## Article

# Automatic Rectification of the Hybrid Stereo Vision System 

Chengtao Cai, Bing Fan *, Xin Liang and Qidan Zhu<br>Department of Automation, 4139, Building No. 61, Harbin Engineering University, Nan Tong No. 145, Harbin 150001, China; caichengtao@hrbeu.edu.cn (C.C.); liang941219@hrbeu.edu.cn (X.L.); zhuqidan@hrbeu.edu.cn (Q.Z.)<br>* Correspondence: fanbing@hrbeu.edu.cn

Received: 18 August 2018; Accepted: 30 September 2018; Published: 8 October 2018


#### Abstract

By combining the advantages of 360-degree field of view cameras and the high resolution of conventional cameras, the hybrid stereo vision system could be widely used in surveillance. As the relative position of the two cameras is not constant over time, its automatic rectification is highly desirable when adopting a hybrid stereo vision system for practical use. In this work, we provide a method for rectifying the dynamic hybrid stereo vision system automatically. A perspective projection model is proposed to reduce the computation complexity of the hybrid stereoscopic 3D reconstruction. The rectification transformation is calculated by solving a nonlinear constrained optimization problem for a given set of corresponding point pairs. The experimental results demonstrate the accuracy and effectiveness of the proposed method.


Keywords: hybrid stereo vision; projection model; automatic rectification

## 1. Introduction

Stereo vision systems have been widely used in tasks such as surveillance [1], search [2], exploration [3], autonomous navigation [4], mapping [5], and obstacle detection [6] for their ability to recover 3D information of real-world scenes. Additional advantages can be derived using omnidirectional cameras in stereo vision systems [7]. These cameras allow the horizontal field of view to be increased to 360 degrees. Although imaging models and camera calibration methods for different types of catadioptric mirrors have been studied [8,9], these cameras have very limited resolution and are unable to provide close observation of particular targets. To improve the applicability of omnidirectional cameras, camera networks consisting of catadioptric and perspective sensing devices [10-12] have been proposed. The combination of these two cameras helps to continuously monitor the entire surveillance area while making detailed observations of specific targets. The hybrid stereo vision system combines the advantage of the 360-degree field of view cameras with high-resolution imaging from the conventional cameras, but it also poses challenges for epipolar geometry and stereo rectification. Omnidirectional images cannot be processed by the pinhole imaging model of conventional images [13], meaning that there is no unified imaging model for omnidirectional and conventional images. Consequently, the geometry of the traditional stereo vision system [14-16] cannot be applied to the hybrid stereo vision system.

Generally, there are two types of methods to solve the aforementioned problem. One is to perform external calibration using the specific relation between the camera pairs or a large amount of known corresponding points. Under a local planar assumption, a non-linear approach for registering images in a hybrid vision system without requiring the calibration of cameras is proposed in [17]. When the position of the omnidirectional camera and the conventional camera are fixed, a checkerboard pattern with two different colors is used to acquire the geometric relation of the stereo camera system
with heterogeneous lenses [18]. A geometric relation between the omnidirectional camera and the conventional camera is derived using manually obtained corresponding points [19]. Several calibration image pairs are obtained under known spatial positions. The extrinsic parameters are extracted via re-projecting known corresponding 2D points into a 3D space [20]. Three different models have been proposed to obtain fundamental matrices for hybrid vision [21]. Although these methods can achieve accurate results, they are limited by the high computational complexity or the priors of the relative position of the hybrid cameras.

Another method is to use pan-tilt-zoom (PTZ) cameras instead of conventional ones in the hybrid vision system, where it is not necessary to calculate the extrinsic parameters explicitly. Spatial mapping can be easily computed between an omnidirectional camera and a PTZ camera. Specifically, the pan-tilt angle of the PTZ camera is acquired by the corresponding points in the omnidirectional camera when the system is operating. This approach assumes that the pan-tilt angles are extremely correlated with the corresponding points. In some studies [12,22], 3D reconstruction is based on data collection and neural network fitting, but in other studies, the assumption of geometry constraints is made. For instance, in [23-25], it is assumed that the omnidirectional camera and PTZ camera are coaxial, or even that they share a common origin [26]. However, these assumptions are violated when the optical axes of both omnidirectional and PTZ cameras do not coincide. In addition, the corresponding points may not be in the same 2D plane without calibration. These problems will significantly affect the accuracy of 3D reconstruction.

In practical applications, for example, a hybrid vision system can be used for target tracking and observation tasks. The conventional camera will move as the position of the target changes to ensure that the target is within the common field of view of the omnidirectional camera and conventional camera. When the hybrid vision system is used in surveillance scenarios, the conventional camera will move as the monitored area changes. Therefore, automatic stereo rectification is highly desirable, while the accuracy is also needed.

In our study, an automatic stereo rectification approach for one omnidirectional camera and one conventional camera is proposed. Compared with state-of-the-art methods [20,27], the main contributions of the proposed approach are as follows:

- A perspective projection model is proposed for the omnidirectional image, which significantly reduces the computational complexity of 3D formulation for mixed-view pairs.
- A method based on a novel, well-defined cost function for optimizing the normalization matrix is employed, which can calculate the rectification transformation more accurately.
- To evaluate the performance of the proposed automatic rectification method and to provide a direct application, a target tracking and odometry hybrid vision system is established based on an automatic rectification approach.

The remainder of this paper is organized as follows: The proposed automatic rectification approach for hybrid stereo vision system is presented in Section 2 . Section 3 describes the configuration of the hybrid imaging system. The methodology, including the novel perspective projection model for the omnidirectional image and the method for optimizing the normalization matrix, is described in Section 4. In Section 5, the experimental results are presented, and a direct application of our rectification method is also provided. In Section 6, we discuss our results, limitations and future work. The conclusion is given in Section 7.

## 2. Proposed Automatic Rectification Approach

The block diagram of automatic stereo rectification for hybrid vision is shown in Figure 1. The proposed approach consists of three parts, including acquisition of the virtual perspective image in Figure 1a, calculation of the fundamental matrix in Figure 1b, and stereo rectification in Figure 1c.

Figure 1a shows how we calculated the direction angle $\alpha$ of the region of interest in the omnidirectional camera, after which the region of interest was projected onto a virtual plane using the proposed projection model.

In Figure 1b, the conventional camera is shown to rotate $\alpha$ degrees to ensure that the common field of view of the two cameras is the region of interest. Affine scale invariant feature transform (ASIFT) [28] was used to extract and match the features in the virtual perspective image and conventional image, after which the proposed optimization method for normalizing corresponding points was applied. Finally, the fundamental matrix was calculated by means of the 8-point algorithm [29].

Figure 1c shows how the fundamental matrix was decomposed into rotation and translation matrices of the virtual perspective image and conventional image. Finally, aligned image pairs could be obtained by remapping.

The main processing modules will be described in detail in the following sections.


Figure 1. Block diagram of the proposed automatic rectification approach. (a) Acquisition of the virtual perspective image. (b) Calculation of the fundamental matrix. (c) Stereo rectification.

## 3. Hybrid Omnidirectional and Conventional Imaging System

The configuration of the hybrid vision system in this paper is shown in Figure 2. It illustrates the point correspondence relation between an omnidirectional image and a conventional image. The hyperbolical mirror was chosen for the omnidirectional camera to ensure that it had a single effective viewpoint, which is a necessary condition for the generation of pure perspective images from the captured images. We placed the omnidirectional camera vertically with the conventional one. The vertical installation not only avoids the own occlusion of the system, but also has a large common field of view. There are three coordinates in the hybrid vision configuration-conventional camera coordinate $X_{p} Y_{p} Z_{p}$, catadioptric coordinate $X_{m} Y_{m} Z_{m}$, and omnidirectional coordinate $X_{c} Y_{c} Z_{c}$. The projection center of the two cameras is $O_{p}$, and $O_{c}$. $d$ is the distance between the catadioptric coordinate $X_{m} Y_{m} Z_{m}$ and projection center $O_{c} . d=2 \sqrt{a^{2}+b^{2}}$. $a$ and $b$ are the long and short axes of the hyperbolic mirror, respectively. The 3D point $M$ was projected to point to $m$ of the conventional image plane by linear mapping. It also projected to point $m^{\prime}$ of the omnidirectional image through the incident and reflected rays, which is nonlinear mapping. If the extrinsic parameters of the two cameras are known, the 3D point $M$ can be determined uniquely by $m$ and $m^{\prime}$.


Figure 2. The configuration of the hybrid vision system. It consists of a perspective camera and a catadioptric camera with a hyperboloidal mirror.

## 4. Methodology

### 4.1. Virtual Image Generation

To avoid the complex geometric relationship between the omnidirectional image and the conventional image, a novel perspective projection model for the omnidirectional image is proposed in this section. Unlike the conventional image, the generation of a virtual perspective image from an omnidirectional image is not one-to-one linear mapping-in this case, a simple perspective projection model is desirable.

As described in [30], a central catadioptric projection is equivalent to two-step mapping via the unit sphere. As shown in Figure 3, $O_{m}$ is the origin of the catadioptric coordinate and $O_{c}$ is the origin of the camera coordinate. For a general omnidirectional camera, the optical axis was aligned to the line defined by $O_{m}$ and $O_{c}$. Point $X$ in the 3D coordinate was projected onto a unit sphere located at the origin of the catadioptric coordinate, $O_{m}$. In the coordinate of the sphere, $X_{s}=\left[X_{s}, Y_{s}, Z_{s}, 1\right]^{T}$. Then, the projection of $X$ on the normalized plane could be given by:

$$
\begin{equation*}
x_{m}=\left(X_{s}, Y_{s}, Z_{s}+\xi\right) \tag{1}
\end{equation*}
$$

where $\xi \in[0,1]$ is the distance between $O_{c}$ and $O_{m}$. Therefore, the corresponding point in the omnidirectional image plane could be obtained by:

$$
\begin{equation*}
m=K_{c} x_{m} \tag{2}
\end{equation*}
$$

where $K_{c}$ is the intrinsic parameter of an omnidirectional camera. According to this, we were able to perform a back projection from the omnidirectional image and then reproject it onto a virtual plane.


Figure 3. The unit sphere model for the catadioptric camera.
As shown in Figure 4a, the view angle $\left(\alpha, \beta, \varphi_{v}, \varphi_{h}\right)$ of the virtual image was selected, where $\alpha$ is the horizontal azimuth. As shown in Figure 4b, where the coordinate of a pixel is $m=[u, v], \alpha$ is:

$$
\begin{equation*}
\alpha=\operatorname{arcos} \frac{u}{\sqrt{u^{2}+v^{2}}}=\operatorname{arsin} \frac{v}{\sqrt{u^{2}+v^{2}}} \tag{3}
\end{equation*}
$$

$\beta$ is the vertical angle of the conventional camera. In our system, we set $\beta$ as $90^{\circ}$. Thus, the optical axis of the virtual image was perpendicular to the baseline defined by the omnidirectional and conventional cameras. In order to acquire a homogenous image pair, the resolution of the virtual perspective image $w \times h$ and the focal length $f$ were set to be the same as parameters of the conventional image. $\varphi_{h}$ and $\varphi_{v}$ stand for the horizontal and vertical field of view, respectively, and can be calculated by:

$$
\begin{align*}
\varphi_{h} & =\arctan \frac{w}{2 f}  \tag{4}\\
\varphi_{v} & =\arctan \frac{h}{2 f} \tag{5}
\end{align*}
$$

Supposing that point $M(i, j)$ in the virtual image plane under the three-dimensional coordinates can be expressed as $M^{\prime}=[i-W / 2, H / 2-j, 0]$, it can be obtained by:

$$
\begin{equation*}
M^{\prime}=M_{3} M_{2} M_{1} X_{s} \tag{6}
\end{equation*}
$$

where $M_{1}, M_{2}, M_{3}$ are:

$$
M_{1}=\left[\begin{array}{cccc}
\cos \alpha & -\sin \alpha & 0 & 0  \tag{7}\\
\sin \alpha & \cos \alpha & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right], M_{2}=\left[\begin{array}{cccc}
\cos \beta & 0 & -\sin \beta & 0 \\
0 & 1 & 0 & 0 \\
\sin \alpha & 0 & \cos \beta & 0 \\
0 & 0 & 0 & 1
\end{array}\right], M_{3}=\left[\begin{array}{llll}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & f
\end{array}\right]
$$

Equation (6) establishes the one-to-one correspondence between $X_{s}$ and $M^{\prime} . M[i, j]$ is the point in 2D coordinates, which can be derived from $M^{\prime}$ by simple coordinate transformation. An overview of virtual perspective image generation is shown in Figure 5. Thus, the virtual image which has the same intrinsic parameters with the conventional camera is derived.


Figure 4. (a) The effective viewpoint of the virtual perspective image; (b) the coordinate of the omnidirectional image.


Figure 5. Overview of virtual perspective image generation.

### 4.2. Automatic Stereo Rectification

After obtaining the virtual perspective image, stereo rectification can be achieved by mapping the virtual and conventional images into the common plane and aligning the columns using rotation and translation matrices. These matrices can be obtained by decomposing the fundamental matrix. How to estimate the fundamental matrix robustly remains a challenging issue. Several methods for estimating the fundamental matrix have been proposed and can be classified into iterative and linear methods. Iterative methods [31,32] are more accurate than linear ones, but have high computational complexity and cannot eliminate the potential outliers. Linear methods contain the 7-point [33] and 8 -point algorithm [34]. The main advantage of the 7-point algorithm is that a fundamental matrix can be estimated by using only seven points, but this fact becomes a drawback when some points are badly located or the corresponding points are redundant. The advantage of the 8-point algorithm is that it permits minimization of the error of estimating the fundamental matrix using redundant points.

The algorithm is fast and easy to implement, but it is sensitive to noise and the solution is unstable. To improve the stability of the 8-point method, Harley et al. [29] normalized the corresponding points before using the 8-point method; in other words, they transformed the data into isotropy.

In our hybrid vision system, there is an affine transformation between the virtual image and the conventional image due to the different shooting angles of the omnidirectional camera and the conventional camera. Thus, ASIFT [28] was used to obtain the exact corresponding points between pairs of images because of its fully affine invariance. The 8-point algorithm with its normalization matrix [29] was adopted to obtain an accurate fundamental matrix through its speed and stability. One of the most important steps in our method was to optimize the normalization matrix. The following is an analysis of the epipolar geometry of the image pair and a detailed description of the optimization method for the normalization matrix.

### 4.2.1. Epipolar Geometry Between Image Pairs

As shown in Figure 6, $O_{p}$ and $O_{v}$ represent the projection center of the conventional camera and virtual perspective camera, respectively. The corresponding imaging planes are $\pi_{p}$ and $\pi_{v} . p$ is a point in the three-dimensional coordinate. $p_{p}$ and $p_{v}$ are the point correspondence of $P$ in the two image planes. The plane defined by $O_{p} O_{v}$ and $p$ is the epipolar plane. The line defined by $O_{p}$ and $O_{v}$ is the baseline. According to [29], the relation of the image pixel coordinates and the fundamental matrix is:

$$
\begin{equation*}
p_{p}^{T} F p_{v}=0 \tag{8}
\end{equation*}
$$

The relation of the essential matrix and the fundamental matrix is:

$$
\begin{equation*}
E=K_{p}^{T} F K_{v} \tag{9}
\end{equation*}
$$

where $K_{v}$ and $K_{p}$ are the intrinsic parameters of the two images. The decomposition of an essential matrix is:

$$
\begin{equation*}
E=[T]_{\times} R \tag{10}
\end{equation*}
$$

where $\bar{T}$ and $T$ differ by a scale factor which can be calculated using two 3D points offline [35]. According to (9) and (10), (8) can be written as:

$$
\begin{equation*}
p_{P}^{T}\left(K_{P}^{T}\right)^{-1}\left([T]_{\times} R\right)\left(K_{v}^{-1}\right) p_{v}=0 \tag{11}
\end{equation*}
$$

Equation (11) establishes the relationship between corresponding points in the image pair and the rotation and translation matrices. Supposing that corresponding points are known, the translation matrix $T$ and rotation matrix $R$ can be calculated. Thus, the image planes $\pi_{p}$ and $\pi_{v}$ are rectified into $\pi_{p}^{\prime}$ and $\pi_{v}^{\prime}$ using $T$ and $R$.


Figure 6. The epipolar geometry of the virtual perspective image and conventional image. $\pi_{p}^{\prime}$ and $\pi_{v}^{\prime}$ are the rectified images. It is obvious that the rectified images are column aligned.

### 4.2.2. Optimization Method of the Normalization Matrix

In order to improve the stability of the 8-point algorithm, raw data was transformed into isotropic data using normalization matrices [29]. However, the normalization matrix was calculated separately without considering the relative position of the image pair.

In our proposed method, we used a cost function to find the optimal normalization matrix and minimize the horizontal distance between corresponding pairs. The main processing algorithm is described in the following.
$N$ pairs of correspondence points $\left(x_{i}, x_{i}^{\prime}\right) i=1,2, \cdots, N$ are obtained by ASIFT [28]. $H$ and $H^{\prime}$ are two normalization matrices for two groups of points, respectively. The normalization can be achieved by:

$$
\begin{equation*}
\bar{x}_{i}=H x_{i}, \bar{x}_{i}^{\prime}=H^{\prime} x_{i}^{\prime} \tag{12}
\end{equation*}
$$

where $\bar{x}$ and $\bar{x}^{\prime}$ are point correspondences after normalization. From (11) and (12), the following expression can be obtained:

$$
\begin{equation*}
x_{i}^{T} H^{T}\left(K_{P}^{T}\right)^{-1}\left([T]_{\times} R\right)\left(K_{v}^{-1}\right) H^{\prime} x_{i}^{\prime}=0 \tag{13}
\end{equation*}
$$

Thus, the error of epipolar geometry between two images can be expressed as:

$$
\begin{equation*}
E\left(H, H^{\prime}\right)=\sum_{i=1}^{N} x_{i}^{T} H^{T}\left(K_{P}^{T}\right)^{-1}\left([T]_{\times} R\right)\left(K_{v}^{-1}\right) H^{\prime} x_{i}^{\prime} \tag{14}
\end{equation*}
$$

In addition to the epipolar geometry constraint, the horizontal distance between two images is another significant factor. The horizontal distance between pair correspondences is:

$$
\begin{equation*}
D\left(H, H^{\prime}\right)=\sum_{i=1}^{N}\left(\left|\left(H x_{i}\right)-\left(H^{\prime} x_{i}^{\prime}\right)\right|\right) \tag{15}
\end{equation*}
$$

According to (14) and (15), we define the objective function of the optimization problem as:

$$
\begin{equation*}
S\left(H, H^{\prime}\right)=\alpha E\left(H, H^{\prime}\right)+\beta D\left(H, H^{\prime}\right) \tag{16}
\end{equation*}
$$

In order to minimize $S\left(H, H^{\prime}\right)$, the iterative Expectation Maximization (EM) [36] is adopted because of its simplicity and effectiveness. $\alpha$ determines the ratio of epipolar geometry error and $\beta$ the ratio of horizontal distance error. They are subject to $\alpha+\beta=1$.

From (16), we can obtain the normalization matrices $H$ and $H^{\prime}$. The corresponding points are normalized into isotropic points using $H$ and $H^{\prime}$, which can obtain a more accurate result than the method proposed in [27].

## 5. Experimental Results and Analysis

### 5.1. Hybrid Stereo Vision System

Figure 7 presents our experimental setup. Improved measurement accuracy can be achieved by adjusting the baseline distance. The hardware configuration of this experiment was a computer equipped with a dual-core Intel Pentium G2020 29 GHz , and 4 GB of RAM, running Windows 10. The system was implemented in VS2015 combined with OpenCV 2.4.9 and OpenGL 4.3. The cameras were synchronized via an external trigger. The parameters of the omnidirectional camera and the conventional camera are shown in Table 1. To verify the accuracy of the proposed rectification method, a stereo rectification experiment was performed.


Figure 7. Experiment platform. The upper camera is omnidirectional, and the lower camera is conventional.

Table 1. Camera parameters given by the manufacturer.

| Hyperbolic Mirror Parameters |  | Omnidirectional Camera Parameters |  | Conventional Camera Parameters |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| a (Major axis) | 31.2888 mm | Part Number | FL2G-50S5C-C | Part Number | FL2G-50S5C-C |
| b (Minor axis) | 51.1958 mm | Resolution | $1360 \times 1360$ pixels | Resolution | $2448 \times 2048$ pixels |
| mapping parameter | 0.82 | Frame rate | 10 frames $/ \mathrm{s}$ | Frame rate | 10 frames $/ \mathrm{s}$ |
| vertical viewing angle | $120^{\circ}$ | Interface | 1394 b | Interface | 1394 b |

### 5.2. Stereo Rectification Experiment with Real Image Pairs

We first evaluated the performance of the proposed stereo rectification approach using 15 image pairs, where three of the original image pairs are shown in Figure 8. The results of the three image pairs with the rectification algorithm from [27] and with the proposed rectification approach are shown in Figure 9a,b. As can be seen from these three pairs of images, the performance improvement obtained by the proposed solution is evident. To show the accuracy of the rectification, we highlighted a few notable regions where results of the rectification method from [27] exhibited misalignments, whereas our results remain aligned in these regions. In addition, it is obvious that the distortion at the edge in Figure 9a is not calibrated with the method proposed in [27].


Figure 8. Three examples of the image pairs used for rectification accuracy comparison.


Figure 9. Stereo rectification results. The first row in each image is from the conventional camera. The second row in each image is from the omnidirectional camera. (a) The image pair with the rectification method in [27]; (b) the image pair with our proposed rectification method.

To represent rectification error quantitatively, we selected four stereo image pairs from the indoor environment that were rectified using the rectification method from [27] and our proposed method. From each image set, we randomly selected 30 corresponding corners and calculated the mean deviation of the horizontal distances. We summarized the average difference of each individual image set along with their overall average (term Average Err.). Table 2 depicts these results in pixels. From the average error, we can calculate that the accuracy of our method increased by $34.78 \%$ compared with the method from [27]. Since rectification gives a pair of images in which corresponding epipolar lines should be collinear and parallel to the vertical axis, the abovementioned criterion is suitable for computing the error in a rectified pair of images.

Table 2. Comparisons of rectification errors for 4 rectified stereo image pairs (in pixels).

|  | Method in [27] | Our Proposed Method |
| :---: | :---: | :---: |
| Set 1 | 2.457 | 1.401 |
| Set 2 | 2.374 | 1.645 |
| Set 3 | 2.621 | 1.831 |
| Set 4 | 1.987 | 1.176 |
| Average Err. | 2.360 | 1.513 |

### 5.3. Odometry in a Simulated Environment

Stereo vision odometry is based on the parallax of two images. The accuracy of odometry represents the accuracy of the rectification method. Thus, we analyzed odometry accuracy in a simulated environment. As shown in Figure 10, we placed the omnidirectional camera at the origin. In other words, the camera coordinate was consistent with the world coordinate, while the conventional camera was put at $(0,0,2)$. The parameters of the cameras were set to the same values as in Table 1. We performed ten groups of experiments where each had a different rotation and translation matrix. We placed ten points in 3D coordinates for each experiment. Captured points of one experiment in an omnidirectional image and perspective image are shown in Figure 11a,b, respectively. Ten points were used to calculate the $R$ and $T$ matrixes. The image pairs were rectified using our proposed approach and the method in [27], respectively. The vertical disparities $Y_{2}-Y_{1}$ of 10 corresponding points were derived. The distances between sample points and cameras were calculated by (17). $f$ is the camera focal length, and L is the length of the vertical baseline. In our simulated experiment, $f=310.57 \mathrm{~mm}$, $L=345.584 \mathrm{~mm}$. The value of the $Y$-axis of each point was used as a ground truth. The mean errors of distance in different orientations between cameras are shown in Figure 12. Compared with the method in [27], the mean error in each experiment decreased by $0.1-0.2 \mathrm{~m}$ using our proposed approach. It can easily be inferred that our proposed rectification approach is more reliable.

$$
\begin{equation*}
d_{i}=\frac{f L}{Y_{2}-Y_{1}} \tag{17}
\end{equation*}
$$



Figure 10. A simulated environment with one omnidirectional image and one conventional image.


Figure 11. Sample omnidirectional (a) and perspective (b) images captured in the simulated environment.


Figure 12. Mean errors of ten experiments with different orientation angles.

### 5.4. Real-Time Target Tracking and Odometry Experiment

To illustrate the performance of the proposed rectification method and provide a direct application, we used the hybrid vision system to track a target with a size of $1.7 \times 0.6 \times 0.3 \mathrm{~m}$ and perform odometry. The tracking algorithm for the omnidirectional camera in [37] was adopted. The cameras shown in Figure 7 was placed in a fixed position. The target tracking and odometry real-time experiment had a total output of 154 frames over a period of 20 s , while the average computation time of each frame was 179.87 ms .

Five frames were randomly selected from the processing results, and are shown in Figure 13. It demonstrates that the angle from the omnidirectional camera can be sent to the conventional camera successfully, and the two cameras can cooperate with each other very well.


Figure 13. The tracking and cooperation result of the two cameras. From left to right, the 17th, 26th, 35th, 43rd, and 85th frame are shown.

We used our proposed method to rectify image pairs, the results of which are shown in Figure 14. The comparison of the odometry results and the ground truth is shown in Figure 15 (the hybrid vision system was fixed at the origin). The average error distance is 0.317 m . The experimental results show that the error is far less than the target scale. We can conclude that the calculated trajectory is consistent with the ground truth, and our hybrid vision system can be used in surveillance tasks.


Figure 14. The rectification result of image pairs in Figure 13.


Figure 15. The comparison of odometry results and ground truth.

## 6. Discussion

In summary, we demonstrated an automatic rectification approach for the hybrid vision system. The geometric relationship between omnidirectional and conventional images was simplified by generating a virtual perspective image from the omnidirectional image. Image pairs were rectified using the 8-point algorithm with an optimized normalization matrix. We showed that the row coordinate parallax of rectified image pairs was within 2 pixels. The mean errors of the odometry based on triangulation were less than $12 \%$. Based on this, we consider that the error of rectification is acceptable when compared with other methods in literature. Deng et al. [20] also used the corresponding points to acquire the extrinsic parameters of the hybrid vision system, but they used 3 D points instead of 2D points on the plane. 3D points were extracted based on 3D Euclidean reconstruction of scene points, which involves high computational complexity due to the establishment of the polynomial approximation model. Lin et al. [27] also generated a virtual perspective image to simplify the epipolar geometry between hybrid cameras, but they synthesized the virtual plane by back-projecting the rays directly from the omnidirectional image. This also suffers from calculation complexity due to the non-linear imaging model of the omnidirectional image. In addition, they derived a 3D point by calculating the intersection of two rays, which cannot obtain an accurate result. Among the previously proposed practical solutions for using an omnidirectional-conventional camera pair, only one of them actually estimates the relative position and orientation of the cameras, which is
given in [27]. Therefore, we numerically compared the accuracy of the parameter estimation of our method only with [27].

It is noteworthy that the number of corresponding points strongly affects the rectification performance, because the fundamental matrix is estimated based on their position. The experiment in no salient feature scene was performed. The error of the fundamental matrix was very large since almost no corresponding points were detected.

In the future, we would like to use our approach for different camera combinations, including omnidirectional-fisheye and fisheye-conventional. Additionally, we plan to develop an improved method which is not limited by various features in the scene.

## 7. Conclusions

In this paper, an active hybrid vision system consisting of an omnidirectional camera and a conventional camera was presented. We provided the key techniques to rectify image pairs automatically. The virtual perspective image from an omnidirectional image was obtained using the proposed perspective projection model. ASIFT and the 8-point algorithm with an optimized normalization matrix were applied to rectify the image pair automatically. Our investigation in the simulated and real environment has demonstrated that the proposed approach not only overcomes the shortcomings of high computational complexity in the hybrid vision system, but is superior to other state-of-the-art methods [27] in regard to accuracy. Therefore, our system can effectively meet the requirements of vision sensors in surveillance tasks.

Author Contributions: B.F. conceived and designed the experiments; B.F. and X.L. performed the experiments; B.F. and C.C. analyzed the data; B.F. wrote the paper; Q.Z. supervised and instructed the experiment.

Acknowledgments: This work has been supported in part by the National Natural Science Foundation of China via grants 61203255 and 61175089. The Fundamental Research Funds for the Central Universities. (HEUCF180405).
Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Klitzke, L.; Koch, C. Robust Object Detection for Video Surveillance Using Stereo Vision and Gaussian Mixture Model. J. WSCG 2016, 24, 9-17.
2. Barry, A.J.; Tedrake, R. Pushbroom Stereo for High-Speed Navigation in Cluttered Environments. arXiv 2014, arXiv:1407.7091.
3. De Wagter, C.; Tijmons, S.; Remes, B.D.W.; de Croon, G.C.H.E. Autonomous flight of a 20-gram Flapping Wing MAV with a 4-gram onboard stereo vision system. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May-7 June 2014; pp. 4982-4987.
4. Marín-Plaza, P.; Beltrán, J.; Hussein, A.; Musleh, B.; Martín, D.; de la Escalera, A.; Armingol, J.M. Stereo Vision-Based Local Occupancy Grid Map for Autonomous Navigation in ROS. In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Rome, Italy, 27-29 February 2016; pp. 701-706.
5. Fu, C.; Carrio, A.; Campoy, P. Efficient visual odometry and mapping for unmanned aerial vehicle using ARM-based stereo vision pre-processing system. In Proceedings of the 2015 International Conference on Unmanned Aircraft Systems (ICUAS), Denver, CO, USA, 9-12 June 2015; pp. 957-962.
6. De La Cruz, C.; Carelli, R. Dynamic model based formation control and obstacle avoidance of multi-robot systems. Robotica 2008, 26, 345-356. [CrossRef]
7. Micusik, B.; Pajdla, T. Estimation of omnidirectional camera model from epipolar geometry. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18-20 June 2003; Volume 1.
8. Wang, Y.; Gong, X.; Lin, Y.; Liu, J. Stereo calibration and rectification for omnidirectional multi-camera systems. Int. J. Adv. Robot. Syst. 2012, 9, 143. [CrossRef]
9. Ramalingam, S.; Sturm, P. A Unifying Model for Camera Calibration. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1309-1319. [CrossRef] [PubMed]
10. Yu, M.-S.; Wu, H.; Lin, H.-Y. A visual surveillance system for mobile robot using omnidirectional and PTZ cameras. In Proceedings of the SICE Annual Conference, Taipei, Taiwan, 18-21 August 2010; pp. 37-42.
11. Cagnoni, S.; Mordonini, M.; Mussi, L.; Adorni, G. Hybrid Stereo Sensor with Omnidirectional Vision Capabilities: Overview and Calibration Procedures. In Proceedings of the ICIAP 2007 14th International Conference on Image Analysis and Processing, Modena, Italy, 10-14 September 2007; pp. 99-104.
12. Bastanlar, Y. A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs. Pattern Recognit. Lett. 2016, 71, 1-7. [CrossRef]
13. Sturm, P. Pinhole Camera Model. In Computer Vision: A Reference Guide; Ikeuchi, K., Ed.; Springer US: Boston, MA, USA, 2014; pp. 610-613. ISBN 978-0-387-31439-6.
14. Lui, W.L.D.; Jarvis, R. Eye-full tower: A gpu-based variable multibaseline omnidirectional stereovision system with automatic baseline selection for outdoor mobile robot navigation. Robot. Auton. Syst. 2010, 58, 747-761. [CrossRef]
15. Schraml, S.; Belbachir, A.N.; Bischof, H. An Event-Driven Stereo System for Real-Time 3-D 360 Panoramic Vision. IEEE Trans. Ind. Electron. 2016, 63, 418-428. [CrossRef]
16. Barone, S.; Neri, P.; Paoli, A.; Razionale, A.V. Catadioptric stereo-vision system using a spherical mirror. Procedia Struct. Integr. 2018, 8, 83-91. [CrossRef]
17. Chen, D.; Yang, J. Image registration with uncalibrated cameras in hybrid vision systems. In Proceedings of the Seventh IEEE Workshops on Application of Computer Vision, WACV/MOTIONS'05, Breckenridge, CO, USA, 5-7 January 2005; Volume 1, pp. 427-432.
18. Rathnayaka, P.; Baek, S.-H.; Park, S.-Y. An Efficient Calibration Method for a Stereo Camera System with Heterogeneous Lenses Using an Embedded Checkerboard Pattern. J. Sens. 2017, 2017, 6742615. [CrossRef]
19. Chen, X.; Yang, J.; Waibel, A. Calibration of a hybrid camera network. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13-16 October 2003.
20. Deng, X.; Wu, F.; Wu, Y.; Duan, F.; Chang, L.; Wang, H. Self-calibration of hybrid central catadioptric and perspective cameras. Comput. Vis. Image Underst. 2012, 116, 715-729. [CrossRef]
21. Puig, L.; Guerrero, J.; Sturm, P. Matching of omnidirectional and perspective images using the hybrid fundamental matrix. In Proceedings of the OMNIVIS 2008-8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France, 17 October 2008.
22. Chen, C.; Yao, Y.; Page, D.; Abidi, B.; Koschan, A.; Abidi, M. Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking. IEEE Trans. Circuits Syst. Video Technol. 2008, 18, 1052-1063. [CrossRef]
23. Liu, Y.; Shi, H.; Lai, S.; Zuo, C.; Zhang, M. A spatial calibration method for master-slave surveillance system. Opt. Int. J. Light Electron Opt. 2014, 125, 2479-2483. [CrossRef]
24. Tan, S.; Xia, Q.; Basu, A.; Lou, J.; Zhang, M. A two-point spatial mapping method for hybrid vision systems. J. Mod. Opt. 2014, 61, 910-922. [CrossRef]
25. Baris, I.; Bastanlar, Y. Classification and tracking of traffic scene objects with hybrid camera systems. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16-19 October 2017; pp. 1-6.
26. Scotti, G.; Marcenaro, L.; Coelho, C.; Selvaggi, F.; Regazzoni, C.S. Dual camera intelligent sensor for high definition 360 degrees surveillance. IEE Proc. Vis. Image Signal Process. 2005, 152, 250-257. [CrossRef]
27. Lin, H.-Y.; Wang, M.-L. HOPIS: Hybrid omnidirectional and perspective imaging system for mobile robots. Sensors 2014, 14, 16508-16531. [CrossRef] [PubMed]
28. Yu, G.; Morel, J.-M. ASIFT: An Algorithm for Fully Affine Invariant Comparison. Image Process. 2011, 1. [CrossRef]
29. Hartley, R.I. In defence of the 8 -point algorithm. In Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, USA, 20-23 June 1995; pp. 1064-1070.
30. Goncalves, N.; Nogueira, A.C.; Miguel, A.L. Forward projection model of non-central catadioptric cameras with spherical mirrors. Robotica 2017, 35, 1378-1396. [CrossRef]
31. Simoncini, V. Computational Methods for Linear Matrix Equations. SIAM Rev. 2016, 58, 377-441. [CrossRef]
32. Toldo, R.; Gherardi, R.; Farenzena, M.; Fusiello, A. Hierarchical structure-and-motion recovery from uncalibrated images. Comput. Vis. Image Underst. 2015, 140, 127-143. [CrossRef]
33. Albl, C.; Kukelova, Z.; Fitzgibbon, A.; Heller, J.; Smid, M.; Pajdla, T. On the Two-View Geometry of Unsynchronized Cameras. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017; pp. 5593-5602.
34. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections. Nature 1981, 293, 133. [CrossRef]
35. Cai, C.; Fan, B.; Weng, X.; Zhu, Q.; Su, L. A target tracking and location robot system based on omnistereo vision. Ind. Robot. 2017, 44, 741-753. [CrossRef]
36. Moon, T.K. The expectation-maximization algorithm. IEEE Signal Process. Mag. 1996, 13, 47-60. [CrossRef]
37. Cai, C.; Weng, X.; Fan, B.; Zhu, Q. Target-tracking algorithm for omnidirectional vision. J. Electron. Imaging 2017, 26, 033014. [CrossRef]
© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
