3D Reconstruction of a Moving Point from a Series of 2D Projections

Hyun Soo Park[1], Takaaki Shiratori[1,2], Iain Matthews[2], and Yaser Sheikh[1]

[1] Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, USA, 15213
[2] Disney Research, Pittsburgh, 4615 Forbes Ave., Pittsburgh, PA, USA, 15213
{hyunsoop,siratori,yaser}@cs.cmu.edu,iainm@disneyresearch.com

Abstract. This paper presents a linear solution for reconstructing the 3D trajectory of a moving point from its correspondence in a collection of 2D perspective images, given the 3D spatial pose and time of capture of the cameras that produced each image. Triangulation-based solutions do not apply, as multiple views of the point may not exist at each instant in time. A geometric analysis of the problem is presented and a criterion, called reconstructibility, is defined to precisely characterize the cases when reconstruction is possible, and how accurate it can be. We apply the linear reconstruction algorithm to reconstruct the time evolving 3D structure of several real-world scenes, given a collection of non-coincidental 2D images.

Keywords: Multiple view geometry, Non-rigid structure from motion, Trajectory basis, and Reconstructibility

1 Introduction

Without making *a priori* assumptions about scene structure, it is impossible to reconstruct a 3D scene from a monocular image. Binocular stereoscopy is a solution used both by biological and artificial systems to localize the position of a point in 3D via correspondences in two views. Classic triangulation used in stereo reconstruction is geometrically well-posed as shown in Figure 1(a). The rays connecting each image location to its corresponding camera center intersect at the true 3D location of the point — this process is called triangulation as the two rays map out a triangle with the baseline that connects the two camera centers. The triangulation constraint does not apply when the point moves in the duration between image capture, as shown in Figure 1(b). This case abounds as most artificial vision systems are monocular and most real scenes contain moving elements.

The 3D reconstruction of a trajectory is directly analogous to monocular image reconstruction: it is impossible to reconstruct a moving point without making some assumptions about the way it moves. In this paper, we represent the 3D trajectory of a moving point as a compact linear combination of a trajectory basis and demonstrate that, under this model, we can recover the 3D motion of the point linearly, and can handle missing data. By posing the problem in

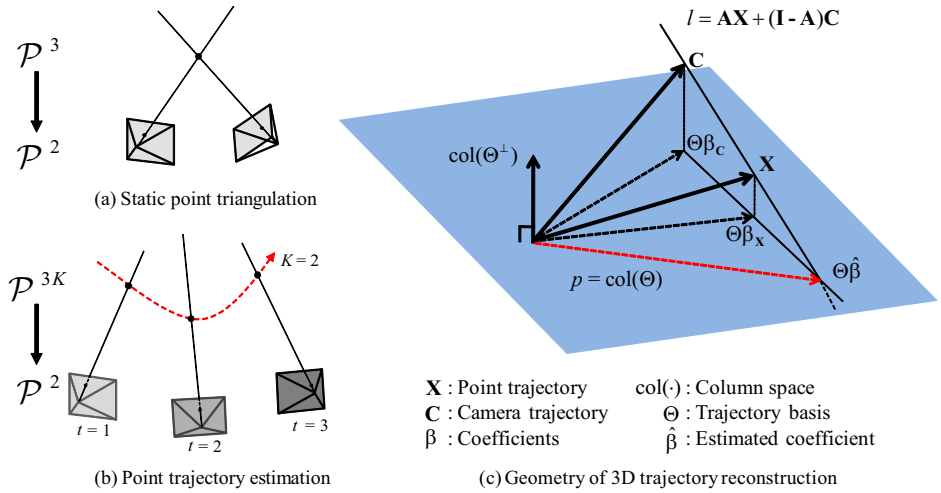


Fig. 1. (a) A point in projective space, $\mathcal{P}^3$, is mapped to $\mathcal{P}^2$. From two views, the 3D point can be triangulated. (b) From a series of images, a point trajectory, $\mathcal{P}^{3K}$, also imaged to $\mathcal{P}^2$. To estimate the trajectory, at least three projections are required when the number of parameters describing the trajectory is 6 (2 for each coordinate, x, y, and z). (c) Geometric illustration of the least squares solution of Equation (4). The estimated trajectory $\boldsymbol{\Theta}\widehat{\boldsymbol{\beta}}$ is placed on the intersection between l containing the camera trajectory space and the point trajectory, and the p space spanned by the column space of the trajectory basis matrix, $\mathrm{col}(\boldsymbol{\Theta})$.

this way, we generalize the problem of triangulation, which is a mapping from $\mathcal{P}^3 \rightarrow \mathcal{P}^2$, to 3D *trajectory* reconstruction, as a mapping $\mathcal{P}^{3K} \rightarrow \mathcal{P}^2$, where $3K$ is the number of the trajectory basis required to represent the 3D point trajectory[3].

The stability of classic triangulation is known to depend on the baseline between camera centers [3]. In this paper, we characterize an instability encountered when interference occurs between the trajectory of the point and the trajectory mapped out by successive cameras centers. We demonstrate that the accuracy of 3D trajectory reconstruction is fundamentally limited by the correlation between the trajectory of the point and the trajectory of successive camera centers. A measure called reconstructibility is defined which can determine the accuracy of reconstruction, given a particular trajectory basis, 3D point trajectory, and 3D camera center trajectory. The linear reconstruction algorithm, in conjunction with this analysis, is used to propose a practical algorithm for the reconstruction of multiple 3D trajectories from a collection of non-coincidental images.

[3] Related observations have been made in [1, 2].

2 Related work

When correspondences are provided across 2D images in static scenes, the method proposed by Longuet-Higgins [4] estimates the relative camera poses and triangulates the point in 3D using epipolar geometry. In subsequent research, summarized in [3, 5, 6], the geometry involved in reconstructing 3D scenes has been developed. While a static point can be estimated by the triangulation method, in the case where the point may move between the capture of both images the triangulation method becomes inapplicable: the line segments mapped out by the baseline and the rays from each camera center to the point no longer form a closed triangle (Figure 1(b)).

The principal work in 'triangulating' moving points from a series of images is by Avidan and Shashua [7], who coined the term *trajectory-triangulation*. They demonstrated two cases where a moving point can be reconstructed: (1) if the point moves along a line, or (2) if the point moves along a conic section. This work inspired a number of papers such as the work by Shashua and Wolf [1], who demonstrated reconstruction for points moving along planes, and the work by Kaminski and Teicher [8] who extended to a general trajectory using the polynomial representation. Wolf and Shashua [9] classified different manifestations of related problems, analyzing them as projections from $\mathcal{P}^N$ to $\mathcal{P}^2$.

In this paper, we investigate the reconstruction of the 3D trajectory of a moving point where the motion of the point can be described as a compact combination of a linear trajectory basis. This generalization allows far more natural motions to be linearly reconstructed. We demonstrate its application in reconstructing dynamic motion of objects from a series of image projections where no two image projections necessarily occur at the same time instant.

The reconstruction of dynamic motion from monocular sequences, or nonrigid structure from motion, is one such domain. The seminal work of Bregler *et al.* [10] introduced linear shape models as a representation for nonrigid 3D structures, and demonstrated their applicability within the factorization-based reconstruction paradigm of Tomasi and Kanade [11]. Subsequently, numerous constraints and techniques have been proposed to specify shape priors depending on models such as facial expressions and articulated body motions [12–16]. In contrast to these methods which represent the instantaneous shape of an object as a linear combination of basis shapes, Akhter *et al.* [17] proposed analyzing each trajectory as a linear combination of basis trajectories. They proposed the use of the Discrete Cosine Transform as a basis, and applied factorization techniques to estimate nonrigid structure. The primary limitation of these factorization-based methods is: (1) the assumption of an orthographic camera, and (2) their inability to handle missing information. Several papers have relaxed the constraint of orthography, such as Hartley and Vidal [2] and Vidal and Abretske [18], and the work by Torresani *et al.* [15] can handle missing data. However, these algorithms remain unstable and have been demonstrated to work only for constrained data like faces or motion capture; studies of this instability have been pursued by Xiao *et al.* [12] and Akhter *et al.* [19].

Unlike previously proposed methods, we do not pursue a factorization based solution. Instead we propose a linear solution to reconstruct a moving point from a series of its image projections inspired by the Direct Linear Transform algorithm [3]. In conjunction with rigid structure from motion estimation, and the trajectory based representation of points, this facilitates the first practical algorithm for dynamic structure reconstruction. It is able to handle problems like missing data (due to occlusion and matching failure) and estimation instability. An analysis is presented which geometrically describes the reconstruction problem as fundamentally restricted by the correlation between the motion of the camera center and the motion of a scene point trajectory. This analysis is leveraged to estimate an optimized trajectory basis to represent scene point motion, given an estimated camera center trajectory. We will assume that scene point correspondences have been provided, and that the relative locations of the view-points have been estimated, and that the basis describing the trajectory are pre-defined: these are reasonable assumptions that will be justified presently.

3 Linear Reconstruction of a 3D Point Trajectory

For a static point in 3D projective space, correspondences across a pair of images enable us to triangulate as shown in Figure 1(a). Traditional triangulation solves for a 3D point from an overconstrained system because there are three unknowns while the number of equations is $2F$, where F is the number of images. For a 3D point trajectory, if it can be represented by K parameters per coordinate, the projection is $\mathcal{P}^{3K} \rightarrow \mathcal{P}^2$ as shown in Figure 1(b). As was the case with static point projection, if $2F \geq 3K$, solving for a 3D trajectory becomes an overconstrained problem. Using this observation, we develop a linear solution for reconstructing a point trajectory given the relative poses of the cameras and the time instances the images were captured.

For a given ith camera projection matrix, $\mathbf{P}_i \in \Re^{3\times 4}$, let a point in 3D, $\mathbf{X}_i = \begin{bmatrix} X_i \; Y_i \; Z_i \end{bmatrix}^\mathsf{T}$, be imaged as $\mathbf{x}_i = \begin{bmatrix} x_i \; y_i \end{bmatrix}^\mathsf{T}$. The index i used in this paper represents the ith time sample. This projection is defined up to scale,

$$\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} \simeq \mathbf{P}_i \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}_\times \mathbf{P}_i \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix} = \mathbf{0}, \tag{1}$$

where $[\cdot]_\times$ is the skew symmetric representation of the cross product [3]. This can be rewritten as an inhomogeneous equation,

$$\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}_\times \mathbf{P}_{i,1:3}\mathbf{X}_i = -\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}_\times \mathbf{P}_{i,4} \quad ,$$

where $\mathbf{P}_{i,1:3}$ and $\mathbf{P}_{i,4}$ are the matrices made of the first three columns and the last column of $\mathbf{P}_i$, respectively, or simply as $\mathbf{Q}_i\mathbf{X}_i = \mathbf{q}_i$, where,

$$\mathbf{Q}_i = \left(\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}_\times \mathbf{P}_{i,1:3} \right)_{1:2}, \; \mathbf{q}_i = \left(\begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix}_\times \mathbf{P}_{i,4} \right)_{1:2},$$

and $(\cdot)_{1:2}$ is the matrix made of two rows from $(\cdot)$. By taking into account all time instants, a closed form for the 3D point trajectory, $\mathbf{X}$, can be formulated as,

$$\begin{bmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_F \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_F \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_F \end{bmatrix}, \quad \text{or } \mathbf{QX} = \mathbf{q}, \tag{2}$$

where F is the number of time samples in the trajectory. Since Equation (2) is an underconstrained system (i.e. $\mathbf{Q} \in \Re^{2F \times 3F}$), there are an infinite number of solutions for a given set of measurements (2D projections). There are many ways to constrain the solution space in which $\mathbf{X}$ lies. One way is approximating the point trajectory using a linear combination of any trajectory basis that can describe it as,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^\mathsf{T} \cdots \mathbf{X}_F^\mathsf{T} \end{bmatrix}^\mathsf{T} \approx \mathbf{\Theta}_1 \beta_1 + \ldots + \mathbf{\Theta}_{3K} \beta_{3K} = \mathbf{\Theta}\boldsymbol{\beta}, \tag{3}$$

where $\mathbf{\Theta}_j \in \Re^{3F}$ is a trajectory basis vector, $\mathbf{\Theta} = \begin{bmatrix} \mathbf{\Theta}_1 \ldots \mathbf{\Theta}_{3K} \end{bmatrix} \in \Re^{3F \times 3K}$ is the trajectory basis matrix, $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \ldots \beta_{3K} \end{bmatrix}^\mathsf{T} \in \Re^{3K}$ are the parameters or coefficients of a point trajectory, and K is the number of bases per coordinate.

If the trajectory basis are known *a priori* [17], this linear map between the point trajectory and basis enables us to formulate a linear solution. By plugging Equation (3) into Equation (2), we can derive an overconstrained system by choosing K such that $2F \geq 3K$,

$$\mathbf{Q}\mathbf{\Theta}\boldsymbol{\beta} = \mathbf{q}. \tag{4}$$

Equation (4) is a linear least squares system for reconstructing a point trajectory, $\boldsymbol{\beta}$, which provides an efficient, numerically stable, and globally optimal solution. $\boldsymbol{\beta}$ is the coefficient of the trajectory based on measurements and known camera poses embedded in $\mathbf{Q}$ and $\mathbf{q}$ and known trajectory basis, $\mathbf{\Theta}$.

4 Geometric Analysis of 3D Trajectory Reconstruction

Empirically, the point trajectory reconstruction approaches the ground truth point trajectory when the camera motion is fast or random. On the other hand, if the camera moves slowly or smoothly, the solution tends to deviate highly from the ground truth. To explain these observations, we decompose the process of solving the linear least squares system into two steps: solving Equation (2) and solving Equation (3).

4.1 Geometry of Point and Camera Trajectories

Let $\mathbf{X}$ and $\widehat{\mathbf{X}}$ be a ground truth trajectory and an estimated point trajectory, respectively. The camera matrix can always be normalized by intrinsic and rotation matrices, $\mathbf{K}$ and $\mathbf{R}$, respectively, because they can be factored out without

loss of generality (as all camera matrices are known), i.e. $\mathbf{R}_i^\mathsf{T}\mathbf{K}_i^{-1}\mathbf{P}_i = \left[\mathbf{I}_3 \middle| -\mathbf{C}_i\right]$, where $\mathbf{P}_i = \mathbf{K}_i\mathbf{R}_i\left[\mathbf{I}_3 \middle| -\mathbf{C}_i\right]$, $\mathbf{C}_i$ is the camera center, and $\mathbf{I}_3$ is a 3×3 identity matrix. This follows from the fact that triangulation and 3D trajectory reconstruction are both geometrically unaffected by the rotation of the camera about its center. All $\mathbf{P}_i$ subsequently used in this analysis are normalized camera matrices, i.e. $\mathbf{P}_i = \left[\mathbf{I}_3 \middle| -\mathbf{C}_i\right]$. Then, a measurement is a projection of $\mathbf{X}$ onto the image plane from Equation (1). Since Equation (1) is defined up to scale, the measurement, $\mathbf{x}$, can be replaced as follows,

$$\left[\mathbf{P}_i \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix}\right]_\times \mathbf{P}_i \begin{bmatrix} \widehat{\mathbf{X}}_i \\ 1 \end{bmatrix} = 0. \tag{5}$$

Plugging in $\mathbf{P}_i = \left[\mathbf{I}_3 \middle| -\mathbf{C}_i\right]$ results in, $[\mathbf{X}_i - \mathbf{C}_i]_\times \left(\widehat{\mathbf{X}}_i - \mathbf{C}_i\right) = 0$, or equivalently,

$$[\mathbf{X}_i - \mathbf{C}_i]_\times \widehat{\mathbf{X}}_i = [\mathbf{X}_i]_\times \mathbf{C}_i. \tag{6}$$

The solution of Equation (6) is

$$\widehat{\mathbf{X}}_i = a_i\mathbf{X}_i + (1 - a_i)\mathbf{C}_i, \tag{7}$$

where a_i is an arbitrary scalar. Geometrically, Equation (7) is the constraint for the perspective camera model due to the fact that it enforces the solution to lie on the ray joining the camera center and the point in 3D. From Equation (3), Equation (7) can be rewritten as $\mathbf{\Theta}_i\widehat{\boldsymbol{\beta}} \approx a_i\mathbf{X}_i + (1 - a_i)\mathbf{C}_i$ where $\widehat{\boldsymbol{\beta}}$ is the estimated parameter and $\mathbf{\Theta}_i$ is the matrix from $\mathbf{\Theta}_{(3(i-1)+1):3i}$.

Figure 1(c) illustrates the geometry of the solution of Equation (4). Let the subspace, p, be the space spanned by the column space of the trajectory basis matrix, $\mathrm{col}(\mathbf{\Theta})$. The solution $\mathbf{\Theta}\widehat{\boldsymbol{\beta}}$, has to simultaneously lie on the hyperplane l, which contains the camera trajectory and the point trajectory, and must lie in $\mathrm{col}(\mathbf{\Theta})$. Thus, $\mathbf{\Theta}\widehat{\boldsymbol{\beta}}$ is the intersection of the hyperplane l and the subspace p where $\mathbf{A} = \mathbf{D} \otimes \mathbf{I}_3$.[4] In the figure, note that the line and the plane are a conceptual 3D vector space representation for the $3F$-dimensional space. The camera center trajectory, $\mathbf{C} = \left[\mathbf{C}_1^\mathsf{T} \ldots \mathbf{C}_F^\mathsf{T}\right]^\mathsf{T}$, and the point trajectory, $\mathbf{X}$, are projected onto $\mathrm{col}(\mathbf{\Theta})$ as $\mathbf{\Theta}\boldsymbol{\beta}_\mathbf{C}$ and $\mathbf{\Theta}\boldsymbol{\beta}_\mathbf{X}$, respectively. From this point of view, we want $\mathbf{\Theta}\widehat{\boldsymbol{\beta}}$ to be as close as possible to $\mathbf{\Theta}\boldsymbol{\beta}_\mathbf{X}$.

4.2 Reconstructibility

When a point trajectory is identical to the camera trajectory, it is not possible to estimate the point trajectory because a series of 2D projections is stationary. This intuition results in the following theorem.

Theorem 1 *Trajectory reconstruction using any linear trajectory basis is impossible if* $\mathrm{corr}(\mathbf{X}, \mathbf{C}) = \pm 1$.[5]

[4] $\otimes$ is the Kronecker product and $\mathbf{D}$ is a diagonal matrix which consists of $\{a_1, \cdots, a_F\}$.

[5] $\mathrm{corr}(X, Y) = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$ where $E[\cdot]$ is the expected value operator and μ and σ are the mean and standard deviation, respectively.

Proof. When $\mathrm{corr}(\mathbf{X}, \mathbf{C}) = \pm 1$, or $\mathbf{X} = c\mathbf{C} + \mathbf{d}$ where c is arbitrary scalar and $\mathbf{d}$ is arbitrary constant vector, we can transform $\mathbf{X}$ and $\mathbf{C}$ to $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{C}}$ such that $\widetilde{\mathbf{X}} = c\widetilde{\mathbf{C}}$ without loss of generality. This linearity causes the RHS of Equation (6) to be zero and the solution $\widehat{\mathbf{X}}_i$ to be the same as $\widetilde{\mathbf{C}}_i$ up to scale. This results in the scale ambiguity of $\widehat{\mathbf{X}}_i$. □

While Theorem 1 shows the reconstruction limitation due to the correlation between the point trajectory and the camera trajectory, solving Equation (3) with respect to $\boldsymbol{\beta}$ provides a measure of the reconstruction accuracy for a given trajectory basis. Solving the least squares, $\widehat{\mathbf{X}} = \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}}$ minimizes the residual error,

$$\underset{\widehat{\boldsymbol{\beta}}, \mathbf{A}}{\operatorname{argmin}} \left\| \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} - \mathbf{A}\mathbf{X} - (\mathbf{I} - \mathbf{A})\,\mathbf{C} \right\|. \tag{8}$$

Let us decompose the point trajectory and the camera trajectory into the column space of $\boldsymbol{\Theta}$ and that of the null space, $\boldsymbol{\Theta}^{\perp}$ as follows, $\mathbf{X} = \boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{X}} + \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{X}}^{\perp}$, $\mathbf{C} = \boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{C}} + \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp}$, where $\boldsymbol{\beta}^{\perp}$ is the coefficient for the null space. Let us also define a measure of *reconstructibility*, η, of the 3D point trajectory reconstruction,

$$\eta = \frac{\left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\|}{\left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{X}}^{\perp} \right\|}. \tag{9}$$

Theorem 2 *As η approaches infinity, $\widehat{\boldsymbol{\beta}}$ approaches $\boldsymbol{\beta}_{\mathbf{X}}$.*

Proof. From the triangle inequality, the objective function of Equation (8) is bounded by,

$$\begin{aligned} &\left\| \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} - \mathbf{A}\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{X}} - (\mathbf{I} - \mathbf{A})\,\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{C}} - \mathbf{A}\boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{X}}^{\perp} - (\mathbf{I} - \mathbf{A})\,\boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\| \qquad (10) \\ \leq &\left\| \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} - \mathbf{A}\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{X}} - (\mathbf{I} - \mathbf{A})\,\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{C}} \right\| + \left\| \mathbf{A}\boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{X}}^{\perp} \right\| + \left\| (\mathbf{I} - \mathbf{A})\,\boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\| \\ \leq &\left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\| \left(\frac{\left\| \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} - \mathbf{A}\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{X}} - (\mathbf{I} - \mathbf{A})\,\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{C}} \right\|}{\left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\|} + \frac{\|\mathbf{A}\|}{\eta} + \|\mathbf{I} - \mathbf{A}\| \right). \qquad (11) \end{aligned}$$

As η approaches infinity, $\|\mathbf{A}\|/\eta$ in Equation (11) becomes zero. In order to minimize Equation (11), $\mathbf{A} = \mathbf{I}$ because it leaves the last term zero and $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{\mathbf{X}}$ because it also cancels the first term. This leads the minimum of Equation (11) to be zero, which bounds the minimum of Equation (10). Thus, as η approaches infinity, $\widehat{\boldsymbol{\beta}}$ approaches $\boldsymbol{\beta}_{\mathbf{X}}$. □

Figure 2(a) shows how reconstructibility is related to the accuracy of the 3D reconstruction error. In each reconstruction, the residual error (null components) of the point trajectory, $e_{\mathbf{X}} = \left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{X}}^{\perp} \right\|$, and the camera trajectory, $e_{\mathbf{C}} = \left\| \boldsymbol{\Theta}^{\perp}\boldsymbol{\beta}_{\mathbf{C}}^{\perp} \right\|$, are measured. Increasing $e_{\mathbf{C}}$ for a given point trajectory enhances the accuracy of the 3D reconstruction, while increasing $e_{\mathbf{X}}$ lowers accuracy. Even though we cannot directly measure the reconstructibility (we never

know the true point trajectory in a real example), it is useful to demonstrate the direct relation with 3D reconstruction accuracy. Figure 2(b) illustrates that the reconstructibility is inversely proportional to the 3D reconstruction error.

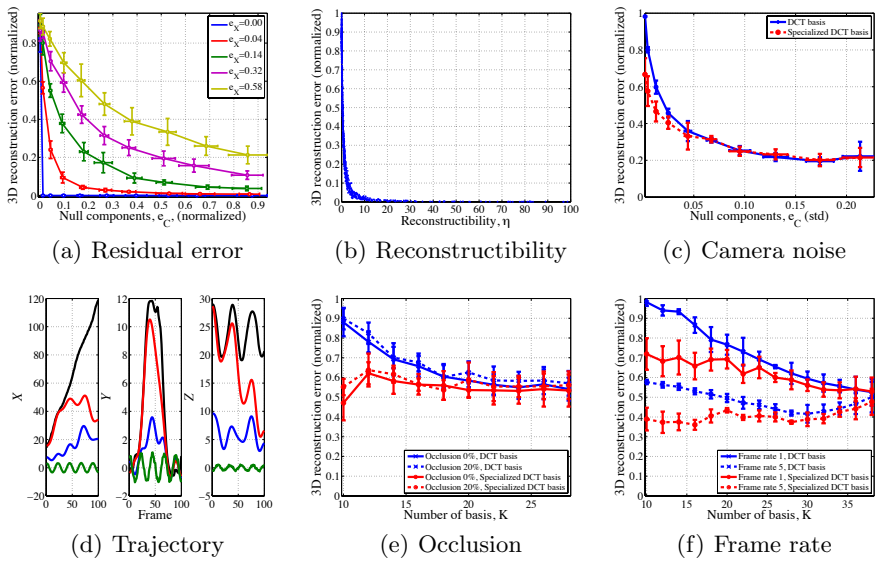
In practice, the infinite reconstructibility criterion is difficult to satisfy because the actual $\mathbf{X}$ is unknown. To enhance reconstructibility we can maximize $e_{\mathbf{C}}$ with constant $e_{\mathbf{X}}$. Thus, the best camera trajectory for a given trajectory basis matrix is the one that lives in the null space, $\text{col}(\boldsymbol{\Theta}^{\perp})$. This explains our observation about slow and fast camera motion described at the beginning of this section. When the camera motion is slow, the camera trajectory is likely to be represented well by the DCT basis, which results in low reconstructibility and vice versa. However, for a given camera trajectory, there is no deterministic way to define a trajectory basis matrix because it is coupled with both the camera trajectory and the point trajectory. If one simply finds an orthogonal space to the camera trajectory, in general, it is likely to nullify space that also spans the point trajectory space. Geometrically, simply changing the surface of p in Figure 1(c) may result in a greater deviation between $\boldsymbol{\Theta}\boldsymbol{\beta}_{\mathbf{X}}$ and $\boldsymbol{\Theta}\widehat{\boldsymbol{\beta}}$. Yet, if we have prior information of a point trajectory, we can enhance the reconstructibility. For example, if one is shooting video while walking, the frequency of the camera trajectory will be concentrated at a certain frequency, say the walking frequency, whereas that of a point trajectory is somewhere else. In such a case, if we find a trajectory basis space that is orthogonal to the walking frequency basis, the point trajectory can be estimated well, as long as it does not contain that frequency. This process allows us to eliminate interference from the camera trajectory.

5 Results

In this section, we evaluate 3D trajectory reconstruction on both synthetic and real data. In all cases, the trajectory bases are the first K discrete cosine transform (DCT) basis in order of increasing frequency. The DCT basis has been demonstrated to accurately and compactly model 1D point trajectories [17]. If a 3D trajectory is continuous and smooth, DCT basis can represent it accurately with relatively few low frequency components. We make the assumption that each point trajectory is continuous and smooth and use the DCT basis as the trajectory basis, $\boldsymbol{\Theta}$. We choose the value of K based on the number of visible points on a trajectory such that the system is overconstrained and $2F \geq 3K$. We consider two choices of DCT bases: the *original DCT* basis set, and the *specialized DCT* basis set. The specialized DCT is a projection of the original DCT onto the null space of the camera trajectory. The idea here is to limit how well the specialized DCT reconstructs the camera trajectory and improve the reconstructibility.

5.1 Simulation

To quantitatively evaluate our method, we generate synthetic 2D images from 3D motion capture data and test it in three perspectives: reconstructibility, ro-



(a) Residual error (b) Reconstructibility (c) Camera noise

(d) Trajectory (e) Occlusion (f) Frame rate

Fig. 2. (a) As the null component of the camera trajectory, $e_{\mathbf{C}}$, decreases, the closed form solution of Equation (4) deviates from the real solution. (b) Reconstructibility, η, provides the degree of interference between the camera trajectory and the point trajectory. (c) Comparisons of reconstruction accuracy of trajectories reconstructed with the specialized and original DCT basis under various camera trajectories, and (d) trajectories between the ground truth and the original and specialized DCT basis under smooth camera trajectory. Black: the ground truth of the point trajectory, green: the camera trajectory, and blue and red: reconstructed trajectory of the motion capture marker from the original and specialized DCT basis, respectively. Comparisons of robustness between the original and specialized DCT basis with regard to (e) occlusion and (f) frame rate.

bustness, and accuracy. For reconstructibility, we compare reconstruction from the original DCT basis with the specialized DCT basis by increasing the null component, $e_{\mathbf{C}}$, of the camera trajectory. Reconstruction error from the original DCT basis is higher when there is small $e_{\mathbf{C}}$. For robustness, we test with missing data and lowered frame rates and we show that the specialized DCT basis performs better. Finally, for accuracy, we compare our algorithm with state-of-the-art algorithms by varying the perspectivity of projection. The results show our method outperforms others, particularly under perspective projection.

Reconstructibility: Earlier, we defined the reconstructibility of a 3D trajectory as the trade off between the ability of the chosen trajectory basis to accurately reconstruct the point trajectory vs. its ability to reconstruct the camera trajectory. To evaluate this effect empirically we generate camera trajectories by varying $e_{\mathbf{C}}$ and measure the error in point trajectory reconstruction in Figure 2(c). Each trajectory is normalized to have zero mean and unit variance so that errors can be compared across different sequences. When $e_{\mathbf{C}}$ is low, there is

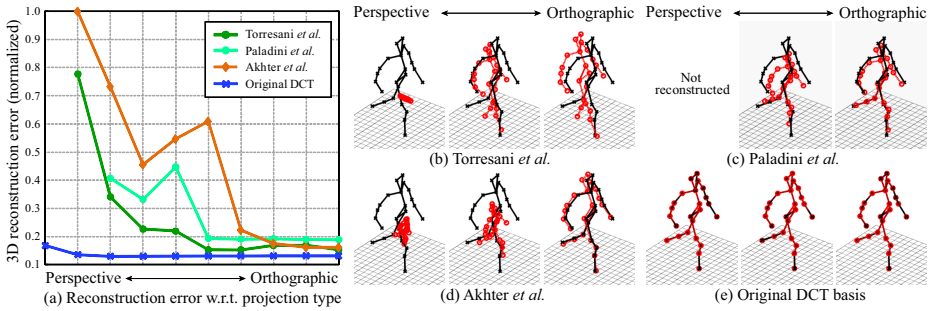


Fig. 3. (a) Quantitative comparisons of reconstruction accuracy with previous methods regarding projection types, and qualitative comparisons of reconstruction errors using the original DCT basis (blue) and the methods by Torresani *et al.* [15](dark green), Paladini *et al.* [16](light green) and Akhter *et al.* [17](orange). (b-e): Qualitative comparison between the ground truth (black) and reconstructed trajectories (red) for each method.

an advantage in using the *specialized* DCT basis. This is expected as the original DCT basis is able to reconstruct both camera and point trajectories well, and the reconstructibility is lower. As $e_{\mathbf{C}}$ increases, this becomes less of an issue, and both original and specialized DCT perform approximately the same. Figure 2(d) shows the comparison of point trajectories reconstructed using the original and specialized DCT basis compared to the ground truth. For this example the reconstructibility using the specialized DCT is 2.45, and for the original DCT basis it is 0.08.

Robustness: In this experiment, we evaluate the robustness of trajectory reconstruction for smooth camera trajectories with missing 2D point samples. Missing samples occur in practice due to occlusion, self-occlusion, or measurement failure. Figure 2(e) shows the normalized trajectory reconstruction error for varying amounts of occlusion (0% and 20% of the sequence) and different numbers of DCT basis. A walking motion capture sequence was used and each experiment was repeated 10 times with random occlusion. As long as the visibility of a point in a sequence is sufficient to overconstrain the linear system of equations, the closed form solution is robust to moderate occlusion. Figure 2(f) evaluates robustness to the frequency of input samples, i.e. varying the effective frame rate of the input sequence. Visibility of the moving points is important to avoid an ill-posed condition of the closed form solution, and intuitively more frequent visibility results in better reconstruction. The results confirm this observation. In both robustness experiment, the specialized DCT basis perform better than the original DCT basis for reduced number of bases. This is due to the (worst case) smooth synthesized camera trajectories. This effect is reduced as the number of DCT basis increases and the reconstructibility of the sequence increases accordingly.

Accuracy: We compare the accuracy of reconstructed trajectories against methods using shape basis reconstruction by Torresani *et al.* [15] and Pala-

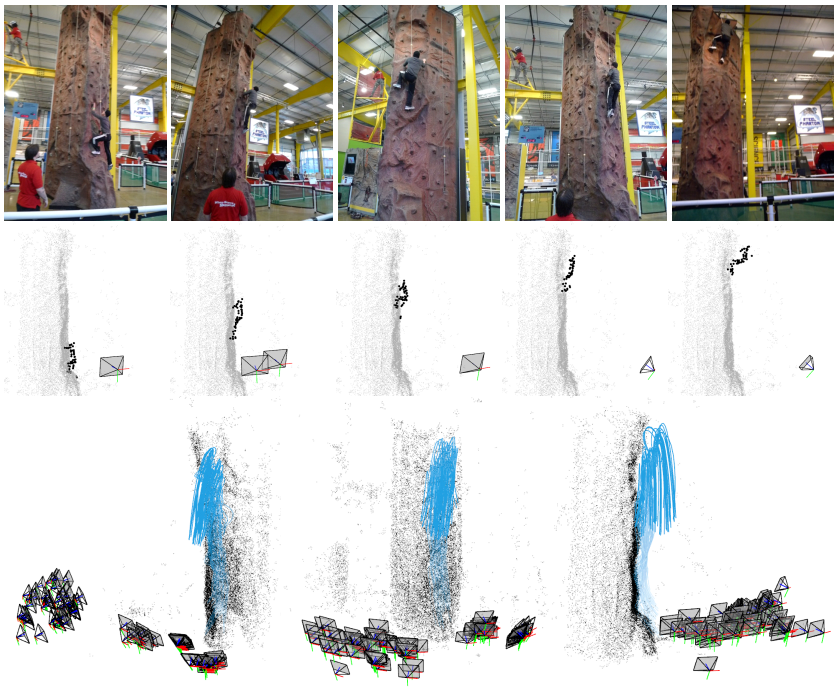


Fig. 4. Results of the rock climbing scene. Top row: sampled image input, second row: five snap shots of 3D reconstruction in different views, and bottom row: reconstructed trajectories (blue line) in different views.

dini *et al.* [16] and the method using trajectory basis reconstruction Akhter *et al.* [17]. To validate that our closed form solution is independent of the camera projection model, we parameterize camera projection as the distance between image plane and the camera center and evaluate across a range that moves progressively from projective at one end to orthographic at the other. Note that we are given all camera poses for the closed form trajectory solution, while the previous methods reconstruct both camera poses and point trajectories simultaneously. We set K to 10 for all methods and use the original DCT basis. Figure 3 compares the normalized reconstruction accuracy for the walking scene under a random camera trajectory. The other methods assume orthographic camera projection and are unable to accurately reconstruct trajectories in the perspective case.

5.2 Experiments with Real Data

The theory of reconstructibility states that it is possible to reconstruct 3D point trajectories using DCT basis precisely if a camera trajectory is random. An interesting real world example of this case occurs when many independent photographers take asynchronous images of the same event from different locations.

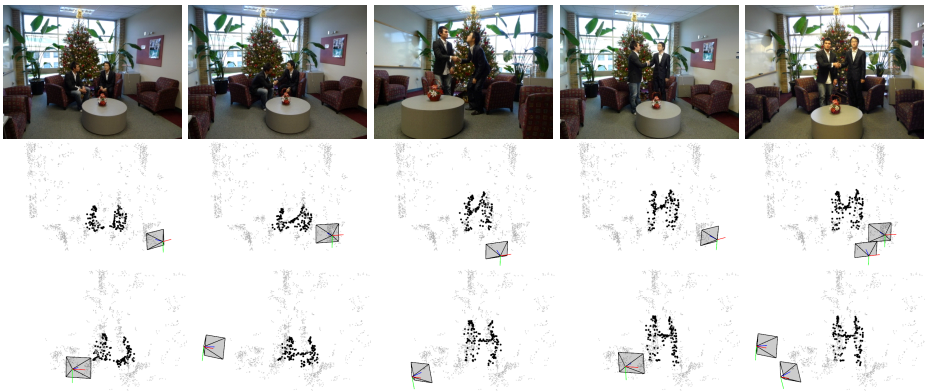


Fig. 5. Results of the handshake scene. Top row: sampled image input, second and third row: five snap shots of 3D reconstruction in different views.

Table 1. Parameters of real data sequences.

	F (sec)	# of photos	# of photographers	K
Rock climbing	39	107	5	12
Handshake	10	32	3	6
Speech	24	67	4	14
Greeting	24	66	4	10

A collection of asynchronous photos can be interpreted as the random motion of a camera center. Using multiple photographers, we collected data in several 'media event' scenarios: a person *rock climbing*, a photo-op *hand shake*, public *speech*, and *greeting*. The static scene reconstruction is based on the structure from motion algorithm described in [20]. We also extracted timing information from image EXIF tags. Correspondences of moving points across images were obtained manually.

The parameters for each scenario are summarized in Table 1. The number of bases was selected empirically for each case. We were able to use the original DCT basis for all scenes. Figures 4, 5, 6, and 7 show some of input images and reconstructed point trajectories. The reconstructed point trajectories look similar to postures of the person.

6 Conclusion

In this paper, we analyze the geometry of 3D trajectory reconstruction and define a measure called reconstructibility to determine the accuracy of 3D trajectory reconstruction. We demonstrate that 3D trajectory reconstruction is fundamentally limited by the correlation between the 3D trajectory of a point and the 3D trajectory of the camera centers. Using this analysis, we propose an algorithm

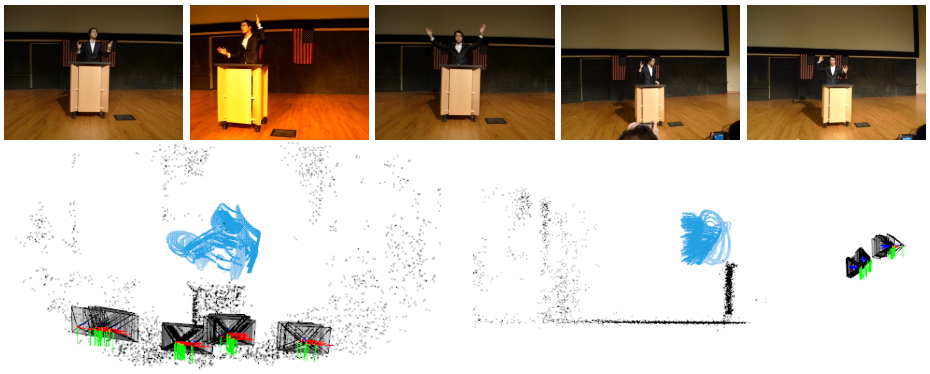


Fig. 6. Results of the speech scene. Top row: sampled image input, and bottom row: reconstructed trajectories (blue line) in different views.

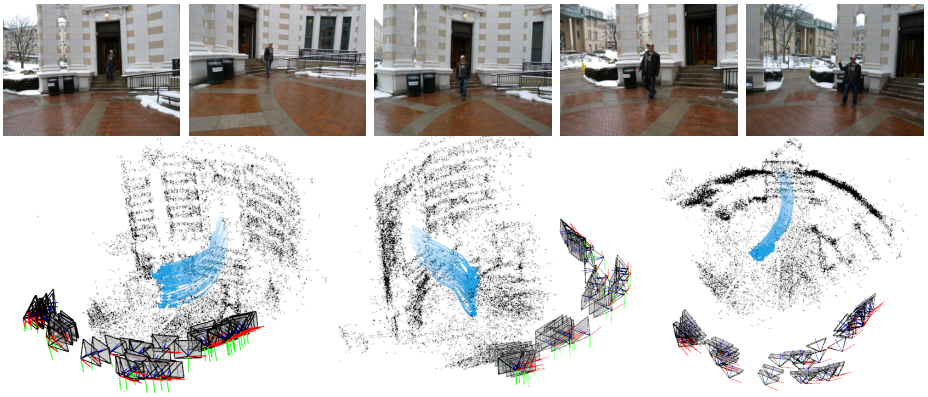


Fig. 7. Results of the greeting scene. Top row: sampled image input, and bottom row: reconstructed trajectories (blue line) in different views.

to reconstruct the 3D trajectory of a moving point from perspective images. By constraining the solution space using a linear trajectory basis, the dimensionality of the solution space can be reduced so that an overconstrained linear least squares system can be formulated. The linear algorithm takes as input the camera pose at each time instant, and a predefined trajectory basis. These requirements are met in our practical application, where we reconstruct dynamic scene from collections of images captured by a number of photographers. We estimate the relative camera pose by applying robust structure from motion to the static points in the scene. The Discrete Cosine Transform is used as a predefined basis. As the effective camera trajectory is quite discontinuous, we are able to obtain accurate 3D reconstructions of the dynamic scenes.

Acknowledgements

The authors wish to thank Natasha Kholgade, Ijaz Akhter, and the anonymous reviewers for their invaluable comments. This work was supported by NSF grant IIS-0916272.

References

1. Shashua, A., Wolf, L.: Homography tensors: On algebraic entities that represent three views of static or moving planar points. In: ECCV. (2000)
2. Hartley, R., Vidal, R.: Perspective nonrigid shape and motion recovery. In: ECCV. (2008)
3. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
4. Longuet-Higgins, H.: A computer algorithm for reconstructing a scene from two projections. Nature (1981)
5. Faugeras, O., Luong, Q.T.: The geometry of multiple images. MIT Press (2001)
6. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An invitation to 3-D vision: From images to Geometric Models. Springer (2001)
7. Avidan, S., Shashua, A.: Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. TPAMI **22** (2000) 348–357
8. Kaminski, J.Y., Teicher, M.: A general framework for trajectory triangulation. J. Math. Imaging Vis. **21** (2004) 27–41
9. Wolf, L., Shashua, A.: On projection matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2, k = 3, \cdots, 6$, and their applications in computer vision. In: ICCV. (2002)
10. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: CVPR. (2000)
11. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. IJCV **9** (1992) 137–154
12. Xiao, J., Kanade, T.: Non-rigid shape and motion recovery: Degenerate deformations. In: CVPR. (2004)
13. Bue, A.D.: A factorization approach to structure from motion with shape priors. In: CVPR. (2008)
14. Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P.: Coarse-to-fine low-rank structure-from-motion. In: CVPR. (2008)
15. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. TPAMI **30** (2008) 878–892
16. Paladini, M., Del Bue, A., Stošić, M., Dodig, M., Xavier, J., Agapito, L.: Factorization for non-rigid and articulated structure using metric projections. In: CVPR. (2009)
17. Akhter, I., Sheikh, Y., Khan, S., Kanade, T.: Nonrigid structure from motion in trajectory space. In: NIPS. (2008)
18. Vidal, R., Abretske, D.: Nonrigid shape and motion from multiple perspective views. In: ECCV. (2006)
19. Akhter, I., Sheikh, Y., Khan, S.: In defense of orthonormality constraints for nonrigid structure from motion. In: CVPR. (2009)
20. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. SIGGRAPH **25** (2006) 835–845