

Image Registration with Uncalibrated Cameras in Hybrid Vision Systems

Datong Chen, Jie Yang
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

This paper addresses the problem of robust registering of images among perspective and omnidirectional cameras in a hybrid vision system (HVS). Nonlinearity in an HVS introduced by omnidirectional cameras poses challenges for computing pixel correspondences among images. In previous HVSs, cameras must be calibrated by performing registration. In this paper, we propose a non-linear approach for registering images in an HVS without requiring calibration of cameras. We first discuss the homographies between omnidirectional and perspective images under a local planar assumption. We then propose a robust patch level registration algorithm by exploiting a constraint on large 3D spatial planes. The proposed approach enables an HVS for applications that require quick deployment or active cameras. Experimental results have demonstrated feasibility of the proposed approach.

1 Introduction

Recent demands on video surveillance in a large area have activated research interest in camera networks. A hybrid vision system (HVS) is a camera network that consists of omnidirectional and perspective cameras. Such a system takes advantage of a large view scope from omnidirectional cameras and higher resolution from perspective cameras. For example, Chen et. al. [3] proposed an HVS architecture in which an omnidirectional was mounted on the ceiling of the center of a large room and several perspective cameras were mounted on surrounding side walls. The omnidirectional camera not only provides a good reference for cameras in the camera network but also minimizes the possibility of occlusions in a tracking process. The perspective camera can capture more detail information in higher resolutions. However, nonlinearity in an HVS introduced by omnidirectional cameras poses challenges for many existing computer vision techniques, including the technique of computing pixel correspondences between omnidirectional and perspective views.

In the previous literature, correspondences of higher res-

olution side-view images and lower resolution top-view images are computed using 2D perspective homography [4]. The approach requires the pre-calibration of the intrinsic parameters of an omnidirectional camera. The most commonly used omnidirectional camera is a catadioptric camera, which is composed of a perspective camera and a mirror and provides a single effective viewpoint [10]. The calibration of a catadioptric camera has been addressed by many other researchers [7, 2, 9, 13]. After an omnidirectional camera is calibrated, 2D perspective homography assumes that the scene in front of each camera is planar and registers perspective view images under the reference of a distorted omnidirectional-view image [8, 4].

A major drawback of the 2D approach is that the calibration step involves manual interaction or specially designed calibration tags with specific patterns or shapes. This limits applications of an HVS system when a quick deployment is required or auto-zooming cameras are employed. In addition, the existing registration methods can make large errors due to the fact of the non-planar scene. Some efforts were made to provide 3D information in an HVS by performing calibration of extrinsic parameters among cameras. Sturm analyzed catadioptric cameras and perspective cameras within a common scene [14]. Chen et al. proposed a manual solution based on pre-measured points in real 3D space [3]. Stereo methods [1, 15, 5, 11, 8] were proposed for object detection and reconstruction in an HVS. Calibration of the omnidirectional camera is required by all these methods.

In this paper, we propose an automatic approach of image registration with uncalibrated cameras in an HVS. We first discuss the geometric correspondence of a planar object between a perspective camera and a catadioptric camera and give two homography matrices from both directions. We then propose an algorithm to register an image of a perspective camera to a catadioptric image under the assumption that local image patches are the projections of planar surfaces. In the proposed algorithm, non-linear 2D registration is performed at a local patch level. A robust estimation methodology is proposed for propagating the homography of patches to their neighborhood. We demonstrate feasibility of the proposed method through experiments.

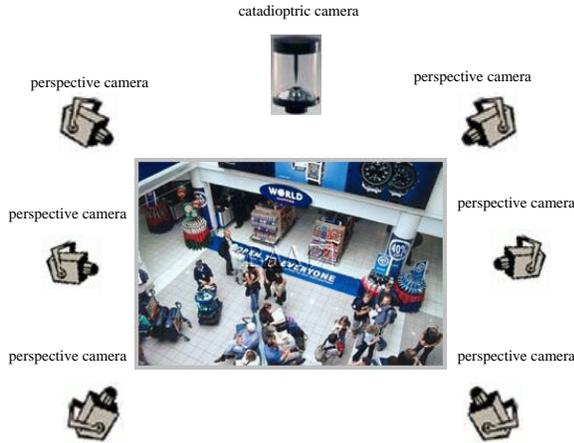


Figure 1: An illustration of an HVS in an indoor public environment

2 Mathematic models of an HVS

Let's consider a simple HVS that consists of only one catadioptric camera and several perspective cameras. Fig. 1 illustrates such a system installed in an indoor public environment. We will limit our discussion to such an HVS in the rest of this paper, though the results can be extended to more complex systems.

2.1 A catadioptric camera model

A commercial catadioptric camera can be modeled as a combination of a paraboloid mirror and lenses (see Fig. 2). To exploit the optical characteristics of a catadioptric system in the spatial domain, we can follow the processes by which a catadioptric acquires an optical signal from a spatial point $P = (X, Y, Z)^T$ shown in Figure 2. Without losing generality, let's select the focus of the paraboloid mirror as the origin O_c . The signal from point $P = (X, Y, Z)^T$ is firstly reflected at $P_m = (X_m, Y_m, Z_m)^T$ on the mirror and then is projected on the image plane at $p = (x, y, Z_c)^T$. To simplify this projection process, we assume the camera is focused on a virtual focal plane \mathcal{F} . The mirror point P_m is first transformed onto focal plane \mathcal{F} at the point $P_f = (X_f, Y_f, Z_f)^T$ and projected on the image plane, which can be modeled by the following equations:

$$\begin{aligned} P_m &= \alpha_P P \\ P_f &= P_m + T_f, \\ p &= R_I P_f \end{aligned} \quad (1)$$

where the scale factor α_P is a function of the spatial point P . $T_f = (0, 0, Z_f - Z_m)^T$ denotes the translation between the focal plane \mathcal{F} and the mirror. R_I models the perspective projection from the focal plane to the image plane. The

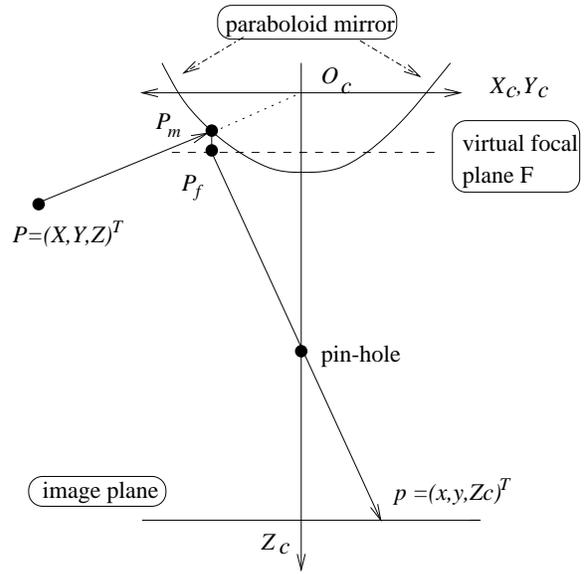


Figure 2: A model of catadioptric camera

focal plane $Z = Z_f$ has only one parameter. In such a paraboloidal mirror based catadioptric system, there are two parameters α_P and T_f that are 3D spatial point dependent. The other parameters, R_I and Z_f , consist only of constant values.

Furthermore, for a point on the paraboloidal mirror $P_m = (X_m, Y_m, Z_m)^T$, the paraboloid can be described as:

$$Z_m = f - \frac{1}{4f}(X_m^2 + Y_m^2), \quad (2)$$

where f is the focal length of the mirror.

2.2 A perspective camera model

There are many different approaches for modeling a perspective camera. In this paper, we use the linear pin-hole model. In this model, the geometric relationship between a spatial point $\hat{P} = (\hat{X}, \hat{Y}, \hat{Z})^T$ and its projection on image plane $\hat{p} = (f_1 \hat{x}, f_1 \hat{y}, f_1)^T$ can be modeled as:

$$\hat{Z} \hat{p} = \hat{P}. \quad (3)$$

To simplify the discussion in this paper, we assume that the principle point is located at the image center, the aspect ratio of the optical axis is 1, and the focus length f_1 (which is a scalar of the system) equals 1. However, the results in this paper can be extended to more complex linear pin-hole models.

2.3 Corresponding points in an HVS

A advantage of an HVS system is that the catadioptric camera can provide a global view of a scene. Therefore, the calibrations of most HVS systems rely on the corresponding point pairs between a perspective image and the catadioptric image. Let us assume that a spatial point P under the catadioptric camera coordinate system corresponds to a point \hat{P} under the coordinate system of a perspective camera. The transform between this point pair can be defined as:

$$\hat{P} = R_e P + T_e. \quad (4)$$

Substituting Eq. 1 and 3 into 4, we have

$$\hat{Z}\hat{p} = \frac{Rp + T}{\alpha_P} + T_e, \quad (5)$$

where $R = R_e R_f^{-1}$, $T = -R_e T_f$, and T_e are homographic related parameters which need to be estimated. p and \hat{p} are projections of the spatial point on the catadioptric and perspective image planes. In general, these homographic related parameters are non-computable since both parameters \hat{Z} and α_P contain unknown depth information of the spatial point. In the next sections, we give a specific solution of estimating these homographic related parameters.

3 Homography of a planar object in an HVS

Without losing generality, we can assume that most of the local regions in images represent planer surfaces in a scene. The homography from catadioptric image to a perspective image can be modeled using a 3×4 matrix as proposed in [14]:

$$\hat{p} = (x, y, 1)^T = H_1(x, y, x^2 + y^2 - \frac{1}{4}, 1)^T. \quad (6)$$

The homography from a perspective image to a catadioptric image, which will be used in a registration task, is a little complex. Suppose that a target object has a planar surface $WP + b = 0$ under the catadioptric coordinate system. Representing this surface using pixels on the catadioptric image plane defined by Eq. 1, we have $\frac{1}{\alpha_P} W R_e^{-1}(Rp + T) + b = 0$. This is a useful constraint on the scale factor α_P :

$$\alpha_P = \frac{W R_e^{-1}(Rp + T)}{-b}. \quad (7)$$

Substitute Eq. 7 into Eq. 5, we obtain a 3×6 homography matrix:

$$p = H_2(\hat{x}^2, \hat{y}^2, \hat{x}\hat{y}, \hat{x}, \hat{y}, 1)^T. \quad (8)$$

This homography actually has a similar constraint to Eq. 6. However, due to the ambiguity when mapping a perspective image back to the the catadioptric surface, we do not have a ‘‘linear’’ equation. We can search the homographic matrix 8 under the constraint 6. The algorithm can be briefly described as:

1. Initialize H_2^t ;
2. Compute all the corresponding $p^t = H_2^t \hat{p}^t$;
3. Register p back to the perspective image to obtain \hat{p}^{t+1} ;
4. Update $H_2^{t+1} = H_2^t + \lambda correlation(\hat{p}^t, \hat{p}^{t+1})$;
5. Loop to step 2 until the stop condition is satisfied.

4 Patch-level image registration

Estimation of the homography matrix for a planar surface is a traditional image registration task [6]. The difficulty arises in that the original scene captured by cameras is not planar as assumed in a 2D approach. To address this problem, we divide an image from a perspective camera into small patches B_i , and assume that each patch corresponds to a planar surface in 3D space. Therefore, the image registration is performed at patch level.

To address this patch level registration, we propose an algorithm consisting of three main iterative steps: patch selection, patch registration, and homography propagation, which is outlined as the following:

Algorithm of robust homography propagation at a patch level

1. Partition a perspective image into n partitions (patches) and label all the patches as unregistered;
2. Select an unregistered patch B with the highest variance;
3. Register the patch B using the technique described in the last section;
4. Propagate the homography of patch B to its unregistered neighbors, which are located in the same 3D spatial plane;
5. If there are un-registered patches, go to step 2; else end.

In this algorithm, we partition an image into patches of the same size. The algorithm iteratively performs from step 2 to step 4 until all the patches are registered. The details of steps 3 and 4 are discussed as the following.

4.1 Registration in a Haar feature space

Patch registration step 3 is performed in a Haar feature space. Haar wavelets are chosen since they are able to model texture in different scales and can be computed very efficiently. The Haar wavelets decompose a given image patch B into four sub-bands: lower frequency band B^l , vertical high frequency band B^v , horizontal high frequency



Figure 3: An image patch and its Haar decomposition. The four bands in the image on the right correspond to B^l (top-left), B^v (top-right), B^h (bottom-left) and B^d (bottom-right)

band B^h , and diagonal high frequency band B^d . Figure 3 illustrates a Haar decomposition of a large image patch. For a point $p = (x, y)$ in the image patch B , its Haar feature values are defined as:

$$B_{x,y}^l = \frac{1}{4} (B_{2x,2y} + B_{2x,2y+1} + B_{2x+1,2y} + B_{2x+1,2y+1}),$$

$$B_{x,y}^v = \frac{1}{4} (B_{2x,2y} - B_{2x,2y+1} + B_{2x+1,2y} - B_{2x+1,2y+1}),$$

$$B_{x,y}^h = \frac{1}{4} (B_{2x,2y} + B_{2x,2y+1} - B_{2x+1,2y} - B_{2x+1,2y+1}),$$

$$B_{x,y}^d = \frac{1}{4} (B_{2x,2y} - B_{2x,2y+1} - B_{2x+1,2y} + B_{2x+1,2y+1}).$$

The registration task is to minimize the following objective function with respect to the homography matrix H and global photometric parameters $\theta = (a_j, b_j)$:

$$f(H, \theta) = \sum_{j \in \{l, v, h, d\}} \sum_{\hat{p} \in B} (B_{\hat{p}}^j - a_j I_{H\hat{p}}^j - b_j)^2, \quad (9)$$

where I^j is the Haar feature image from the catadioptric camera. This minimization involves non-linear constraints and can be solved by the Levenverg-Marquardt [6, 12] technique.

4.2 Robust homography propagation

Homography propagation step 4 propagates the homography H obtained from the registration of patch B in step 3 to its neighbors. From Eq. 7, we can observe that a homography matrix contains the factor of W and b associated with a spatial plane. Therefore, patches coming from the same 3D spatial plane should share the same homography. Firstly, we use H to initialize a set of homographies $S_H = H^1 = H$ and set the seed homography as $H^* = H$. Then, we register the unregistered patches in the 8 neighbors of the patch B . The seed homography H^* is used as an initialization in the Levenverg-Marquardt based registration algorithm. The resulted homographies are added into the set

$S_H = H^1, \dots, H^n$. The new seed homography $H^* \equiv h_{ij}^*$ is then updated as:

$$h_{ij}^* = \frac{\sum_{k=1}^n \beta_{ij}^k h_{ij}^k}{\sum_{k=1}^n \beta_{ij}^k},$$

where $\beta_{ij}^k = \begin{cases} 1 & \frac{|h_{ij}^k - m_{ij}|}{\sigma_{ij}} < \delta \\ 0 & \text{otherwise} \end{cases}$, and the mean m_{ij} is defined as:

$$m_{ij} = \arg \min_m r_{ij}^m,$$

$$r_{ij}^m = \text{Med}_{k=1, \dots, n} |h_{ij}^k - m|.$$

The Med denotes the median operator. The variance σ_{ij} is computed as: $\sigma_{ij} = 1.48 \times \left(1 + \frac{5}{n-1}\right) r_{ij}^{m_{ij}}$. The threshold δ is a tradeoff between precision and robustness of the registration performance.

We iteratively propagate the registration to the 8 neighbors of each new registered patch until the weights β_{ij}^k of all the new registered patches are equal to zero.

This registration algorithm mainly focuses on registering large background planes such as walls and ground floor. Small foreground objects are usually not planar enough and have too low resolution. Therefore the registration results can be noisier.

5 Experimental results

The proposed approach is evaluated on images obtained from a catadioptric camera and a perspective camera. Fig. 4 shows two of these images: (a) a top-view image from a Cyclovision's catadioptric camera; (b) a side-view image from a SONY perspective camera. The catadioptric image has a resolution of 640×480 . The resolution of the perspective image is 800×600 .

During the patch level registration process, we first estimate the translation and the scale parameters and then the global photometric parameters. Finally, we estimate the homography. In Fig. 5, we display the registration results for the proposed methods in different patch sizes: (a) image registration results using 2D perspective homography; (b) image registration results using non-linear homography with only one partition; (c) image registration results using non-linear homography with 2×2 partitions; (d) image registration results using non-linear homography with 4×4 partitions. The traditional 2D perspective registration does not work well without pre-calibrating and warping the catadioptric image. The proposed approach gives better registration results using non-linear homography. Comparing with the result in (b), the result in (c) illustrates clearly that

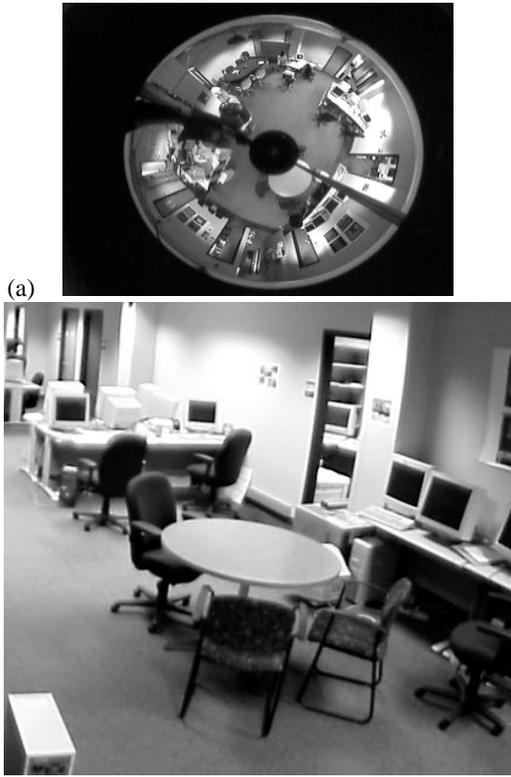


Figure 4: Experiment images: (a) a top-view image from a Cyclovision's catadioptric camera; (b) a side-view image from a SONY perspective camera.

there are three 3D spatial “planes” in the scene: the wall combined with foreground objects, the wall (on the right), and the ground floor. When using 16 partitions, the wall combined with foreground objects part is better registered. However, the registrations of the patches on the wall (on the right) become noisy due to not enough texture in some partitions.

To evaluate the registration results more precisely, we manually label 60 corresponding points $\{(p_i, \hat{p}_i)\}$ in both catadioptric image and perspective image, which are shown in Fig. 6. According to this ground truth, the registration error is measured by the sum of the translations of the 60 points between the registered coordinates and the coordinates in the ground truth:

$$E = \sum_{i=1}^{60} \|p_i - H\hat{p}_i\|. \quad (10)$$

Fig. 7 shows the registration errors using different number of partitions. We can observe that the registration error decreases as the local patch size decreases.

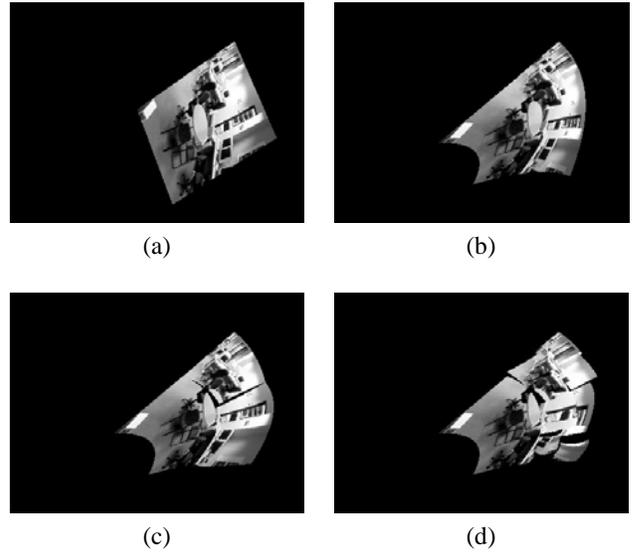


Figure 5: Image registration result comparisons: (a) image registration result using 2D perspective homography; (b) image registration result using the proposed method with only 1 partition; (c) image registration results using the proposed method with 4 partitions; (d) image registration results using the proposed method with 16 partitions.

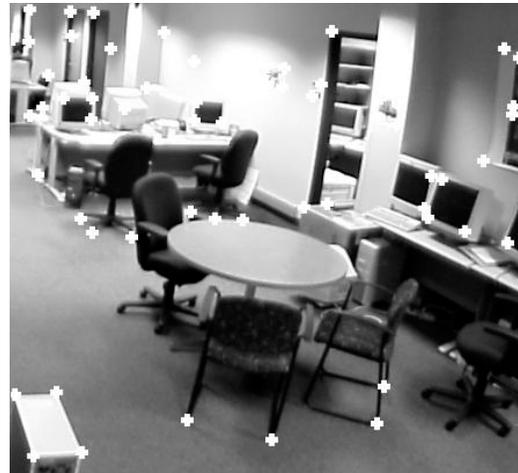


Figure 6: Corresponding points in the ground truth.

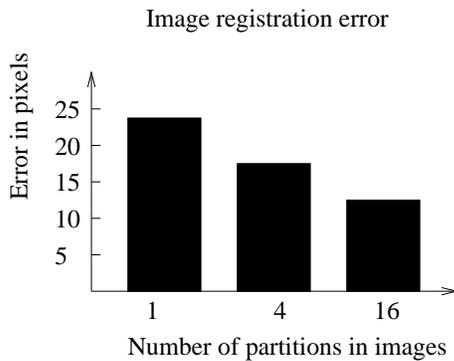


Figure 7: Image registration errors for the proposed methods with different numbers of partitions.

6 Conclusions

In this paper, we have proposed an automatic image registration approach for an uncalibrated hybrid omnidirectional/perspective vision system. Using the proposed approach, omnidirectional and perspective images can be registered with uncalibrated cameras. Therefore, an HVS can be applied to those applications that require quick deployment or active cameras.

The proposed approach is based on non-linear homographies deduced from the 3D geometric model of an HVS and the models of both catadioptric and perspective cameras under a local planar assumption. We assume that local patches of the images represent planar 3D surfaces, which is reasonable in most general cases.

Furthermore, we also proposed a robust patch level registration algorithm by exploiting the constraint that patches from the same 3D planar surface share the same homography. The algorithm propagates the registration of a local image patch to its neighborhood using a robust estimation of the shared homography. Experiments have demonstrated the registration results using the proposed approach. Comparisons between traditional 2D perspective approach and the proposed approach with different partitions illustrated the efficiency of the proposed approach. The dependence of the registration results to the size and properties of local patches indicates that more work needs to be focused on irregular image partition in the future.

Acknowledgements

This research is partially supported by of the Department of Defense (USA) through award numbers N41756-03-C4024, the National Science Foundation (USA) through award No. 0205219, and the European Commission within the project CHIL under contract No. 506909.

References

- [1] G. Adorni, L. Bolognini, S. Cagnoni, and M. Mordonini. Stereo obstacle detection method for a hybrid omnidirectional/pin-hole vision system. In *Proc. RoboCup 2001: Robot Soccer. World Cup V*, pages 244–250, Aug. 2001.
- [2] C. Cauchois, E. Brassart, L. Delahoche, and T. Delhommele. Reconstruction with the calibrated syclop sensor. In *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, pages 1493–1498, 2000.
- [3] X. Chen, J. Yang, and A. Waibel. Calibration of a hybrid camera network. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 000–000, 2003.
- [4] F. Dornaika and J. Elder. Image registration for foveated omnidirectional sensor. In *Proc. ECCV*, pages 606–620, 2002.
- [5] M. Fiala and A. Basu. Feature extraction and calibration for stereo reconstruction using non-svp optics in a panoramic stereo-vision sensor. In *Proc. IEEE Int. Conf. on Pattern Recognition*, volume 4, pages 27–30, Aug. 2002.
- [6] R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 1990.
- [7] C. Geyer and K. Daniilidis. Para-catadioptric calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):687–695, 2002.
- [8] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2(14):257–262, 1992.
- [9] S. Kang. Catadioptric self-calibration. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 201–207, 2000.
- [10] S. K. Nayar. Catadioptric omnidirectional camera. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 482–480, 1997.
- [11] S. Peleg, Y. Pritch, and M. Ben-Ezra. Cameras for stereo panoramic imaging. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 208–214, Jun. 2000.
- [12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, New York, 1992.
- [13] D. Strelow, J. Mishler, D. Koes, and S. Singh. Precise omnidirectional camera calibration. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 689–694, 2000.
- [14] P. Sturm. Mixing catadioptric and perspective cameras. In *Proc. IEEE Workshop on Omnidirectional Vision*, pages 37–44, 2002.
- [15] J.-I. Takiguchi, M. Yoshida, A. Takeya, J.-I. Eino, and T. Hashizume. High precision range estimation from an omnidirectional stereo system. In *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, volume 1, pages 263–268, Sept. 2002.