Camera Pose Estimation Using First-Order Curve Differential Geometry

Ricardo Fabbri, Benjamin B. Kimia
Brown University
Division of Engineering
Providence RI 02912, USA
{rfabbri,kimia}@lems.brown.edu
and
Peter J. Giblin
University of Liverpool
Liverpool, UK
pjgiblin@liverpool.ac.uk

ECCV 2012 Accepted Paper Draft

Abstract. This paper considers and solves the problem of estimating camera pose given a pair of point-tangent correspondences between the 3D scene and the projected image. The problem arises when considering curve geometry as the basis of forming correspondences, computation of structure and calibration, which in its simplest form is a point augmented with the curve tangent. We show that while the standard resectioning problem is solved with a minimum of three points given the intrinsic parameters, when points are augmented with tangent information only two points are required, leading to substantial computational savings, *e.g.*, when used as a minimal engine within a RANSAC strategy. In addition, computational algorithms are developed to find a practical and efficient solution which is shown to effectively recover camera pose using synthetic and realistic datasets. The resolution of this problem is intended as a basic building block of future curve-based structure from motion systems, allowing new views to be incrementally registered to a core set of views for which relative pose has already been computed.

1 Introduction

A key problem in the reconstruction of structure from multiple views is the determination of relative pose among cameras as well as the intrinsic parameters for each camera. The classical method is to rely on a set of corresponding points across views to determine each camera's intrinsic parameter matrix $\mathcal{K}_{im}$ as well as the relative pose between pairs of cameras [11]. The set of corresponding points can be determined using a calibration jig, but, more generally, using isolated keypoints such as Harris corners [10] or SIFT/HOG [17] features which remain somewhat stable over view and other variations. As long as there is a sufficient number of keypoints between two views, a random selection of a few feature correspondences using RANSAC [7, 11] can be verified by measuring the number of inlier features. This class of isolated feature point-based methods are currently in popular and successful use through packages such as the Bundler and used in applications such as Phototourism [1].

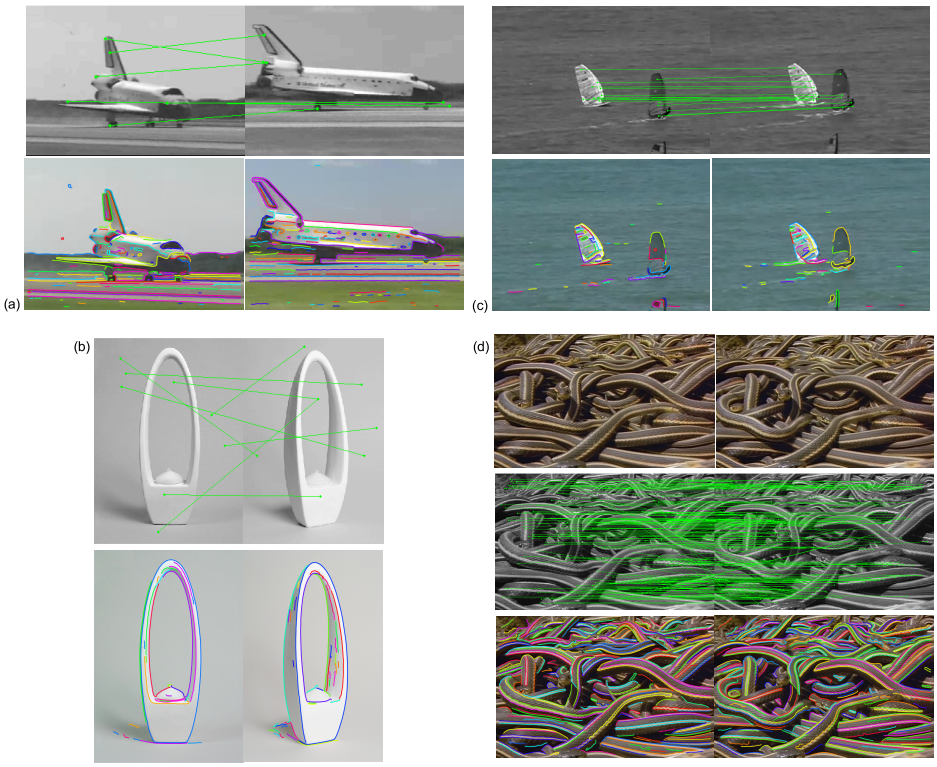

Fig. 1. (a) Views with wide baseline separation may not have a sufficient number of interest points in common, but they often do share common curve structure. (b) There may not always be sufficient interest points matching across views of homogeneous objects, such as for the sculpture, but there is sufficient curve structure. (c) Each moving object requires its own set of features, but they may not be sufficient without a richly textured surface. (d) Non-rigid structures face the same issue.

Two major drawbacks limit the applicability of automatic methods based on interest points. First, it is well-known that in practice the correlation of interest points works for views with a *limited baseline*, according to some estimates no greater than 30° [18], Figure 1(a). In contrast, certain image curve fragments, *e.g.*, those corresponding to sharp ridges, reflectance curves, *etc*, persist stably over a much larger range of views. Second, the success of interest point-based methods is based on the presence of an abundance of features so that a sufficient number of them survive the various variations between views. While this is true in many scenes, as evidenced by the popularity of this approach, in a non-trivial number of scenes this is not the case, such as *(i)* Homogeneous regions, *e.g.*, from man-made objects, corridors, *etc*., Figure 1(b); *(ii)* Multiple moving objects require their own set of features which may not be sufficiently abundant without sufficient texture, Figure 1(c); *(iii)* Non-rigid objects require a rich set of features per roughly non-deforming patch, Figure 1(d). In all these cases, however, there is often

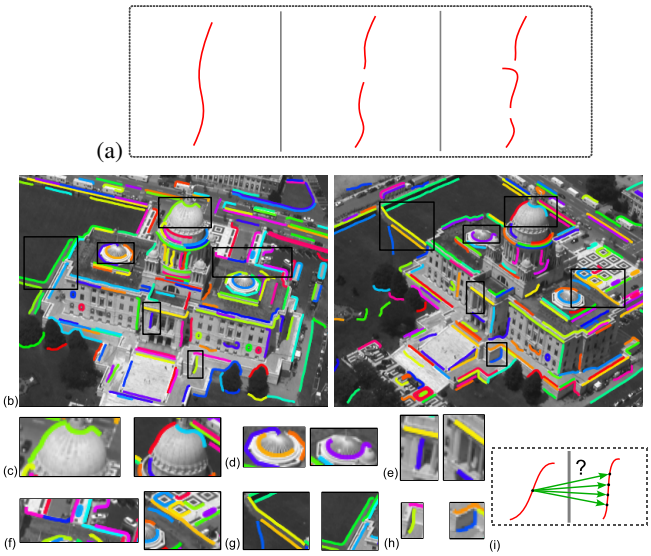

Fig. 2. Challenges in using curve fragments in multiview geometry: (a) instabilities with slight changes in viewpoint, as shown for two views in (b) and zoomed in selectively in (c-h) showing real examples of edge grouping instabilities, such as a curve in one being broken into two in another view, a curve being linked onto background, a curve being detected in one view but absent in another, a curve being fragmented into various pieces at junctions in one view but fully linked in another view, different parts of a curve being occluded in different views, and a curve undergoing shape deformation from one view to the other. (i) Point correspondence ambiguity along the curve.

sufficient **image curve structure**, motivating augmenting the use of interest points by developing a parallel technology for the use of image curve structure.

The use of image curves in determining camera pose has generally been based on epipolar tangencies, but these techniques assume that curves are closed or can be described as conics or other algebraic curves [14, 15, 19, 21]. The use of image curve fragments as the basic structure for auto-calibration under general conditions is faced with two significant challenges. First, current edge linking procedures do not generally produce curve segments which persist stably across images. Rather, an image curve fragment in one view may be present in broken form and/or or grouped with other curve fragments. Thus, while the underlying curve geometry correlates well across views, the individual curve fragments do not, Figure 2(a-h). Second, even when the image curve fragments correspond exactly, there is an intra-curve correspondence ambiguity, Figure 2(i). This ambiguity prevents the use of corresponding curve points to solve for the unknown pose and intrinsic parameters. Both these challenges motivate the use of *small curve fragments*.

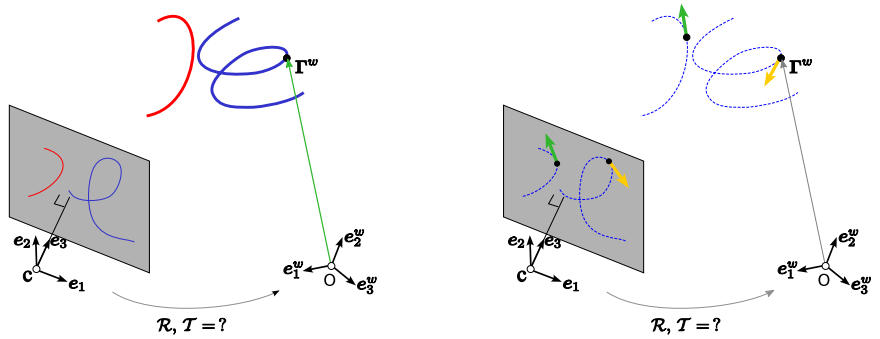

Fig. 3. The problem of finding the camera pose $\mathcal{R}$, $\mathcal{T}$ given space curves in a world coordinate system and their projections in an image coordinate system (left), and an approach to that consisting of (right) finding the camera pose $\mathcal{R}$, $\mathcal{T}$ given 3D point-tangents (*i.e.*, local curve models) in a world coordinate system and their projections in an image coordinate system.

The paradigm explored in this paper is that small curve fragments, or equivalently points augmented with differential-geometric attributes[1], can be used as the basic image structure to correlate across views. The intent is to use curve geometry as a complementary approach to the use of interest points in cases where these fail or are not available. The value of curve geometry is in correlating structure across three frames or more since the correspondence geometry in two views is unconstrained. However, the differential geometry at two corresponding points in two views reconstruct the differential geometry of the space curve they arise from [4] and this constrains the differential geometry of corresponding curves in a third view.

The fundamental questions underlying the use of points augmented with differential-geometric attributes are: how many such points are needed, what order of differential geometry is required, *etc*. This paper explores the use of first-order differential geometry, namely points with tangent attributes, for determining the pose of a single camera with respect to the coordinates of observed 3D point-tangents. It poses and solves the following:

Problem: For a camera with known intrinsic parameters, how many corresponding pairs of point-tangents in space specified in the world coordinates, and point-tangents in 2D specified in the image coordinates, are required to establish the pose of the camera with respect to the world coordinates, Figure 3.

The solution to the above problem is useful under several scenarios. First, in situations where many views of the scene are available and there is a reconstruction available from two views, *e.g.*, as in [5]. In this case a pair of point-tangents in the reconstruction can be matched under the RANSAC strategy to a pair of point-tangents in the image to determine camera pose. The advantage as compared to using three points from unorganized point reconstruction and resectioning is that *(i)* there are fewer edges than surface

[1] Previous work in exploring local geometric groupings [22] has shown that tangent and curvature as well as the sign of curvature derivative can be reliably estimated.

points and *(ii)* the method uses two rather than three points in RANSAC, requiring about half the number of runs for the same level of robustness, *e.g.*, 32 runs instead of 70 to achieve 99.99% probability of not hitting an outlier in at least one run, assuming 50% outliers (in practical systems it is often necessary to do as many runs as possible, to maximize robustness). Second, the 3D model of the object may be available from CAD or other sources, *e.g.*, civilian or military vehicles. In this case a strategy similar to the first scenario can be used. Third, in stereo video sequences obtained from precisely calibrated binocular cameras, the reconstruction from one frame of the video can be used to determine the camera pose in subsequent frames.

In general, this is a basic problem of interest in pose estimation, camera calibration, triangulation, *etc*., in computer vision, robotics, computer graphics, photogrammetry and cartography.

2 Related Work

Previous work generally has relied on the concept of matching *epipolar tangencies* on *closed curves*. Two corresponding points $\boldsymbol{\gamma}^1$ in image 1 and $\boldsymbol{\gamma}^2$ in image 2 are related by $\boldsymbol{\gamma}^{2\top} E \boldsymbol{\gamma}^1 = 0$, where E is the well-known *essential matrix* [16]. This can be extended to the relationship between the differential geometry of two curves, $\boldsymbol{\gamma}^1(s)$ in the first view and a curve $\boldsymbol{\gamma}^2(s)$ in a second view, *i.e.*,

$$\boldsymbol{\gamma}^{1\top}(s) E \boldsymbol{\gamma}^2(s) = 0. \tag{2.1}$$

The tangents $\boldsymbol{t}^1(s)$ and $\boldsymbol{t}^2(s)$ are related by differentiation

$$g^1(s) \boldsymbol{t}^{1^\top}(s) E \boldsymbol{\gamma}^2(s) + \boldsymbol{\gamma}^{1\top}(s) E g^2(s) \boldsymbol{t}^2(s) = 0, \tag{2.2}$$

where $g^1(s)$ and $g^2(s)$ are the respective speeds of parametrization of the curves $\boldsymbol{\gamma}^1(s)$ and $\boldsymbol{\gamma}^2(s)$. It is then clear that when one of the tangents $\boldsymbol{t}^1(s)$ is along the epipolar plane also, *i.e.*, $\boldsymbol{t}^{1^\top}(s) E \boldsymbol{\gamma}^2(s) = 0$ at a point s, then by necessity $\boldsymbol{\gamma}^{1\top}(s) E \boldsymbol{t}^2(s) = 0$. Thus, epipolar tangency in image 1 implies epipolar tangency in image 2 at the corresponding point, Figure 4.

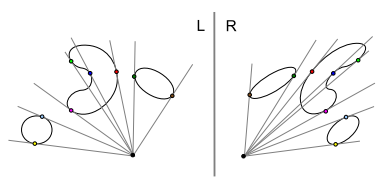

Fig. 4. Correspondence of epipolar tangencies used in curve-based camera calibration. An epipolar line on the left, whose tangency at a curve is marked in a certain color, must correspond to the epipolar line on the right having tangency on the corresponding curve, marked with the same color. This concept works for both static curves and occluding contours.

The epipolar tangency constraint was first shown in [19] who use linked edges and a coarse initial estimate E to find a sparse set of epipolar tangencies, including those at

corners, in each view. They are matched from one view to another manually. This is then used to refine the estimate E, see Figure 5, by minimizing the residual $\boldsymbol{\gamma}^{1\top}(s)E\boldsymbol{\gamma}^{2}(s)$

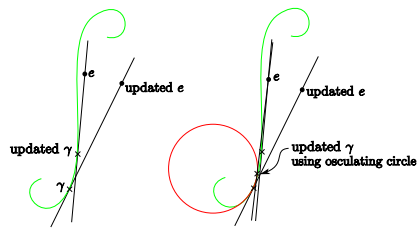

Fig. 5. Illustrating the differential update of epipolar tangencies through the use of the osculating circle or curvature information.

over all matches in an iterative two-step scheme: the corresponding points are kept fixed and E is optimized in the first step and then E is kept fixed and the points are updated in a second step using a closed form solution based on an approximation of the curve as the osculating circle. This approach assumes that closed curves are available.

Kahl and Heyden [14] consider the special case when four corresponding conics are available in two views, with unknown intrinsic parameters. In this algebraic approach, each pair of corresponding conics provides a pair of tangencies and therefore two constraints. Four pairs of conics are then needed. If the intrinsic parameters are available, then the absolute conic is known giving two constraints on the epipolar geometry, so that only 3 conic correspondences are required. This approach is only applied to synthetic data which shows the scheme to be extremely sensitive to even when a large number of conics (50) is used. Kaminski and Shashua [15] extended this work to general algebraic curves viewed in multiple uncalibrated views. Specifically, they extend Kruppa's equations to describe the epipolar constraint of two projections of a general algebraic curve. The drawback of this approach is that algebraic curves are restrictive.

Sinha *et. al.* [21] consider a special configuration where multiple static cameras view a moving object with a controlled background. Since the epipolar geometry between any pair of cameras is fixed, each hypothesized pair of epipoles representing a point in 4D is then probed for a pair of epipolar tangencies across video frames. Specifically, two pairs of tangencies in one frame in time and a single pair of tangencies in another frame provide a constraint in that they must all intersect in the same point. This allows for an estimation of epipolar geometry for each pair of cameras, which are put together for refinement using bundle adjustment, providing intrinsic parameters and relative pose. This approach, however, is restrictive in assuming well-segmentable silhouettes.

We should briefly mention the classic results that only three 2D-3D point correspondences are required to determine camera pose [7], in a procedure known as *camera resectioning* in the photogrammetry literature (and by Hartley and Zisserman [11]), also known as *camera calibration* when this is used with the purpose of obtaining the intrinsic parameter matrix $\mathcal{K}_{im}$ where the camera pose relative to the calibration jig is

not of interest. This is also related to the perspective n-point problem (PnP) originally introduced in [7] which can be stated as the recovery of the camera pose from n corresponding 3D-2D point pairs [12] or alternatively of depths [9].

Notation: Consider a sequence of n 3D points $(\boldsymbol{\Gamma}_1^w, \boldsymbol{\Gamma}_2^w, \ldots, \boldsymbol{\Gamma}_n^w)$, described in the world coordinate system and their corresponding projected image points $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_n)$ described as points in the 3D camera coordinate system. Let the rotation $\mathcal{R}$ and translation $\mathcal{T}$ relate the camera and world coordinate systems through

$$\boldsymbol{\Gamma} = \mathcal{R}\boldsymbol{\Gamma}^w + \mathcal{T}, \tag{2.3}$$

where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}^w$ are the coordinates of a point in the camera and world coordinate systems, respectively. Let $(\rho_1, \rho_2, \ldots, \rho_n)$ be the depth defined by

$$\boldsymbol{\Gamma}_i = \rho_i \boldsymbol{\gamma}_i, \qquad i = 1, \ldots, n. \tag{2.4}$$

In general we assume that each point $\boldsymbol{\gamma}_i$ is a sample from an image curve $\boldsymbol{\gamma}_i(s_i)$ which is the projection of a space curve $\boldsymbol{\Gamma}_i(S_i)$, where s_i and S_i are length parameters along the image and space curves, respectively.

The direct solution to P3P, also known as the *triangle pose problem*, and given in 1841 [8], equates the sides of the triangle formed by the three points with those of the vectors in the camera domain, *i.e.*,

$$\begin{cases} \|\rho_1\boldsymbol{\gamma}_1 - \rho_2\boldsymbol{\gamma}_2\|^2 = \|\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w\|^2 \\ \|\rho_2\boldsymbol{\gamma}_2 - \rho_3\boldsymbol{\gamma}_3\|^2 = \|\boldsymbol{\Gamma}_2^w - \boldsymbol{\Gamma}_3^w\|^2 \\ \|\rho_3\boldsymbol{\gamma}_3 - \rho_1\boldsymbol{\gamma}_1\|^2 = \|\boldsymbol{\Gamma}_3^w - \boldsymbol{\Gamma}_1^w\|^2 \end{cases} \tag{2.5}$$

This gives a system of three quadratic equations (conics) in unknowns ρ_1, ρ_2, and ρ_3. Following traditional methods going back to the German mathematician Grunert in 1841 [8] and later Finsterwalder in 1937 [6], by factoring out one depth, say ρ_1, this can be reduced to a system of two quadratic equations in two unknowns which are depth ratios $\frac{\rho_2}{\rho_1}$ and $\frac{\rho_3}{\rho_1}$. Grunert further reduced this to a single quartic equation and Finsterwalder proposed an analytic solution.

Case	Unknowns	Min. # of Point Corresp.	Min. # of Pt-Tgt Corresp.
Calibrated ($\mathcal{K}_{im}$ known)	Camera pose $\mathcal{R}$, $\mathcal{T}$	3	2 (this paper)
Focal length unknown	Pose $\mathcal{R}$, $\mathcal{T}$ and f	4	3 (conjecture)
Uncalibrated ($\mathcal{K}_{im}$ unknown)	Camera model $\mathcal{K}_{im}$, $\mathcal{R}$, $\mathcal{T}$	6	4 (conjecture)

Table 1. The number of 3D–2D point correspondences needed to solve for camera pose and intrinsic parameters.

In general, the camera resectioning problem can be solved using **three** 3D $\leftrightarrow$ 2D point correspondences when the intrinsic parameters are known, and **six** points when the intrinsic parameters are not known. The camera pose can be solved using **four** point correspondences when only the focal length is unknown, but all the other intrinsic parameters are known [3], Table 1. We now show that when intrinsic parameters

are known, **only a pair of point-tangent correspondences are required to estimate camera pose.** We conjecture that future work will show that 3 and 4 points, respectively, are required for the other two cases, Table 1. This would represent a significant reduction for a RANSAC-based computation.

3 Determining Camera Pose from a Pair of 3D–2D Point-Tangent Correspondences

Theorem 1. *Given a pair of 3D point-tangents* $\{(\boldsymbol{\Gamma}_1^w, \boldsymbol{T}_1^w), (\boldsymbol{\Gamma}_2^w, \boldsymbol{T}_2^w)\}$ *described in a world coordinate system and their corresponding perspective projections, the 2D point-tangents* $(\boldsymbol{\gamma}_1, \boldsymbol{t}_1)$, $(\boldsymbol{\gamma}_2, \boldsymbol{t}_2)$*, the pose of the camera* $\mathcal{R}$, $\mathcal{T}$ *relative to the world coordinate system defined by* $\boldsymbol{\Gamma} = \mathcal{R}\boldsymbol{\Gamma}^w + \mathcal{T}$ *can be solved up to a finite number of solutions*[2]*, by solving the system*

$$\begin{cases} \boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 \, \rho_1^2 - 2\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_2 \, \rho_1 \rho_2 + \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2 \, \rho_2^2 = \|\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w\|^2, \\ Q(\rho_1, \rho_2) = 0, \end{cases} \tag{3.1}$$

where $\mathcal{R}\boldsymbol{\Gamma}_1^w + \mathcal{T} = \boldsymbol{\Gamma}_1 = \rho_1 \boldsymbol{\gamma}_1$ *and* $\mathcal{R}\boldsymbol{\Gamma}_2^w + \mathcal{T} = \boldsymbol{\Gamma}_2 = \rho_2 \boldsymbol{\gamma}_2$*, and* $Q(\rho_1, \rho_2)$ *is an eight degree polynomial. This then solves for* $\mathcal{R}$ *and* $\mathcal{T}$ *as*

$$\begin{cases} \mathcal{R} = \left[(\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w) \; \boldsymbol{T}_1^w \; \boldsymbol{T}_2^w \right]^{-1} \cdot \\ \qquad \left[\rho_1 \boldsymbol{\gamma}_1 - \rho_2 \boldsymbol{\gamma}_2 \;\; \rho_1 \frac{g_1}{G_1} \boldsymbol{t}_1 + \frac{\rho_1'}{G_1} \boldsymbol{\gamma}_1 \;\; \rho_2 \frac{g_2}{G_2} \boldsymbol{t}_2 + \frac{\rho_2'}{G_2} \boldsymbol{\gamma}_2 \right] \\ \mathcal{T} = \rho_1 \boldsymbol{\gamma}_1 - \mathcal{R}\boldsymbol{\Gamma}_1^w, \end{cases}$$

where expressions for four auxiliary variables $\frac{g_1}{G_1}$ *and* $\frac{g_2}{G_2}$*, the ratio of speeds in the image and along the tangents, and* ρ_1 *and* ρ_2 *are available.*

Proof. We take the 2D-3D point-tangents as samples along 2D-3D curves, respectively, where the speed of parametrization along the image curves are g_1 and g_2 and along the space curves G_1 and G_2. The proof proceeds by *(i)* writing the projection equations for each point and its derivatives in the simplest form involving $\mathcal{R}$, $\mathcal{T}$, depths ρ_1 and ρ_2, depth derivatives ρ_1' and ρ_2', and speed of parametrizations G_1 and G_2, respectively; *(ii)* eliminating the translation $\mathcal{T}$ by subtracting point equations; *(iii)* eliminating $\mathcal{R}$ using dot products among equations. This gives six equations in six unknowns: $(\rho_1, \rho_2, \rho_1 \frac{g_1}{G_1}, \rho_2 \frac{g_2}{G_2}, \frac{\rho_1'}{G_1}, \frac{\rho_2'}{G_2})$; *(iv)* eliminating the unknowns ρ_1' and ρ_2' gives four quadratic equations in four unknowns: $(\rho_1, \rho_2, \rho_1 \frac{g_1}{G_1}, \rho_2 \frac{g_2}{G_2})$. Three of these quadratics can be written in the form:

$$\begin{cases} Ax_1^2 + Bx_1 + C = 0 & (3.2) \\ Ex_2^2 + Fx_2 + G = 0 & (3.3) \\ H + Jx_1 + Kx_2 + Lx_1x_2 = 0, & (3.4) \end{cases}$$

[2] assuming that the intrinsic parameters $\mathcal{K}_{im}$ are known

where $x_1 = \rho_1 \frac{g_1}{G_1}$ and $x_2 = \rho_2 \frac{g_2}{G_2}$ and where A through L are only functions of the two unknowns ρ_1 and ρ_2. Now, Equation 3.4 represents a rectangular hyperbola, Figure 6, while Equations 3.2 and 3.3 vertical and horizontal lines in the (x_1, x_2) space. Figure 6 illustrates that only one solution is possible which is then analytically written in terms of variables A–L (not shown here). This allows an expression for $\rho_1 \frac{g_1}{G_1}$ and $\rho_2 \frac{g_2}{G_2}$ in terms of ρ_1 and ρ_2 which is a degree 16 polynomial, but this is in fact divisible by $\rho_1^4 \rho_2^4$, leaving a polynomial Q of degree 8. Furthermore, we find that $Q(-\rho_1, -\rho_2) = Q(\rho_1, \rho_2)$, using the symmetry of the original equations. This, together with the unused equation (the remaining one of four) gives the system of equations 3.1. Please check

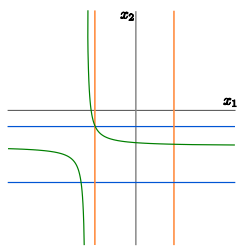

Fig. 6. Diagram of the mutual intersection of Equations 3.2–3.4 in the x_1–x_2 plane.

the detailed proof in the supplementary material.

Proposition 1. *The algebraic solutions to the system* (3.1) *of Theorem 1 are also required to satisfy the following inequalities arising from imaging and other requirements enforced by*

$$\rho_1 > 0,\ \rho_2 > 0 \tag{3.5}$$

$$\frac{g_1}{G_1} > 0,\ \frac{g_2}{G_2} > 0 \tag{3.6}$$

$$\frac{\det[\rho_1 \boldsymbol{\gamma}_1 - \rho_2 \boldsymbol{\gamma}_2 \ \ \rho_1 \frac{g_1}{G_1} \boldsymbol{t}_1 + \frac{\rho_1'}{G_1} \boldsymbol{\gamma}_1 \ \ \rho_2 \frac{g_2}{G_2} \boldsymbol{t}_2 + \frac{\rho_2'}{G_2} \boldsymbol{\gamma}_2]}{\det\left[\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w \ \boldsymbol{T}_1^w \ \boldsymbol{T}_2^w\right]} > 0. \tag{3.7}$$

Proof. There are multiple solutions for ρ_1 and ρ_2 in Equation 3.1. Observe first that if $\rho_1, \rho_2, \mathcal{R}, \mathcal{T}$ are a solution, then so are $-\rho_1$, $-\rho_2$, $-\mathcal{R}$, and $-\mathcal{T}$. Only one of these two solutions are valid, however, as the camera geometry enforces positive depth, $\rho_1 > 0$ and $\rho_2 > 0$, so that solutions are sought only in the top right quadrant of the ρ_1–ρ_2 space. In fact, the imaging geometry further restricts the points to lie in front of the camera.

Second, observe that the matrix $\mathcal{R}$ can only be a rotation matrix if it has determinant $+1$ and is a reflection rotation matrix if it has determinant -1. Using (3.2), $\det(\mathcal{R})$ can be written as

$$\det \mathcal{R} = \frac{\det\left[\rho_1 \boldsymbol{\gamma}_1 - \rho_2 \boldsymbol{\gamma}_2 \ \rho_1 \frac{g_1}{G_1} \boldsymbol{t}_1 + \frac{\rho_1'}{G_1} \boldsymbol{\gamma}_1 \ \rho_2 \frac{g_2}{G_2} \boldsymbol{t}_2 + \frac{\rho_2'}{G_2} \boldsymbol{\gamma}_2\right]}{\det\left[\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w \ \boldsymbol{T}_1^w \ \boldsymbol{T}_2^w\right]}.$$

Finally, the space curve tangent $\boldsymbol{T}$ and the image curve tangent $\boldsymbol{t}$ must point in the same direction, *i.e.*, $\boldsymbol{T} \cdot \boldsymbol{t} > 0$, or, as detailed in the supplementary material, $\frac{g_1}{G_1} > 0$ and $\frac{g_2}{G_2} > 0$.

4 A Practical Approach to Computing a Solution

Equations 3.1 can be viewed as the intersection of two curves in the $\rho_1 - \rho_2$ space. Since one of the curves to be intersected is shown to be an ellipse, it is possible to parametrize it by a bracketed parameter and then look for intersections with the second curve which is of degree 8. This gives a higher-order polynomial in a *single* unknown which can be solved more readily than simultaneously solving the two equations of degree 2 and 8.

Proposition 2. *Solutions ρ_1 and ρ_2 to the quadratic equation in* (3.1) *can be parametrized as*

$$\begin{cases} \rho_1(t) = \dfrac{2\alpha t \cos\theta + \beta(1-t^2)\sin\theta}{1+t^2} \\ \rho_2(t) = \dfrac{-2\alpha t \sin\theta + \beta(1-t^2)\cos\theta}{1+t^2}, \end{cases} \qquad -1 \le t \le 1$$

where

$$\tan(2\theta) = \frac{2(1+\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_2)}{\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2}, \qquad 0 \le 2\theta \le \pi,$$

and

$$\alpha = \frac{\sqrt{2}\|\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w\|}{\sqrt{(\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2) + (\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2)\cos(2\theta) + 2\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_2 \sin(2\theta)}}, \qquad \alpha > 0,$$

$$\beta = \frac{\sqrt{2}\|\boldsymbol{\Gamma}_1^w - \boldsymbol{\Gamma}_2^w\|}{\sqrt{(\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2) - (\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2^\top \boldsymbol{\gamma}_2)\cos(2\theta) - 2\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_2 \sin(2\theta)}}, \qquad \beta > 0.$$

Proof. An ellipse centered at the origin with semi-axes of lengths $\alpha > 0$ and $\beta > 0$ and parallel to the coordinates x and y can be parametrized as

$$x = \frac{2t}{1+t^2}\alpha, \;\; y = \frac{(1-t^2)}{1+t^2}\beta, \qquad t \in (-\infty, \infty), \tag{4.1}$$

with ellipse vertices identified at $t = -1,\ 0,\ 1$ and ∞, as shown in Figure 7. For a general ellipse centered at the origin, the coordinates must be multiplied with the rotation matrix for angle θ, obtaining

$$\begin{cases} \rho_1 = \dfrac{2\alpha t \cos\theta + \beta(1-t^2)\sin\theta}{1+t^2} \\ \rho_2 = \dfrac{-2\alpha t \sin\theta + \beta(1-t^2)\cos\theta}{1+t^2}. \end{cases} \qquad -1 \le t \le 1$$

Figure 7 illustrates this parametrization. Notice that the range of values of t which we need to consider certainly lies in the interval $[-1, 1]$ and in fact in a smaller interval

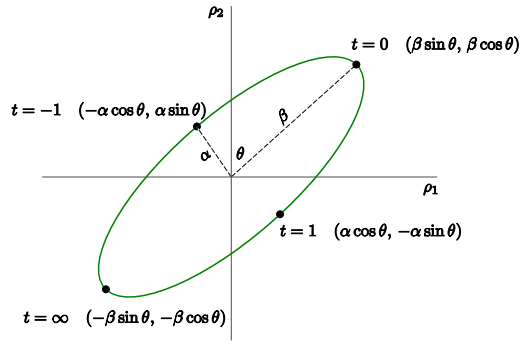

Fig. 7. Diagram illustrating a parametrization of the ellipse by a parameter t.

where $\rho_1 > 0$ and $\rho_2 > 0$. Note that t and $-\frac{1}{t}$ correspond to opposite points on the ellipse.

The parameters α, β, and θ for the ellipse in (3.1) can then be found by substitution of ρ_1 and ρ_2, details of which are found in the supplementary material.

Both equations in (3.1) are symmetric with respect to the origin in the (ρ_1, ρ_2)-plane and the curves will intersect in at most $2 \times 8 = 16$ real points, at most 8 of which will be in the positive quadrant, as we in fact require $\rho_1 > 0$ and $\rho_2 > 0$.

The parametrization of the ellipse given in Proposition 2 allows us to reduce the two Equations 3.1 to a single polynomial equation in t. Substituting for ρ_1, ρ_2 in terms of t into $Q = 0$ gives an equation in t for which, in fact, all the denominators are $(1+t^2)^{12}$, so that these can be cleared leaving a polynomial in $\tilde{Q}(t)$ of degree 16. The symmetry with respect to the origin in the (ρ_1, ρ_2)-plane becomes, in terms of t, a symmetry with respect to the substitution $t \to -1/t$, which gives diametrically opposite points of the ellipse. This implies that $\tilde{Q}$ has the special form

$$\tilde{Q}(t) = q_0 + q_1 t + q_2 t^2 + \cdots + q_{16} t^{16}, \tag{4.2}$$

where $q_i = -q_{16-i}$ for i odd. At most 8 solutions will lie in the range $-1 < t \leq 1$, and indeed we are only interested in solutions which make $\rho_1 > 0$ and $\rho_2 > 0$.

5 Experiments

We use two sets of experiments to probe camera pose recovery using 2D-3D point-tangent correspondences. First, we use a set of synthetically generated 3D curves consisting of a variety of curves (helices, parabolas, ellipses, straight lines, and saddle curves), as shown in Figure 8. Second, we use realistic data.

The synthetic 3D curves of Figure 8 are densely sampled and projected to a single 500×400 view, and their location and tangent orientation are perturbed to simulate measurement noise in the range of $0 - 2$ pixels in location and $0 - 10°$ in orientation. Our expectation in practice using the publically available edge detector [22] is that the

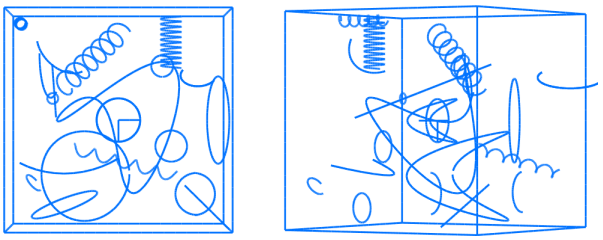

Fig. 8. Sample views of the synthetic dataset. Real datasets have also been used in our experiments, reported in further detail in the **supplemental material.**

edges can be found with subpixel accuracy and edge orientations are accurate to less than $5°$.

In order to simulate the intended application, pairs of 2D-3D point-tangent correspondences are selected in a RANSAC procedure from among 1000 veridical ones, to which 50% random spurious correspondences were added. The practical method discussed in Section 4 is used to determine the pose of the camera $(\mathcal{R}, \mathcal{T})$ inside the RANSAC loop. Each step takes $90ms$ in Matlab on a standard 2GHz dual-core laptop. What is most significant, however, is that only 17 runs are sufficient to get 99% probability of hitting an outlier-free correspondence pair, or 32 runs for 99.99% probability. In practice more runs can easily be used depending on computational requirements. To assess the output of the algorithm, we could have measured the error of the estimated pose compared to the ground truth pose. However, what is more meaningful is the impact of the measured pose on the *measured* reprojection error, as commonly used in the field to validate the output of RANSAC-based estimation. Since this is a controlled experiment, we measure final reprojection error not just to the inlier set, but to the entire pool of 1000 true correspondences. In practice, a bundle-adjustment would be run to refine the pose estimate using all inliers, but we chose to report the raw errors without nonlinear least-squares refinement. The distribution of reprojection error is plotted for various levels of measurement noise, Figure 9. These plots show that the relative camera pose can be effectively determined for a viable range of measurement errors, specially since these results are typically optimized in practice through bundle adjustment. Additional information can be found in the supplemental material.

Second, we use data from a real sequence, the "Capitol sequence", which is a set of 256 frames covering a $90°$ helicopter fly-by from the Rhode Island State Capitol, Figure 2, using a High-Definition camera (1280×720). Intrinsic parameters were initialized using the Matlab Calibration toolbox from J. Bouguet (future extension of this work would allow for an estimation of intrinsic parameters as well). The camera parameters were obtained by running Bundler [1] essentially out-of-the-box, with calibration accuracy of $1.3px$. In this setup, a pair of fully calibrated views are used to reconstruct a 3D cloud of 30 edges from manual correspondences. Pairs of matches from 3D edges to observed edges in novel views are used with RANSAC to compute the camera pose with respect to the frame of the 3D points, and measure reprojection error. One can then either use multiple pairs or use bundle adjustment to improve the reprojection error re-

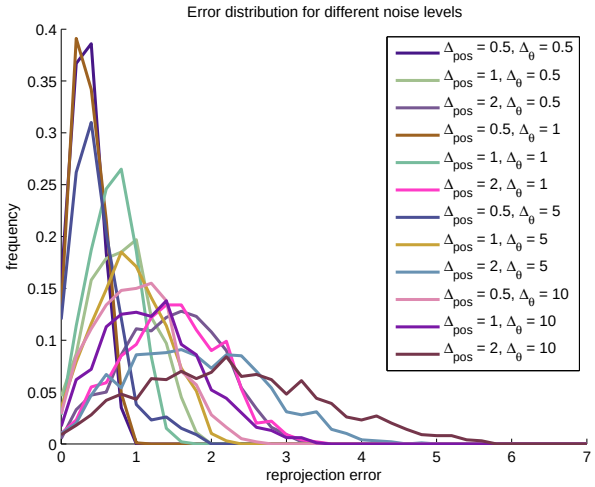

Fig. 9. Distributions of reprojection error for synthetic data without running bundle adjustment, for increasing levels of positional and tangential perturbation in the measurements. **Additional results are reported in the supplemental material.**

sulting from our initial computation of relative pose. Figure 10 shows the reprojection error distribution of our method for a single point-tangent pair after RANSAC, before and after running bundle-adjustment, versus the dataset camera from bundler (which is bundle-adjusted), for the Capitol sequence. The proposed approach achieved an average error of $1.1px$ and $0.76px$ before and after a metric bundle adjustment, respectively, as compared to $1.3px$ from Bundler. Additional information and results can be found in the supplemental material.

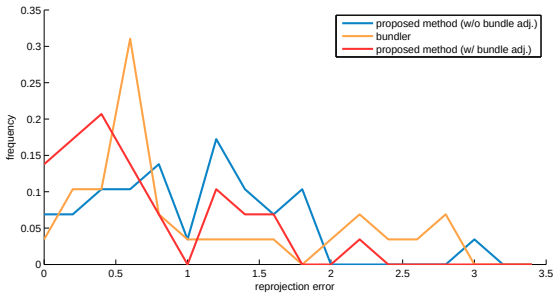

Fig. 10. The reprojection error distribution for real data (Capitol sequence) using only two point-tangents, before and after bundle adjustment. **Additional results are reported in the supplemental material.**

6 Future Directions

The paper can be extended to consider the case when intrinsic parameters are unknown. Table 1 conjectures that four pairs of corresponding 3D-2D point-tangents are sufficient to solve this problem. Also, we have been working on the problem of determining trinocular relative pose from corresponding point-tangents across 3 views. We conjecture that three triplets of correspondences among the views are sufficient to establish relative pose. This would allow for a complete curve-based structure from motion system starting from a set of images without any initial calibration.

References

1. S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In *ICCV '09*. 1, 12
2. N. Ayache and L. Lustman. Fast and reliable passive trinocular stereovision. In *ICCV'87*.
3. M. Bujnak, Z. Kukelova, and T. Pajdla. A general solution to the p4p problem for camera with unknown focal length. In *CVPR'08*. 7
4. R. Fabbri and B. B. Kimia. High-order differential geometry of curves for multiview reconstruction and matching. In *EMMCVPR'05*. 4
5. R. Fabbri and B. B. Kimia. 3D curve sketch: Flexible curve-based stereo reconstruction and calibration. In *CVPR'10*. 4
6. S. Finsterwalder and W. Scheufele. Das ruckwartseinschneiden im raum. *Sebastian Finsterwalder zum 75*, pages 86–100, 1937. 7
7. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981. 1, 6, 7
8. J. A. Grunert. Das pothenotische problem in erweiterter gestalt nebst Über seine anwendungen in der geodäsie. *Archiv der für Mathematik and Physik*, 1:238–248, 1841. 7
9. R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV*, 13(3):331–356, 1994. 7
10. C. Harris and M. Stephens. A combined edge and corner detector. In *Alvey Vision Conference*, 1988. 1
11. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000. 1, 6
12. R. Horaud, B. Conio, O. Leboulleux, and B. Lacolle. An analytic solution for the p4p problem. *CVGIP*, 47(1):33–44, 1989. 7
13. Z. Y. Hu and F. C. Wu. A note on the number of solutions of the noncoplanar p4p problem. *PAMI*, 24(4):550–555, 2002.
14. F. Kahl and A. Heyden. Using conic correspondence in two images to estimate the epipolar geometry. In *ICCV'98*. 3, 6
15. J. Y. Kaminski and A. Shashua. Multiple view geometry of general algebraic curves. *IJCV*, 56(3):195–219, 2004. 3, 6
16. H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981. 5
17. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 1
18. P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3D objects. *IJCV*, 73(3):263–284, 2007. 2
19. J. Porrill and S. Pollard. Curve matching and stereo calibration. *IVC*, 9(1):45–50, 1991. 3, 5

20. L. Robert and O. D. Faugeras. Curve-based stereo: figural continuity and curvature. In *CVPR'91*.
21. S. N. Sinha, M. Pollefeys, and L. McMillan. Camera network calibration from dynamic silhouettes. In *CVPR'04*. 3, 6
22. A. Tamrakar and B. B. Kimia. No grouping left behind: From edges to curve fragments. In *ICCV '07*. 4, 11