In Defense of Orthonormality Constraints for Nonrigid Structure from Motion

Ijaz Akhter[1]
akhter@lums.edu.pk

Yaser Sheikh[2]
yaser@cs.cmu.edu

Sohaib Khan[1]
sohaib@lums.edu.pk

[1]LUMS School of Science and Engineering, Lahore, Pakistan
[2]The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA
http://cvlab.lums.edu.pk

Abstract

In factorization approaches to nonrigid structure from motion, the 3D shape of a deforming object is usually modeled as a linear combination of a small number of basis shapes. The original approach to simultaneously estimate the shape basis and nonrigid structure exploited orthonormality constraints for metric rectification. Recently, it has been asserted that structure recovery through orthonormality constraints alone is inherently ambiguous and cannot result in a unique solution. This assertion has been accepted as conventional wisdom and is the justification of many remedial heuristics in literature. Our key contribution is to prove that orthonormality constraints are in fact sufficient to recover the 3D structure from image observations alone. We characterize the true nature of the ambiguity in using orthonormality constraints for the shape basis and show that it has no impact on structure reconstruction. We conclude from our experimentation that the primary challenge in using shape basis for nonrigid structure from motion is the difficulty in the optimization problem rather than the ambiguity in orthonormality constraints.

1. Introduction

Given 2D trajectories of world points in a video sequence, the purpose of nonrigid structure from motion (NRSFM) is to estimate the 3D trajectories and the camera projection matrices corresponding to each frame. In their seminal work on NRSFM, Bregler *et al.* [4] proposed that the nonrigid 3D structure can be approximated by the linear combination of a small set of basis shapes. They used an SVD-based factorization approach similar to that of Tomasi and Kanade [7], and exploited the orthonormality of camera rotations to compute the metric rectification. In subsequent work, Xiao *et al.* in [9], asserted that orthonormality constraints alone are not enough to estimate a unique solution of 3D structure and characterized the nature of ambiguity in using these constraints.

The ambiguity theorem introduced by Xiao *et al.* has greatly influenced research on NRSFM. They proposed additional constraints termed basis constraints to resolve this ambiguity to get a unique linear solution. In addition, several new constraints have also been proposed. For example, Torresani *et al.* [8] proposed a Gaussian prior for the shape coefficients. Del Bue *et al.* [5] assumed that nonrigid shape contained a significant number of points which behave in a rigid fashion. Recently, Bartoli *et al.* [2] proposed that we can iteratively find the shape basis based on the assumption that each basis shape should decrease the reprojection error left due to the earlier basis. While good results have been demonstrated for specific cases, each of these additional constraints introduces restrictions on the nature of the nonrigid deformation or makes additional assumptions about the deforming 3D structure, impacting the generality of the method.

Our key contribution in this paper is to show that by exploiting an inherent rank three constraint, the recovery of structure from orthonormality constraints alone is infact *unambiguous*, other than a 3×3 rotation. The general solution of the orthonormality constraints given by Xiao *et al.* [9] is simplified considerably when this rank constraint is imposed. We show that the ambiguity in orthonormality constraints is based on an intrinsic ambiguity of vector basis: any linear transformation on the basis vectors gives another valid basis. While an ambiguity in the solution of the constraints does exist, it does not result in ambiguous structure. In order to support this claim, we provide analytical proof, as well as empirical evidence. Additional empirical evidence is the optimization method proposed by Brand in [3], which uses only orthonormality constraints to resolve the underlying optimization problem. Our paper provides a principled explanation of the apparent contradiction between Brand's approach and the assertion by Xiao *et al.* that orthonormality constraints alone cannot be sufficient for the correct recovery of nonrigid structure.

2. Summary of Results

The key results of this paper are on characterizing the ambiguity of orthonormality constraints for nonrigid structure reconstruction. For the demonstration of results, we use the cube and moving points dataset which Xiao *et al.* used in [9] as an illustrative example. It consists of a rigid cube in the middle and three points moving along the coordinate axes with different speeds. The overall structure of these points is nonrigid because of the motion of the three points relative to the cube. We create a synthetic orthographic camera moving randomly around the cube to generate images.

The first result in this paper is to show that the ambiguity of the orthonormality constraints does not matter for structure reconstruction. Perfect structure reconstruction is possible by using orthonormality constraints alone, without using any additional constraints. The reason for this is that the ambiguity in the solution space of orthonormality constraints results in a different solution of the shape coefficients. However, the solution for the shape basis is affected exactly by the inverse of this ambiguity, hence having no impact on the recovery of structure. The ambiguous shape basis are still a valid set of basis to represent structure, and are, in fact, just a linearly transformed version of the original basis. In Figure 1(a), we show the original structure and the reconstructed structure at one time instance, computed by the method of Bregler *et al.* [4] using two shape basis. Figure 1(b) shows the structure obtained from our optimization, which is based on orthonormality constraints alone. Figure 1(c) compares the structure reconstruction between these two methods. The error is significantly larger in Bregler's Method as noted correctly by Xiao *et al.* [9]. Figure 1(d-f) shows three possible solutions of orthonormality constraints, demonstrating that the ambiguity in orthonormality constraints does not translate to an ambiguity in structure reconstruction, other than a 3×3 rotation. This indicates that orthonormality constraints are sufficient for perfect structure reconstruction.

Our second result is to describe the true ambiguity of orthonormality constraints. Xiao *et al.* in [9] characterized the ambiguity in these constraints and described a general form of the solution space of orthonormality constraints. However, we note that they did not impose a necessary rank three constraint on this solution space. If we impose this additional constraint, we show that the form of ambiguity is considerably simplified. Moreover, we also show that this simplified form of ambiguity is intuitively understandable from the nature of orthonormality constraints themselves. This gives an insight into the source ambiguity even without going through the analytical proofs.

Our third contribution is the insight that the real problem in NRSFM is the difficulty in optimization and not the ambiguity. This suggests that we should look for better ways of
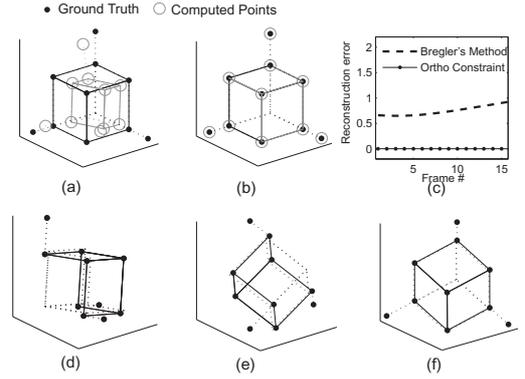

Figure 1. Structure reconstruction on the cube and moving points example of [9]. (a) The original structure and the aligned reconstruction using Bregler *et al.*'s method [4] on one of the frames. (b) The result obtained from the optimization of a cost function based only on orthonormality constraints. (c) The average reconstruction error over 16 frames for both methods. (d-f) Three structure reconstructions using different ambiguous solutions of orthonormality constraints. All three solutions are identical except for a rotation with respect to each other.

optimization rather than introducing the constraints which restrict the generality of the idea of shape basis. Although, it is true that additional constraints reshape the optimization surface and help in estimating the solution, they should not shift the location of the global minimum given by the orthonormality constraints.

A dual approach of solving for nonrigid structure is to use the trajectory basis instead of the shape basis, as formulated recently by Akhter *et al.* [1]. They also use just the orthonormality constraints to solve for structure. Our final result is to show that unlike the shape basis, the solution space of orthonormality constraints for trajectory basis is not ambiguous at all. Thus, orthonormality constraints are sufficient for both shape and trajectory basis approaches of NRSFM.

3. Shape Basis Formulation

The nonrigid 3D structure at any time instance t can be denoted by a $3 \times P$ matrix $S(t)$, which is a horizontal concatenation of 3D locations of P points, with t ranging from 1 to F frames. In Bregler *et al.*'s shape basis approach [4] for nonrigid structure from motion, the structure at each time instance is represented as a linear combination of a small number of shape basis $S_1, S_2, \ldots, S_K$,

$$S(t) = \sum_{i=1}^{K} c_{ti} S_i \qquad S(t), S_i \in \mathbb{R}^{3 \times P}, c_{ti} \in \mathbb{R}, \quad (1)$$

where $K \ll F$. This can be written more compactly as $S(t) = (c_{t1} I \cdots c_{tK} I) \mathbf{B}$, where I is a 3×3 identity ma-

trix and $\mathbf{B} = \left(S_1^T \cdots S_K^T\right)^T$ is a $3 \times P$ matrix containing all shape basis. The overall structure of the F frames can be written as $\mathbf{S} = \mathbf{CB}$, where $\mathbf{S}_{3F \times P}$ and $\mathbf{C}_{3F \times 3K}$ are a concatenation of $3 \times P$ structures and shape coefficients respectively. This structure, when imaged by an orthographic camera, can be compactly represented as $\mathbf{W} = \mathbf{RS}$, where $\mathbf{W}$ represents a $2F \times P$ image observation matrix and $\mathbf{R}$ is a $2F \times 3F$ block diagonal camera matrix containing the first two rows of a rotation matrix on each 2×3 block on the diagonal. $\mathbf{W}$ can be written using $\mathbf{S} = \mathbf{CB}$ as

$$\mathbf{W} = \mathbf{RCB} = \mathbf{MB}, \tag{2}$$

where

$$\mathbf{M}_{2F \times 3K} = \mathbf{RC} = \begin{pmatrix} c_{11}R_1 & \cdots & c_{1K}R_1 \\ \vdots & \ddots & \vdots \\ c_{F1}R_F & \cdots & c_{FK}R_F \end{pmatrix}. \tag{3}$$

Equation 2 shows that the rank of matrix $\mathbf{W}$ will be $3K$ under the compact shape basis model. Therefore, a rank $3K$ factorization of $\mathbf{W}$ can be performed using SVD, given as $\mathbf{W} = \hat{\mathbf{M}}\hat{\mathbf{B}}$. This factorization, however, is not unique and a rectifying transform $\mathbf{G}$ is required to impose the orthonormality in the structure of $\mathbf{M}$, i.e.

$$\mathbf{M} = \hat{\mathbf{M}}\mathbf{G} \quad \text{and} \quad \mathbf{B} = \mathbf{G}^{-1}\hat{\mathbf{B}}. \tag{4}$$

In order to find the constraints on the rectifying transform $\mathbf{G}$ for its estimation, let us denote $\mathbf{G}$ as the horizontal concatenation of $\mathbf{g}_k$ matrices, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K]$, where each column triple $\mathbf{g}_k$ is of size $3K \times 3$. Moreover, let $\hat{\mathbf{M}}_{2i-1:2i}$ denote two consecutive rows in $\hat{\mathbf{M}}$ at position $2i-1$ and $2i$. Then $\hat{\mathbf{M}}_{2i-1:2i}\mathbf{g}_k$ should be a scaled orthonormal matrix i.e. $\hat{\mathbf{M}}_{2i-1:2i}\mathbf{g}_k = c_{ik}R_i$. This results in constraints of the form,

$$\hat{\mathbf{M}}_{2i-1:2i}\mathbf{g}_k\mathbf{g}_k^T\hat{\mathbf{M}}_{2i-1:2i}^T = c_{ik}^2\mathbf{I}_{2\times 2}. \tag{5}$$

In the above equation, c_{ik} are unknowns. However, we can constrain the solution of $\mathbf{g}_k$ by observing that the two diagonal terms on the R.H.S. of Equation 5 are equal, and the two off-diagonal terms are zero. This implies the following set of constraints on $\mathbf{g}_k$:

$$\begin{aligned} \hat{\mathbf{M}}_{2i-1}\mathbf{g}_k\mathbf{g}_k^T\hat{\mathbf{M}}_{2i-1}^T &= \hat{\mathbf{M}}_{2i}\mathbf{g}_k\mathbf{g}_k^T\hat{\mathbf{M}}_{2i}^T \\ \hat{\mathbf{M}}_{2i-1}\mathbf{g}_k\mathbf{g}_k^T\hat{\mathbf{M}}_{2i}^T &= 0. \end{aligned} \tag{6}$$

The constraints given in Equations 6 are called orthonormality constraints because they exploit the orthonormality property of rotations to constrain the solution of $\mathbf{g}_k$. The constraints from all rows of $\mathbf{M}$ can be combined to yield a squared-error cost function

$$\begin{aligned} cost(\mathbf{g}_k) = \sum_{i=1}^{F} &\left(\mathbf{M}_{2i-1}\mathbf{g}_k\mathbf{g}_k^T\mathbf{M}_{2i-1}^T - \mathbf{M}_{2i}\mathbf{g}_k\mathbf{g}_k^T\mathbf{M}_{2i}^T\right)^2 \\ &+ \left(\mathbf{M}_{2i-1}\mathbf{g}_k\mathbf{g}_k^T\mathbf{M}_{2i}^T\right)^2. \end{aligned} \tag{7}$$

Equation 7, in its present form, cannot be used for the estimation of more than one column triple $\mathbf{g}_k$, because $\mathbf{M}$ is independent of k. If one column triple is known then other triples can be estimated as shown in [3] and $\mathbf{G}$ can be computed simply as their concatenation. However, this requires that these column triples are aligned, i.e. the multiplication of $\hat{\mathbf{M}}$ with each column triple should give the same camera rotations. Equation 6 shows that orthonormality constraints do not ensure this, as an arbitrary 3×3 rotation applied to a column triple also satisfies the orthonormality constraints. Alignment of the column triples can be achieved using the method discussed by Xiao *et al*. [9]. It involves first finding the camera rotations from each column triple separately by multiplying it with $\hat{\mathbf{M}}$ and then aligning the rotations corresponding to each column triple with a 3×3 rotation, using the Procrustes method. These rotations, when applied to the corresponding column triples, align them properly.

Once $\mathbf{G}$ is known, it can be used to compute $\mathbf{M}$ and $\mathbf{B}$ using Equation 4. The shape coefficients $\mathbf{C}$ and the camera rotations $\mathbf{R}$ can be computed by the factorization of $\mathbf{M}$, by dividing by the norm of rows in each of the $3K \times 3$ column triples. However, this procedure will leave a sign ambiguity between camera rotations and the shape coefficients. This ambiguity can be resolved by restricting the camera movement between frames to at most $\pm 90°$. Hence, in summary, the problem of reconstruction of nonrigid 3D structure reduces to the problem of finding an appropriate rectifying transform $\mathbf{G}$.

A major question about orthonormality constraints is whether they are sufficient to result in an appropriate rectifying transform $\mathbf{G}$ or do invalid solutions also exist? In the next sections, we will analyze the ambiguity of these constraints and their impact on structure reconstruction.

4. Ambiguity in Orthonormality Constraints

In this section, we will build upon the ambiguity theorem proposed by Xiao *et al*. in [9] to develop a general solution for orthonormality constraints. We will show that by imposing an additional rank three constraint inherent to the problem, the form of ambiguity identified by Xiao *et al*. becomes considerably simplified and intuitively easier to grasp. To understand the ambiguity theorem of Xiao *et al*., the following two definitions are necessary.

Definition 1: A $3K \times 3K$ symmetric matrix $\mathbf{Z}$ is called a *block-scaled-identity* matrix, if each of its 3×3 sub-blocks is a scaled identity, i.e. $Z_{ij} = \lambda_{ij}I_{3\times 3}$.

Definition 2: A $3K \times 3K$ symmetric matrix $\mathbf{Y}$ is called a *block-skew-symmetric* matrix, if all of its 3×3 sub-blocks

are skew symmetric, i.e.

$$Y_{ij} = \begin{pmatrix} 0 & y_{ij1} & y_{ij2} \\ -y_{ij1} & 0 & y_{ij3} \\ -y_{ij2} & -y_{ij3} & 0 \end{pmatrix} = -Y_{ij}^T = Y_{ji}, \quad i \neq j,$$

$$Y_{ii} = 0_{3\times3}, \quad i = 1, \cdots, K.$$

Recall that $\mathbf{g}_k$ is the kth column triple of the metric rectification matrix $\mathbf{G}$ given in Equation 4. Let $\mathcal{I}_k$ be a $3K \times 3K$ matrix whose k^{th} 3×3 diagonal block is identity and all other elements are zero. Then $\mathbf{g}_k\mathbf{g}_k^T = \mathbf{G}\mathcal{I}_k\mathbf{G}^T$. The ambiguity theorem given by Xiao *et al.* in [9] states that for any column triple $\mathbf{g}_k$, if $\mathbf{g}_k$ satisfies the orthonormality constraints given in Equation 6, then there will exist a $\tilde{\mathbf{g}}_k$, which will also satisfy these constraints, provided $\tilde{\mathbf{g}}_k\tilde{\mathbf{g}}_k^T = \mathbf{G}\mathbf{H}_k\mathbf{G}^T$, where $\mathbf{H}_k$ is the sum of an arbitrary block skew matrix and a block scaled identity matrix, that is, $\mathbf{H}_k = \mathbf{Y}_k + \mathbf{Z}_k$.

The above theorem characterizes the ambiguity in $\tilde{\mathbf{g}}_k$, the general solution of the orthonormality constraints. We note that in this theorem, the inherent rank three constraint on the matrix product $\mathbf{G}\mathbf{H}_k\mathbf{G}^T$ has been overlooked. $\mathbf{G}\mathbf{H}_k\mathbf{G}^T$ necessarily has rank of 3 because it is equal to $\tilde{\mathbf{g}}_k\tilde{\mathbf{g}}_k^T$. Modifying the general solution of the orthonormality constraints proposed by Xiao *et al.*, by imposing this rank 3 constraint, results in the following theorem.

Theorem 1: *Let $\mathbf{g}_k$ denote column triples of a rectifying transform which results in correct structure recovery and let $\tilde{\mathbf{g}}_k$ denote the general solution of orthonormality constraints. Then $\tilde{\mathbf{g}}_k$ can be written as a linear combination of column triples $\{\mathbf{g}_i\}_{i=1,..,K}$ up to a 3×3 rotation V_k, i.e. $\tilde{\mathbf{g}}_k = (\sum_{i=1}^{K} a_i\mathbf{g}_i)V_k$, where a_i are arbitrary scalars.*

The proof of Theorem 1 builds upon the proof of the ambiguity theorem given in [9] by imposing the rank of $\mathbf{G}\mathbf{H}_k\mathbf{G}^T$, and is given in Appendix A. While the complete analytical proof is rather involved, the final result is satisfyingly intuitive. It says that for any valid rectifying transform $\mathbf{G}$, a linear combination of its column triples also forms a valid solution. This result could be understood just by analyzing the form of orthonormality constraints given in Equation 6. Since c_{ik} in Equation 5 are unknown, the two diagonal terms of the identity matrix are equated to yield the first constraint in Equation 6. If we consider a linear combination of $\mathbf{g}_k$, they must also satisfy the same constraints since, in this case, the right hand side of Equation 5 will turn out to be a linear combination of several scaled identity matrices,

$$\hat{\mathbf{M}}_{2i-1:2i} \left(\sum_{k=1}^{K} a_k^2 \mathbf{g}_k\mathbf{g}_k^T \right) \hat{\mathbf{M}}_{2i-1:2i}^T = \left(\sum_{k=1}^{K} a_k^2 c_{ik}^2 \right) \mathbf{I}_{2\times2}.$$
(8)

This yields exactly the same form of constraints as given in Equation 6, hence giving us an insight into the source of
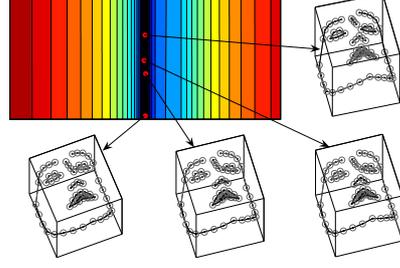

Figure 2. Visualizing a 2D slice of the $9K$ dimensional cost function given in Equation 7, for a face dataset having $K = 2$. This 2D slice contains a subspace of valid solutions of the orthonormality constraints (dark blue line). Four random solutions are shown (red dots) with corresponding structure reconstruction for one frame (gray circles). The structure is equivalent for each solution, and is aligned to the ground truth (black dots), indicating the multiplicity of solutions.

the ambiguity in the use of orthonormality constraints for nonrigid structure from motion.

In order to validate Theorem 1, we analyzed the surface of the cost function given in Equation 7, whose minima gives the solution of $\tilde{\mathbf{g}}_k$ under the orthonormality constraints. The search space for this minimization is $9K$ dimensional because $\tilde{\mathbf{g}}_k$ is of size $3K \times 3$. Note that according to Theorem 1, a linear combination of column triples is also a valid solution. Hence, for $K = 2$ case, the points lying on the line connecting any two column triples are also valid solutions.

For a demonstration of this space of valid solutions, we chose a face dataset with known 3D structure lying in a subspace spanned by two shape basis. We created images of this dataset with random camera rotations. We precomputed the column triples $\mathbf{g}_1$ and $\mathbf{g}_2$ using ideal rotations and shape coefficients, and found a plane (in $9K$ dimensions) passing through them. Figure 2 provides the visualization of a space of valid solutions of orthonormality constraints for this dataset. This figure shows that a line of valid solutions is clearly visible. We chose some random points on this line and for each point computed the full rectifying transform matrix $\mathbf{G}$ using the method by Brand [3]. Then using $\mathbf{G}$, we estimated the structure. Figure 2 shows the chosen points (red dots) and the corresponding structure reconstruction superimposed on the original structure. The equivalence between the estimated structure and the known structure shows that all points lying on the line are valid solutions of the orthonormality constraints and give perfect structure reconstruction.

In the next section, we will show that while an ambiguity exists in the solution of $\mathbf{g}_k$, as given by Theorem 1, the structure recovery using these column triples does not suffer from a similar ambiguity.

5. Ambiguity in Structure Reconstruction

According to Theorem 1, there exists an ambiguity of a linear transformation in the solution space of orthonormality constraints. In the following theorem, we will show that this ambiguity is due to the fact that the shape basis are ambiguous. However, this ambiguity has no impact on structure reconstruction, because the transformed basis are also an equally valid basis. Therefore, every solution of Equation 6 is a valid solution from the point of view of structure reconstruction.

Theorem 2: *Let $\mathbf{G}$ be a rectifying transform which results in correct structure recovery, and let $\tilde{\mathbf{G}}$ be an ambiguous solution of orthonormality constraints. Let the column triples of $\tilde{\mathbf{G}}$ be linearly independent and aligned to the column triples of $\mathbf{G}$ by the same rotation V_0. Then the structure recovered by $\tilde{\mathbf{G}}$ will be the same as the structure recovered from $\mathbf{G}$ up to a rotation V_0^T, provided that the motion between any two consecutive frames is not more than $\pm 90°$.*

The proof of this theorem is given in Appendix B.

Theorem 2 shows that the structure recovered by any solution of orthonormality constraints is equal to the ideal structure by a 3×3 rotation. The linear ambiguity in the solution of $\mathbf{G}$ only results in a linear ambiguity of dimension $K \times K$ in finding the shape basis. However, the shape coefficients get affected by the inverse of this ambiguity which results in unique structure reconstruction. Hence, while there are multiple solutions of $\tilde{\mathbf{G}}$ under orthonormality constraints, all of them yield the same unique 3D structure, differing from each other just by an overall rotation.

To provide empirical evidence of this theorem, we are able to perfectly reconstruct the cube and moving points dataset based on just orthonormality constraints (discussed in Section 7). The results of this experiment are given in Figure 1 (b,c). Figure 1 (d-f) show the reconstruction using three valid solutions of orthonormality constraints, showing perfect recovery of structure but at a different overall rotation.

6. Ambiguity for Trajectory Basis

The trajectory basis approach proposed by Akhter *et al.* [1] is a dual representation of the nonrigid structure. They proposed to use predefined basis to serve as trajectory basis. This approach also exploits orthonormality constraints to estimate the rectifying transform. Since, we are characterizing the nature of ambiguity due to orthonormality constraints, it is instructive to see how this ambiguity applies to the trajectory basis approach.

The formulation under trajectory basis approach is dual to shape basis approach. It results in the same form of matrix $\mathbf{M}$ and same type of orthonormality constraints for rectifying transform, which came across in shape basis approach. given as,

$$\mathbf{M}_{2F \times 3K} = \begin{pmatrix} \theta_{11}R_1 & \cdots & \theta_{1K}R_1 \\ \vdots & \ldots & \vdots \\ \theta_{F1}R_F & \cdots & \theta_{FK}R_F \end{pmatrix}, \quad (9)$$

and

$$\hat{\mathbf{M}}_{2i-1:2i}\mathbf{g}_k\mathbf{g}_k^T\hat{\mathbf{M}}_{2i-1:2i}^T = \theta_{ik}^2\mathbf{I}_{2\times 2}, \quad (10)$$

where θ_{ik} are the coefficients of a predefined trajectory basis such as the DCT. Although Equation 10 looks similar to Equation 6, however, there is an important difference. In the shape basis case, each image gives two constraints for a column triple, while in the trajectory basis case, each image gives three constraints, because now the basis coefficients, θ_{ij}, are known. Secondly, in the shape basis approach, all column triples need to be estimated for structure reconstruction, while for the trajectory basis, only a single column triple is enough for metric rectification as shown in [1].

In the next theorem, we will show that unlike the shape basis case, there is no ambiguity in the solution of Equation 10. This is because, by specifying the coefficients, the ambiguity in choosing the basis is no longer there. Hence, the structure reconstruction is still unique up to a Euclidean transformation.

Theorem 3: *Let $\mathbf{G}$ be a rectifying transform which results in the correct structure recovery up to a rotation under the trajectory basis approach and let $\mathbf{g}_k$ be the kth column triple of $\mathbf{G}$. Let $\tilde{\mathbf{g}}_k$ denote the general solution of orthonormality constraints given in Equation 10 for the kth column triple, then $\tilde{\mathbf{g}}_k = \mathbf{g}_k V_k$, where V_k is a 3×3 rotation.*

The proof of this theorem is given in Appendix C.

The above theorem shows that orthonormality constraints for trajectory basis do not give an ambiguous solution other than a 3×3 rotation. Therefore structure recovery will also be up to a 3×3 rotation. This is also an intuitively satisfying result, considering that the linear combination of $\mathbf{g}_i$s arose in the shape basis because c_{ik} were unknown. In the case of trajectory basis, since θ_{ik}s are known, the linear combination ambiguity should not exist.

7. Quantitative Evaluation

In order to compute the rectifying transform $\tilde{\mathbf{G}}$, we follow a two step procedure. In the first step, we estimate just one column triple $\tilde{\mathbf{g}}_1$ using nonlinear least-squares on the cost function given in Equation 7. We add one additional constraint to this equation to avoid the trivial solution of $\tilde{\mathbf{g}}_1 = \mathbf{0}$. This constraint is written simply as an additional term of $\left(\|\tilde{\mathbf{g}}_1\|_F^2 - 1\right)^2$ in Equation 7, where $\|\cdot\|_F$ denotes the Frobenius norm.

The second step involves using the estimated $\tilde{\mathbf{g}}_1$ to compute the rest of the column triples in $\tilde{\mathbf{G}}$, following the method of Brand [3]. To improve stability, we use the result
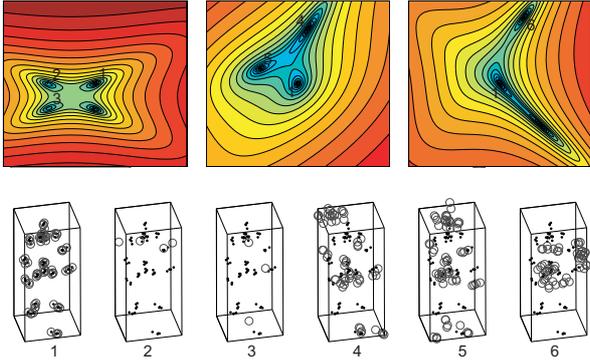
Figure 3. 2D slices of the $9K$ dimensional cost function for the walk dataset with $K = 2$. Top: Three random slices through the global minimum (marked as 1), showing the presence of several local minima (marked 2-6). Bottom Row: The reconstruction of nonrigid structure corresponding to different minima. The result for the global minimum matches the ground truth, while for local minima deviates significantly from ground truth.

given by Brand's method as an initial guess in a nonlinear least squares procedure, with the cost being the addition of the cost functions of all column triples. Again, to avoid the trivial solution and to impose linear independence between column triples of $\tilde{\mathbf{G}}$, we add the term $(|\tilde{\mathbf{G}}|^2 - 1)^2$ to the cost function, where $|\cdot|$ denotes the determinant. Finally, as suggested by Xiao *et al.* [9], the column triples given by this optimization procedure are aligned using the Procrustes method, as described in Section 3. The $\tilde{\mathbf{G}}$ thus computed is used to recover the structure and the camera matrices as discussed in Section 3.

We tested the above procedure on the cube and moving points dataset and the result of the reconstruction is given in Figure 1 (b). For this simple nonrigid dataset, using $K = 2$, the results are almost perfect, with very low reconstruction error. This experiment validates our assertion that in the ideal case, orthonormality constraints are sufficient for the recovery of structure.

To test the multiplicity of solutions up to a rotation, we took the solution of $\tilde{\mathbf{g}}$ which results in the global minimum of the cost function and applied a random rotation. We added noise to this new $\tilde{\mathbf{g}}$ to generate an initialization and using it as the initial guess, ran our optimization routine. The optimization converged to a different, yet correct, solution where the overall structure was simply rotated. Three such experiments with different rotations are shown in Figure 1 (d-f). This confirms our assertion that multiple solutions differing in overall rotation exist in the solution space.

Though the cube dataset is nonrigid, it contains a significant number of points which behave in a rigid fashion. The optimization problem gets much more complicated when no reasonable sized subset of the dataset behaves rigidly.

To show this, we took motion captured data of the walk action, projected it on two PCA basis and synthetically generated images from random locations. This forms the $K = 2$ case, and hence the cost function is being optimized over $9K = 18$ dimensional space. The optimization procedure did not converge on this dataset. To analyze this case further, we used our known camera rotations to compute the ideal solution of $\mathbf{g}_1$, and then evaluated the cost function in the vicinity of this ideal solution. Figure 3 shows random as well as axis-aligned 2D slices of the 18 dimensional space around the cost function. The presence of several local minima is clearly visible in these slices. Note that since our cost function is quartic, the maximum number of minima in any 18 dimensional space is to the order of 2^{18}. The maximum number of minima that can be seen in a two dimensional slice is four, a case which occurs frequently in Figure 3, indicating the presence of a high number of local minima in the 18-dimensional space. Our understanding, derived from experiments, is that the higher the level of non-rigidity in a dataset, the more complicated the cost surface is, with higher number of local minimas. This experiment leads us to conclude that the real difficulty in achieving good 3D reconstructions for nonrigid structures, based on orthonormality constraints alone, is not the ambiguity of the constraints but the complexity of the underlying non-linear optimization.

Acknowledgements

References

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid Structure from Motion in Trajectory Space. *NIPS*, 2008.

[2] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-Fine Low-Rank Structure-from-Motion. *CVPR*, 2008.

[3] M. Brand. A Direct Method for 3D Factorization of Nonrigid Motion Observed in 2D. *CVPR*, 2005.

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering Non-Rigid 3D Shape from Image Streams. *CVPR*, 2000.

[5] A. D. Bue, X.Liao, and L. Agapito. Non-Rigid Metric Shape and Motion Recovery from Uncalibrated Images using Priors. *CVPR*, 2006.

[6] L. Mirsky. *An Introduction to Linear Algebra*. Courier Dover Publications, 1990.

[7] C. Tomasi and T. Kanade. Shape and Motion from Image Streams Under Orthography: A Factorization Method. *IJCV*, 9(2), 1992.

[8] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid Structure from Motion: Estimating Shape and Motion with Hierarchical Priors. *IEEE Trans. on PAMI*, 30(5), 2008.

[9] J. Xiao, J. Chai, and T. Kanade. A Closed Form Solution to Non-Rigid Shape and Motion Recovery. *IJCV*, 67(2), 2006.

Appendix A

Proof of Theorem 1: According to the Ambiguity Theorem of Xiao *et al.*, $\tilde{\mathbf{g}}_k \tilde{\mathbf{g}}_k^T = \mathbf{G}\mathbf{H}_k\mathbf{G}^T$. Since $\mathbf{G}$ is a full rank matrix, the rank of any matrix product $\mathbf{A}\mathbf{G}$ or $\mathbf{G}\mathbf{A}$ is equal to rank of $\mathbf{A}$ [6]. Hence, $\mathbf{H}_k$ must be of rank 3 for $\mathbf{G}\mathbf{H}_k\mathbf{G}^T$ to be of rank 3. In order to see the effect of this rank constraint on $\mathbf{H}_k$, we define $\mathbf{h}_{km}$ as the mth $3K \times 3$ column triple of $\mathbf{H}_k$.

Let us consider two column triples $\mathbf{h}_{ki}$ and $\mathbf{h}_{kj}$. We also consider four 3×3 blocks in $\mathbf{h}_{ki}$ and $\mathbf{h}_{kj}$, two blocks on the block diagonal of $\mathbf{H}_k$ and two blocks off the block diagonal, as follows:

1. i^{th} block in $\mathbf{h}_{ki}$: $H_{kii} = \lambda_{ii} I_{3 \times 3}$

2. j^{th} block in $\mathbf{h}_{ki}$: $H_{kji} = \lambda_{ij} I_{3 \times 3} - Y_{ij}$

3. i^{th} block in $\mathbf{h}_{kj}$: $H_{kij} = \lambda_{ij} I_{3 \times 3} + Y_{ij}$

4. j^{th} block in $\mathbf{h}_{kj}$: $H_{kjj} = \lambda_{jj} I_{3 \times 3}$

We define

$$\mathcal{P} = \left(\begin{array}{cc} H_{kii} & H_{kji} \\ H_{kij} & H_{kjj} \end{array} \right) = \left(\begin{array}{cc} \lambda_{ii} I & \lambda_{ij} I + Y_{ij} \\ \lambda_{ij} I - Y_{ij} & \lambda_{jj} I \end{array} \right).$$

$\mathcal{P}$ is a symmetric matrix and its rank is at most 3, because $\mathbf{H}_k$ is of rank 3. Hence $\mathcal{P}$ is a semi-positive definite matrix and can be factorized as $\mathbf{p}\mathbf{p}^T$, where $\mathbf{p}$ is a 6×3 matrix. We denote $\mathbf{p} = [P_1^T, P_2^T]^T$, where P_1 and P_2 are 3×3 matrices. Using these terms $\mathcal{P}$ can be written as,

$$\left(\begin{array}{cc} P_1 P_1^T & P_1 P_2^T \\ P_2 P_1^T & P_2 P_2^T \end{array} \right) = \left(\begin{array}{cc} \lambda_{ii} I & \lambda_{ij} I + Y_{ij} \\ \lambda_{ij} I - Y_{ij} & \lambda_{jj} I \end{array} \right). \tag{11}$$

We will now show that $Y_{ij} = 0_{3 \times 3}$ by considering the following two cases.

Case 1: $\lambda_{ii} = 0$
From Equation 11, if $\lambda_{ii} = 0$ then $P_1 = 0$ and $P_1 P_2^T = \lambda_{ij} I + Y_{ij} = 0$. This implies that $\lambda_{ij} = 0$ and $Y_{ij} = 0$, because I is diagonal and Y_{ij} is skew symmetric.

Case 2: $\lambda_{ii} \neq 0$
We consider the multiplication of diagonal blocks in $\mathcal{P}$. Using Equation 11 we can write,

$$\begin{aligned} P_1 P_2^T P_2 P_1^T &= \lambda_{ij}^2 I - Y_{ij}^2 \\ \lambda_{ii}\lambda_{jj} I &= \lambda_{ij}^2 I - Y_{ij}^2 \\ (\lambda_{ii}\lambda_{jj} - \lambda_{ij}^2) I &= -Y_{ij}^2 \end{aligned} \tag{12}$$

Since the RHS of Equation 12 is a scaled identity matrix, its rank can only be 0 or 3. However, the rank of LHS can only be 0 or 2, since Y_{ij} is a skew symmetric matrix. Hence, it

follows, that for the equality to hold, the rank of both sides of Equation 12 should be zero. Therefore we can write, $Y_{ij} = 0_{3 \times 3}$ and $\lambda_{ij} = \sqrt{\lambda_{ii}\lambda_{jj}}$.

Thus, the both cases given above lead to the conclusion that Y_{ij} is always $0_{3 \times 3}$. Extending this idea to all other column triples in matrix $\mathbf{H}_k$, we conclude that $\mathbf{Y}_k = \mathbf{0}$. Furthermore, from case 2, $\lambda_{ij} = \sqrt{\lambda_{ii}\lambda_{jj}}$. This also holds true for case 1, since $\lambda_{ij} = \lambda_{ii} = 0$. Hence, because of the rank 3 constraint on $\mathbf{H}_k$, it will have scaled identity blocks of the following form:

$$\mathbf{H}_k = \left(\begin{array}{cccc} \lambda_{11} I & \sqrt{\lambda_{11}\lambda_{22}} I & \cdots & \sqrt{\lambda_{11}\lambda_{KK}} I \\ \sqrt{\lambda_{11}\lambda_{22}} I & \lambda_{22} I & \cdots & \sqrt{\lambda_{22}\lambda_{KK}} I \\ \vdots & \vdots & \cdots & \vdots \\ \sqrt{\lambda_{11}\lambda_{KK}} I & \sqrt{\lambda_{22}\lambda_{KK}} I & \cdots & \lambda_{KK} I \end{array} \right). \tag{13}$$

Equation 13 shows that all of λ_{ii}s are of the same sign, because $\mathbf{H}_k$ does not, by definition, have complex numbers. For simplicity of notation, let us define $a_i = \sqrt{\lambda_{ii}}$ if $\lambda_{ii} > 0$ and $\sqrt{-\lambda_{ii}}$ otherwise. Since $\mathbf{H}_k$ is of rank 3, it can be factorized as, $\mathbf{H}_k = \mathbf{h}_k \mathbf{h}_k^T$, where $\mathbf{h}_k = \left(\begin{array}{ccc} a_1 I \cdots a_K I \end{array} \right)^T$. This factorization, however, is unique only up to a 3×3 rotation, say V_K. Therefore the general solution $\tilde{\mathbf{g}}_k$ can be written as,

$$\tilde{\mathbf{g}}_k = \mathbf{G}\mathbf{h}_k V_k = (\sum_{i=1}^{K} a_i \mathbf{g}_i) V_k. \tag{14}$$

$\square$

Lemma 1: *Let H denote a 3×3 matrix, which is the summation of a scaled identity matrix λI and a skew-symmetric matrix Y. Then $\text{rank}(H) < 3$ implies $\lambda = 0$.*

Proof: Consider a matrix $H = \lambda I + Y$,

$$H = \left(\begin{array}{ccc} \lambda & y_1 & y_2 \\ -y_1 & \lambda & y_3 \\ -y_2 & -y_3 & \lambda \end{array} \right). \tag{15}$$

For the sake of contradiction, let us assume that $\text{rank}(H) < 3$ but $\lambda \neq 0$. The rank deficiency implies the linear dependence between columns (or rows) in H, which can be written as,

$$\left(\begin{array}{c} y_2 \\ y_3 \\ \lambda \end{array} \right) = \left(\begin{array}{cc} \lambda & y_1 \\ -y_1 & \lambda \\ -y_2 & -y_3 \end{array} \right) \left(\begin{array}{c} a_1 \\ a_2 \end{array} \right),$$

or

$$\left(\begin{array}{c} y_2 \\ y_3 \end{array} \right) = \left(\begin{array}{c} a_1\lambda + a_2 y_1 \\ -a_1 y_1 + a_2\lambda \end{array} \right), \tag{16}$$

$$\lambda = -a_1 y_2 - a_2 y_3. \tag{17}$$

Equations 16 and 17 imply that

$$\begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} y_2 \\ y_3 \end{pmatrix} = (a_1^2 + a_2^2)\lambda = -\lambda$$
$$\Rightarrow a_1^2 + a_2^2 = -1 \quad \because \lambda \neq 0. \quad (18)$$

However, since a_1 and a_2 are real numbers, this results in a contradiction in the form of Equation 18. Therefore, we must have $\lambda = 0$. □

Appendix B

Proof of Theorem 2: The ambiguity in the rectifying transform, given in Equation 14, can be written as $\tilde{\mathbf{G}} = \mathbf{GAV}$, where

$$\mathbf{A} = \begin{pmatrix} a_{11}I & \cdots & a_{1K}I \\ \vdots & \cdots & \vdots \\ a_{K1}I & \cdots & a_{KK}I \end{pmatrix}, \mathbf{V} = \begin{pmatrix} V_0 & & \\ & \ddots & \\ & & V_0 \end{pmatrix}.$$

$\tilde{\mathbf{G}}$ can be used for the rectification of matrix $\hat{\mathbf{M}}$ given in Equation 4, as follows:

$$\tilde{\mathbf{M}} = \hat{\mathbf{M}}\tilde{\mathbf{G}} = \hat{\mathbf{M}}\mathbf{GAV}. \quad (19)$$

Since $\mathbf{V}$ has block diagonal structure and $\mathbf{A}$ consists of scaled identities, $\mathbf{VA} = \mathbf{AV}$. Substituting $\mathbf{M} = \hat{\mathbf{M}}\mathbf{G}$ in above equation gives $\tilde{\mathbf{M}} = \mathbf{MVA}$. Further substituting the value of $\mathbf{M}$ from Equation 3, we can write $\tilde{\mathbf{M}}$ as $\tilde{\mathbf{M}} = \mathbf{RCVA}$. Let $\mathbf{V}'$ be a $3F \times 3F$ matrix which has structure similar to $\mathbf{V}$, i.e. it has V_0 on its diagonal blocks and zero elsewhere. Then we can rewrite $\tilde{\mathbf{M}}$ as $\tilde{\mathbf{M}} = \mathbf{RV}'\mathbf{CA}$. This shows that the recovered camera rotations will be same as the ones in $\mathbf{R}$ up to a 3×3 rotation V_0. The recovered shape coefficients will be a linear transform of the original shape coefficients $\mathbf{C}$. Therefore, the following equation describes the ambiguity in recovered rotations and shape coefficients through orthonormality constraints,

$$\mathbf{R}_{V_0} = \mathbf{RV}' \quad \text{and} \quad \tilde{\mathbf{C}} = \mathbf{CA}. \quad (20)$$

In general, there will a sign ambiguity as well between the recovered rotations and recovered coefficients by factorizing matrix $\tilde{\mathbf{M}}$ into $\mathbf{R}_{V_0}$ and $\tilde{\mathbf{C}}$. However, we can fix this sign, if camera movement is restricted and it does not move more than $90°$ in two consecutive frames.

The ambiguous basis can be computed as,

$$\tilde{\mathbf{B}} = \tilde{\mathbf{G}}^{-1}\hat{\mathbf{B}} = \mathbf{A}^{-1}\mathbf{V}^T\mathbf{G}^{-1}\hat{\mathbf{B}}$$
$$= \mathbf{A}^{-1}\mathbf{V}^T\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}_{V_0}, \quad (21)$$

where the subscript V_0 in $\mathbf{B}_{V_0}$ shows that the original basis are transformed by 3×3 rotation V_0^T.

Combining Equations 20 and 21, the expression for recovered structure $\tilde{\mathbf{S}}$ can be written as,

$$\tilde{\mathbf{S}} = \tilde{\mathbf{C}}\tilde{\mathbf{B}} = (\mathbf{CA})(\mathbf{A}^{-1}\mathbf{B}_{V_0})$$
$$= \mathbf{CV}^T\mathbf{B} = \mathbf{V}'^T\mathbf{CB}$$
$$= \mathbf{V}'^T\mathbf{S} = \mathbf{S}_{V_0}. \quad (22)$$

□

Appendix C

Proof of Theorem 3: According to Theorem 1, the general solution of orthonormality constraints is given by

$$\tilde{\mathbf{g}}_k = \left(\sum_{i=1}^{K} a_i\mathbf{g}_i\right)V_k = \mathbf{G}\begin{pmatrix} a_iI \\ \vdots \\ a_KI \end{pmatrix}V_k. \quad (23)$$

Substituting the above expression in Equation 10 gives

$$\hat{\mathbf{M}}_{2i-1:2i}\hat{\mathbf{g}}_k\tilde{\mathbf{g}}_k^T\tilde{\mathbf{M}}_{2i-1:2i}^T = \theta_{ik}^2 I_{2\times2}$$

$$= \hat{\mathbf{M}}_{2i-1:2i}\mathbf{G}\begin{pmatrix} a_iI \\ \vdots \\ a_KI \end{pmatrix}\begin{pmatrix} a_iI & \cdots & a_KI \end{pmatrix}\mathbf{G}^T\hat{\mathbf{M}}_{2i-1:2i}^T$$

$$= \mathbf{M}_{2i-1:2i}\begin{pmatrix} a_iI \\ \vdots \\ a_KI \end{pmatrix}\begin{pmatrix} a_iI & \cdots & a_KI \end{pmatrix}\mathbf{M}_{2i-1:2i}^T$$

$$= \left(\sum_{j=1}^{K}\theta_{ij}a_j\right)R_iR_i^T\left(\sum_{j=1}^{K}\theta_{ij}a_j\right) = \left(\sum_{j=1}^{K}\theta_{ij}a_j\right)^2 I_{2\times2}.$$

This implies that

$$\theta_{ik} = \sum_{j=1}^{K}\theta_{ij}a_j. \quad (24)$$

Since i ranges from $1, \ldots, F$, this system of equations can be written as,

$$\begin{pmatrix} \theta_{1k} \\ \vdots \\ \theta_{Fk} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1K} \\ \vdots & \ddots & \vdots \\ \theta_{F1} & \cdots & \theta_{FK} \end{pmatrix}\begin{pmatrix} a_1 \\ \vdots \\ a_K \end{pmatrix} \quad (25)$$

This equation shows that in order for the above equality to hold, one obvious solution is to take $a_j = 1$ for $j = k$ and 0 otherwise. In fact this is the only possible solution because, since $K < F$ and the matrix of K DCT basis has rank K, we have an overconstrained system which has to have a unique solution. Substituting above values of a_j in Equation 23, we get

$$\tilde{\mathbf{g}}_k = \mathbf{g}_k V_k.$$

□