On the Multiple-view Triangulation Problem with Perspective and Non-perspective Cameras
A Virtual Reprojection-based Approach

Graziano Chesi
Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam Road, Hong Kong

Keywords: Vision System, Multiple-view, Perspective Camera, Non-perspective Camera, Triangulation.

Abstract: This paper considers the multiple-view triangulation problem in a vision system with perspective and non-perspective cameras. In particular, cameras that can be modeled through a spherical projection followed by a perspective one, such as perspective cameras and fisheye cameras, are considered. For this problem, an approach based on reprojecting the available image points onto virtual image planes is proposed, which has the advantage of transforming the original problem into a new one for which the existing methods for multiple-view triangulation with perspective cameras can be used. In particular, algebraic and geometric errors of such methods are now evaluated on the virtual image planes, and the solution of the new problem exactly approaches the sought scene point as image noise and calibration errors tend to zero. The proposed approach is illustrated by several numerical investigations with synthetic and real data.

1 INTRODUCTION

It is well-known that the multiple-view triangulation problem is of fundamental importance in computer vision and robotics. Specifically, this problem consists of recovering a scene point from its available image projections on two or more cameras located in the scene. Unfortunately, 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. The multiple-view triangulation problem has numerous key applications, such as 3D object reconstruction, map estimation, and visual servo control, see for instance (Hartley and Zisserman, 2000; Faugeras and Luong, 2001; Chesi and Vicino, 2004; Chesi and Hung, 2007).

The multiple-view triangulation problem with perspective cameras has been studied for a long time, and numerous contributions can be found in the literature. Pioneering contributions have considered the minimization of algebraic errors for defining the estimate of the sought point, since the resulting optimization problems can be solved via linear least-squares, while later contributions have proposed the minimization of geometric errors since they can generally provide more accurate estimates, see for instance (Hartley and Zisserman, 2000) about the definition of algebraic and geometric errors. A commonly adopted geometric error is the L2 norm of the reprojection error, for which several solutions have been proposed. In (Hartley and Sturm, 1997; Hartley and Zisserman, 2000), 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 (Stewenius et al., 2005) by solving a system of polynomial equations through methods from computational commutative algebra, and in (Byrod et al., 2007) through Groebner basis techniques. Multiple-view triangulation is considered also in (Lu and Hartley, 2007) via branch-and-bound algorithms, and in (Chesi and Hung, 2011) via convex programming. Other geometric errors include the infinity norm of the reprojection error, see for instance (Hartley and Schaffalitzky, 2004).

This paper considers the multiple-view triangulation problem in a vision system with perspective and non-perspective cameras, hereafter simply denoted as generalized cameras. In particular, cameras that can be modeled through a spherical projection followed by a perspective one, such as perspective cameras and fisheye cameras, are considered by exploiting a unified camera model. An approach based on reprojecting the available image points onto virtual image

Chesi G..
On the Multiple-view Triangulation Problem with Perspective and Non-perspective Cameras - A Virtual Reprojection-based Approach.
DOI: 10.5220/0003981300050013
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2012), pages 5-13
ISBN: 978-989-8565-22-8

planes is hence proposed for the multiple-view triangulation problem, which has the advantage of transforming such a problem into a new one for which the existing methods for multiple-view triangulation with perspective cameras can be used. In particular, algebraic and geometric errors of such methods are now evaluated on the virtual image planes, and the solution of the new problem exactly approaches the sought scene point as image noise and calibration errors tend to zero. The proposed approach is illustrated by several numerical investigations with synthetic and real data.

The paper is organized as follows. Section 2 provides some preliminaries and the problem formulation. Section 3 describes the proposed approach. Section 4 shows the results with synthetic and real data. Lastly, Section 5 concludes the paper with some final remarks.

2 PRELIMINARIES

The notation adopted throughout the paper is as follows:

- $\mathbf{M}^T$: 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 us denote the coordinate frame of the i-th generalized camera as

$$F_i = (\mathbf{O}_i, \mathbf{c}_i) \in SE(3) \tag{1}$$

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 generalized camera consists of a spherical projection followed by a perspective projection. The center of the sphere coincides with $\mathbf{c}_i$ while the center of the perspective camera is given by

$$\mathbf{d}_i = \mathbf{c}_i - \xi_i \mathbf{O}_i \mathbf{e}_3 \tag{2}$$

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

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{3}$$

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 generalized camera in pixel coordinates is denoted by $\mathbf{p}_i \in \mathbb{R}^{3\times 3}$ and is given by

$$\mathbf{p}_i = \mathbf{K}_i \mathbf{x}_i \tag{4}$$

where $\mathbf{K}_i \in \mathbb{R}^{3\times 3}$ is the upper triangular matrix containing the intrinsic parameters of the i-th generalized 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

$$\mathbf{X}_i = \mathbf{A}_i(\mathbf{X}) \tag{5}$$

where

$$\mathbf{A}_i(\mathbf{X}) = \frac{\mathbf{O}_i^T (\mathbf{X} - \mathbf{c}_i)}{\left\| \mathbf{O}_i^T (\mathbf{X} - \mathbf{c}_i) \right\|}, \tag{6}$$

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

$$\mathbf{x}_i = \mathbf{B}_i(\mathbf{X}_i) \tag{7}$$

where

$$\mathbf{B}_i(\mathbf{X}_i) = \frac{1}{\mathbf{e}_3^T \mathbf{X}_i + \xi_i \|\mathbf{X}_i\|} \begin{pmatrix} \mathbf{e}_1^T \mathbf{X}_i \\ \mathbf{e}_2^T \mathbf{X}_i \\ \mathbf{e}_3^T \mathbf{X}_i + \xi_i \|\mathbf{X}_i\| \end{pmatrix}. \tag{8}$$

The solution for $\mathbf{p}_i$ in (4) as a function of $\mathbf{X}$ is denoted by

$$\mathbf{p}_i = \Phi_i(\mathbf{X}). \tag{9}$$

Figure 1 illustrates the spherical projection and the perspective projection just described for the i-th generalized camera of the vision system.

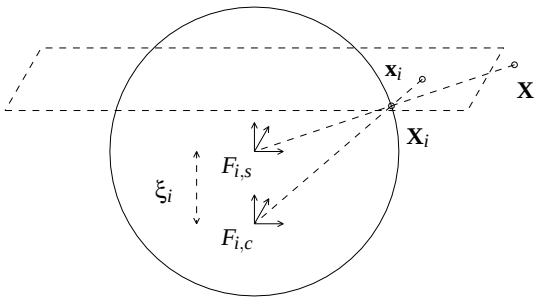

Figure 1: A point $\mathbf{X}$ is firstly projected on the the point $\mathbf{X}_i$ according to a spherical projection (frame $F_{i,s}$). Then, the point $\mathbf{X}_i$ is projected on the image point $\mathbf{x}_i$ (in normalized coordinates) according to a perspective projection (frame $F_{i,c}$). The image point $\mathbf{p}_i$ (in pixel coordinates) is hence obtained as $\mathbf{p}_i = \mathbf{K}_i \mathbf{x}_i$. The distance between $F_{i,s}$ and $F_{i,c}$ is ξ_i, while the distance between $F_{i,c}$ and the plane where $\mathbf{x}_i$ lies is 1.

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

$$\text{given } \left\{ \left(\hat{\mathbf{p}}_i, \hat{\Phi}_i(\cdot) \right), \; i = 1, \ldots, N \right\}, \text{ estimate } \mathbf{X}. \tag{10}$$

3 PROPOSED APPROACH

Let us start by observing that existing methods for triangulation with perspective cameras cannot be used to estimate $\mathbf{X}$ with the image points $\mathbf{p}_i$ (clearly, unless $\xi_i = 0$ for all generalized cameras, since in such a case the cameras are perspective ones). This is due to the fact that, as it can be seen from Figure 1, the scene point $\mathbf{X}$ does not lie on the line connecting the image point $\mathbf{x}_i$ (i.e., $\mathbf{p}_i$ expressed in normalized coordinates instead of pixel coordinates) to the center of its projection (i.e., $F_{i,c}$).

The idea proposed in this paper consists of reprojecting the image points $\mathbf{p}_i$ onto virtual image planes, one per camera, in order to obtain new image points for which this problem does not occur. This can be done by determining the intersections of the lines connecting the scene point $\mathbf{X}$ to the centers of the spherical projections (i.e., $F_{i,s}$) with virtual image planes V_i. Figure 2 illustrates this procedure for the i-th camera. The virtual image plane V_i is chosen for convenience parallel to the true image plane of the camera and at a unitary distance from the center of $F_{i,s}$. The new image point is denoted by $\mathbf{y}_i$. Let us observe that $\mathbf{y}_i$ exists whenever $\mathbf{X}$ has a positive depth in the frame $F_{i,s}$, while $\mathbf{y}_i$ tends to infinity as this depth tends to zero.

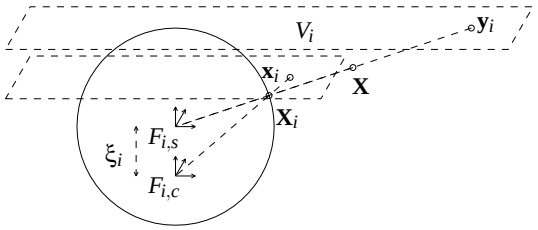

Figure 2: The image point $\mathbf{x}_i$ (in normalized coordinates) is reprojected onto the virtual image plane V_i, firstly, by determining the intersection $\mathbf{X}_i$ of the line connecting $\mathbf{x}_i$ to $F_{i,c}$ with the sphere, and secondly, by determining the intersection of the line connecting $\mathbf{X}_i$ to $F_{i,s}$ with V_i. The new image point is denoted by $\mathbf{y}_i$.

In order to derive the expression of the new image point $\mathbf{y}_i$ on the virtual image plane V_i, let us proceed as follows. First, let us recover the expression of $\mathbf{p}_i$ in normalized coordinates, i.e. $\mathbf{x}_i$. This is given by

$$\mathbf{x}_i = \mathbf{K}_i^{-1}\mathbf{p}_i. \tag{11}$$

Second, let us express $\mathbf{x}_i$ as

$$\mathbf{x}_i = \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \tag{12}$$

where $u_i, v_i \in \mathbb{R}$. The line connecting $\mathbf{x}_i$ to $F_{i,c}$ can be parametrized with respect to the frame $F_{i,s}$ as

$$\mathbf{l}_i(\alpha) = \begin{pmatrix} \alpha u_i \\ \alpha v_i \\ -\xi_i + \alpha \end{pmatrix} \tag{13}$$

where $\alpha \in \mathbb{R}$. The spherical projection of $\mathbf{X}$ is hence given by the intersection of l_i with the sphere, i.e.

$$\mathbf{X}_i = \mathbf{l}_i(\alpha^*) \tag{14}$$

where α^* is the solution of

$$\alpha^* : \begin{cases} \|\mathbf{l}_i(\alpha)\| = 1 \\ \begin{pmatrix} u_i \\ v_i \\ 1-\xi_i \end{pmatrix}^T \mathbf{l}_i(\alpha) > 0. \end{cases} \tag{15}$$

This solution is given by

$$\alpha^* = \frac{\xi_i + \delta_i}{1 + u_i^2 + v_i^2} \tag{16}$$

where

$$\delta_i = \sqrt{1 + (1 - \xi_i^2)(u_i^2 + v_i^2)}. \tag{17}$$

Then, the line connecting $\mathbf{X}_i$ to $F_{i,s}$ can be parametrized with respect to the frame $F_{i,s}$ as

$$\mathbf{m}_i(\beta) = \beta \mathbf{X}_i \tag{18}$$

where $\beta \in \mathbb{R}$. The intersection of this line with the virtual image plane V_i is hence given by

$$\mathbf{y}_i = \mathbf{m}_i(\beta^*) \tag{19}$$

where β^* is the solution of

$$\beta^* : \ \mathbf{e}_3^T \mathbf{m}_i(\beta) = 1. \tag{20}$$

This solution is given by

$$\beta^* = \frac{1 + u_i^2 + v_i^2}{\delta_i - \xi_i(u_i^2 + v_i^2)}. \tag{21}$$

It is possible to verify that the overall expression of $\mathbf{y}_i$ in normalized coordinates is given by

$$\mathbf{y}_i = \begin{pmatrix} \gamma_i u_i \\ \gamma_i v_i \\ 1 \end{pmatrix} \tag{22}$$

where γ_i is defined as

$$\gamma_i = \frac{1 + \xi_i \delta_i}{1 - \xi_i^2(u_i^2 + v_i^2)}. \tag{23}$$

We denote the expression of $\mathbf{y}_i$ as a function of $\mathbf{p}_i$ according to

$$\mathbf{y}_i = \Omega_i(\mathbf{p}_i). \tag{24}$$

Let us observe that $\mathbf{y}_i$ exists whenever

$$\xi_i^2(u_i^2 + v_i^2) \neq 1 \tag{25}$$

i.e. whenever $\mathbf{X}$ has a positive depth in the frame $F_{i,s}$.

The procedure just described assumes that $\mathbf{p}_i$ and $\Omega_i(\cdot)$ are known. However, in real situations this is clearly not true due to the presence of uncertainties, and hence the new image points have to be defined using the available data. In particular, $\mathbf{p}_i$ is replaced by $\hat{\mathbf{p}}_i$, while $\Omega_i(\cdot)$ is replaced by $\hat{\Omega}_i(\cdot)$ which is obtained as in (22)–(24) by replacing u_i, v_i, γ_i, δ_i and ξ_i with their available estimates $\hat{u}_i$, $\hat{v}_i$, $\hat{\gamma}_i$, $\hat{\delta}_i$ and $\hat{\xi}_i$, respectively. The estimates of the new image points are given by

$$\hat{\mathbf{y}}_i = \hat{\Omega}_i(\hat{\mathbf{p}}_i). \tag{26}$$

In order to estimate the scene point $\mathbf{X}$, let us observe that the new image points satisfy the relationship

$$\lambda_i \mathbf{y}_i = \mathbf{P}_i \mathbf{X} \tag{27}$$

where $\lambda_i \in \mathbb{R}$ and $\mathbf{P}_i \in \mathbb{R}^{3\times 4}$ is the projection matrix given by

$$\mathbf{P}_i = \begin{pmatrix} \mathbf{R}_i & \mathbf{t}_i \end{pmatrix} \tag{28}$$

where the rotation matrix $\mathbf{R}_i \in SO(3)$ and the translation vector $\mathbf{t}_i \in \mathbb{R}^3$ are given by

$$\begin{aligned} \mathbf{R}_i &= \mathbf{O}_i^T \\ \mathbf{t}_i &= -\mathbf{O}_i^T \mathbf{c}_i. \end{aligned} \tag{29}$$

Hence, (27) can be rewritten as

$$\mathbf{y}_i = \frac{1}{\mathbf{e}_3^T \mathbf{P}_i \mathbf{X}} \mathbf{P}_i \mathbf{X}. \tag{30}$$

The relationship (30) can be exploited to estimate $\mathbf{X}$ from the available estimates of the new image points $\hat{\mathbf{y}}_i$. In the sequel we discuss two criteria for this estimation.

The first method that we consider is based on the estimation of $\mathbf{X}$ by minimizing the algebraic error in the relationship (30). Specifically, according to this method, the estimate of $\mathbf{X}$ is obtained through the linear least-squares problem

$$\min_{\mathbf{Y}} \ \hat{c}_{alg}(\mathbf{Y}) \tag{31}$$

where $\mathbf{Y} \in \mathbb{R}^3$ and

$$\hat{c}_{alg}(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \begin{pmatrix} \mathbf{e}_1^T \hat{\mathbf{P}}_i \mathbf{Y} - \hat{\gamma}_i \hat{u}_i \mathbf{e}_3^T \hat{\mathbf{P}}_i \mathbf{Y} \\ \mathbf{e}_2^T \hat{\mathbf{P}}_i \mathbf{Y} - \hat{\gamma}_i \hat{v}_i \mathbf{e}_3^T \hat{\mathbf{P}}_i \mathbf{Y} \end{pmatrix} \right\|^2 \tag{32}$$

and $\hat{\mathbf{P}}_i$ is the available estimate of $\mathbf{P}_i$. The solution of linear least-squares problems can be obtained either in closed form or through a singular value decomposition (SVD). Indeed, let us define

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{e}_1^T \hat{\mathbf{R}}_1 \\ \mathbf{e}_2^T \hat{\mathbf{R}}_1 \\ \vdots \\ \mathbf{e}_N^T \hat{\mathbf{R}}_N \\ \mathbf{e}_N^T \hat{\mathbf{R}}_N \end{pmatrix}, \ \hat{\mathbf{b}} = \begin{pmatrix} \hat{\gamma}_1 \hat{u}_1 \mathbf{e}_3^T \hat{\mathbf{t}}_1 \\ \hat{\gamma}_1 \hat{v}_1 \mathbf{e}_3^T \hat{\mathbf{t}}_1 \\ \vdots \\ \hat{\gamma}_N \hat{u}_N \mathbf{e}_3^T \hat{\mathbf{t}}_N \\ \hat{\gamma}_N \hat{v}_N \mathbf{e}_3^T \hat{\mathbf{t}}_N \end{pmatrix}. \tag{33}$$

It follows that (31) can be rewritten as

$$\min_{\mathbf{Y}} \left\| \hat{\mathbf{A}} \mathbf{Y} - \hat{\mathbf{b}} \right\|^2. \tag{34}$$

The minimizer of (34), denoted by $\hat{\mathbf{X}}_{alg}$, is given by

$$\hat{\mathbf{X}}_{alg} = \left(\hat{\mathbf{A}}^T \hat{\mathbf{A}} \right)^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{b}}. \tag{35}$$

Alternatively, one can get this minimizer by introducing the SVD

$$\hat{\mathbf{U}} \hat{\mathbf{S}} \hat{\mathbf{V}}^T = \begin{pmatrix} \hat{\mathbf{A}} & -\hat{\mathbf{b}} \end{pmatrix} \tag{36}$$

and by defining

$$\hat{\mathbf{X}}_{alg} = \frac{\hat{\mathbf{v}}_a}{\hat{v}_b} \tag{37}$$

where $\hat{\mathbf{v}}_a \in \mathbb{R}^3$ is the vector with the first three entries of the last column of $\hat{\mathbf{V}} \in \mathbb{R}^4$ and $\hat{v}_b \in \mathbb{R}$ is the fourth entry of such a column.

The second method that we consider is based on the estimation of $\mathbf{X}$ by minimizing the L2 norm of the reprojection error in the relationship (30). Specifically, according to this method, the estimate of $\mathbf{X}$ is obtained through the optimization problem

$$\min_{\mathbf{Y}} \ \hat{c}_{L2}(\mathbf{Y}) \tag{38}$$

where

$$\hat{c}_{L2}(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \begin{pmatrix} \mathbf{e}_1^T \hat{\Psi}_i(\mathbf{Y}) - \hat{\gamma}_i \hat{u}_i \\ \mathbf{e}_2^T \hat{\Psi}_i(\mathbf{Y}) - \hat{\gamma}_i \hat{v}_i \end{pmatrix} \right\|^2 \tag{39}$$

and the function $\hat{\Psi}_i(\cdot)$ is the available estimate of the function $\Psi_i(\cdot)$ which defines the solution for $\mathbf{y}_i$ in (30) as a function of $\mathbf{X}$, i.e.

$$\mathbf{y}_i = \Psi_i(\mathbf{X}). \tag{40}$$

We denote the minimizer of (38) as

$$\hat{\mathbf{X}}_{L2} = \arg\min_{\mathbf{Y}} \ \hat{c}_{L2}(\mathbf{Y}). \tag{41}$$

The computation of this minimizer can be addressed in various ways. For instance, in (Chesi and Hung, 2011) a technique based on convex programming has been proposed recently, which provides a candidate of the sought solution and a simple test for establishing its optimality. See also the other techniques described in the introduction.

It is important to observe that the two methods just described provide estimates of the sought scene point by minimizing an error (either algebraic or geometric) defined for the new image points $\hat{\mathbf{y}}_i$. This means that such an error is evaluated on the virtual image planes V_i unless the cameras are perspective (in such a case, in fact, the virtual image planes V_i coincide with the image planes of the cameras). Let us also

observe that the estimates provided by these methods approach the sought scene point as image noise and calibration errors tend to zero (clearly, if enough information is available for triangulation).

In the sequel we denote the 3D estimation errors achieved by minimizing the algebraic error in (31) and by minimizing the L2 norm of the reprojection error in (38) as

$$\begin{array}{rcl} d_{alg} & = & \|\hat{\mathbf{X}}_{alg} - \mathbf{X}\| \\ d_{L2} & = & \|\hat{\mathbf{X}}_{L2} - \mathbf{X}\|. \end{array} \quad (42)$$

4 EXAMPLES

In this section we present some results obtained with synthetic and real data. The minimization of the algebraic error in (31) is solved through (35), while the minimization of the L2 norm of the reprojection error in (38) is solved with the TFML method described in (Chesi and Hung, 2011) available at http://www.eee.hku.hk/∼chesi. In both cases, the data are pre-elaborated in order to work with normalized data.

4.1 Example 1

Let us consider a vision system composed by three generalized cameras with 180 degrees-field of view defined by

$$\forall i = 1,2,3 \quad \left\{ \begin{array}{rcl} \mathbf{K}_i & = & \begin{pmatrix} 200 & 0 & 400 \\ 0 & 200 & 400 \\ 0 & 0 & 1 \end{pmatrix} \\ \xi_i & = & 0.5 \\ \mathbf{O}_i & = & e^{[\theta_i]_\times} \end{array} \right.$$

and

$$\begin{array}{rcl} \theta_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & , & \mathbf{c}_1 = \begin{pmatrix} -9 \\ 4 \\ 1 \end{pmatrix} \\ \theta_2 = \begin{pmatrix} -\pi/2 \\ 0 \\ 0 \end{pmatrix} & , & \mathbf{c}_2 = \begin{pmatrix} 3 \\ -1 \\ -7 \end{pmatrix} \\ \theta_3 = \begin{pmatrix} 0 \\ -\pi/3 \\ \pi/2 \end{pmatrix} & , & \mathbf{c}_3 = \begin{pmatrix} 1 \\ 7 \\ 6 \end{pmatrix}. \end{array}$$

The scene point is

$$\mathbf{X} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

and the corresponding image points in pixel coordinates are given by

$$\mathbf{p}_1 = \begin{pmatrix} 677.926 \\ 344.415 \\ 1 \end{pmatrix}, \ \mathbf{p}_2 = \begin{pmatrix} 351.895 \\ 159.473 \\ 1 \end{pmatrix}$$
$$\mathbf{p}_3 = \begin{pmatrix} 133.527 \\ 465.346 \\ 1 \end{pmatrix}.$$

Figure 3 shows the three cameras and the scene point, while Figures 4a–4c show the image points and the boundary of the visible region in each camera.

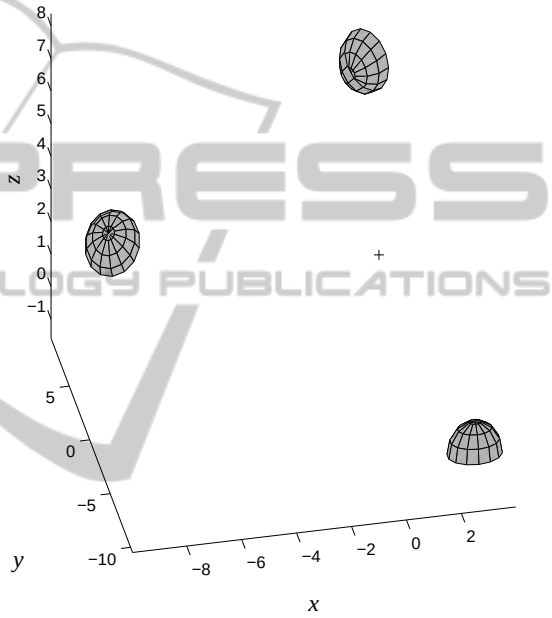

Figure 3: Example 1: the three cameras and the scene point ("+" mark).

In this example we want to consider the presence of image noise on the available image points. To this end, we define the available image points as

$$\hat{\mathbf{p}}_i = \mathbf{p}_i + \eta \mathbf{n}_i \ \ \forall i = 1,2,3$$

where η is a parameter defining the image noise intensity and

$$\mathbf{n}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{n}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \ \mathbf{n}_3 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

The problem consists of estimating $\mathbf{X}$ for η varying in the interval $[0,6]$ pixels.

We repeat the multiple-view triangulation procedure described in the previous section for a grid of values η in $[0,6]$. Figure 5 shows the obtained estimates by minimizing the algebraic error and by minimizing the L2 norm of the reprojection error. In particular, for $\eta = 6$, the 3D estimation errors achieved

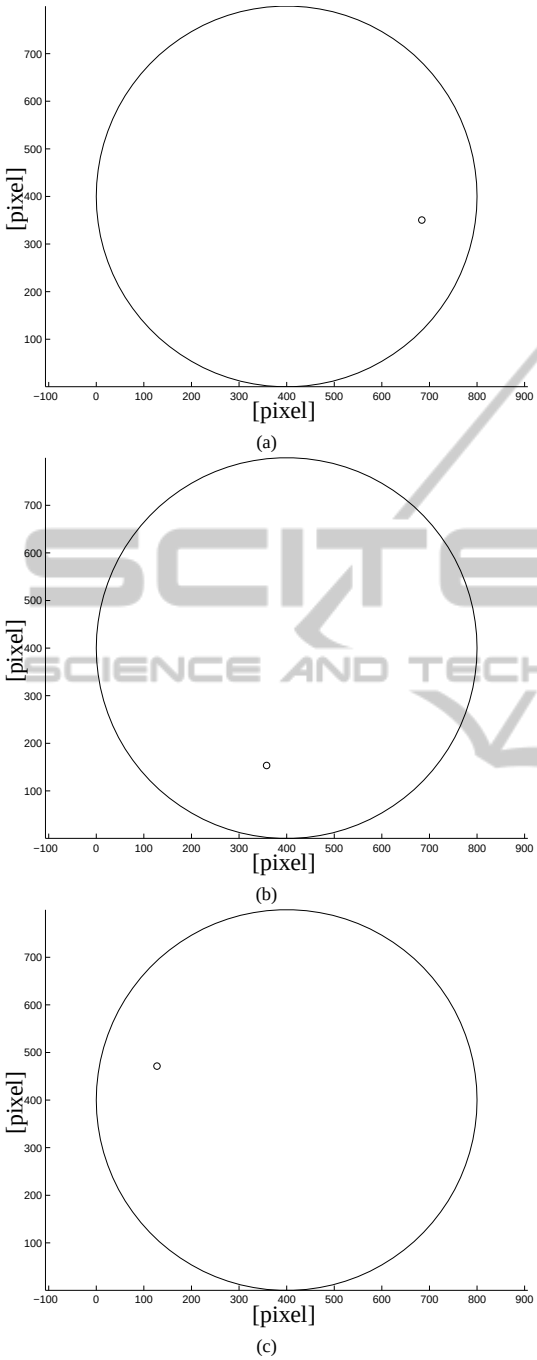

Figure 4: Example 1: image points ("o" marks) and boundary of the visible region (solid line) for each camera.

by the two methods are

$$d_{alg} = 0.231, \quad d_{L2} = 0.139.$$

As we can see, minimizing the algebraic error provides quite worse estimates than minimizing the L2 norm of the reprojection error in this example. Interesting, the next examples will show that the situation is generally different.

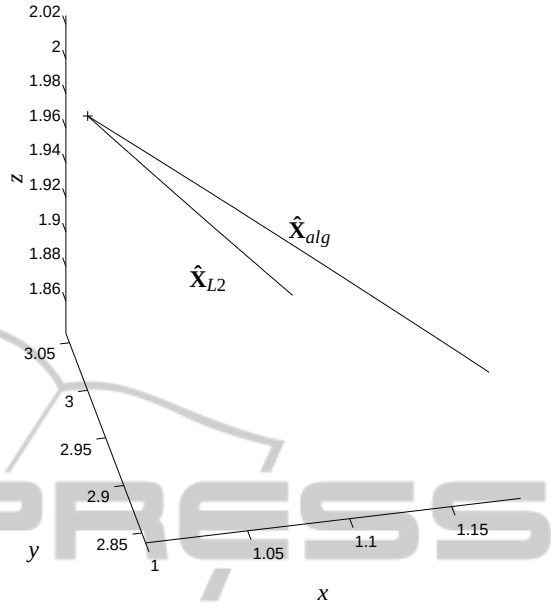

Figure 5: Example 1: solutions $\hat{\mathbf{X}}_{alg}$ and $\hat{\mathbf{X}}_{L2}$ for $\eta \in [0,6]$.

4.2 Example 2: Statistics with 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 generalized cameras with 180 degrees-field of view, in particular with intrinsic parameters given by

$$\forall i = 1,\ldots,4 \begin{cases} \mathbf{K}_i &= \begin{pmatrix} 300 & 0 & 600 \\ 0 & 200 & 400 \\ 0 & 0 & 1 \end{pmatrix} \\ \xi_i &= 0.5. \end{cases}$$

For each vision system, $\mathbf{X}$ and the centers of the cameras are randomly chosen in a sphere of radius 500 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 6a shows the scene points and the generalized cameras for 10 of the 500 vision systems, while Figure 6b shows the image points and the boundary of the visible region in these cameras.

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 $[1-\eta/100, 1+\eta/100]$;

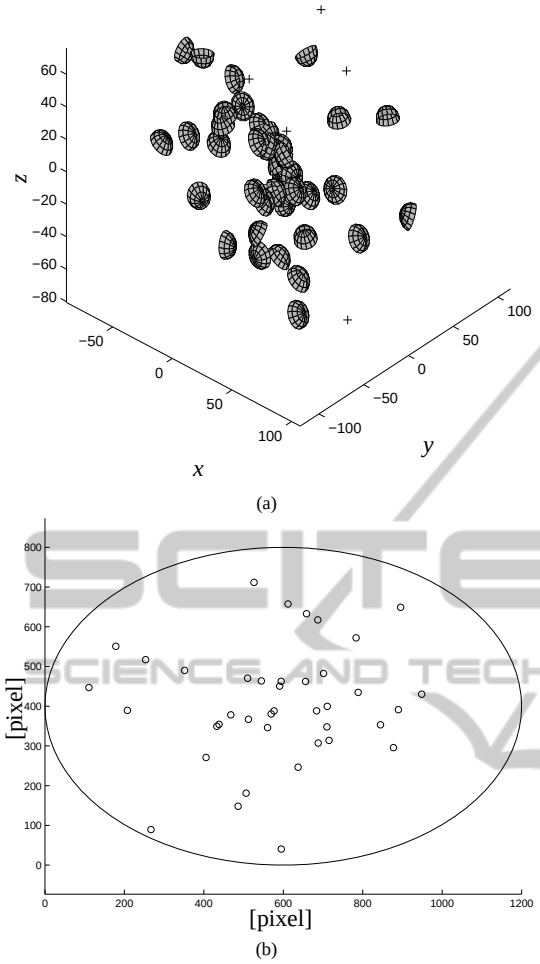

Figure 6: Example 2 (synthetic data): (a) scene points ("+" marks) and generalized cameras for 10 of the 500 vision systems; (b) image projections of such scene points ("o" marks) and boundary of the visible region (solid line).

- multiplied the camera centers and the angles of the rotation matrices times random variables in the interval $[1-\eta/100, 1+\eta/100]$.

Hence, 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. Table 1 shows the average values of d_{alg} and d_{L2} denoted by "alg" and "L2", respectively.

4.3 Example 3: Wadham College Sequence

Lastly, we present some results obtained with almost real data. In fact, we do not have real data for a non-perspective camera, moreover with real data it is impossible to know the true scene points that we would like to use for evaluation. Hence, we have considered the Wadham college sequence available at the webpage of the Visual Geometry Group of Oxford University, 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 of such views, and 3019 image points corresponding to 1331 scene points visible in at least 2 of such views (with known correspondence).

Table 1: Example 2 (synthetic data): average 3D error for different number of generalized cameras (N) and noise intensity (η).

$N = 2$ (2000 points)

method \ η	0.5	1	1.5	2
alg	2.021	3.6212	6.1174	8.1519
L2	2.0181	4.0248	6.1692	8.42

$N = 3$ (2000 points)

method \ η	0.5	1	1.5	2
alg	1.0846	2.1688	3.0144	4.1615
L2	1.0817	2.0342	3.047	4.2966

$N = 4$ (2000 points)

method \ η	0.5	1	1.5	2
alg	0.9066	1.7793	2.4941	3.4809
L2	0.94721	1.7375	2.5739	3.8066

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, which are shown in Figure 9. Second, we have computed the projections of these scene points onto generalized cameras with same orientation, same center except for a translation along the optical axis in order to enlarge the spanned image area, and intrinsic parameters given by

$$\forall i = 1, \ldots, 5 \quad \begin{cases} \mathbf{K}_i & = & \begin{pmatrix} 256 & 0 & 512 \\ 0 & 192 & 384 \\ 0 & 0 & 1 \end{pmatrix} \\ \xi_i & = & 0.5. \end{cases}$$

Figures 7–8 show the first image and last one of the 5 images, the corresponding extracted points, and the same points after transforming and shifting the cameras.

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

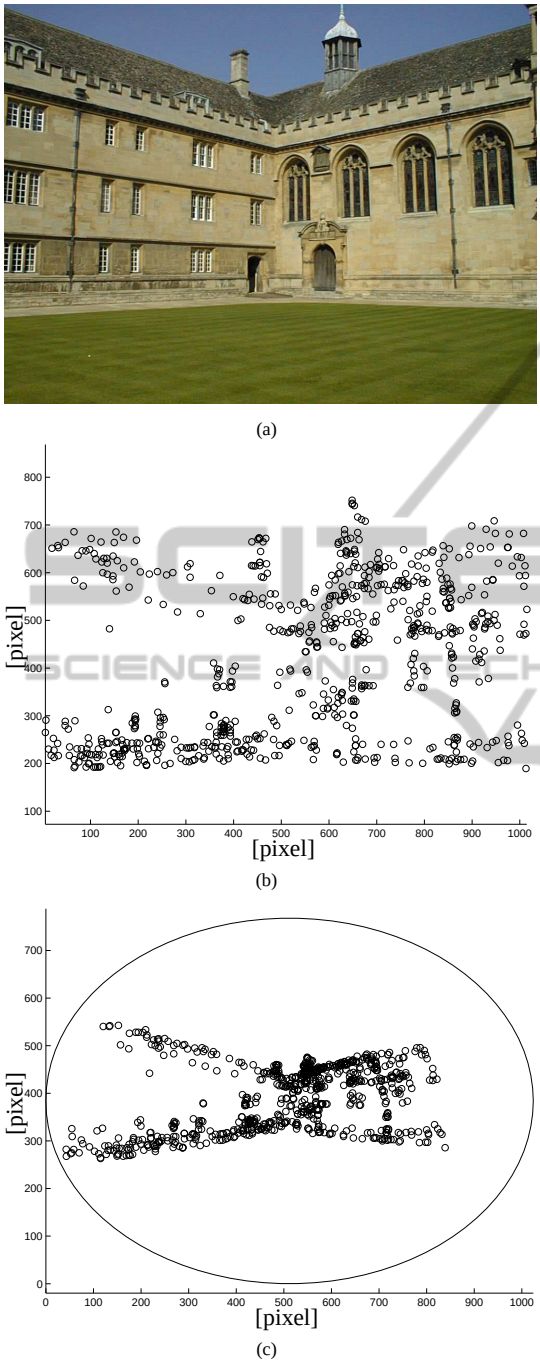

Figure 7: Example 3 (Wadham college sequence): (a) first image of the sequence; (b) points extracted in such an image; (c) same points after transforming and shifting the cameras.

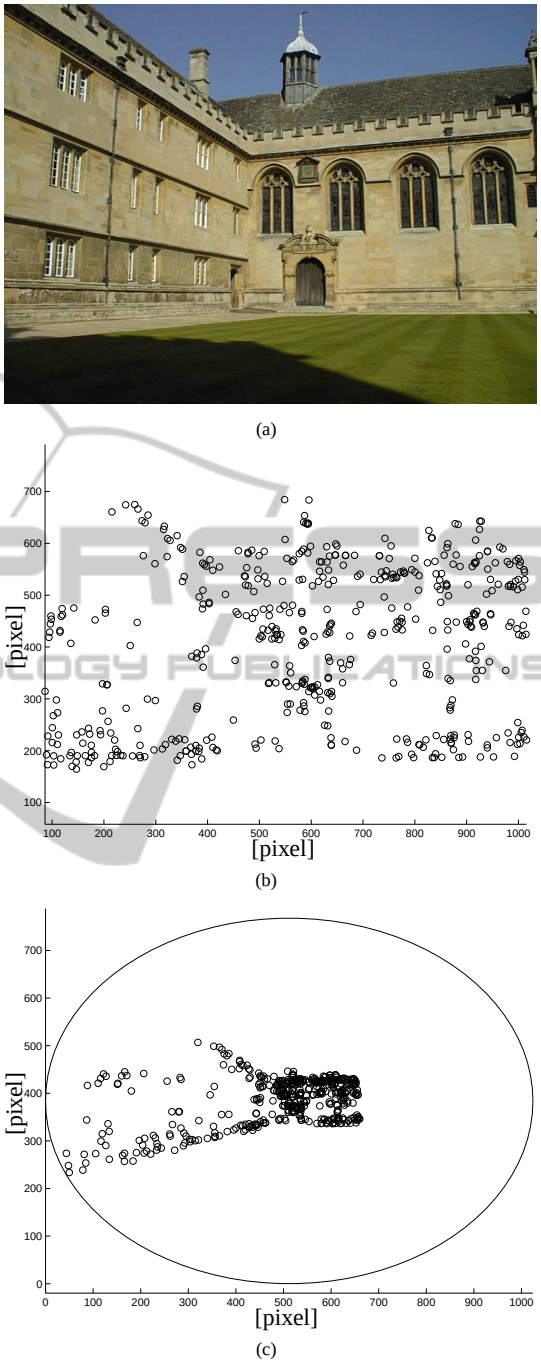

Figure 8: Example 3 (Wadham college sequence): (a) last image of the sequence; (b) points extracted in such an image; (c) same points after transforming and shifting the cameras.

the triangulation using for each scene point the maximum number of cameras where the point is visible. Table 2 shows the average values of d_{alg} and d_{L2} denoted by "alg" and "L2", respectively.

5 CONCLUSIONS

We have addressed the multiple-view triangulation problem in a vision system with perspective and

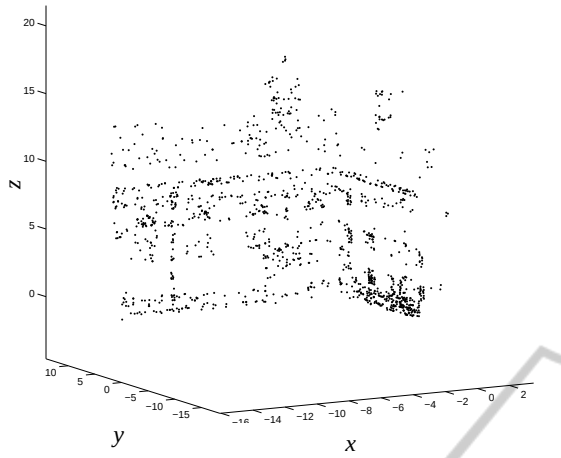

Figure 9: Example 3 (Wadham college sequence): estimated scene points.

Table 2: Example 3 (Wadham college sequence): average 3D error for different number of generalized cameras (N).

N	alg	L2
2	6.2038	8.4202
3	1.2768	1.3405
4	1.0436	0.93116
5	0.34739	0.3462

non-perspective cameras, and we have proposed an approach based on reprojecting the available image points onto virtual image planes. This approach has the advantage of transforming the original problem into a new one for which the existing methods for multiple-view triangulation with perspective cameras can be used. In particular, algebraic and geometric errors of such methods are now evaluated on the virtual image planes, and the solution of the new problem exactly approaches the sought scene point as image noise and calibration errors tend to zero.

The obtained numerical results suggest that minimizing the simple algebraic error on the virtual image planes can provide competitive estimates compared with those provided by the minimization of the L2 norm of the reprojection error on such planes. This is indeed interesting, and it is probably due to the different meaning that the L2 norm assumes when evaluated for the new image points. Future work will investigate this aspect.

REFERENCES

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

Chesi, G. and Hung, Y. S. (2007). Global path-planning for constrained and optimal visual servoing. *IEEE Trans. on Robotics*, 23(5):1050–1060.

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

Chesi, G. and Vicino, A. (2004). Visual servoing for large camera displacements. *IEEE Trans. on Robotics*, 20(4):724–735.

Faugeras, O. and Luong, Q.-T. (2001). *The Geometry of Multiple Images*. MIT Press, Cambridge (Mass.).

Hartley, R. and Schaffalitzky, F. (2004). l_∞ minimization in geometric reconstruction problems. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 504–509, Washington, USA.

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

Hartley, R. and Zisserman, A. (2000). *Multiple view in computer vision*. Cambridge University Press.

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

Stewenius, H., Schaffalitzky, F., and Nister, D. (2005). How hard is 3-view triangulation really? In *Int. Conf. on Computer Vision*, pages 686–693, Beijing, China.