A Geometric Approach to Multiple-View Triangulation for Fisheye Cameras

Graziano Chesi
The University of Hong Kong
Pokfulam Road, Hong Kong
chesi@eee.hku.hk

Abstract

It is well-known that fisheye cameras can be very useful in real applications due to their large field of view when compared to classic perspective cameras. This paper considers the fundamental problem of multiple-view triangulation for fisheye cameras. A geometric criterion is introduced in this paper for this problem, which consists of determining the scene point that minimizes the angles between the projections of the available image points and of the sought point itself. A solution based on convex optimization is proposed in this paper for this criterion, which consists of solving an eigenvalue problem. Numerical investigations carried out with synthetic and real data suggest that the proposed criterion provides significantly better estimates than algebraic criteria.

1 Introduction

Multiple-view triangulation is the process that attempts to recover a scene point from its available image projections on two or more cameras located in the scene. Due to image noise and calibration errors, this process generally provides an estimate only of the sought point, which depends on the criterion chosen to match the available image points with the image projections of the estimate on all the cameras.

For perspective cameras, multiple-view triangulation has been studied for a long time. The contributions typically consider a geometric criterion for defining the estimate of the sought point since geometric criteria generally provide more accurate results than algebraic ones. A commonly adopted geometric criterion is the minimization of the reprojection error in the L_2 norm, for which several solutions have been proposed. In [4], the authors show how the exact solution of triangulation with two-views can be obtained by computing the roots of a one-variable polynomial of degree six. For triangulation with three-views, the exact solution is obtained in [6] by solving a system of polynomial equations through methods from computational commutative algebra, and in [1] through Groebner basis techniques. Multiple-view triangulation is considered also in [5] via branch-and-bound algorithms and in [2] via semidefinite programming. Other geometric criteria include the minimization of the reprojection error in the L_∞ norm.

This paper considers the problem of multiple-view triangulation in a vision system with fisheye cameras. Indeed, as it is well-known, fisheye cameras can be very useful in real applications due to their large field of view when compared to classic perspective cameras. Fisheye cameras can be modeled through a spherical projection followed by a perspective one. Hence, a geometric criterion for defining the solution of the multiple-view triangulation problem is introduced in this paper, which consists of determining the scene point that minimizes the angles between the projections on the sphere of the available image points and of the sought point itself. This criterion can be regarded as an alternative L_2 norm criterion for perspective and non-perspective cameras, since the angles are proportional to the euclidean distances between the available and the sought projections on the spheres. A solution based on convex optimization is proposed in this paper for this criterion, which consists of solving an eigenvalue problem, i.e. the minimization of a linear function subject to linear matrix inequality constraints. Numerical investigations carried out with synthetic and real data suggest that the proposed criterion provides significantly better estimates than algebraic criteria.

2 Proposed Approach

Notation: $\mathbf{M}^T$: transpose of matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$; $\mathbf{I}_n$: $n \times n$ identity matrix; $\mathbf{0}_n$: $n \times 1$ null vector; $\mathbf{e}_i$: i-th column of $\mathbf{I}_3$; $SO(3)$: set of all 3×3 rotation matrices; $SE(3)$: $SO(3) \times \mathbb{R}^3$; $\|\mathbf{v}\|$: 2-norm of $\mathbf{v} \in \mathbb{R}^n$; s.t.: subject to.

Let $F_i = (\mathbf{O}_i, \mathbf{c}_i) \in SE(3)$ denote the coordinate frame of the i-th fisheye camera, where the rotation matrix $\mathbf{O}_i \in SO(3)$ defines the orientation and the vector $\mathbf{c}_i \in \mathbb{R}^3$ defines the position expressed with respect to a common reference coordinate frame $F^{ref} \in SE(3)$. Each fisheye camera consists of a spherical projection followed by a perspective projection, see e.g. [3, 7]. The center of the sphere coincides with $\mathbf{c}_i$ while the center of the perspective camera is given by

$$\mathbf{d}_i = \mathbf{c}_i - \xi_i \mathbf{O}_i \mathbf{e}_3 \tag{1}$$

where $\xi_i \in \mathbb{R}$ is the distance between $\mathbf{c}_i$ and $\mathbf{d}_i$. Let

$$\mathbf{X} = (x, y, z)^T \tag{2}$$

denote a generic scene point, where $x, y, z \in \mathbb{R}$ are expressed with respect to F^{ref}. The projection of $\mathbf{X}$ onto the image plane of the i-th fisheye camera in pixel coordinates is denoted by $\mathbf{p}_i \in \mathbb{R}^{3\times 3}$ and is given by

$$\mathbf{p}_i = \mathbf{K}_i \mathbf{x}_i \tag{3}$$

where $\mathbf{K}_i \in \mathbb{R}^{3\times 3}$ is the upper triangular matrix containing the intrinsic parameters of the i-th fisheye camera, and $\mathbf{x}_i \in \mathbb{R}^{3\times 3}$ is $\mathbf{p}_i$ expressed in normalized coordinates. The image point $\mathbf{x}_i$ is the perspective projection of the spherical projection of $\mathbf{X}$. Specifically, the spherical projection of $\mathbf{X}$ is given by

$$\mathbf{X}_i = \mathbf{A}_i(\mathbf{X}) \tag{4}$$

where

$$\mathbf{A}_i(\mathbf{X}) = \frac{\mathbf{O}_i^T(\mathbf{X}-\mathbf{c}_i)}{\left\|\mathbf{O}_i^T(\mathbf{X}-\mathbf{c}_i)\right\|}, \tag{5}$$

while the perspective projection of $\mathbf{X}_i$ is given by

$$\mathbf{x}_i = \mathbf{B}_i(\mathbf{X}_i) \tag{6}$$

where

$$\mathbf{B}_i(\mathbf{X}_i) = \frac{1}{\mathbf{e}_3^T\mathbf{X}_i + \xi_i\|\mathbf{X}_i\|}\begin{pmatrix} \mathbf{e}_1^T\mathbf{X}_i \\ \mathbf{e}_2^T\mathbf{X}_i \\ \mathbf{e}_3^T\mathbf{X}_i + \xi_i\|\mathbf{X}_i\| \end{pmatrix}. \tag{7}$$

The solution for $\mathbf{p}_i$ in (3) as a function of $\mathbf{X}$ is denoted by

$$\mathbf{p}_i = \mathbf{\Phi}_i(\mathbf{X}). \tag{8}$$

The multiple-view triangulation problem for fisheye cameras consists of estimating $\mathbf{X}$ from estimates of the image points $\mathbf{p}_i$ (denoted by $\hat{\mathbf{p}}_i$) and functions $\mathbf{\Phi}_i$ (denoted by $\hat{\mathbf{\Phi}}_i$), $i=1,\ldots,N$, where N is the number of fisheye cameras:

$$\text{given } \left\{\left(\hat{\mathbf{p}}_i, \hat{\mathbf{\Phi}}_i\right),\ i=1,\ldots,N\right\},\ \text{estimate } \mathbf{X}. \tag{9}$$

In order to address this problem, we introduce a geometric criterion which consists of determining the scene point that minimizes the angles between the projections on the sphere of the available image points and of the sought point itself. In particular, we define this criterion according to

$$\begin{array}{rcl} \hat{\mathbf{X}} & = & \arg\max_{\mathbf{X}} g \\ & & \text{s.t. } \mathbf{e}_3^T\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i) \geq 0 \ \ \forall i=1,\ldots,N \end{array} \tag{10}$$

where the constraint ensures that the estimated point lies in front of the cameras, and $g \in \mathbb{R}$ is the reprojection error

$$g = \frac{1}{N}\sum_{i=1,\ldots,N} \hat{\mathbf{a}}_i^T\hat{\mathbf{b}}_i \tag{11}$$

where $\hat{\mathbf{a}}_i \in \mathbb{R}^3$ is the estimated spherical projection of $\mathbf{X}$ on the i-th fisheye camera (expressed in F_i), and $\hat{\mathbf{b}}_i \in \mathbb{R}^3$ is the back projection of $\hat{\mathbf{p}}_i$ on the sphere of this camera (expressed in F_i), i.e.

$$\begin{array}{rcl} \hat{\mathbf{a}}_i & = & \hat{\mathbf{A}}_i(\mathbf{X}) \\ \hat{\mathbf{b}}_i & = & \hat{\mathbf{B}}_i^{-1}(\hat{\mathbf{p}}_i). \end{array} \tag{12}$$

In fact, since $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ are unitary norm vectors, it follows that

$$\hat{\mathbf{a}}_i^T\hat{\mathbf{b}}_i = \cos\theta_i \tag{13}$$

where $\theta_i \in [-\pi, \pi)$ is the angle between $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{b}}_i$ in the plane containing these vectors. The expression for $\hat{\mathbf{a}}_i$ is given by (5) replacing $\mathbf{O}_i$ and $\mathbf{c}_i$ with their available estimates, i.e.

$$\hat{\mathbf{a}}_i = \frac{\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i)}{\left\|\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i)\right\|}, \tag{14}$$

while the expression for $\hat{\mathbf{b}}_i$ is obtained inverting (3) and (7). The criterion (10) can be regarded as an alternative L_2 norm criterion for perspective and non-perspective cameras, since the angles are proportional to the euclidean distances between the available and the sought projections on the spheres.

Let us now address the solution of (10). The optimal cost of (10) is given by

$$\begin{array}{l} \max_{\mathbf{X}} \dfrac{1}{N}\displaystyle\sum_{i=1,\ldots,N}\left(\frac{\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i)}{\left\|\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i)\right\|}\right)^T\hat{\mathbf{b}}_i \\ \text{s.t. } \mathbf{e}_3^T\hat{\mathbf{O}}_i^T(\mathbf{X}-\hat{\mathbf{c}}_i) \geq 0 \ \ \forall i=1,\ldots,N. \end{array} \tag{15}$$

As it can be seen from (15), finding the solution of the proposed criterion requires the maximization of an irrational function in the variable $\mathbf{X}$. In order to address this optimization problem, we propose the use of convex programming as follows.

First, let us rewrite (15) in the equivalent form

$$\begin{array}{l} \sup_{\mathbf{z}} h(\mathbf{z}) \\ \text{s.t. } \left\{\begin{array}{rcl} l_i(\mathbf{z}) & = & 0 \\ m_j(\mathbf{z}) & \geq & 0 \\ \forall i=1,\ldots,N & & \\ \forall j=1,\ldots,2N & & \end{array}\right. \end{array} \tag{16}$$

where $\mathbf{z}$ is the new variable defined by $\mathbf{X}$ and slack variables $y_1,\ldots,y_N \in \mathbb{R}$ as

$$\mathbf{z} = (\mathbf{X}, y_1, \ldots, y_N)^T, \tag{17}$$

and $h(\mathbf{z})$, $l_i(\mathbf{z})$ and $m_j(\mathbf{z})$ are polynomial functions of $\mathbf{z}$. In order to solve (16), we define the new optimization problem

$$\begin{array}{l} \min_{q,\mathbf{r}_i,\mathbf{s}_j} q \\ \text{s.t. } \left\{\begin{array}{rcl} p(\mathbf{z}) & \geq & 0 \\ s_j(\mathbf{z}) & \geq & 0 \\ \forall j=1,\ldots,2N & & \\ \forall \mathbf{z} \in \mathbb{R}^{3+N} & & \end{array}\right. \end{array} \tag{18}$$

with

$$p(\mathbf{z}) = q - h(\mathbf{z}) - \sum_{i=1}^{N} r_i(\mathbf{z})l_i(\mathbf{z}) - \sum_{j=1}^{2N} s_j(\mathbf{z})m_j(\mathbf{z}) \tag{19}$$

where $q \in \mathbb{R}$ is an auxiliary variable, and $r_1(\mathbf{z}),\ldots,r_N(\mathbf{z})$ and $s_1(\mathbf{z}),\ldots,s_{2N}(\mathbf{z})$ are auxiliary polynomial variables of fixed degrees.

Second, let us express the polynomial $p(\mathbf{z})$ as

$$p(\mathbf{z}) = \mathbf{t}_0(\mathbf{z})^T(\mathbf{P} + \mathbf{L}(\boldsymbol{\alpha}))\mathbf{t}_0(\mathbf{z}) \tag{20}$$

where $\mathbf{t}_0(\mathbf{z})$ is a vector containing monomials in $\mathbf{z}$, $\mathbf{P}$ is a symmetric matrix, and $\mathbf{L}(\boldsymbol{\alpha})$ is a linear parametrization of the set

$$\mathcal{L} = \left\{\mathbf{L} = \mathbf{L}^T:\ \mathbf{t}_0(\mathbf{z})^T\mathbf{L}\mathbf{t}_0(\mathbf{z}) = 0\right\}. \tag{21}$$

Similarly, the polynomials $r_i(\mathbf{z})$ and $s_j(\mathbf{z})$ can be expressed as

$$\begin{array}{rcl} r_i(\mathbf{z}) & = & \mathbf{R}_i^T\mathbf{t}_i(\mathbf{z}) \\ s_j(\mathbf{z}) & = & \mathbf{t}_{N+j}(\mathbf{z})^T\mathbf{S}_j\mathbf{t}_{N+j}(\mathbf{z}) \end{array} \tag{22}$$

where $\mathbf{R}_i$ is a vector, $\mathbf{S}_j$ is a symmetric matrix, $\mathbf{t}_1(\mathbf{z}),\ldots,\mathbf{t}_{3N}(\mathbf{z})$ are vectors containing monomials in

$\mathbf{z}$. In order to get the sought candidate estimate of the criterion (15), we define the final optimization problem

$$\begin{array}{ll} \min\limits_{q,\mathbf{R}_i,\mathbf{S}_j,\boldsymbol{\alpha}} q & \\ \text{s.t.} & \left\{ \begin{array}{rcl} \mathbf{P}+\mathbf{L}(\boldsymbol{\alpha}) & \geq & 0 \\ \mathbf{S}_j & \geq & 0 \\ \forall j = 1,\ldots,2N. & & \end{array} \right. \end{array} \tag{23}$$

Problem (23) consists of minimizing a linear cost function subject to linear matrix inequality constraints, and belongs to the class of eigenvalue problems. Eigenvalue problems are convex optimization problems since the cost function is linear and hence convex, and the feasible set is the intersection of feasible sets of LMIs which are convex. Several toolboxes have been developed for solving eigenvalue problems, based for instance on interior-point methods.

Solving the eigenvalue problem (23) provides the sought candidate optimal cost of the criterion (15). From the optimal values of the variables of (23), one can obtain the sought candidate estimate of the criterion (15). Indeed, let $\mathbf{T}$ be the matrix $\mathbf{P}+\mathbf{L}(\boldsymbol{\alpha})$ evaluated for such optimal values, and let $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ be the SVD of $\mathbf{T}$, i.e.

$$\mathbf{T} = \mathbf{U}\mathbf{S}\mathbf{V}^T. \tag{24}$$

Let $\mathbf{v}$ be the last column of $\mathbf{V}$, and let be $v_0,\ldots,v_3$ be the entries of $\mathbf{v}$ in the positions of the monomials 1, x, y and z in the vector $\mathbf{t}_0(\mathbf{z})$ (remember that $x = \mathbf{e}_1^T\mathbf{X}$, $y = \mathbf{e}_2^T\mathbf{X}$ and $z = \mathbf{e}_3^T\mathbf{X}$ from (2)). The candidate for $\hat{\mathbf{X}}$ in (10) is given by

$$\bar{\mathbf{X}} = \frac{1}{v_0}(v_1,v_2,v_3)^T. \tag{25}$$

Indeed, at optimality, one obtains $p(\mathbf{z}) = 0$ for $\mathbf{X}$ replaced by $\hat{\mathbf{X}}$ since $p(\mathbf{z})$ and $s_j(\mathbf{z})m_j(\mathbf{z})$ are nonnegative. Consequently, one has that $\mathbf{t}_0(\mathbf{z})^T\mathbf{T}\mathbf{t}_0(\mathbf{z}) = 0$, and since $\mathbf{T}$ is positive semidefinite, it follows that $\mathbf{t}_0(\mathbf{z})$ is an eigenvector of the null eigenvalue of $\mathbf{T}$.

3 Examples

3.1 Synthetic Data

Here we present some results obtained with synthetic data. Specifically, we have generated 500 vision systems, each of them composed by a scene point to reconstruct (denoted hereafter as $\mathbf{X}$) and 4 fisheye cameras with 180 degrees-field of view, in particular with parameter $\xi = 0.5$. For each vision system, $\mathbf{X}$ and the centers of the cameras are randomly chosen in a sphere of radius 100 centered in the origin of the reference frame, while the orientation matrices of the cameras are randomly chosen under the constraint that $\mathbf{X}$ is visible by the cameras. Figure 1 shows some of the vision systems and corresponding image projections.

In order to generate the corrupted data, we have:

- added random variables in the interval $[-\eta,\eta]$ pixels to each coordinate of the image points, where $\eta \in \mathbb{R}$ defines the noise intensity;
- multiplied ξ and each intrinsic parameter times random variables in the interval $[1-\eta/100, 1+\eta/100]$;
- multiplied the camera centers and the angles of the rotation matrices times random variables in the interval $[1-\eta/100, 1+\eta/100]$.

Then, we have repeated the triangulation for 3 numbers of available cameras (i.e., 2, 3 and 4) and for 4 values of noise intensity (i.e., $\eta = 0.5, 1, 1.5, 2$), hence solving a total number of $3 \times 4 \times 500 = 6000$ triangulation problems.

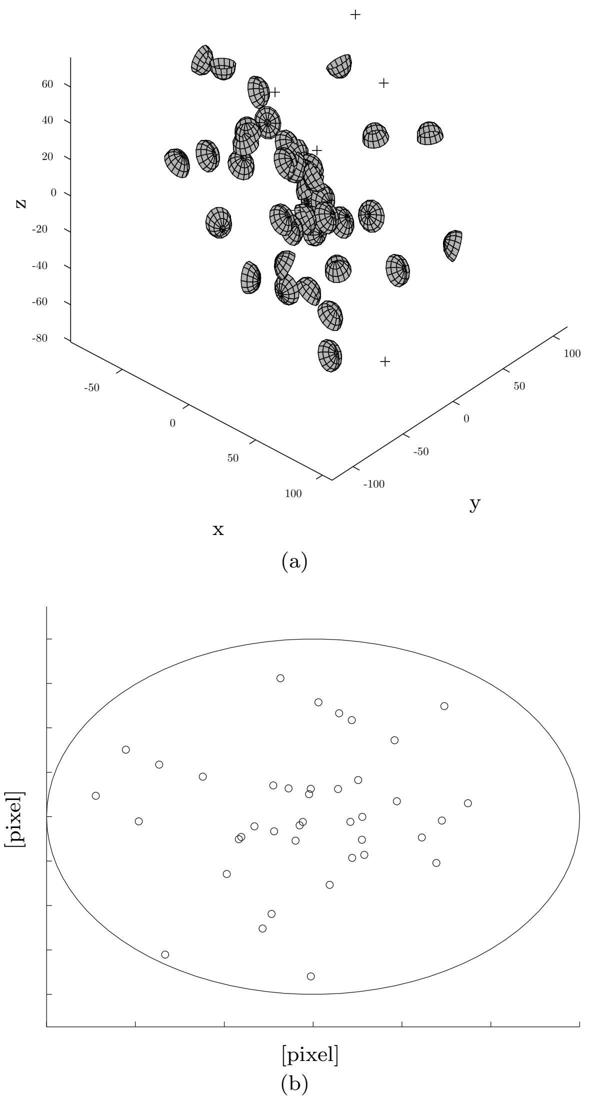


Figure 1. Synthetic data: (a) scene points ("+" marks) and fisheye cameras for 10 of the 500 vision systems; (b) image projections of such scene points ("o" marks) and boundary of the visible region (solid line).

Table 1 shows the average 3D error obtained with the proposed method (denoted by "this"), i.e. $\|\bar{\mathbf{X}} - \mathbf{X}\|$ where $\bar{\mathbf{X}}$ is given by (25). For comparison, we have computed also the average 3D error obtained by minimizing the algebraic error with standard linear least-squares (denoted by "algebraic"). As we can see, the proposed method provides quite better estimates.

method \ η	0.5	1	1.5	2
	$N = 2$ (2000 points)			
this	1.9145	3.592	5.8116	7.8517
algebraic	2.021	3.6212	6.1174	8.1519

method \ η	0.5	1	1.5	2
	$N = 3$ (2000 points)			
this	0.96464	1.9393	2.7042	3.6104
algebraic	1.0846	2.1688	3.0144	4.1615

method \ η	0.5	1	1.5	2
	$N = 4$ (2000 points)			
this	0.80961	1.5383	2.1309	2.9188
algebraic	0.9066	1.7793	2.4941	3.4809

Table 1. Results with synthetic data: average 3D error for different number of fisheye cameras (N) and noise intensity (η).

3.2 Real Data

Here we present some results obtained using the Wadham college sequence available at `http://www.robots.ox.ac.uk/~vgg/data/data-mview.html`. This sequence consists of 5 views taken with a perspective camera, the projection matrices, and 3019 image points corresponding to 1331 scene points. In particular:

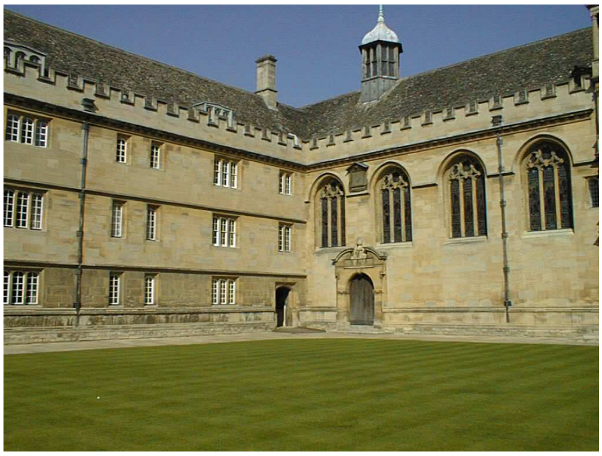
- 1052 points are visible in 2 views;
- 215 points are visible in 3 views;
- 50 points are visible in 4 views;
- 14 points are visible in 5 views.

First, we have estimated the 1331 scene points using standard triangulation for perspective cameras. Second, we have computed the projections of these scene points onto fisheye cameras with same orientation and same center except for a translation along the optical axis in order to enlarge the spanned image area. Figure 2 shows one image of the sequence and the corresponding projections onto the fisheye camera. The data obtained so far will be used as "true" data. Third, we have corrupted the true data as done in the previous subsection for the case of synthetic data with noise intensity $\eta = 1$. Fourth, we have repeated the triangulation using for each scene point the maximum number of cameras where the point is visible.

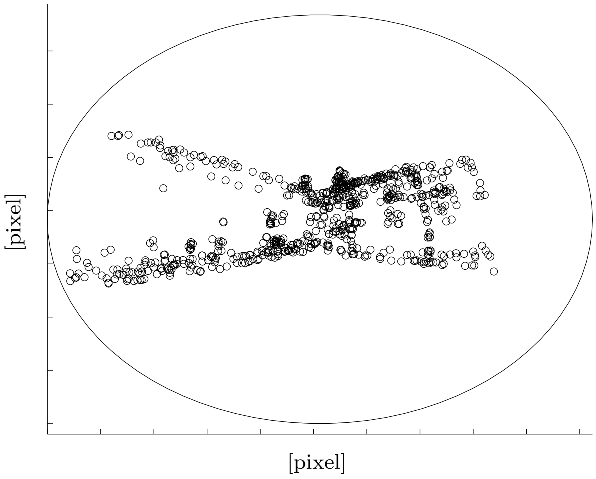
The average 3D error obtained with the proposed method is 3.3579, while the one obtained by minimizing the algebraic error is 5.1525.

4 Conclusion

We have introduced a geometric criterion for multiple-view triangulation in a vision system with fisheye cameras, which consists of determining the scene point that minimizes the angles between the projections of the available image points and of the sought point itself. For this problem we have proposed a solution based on convex optimization, which consists of solving an eigenvalue problem. The obtained results indicate that the proposed criterion provides significantly better estimates than algebraic criteria.



(a)



(b)

Figure 2. Wadham college sequence: (a) one of the 5 images; (b) projections on fisheye camera.

References

[1] M. Byrod, K. Josephson, and K. Astrom. Fast optimal three view triangulation. In *Asian Conference on Computer Vision*, volume 4844 of *LNCS*, pages 549–559, Tokyo, Japan, 2007.

[2] G. Chesi and Y. S. Hung. Fast multiple-view L2 triangulation with occlusion handling. *Computer Vision and Image Understanding*, 115(2):211–223, 2011.

[3] C. Geyer and K. Daniilidis. Mirrors in motion: epipolar geometry and motion estimation. *International Journal on Computer Vision*, 45(3):766–773, 2003.

[4] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[5] F. Lu and R. Hartley. A fast optimal algorithm for l_2 triangulation. In *Asian Conference on Computer Vision*, volume 4844 of *LNCS*, pages 279–288, Tokyo, Japan, 2007.

[6] H. Stewenius, F. Schaffalitzky, and D. Nister. How hard is 3-view triangulation really? In *International Conference on Computer Vision*, pages 686–693, Beijing, China, 2005.

[7] S. Thirthala and M. Pollefeys. Multi-view geometry of 1D radial cameras and its application to omnidirectional camera calibration. In *International Conference on Computer Vision*, pages 1539–1546, 2005.