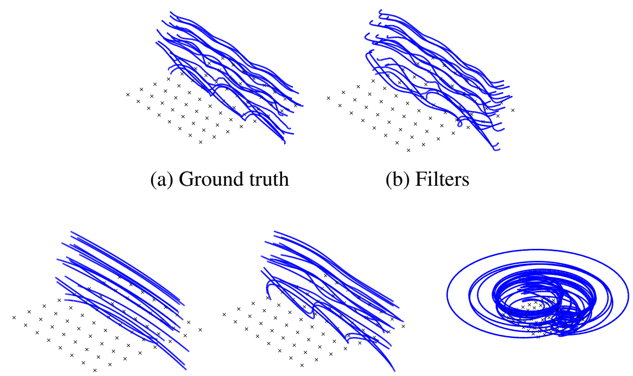
General Trajectory Prior for Non-Rigid Reconstruction

Jack Valmadre[1,2] Simon Lucey[2,1]
[1]Queensland University of Technology, Brisbane, Australia
[2]Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia
{jack.valmadre, simon.lucey}@csiro.au

Abstract

Trajectory basis Non-Rigid Structure From Motion (NRSFM) currently faces two problems: the limit of reconstructability and the need to tune the basis size for different sequences. This paper provides a novel theoretical bound on 3D reconstruction error, arguing that the existing definition of reconstructability is fundamentally flawed in that it fails to consider system condition. This insight motivates a novel strategy whereby the trajectory's response to a set of high-pass filters is minimised. The new approach eliminates the need to tune the basis size and is more efficient for long sequences. Additionally, the truncated DCT basis is shown to have a dual interpretation as a high-pass filter. The success of trajectory filter reconstruction is demonstrated quantitatively on synthetic projections of real motion capture sequences and qualitatively on real image sequences.



(a) Ground truth (b) Filters

(c) $K = 3$ bases (d) $K = 5$ bases (e) $K = 8$ bases

Figure 1: Ground truth and reconstructions of a person walking. If the basis dimension K is chosen too small (c), then it cannot accurately represent the trajectory. If chosen too large (e), then the system may be ill-conditioned. Even when chosen well (d), the result can be over-smoothed. Trajectory filters (b) provide increased model capacity and remove the need to tune this parameter.

1. Introduction

Trajectory basis NRSFM [1] proposes to reconstruct deformable objects from video by restricting the solution to a known low-dimensional subspace of smooth 3D trajectories. It currently faces two major problems. The first is the limit of reconstructability [12] which states that in order to obtain a good reconstruction, the point trajectory must be well-described by the basis while the camera trajectory must not. The second problem is that, while the basis type is sequence agnostic, the basis dimensionality K depends on camera motion, object motion and sequence length, and therefore must be tuned to each sequence.

This paper argues that the definition of reconstructability given in [12] is fundamentally flawed in that it critically fails to consider the condition of the resulting system of equations. We derive a novel theoretical bound on 3D reconstruction error which highlights the hazard of a poorly-conditioned system. Although the new bound offers an effective strategy for automatically choosing K, this insight into the reconstructability problem leads us to explore alternatives to subspace prior. In particular, we propose to minimise the response of the trajectory to one or more high-pass filters, which do not need to be tuned or adjusted for different sequences. The compact temporal support of these filters yields better computational efficiency for long sequences and may enable exploration of online solutions.

This paper is structured as follows. Related work is reviewed in §2. The problem is formally defined in §3 before briefly revising the truncated basis solution in §4. In §5 we introduce our general form and revisit the definition of reconstructability in [12] before developing our novel bound. A simple adaptive strategy which uses this definition is compared to a sparse coding approach in §6. In §7 we introduce our filter-based prior and highlight the connection to a DCT basis. Experimental validation is described in §8 and practical examples are given. Conclusions and future work are discussed in §9.

2. Related Work

Bregler et al. introduced shape basis NRSFM in their seminal paper [3], learning a low-dimensional, object-specific subspace from a sequence of projections. Further work by Torresani et al. adopted the low-rank constraint to assist in non-rigid tracking [19], incorporated temporal constraints by modelling the shape basis coefficients as a linear dynamical system [17] and modelled the distribution of non-rigid deformation by a hierarchical prior [18]. Bartoli et al. [2] used a temporal smoothness prior to reduce the sensitivity of their solution to the number of shape bases. Rabaud and Belongie [14] shifted away from the linear basis interpretation, proposing to learn a smooth manifold of shape configurations from video. They incorporated temporal regularisation to prevent the camera and structure from changing excessively between frames.

More recently, Akhter et al. [1] proposed that the trajectory of each point could instead be restricted to a low-dimensional subspace. The key advantage of this approach over a shape basis is that an object-agnostic basis such as the Discrete Cosine Transform (DCT) can be employed. Gotardo and Martinez [6] recently combined shape and trajectory basis approaches, describing the shape basis coefficients with a DCT basis over time. In their subsequent notable work [5], they extend this approach to non-linear shape models using kernels.

Park et al. [12] examined the limitations of trajectory basis NRSFM in solving for structure given known cameras. They recognised that it is difficult to obtain a good reconstruction when the motion of the point and the camera are correlated, introducing the notion of reconstructability. In subsequent work, Park and Sheikh [11] note that reconstructability is not a problem for the special case where the point correspondences belong to a known articulated structure. Zhu et al. [20] showed that sequences with poor reconstructability could be salvaged by injecting rigid keyframes. They also addressed the need to choose K by instead using the full DCT basis and applying ℓ_1-norm regularisation to the vector of coefficients to find a sparse solution. However, their approach does not take advantage of the known distribution of DCT coefficients in natural signals [4].

Salzmann and Urtasun [15] recently proposed a solution for “3D tracking” which is similar to our method of trajectory reconstruction. They imposed ℓ_1, ℓ_2 and grouped sparsity priors on particle accelerations, which were estimated by second-order finite differences. They also account for gravity by assuming an upright camera. However, they limit their discussion to a stationary camera and hence do not make the connection to trajectory basis reconstruction and the issue of reconstructability. Without a moving camera, they are limited to reconstruction on a plane or require the use of data-driven models to obtain a 3D solution. We do not consider reconstruction under the ℓ_1-norm of filter responses here as it would greatly complicate the reconstructability conditions.

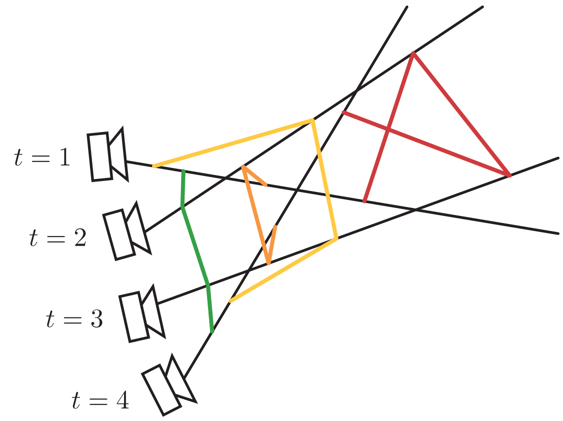


Figure 2: The back-projected rays through an observed point from a known moving camera define a hyperplane of infinite solutions, however we intuitively understand that trajectories are more likely to be slow and smooth (green) than fast and erratic (yellow to red). Trajectory prior defines a likelihood over the space of all trajectories.

3. Problem Formulation

Let $\mathbf{x}_1, \dots, \mathbf{x}_F \in \mathcal{R}^3$ be equally-spaced samples of the position of a 3D point over time, observed by a moving camera as projections $\mathbf{w}_1, \dots, \mathbf{w}_F \in \mathcal{R}^2$. Assuming a pinhole camera model, these are related by the projective equality

$$\begin{bmatrix} \mathbf{w}_t \\ 1 \end{bmatrix} \simeq \mathbf{P}_t \begin{bmatrix} \mathbf{x}_t \\ 1 \end{bmatrix}. \tag{1}$$

Partitioning the projection matrix

$$\mathbf{P}_t = \begin{bmatrix} \mathbf{A}_t & \mathbf{b}_t \\ \mathbf{c}_t^T & d_t \end{bmatrix}, \tag{2}$$

the projective equality in (1) yields the under-determined 2×3 system of linear equations

$$\mathbf{Q}_t \mathbf{x}_t = \mathbf{u}_t, \tag{3}$$

where $\mathbf{Q}_t = \mathbf{A}_t - \mathbf{w}_t \mathbf{c}_t^T$ and $\mathbf{u}_t = d_t \mathbf{w}_t - \mathbf{b}_t$. Each $\mathbf{Q}_t$ matrix has a 1D right nullspace corresponding to the ray connecting the camera center and the projection on the image plane. When $\mathbf{P}_t$ represents an affine camera,

$$\mathbf{P}_t = \begin{bmatrix} \mathbf{R}_t & \mathbf{d}_t \\ \mathbf{0} & 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{R}_t \mathbf{x}_t = \mathbf{w}_t - \mathbf{d}_t. \tag{4}$$

4. Reconstruction Using a Basis

Park et al. [12] constrained the trajectory of each point to lie on a low-dimensional subspace. Let $\mathbf{X} \in \mathcal{R}^{F \times 3}$ be a matrix whose rows are the 3D positions and $\mathbf{\Phi} \in \mathcal{R}^{F \times K}$ be a matrix whose columns are an orthonormal basis for the space of possible trajectories. Trajectory basis approaches assume that there exists some $\mathbf{B} \in \mathcal{R}^{K \times 3}$ such that

$$\mathbf{X} = \mathbf{\Phi}\mathbf{B}, \tag{5}$$

where $\mathbf{B}$ is a matrix of coefficients representing the trajectory in terms of the basis. Defining $\mathbf{x} = \text{vec}(\mathbf{X}^T)$,[1] $\boldsymbol{\beta} = \text{vec}(\mathbf{B}^T)$ and $\mathbf{\Theta} = \mathbf{\Phi} \otimes \mathbf{I}_3$,[2] this can be written

$$\mathbf{x} = \mathbf{\Theta}\boldsymbol{\beta}, \tag{6}$$

while the system of projection equations can be written

$$\mathbf{Q}\mathbf{x} = \mathbf{u}, \tag{7}$$

defining

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_F \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_F \end{bmatrix}. \tag{8}$$

Substituting for $\mathbf{x}$ in (7) yields

$$\mathbf{Q}\mathbf{\Theta}\boldsymbol{\beta} = \mathbf{u}, \tag{9}$$

which is over-determined provided that $\text{rank}(\mathbf{Q}\mathbf{\Theta}) > 3K$, which implies $3K < 2F$. The solution which minimises residual error in the projection equations is

$$\tilde{\mathbf{x}} = \mathbf{\Theta}\tilde{\boldsymbol{\beta}}, \quad \tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Q}\mathbf{\Theta}\boldsymbol{\beta} - \mathbf{u}\|_2^2. \tag{10}$$

5. General Trajectory Prior

Let $\mathbf{\Phi}_\perp \in \mathcal{R}^{F \times (F-K)}$ be a matrix whose columns are an orthonormal basis for the left nullspace of $\mathbf{\Phi}$, such that

$$\mathbf{\Phi}^T \mathbf{\Phi}_\perp = \mathbf{0}, \quad \mathbf{\Phi}_\perp^T \mathbf{\Phi}_\perp = \mathbf{I}. \tag{11}$$

Similarly, let $\mathbf{\Theta}_\perp = \mathbf{\Phi}_\perp \otimes \mathbf{I}_3$ be a matrix whose columns are an orthonormal basis for $\text{null}(\mathbf{\Theta}^T)$. It is trivial to show that

$$\exists\, \boldsymbol{\beta} \text{ s.t. } \mathbf{x} = \mathbf{\Theta}\boldsymbol{\beta} \quad \Leftrightarrow \quad \mathbf{\Theta}_\perp^T \mathbf{x} = \mathbf{0}. \tag{12}$$

Therefore, the problem in (10) can equivalently be stated

$$\begin{aligned} \tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\|\mathbf{Q}\mathbf{x} - \mathbf{u}\right\|_2^2 \\ \text{subject to } \mathbf{\Theta}_\perp^T \mathbf{x} = \mathbf{0}. \end{aligned} \tag{13}$$

[1] The $\text{vec}(\cdot)$ operator stacks the columns of a matrix.

[2] The $\otimes$ operator denotes the Kronecker (tiled) product.

Rather than seeking $\mathbf{x}$ in a low-dimensional space such that the residual error in the projection equations is minimised, we consider seeking $\mathbf{x}$ nearest to the subspace such that projection constraints are satisfied, finding

$$\begin{aligned} \tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\|\mathbf{\Theta}_\perp^T \mathbf{x}\right\|_2^2 \\ \text{subject to } \mathbf{Q}\mathbf{x} = \mathbf{u}. \end{aligned} \tag{14}$$

This formulation may in fact be better motivated, as we are otherwise forced to violate the image measurements in order to adhere to the basis, which is counter-intuitive since the original problem is *under*-constrained. Furthermore, while $\mathbf{Q}\mathbf{x} = \mathbf{u}$ defines the projection constraints in the noiseless case, minimising $\|\mathbf{Q}\mathbf{x} - \mathbf{u}\|_2$ does not necessarily correspond to the optimal solution under an assumption of Gaussian projection noise for a general perspective camera. This is precisely the reason that linear triangulation methods should only be used to provide initialisation for an algorithm which minimises a non-linear geometric error cost function in multiple view reconstruction [7]. It's worth noting, however, that the formulation in (10) remains a more computationally attractive approach to approximately solve this problem.

Based on this reformulation, we propose a general form for linear trajectory reconstruction,[3]

$$\begin{aligned} \tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathbf{M}}^2 \\ \text{subject to } \mathbf{Q}\mathbf{x} = \mathbf{u}, \end{aligned} \tag{15}$$

where $\mathbf{M} = \mathbf{E} \otimes \mathbf{I}_3$ assumes independent, identical prior in each dimension, and $\mathbf{E} \succeq 0$ is the precision (inverse covariance) matrix of a zero-mean Gaussian distribution over one-dimensional trajectories. Reconstruction using a truncated basis is the special case

$$\mathbf{E} = \mathbf{\Phi}_\perp \mathbf{\Phi}_\perp^T = \mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^T. \tag{16}$$

Note that $\mathbf{E}$ (and therefore $\mathbf{M}$) is generally rank-deficient, since at least the DC trajectory should lie in $\text{null}(\mathbf{E})$ to ensure that the prior is invariant to global translation. In the case of a trajectory basis, $\text{rank}(\mathbf{E}) = F - K$.

5.1. Canonical Reconstructability

Park et al. [12] define reconstructability in terms of the trajectory of the camera center $\mathbf{c} \in \mathcal{R}^{3F}$,

$$\eta(\mathbf{x}, \mathbf{c}, \mathbf{\Theta}) = \frac{\left\|\mathbf{\Theta}_\perp^T \mathbf{c}\right\|}{\left\|\mathbf{\Theta}_\perp^T \mathbf{x}\right\|}, \tag{17}$$

such that as $\eta \to \infty$, the basis space is guaranteed to intersect the hyperplane defined by the projection equations

$$\left(\min_{\mathbf{x}, \boldsymbol{\beta}} \|\mathbf{x} - \mathbf{\Theta}\boldsymbol{\beta}\|_2 \text{ subject to } \mathbf{Q}\mathbf{x} = \mathbf{u} \right) \to 0. \tag{18}$$

[3] We adopt $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$.

However, this is a necessary and *not* a sufficient condition for reconstruction error to approach zero. If the intersection of the nullspace of $\mathbf{Q}$ and the column-space of $\mathbf{\Theta}$ is itself a space having non-zero dimension, there will be infinite solutions which satisfy both the projection equations and the basis constraint. Even when the two spaces do not exactly intersect, the solution can be extremely unstable if they are "close" to intersecting. Furthermore, it is unintuitive for the expression in (17) to depend on camera position when it did not appear in the derivation. This prohibits analysis of affine cameras since they do not have a center. In this section we propose an alternative reconstructability for our general form which bounds the reconstruction error, taking into account the system condition and not requiring a camera center.

5.2. Our Measure

While (15) can be solved using Lagrange multipliers, for the purpose of deriving a bound we instead consider parametrising the F-dimensional hyperplane of feasible solutions

$$\mathbf{x} = \mathbf{x}' + \mathbf{Q}_\perp \mathbf{z} \tag{19}$$

where $\mathbf{x}'$ is any solution to $\mathbf{Q}\mathbf{x}' = \mathbf{u}$, $\mathbf{Q}_\perp \in \mathcal{R}^{3F\times F}$ is a matrix whose columns are an orthonormal basis for $\operatorname{null}(\mathbf{Q})$ such that

$$\mathbf{Q}\mathbf{Q}_\perp = \mathbf{0}, \quad \mathbf{Q}_\perp^T \mathbf{Q}_\perp = \mathbf{I}, \tag{20}$$

and $\mathbf{z} \in \mathcal{R}^F$ defines a component of the solution in the nullspace of $\mathbf{Q}$. Solving (15) is exactly equivalent to finding

$$\tilde{\mathbf{x}}(\mathbf{x}') = \mathbf{x}' + \mathbf{Q}_\perp \tilde{\mathbf{z}}(\mathbf{x}') \tag{21}$$

where

$$\begin{aligned} \tilde{\mathbf{z}}(\mathbf{x}') &= \arg\min_{\mathbf{z}} \|\mathbf{x}' + \mathbf{Q}_\perp \mathbf{z}\|_{\mathbf{M}}^2 \\ &= -(\mathbf{Q}_\perp^T \mathbf{M} \mathbf{Q}_\perp)^{-1} \mathbf{Q}_\perp^T \mathbf{M} \mathbf{x}' \end{aligned} \tag{22}$$

regardless of the choice of $\mathbf{x}'$, since (21) is an affine transform and the objective in (15) is convex. Considering the case where $\mathbf{x}'$ is the ground truth trajectory $\mathbf{x}$, we obtain an expression for the reconstruction error

$$\|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{x})\|_2 = \left\|(\mathbf{Q}_\perp^T \mathbf{M} \mathbf{Q}_\perp)^{-1} \mathbf{Q}_\perp^T \mathbf{M} \mathbf{x}\right\|_2. \tag{23}$$

This facilitates the definition of an upper bound υ on reconstruction error[4]

$$\begin{aligned} \upsilon(\mathbf{x}, \mathbf{Q}, \mathbf{M}) &= \left\|(\mathbf{Q}_\perp^T \mathbf{M} \mathbf{Q}_\perp)^{-1}\right\|_2 \left\|\mathbf{Q}_\perp^T \mathbf{M} \mathbf{x}\right\|_2 \\ &= \underbrace{\operatorname{cond}(\mathbf{Q}_\perp^T \mathbf{M} \mathbf{Q}_\perp)}_{\text{gain } \gamma} \underbrace{\frac{\left\|\mathbf{Q}_\perp^T \mathbf{M} \mathbf{x}\right\|_2}{\left\|\mathbf{Q}_\perp^T \mathbf{M} \mathbf{Q}_\perp\right\|_2}}_{\text{contradiction } \varepsilon}, \end{aligned} \tag{24}$$

[4]We adopt the convention that the norm of a matrix is that induced by the corresponding vector norm, $\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{0}} \|\mathbf{A}\mathbf{x}\|/\|\mathbf{x}\|$.

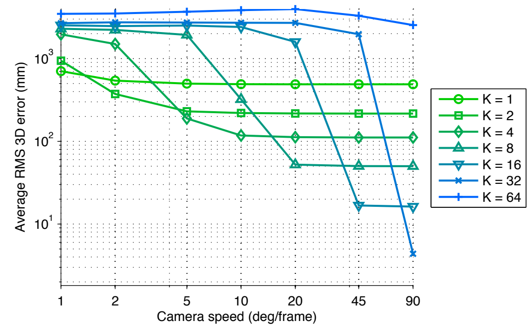


Figure 3: Reconstruction error versus orbiting camera speed for DCT basis of varying size. When using a truncated DCT basis to constrain the solution, the motion of the camera critically affects the choice of basis size K. The true limit of reconstructability is a lower bound on these curves and should be independent of K.

such that

$$\|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{x})\|_2 \le \upsilon(\mathbf{x}, \mathbf{Q}, \mathbf{M}), \tag{25}$$

where the condition number

$$\operatorname{cond}(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})} \tag{26}$$

provides a bound on the perturbed solution of a system of linear equations. Note that our measure extends to affine cameras, as it does not depend on a camera center.

The contradiction term $\varepsilon \ge 0$ reflects how much the unobserved component disagrees with the prior. When the prior is a truncated basis, the numerator becomes

$$\left\|\mathbf{Q}_\perp^T(\mathbf{I} - \mathbf{\Theta}\mathbf{\Theta}^T)\mathbf{x}\right\|_2, \tag{27}$$

the component of the trajectory which is orthogonal to both the basis and the projection matrix.

The gain term $\gamma \ge 1$, dependent solely on the cameras and the prior, becomes large for a poorly-conditioned system of equations. When the nullspace of $\mathbf{Q}$ and the nullspace of $\mathbf{M}$ intersect non-trivially (at a subspace with non-zero dimension), this term approaches infinity and the reconstruction error is unbounded. This situation corresponds to a set of cameras for which there exists a non-zero trajectory in the nullspace of the prior which would go unobserved (and therefore at least a 1D space of such trajectories exists). When using a trajectory basis, the gain term is monotonically increasing with K.

Our definition of reconstructability highlights the critical nature of choosing the basis size K in trajectory basis reconstruction. If K is chosen too small, the trajectory is poorly represented by the basis, but if it is chosen too large, the system is ill-conditioned and the reconstruction error becomes unbounded. Figure 3 confirms that this trade-off is

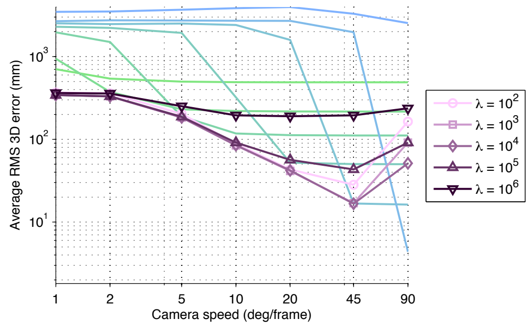


Figure 4: Reconstruction error versus orbiting camera speed for the sparse coding algorithm with varying regularisation coefficients. Promoting sparsity in the vector of coefficients has the effect of automatically selecting which basis vectors to use.

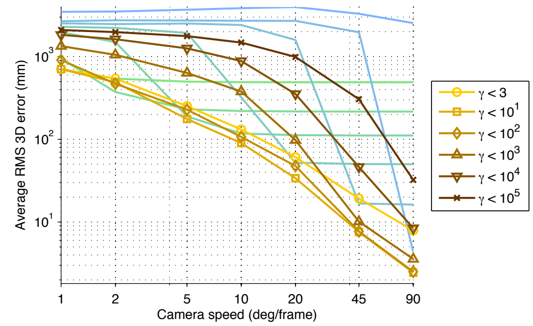


Figure 5: Reconstruction error versus orbiting camera speed using a DCT basis with size automatically determined by a gain threshold. This simple strategy is at least as effective as sparse coding for non-negligible camera motion.

observed in practice. The full details of this experiment are found in §8.1.

6. Sparse Coding and an Adaptive Strategy

Zhu et al. [20] recently used a complete DCT basis and applied ℓ_1-norm regularisation to the vector of coefficients to promote sparsity, finding

$$\tilde{\mathbf{x}} = \boldsymbol{\Theta}\tilde{\boldsymbol{\beta}}, \quad \tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Q}\boldsymbol{\Theta}\boldsymbol{\beta} - \mathbf{u}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 . \quad (28)$$

The feature-sign search algorithm [9] was used to minimise this objective. Figure 4 shows that this is an effective strategy for automatically selecting basis vectors. We postulate that the reason sparse coding is effective in this context is that it indirectly controls the conditioning problem.

While the assumption of sparse coefficients is well-founded because the DCT tends to concentrate non-zero entries at lower frequencies for natural signals [4], a sparse coding approach completely ignores this known pattern of sparsity. Examining the bound in (24), one can imagine a simple adaptive strategy which chooses the largest K such that $\gamma(\mathbf{Q}, \mathbf{M}) < \gamma_{\max}$. Figure 5 shows that this is at least as effective as the sparse coding approach. The efficacy of this approach is encouraging evidence for the bound from the previous section being reasonably tight.

7. Filters Instead of Bases

Truncated basis approaches are especially prone to the conditioning problem because the nullspace of $\mathbf{M}$ has dimension $3K$, making intersection with the nullspace of $\mathbf{Q}$ increasingly likely with larger K. An alternative way to encourage smooth motion is to penalise the response of the trajectory to a set of compact high-pass filters. Filters are elegant in that they enforce temporal constraints locally rather than globally, extending trivially to sequences of different length. Some simple filters have physical motivation. Minimising the ℓ_2-norm of the second derivative corresponds to an assumption of constant mass subject to isotropic Gaussian distributed forces [15] and minimising the ℓ_2-norm of the first derivative corresponds to finding the solution with the least kinetic energy.

Convolution is a linear operation and can therefore be written as a matrix multiplication

$$\mathbf{g} * \mathbf{h} = \mathbf{G}\mathbf{h}, \quad (29)$$

where

$$\mathbf{G} = \begin{bmatrix} g_M & \cdots & g_1 & & \\ & \ddots & \ddots & \ddots & \\ & & g_M & \cdots & g_1 \end{bmatrix}, \quad (30)$$

$\mathbf{g} \in \mathcal{R}^M$, $\mathbf{h} \in \mathcal{R}^N$ and $M \leq N$ is the support of the filter. Since we use the variant of convolution in which only the entirely overlapping segments of the response are taken, the convolution matrix $\mathbf{G}$ is of size $(N - M + 1) \times N$. Minimising the magnitude of the response of each dimension of the trajectory to a filter $\mathbf{g}$ is a special case of the general form in (15) where $\mathbf{E} = \mathbf{G}^T\mathbf{G}$. Multiple filters can be realised by taking a linear combination of such matrices.

When $\mathbf{M}$ is constructed using filters, its nullspace has dimension $3(M - 1)$. For compact filters, this leads to a much lower-dimensional nullspace than the truncated basis approach, reducing the likelihood of an intersection with the nullspace of the projection matrix. This contrast is clearly observed by comparing the eigenspectra of first- and second-difference filters (Figures 7b and 7d) to that of a truncated basis (Figure 8a). The generalisation ability of trajectory filters is partially attributed to the low-dimensional nullspace of $\mathbf{M}$. Figure 6 demonstrates empirically that simple filters achieve reconstruction error at the

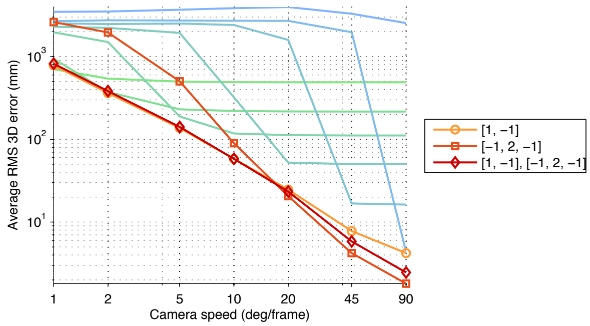


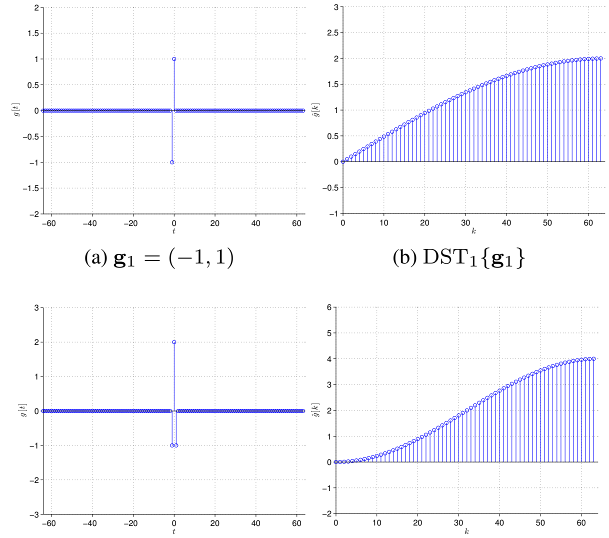
Figure 6: Reconstruction error versus orbiting camera speed for simple trajectory filters. Superior results are achieved over all sequences without having to tune any parameters.

lower limit of other methods without requiring any parameter tuning. In our experiments we only consider the filters $(-1, 1)$ and $(-1, 2, -1)$ since these estimate the first- and second-derivative respectively, each with minimal support.

7.1. Symmetric Convolution and the DCT

The basis which was first used by Akhter et al. [1] and has generally been adopted in literature [6, 12, 20] is the DCT. It was chosen for its ability to compactly describe natural signals in lossy compression [4]. In the same way that the DFT (Discrete Fourier Transform) diagonalises periodic convolution, the DCT and DST (Discete Sine Transform) diagonalise symmetric-periodic convolution [10], their bases forming the eigenvectors of symmetric convolution matrices. Using a generalised form of Parseval's theorem, it is possible to examine the spectra of the proposed filters (Figure 7) and find an equivalent filter to the truncated DCT basis (Figure 8).

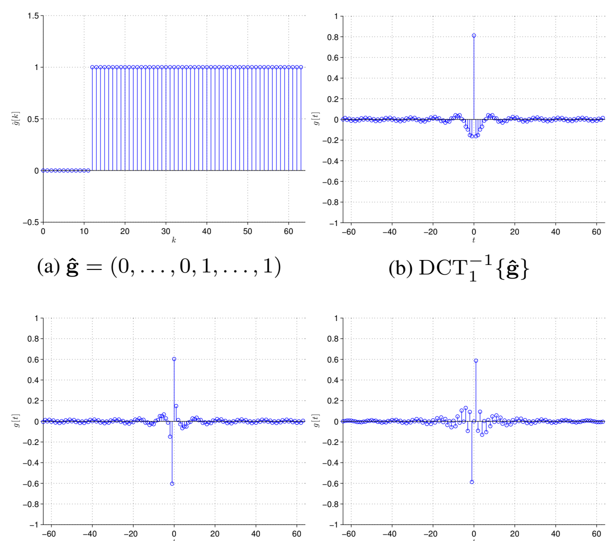
It is worth noting that multiplication in the DCT domain actually computes convolution with the symmetric extension of the trajectory. Combining this with a smooth-motion prior unavoidably induces stationary boundary conditions at either end. Although past work has noted that the type of transform is not important [1], we point out that the DFT and DST imply periodic and asymmetric extension of a signal, respectively. Optimising for smooth motion over these extensions requires that trajectories start and end at the same place and at the origin, respectively. These would clearly be poor assumptions, thus motivating the specific use of the DCT basis. Trajectory filters do not necessarily impose any constraints on boundary conditions, although we find a small component of $(-1, 1)$ filter response can be added to ensure proper regularisation when the camera is slow (observed in Figure 6), which tends to induce similar stationary boundary conditions.



(a) $\mathbf{g}_1 = (-1, 1)$ (b) $\mathrm{DST}_1\{\mathbf{g}_1\}$

(c) $\mathbf{g}_2 = (-1, 2, -1)$ (d) $\mathrm{DCT}_1\{\mathbf{g}_2\}$

Figure 7: Impulse responses and appropriate transforms of first- and second-difference filters. The transform gives the eigenvalues of the symmetric convolution matrix.



(a) $\hat{\mathbf{g}} = (0, \ldots, 0, 1, \ldots, 1)$ (b) $\mathrm{DCT}_1^{-1}\{\hat{\mathbf{g}}\}$

(c) $\mathrm{DST}_2^{-1}\{\hat{\mathbf{g}}\}$ (d) $\mathrm{DST}_1^{-1}\{\hat{\mathbf{g}}\}$

Figure 8: The truncated basis approach is equivalent to applying a "brick-wall" high-pass filter in the DCT domain. This transform has multiple interpretations as a symmetric convolution, each approximating a compact-support filter.

7.2. Computational Considerations

From the previous section, we can still entertain a truncated basis representation to reduce the dimensionality, provided that symmetric extension of the trajectory can be assumed and an ℓ_2-norm regularisation term is included, weighted by the transform of the desired filter $\hat{\mathbf{g}} = \mathcal{T}\{\mathbf{g}\}$. The resulting objective is

$$\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Q}\boldsymbol{\Theta}\boldsymbol{\beta} - \mathbf{u}\|_2^2 + \lambda \|\mathrm{diag}(\hat{\mathbf{g}})\boldsymbol{\beta}\|_2^2 . \tag{31}$$

As in the original trajectory reconstruction paper [12], this requires inversion of a dense $3K \times 3K$ matrix. Assuming that the basis size must be chosen as some small constant fraction of the number of frames $K = kF$ (as in [6]), the asymptotic time complexity of inverting this matrix is $\mathcal{O}(F^3)$.

We instead consider solving (15) directly for the case where $\mathbf{M}$ is constructed using compact filters. The method of Lagrange multipliers yields the sparse linear system

$$\begin{bmatrix} \mathbf{M} & \mathbf{Q}^T \\ \mathbf{Q} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{u} \end{bmatrix} . \tag{32}$$

Since $\mathbf{M}$ is banded-diagonal and $\mathbf{Q}$ is block-diagonal, the columns and rows can be re-ordered to give a matrix with band $\mathcal{O}(M)$. Solving this banded-diagonal system of equations has time complexity $\mathcal{O}(F^2 M)$ [13]. Therefore, applying compact filters in the time domain is more computationally attractive than using a low-dimensional basis for long sequences. Furthermore, efficient parallel solvers exist for diagonally banded systems [8].

8. Experimental Validation

8.1. Synthetic Projection of MoCap Sequences

The experiment in Figures 3, 4, 5 and 6 consists of a 100-frame sequence from the CMU MoCap database,[5] observed by a perspective camera orbiting the subject on a horizontal plane at speeds varying from 1 to 90 degrees per frame, with results averaged over 100 different sequences. Error is measured by the RMS 3D distance between points over all frames.

8.2. Real Experiments

Our filter algorithm was qualitatively compared to the orthonormal truncated basis algorithm of Park et al. [12] across a number of sequences. A representative example is shown in Figure 9. We typically observe that while the low-dimensional DCT solution is smoother, the filter algorithm produces a more realistic trajectory. For example, note the triangular path of the feet and the more complex path of the swinging arm. More importantly, however, our filter-based method did not require any hand-tuning.

[5] `http://mocap.cs.cmu.edu/`

9. Conclusion and Future Work

This paper has brought to light the critical oversight of system conditioning in trajectory basis NRSFM, which has confused the issue of reconstructability. A theoretical bound on reconstruction error which considers system condition and extends to affine cameras has been established. Based on this insight, a novel approach to trajectory reconstruction using high-pass filters was proposed which eliminates the need to choose basis size and has better asymptotic time complexity.

This work motivates a number of possible future directions. The compact support of the filters may assist integration of non-rigid reconstruction into real-time monocular pose estimation techniques. The duality with the DCT could be combined with Probabilistic Principal Component Analysis [16] to estimate the inertial reference frame from a sequence of projections. Although gravity, electromagnetic and biomechanical forces roughly follow an isotropic Gaussian distribution, collision forces typically act in sparse impulses, motivating investigation into combined ℓ_1-ℓ_2 strategies for the second derivative and alternative norms in general. Finally, while existing approaches to unify shape and trajectory bases apply temporal prior in shape space, trajectory filters provide a way to enforce temporal prior in world space, by filtering the output of a shape basis.

10. Acknowledgements

This research was supported under an Australian Research Council Future Fellowship (project FT0991969).

References

[1] I. Akhter, Y. A. Sheikh, S. Khan, and T. Kanade. Non-rigid structure from motion in trajectory space. In *NIPS*. MIT Press, 2008.

[2] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. I. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *CVPR*. IEEE, June 2008.

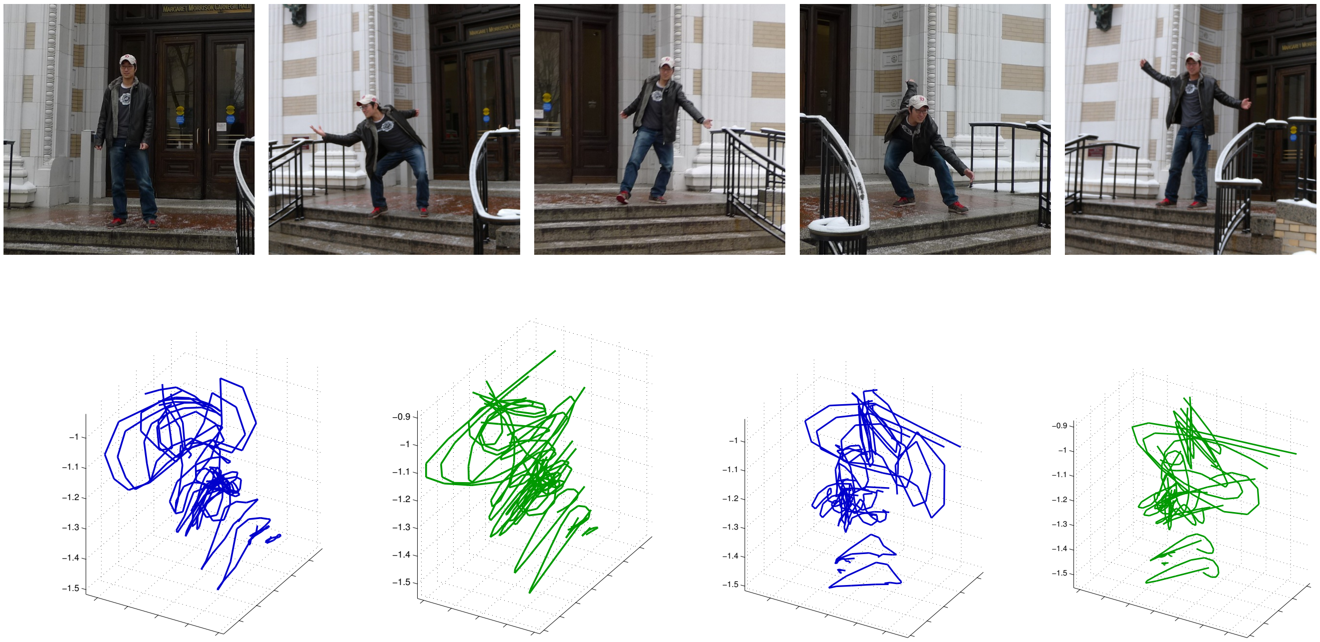
[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, volume 2, pages 690–696. IEEE, 2000.

[4] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2002.

[5] P. F. U. Gotardo and A. M. Martinez. Kernel non-rigid structure from motion. In *ICCV*. IEEE, 2011.

[6] P. F. U. Gotardo and A. M. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *CVPR*, pages 3065–3072. IEEE, June 2011.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.



(a) (b) (c) (d)

Figure 9: Trajectory filtering (a, c) and a truncated DCT basis (b, d) achieve similarly plausible reconstruction on real-world examples where no ground-truth is available. The filter-based reconstruction used the sum of the response of $(-1, 1)$ and $(-1, 2, -1)$ filters and did not require any hand-tuning. A $K = 6$ dimensional DCT basis had to be chosen by trial and error.

[8] D. H. Lawrie and A. H. Sameh. The computation and communication complexity of a parallel banded system solver. *ACM Transactions on Mathematical Software*, 10(2):185–195, May 1984.

[9] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *NIPS*, pages 801–808. MIT Press, 2007.

[10] S. A. Martucci. Symmetric convolution and the discrete sine and cosine transforms. *IEEE Transactions on Signal Processing*, 42(5):1038–1051, May 1994.

[11] H. S. Park and Y. A. Sheikh. 3D reconstruction of a smooth articulated trajectory from a monocular image sequence. In *ICCV*, pages 201–208. IEEE, Nov. 2011.

[12] H. S. Park, T. Shiratori, I. Matthews, and Y. A. Sheikh. 3D reconstruction of a moving point from a series of 2D projections. In *ECCV*, volume 6313, pages 158–171. Springer, 2010.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.

[14] V. Rabaud and S. J. Belongie. Re-thinking non-rigid structure from motion. In *CVPR*. IEEE, June 2008.

[15] M. Salzmann and R. Urtasun. Physically-based motion models for 3D tracking: A convex formulation. In *ICCV*, pages 2064–2071. IEEE, Nov. 2011.

[16] M. E. Tipping and C. M. Bishop. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society*, 61(3):611–622, 1999.

[17] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3D shape from 2D motion. In *NIPS*. MIT Press, 2004.

[18] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *PAMI*, 30(5):878–92, May 2008.

[19] L. Torresani, D. Yang, E. J. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *CVPR*, volume 1, pages 493–500. IEEE, 2001.

[20] Y. Zhu, M. Cox, and S. Lucey. 3D motion reconstruction for real-world camera motion. In *CVPR*. IEEE, June 2011.