a feature point using a monocular camera. The algorithm has three inputs: camera linear velocity, camera angular velocity, and image point coordinates. The velocities are the robot velocities transformed to camera coordinate frame. The robot publishes its linear and angular velocities on "/odom" and "/imu/data" ROS topics, respectively.

## 3. Results

The experiments were carried using two different camera resolutions: HD (1280 × 720) and full HD (1920 × 1080) with a frame rate of 15 fps. In the experiments with lower resolution we considered three cases: two cases with favorable geometry for motion parallax with the features far from the focus of expansion (peripheral features) and one case with poor geometry with the features closed to the focus of expansion (head-on features).

In the case with peripheral features, first, the robot was placed at 3.3 m from the feature point and was moving back in an, approximately, straight line up to 8.2 m from the feature point with a nominal speed of approximately 0.5 m/s. The angle between the line-of-sight (LOS) and velocity vector is about 41° at the closest point and 15° at the farthest point. Then, the robot was placed at 2.2 m from the feature point and was moving backward in an approximately straight line up to 7.26 m from the feature point with a speed of about 0.5 m/s. The angle between the LOS and velocity vector is ~31° at the closest point and ~12° at the farthest point.

In the case with head-on features, the robot was placed at 3.2 m from the feature point and was moving backward in an approximately straight line up to 7.26 m from the feature point with a speed of about 0.5 m/s. The angle between the LOS and velocity vector is ~15° at the closest point and ~5° at the farthest point.

The experiments with full HD resolution consider only the case of head-on features when at the beginning the feature point is at 25 m distance and the camera is moving forward, approximately, along the straight line with a nominal speed of ~0.5 m/s. The angle between the LOS and velocity vector is ~4° at 10 m and ~1.1° at the farthest point.

For the tests with lower resolution, the effective range of the stereo camera is only about 7 m. The median error in depth estimation of the monocular camera is ~0.15–0.3 m (Figure 6) for the tests with peripheral features and 0.35 m (Figure 6) for the tests with head-on features. The performance of the stereo camera is mostly affected by the distance and slightly by the geometry. Depending on the geometry, the motion parallax approach outperforms stereo vision for the depths larger than 4–7 m, and the binocular disparity is clearly better for the distances smaller than 4 m (Figures 7 and 8).

The tests with full HD resolution cover only the scenarios with poor geometry. In the beginning the angle between the direction of camera movement and the line-of-sight (LOS) is only ~1.1°. This is very difficult case for the motion parallax-based distance estimation since the feature is very close to the focus of expansion. Therefore, the convergence is slow (Figure 9). After 25 s, the filter settles and the depth estimation error is ~0.5 m. For this scenario, the motion parallax approach outperforms binocular disparity for depths larger than 10 m while binocular disparity is better for depths less than 7 m.

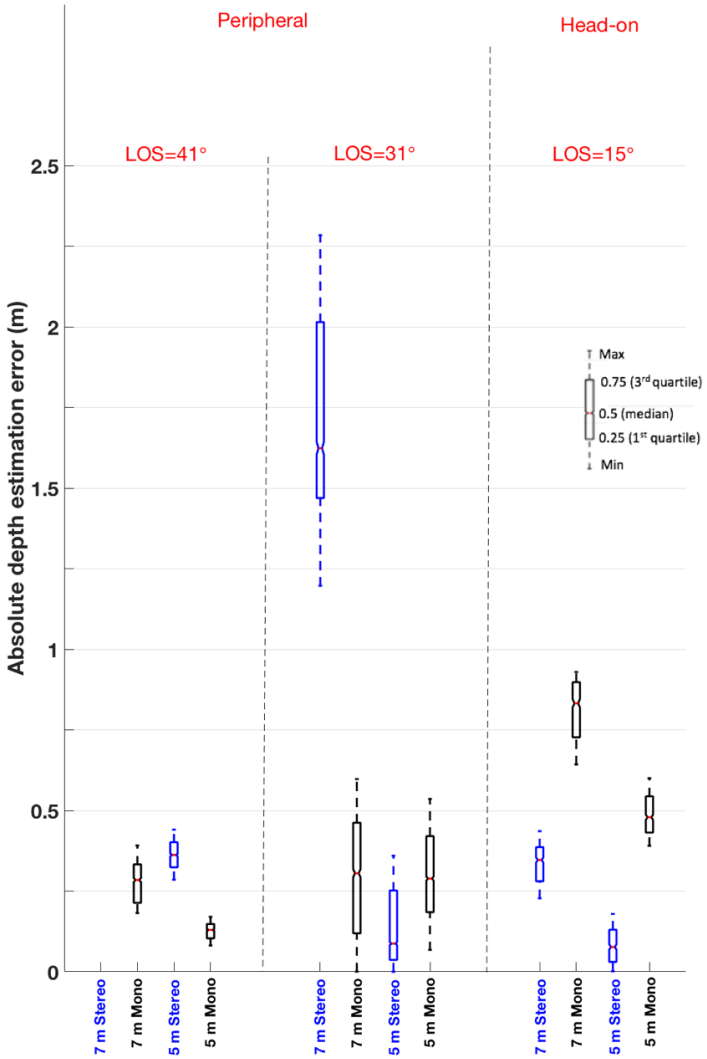


**Figure 6.** Absolute depth estimation error for both approaches at different distances. The errors are represented by the minimum, first quartile, median, third quartile, and maximum of a set of data and shown by box-and-whisker plots. For peripheral point features when line-of-sight (LOS) is 41° the effective range of stereo camera is less than 7 m, therefore the corresponding box is not shown.

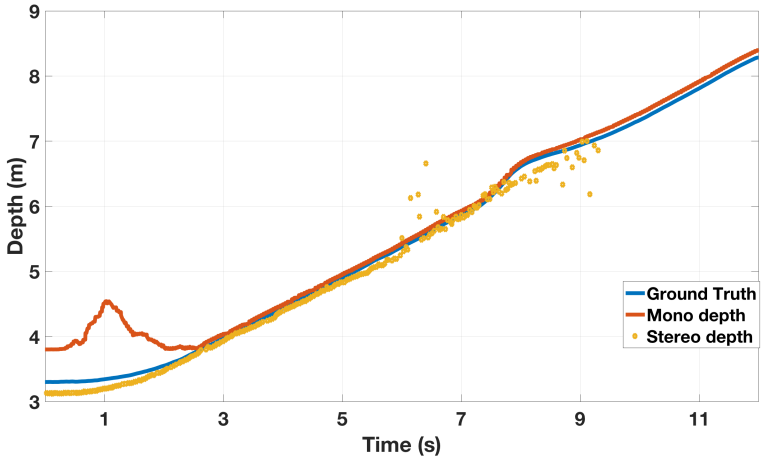


**Figure 7.** Depth estimation results for the stereo vision (amber) and motion parallax (red) approaches in the case of a peripheral point feature when the LOS angle is between 15 and 41°. The ground truth is shown by the blue line. The camera resolution is set to HD.

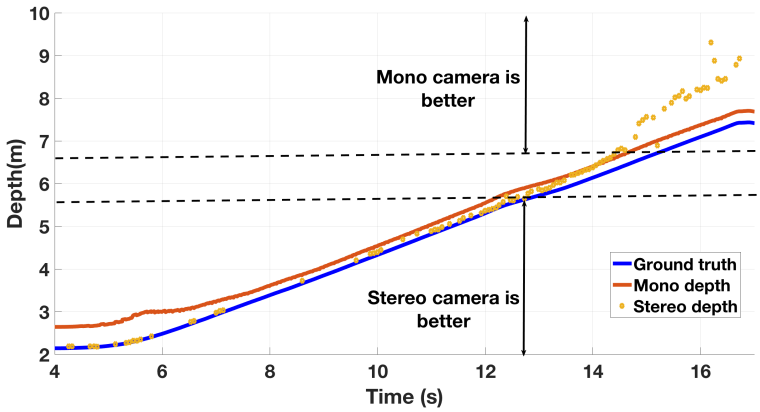


**Figure 8.** Depth estimation results for the stereo vision (amber) and motion parallax (red) approaches in the case of a peripheral point feature when the LOS angle is between 12 and 31°. The ground truth is shown by the blue line. The camera resolution is set to HD.

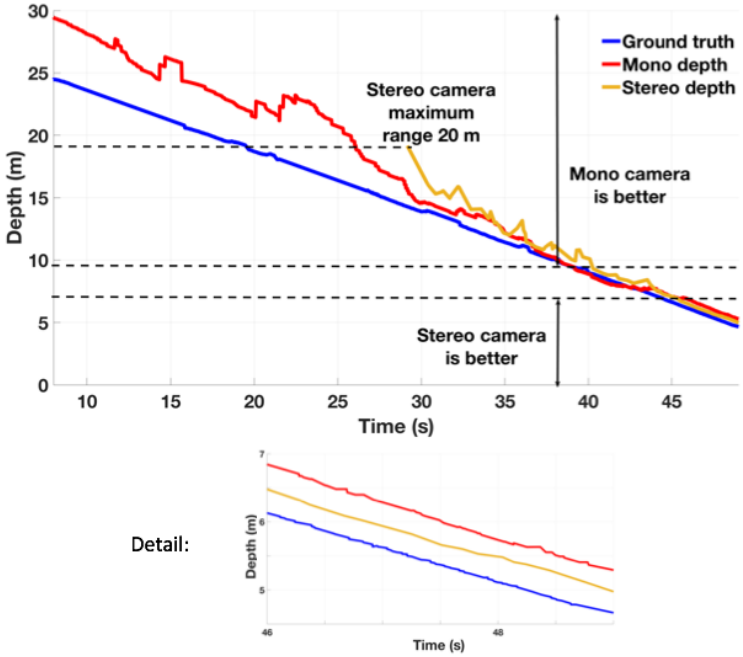


**Figure 9.** Depth estimation results for the stereo vision (amber) and motion parallax (red) approaches in the case of a head-on point feature when the LOS angle is between 1.1 and 4°. The ground truth is shown by the blue line. The camera resolution is set to full HD. The lower part emphasizes the results for depths less than 7 m.

## 4. Discussion

The performance of stereo camera in depth estimation depends on the following factors; distance and angle to the point features, texture, and camera resolution. Distance to the point feature has the biggest impact on the depth estimation performance because according to Equation (2) the error in the estimated depth increases quadratically with the estimated depth. For any stereo camera, there is a maximum distance for which the depth estimation can be produced. Camera resolution also affects the depth estimation performance because the disparity error ($\delta d$) in Equation (2) decreases with the increase in camera resolution causing the error of the estimated depth to be reduced. The field of view (FOV) of a stereo camera is smaller than the FOV of an equivalent monocular camera. Besides the accuracy and maximum range of a stereo camera in depth estimation is better for head-on point features than for peripheral ones.

Performance of the motion parallax-based algorithm in depth estimation can be evaluated in terms of accuracy, convergence time, consistency, and influence of the initial guess on the depth estimation. The geometry of the camera and feature point makes the biggest impact on the accuracy and convergence time. If the camera motion creates significantly different viewpoints of an object, the distance to this object can be estimated more accurately.

In a previous work [27], we showed that the accuracy of depth estimation and speed of filter convergence deteriorate significantly when the angle between the LOS and velocity vector is less than 5°. If there is no significant change in the LOS, the disparity will be close to zero. We have found that for HD resolution if the Euclidean distance in disparity is less than 10 pixels, the algorithm will behave poorly and diverge. The convergence time for head-on features may be several times longer than the convergence time for peripheral features. Figures 7 and 10 show that the convergence time for peripheral points is ~2 s and for head-on points (Figure 9) is ~25 s.

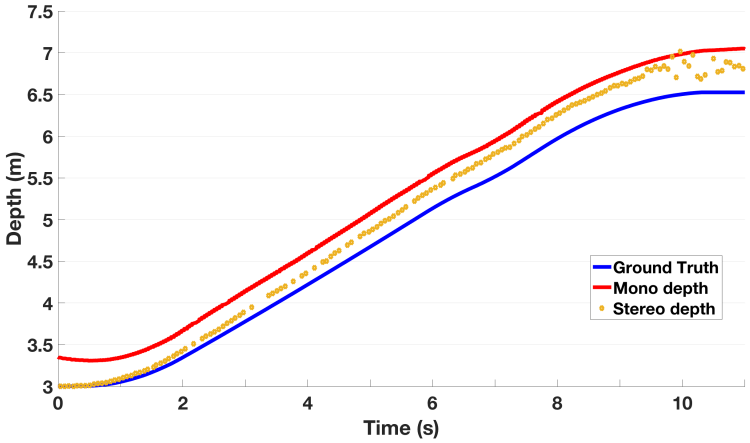


**Figure 10.** Depth estimation results for the stereo vision (amber) and motion parallax (red) approaches in the case of a head-on point feature when the LOS angle is between 5 and 15°. The ground truth is shown by the blue line. The camera resolution is set to HD.

The impact of initial guess is shown in Figure 11. It can be seen that the algorithm can tolerate initial errors that are 3–4 times larger than the actual depth. The large errors increase slightly the convergence time. The measurement errors in proprioceptive sensors also have an influence on the accuracy of this approach. However, it was found that for wheeled vehicles equipped with an odometer and reasonably good MEMS IMU, the influence of these errors on depth estimation accuracy is small [27].

The results presented in Figures 8 and 9 show that the motion parallax-based approach outperforms binocular disparity based approach for large distances. Therefore we considered an algorithm that can extend range of a stereo camera by switching to the motion parallax-based approach for large distances. For a given resolution the switching point depends on the distance and angle to the feature.

Integrity monitoring of a stereo camera is provided by a confidence map that includes the estimated depth for each pixel described by a value from 0 to 100 where 0 corresponds to completely unreliable depth estimation. To improve performance of a stereo camera we switched to motion parallax-based approach when the confidence value of the pixel of interest becomes less than 80. In the experiments using HD resolution, we found that the confidence value drops below 80 for depths greater than 6 m. Figure 12 shows the errors in depth estimation for a stereo camera and the corresponding confidence values. For example, at 6 m the confidence value is less than 80 and the depth estimation error is 0.2 m. At this point, the algorithm is switched from stereo camera to monocular camera for better depth estimation as shown in Figure 13.

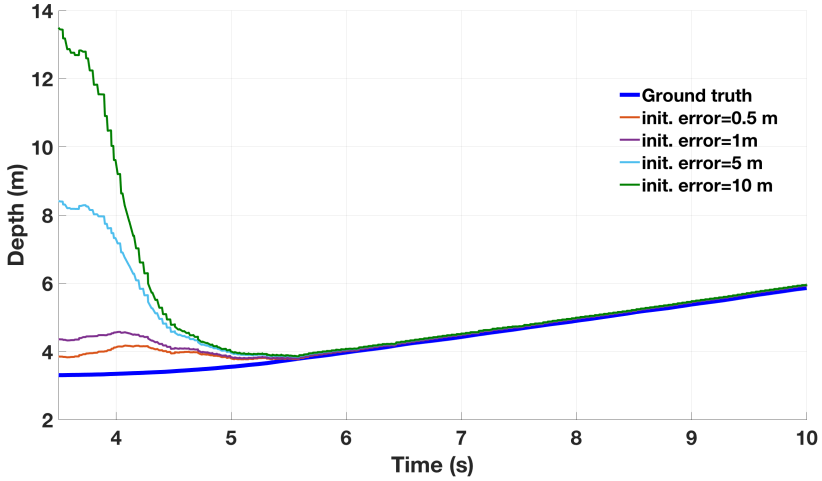


**Figure 11.** Influence of initial guess on accuracy and convergence. The depth estimation algorithm based on motion parallax is robust against initialization errors.

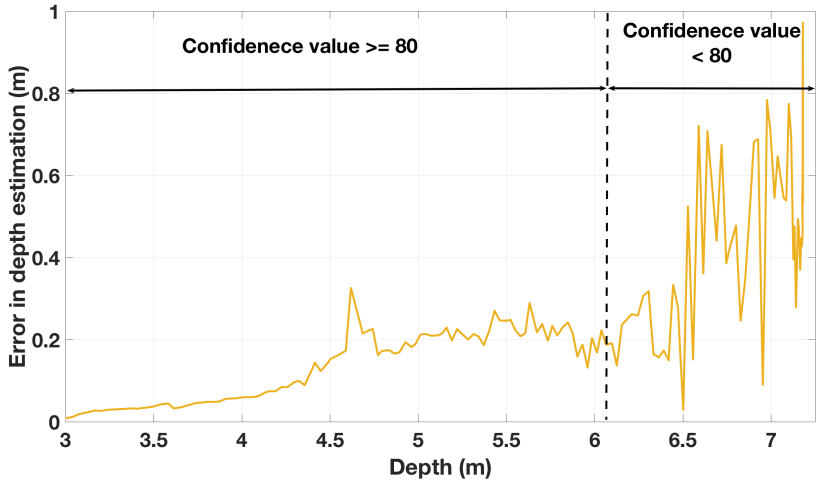


**Figure 12.** Estimation error using stereo camera.

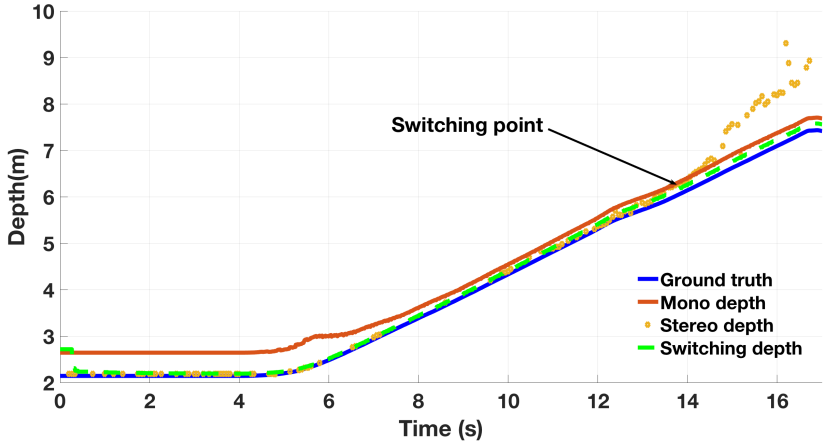


**Figure 13.** Switching between binocular disparity and motion parallax for depth estimation. The camera resolution is set to HD.

## 5. Conclusions

In this paper, we compared depth estimation performance of motion parallax and binocular disparity visual cues. We used two different camera resolutions and feature points locations. The field

tests confirmed that motion parallax outperforms binocular vision for distant features. However, binocular vision is more accurate for near ranges. We proposed a method to extend the range of stereo cameras by switching to motion parallax when binocular disparity becomes unreliable. This approach can overcome the limitations of the stereo camera as follows.

- The image points in successive frames are tracked as a part of the state vector and camera noises will be presented by a measurement error. By doing so, we have the ability to estimate them more accurately and decrease the error in the disparity between the image points recursively in the successive frames.
- It is a multiview-based approach. It means that the baseline is increased from frame to frame in a dynamic way due to the camera motion. It leads to increasing the baseline–depth ratio, making depth estimation more accurate.
- The proposed approach does not depend, as in plain stereo vision, on how far the object is. Instead, it depends on how wider the triangulation angle between the first view and the last view is.

Our future work will address the possibility of fusing stereo visual odometer with monocular camera measurements for depth estimation. By doing so, we relax the requirements for using IMU and odometer to measure motion parameters. This can be useful in many cases where the odometer is not available like in UAVs.

**Author Contributions:** Study conception, P.D.; hardware and software development, M.M.; data acquisition, M.M.; data processing, M.M.; visualization, M.M.; writing the original draft, M.M.; development of methodology, P.D.; validation, P.D., O.S. and R.P.; manuscript drafting and revision, P.D., O.S. and R.P.; project administration P.D.; data interpretation, R.P. and O.S.; supervision of the project, R.P.; funding acquisition, R.P. and O.S.

**Funding:** This research was partially funded by the Government of the Russian Federation grant number 08-08.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Landy, M.S.; Maloney, L.T.; Johnston, E.B.; Young, M. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vis. Res.* **1995**, *35*, 389–412. [CrossRef]
2. Smolyanskiy, N.; Kamenev, A.; Birchfield, S. On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1007–1015.
3. Saxena, A.; Schulte, J.; Ng, A.Y. Depth Estimation Using Monocular and Stereo Cues. In Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007; Volume 7, pp. 2197–2203.
4. Bradshaw, M.F.; Rogers, B.J. The interaction of binocular disparity and motion parallax in the computation of depth. *Vis. Res.* **1996**, *36*, 3457–3468. [CrossRef]
5. Durgin, F.H.; Proffitt, D.R.; Olson, T.J.; Reinke, K.S. Comparing depth from motion with depth from binocular disparity. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 679. [CrossRef] [PubMed]
6. Ono, H.; Wade, N.J. Depth and motion perceptions produced by motion parallax. *Teach. Psychol.* **2006**, *33*, 199–202.
7. McKee, S.P.; Taylor, D.G. The precision of binocular and monocular depth judgments in natural settings. *J. Vis.* **2010**, *10*, 5. [CrossRef] [PubMed]
8. Sousa, R.; Brenner, E.; Smeets, J. A new binocular cue for absolute distance: Disparity relative to the most distant structure. *Vis. Res.* **2010**, *50*, 1786–1792. [CrossRef]
9. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
10. Lazaros, N.; Sirakoulis, G.C.; Gasteratos, A. Review of stereo vision algorithms: From software to hardware. *Int. J. Optomechatron.* **2008**, *2*, 435–462. [CrossRef]
11. Hamzah, R.A.; Ibrahim, H. Literature survey on stereo vision disparity map algorithms. *J. Sens.* **2016**, *2016*, 8742920. [CrossRef]

12. Vishnyakov, B.V.; Vizilter, Y.V.; Knyaz, V.A.; Malin, I.K.; Vygolov, O.V.; Zheltov, S.Y. Stereo sequences analysis for dynamic scene understanding in a driver assistance system. In Proceedings of the Automated Visual Inspection and Machine Vision, Munich, Germany, 21–25 June 2015; Volume 9530.
13. Wu, B.; Zhang, Y.; Zhu, Q. A triangulation-based hierarchical image matching method for wide-baseline images. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 695–708. [CrossRef]
14. Milella, A.; Reina, G. 3D reconstruction and classification of natural environments by an autonomous vehicle using multi-baseline stereo. *Intell. Serv. Robot.* **2014**, *7*, 79–92. [CrossRef]
15. Tippetts, B.; Lee, D.J.; Lillywhite, K.; Archibald, J. Review of stereo vision algorithms and their suitability for resource-limited systems. *J. Real-Time Image Process.* **2016**, *11*, 5–25. [CrossRef]
16. Kytö, M.; Nuutinen, M.; Oittinen, P. Method for measuring stereo camera depth accuracy based on stereoscopic vision. In Proceedings of the SPIE Three-Dimensional Imaging, Interaction, and Measurement Conference, San-Francisco, CA, USA, 24–27 January 2011; p. 7864.
17. Sabattini, L.; Levratti, A.; Venturi, F.; Amplo, E.; Fantuzzi, C.; Secchi, C. Experimental comparison of 3D vision sensors for mobile robot localization for industrial application: Stereo-camera and RGB-D sensor. In Proceedings of the 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV), Guangzhou, China, 5–7 December 2012; pp. 823–828.
18. Ortiz, L.E.; Cabrera, E.V.; Gonçalves, L.M. Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs. *Electron. Lett. Comput. Vis. Image Anal.* **2018**, *17*, 1–15. [CrossRef]
19. Hanes, D.A.; Keller, J.; McCollum, G. Motion parallax contribution to perception of self-motion and depth. *Biol. Cybern.* **2008**, *98*, 273–293. [CrossRef] [PubMed]
20. Holmin, J.; Nawrot, M. Motion parallax thresholds for unambiguous depth perception. *Vis. Res.* **2015**, *115*, 40–47. [CrossRef] [PubMed]
21. Regan, D.; Beverley, K.; Cynader, M. The visual perception of motion in depth. *Sci. Am.* **1979**, *241*, 136–151. [CrossRef] [PubMed]
22. Gordon, D.A. Static and dynamic visual fields in human space perception. *J. Opt. Soc. Am.* **1965**, *55*, 1296–1303. [CrossRef] [PubMed]
23. Gibson, J.J. *The Ecological Approach to Visual Perception: Classic Edition*; Psychology Press: Abingdon, UK, 2014.
24. Gibson, J.J. *The Perception of the Visual World*; Houghton Mifflin: Boston, MA, USA, 1950.
25. Grabe, V.; Bulthoff, H.H.; Giordano, P.R. A comparison of scale estimation schemes for a quadrotor UAV based on optical flow and IMU measurements. In Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013, pp. 5193–5200.
26. Schmid, S.; Fritsch, D. Precision analysis of triangulations using forward-facing vehicle-mounted cameras for augmented reality applications. In Proceedings of the Videometrics, Range Imaging, and Applications XIV, Munich, Germany, 26–27 June 2017; Volume 10332.
27. Davidson, P.; Raunio, J.P.; Piché, R. Monocular vision-based range estimation supported by proprioceptive motion. *Gyroscopy Navig.* **2017**, *8*, 150–158. [CrossRef]
28. Mansour, M.; Davidson, P.; Stepanov, O.; Aref, M.; Raunio, J.P.; Piché, R. Depth estimation with egomotion assisted monocular camera. *Gyroscopy Navig.* **2019**, *10*, 111–123.
29. Oshman, Y.; Davidson, P. Optimal observer trajectories for passive target localization using bearing-only measurements. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, San-Diego, CA, USA, 29–31 July 1996; p. 3740.
30. Oshman, Y.; Davidson, P. Optimization of observer trajectories for bearings-only target localization. *IEEE Trans. Aerosp. Electron. Syst.* **1999**, *35*, 892–902. [CrossRef]
31. Kaehler, A.; Bradski, G. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2016.
32. Corke, P. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB®*; Springer-Verlag: Berlin/Heidelberg, Germany, 2011; Volume 73, pp. 329–340.
33. Longuet-Higgins, H.C.; Prazdny, K. The interpretation of a moving retinal image. *Proc. R. Soc. Lond. B* **1980**, *208*, 385–397.

34. Davidson, P.; Mansour, M.; Stepanov, O.; Piché, R. Depth estimation from motion parallax: Experimental evaluation. In Proceedings of the 2019 26th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), St. Petersburg, Russia, 27–29 May 2019.
35. Cooper, M.A.; Raquet, J.F.; Patton, R. Range Information Characterization of the Hokuyo UST-20LX LiDAR Sensor. *Photonics* **2018**, *5*, 12. [CrossRef]