

Article

Three-Dimensional Reconstruction of Indoor and Outdoor Environments Using a Stereo Catadioptric System

Diana-Margarita Córdova-Esparza ^{1,*}, Juan Terven ², Julio-Alejandro Romero-González ¹ and Alfonso Ramírez-Pedraza ³

¹ Faculty of Informatics, Autonomous University of Queretaro, Av. de las Ciencias S/N, Campus Juriquilla, Queretaro C.P. 76230, Mexico; julio.romero@uaq.mx

² AiFi Inc., 2388 Walsh Av., Santa Clara, CA 95051, USA; juan@aifi.io

³ Research Center in Optics A.C., Lomas del Bosque 115, Col. Lomas del Campestre, Leon C.P. 37150, Mexico; pedro.ramirez@cio.mx

* Correspondence: diana.cordova@uaq.mx

Received: 19 October 2020; Accepted: 7 December 2020; Published: 10 December 2020



Abstract: In this work, we present a panoramic 3D stereo reconstruction system composed of two catadioptric cameras. Each one consists of a CCD camera and a parabolic convex mirror that allows the acquisition of catadioptric images. We describe the calibration approach and propose the improvement of existing deep feature matching methods with epipolar constraints. We show that the improved matching algorithm covers more of the scene than classic feature detectors, yielding broader and denser reconstructions for outdoor environments. Our system can also generate accurate measurements in the wild without large amounts of data used in deep learning-based systems. We demonstrate the system's feasibility and effectiveness as a practical stereo sensor with real experiments in indoor and outdoor environments.

Keywords: reconstruction 3D; stereo vision; DeepMatching

1. Introduction

Conventional stereoscopic reconstruction systems based on perspective cameras have a limited field-of-view that restricts them in applications that involve wide areas such as large-scale reconstruction or self-driving cars. To compensate for the narrow field-of-view, these systems acquire multiple images and combine them to cover the desired scene [1]. A different way to increase the field-of-view is to use fish-eye lenses [2] or mirrors in conjunction with lenses to constitute a panoramic catadioptric system [3,4].

Catadioptric omnidirectional stereo vision can be achieved using one camera with multiple mirrors [5,6]; two cameras and a single mirror [7] or two cameras and two mirrors [8]. The common types of mirrors used are hyperbolic double mirrors [9], multiple conic mirrors [10,11], multiple parabolic mirrors [12], concave lens with a convex mirrors [13,14], multi-planar mirrors [15], and mirror pyramids [16]. The large field-of-view of these systems makes them useful in applications such as driver assistance tasks [17,18], unmanned vehicle aerial navigation [19–21], visual odometry [22], and 3D reconstruction for metrology applications [23,24].

To perform 3D reconstruction with a stereo system, we need to find matching points on the set of cameras. Typically, there are two methods for matching image pairs in passive stereo systems: dense methods based on finding a disparity map by pixel-wise comparison of rectified images along the image rows, and sparse methods using matching features based on pixel region descriptors.

One challenge of the dense methods with panoramic images is that the rectified images are heavily distorted, and their appearance depends on the relative configuration of the two cameras. Therefore, many pixels are interpolated in some image regions, influencing pixel mapping precision and 3D reconstruction. A limitation of dense omnidirectional stereo matching is the running time making it not suitable for real-time applications [25].

On the other hand, sparse methods with panoramic images provide a confined number of matching features for triangulation [26,27] due to the reduced spatial resolution, wasted image-plane pixels, and exposure control [28]. Due to this, it is a common practice to manually select matching points for calibration or to validate the measurements [5,29]. Another prevalent limitation of these systems is the validation in indoor environments only. For example, Mao et al. [26] present a reconstruction algorithm for multiple spherical images; they use a combination of manually selected points and SIFT matching to estimate the fundamental matrix for each image pairs and present sparse reconstruction results in indoor environments. Fiala et al. [5], present a bi-lobed panoramic sensor with the matching features detected using the Panoramic Hough Transform. The results show sparse points of scene polygons in indoor environments. The work presented by [29] similar to our research proposal, shows a sensor with the same amount of components but with the use of hyperbolic mirrors. The 3D reconstruction was done on manually selected points to avoid possible errors due to an automatic matching. To validate the method, they selected four corners of the ceiling to compute the size of the room with an estimated error between 1 and 6 cm. Zhou et al. [30] propose an omnidirectional stereo vision sensor composed of one CCD camera and two pyramid mirrors. To evaluate the performance of the proposed sensor, they tested a calibration plane. Although this sensor cannot accomplish a whole panoramic view, it nevertheless provides a useful vision measurement approach for indoor scenes. Jang et al. [31] propose a catadioptric stereo camera system that uses a single camera and a single lens with conic mirrors. Stereo matching was conducted with a window-based correlation search. The disparity map and the depth map show good reconstruction accuracy indoors except for ambiguous regions like repetitive wall patterns where the intensity-based matching does not work. Ragot et al. [32] propose a dense matching algorithm based on interpolation of sparse points. The matchings points are processed during calibration, and results are stored in Look-Up Tables. The reconstruction algorithms are validated on synthetic images and real indoor images. Chen et al. [4] present a catadioptric multi-stereo system composed of a single camera and five spherical mirrors. They implemented a dense stereo matching and multi point cloud fusion. The results presented in indoor scenes improved between 15% to 30% when combining the stereo results.

Alternatively, deep learning techniques have recently obtained remarkable results in estimating depth from a single perspective image in indoor and outdoor environments. Depth estimation can be defined as a pixel-level regression, and the model architectures usually use Convolutional Neural Networks (CNNs) architectures. Xu et al. [33] proposed predicting depth maps from RGB inputs in a multi-scale framework combining CNNs with Conditional Random Fields (CRF). Later in [34] they propose a multi-task architecture called PAD-Net capable of simultaneously perform depth estimation and scene parsing. Fu et al. [35] introduced a depth estimation network that uses dilated convolutions and a full-image encoder to directly obtain a high-resolution depth map and improve the training time through depth discretization and the use of ordinal regression loss. More related to our work is the one from Won et al. [36], where an end-to-end deep learning model estimates depth from multiple fisheye cameras. Their network can learn global context information and reconstructs accurate omnidirectional depth estimates. However, all these methods require large training sets of RGB-depth pairs, which can be expensive to obtain, and the quality of the results is limited by the data used for training.

Deep Learning techniques have also been applied in feature matching, and optical flow applications [37,38]. Sarlin et al. [39] proposed a graph neural network along with a context aggregation mechanism based on attention to allow their model to reason about the 3D information and the feature assignments. However, the performance is bounded by the training data used and the augmentations performed during training.

To combine the best of both approaches (panoramic imaging and Deep Learning), in this work, we present a stereo catadioptric system that uses a deep learning-inspired feature matching algorithm called DeepMatching [40] augmented with stereo epipolar constraints. With this, our system can produce wide field-of-view 3D reconstructions in indoor and outdoor scenes without the need of specific training data.

The main contributions of this work are twofold. First, we propose a catadioptric stereo system capable of generating semi-dense reconstructions using deep learning matching methods such as DeepMatching [40] with stereo epipolar constraints. Second, the system can produce 3D reconstructions in indoor or outdoor environments at higher framerate than dense methods without new training data.

The rest of the paper is organized as follows: Section 2 illustrates the catadioptric vision system, Section 3 describes the methodology, Section 4 discusses the results, and Section 5 presents the conclusion.

2. Catadioptric Vision System

This section presents the catadioptric vision system's experimental arrangement, which consists of two omnidirectional cameras, including a CCD camera and a parabolic mirror.

Experimental Setup

Figure 1 shows a schematic overview of the experimental setup for the proposed panoramic 3D reconstruction system, which consists of two catadioptric cameras; an upper camera and a lower camera, each composed of a 0–360 parabolic mirror and a Marlin F-080 CCD camera on a back-to-back configuration.

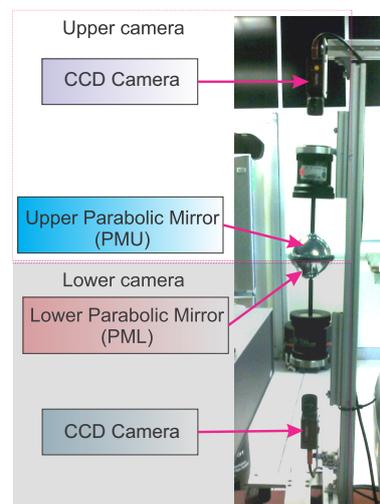


Figure 1. Panoramic 3D reconstruction system formed by two catadioptric cameras (upper camera and lower camera), each composed by a parabolic mirror and a Marlin F-080c camera.

We assembled the catadioptric cameras to capture the full environment that is reflected in the mirrors. For this purpose, we aligned a mirror with the lower CCD camera to compose the lower parabolic mirror (PML). Then we placed another mirror back-to-back with PML aligned with the upper CCD camera to produce the upper parabolic mirror (PMU). Using a chessboard calibration pattern, we calibrate both catadioptric cameras obtaining their intrinsic and extrinsic parameters, as explained in the next section.

3. Methodology

This section presents the methodology to perform 3D reconstruction of an object or scene using the proposed catadioptric stereo system. Section 3.1 describes the calibration procedure, Section 3.2 explains the epipolar geometry for panoramic cameras, and Section 3.3 describes the stereo reconstruction.

3.1. Catadioptric Camera Calibration

To perform 3D reconstructions with metric dimensions, it is necessary to know the cameras' intrinsic parameters, such as the focal length, optical center, mirror curvature, and extrinsic parameters composed by the rotation and translation between them. These parameters are obtained with the camera calibration. This section briefly describes the process to calibrate the catadioptric stereo vision system based on the work presented in [41].

To calibrate the catadioptric cameras, we used the geometric model proposed in [42]; this model uses a checkerboard pattern (shown in Figure 2a) to obtain the intrinsic parameters of each camera and the extrinsic parameters between the reference system of the cameras and that of the pattern. The pixel coordinates of a point in the image are represented by (u, v) . Equation (1) shows the relation between a 3D point X_W in the mirror $x_{PM} = [x, y, z]^T$ and a 2D point in the image.

$$x_{PM} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(u, v) \end{bmatrix} \tag{1}$$

Since the mirror is rotationally symmetric, the function $f(u, v)$ only depends on the distance $\rho = \sqrt{u^2 + v^2}$ leading to Equation (2).

$$x_{PM} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(\rho) \end{bmatrix} \tag{2}$$

where $f(\rho)$ is a fourth degree polynomial that defines the curvature of the parabolic mirror:

$$f(\rho) = a_0 + a_1\rho + a_2\rho^2 + a_3\rho^3 + a_4\rho^4 \tag{3}$$

the coefficients a_0, a_1, a_2, a_3, a_4 correspond to the intrinsic parameters of the mirror.

To calibrate the catadioptric camera, we projected the checkerboard pattern shown in Figure 2a at different positions around the catadioptric stereo system. Figure 2b shows images from the pattern acquired with the upper camera and Figure 2c shows the images acquired with the lower camera.

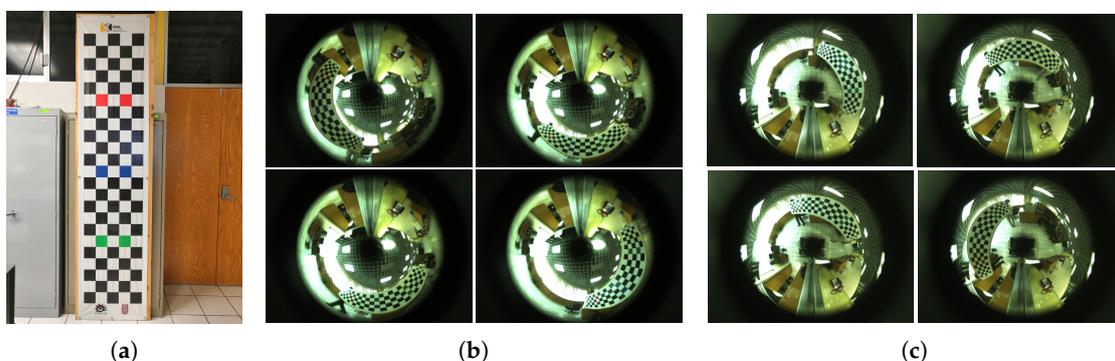


Figure 2. Calibration pattern. (a) Chessboard pattern (b) Images of the pattern using the upper camera. (c) Images of the pattern using the lower camera.

3.2. Epipolar Geometry for Panoramic Cameras

Epipolar geometry describes the relationship between positions of corresponding points in a pair of images acquired by different cameras [41,43,44]. Given a feature point in one image, its matching view in the other image must lie along the corresponding epipolar curve. This fact is known as the epipolar constraint. Once we know the stereo system’s epipolar geometry, the epipolar constraint allows us to reject feature points that could otherwise lead to incorrect correspondences.

In our system, the images are acquired by the catadioptric system formed by the upper camera-mirror (PMU) and the lower camera-mirror (PML), as seen in Figure 3. The projections of a 3D point X_W over the mirrors PMU and PML are denoted by x_{PMU}^i and x_{PML}^i respectively. The coplanarity of the vectors $x_{PMU}^i, x_{PML}^i, T_{PMU \rightarrow PML} = [tx, ty, tz]^T$ is expressed by Equation (4)

$$(x_{PML}^i)^T R(T_{PMU \rightarrow PML} \times x_{PMU}^i) = 0 \tag{4}$$

where \times corresponds to the cross product and R and T are the rotation matrix and translation vector between the upper and lower mirrors. The coplanarity restriction is simplified in Equation (5)

$$(x_{PML}^i)^T E x_{PMU}^i = 0 \tag{5}$$

where:

$$E = RS \tag{6}$$

E is the essential matrix. The skew symmetric matrix S is given by Equation (7).

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{7}$$

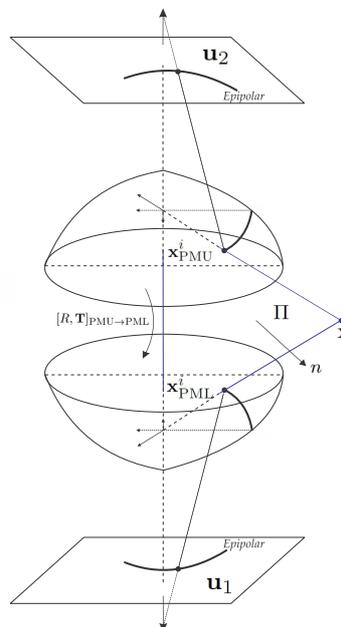


Figure 3. Epipolar geometry between two catadioptric cameras with parabolic mirrors. R is the rotation matrix and T is the translation vector, between the upper and lower mirrors (PMU→PML) expressed as a skew symmetric matrix. Π is the epipolar plane.

For every point \mathbf{u}_1 in the catadioptric image, an epipolar conic is uniquely assigned to the other image \mathbf{u}_2 , the search space between corresponding points is limited by Equation (8).

$$\mathbf{u}_2^T A_2(E, \mathbf{u}_1) \mathbf{u}_2 = 0 \quad (8)$$

In the general case, the matrix $A_2(E, \mathbf{u}_1)$ is a non-linear function of the essential matrix E , the points \mathbf{u}_1 , and the calibration parameters of the catadioptric camera.

The vector $\mathbf{T}_{\text{PMU} \rightarrow \text{PML}}$ defines the epipolar plane Π . Equation (9) describes the vector normal to the plane Π in the camera coordinates system.

$$n_1 = \mathbf{T}_{\text{PMU} \rightarrow \text{PML}} \times \mathbf{x}_{\text{PMU}}^i \quad (9)$$

The normal vector n_1 can be expressed in the second camera coordinate system using the essential matrix as follows:

$$\begin{aligned} n_2 &= Rn_1 \\ &= R(\mathbf{T}_{\text{PMU} \rightarrow \text{PML}} \times \mathbf{x}_{\text{PMU}}^i) \\ &= RS\mathbf{x}_{\text{PMU}}^i \\ &= E\mathbf{x}_{\text{PMU}}^i \end{aligned} \quad (10)$$

Then Equation (10) is represented as:

$$n_2 = [p, q, s]^T \quad (11)$$

The plane equation Π described in the second camera coordinate system is given by Equation (12).

$$px + qy + sz = 0 \quad (12)$$

The derivation of $A_2(E, \mathbf{u}_1)$ from Equation (8) based on a parabolic mirror, z is substituted into Equation (12) resulting in Equation (13) that allows the calculation of the epipolar conics for this type of mirror:

$$sx^2 + 2b_2px + sy^2 + 2b_2qy - sb_2^2 \quad (13)$$

where b_2 is the mirror parameter and p, q, s are the components of the normal vector n_2 computed in Equation (10).

3.3. Stereo Reconstruction

Once we have the system calibrated, the next step is to acquire images and reconstruct a scene using the following procedure (see Figure 4):

1. Capture one image of the environment from each camera of the calibrated system.
2. Transform the catadioptric images to panoramic images using Algorithm 1.
3. Extract the features and descriptors from the panoramic images using a feature point detector such as SIFT, SURF, KAZE, a corner detector such as Harris, or more advanced feature detectors such as DeepMatching [40], CPM [45], or SuperGlue [39]. Match the points between the upper and lower camera features as described in Section 3.3.1.
4. Filter the wrong matches using epipolar constraints as described in Section 3.2.
5. Map the matching points coordinates from panoramic back to catadioptric images coordinates, using Algorithm 2.
6. Transform the catadioptric image points to the corresponding mirror PMU and PML.
7. Obtain the 3D reconstruction by triangulating the mirror points points using Equation (15) describe in Section 3.3.2.

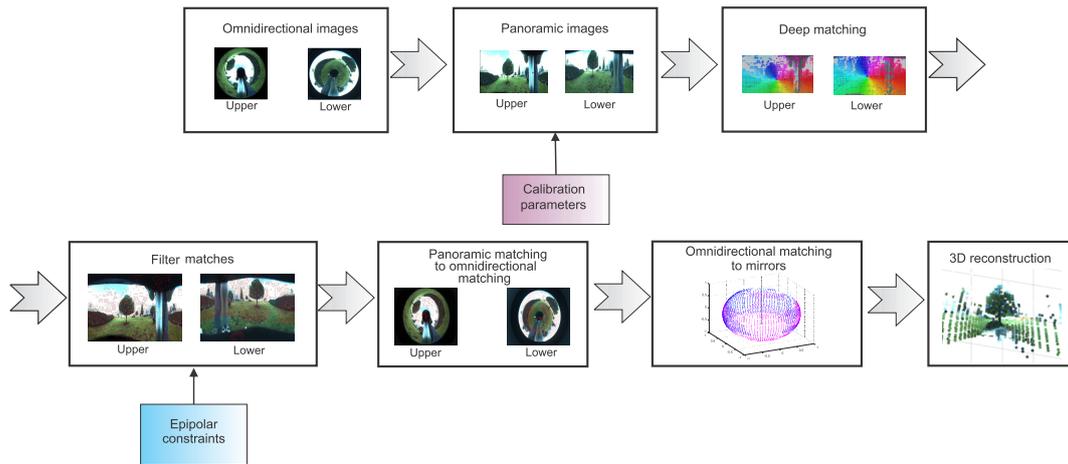


Figure 4. Panoramic 3D reconstruction (best seen in color). The procedure is the following: (1) Capture images from the catadioptric camera system, (2) unwrap the omnidirectional images to obtain panoramic images (3) to search for feature matches and (4) filter the matches using epipolar constraints. (5) Convert features back to catadioptric image coordinates and (6) to mirror coordinates. (7) Perform 3D reconstruction.

Algorithm 1: Transforming catadioptric images to panoramic images

Input : img_cat , catadioptric image
 x_c, y_c , optical center
Output: img_pan , panoramic image
 LUT_x , Look up table X-coordinates
 LUT_y , Look up table Y-coordinates

Procedure ;
 $h, w \leftarrow size(img_cat)$;
 $radius \leftarrow h/2$;
 $angle \leftarrow w/2$;
 $r \leftarrow 0$ to $radius$;
 $a \leftarrow 0$ to $2\pi - \frac{2\pi}{angle}$;
 $Y \leftarrow \sin(a)$;
 $X \leftarrow \cos(a)$;
 $LUT_x \leftarrow round(kron(r, X) + x_c)$;
 $LUT_y \leftarrow round(kron(r, Y) + y_c)$;
 $img_pan \leftarrow interpolate(img_cat, LUT_x, LUT_y)$;

Algorithm 2: Mapping between catadioptric to panoramic images

Input : pan_points , points in the panoramic image
 LUT_x , Look up table X-coordinates
 LUT_y , Look up table Y-coordinates
Output: $cata_points$, points in the catadioptric image

Procedure ;
 $cata_points \leftarrow []$;
 $num_points \leftarrow size(pan_points)$;
for $i \leftarrow 1$ to num_points **do**
 $Pt_{xy} \leftarrow round(pan_points(i))$;
 $x \leftarrow LUT_x[Pt_{xy}]$;
 $y \leftarrow LUT_y[Pt_{xy}]$;
end
 $cata_points \leftarrow concatenate(cata_points, [x, y])$;

3.3.1. Feature Detection and Matching

A feature point in an image is a small patch with enough information to be found in another image. Feature points can be corners, which are image patches with significant changes in both directions and are rotation-invariant [46,47]. They can be keypoints that encode texture information in vectors called descriptors, which besides being rotation-invariant, are also scale-invariant [48–52]. Additionally, they can be Deep Learning-based features that can also be invariant to repetitive textures and non-rigid deformations [39,40].

Feature matching entails finding the same feature point in both images, for this, we compared corner detectors such as Harris [46] and Shi-Tomasi [47] along with feature point detectors such SIFT [48], SURF [49], BRISK [50], FAST [51], KAZE [52], a state-of-the-art Deep Learning method called SuperGlue [39] and DeepMatching [40]. In the case of a corner detector, we use epipolar geometry described in Section 3.2 to reduce the search along the epipolar curve and match the points using image-patch correlation.

For the case of a feature point detector, we compute the pairwise distance between the descriptor vectors. Two descriptor vectors match when the distance between them is less than a defined threshold t ; $t = 10.0$ for BRISK and $t = 1.0$ for SIFT, SURF, KAZE, and FAST. For the case of SuperGlue, we used the outdoor trained model provided by the authors for the outdoor scene and the indoor trained model for the indoor scene. For DeepMatching, after finding the matches using the original implementation, we filter out the wrong matches using epipolar constraints. For this, we measure the Euclidean distance to the epipolar curve and keep the corresponding point in the other image if it lies d pixels from the epipolar. Figure 5c shows a selected feature as a red mark in the upper panoramic image, and Figure 5d shows the matched point in the lower panoramic image as a red mark along with the epipolar line through which the point should lie.

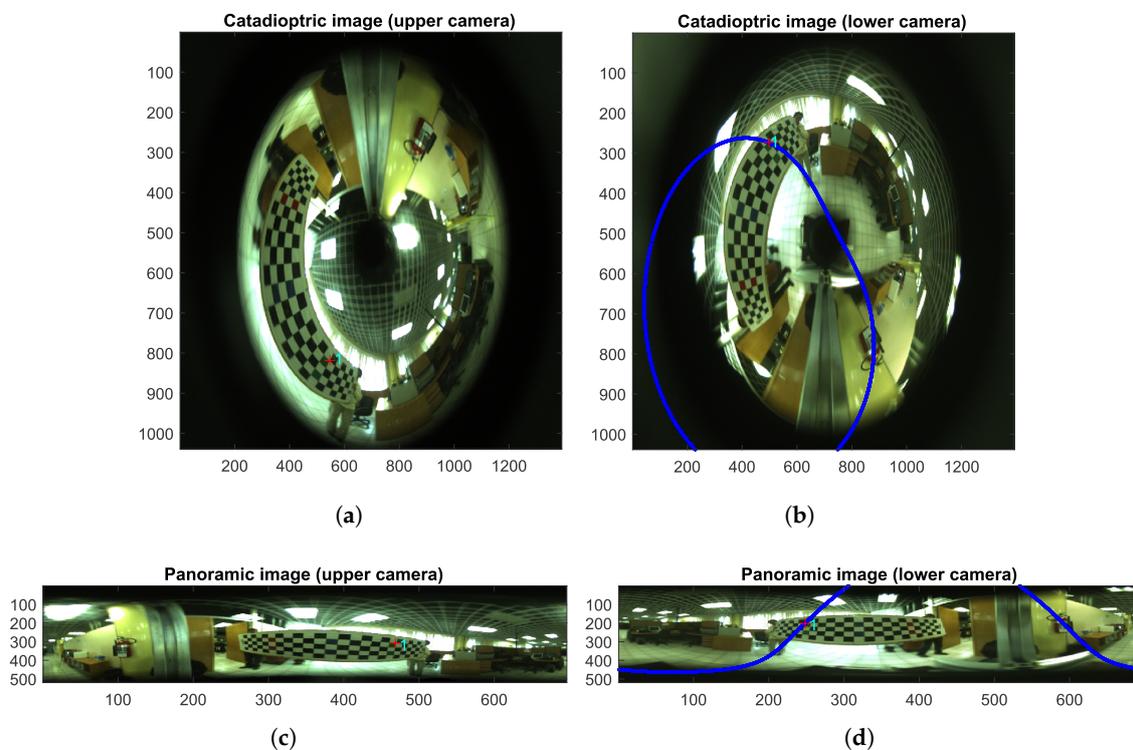


Figure 5. Feature point and epipolar curve (best seen in color). (a) Feature point (red mark) in the upper catadioptric image, (b) Feature point (red mark) and epipolar curve in the lower catadioptric image, (c) Feature point (red mark) in the upper panoramic image, (d) Feature point (red mark) and epipolar curve in the lower panoramic image.

3.3.2. 3D Reconstruction

Once we have the feature points between the two catadioptric cameras, we convert those coordinates to mirror coordinates to perform the 3D reconstruction.

Given point pairs \mathbf{x}_{PMU}^i and \mathbf{x}_{PML}^i on the mirrors surfaces and the coordinates $[x, y, z]$ of a point \mathbf{x}_{PMU}^i , we obtained Equation (14) [41,43].

$$\begin{aligned}
 \phi_1 &= \arctan\left(\frac{y}{x}\right) \\
 \theta_1 &= \arctan\left(\frac{\sqrt{x^2 + y^2}}{z}\right) \\
 \omega_1 &= \arccos\left(\frac{(\mathbf{x}_{PMU}^i)(-R_{PMU \rightarrow PML} \mathbf{T}_{PMU \rightarrow PML})}{|\mathbf{x}_{PMU}^i| |\mathbf{T}_{PMU \rightarrow PML}|}\right) \\
 \omega_2 &= \arccos\left(\frac{(\mathbf{x}_{PML}^i) \mathbf{T}_{PMU \rightarrow PML}}{|\mathbf{x}_{PML}^i| |\mathbf{T}_{PMU \rightarrow PML}|}\right)
 \end{aligned} \tag{14}$$

From the distance $|D|$ between two points \mathbf{x}_{PMU}^i and \mathbf{x}_{PML}^i , we determined the coordinates of a point \mathbf{X} using Equation (15).

$$\begin{aligned}
 |D| &= \frac{|\mathbf{T}_{PMU \rightarrow PML}| \sin \omega_2}{\sin(\omega_1 + \omega_2)} \\
 \mathbf{X} &= \begin{bmatrix} |D| \sin \theta_1 \cos \phi_1 \\ |D| \sin \theta_1 \sin \phi_1 \\ |D| \cos \theta_1 \end{bmatrix}
 \end{aligned} \tag{15}$$

4. Results

This section presents the results for the system calibration, catadioptric epipolar geometry, feature matching, and 3D reconstruction.

4.1. Calibration Results

Table 1 shows the intrinsic parameters of the upper and lower cameras, respectively. As described in Section 3.1, a_0, a_1, a_2, a_3, a_4 are the coefficients of a fourth degree polynomial that defines the curvature of the parabolic mirror representing the intrinsic parameters of the mirror. x_c and y_c define the optical center of the catadioptric cameras with a resolution of 1392×1038 .

Table 1. Intrinsic parameters of upper and lower camera.

	Polynomial Coefficients $f(\rho)$					Optical Center	
	a_0	a_1	a_2	a_3	a_4	x_c	y_c
Upper camera	-347.7309	0	0.0056	0	0	696.3600	474.1928
Lower camera	-162.3562	0	0.0023	0	0	698.3097	524.4199

We obtain the extrinsic parameters of the upper and lower catadioptric cameras with respect to each calibration plane (Figure 2a). That is, six extrinsic parameters (rotation and translation) for each of the 32 images acquired during calibration. The calibration errors of each catadioptric camera are shown in Table 2 which are inside the ranges of typical panoramic cameras [53].

Table 2. Calibration error on the catadioptric camera.

	Error [Pixels]
Upper camera	0.86
Lower camera	0.52

4.2. Epipolar Geometry Results

In this section, we show how the epipolar curves can be used to filter wrong matching points. Figure 5a shows a feature point (red mark) on the upper catadioptric image, and Figure 5b shows the corresponding point on the lower catadioptric image along with the epipolar curve shown in blue. Figures 5c,d show the same information but for the unwrapped panoramic images. For a corresponding point in the lower camera to be correct, it must lie along the epipolar curve, as shown in the images.

4.3. Features Matching Results

We compared the feature detection and matching using multiple methods. Table 3 shows the number of feature matches found with each method in descending order as well as the running time. We use MATLAB 2019b built-in functions for all the methods except SIFT, DeepMatching, and SuperGlue. For SIFT, we use the Machine Vision Toolbox from [54], and for DeepMatching and SuperGlue we use the original implementations [55,56]. For these last two methods, we report the inference time on a CPU Intel Xeon ES-1630 at 3.7 GHz and on a GPU GTX1080. Although these two methods are slow on CPU due to the computationally expensive nature of deep neural networks, the parallel GPU implementation can achieve running times comparable to the other CPU-based methods.

Figures 6 and 7 show the feature matches for the first six methods shown in Table 3. From these images, we can see that DeepMatching has a significantly broader coverage and density at the expense of higher computational cost. The second best feature detector in terms of the number of features was the Harris corner detector; however, we can see that the features cover only the trees' borders but not the entire image. KAZE and SuperGlue matches have more image coverage than Harris, but the features are sparse compared to DeepMatching.

Table 3. Number of matches and running times obtained with each method. For the Deep learning-based methods, we also report the running times obtained in a GTX 1080.

Method	Number of Matches	Time (Msec) CPU/GPU
DeepMatching ($d = 20$)	1114	3400/210
Harris corner detector	845	100
KAZE	463	125
SuperGlue	200	900/70
SURF	105	150
SIFT	76	350
MIN-EIGEN	62	120
BRISK	52	110
FAST	36	80

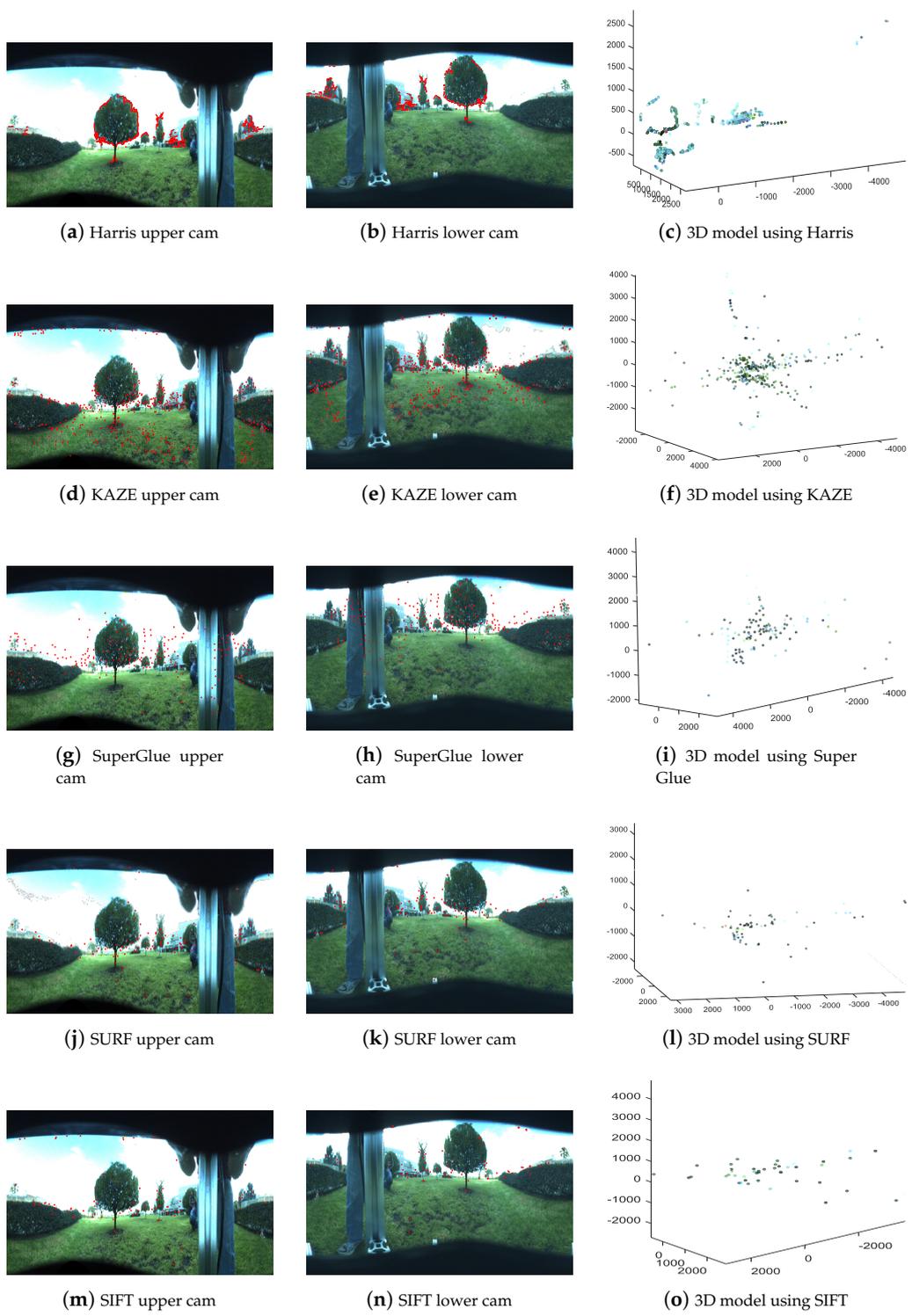


Figure 6. Feature matching points on the upper and lower cameras and 3D reconstruction for each comparing method.

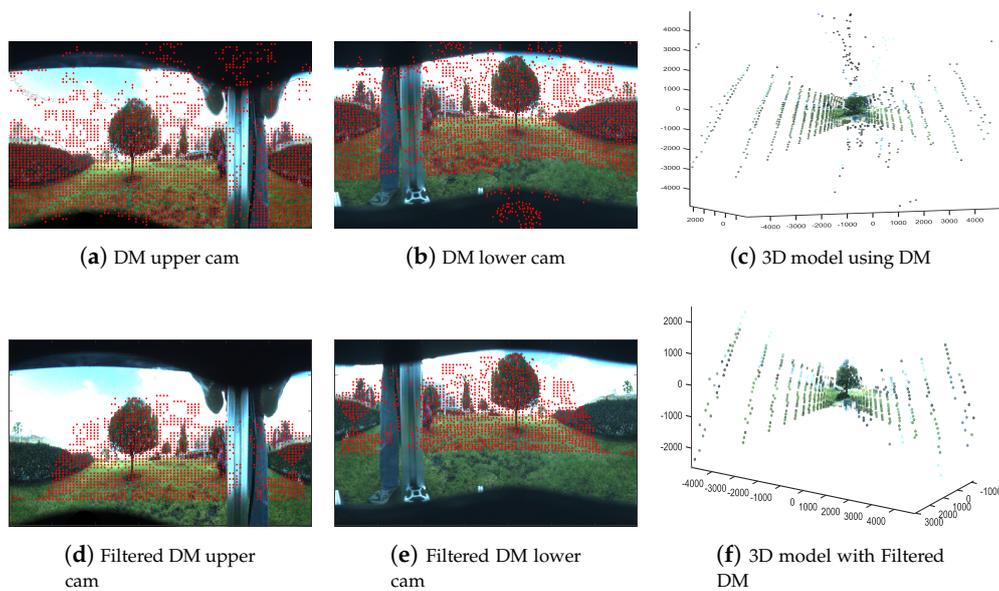


Figure 7. DeepMatching features and reconstruction. (a,b) show the original DeepMatching on the upper and lower camera, respectively. (c) shows the 3D reconstruction using the original DeepMatching. (d,e) show the filtered DeepMatching using epipolar constraints with $d = 30$ on the upper and lower cameras. (f) shows the 3D reconstruction using filtered DeepMatching.

As described in Section 3.3.1, we filtered the DeepMatching results using epipolar constraints by keeping the matches whose distance from the epipolar curve is less than a defined threshold d . Figure 8 shows the number of DeepMatching features obtained with different filtering levels. The more we increase d , the more matches we get, but also, the more error we are allowing in the reconstruction. Empirically, we found that a value of d between 20 and 30 pixels from the epipolar curve gives the best compromise between features and features' quantity and quality. The effects of d in the reconstruction are described in the next section.

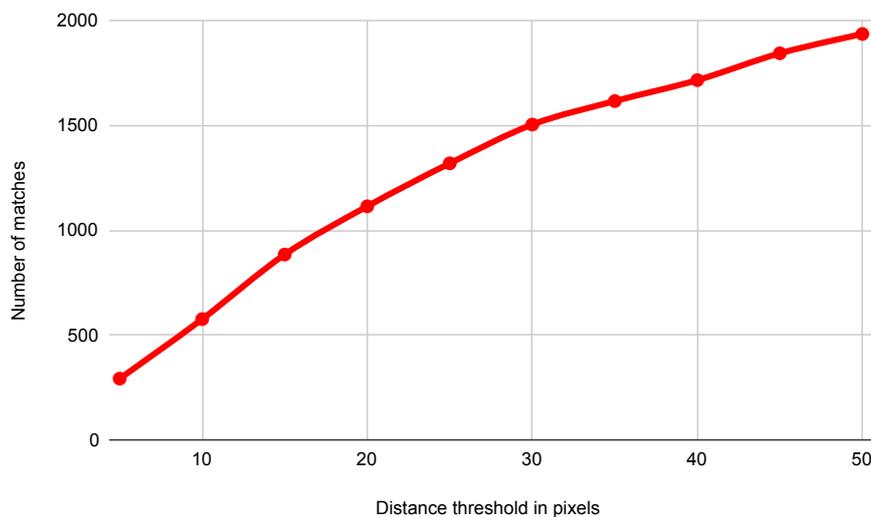


Figure 8. Amount of DeepMatching features obtained with different levels of filtering. The more we increase the distance threshold, the more matches we get, but also, the more error we are allowing in the reconstruction.

4.4. 3D Reconstruction Results

Once we have the matching points in both catadioptric cameras, the next step is to transform those points into the mirrors. Figure 9a shows the Harris corners (which was the runner-up method in terms of the number of features) on the upper parabolic mirror (PMU) and the lower panoramic mirror (PML). Similarly, Figure 9b shows the DeepMatching points on each of the mirrors. As described in Section 4.3, DeepMatching provides broader coverage and density than all the other methods (see Figures 6 and 7) shown in Table 3 resulting in a more well-rounded reconstruction.

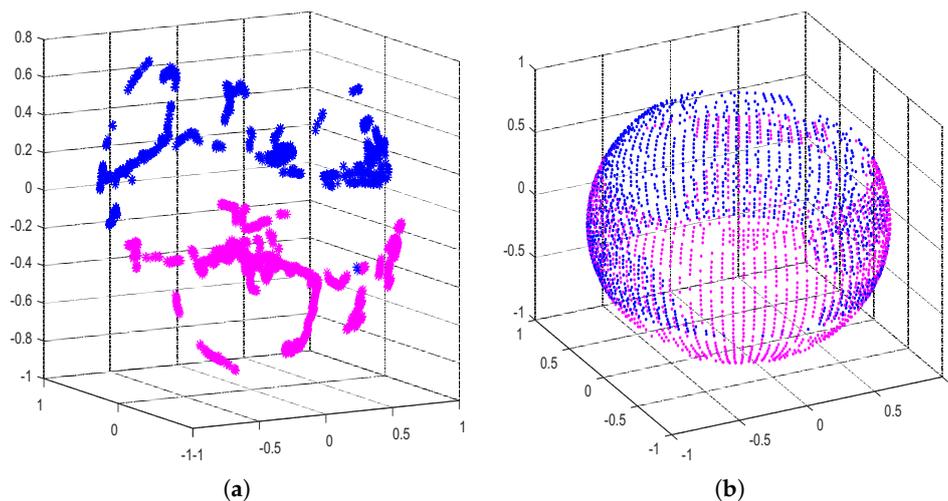


Figure 9. Matching points in the panoramic mirrors (best seen in color). Pink points denote the features in the panoramic upper mirror PMU, and blue points are the features on the panoramic lower mirror PML. (a) Shows Harris corners on the mirrors, and (b) shows the DeepMatching points on the mirrors.

The third columns of Figures 6 and 7 show the reconstruction of an outdoor scenario with features obtained with the methods presented in Table 3. Figure 6a–c show the Harris corners matches and the reconstruction, Figure 6d–f using KAZE, Figure 6g–i applying SuperGlue, Figure 6j–l employing SURF, Figure 6m–o with SIFT, and Figure 7 with DeepMatching.

For this challenging outdoor environment, we see that the reconstructions generated with the feature matches from all these methods are poor due to the lack of feature density and coverage except for the DeepMatching method, where the features can cover most of the image as shown in Figure 7.

Although DeepMatching returns many feature matches, not all of them are correct. To fix this, we use epipolar constraints as described in Section 3.2. Figure 7a,b show the unfiltered DeepMatching results along with the 3D reconstruction in Figure 7c. Figure 7d,e show the DeepMatching results filtered with $d = 30$ pixels. Figure 7f shows the reconstruction with these filtered features points.

Figure 10 shows the point clouds reconstructed using the Deep features matches with the different filtering levels. As the images show, when we relax the filter (when d is larger), we get more features but also more errors in the background.

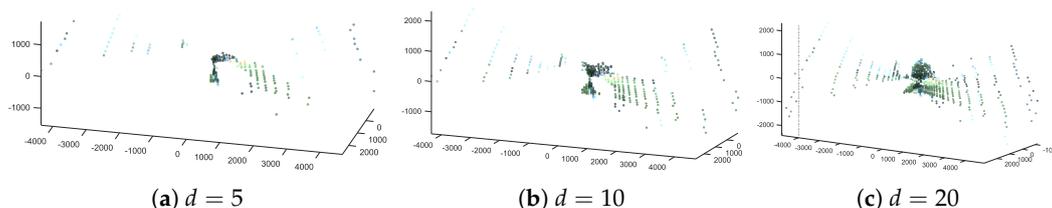


Figure 10. 3D reconstruction using DeepMatching. Figures (a–c) show the reconstruction with features filtered at $d = 5$, $d = 10$, and $d = 20$ respectively.

To quantify the reconstruction error, we reconstruct an object with known dimensions, in this case a rectangular pattern of size 50 cm × 230 cm. Figure 11 shows the results obtained with DeepMatching. Figures 11a,b show the unfiltered DeepMatching features and Figure 11c shows the 3D reconstruction. Figures 11d,e show the filtered matches with $d = 15$ and Figure 11f shows the 3D reconstruction. From Figure 11 we see that the filtering of the features using epipolar constraints produces a cleaner reconstruction without compromising the density.

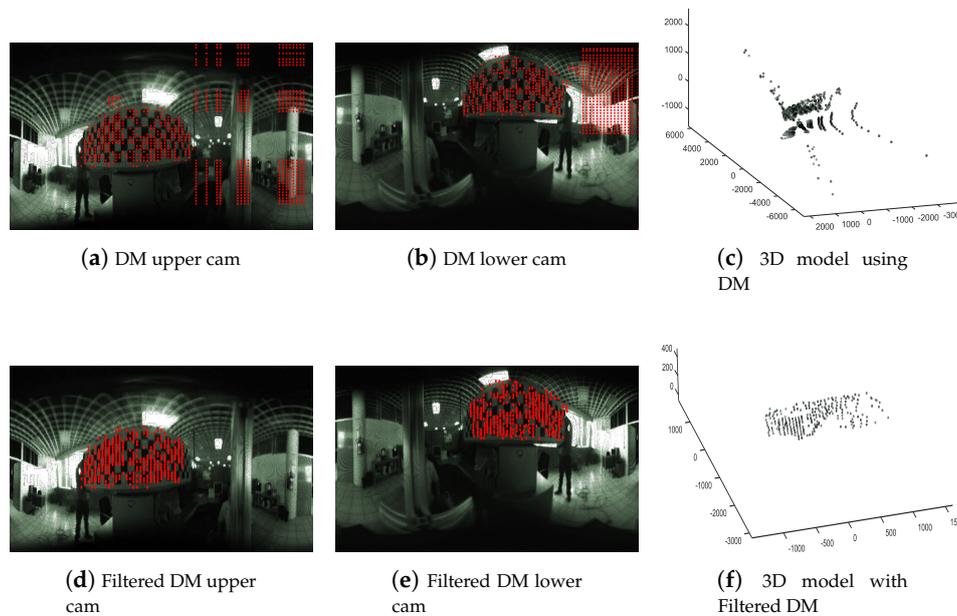


Figure 11. DeepMatching and filtered DeepMatching for 3D reconstruction of a known pattern.

Using the rectangular pattern with known dimensions, we calculate the reconstruction’s error at the center of the pattern and the extremes. At the right extreme, the error correlates with a larger distortion on the periphery of the mirror. Table 4 shows the mean reconstruction error in millimeters and the standard deviation.

Table 4. Reconstruction error in millimeters.

	Mean Error (mm)	Standard Deviation (mm)
Pattern right	20.59	2.23
Pattern center	8.74	4.89
Pattern left	9.57	5.56

To evaluate the qualitative results of the 3D reconstructions, we reconstructed three more objects shown in Figure 12. A squared box, a hat, and a clay pot. For the squared box shown in Figure 12a, we computed the angles between normal vectors of adjacent planes compared with 90°. The results are shown in Table 5. In this table, we compared the angle errors with [4] achieving slightly better results with one stereo pair.

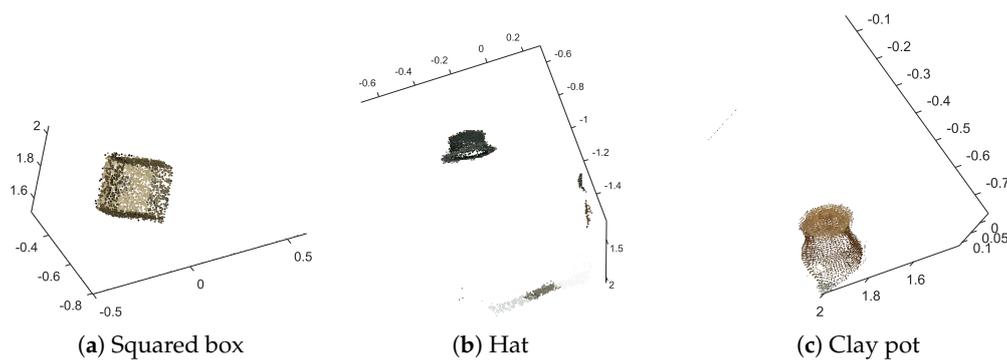


Figure 12. Qualitative reconstructions results using filtered DeepMatching. (a) Shows the reconstruction of a squared box, (b) shows the reconstruction of a hat, and (c) shows the reconstruction of a clay pot.

Table 5. Angle errors of the squared box.

Method	Angle Errors ($^{\circ}$)			
	θ_1	θ_2	θ_3	$\bar{\theta}$
With one stereo pair [4]	6.52	9.88	8.32	8.24
Fusion with two stereo pairs [4]	4.55	9.57	6.98	7.03
Proposed (one stereo pair)	6.23	8.17	7.10	7.16

5. Conclusions

We introduced the development of a stereo catadioptric 3D reconstruction system capable of generating semi-dense reconstructions based on epipolar constrained DeepMatching. The proposed method generates accurate 3D reconstructions for indoor and outdoor environments with significantly more matches than sparse methods, producing broader and denser 3D reconstructions and gracefully removing incorrect correspondences provided by the DeepMatching algorithm. Our system's current limitations in terms of hardware are its large size and fragility, making it unsuitable for real-life situations. In terms of the method, although DeepMatching provides significantly more feature points than corner or feature-point detectors, it is still relatively sparse compared to dense 3D reconstruction deep learning techniques at the expense of faster and more accurate measurements. In future work, we plan to increase the reconstruction's density by combining the current approach with dense 3D reconstruction methods and improving the system's size and robustness.

Author Contributions: Conceptualization, D.M.C.-E. and J.T.; Methodology, D.M.C.-E.; Resources, D.M.C.-E., J.T., J.A.R.-G., A.R.-P.; Software, D.M.C.-E., J.T., J.A.R.-G., A.R.-P.; Writing—original draft, D.M.C.-E. and J.T.; Writing—review and editing, D.M.C.-E., J.T., J.A.R.-G., A.R.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors wish to acknowledge the support for this work by the Consejo Nacional de Ciencia y Tecnología (CONACYT) through Sistema Nacional de Investigadores (SNI). We also want to thank CICATA-IPN, Queretaro for providing the facilities and support during the development of this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jiang, W.; Okutomi, M.; Sugimoto, S. Panoramic 3D reconstruction using rotational stereo camera with simple epipolar constraints. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 1, pp. 371–378.
- Deng, X.; Wu, F.; Wu, Y.; Wan, C. Automatic spherical panorama generation with two fisheye images. In Proceedings of the 2008 7th World Congress on Intelligent Control and Automation, Chongqing, China, 25–27 June 2008; pp. 5955–5959.

3. Sagawa, R.; Kurita, N.; Echigo, T.; Yagi, Y. Compound catadioptric stereo sensor for omnidirectional object detection. In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566), Sendai, Japan, 28 September–2 October 2004; Volume 3, pp. 2612–2617.
4. Chen, S.; Xiang, Z.; Zou, N.; Chen, Y.; Qiao, C. Multi-stereo 3D reconstruction with a single-camera multi-mirror catadioptric system. *Meas. Sci. Technol.* **2019**, *31*, 015102. [[CrossRef](#)]
5. Fiala, M.; Basu, A. Panoramic stereo reconstruction using non-SVP optics. *Comput. Vis. Image Underst.* **2005**, *98*, 363–397. [[CrossRef](#)]
6. Jaramillo, C.; Valenti, R.G.; Xiao, J. GUMS: A generalized unified model for stereo omnidirectional vision (demonstrated via a folded catadioptric system). In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 2528–2533.
7. Lin, S.S.; Bajcsy, R. High resolution catadioptric omni-directional stereo sensor for robot vision. In Proceedings of the 2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422), Taipei, Taiwan, 14–19 September 2003; Volume 2, pp. 1694–1699.
8. Ragot, N.; Ertaud, J.; Savatier, X.; Mazari, B. Calibration of a panoramic stereovision sensor: Analytical vs. interpolation-based methods. In Proceedings of the Iecon 2006-32nd Annual Conference on IEEE Industrial Electronics, Paris, France, 6–10 November 2006; pp. 4130–4135.
9. Cabral, E.L.; De Souza, J.; Hunold, M.C. Omnidirectional stereo vision with a hyperbolic double lobed mirror. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 26 August 2004; Volume 1, pp. 1–9.
10. Jang, G.; Kim, S.; Kweon, I. Single-camera panoramic stereo system with single-viewpoint optics. *Opt. Lett.* **2006**, *31*, 41–43. [[CrossRef](#)]
11. Su, L.; Luo, C.; Zhu, F. Obtaining obstacle information by an omnidirectional stereo vision system. In Proceedings of the 2006 IEEE International Conference on Information Acquisition, Weihai, China, 20–23 August 2006; pp. 48–52.
12. Caron, G.; Marchand, E.; Mouaddib, E.M. 3D model based pose estimation for omnidirectional stereovision. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 5228–5233.
13. Yi, S.; Ahuja, N. An omnidirectional stereo vision system using a single camera. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; Volume 4, pp. 861–865.
14. Li, W.; Li, Y.F. Single-camera panoramic stereo imaging system with a fisheye lens and a convex mirror. *Opt. Express* **2011**, *19*, 5855–5867. [[CrossRef](#)] [[PubMed](#)]
15. Xu, J.; Wang, P.; Yao, Y.; Liu, S.; Zhang, G. 3D multi-directional sensor with pyramid mirror and structured light. *Opt. Lasers Eng.* **2017**, *93*, 156–163. [[CrossRef](#)]
16. Tan, K.H.; Hua, H.; Ahuja, N. Multiview panoramic cameras using mirror pyramids. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 941–946. [[CrossRef](#)]
17. Schönbein, M.; Kitt, B.; Lauer, M. Environmental Perception for Intelligent Vehicles Using Catadioptric Stereo Vision Systems. In Proceedings of the 5th European Conference on Mobile Robots (ECMR), Berlin, Germany, 1 January 2018; pp. 189–194.
18. Ehlgren, T.; Pajdla, T.; Ammon, D. Eliminating blind spots for assisted driving. *IEEE Trans. Intell. Transp. Syst.* **2008**, *9*, 657–665. [[CrossRef](#)]
19. Xu, J.; Gao, B.; Liu, C.; Wang, P.; Gao, S. An omnidirectional 3D sensor with line laser scanning. *Opt. Lasers Eng.* **2016**, *84*, 96–104. [[CrossRef](#)]
20. Lauer, M.; Schönbein, M.; Lange, S.; Welker, S. 3D-objecttracking with a mixed omnidirectional stereo camera system. *Mechatronics* **2011**, *21*, 390–398. [[CrossRef](#)]
21. Jaramillo, C.; Valenti, R.; Guo, L.; Xiao, J. Design and Analysis of a Single—Camera Omnistereo Sensor for Quadrotor Micro Aerial Vehicles (MAVs). *Sensors* **2016**, *16*, 217. [[CrossRef](#)]
22. Jaramillo, C.; Yang, L.; Muñoz, J.P.; Taguchi, Y.; Xiao, J. Visual odometry with a single-camera stereo omnidirectional system. *Mach. Vis. Appl.* **2019**, *30*, 1145–1155. [[CrossRef](#)]
23. Almaraz-Cabral, C.C.; Gonzalez-Barbosa, J.J.; Villa, J.; Hurtado-Ramos, J.B.; Ornelas-Rodriguez, F.J.; Córdova-Esparza, D.M. Fringe projection profilometry for panoramic 3D reconstruction. *Opt. Lasers Eng.* **2016**, *78*, 106–112. [[CrossRef](#)]

24. Flores, V.; Casaletto, L.; Genovese, K.; Martinez, A.; Montes, A.; Rayas, J. A panoramic fringe projection system. *Opt. Lasers Eng.* **2014**, *58*, 80–84. [[CrossRef](#)]
25. Kerkaou, Z.; Alioua, N.; El Ansari, M.; Masmoudi, L. A new dense omnidirectional stereo matching approach. In Proceedings of the 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2–4 April 2018; pp. 1–8.
26. Ma, C.; Shi, L.; Huang, H.; Yan, M. 3D reconstruction from full-view fisheye camera. *arXiv* **2015**, arXiv:1506.06273.
27. Song, M.; Watanabe, H.; Hara, J. Robust 3D reconstruction with omni-directional camera based on structure from motion. In Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 7–9 January 2018; pp. 1–4.
28. Corke, P. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB® Second, Completely Revised*; Springer: New York, NY, USA, 2017; Volume 118.
29. Boutteau, R.; Savatier, X.; Ertaud, J.Y.; Mazari, B. An omnidirectional stereoscopic system for mobile robot navigation. In Proceedings of the 2008 International Workshop on Robotic and Sensors Environments, Ottawa, ON, Canada, 17–18 October 2008; pp. 138–143.
30. Zhou, F.; Chai, X.; Chen, X.; Song, Y. Omnidirectional stereo vision sensor based on single camera and catoptric system. *Appl. Opt.* **2016**, *55*, 6813–6820. [[CrossRef](#)]
31. Jang, G.; Kim, S.; Kweon, I. Single camera catadioptric stereo system. In Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China, 1 January 2006.
32. Ragot, N.; Rossi, R.; Savatier, X.; Ertaud, J.; Mazari, B. 3D volumetric reconstruction with a catadioptric stereovision sensor. In Proceedings of the 2008 IEEE International Symposium on Industrial Electronics, Cambridge, UK, 30 June–2 July 2008; pp. 1306–1311.
33. Ricci, E.; Ouyang, W.; Wang, X.; Sebe, N. Monocular depth estimation using multi-scale continuous CRFs as sequential deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1426–1440.
34. Xu, D.; Ouyang, W.; Wang, X.; Sebe, N. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 675–684.
35. Fu, H.; Gong, M.; Wang, C.; Batmanghelich, K.; Tao, D. Deep ordinal regression network for monocular depth estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2002–2011.
36. Won, C.; Ryu, J.; Lim, J. Omnimvs: End-to-end learning for omnidirectional stereo matching. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–3 November 2019; pp. 8987–8996.
37. Ma, W.C.; Wang, S.; Hu, R.; Xiong, Y.; Urtasun, R. Deep rigid instance scene flow. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–21 June 2019; pp. 3614–3622.
38. Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; Brox, T. FlowNet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2758–2766.
39. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–18 June 2020.
40. Revaud, J.; Weinzaepfel, P.; Harchaoui, Z.; Schmid, C. Deepmatching: Hierarchical deformable dense matching. *Int. J. Comput. Vis.* **2016**, *120*, 300–323. [[CrossRef](#)]
41. Córdova-Esparza, D.M.; Gonzalez-Barbosa, J.J.; Hurtado-Ramos, J.B.; Ornelas-Rodriguez, F.J. A panoramic 3D reconstruction system based on the projection of patterns. *Int. J. Adv. Robot. Syst.* **2014**, *11*, 55. [[CrossRef](#)]
42. Scaramuzza, D.; Martinelli, A.; Siegwart, R. A toolbox for easily calibrating omnidirectional cameras. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 5695–5701.
43. Gonzalez-Barbosa, J.J.; Lacroix, S. Fast dense panoramic stereovision. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 1210–1215.
44. Svoboda, T.; Pajdla, T. Epipolar geometry for central catadioptric cameras. *Int. J. Comput. Vis.* **2002**, *49*, 23–37. [[CrossRef](#)]

45. Hu, Y.; Song, R.; Li, Y. Efficient coarse-to-fine patchmatch for large displacement optical flow. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5704–5712.
46. Harris, C.G.; Stephens, M. A combined corner and edge detector. In Proceedings of the Fourth Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.
47. Shi, J. Good features to track. In Proceedings of the 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
48. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
49. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision*; Springer: New York, NY, USA, 2006; pp. 404–417.
50. Leutenegger, S.; Chli, M.; Siegwart, R. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 IEEE international conference on computer vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
51. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*; Springer: New York, NY, USA, 2006; pp. 430–443.
52. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE features. In *European Conference on Computer Vision*; Springer: New York, NY, USA, 2012; pp. 214–227.
53. Schönbein, M. *Omnidirectional Stereo Vision for Autonomous Vehicles*; KIT Scientific Publishing: Karlsruhe, Germany, 2015; Volume 32.
54. Corke, P. The Machine Vision Toolbox: A MATLAB toolbox for vision and vision-based control, Omnidirectional Stereo Vision for Autonomous Vehicles. *IEEE Robot. Autom. Mag.* **2005**, *12*, 16–25. [[CrossRef](#)]
55. MagicLeap. SuperGlue Inference and Evaluation Demo Script. Available online: <https://github.com/magicleap/SuperGluePretrainedNetwork> (accessed on 9 July 2020).
56. Revaud, J. DeepMatching: Deep Convolutional Matching. Available online: <https://thoth.inrialpes.fr/src/deepmatching/> (accessed on 12 April 2020).

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).