MULTI-CAMERA TOPOLOGY RECOVERY USING LINES

Sang Ly[1], Cédric Demonceaux[1] and Pascal Vasseur[1,2]
[1]*MIS laboratory, University of Picardie Jules Verne, 7 rue du Moulin Neuf, 80000 Amiens, France*
[2]*Heudiasyc laboratory, University of Technology of Compiègne, Centre de Recherches de Royallieu BP 20529, 60205 Compiègne, France*

Keywords: Multi-view reconstruction, Topology recovery, Extrinsic calibration.

Abstract: We present a topology estimation approach for a system of single view point (SVP) cameras using lines. Images captured by SVP cameras such as perspective, central catadioptric or fisheye cameras are mapped to spherical images using the unified projection model. We recover the topology of a multiple central camera setup by rotation and translation decoupling. The camera rotations are first recovered from vanishing points of parallel lines. Next, the translations are estimated from known rotations and line projections in spherical images. The proposed algorithm has been validated on simulated data and real images from perspective and fisheye cameras. This vision-based approach can be used to initialize an extrinsic calibration of a hybrid camera network.

1 INTRODUCTION

Multi-camera setups, or camera networks are widely used in vision-based surveillance activities as it possesses a larger monitoring area than a single camera. Calibration is generally a critical step for any further employ of the cameras. The extrinsic calibration of a multi-camera system in order to estimate the transformations (or topology) among cameras can be divided into three steps: 1. feature detection and matching among different views, 2. initial reconstruction of multi-camera topology and 3. optimization of the reconstruction using bundle adjustment. We present in this paper a topology reconstruction approach for a system of multiple SVP cameras which can be used in the second step of a general calibration. We therefore review some related works on multi-view reconstruction approaches.

Multi-view reconstruction methods can be started with *factorization* technique. Tomasi and Kanade (Tomasi and Kanade, 1992) have proposed a factorization method to recover the scene structure and camera motion from a sequence of images. The implementation of this method is simple and provides reliable results. However, its use is limited to affine camera model and it requires that all point features be visible in all images (Hartley and Zisserman, 2003). The projective factorization, an extension of the previous one to projective camera model, has been developed in (Sturm and Triggs, 1996; Mahamud and M. Hebert, 2000). It is usually employed as an initialization for *bundle adjustment* (Triggs et al., 1999), which should be the final stage of any reconstruction algorithm (Hartley and Zisserman, 2003).

Recently, *L∞ optimization* has been proposed to solve the structure and motion problem. In (Kahl, 2005), Kahl has presented an L∞ approach based on second-order cone programming (SOCP) to estimate the camera translations and 3D points assuming known rotations. Martinec and Pajdla (Martinec and Pajdla, 2007) have solved the reconstruction problem in two stages: estimated first camera rotations linearly in least squares and then camera translations using SOCP. The main disadvantage of L∞-norm is that it is not robust to outliers (Kahl and Hartley, 2008). Method proposed in (Kahl, 2005) may fail due to a single wrong correspondence (Martinec and Pajdla, 2007).

Omnidirectional vision systems possess a wider field of view than conventional cameras. Such devices can be built up from an arrangement of several cameras or a single camera with fisheye lens or with mirrors of particular curvatures. In structure and motion problem, omnidirectional sensors play an important role as they overcome several disadvantages when working with perspective cameras, such as translation/rotation ambiguity, lack of features and the large number of views in use. In (Antone and

Ly S., Demonceaux C. and Vasseur P. (2010).
MULTI-CAMERA TOPOLOGY RECOVERY USING LINES.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 245-250
DOI: 10.5220/0002843302450250
Copyright © SciTePress

Teller, 2002), they have first estimated camera rotations using vanishing points of 3D parallel lines and then extracted camera translations using Hough transform. This method provided interesting results but might be time consuming. Moreover, two stages of their algorithm require different feature types, i.e. lines for rotation and points for translation estimations. In (Kim and Hartley, 2005), the translations among omnidirectional cameras have been estimated from known rotations and point correspondences using a constrained minimization.

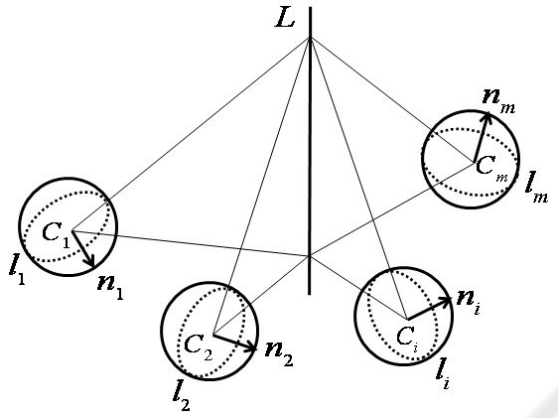


Figure 1: Multi-view geometry of spherical cameras.

In this paper, we propose a multi-view reconstruction approach in which the rotations are recovered from bundle of parallel lines and the translations are estimated from known rotations and line correspondences across multiple views. The two main contributions of this algorithm are as follows:

1. We use the unified projection model proposed by Mei (Mei, 2007). This model encompasses a large range of central projection devices including fisheye lenses. Hence, our method can be applied to any kind of SVP cameras such as perspective, central catadioptric and fisheye cameras. We can recover the topology of a hybrid camera network built up from different types of central cameras.
2. Lines are used as the primitive features. Such features are typically more stable than points and less likely to be produced by clutter or noise, especially in man-made environment (David et al., 2003). Compared to point features, lines are less numerous but more informative. They have geometrical and topological characteristics which are useful for matching (Gros et al., 1998; Bay et al., 2005). Moreover, we use uniquely lines for both rotation and translation estimations, hence optimizing the computation time of such two-stage technique.

Our approach is slightly similar to the motion recovery from multi-view tensor using lines proposed in (Gasparini and Sturm, 2008) except that we do not need to estimate such tensors but recover directly the transformations by decoupling rotation and translation.

In the following section, we develop the multi-view geometry for SVP cameras. Next, we present our topology reconstruction algorithm using lines. We show then the experimental results from simulated data and real images before the conclusions.

2 MULTI-VIEW GEOMETRY

Central imaging systems including fisheye lenses can be modelled by the unitary sphere, hence considered equivalent to spherical cameras. Noting that line correspondences can be used only in more than two views (Hartley and Zisserman, 2003), we consider a multi-camera setup composed of at least three central cameras. In (Torii et al.,), they demonstrated the bilinear and trilinear relations among spherical cameras but did not discuss any further application. In this section, we develop a similar trilinear relation which permits the computation of multi-camera topology from line correspondences.

Notation: Matrices are denoted using Sans Serif fonts, vectors in bold fonts and scalars in italics.

Consider m spherical cameras with projection centers $\mathbf{C}_i$ $(i = 1,...,m)$ as illustrated in figure 1. A line $\mathbf{L}$ in 3D space is projected to spherical images as great circles $\mathbf{l}_i$ with corresponding normals $\mathbf{n}_i$. $\mathbf{L}$ can be expressed vectorially by $\mathbf{L} = \mathbf{X}_0 + \mu\mathbf{d}$ where $\mathbf{L}, \mathbf{X}_0, \mathbf{d} \in \mathbb{R}^3$ and $\mu \in \mathbb{R}$. And $\mathbf{n}_i \in \mathbb{R}^3$ are normal correspondences in spherical images.

Let $[\mathsf{R}_i|\mathbf{t}_i]$ be the [Rotation|translation] between $\mathbf{C}_i$ and the coordinate system origin $\mathbf{O}$. Assuming that $\mathbf{C}_1$ is at $\mathbf{O}$, we have $[\mathsf{R}_1|\mathbf{t}_1] = [\mathrm{I}|0]$. As the line $\mathbf{L}$ lies on the projective planes passing through great circles $\mathbf{l}_i$ and perpendicular to normals $\mathbf{n}_i$, we obtain the next relations in which we express $\mathbf{L}$ in $\{\mathbf{O}\}$ and $\mathbf{n}_i$ in $\{\mathbf{C}_i\}$:

$$\mathbf{n}_i^T(\mathsf{R}_i\mathbf{L} + \mathbf{t}_i) = 0 \tag{1}$$

Consider a triplet of views consisting of the view 1 and two other different views a and b. We denote such triplet by (1,a,b) where $2 \leq a,b \leq m$ and $a \neq b$. The trilinear relation among three views 1, a and b is built up from equation 1 with $i = 1,a,b$ and can be rewritten as follows:

$$\mathsf{A}\hat{\mathbf{L}} = 0 \tag{2}$$

where

$$\mathsf{A} = \begin{bmatrix} \mathbf{n}_1^T & 0 \\ \mathbf{n}_a^T \mathsf{R}_a & \mathbf{n}_a^T \mathbf{t}_a \\ \mathbf{n}_b^T \mathsf{R}_b & \mathbf{n}_b^T \mathbf{t}_b \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{L}} = (\mathbf{L}^T, 1)^T$$

The existence of at least a non-zero solution in equation 2 requires that the 3x4 matrix A have rank 2. It results in a linear dependence among three rows of A. Denoting $\mathsf{A} = (\mathbf{r}_1^T, \mathbf{r}_2^T, \mathbf{r}_3^T)$, the linear relation can be written as $\mathbf{r}_1 = \alpha \mathbf{r}_2 + \beta \mathbf{r}_3$. Noting that $r_{14} = 0$, we can select the scalars $\alpha = k\mathbf{t}_b^T \mathbf{n}_b$ and $\beta = -k\mathbf{t}_a^T \mathbf{n}_a$ for some scalar k. This can be applied to the first three columns of A to obtain the next relation:

$$\mathbf{n}_1^T = \alpha \mathbf{n}_a^T \mathsf{R}_a + \beta \mathbf{n}_b^T \mathsf{R}_b \tag{3}$$

$$\mathbf{n}_1 = \alpha R_a^T \mathbf{n}_a + \beta \mathsf{R}_b^T \mathbf{n}_b \tag{4}$$

$$\mathbf{n}_1 = k\mathbf{t}_b^T \mathbf{n}_b \mathsf{R}_a^T \mathbf{n}_a - k\mathbf{t}_a^T \mathbf{n}_a \mathsf{R}_b^T \mathbf{n}_b \tag{5}$$

$$\mathsf{R}_a^T \mathbf{n}_a \mathbf{n}_b^T \mathbf{t}_b - \mathsf{R}_b^T \mathbf{n}_b \mathbf{n}_a^T \mathbf{t}_a + k_{1ab} \mathbf{n}_1 = 0 \tag{6}$$

with scalar $k_{1ab} = -1/k$. Note that k is definitely nonzero.

Equation 6 relates the normal correspondences in a triplet of views (1,*a*,*b*) to each other through the transformations $[\mathsf{R}_a|\mathbf{t}_a]$ and $[\mathsf{R}_b|\mathbf{t}_b]$ among those views.

3 TOPOLOGY RECOVERY

In this section, we present our algorithm to recover the topology of a multi-camera system by decoupling rotations and translations.

3.1 Rotation Estimation

Rotation between two SVP cameras can be estimated using vanishing point correspondences (Bazin et al., 2009). We first detect vanishing points $\mathbf{V}_i$ ($i = 1,...,m$) in all views from bundles of parallel lines and then recover all rotations R_a ($a = 2,...,m$) using the closed-form solution proposed by Horn in (Horn, 1987).

$$\mathbf{V}_a = \mathsf{R}_a \mathbf{V}_1 \tag{7}$$

3.2 Translation Estimation

The trilinear relation among three views 1, a and b in equation 6 allows the estimation of translations ($\mathbf{t}_a$ and $\mathbf{t}_b$) from rotations (R_a and R_b) and normal correspondences ($\mathbf{n}_1$, $\mathbf{n}_a$ and $\mathbf{n}_b$). With an m-camera setup, there are C_{m-1}^2 triplets of views $(1,a,b)$ or trilinear relations where C_q^p means the number of p-combinations from a set of q elements. These trilinear relations can be concatenated in a single linear system that permits the estimation of all translation $\mathbf{t}_a$ ($a = 2,...,m$) from rotation R_a and normal correspondences $\mathbf{n}_i$ ($i = 1,...,m$).

$$\mathsf{Q}\mathbf{X} = 0 \tag{8}$$

where Q is a $3\mathsf{C}_{m-1}^2$ by $(3m - 3 + \mathsf{C}_{m-1}^2)$ matrix as follows:

$$\mathsf{Q} = [\mathsf{Q}_1|\mathsf{Q}_2]$$

$$\mathsf{Q}_1 = \begin{bmatrix} -\mathsf{R}_3^T \mathbf{n}_3 \mathbf{n}_2^T & \mathsf{R}_2^T \mathbf{n}_2 \mathbf{n}_3^T & \dots & 0 \\ \dots & \dots & \dots & \\ 0 \quad \dots & -\mathsf{R}_m^T \mathbf{n}_m \mathbf{n}_{m-1}^T & \mathsf{R}_{m-1}^T \mathbf{n}_{m-1} \mathbf{n}_m^T & \end{bmatrix}$$

$$\mathsf{Q}_2 = \mathtt{diag}(\mathbf{n}_1, ..., \mathbf{n}_1)$$

$$\mathbf{X} = (\mathbf{t}_2^T, \mathbf{t}_3^T, ...\mathbf{t}_m^T, k_{123}, k_{124}, ...k_{1(m-1)m})^T$$

It can be noticed that each trilinear relation permits the estimation of two translations and different trilinear relations may contain the same translations. However, we use all C_{m-1}^2 trilinear relations as they are independent of each other. Obviously, from the diagonal part Q_2 of matrix Q, it is impossible that a trilinear relation is dependent on the others.

Given a line/normal correspondence $\mathbf{n}_i$ in m spherical views, equation 8 is a linear system in translations $\mathbf{t}_a$ ($a = 2,...,m$) and C_{m-1}^2 scalars. Each extra correspondence enlarges the matrix Q by $3\mathsf{C}_{m-1}^2$ lines and C_{m-1}^2 columns, and the unknown vector $\mathbf{X}$ by C_{m-1}^2 scalars. Therefore, n correspondences provide the following linear system:

$$\hat{\mathsf{Q}}\hat{\mathbf{X}} = 0 \tag{9}$$

where $\hat{\mathsf{Q}}$ is a $3\mathsf{C}_{m-1}^2 n$ by $(3m - 3 + \mathsf{C}_{m-1}^2 n)$ matrix and

$$\hat{\mathbf{X}} = (\mathbf{t}_2^T, ...\mathbf{t}_m^T, k_{123}^1, ...k_{1(m-1)m}^1, ...k_{123}^n, ...k_{1(m-1)m}^n)^T$$

4 EXPERIMENTAL RESULTS

4.1 Simulated Data

Since the proposed algorithm is based on line projections in spherical images, we first create 3D lines surrounding six spherical cameras $\mathbf{C}_i$ (i=1,...,6) with $\mathbf{C}_1$ at the origin of the coordinate system. The average baseline among these cameras is 2000 mm and the 3D lines are at the distance of 5000 mm to 11000 mm from the origin. These lines are mapped to spherical images as great circles and normals. Estimation algorithm in the previous section is used to recover the transformations among these cameras. The rotation estimation has been already evaluated in state of the art, therefore we focus on our translation estimation approach.

The normals are on unitary spheres, thus may be specified by elevation and azimuth angles. Gaussian noise of zero mean and varying standard deviations from 0.25 to 1.00 degrees is added to two angles of every normal. To simulate the inaccuracy in rotation estimation, the roll, pitch and yaw angles of each rotation are perturbed by Gaussian noise of zero mean and standard deviations from 0.25 to 1.00 degrees. Figure 2 shows the average angular error of the translation estimation after 1000 runs.

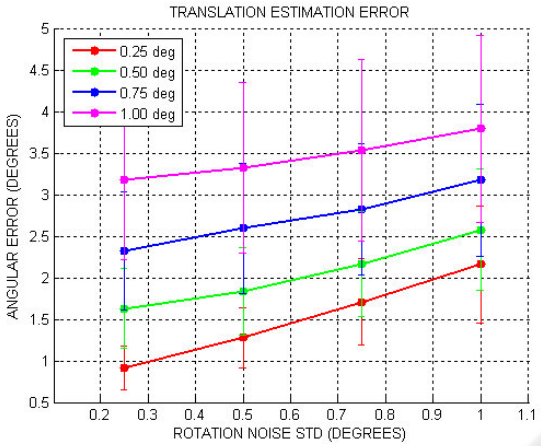


Figure 2: Translation estimation error. Normals are perturbed by Gaussian noise of zero mean and standard deviations of 0.25, 0.50, 0.75 and 1.00 degrees (corresponding to 4 curves).

4.2 Real Data

We show in this section the topology recovery of a multiple SVP camera system using line projections. In order to evaluate the topology recovery algorithm, we have used two sets of images: one captured by a perspective camera and the other by a fisheye camera.

1. *Camera calibration using the checker pattern*: we have calibrated the perspective camera using the Camera Calibration Toolbox (Bouguet,) and the fisheye camera using the Omnidirectional Calibration Toolbox (Mei,). The calibration provides not only intrinsic parameters but also extrinsic information, i.e. transformations among camera views which is useful for the evaluation of our estimation.
2. *Line extraction and matching*: in each image set, we have selected six images and performed the line detection. A fast central catadioptric line extraction method has been proposed in (Bazin et al., 2007). The extraction is composed of a splitting step and a merging step in both original and spherical images. Modifying the projection model, we extend this approach to a line detection algorithm applicable to any SVP cameras. To focus on our estimation that requires just a few number of line correspondences, line matching has been done offine and manually.
3. *Topology recovery from lines*: we have estimated the transformations among six camera views using our algorithm and then compared the recovery results with the transformations provided by the calibration in the first step.

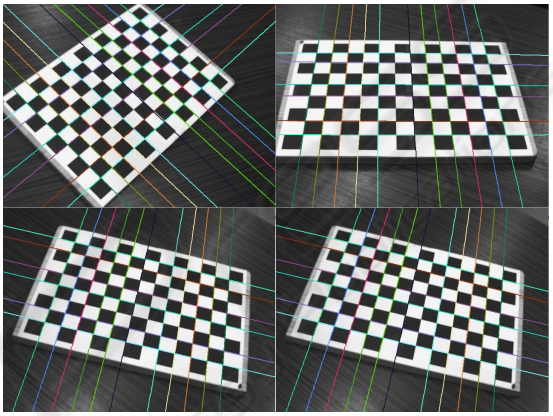


Figure 3: Four sample views captured by the perspective camera with line detection and matching.

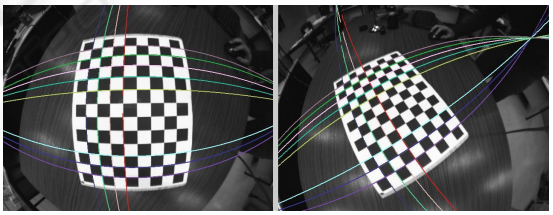


Figure 4: Two sample views captured by the fisheye camera with line detection and matching.

Figure 3 illustrates four sample views captured by a perspective camera and figure 4 illustrates two sample views captured by a fisheye camera. Line detection and matching are also illustrated. Line correspondences across multiple views are displayed in same color.

The estimated rotations among perspective and fisheye views are given in tables 1 and 2 respectively in which each rotation is represented by axis and angle of rotation.

The estimation error of translations among perspective and fisheye views is given in table 3. We have compared the direction of each recovered translation to the calibration data.

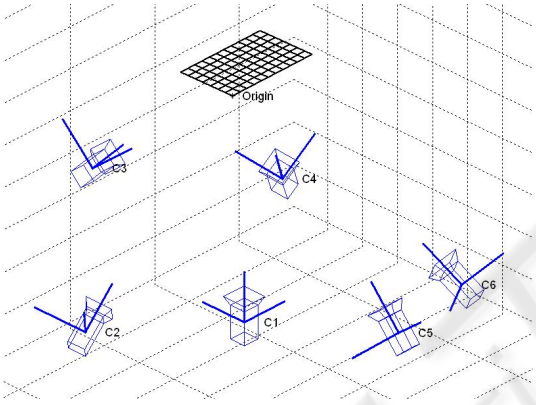
It can be seen from these tables that our recovery algorithm provides very satisfactory results. The

Table 1: Rotation estimation for perspective views.

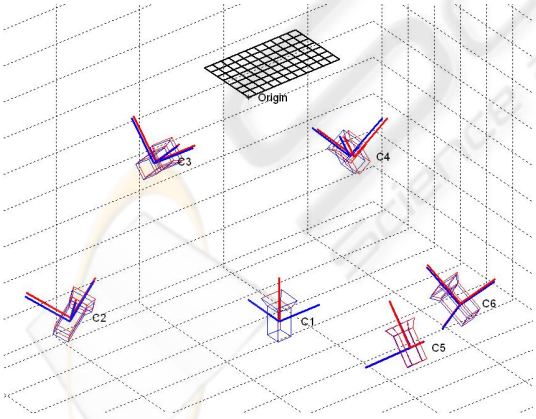
R_i	Recovery: Axis, Angle (deg)	Error (deg)
R_2	[-0.431,-0.391,-0.813]',51.721	0.585
R_3	[-0.998,-0.064,-0.017]',43.846	0.174
R_4	[-0.470,0.862,-0.189]',36.827	0.538
R_5	[0.053,-0.333,-0.942]',167.211	0.211
R_6	[-0.0866,0.377,0.922]',149.417	0.304

Table 2: Rotation estimation for fisheye views.

R_i	Recovery: Axis, Angle (deg)	Error (deg)
R_2	[-0.308,0.045,0.950]',48.029	0.052
R_3	[-0.110,0.105,0.988]',99.066	0.029
R_4	[0.060,0.096,0.994]',89.702	0.114
R_5	[-0.026,-0.117,-0.993]',131.643	0.075
R_6	[-0.477,-0.279,-0.833]',38.405	0.023



(a)



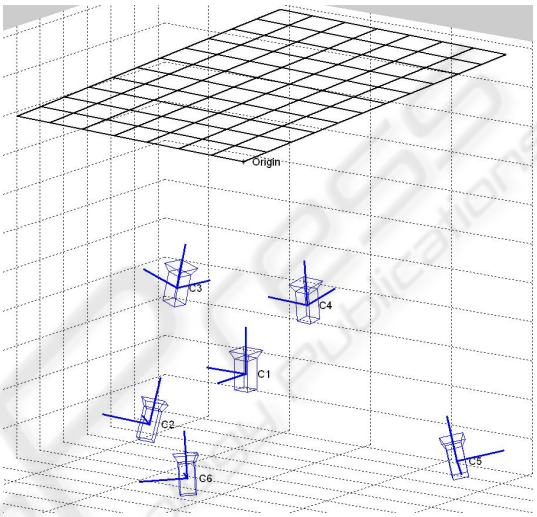
(b)

Figure 5: (a) Topology recovery of perspective cameras - (b) Comparison of our recovery result (in blue) and the extrinsic calibration data (in red).

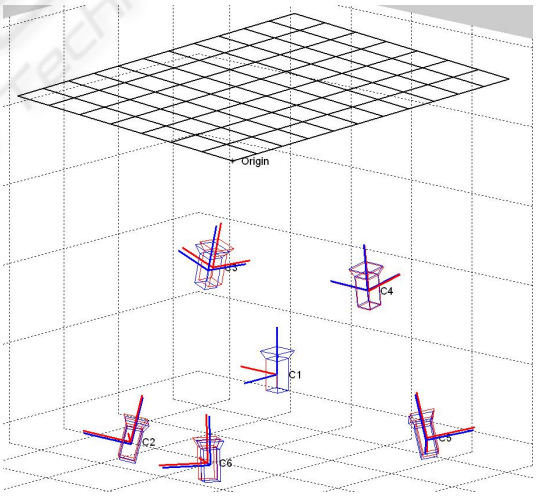
translation error is more significant than the rotation error as the translation calculation suffers from the inaccuracy of the rotation estimation and line detection.

Table 3: Translation estimation error for perspective (second column) and fisheye (third column) views.

t_i	Error (deg)-Perspective	Error (deg)-Fisheye
t_2	0.818	1.416
t_3	0.940	1.756
t_4	1.531	0.968
t_5	0.194	2.620
t_6	1.024	1.292



(a)



(b)

Figure 6: (a) Topology recovery of fisheye cameras - (b) Comparison of our recovery result (in blue) and the extrinsic calibration data (in red).

In figures 5 and 6 are the topology recovery of six perspective cameras and six fisheye cameras respectively. We have also reconstructed the calibration pattern. To illustrate the comparison of our topology recovery and the extrinsic data obtained using the cal-

ibration toolbox, we display our recovery results in blue and the extrinsic calibration results in red.

5 CONCLUSIONS

We have presented in this paper a topology recovery approach for a setup of multiple SVP cameras. We have validated our method using simulated data and real images captured by perspective and fisheye cameras. To recovery the transformations among central camera views, we first estimate the rotations using vanishing points of parallel line bundles and then the translations from known rotations and line correspondences by a linear algorithm. Using the unified projection model, this approach can be applied to a hybrid camera network built up from any kind of SVP cameras. Moreover, using line feature for both rotation and translation estimations, the proposed method promises a fast transformation recovery. We have applied this method to dissimilar types of SVP cameras and obtained very satisfied results. This would be a good initial solution for a later non-linear phase such as bundle adjustment to complete the reconstruction.

REFERENCES

M. E. Antone and S. J. Teller. Scalable extrinsic calibration of omnidirectional image networks. In *International Journal of Computer Vision (IJCV)*, vol. 49, pp. 143-174, 2002.

H. Bay, V. Ferrari and L.J. Van Gool. Wide-baseline stereo matching with line segments. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 329-336, 2005.

J. C. Bazin, C. Demonceaux and P. Vasseur. Fast Central Catadioptric Line Extraction. In *3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA*, Lecture Notes in Computer Science, vol. 4478, pp. 25-32, 2007.

J.C. Bazin, C. Demonceaux, P. Vasseur and I.S. Kweon. Motion estimation by decoupling rotation and translation in catadioptric vision. In *Computer Vision and Image Understanding (CVIU)*, 2009.

J. Y. Bouguet. Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/

P. David, D. Dementhon, R. Duraiswami and H. Samet. Simultaneous pose and correspondence determination using line features. In *CVPR*, vol. 2, pp. 424, 2003.

S. Gasparini and P. Sturm. Multi-view matching tensors from lines for general camera models. In *CVPR Workshops CVPRW '08*, pp. 1-6, 2008.

P. Gros, O. Bournez and E. Boyer. Using local planar geometric invariants to match and model images of line segments. In *CVIU*, vol. 69, no. 2, pp. 135-155, 1998.

R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2nd edition, 2003.

B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. In *Journal of the Optical Society of America. A*, vol. 4, no. 4, pp. 629-642, 1987.

F. Kahl. Multiple view geometry and the L∞-norm. In *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pp. II: 1002-1009, 2005.

F. Kahl and R. Hartley. Multiple-View Geometry Under the L_∞-Norm. In *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 30, pp. 1603-1617, 2008.

J.H. Kim and R. Hartley. Translation estimation from omnidirectional images. In *Digital Image Computing: Techniques and Applications (DICTA)*, pp. 22, 2005.

S. Mahamud and M. Hebert. Iterative projective reconstruction from multiple views. In *CVPR*, vol. 2, pp. 430-437, 2000.

D. Martinec and T. Pajdla. Robust Rotation and Translation Estimation in Multiview Reconstruction. In *CVPR*, pp. 1-8, 2007.

C. Mei. Laser-augmented omnidirectional vision for 3D localisation and mapping. PhD Thesis, INRIA Sophia Antipolis, 2007.

C. Mei. Omnidirectional Calibration Toolbox. http://www.robots.ox.ac.uk/∼cmei/Toolbox.html

C. Olson, L. Matthies, M. Schoppers and M. Maimone. Robust stereo ego-motion for long distance navigation. In *CVPR*, vol. 2, pp. 453-458, 2000.

K. Sim and R. Hartley. Recovering camera motion using the L∞-Norm. In *CVPR*, pp. 1230-1237, 2006.

P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. European Conference on Computer Vision (ECCV)*, pp. 709-720, 1996.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. In *IJCV(9)*, no. 2, pp. 137-154, 1992.

A. Torii, A. Imiya and N. Ohnishi. Two- and three- view geometry for spherical cameras. In *Proc. IEEE Workshop on Omnidirectional Vision (OMNIVIS05)*.

B. Triggs, P.F. McLauchlan, R.I. Hartley and A.W. Fitzgibbon. Bundle Adjustment - A Modern Synthesis. In *ICCV*, pp. 298-372, 1999.