Perturbation Estimation of the Subspaces for Structure from Motion with Noisy and Missing Data

Hongjun Jia, Jeff Fortuna and Aleix M. Martinez
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH 43210

Abstract

It is common when analyzing experimental data to encounter matrices that have been contaminated by noise and have missing elements. Missing data can be recovered with imputation methods if the measurement data matrix is of low rank and the data is noise-free. However, iterative imputation can produce poor results for cases of large noise or a large proportion of missing data. Non-imputing methods rely on the use of existent data and require a selection of complete submatrices. Jacobs introduced a non-imputing method which can produce good results, but the randomness in selecting the submatrices cannot guarantee a consistently accurate recovery of missing data. Chen and Suter's method chooses the most reliable submatrix based on the number of missing data elements only, which fails to consider the effect of noise on the selected data. Herein, a new criterion based on an estimate of the sensitivity of submatrices to perturbation is introduced which takes into consideration that in some cases a column with more missing data could provide more useful information than one with less missing data. Experimental results for the problem of Structure from Motion with noisy point correspondences and missing data show that our criterion can sort submatrices properly in terms of their possible perturbation and recover the 3D structure of the scene more accurately than other non-imputing methods.

1. Introduction

Several problems in computer vision can be reduced to finding a low-dimensional linear subspace that provides the scene structure, such as the factorization method of Structure from Motion (SfM) [1], optical flow estimation in multi-frame video [2] and Shape from Shading [3]. These approaches require fitting a large measurement matrix to its closest approximation subject to rank constraints which arise from the geometry or radiometry of the problem. Herein, we will examine matrix factorization as it applies to SfM. Ullman and Basri [4] also propose using the SVD method to fit a low rank matrix to a larger matrix when extra measurement data can be obtained.

When all measured 2D points can be tracked throughout an image sequence, a measurement matrix can be formed from the pixel coordinates of corresponding points in each image. The method of Tomasi and Kanade [1] uses the Singular Value Decomposition (SVD) to provide a rank 3 factorization of the measurement matrix and recover the 3D structure.

In most real world SfM problems, we cannot obtain the full measurement data matrix. This is due to ambiguity inherent in obtaining point correspondences for tracking. This ambiguity can result from changes in the appearance of object features under motion or where previously visible features become invisible in subsequent images. Additionally, the incomplete data matrix contains noise due to inaccurately located point correspondences. Hence, any method which tries to solve the SfM problems must deal with both the missing data and noisy data [1, 5, 6, 7, 8]. An iterative method is applied by Shum, Ikeuchi and Reddy [5] on the missing data problem, which can minimize the sum of square differences between the elements that are not missing in the fitting low-rank matrix and those in the data matrix. This method can converge to a locally optimal solution, but it does not guarantee a globally optimal one. Jacobs [6] provides another effective method to solve the missing data problem which doesn't have to start with a complete submatrix. In this algorithm, several triple-columns are selected to form a large complementary matrix, and SVD is then utilized to find the closest 3D linear subspace. Unfortunately, due to the randomness of the selected triple-columns, the recovered results vary extensively among different trials. In Chen and Suter's paper [7], a hypothesis about the denoising capacity of the incomplete matrix is proposed and the Minimal Unreliability Ratio (MUR) is treated as the criterion when choosing the subset of data matrix. This selection is based on the number of missing data elements in each column, and the columns with less missing data are chosen. *The problem with this technique is that in some cases, a column with more missing data could provide more useful shape recovery information than one with less missing data.*

Our key idea is to derive a new criterion for the selection of triple-columns. Specifically, we use a measure of the sensitivity of each triple-column space to perturbations resulting from measurement noise in the data matrix. The measure comes from matrix perturbation theory which can define bounds on the sensitivity of singular values and singular vectors to additive noise. We will show that our criterion for column selection provides a more accurate recovery of 3D shape than previous selection methods.

2. Structure from motion with missing data

In this section, we first review the factorization method of SfM when there is no missing data. We then review the methods of Jacobs and Chen and Suter for recovering the full data matrix in the missing data case. For simplicity, we assume that the data matrix is composed of mean-centered points. This removes any translation component present in the motion.

2.1. Factorization method in SfM

A sequence of F images with P annotated points can be represented as a data matrix $\mathbf{Q} \in \mathbb{R}^{2F \times P}$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_F \end{bmatrix} \quad \text{, where} \quad \mathbf{X}_f = \begin{bmatrix} \mathbf{x}_f \\ \mathbf{y}_f \end{bmatrix} \tag{1}$$

with each $\mathbf{X}_f \in \mathbb{R}^{2\times P}, f = \{1, 2, \cdots, F\}$, composed of the mean-centered x_p and y_p coordinates of the points ($p = \{1, 2, \cdots, P\}$) in the image sequence composed of row vectors $\mathbf{x}_f, \mathbf{y}_f \in \mathbb{R}^{1\times P}$:

$$\begin{array}{rcl} \mathbf{x}_f & = & \begin{bmatrix} x_1 & x_2 & \cdots & x_P \end{bmatrix} \\ \mathbf{y}_f & = & \begin{bmatrix} y_1 & y_2 & \cdots & y_P \end{bmatrix} \end{array} . \tag{2}$$

In the structure from motion problem, this set of points is assumed to have resulted from the projection of a set of 3D point coordinates described by $\mathbf{P} \in \mathbb{R}^{3\times P}$:

$$\mathbf{P} = \begin{bmatrix} x_1 & x_2 \cdots & x_P \\ y_1 & y_2 \cdots & y_P \\ z_1 & z_2 \cdots & z_P \end{bmatrix}, \tag{3}$$

which undergoes an affine transformation between images

$$\mathbf{Q} = \mathbf{SP}, \tag{4}$$

where $\mathbf{S} \in \mathbb{R}^{2F\times 3}$ is composed of 2×3 orthogonal affine transformation matrices $\mathbf{R}_f$ for each image:

$$\mathbf{S} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_F \end{bmatrix} . \tag{5}$$

In the noise-free case, the motion and shape matrices can be recovered exactly (up to an arbitrary affine transformation) from the SVD of $\mathbf{Q}$ for all images by

$$\mathbf{Q} = \mathbf{UDV}^T. \tag{6}$$

As such, $\mathbf{S}$ and $\mathbf{P}$ can be found by the rank 3 approximation of $\mathbf{Q}$. This defines the factorization method of Tomasi and Kanade [1].

When the points in the 2D images are inaccurately measured, the SfM model becomes

$$\widehat{\mathbf{Q}} = \mathbf{Q} + \mathbf{E} = \mathbf{SP} + \mathbf{E}, \tag{7}$$

where $\mathbf{E} \in \mathbb{R}^{2F\times P}$ is a matrix of the noise in measurement at each point for each image. If the SVD is used to factorize $\widehat{\mathbf{Q}}$, a least squares solution is obtained.

2.2. Missing data and data selection strategies

If there are some missing elements in the data matrix, it is necessary to recover the missing data before applying factorization methods to recover the 3D structure. In Tomasi and Kanade [1], a rectangular subset of the matrix with no missing data is obtained prior to factorization. The missing data is recovered by adding new columns (or rows) and finding the linear combination of a basis for the columns (or rows) of the submatrix that best fits the non-missing elements of the new columns (or rows).

In Jacobs' approach [6], constraints derived from some submatrices of the original matrix are combined and the linear fitting method is used to recover the missing data. Each column of $\mathbf{Q}$ can be regarded as the coordinate of a point in a 2F-dimensional space. SVD finds the 3D linear subspace $\mathcal{L}$ which is closest to the P $2F$-dimensional points in $\widehat{\mathbf{Q}}$. When there is neither noise nor missing data, $\mathcal{L}$ is the space spanned by any three linearly independent columns of $\mathbf{Q}$. If there are some missing data in $\mathbf{Q}$, each column can span an affine subspace containing all the possible points, and $\mathcal{L}$ should lie in the space spanned by three such affine subspaces. Let $\mathbf{M}_i$ be the ith affine subspace and $\mathbf{F}_k$ be the space spanned by the kth triple affine subspaces group which is randomly selected, $\mathcal{L} \subseteq \mathbf{F}_k$. Thus, $\mathcal{L}$ should be a subset of the intersection of all possible $\mathbf{F}_k$, i.e.

$$\mathcal{L} \subseteq \mathbf{F} = \cap \mathbf{F}_k, k = 1, 2, \cdots, l. \tag{8}$$

If noise is introduced in $\mathbf{Q}$, we have the measurement matrix $\widehat{\mathbf{Q}}$. $\mathbf{F}$ will become the null-space because $\mathcal{L}$ cannot accurately lie in any $\mathbf{F}_k$. Jacobs uses a null-space based method to solve this problem. $\mathbf{N}_k$ gives the matrix representation of the orthogonal complementary space of $\mathbf{F}_k$, and then $\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 & \mathbf{N}_2 & \cdots & \mathbf{N}_l \end{bmatrix}$ is the matrix representation of $\mathbf{F}^\perp$ which is the final matrix in Jacobs' algorithm to be decomposed by SVD. In the presence of error, the matrix $\mathbf{N}$

will typically have full rank. The three singular vectors corresponding to the three smallest singular values are selected to form a 3D linear space $\mathcal{L}$ to be orthogonal to the matrix closest to $\mathbf{N}$ in the Frobenius norm sense. The affine shape of the original structure is recovered from $\mathcal{L}$. The Euclidean shape that best approximates the original structure can be recovered if the Euclidean constraints are included [1].

In Jacobs' algorithm, the triple-columns are randomly selected taking no account of the number of missing data elements in each column. Jacobs uses two methods to avoid the complete blind search: one is to set a maximum value on the number of columns in $\mathbf{N}$, the other is to bias the selection of column triples to those having full entries in the rows which we have no information about in $\mathbf{N}$. Unfortunately, in practice we see that the recovered results vary extensively when measured by the Mean Square Error (MSE). This indicates that the performance of Jacobs' algorithm really depends on the data (or the triple-columns) used. If a "good" set of triple-columns happen to be selected, the recovered shape of the rigid object can be very close to the ground-truth, but the worst case produces very poor results.

In Chen and Suter's method [7], they define the reliability of a submatrix, which focuses on the number of missing data in each column. The Minimal Unreliability Ratio (MUR) is proposed as the criterion to select the most reliable submatrix whose missing entries are recovered first by iterative imputation. Generally, the column with fewer missing data elements will be favored for selection. If the existent data in the column with more missing entries are measured with less error, this column may provide more useful information for shape recovery than a more complete column of very noisy measurements. In general, the problem of finding a good approximation to a matrix with missing elements appears to be much harder than that of approximating a noisy but complete matrix. A selection step is necessary in non-imputing methods and we will next describe our selection criterion.

3. Proposed Selection Criterion

In Jacobs' algorithm, three columns of the measurement data matrix $\widehat{\mathbf{Q}}$ are selected randomly to form a submatrix. Now, consider the submatrix $\widehat{\mathbf{A}}_i$ which consists of the i^{th} triple-columns of $\widehat{\mathbf{Q}}$. Then, let $\widehat{\mathbf{A}}'_i$ be the corresponding reduced form of $\widehat{\mathbf{A}}_i$ by removing all the rows having at least one missing entry. The spanning space of $\widehat{\mathbf{A}}'_i$ is $\widehat{\mathbf{S}}'_i$, and $\widehat{\mathbf{N}}'_i$ is the complementary subspace of $\widehat{\mathbf{S}}'_i$. Then we expand $\widehat{\mathbf{N}}'_i$ by adding zero in all the rows which are removed from $\widehat{\mathbf{A}}_i$ to $\widehat{\mathbf{A}}'_i$ to get the complementary space $\widehat{\mathbf{N}}_i$. All such $\widehat{\mathbf{N}}_i$ will be packed together to form a big matrix $\widehat{\mathbf{N}} = [\widehat{\mathbf{N}}_1\widehat{\mathbf{N}}_2\cdots\widehat{\mathbf{N}}_l]$, and the low-dimensional linear space which we are seeking should be orthogonal to the matrix closest to $\widehat{\mathbf{N}}$ according to the Frobenius norm (such that this close matrix has a rank 3 complementary space). If all the above process is considered under the noise-free condition, we will get the matrices $\mathbf{A}_i$, $\mathbf{A}'_i$, $\mathbf{S}'_i$, $\mathbf{N}'_i$, $\mathbf{N}_i$ and $\mathbf{N}$ corresponding to $\widehat{\mathbf{A}}_i$, $\widehat{\mathbf{A}}'_i$, $\widehat{\mathbf{S}}'_i$, $\widehat{\mathbf{N}}'_i$, $\widehat{\mathbf{N}}_i$ and $\widehat{\mathbf{N}}$ respectively. Since $\widehat{\mathbf{N}}$ is the final matrix to be decomposed by SVD, the perturbation between $\widehat{\mathbf{N}}$ and $\mathbf{N}$ should be as small as possible to get the most precisely recovered result. In this sense, we also require the perturbation between each $\widehat{\mathbf{N}}_i$ and $\mathbf{N}_i$ to be as small as possible.

The perturbation $pb(\mathbf{X}, \mathbf{X}')$ between two matrices $\mathbf{X}$ and $\mathbf{X}'$ of the same dimension can be measured by the distance, $dist(\mathbf{X}, \mathbf{X}')$, between two spaces spanned by the columns of $\mathbf{X}$ and $\mathbf{X}'$ respectively. From [9] we know that the largest principal angle between two corresponding spanning spaces $\theta(span(\mathbf{X}), span(\mathbf{X}'))$ and the distance $dist(\mathbf{X}, \mathbf{X}')$ are related to each other by

$$dist(\mathbf{X}, \mathbf{X}') = \sin\theta(span(\mathbf{X}), span(\mathbf{X}')). \qquad (9)$$

This allows us to define the perturbation $pb(\mathbf{X}, \mathbf{X}')$ as the largest principal angle $\theta(span(\mathbf{X}), span(\mathbf{X}'))$. In the above process, it is easy to show that

$$\theta(\mathbf{N}_i, \widehat{\mathbf{N}}_i) = \theta(\mathbf{N}'_i, \widehat{\mathbf{N}}'_i) = \theta(\mathbf{S}'_i, \widehat{\mathbf{S}}'_i) = pb(\mathbf{A}'_i, \widehat{\mathbf{A}}'_i). \quad (10)$$

3.1. Perturbation estimation of a submatrix

Using results from matrix perturbation theory, we now consider the problem of determining a submatrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ of a matrix $\mathbf{Q} \in \mathbb{R}^{m\times p}$ with rank $n < \min(m, p)$, where $\mathbf{Q}$ has been perturbed by a matrix $\mathbf{E}_Q \in \mathbb{R}^{m\times p}$. The perturbing matrix is considered to be noise in the observation of $\mathbf{Q}$. As such, it is generally assumed that $\|\mathbf{E}_Q\|_F << \|\mathbf{Q}\|_F$ where $\|\cdot\|_F$ denotes the Frobenius norm. We can define the noisy matrices as:

$$\widehat{\mathbf{Q}} = \mathbf{Q} + \mathbf{E}_Q \quad \text{and} \quad \widehat{\mathbf{A}} = \mathbf{A} + \mathbf{E} \qquad (11)$$

where $\mathbf{E}$ is the perturbation on $\mathbf{A}$ caused by $\mathbf{E}_Q$.

Ultimately, we would like to find submatrices which are the most insensitive to perturbation. To this end, we begin by determining a bound on the distance between $\mathcal{R}(\widehat{\mathbf{A}})$ and $\mathcal{R}(\mathbf{A})$, where $\mathcal{R}(\cdot)$ denotes the range, in terms of some $f(\widehat{\mathbf{A}})$ and $\|\mathbf{E}\|_2$, where $\|\cdot\|_2$ denotes the 2-norm. Our idea is to use a perturbation theorem provided by Wedin [10] to determine a bound on the sine of the angle between the spanning spaces of $\mathbf{A}$ and $\widehat{\mathbf{A}}$. Wedin's theorem is general, in that it actually defines a bound between *subspaces* of $\mathbf{A}$ and $\widehat{\mathbf{A}}$. Before the theorem is presented, we begin with some definitions. Let

$$\mathbf{A} \in \mathbb{R}^{m\times n} \quad \text{and} \quad \widehat{\mathbf{A}} = \mathbf{A} + \mathbf{E} \qquad (12)$$

with SVDs $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and $\widehat{\mathbf{A}} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^T$. $\mathbf{A}$ and $\widehat{\mathbf{A}}$ can be decomposed as

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{bmatrix} \begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{V}_1 & \mathbf{V}_2\end{bmatrix}^T \\ \widehat{\mathbf{A}} &= \begin{bmatrix}\widehat{\mathbf{U}}_1 & \widehat{\mathbf{U}}_2\end{bmatrix} \begin{bmatrix}\widehat{\mathbf{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Sigma}}_2 \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\widehat{\mathbf{V}}_1 & \widehat{\mathbf{V}}_2\end{bmatrix}^T \quad , \end{aligned} \tag{13}$$

where $\mathbf{U}_1, \widehat{\mathbf{U}}_1 \in \mathbb{R}^{m\times k}$, $\mathbf{U}_2, \widehat{\mathbf{U}}_2 \in \mathbb{R}^{m\times(m-k)}$, $\mathbf{V}_1, \widehat{\mathbf{V}}_1 \in \mathbb{R}^{n\times k}$, $\mathbf{V}_2, \widehat{\mathbf{V}}_2 \in \mathbb{R}^{n\times(n-k)}$, $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \cdots, \sigma_k)$, $\mathbf{\Sigma}_2 = \text{diag}(\sigma_{k+1}, \cdots, \sigma_n)$, $\widehat{\mathbf{\Sigma}}_1 = \text{diag}(\widehat{\sigma}_1, \cdots, \widehat{\sigma}_k)$ and $\widehat{\mathbf{\Sigma}}_2 = \text{diag}(\widehat{\sigma}_{k+1}, \cdots, \widehat{\sigma}_n)$.

In the theorem that follows, the representation of $\mathbf{E}$ in the orthonormal subspace $\widehat{\mathbf{V}}_1$ is used, rather than $\mathbf{E}$ directly, since we are defining bounds for subspaces. As such, define

$$\begin{aligned} \mathbf{R} &= \mathbf{A}\widehat{\mathbf{V}}_1 - \widehat{\mathbf{U}}_1\widehat{\mathbf{\Sigma}}_1 = -\mathbf{E}\widehat{\mathbf{V}}_1 \\ \mathbf{S} &= \mathbf{A}^T\widehat{\mathbf{U}}_1 - \widehat{\mathbf{V}}_1\widehat{\mathbf{\Sigma}}_1 = -\mathbf{E}^T\widehat{\mathbf{U}}_1 \quad . \end{aligned} \tag{14}$$

Note that

$$\begin{aligned} \|\mathbf{R}\| &\leq \|\mathbf{E}\widehat{\mathbf{V}}_1\| \leq \|\mathbf{E}\| \\ \|\mathbf{S}\| &\leq \|\mathbf{E}^T\widehat{\mathbf{U}}_1\| \leq \|\mathbf{E}\| \quad . \end{aligned} \tag{15}$$

Here $\|\cdot\|$ represents 2-norm or Frobenius norm.

Given the aforementioned definitions, Wedin's theorem states [10]:

Theorem 1. *If $\exists \alpha, \delta > 0$ such that*

$$\min\sigma(\widehat{\mathbf{\Sigma}}_1) \geq \alpha + \delta \quad \textit{and} \quad \max\sigma(\mathbf{\Sigma}_2) \leq \alpha$$

then

$$\max\{\|\sin\mathbf{\Phi}\|, \|\sin\mathbf{\Theta}\|\} \leq \frac{\max\{\|\mathbf{R}\|, \|\mathbf{S}\|\}}{\delta}$$

where $\mathbf{\Phi}$ is a matrix of angles between $\mathcal{R}(\mathbf{U}_1)$ and $\mathcal{R}(\widehat{\mathbf{U}}_1)$, $\mathbf{\Theta}$ is a matrix of angles between $\mathcal{R}(\mathbf{V}_1)$ and $\mathcal{R}(\widehat{\mathbf{V}}_1)$, and the operator $\sigma(\cdot)$ denotes the singular value spectrum.

Returning to the original problem, to determine a bound on the distance between the spanning spaces $\mathcal{R}(\widehat{\mathbf{A}})$ and $\mathcal{R}(\mathbf{A})$, we don't need to refer to subspaces of these matrices. Further, using the sine of the largest angle between $\mathcal{R}(\widehat{\mathbf{A}})$ and $\mathcal{R}(\mathbf{A})$, defined as $\theta(\mathbf{A}, \widehat{\mathbf{A}})$, and the inequality in (15),

$$\begin{aligned} \theta(\mathbf{A}, \widehat{\mathbf{A}}) &\leq \max\{\|\sin\mathbf{\Phi}\|, \|\sin\mathbf{\Theta}\|\} \\ &\leq \frac{\max\{\|\mathbf{R}\|, \|\mathbf{S}\|\}}{gap} \leq \frac{\|\mathbf{E}\|}{gap} \end{aligned} \tag{16}$$

where $\mathbf{\Phi}$ is a matrix of angles between $\mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\widehat{\mathbf{A}})$ and $\mathbf{\Theta}$ is a matrix of angles between $\mathcal{R}(\mathbf{A}^T)$ and $\mathcal{R}(\widehat{\mathbf{A}}^T)$. From the conditions in the theorem, we can define $gap = \min\sigma(\widehat{\mathbf{\Sigma}}_1) - \max\sigma(\mathbf{\Sigma}_2)$, with $\mathbf{\Sigma}_2$ and $\widehat{\mathbf{\Sigma}}_1$ diagonal matrices of the singular values of $\mathbf{A}$ and $\widehat{\mathbf{A}}$ respectively.

Finally, for the case where $m > n$, the first n left singular vectors of $\widehat{\mathbf{U}}$ will span $\mathcal{R}(\widehat{\mathbf{A}})$ and the remaining singular vectors will have corresponding singular vales equal to 0 (the so-called "ghost" singular values). To illustrate, consider the following SVD for $m > n$, rewritten with a square matrix of singular values:

$$\mathbf{A} = \begin{bmatrix}\mathbf{U}'_1 & \mathbf{U}'_2\end{bmatrix} \begin{bmatrix}\mathbf{\Sigma}'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}'_2\end{bmatrix} \begin{bmatrix}\mathbf{V}'_1 & \mathbf{V}'_2\end{bmatrix}^T \tag{17}$$

where $\mathbf{U}'_1 \in \mathbb{R}^{m\times n}$, $\mathbf{U}'_2 \in \mathbb{R}^{m\times(m-n)}$, $\mathbf{V}'_1 \in \mathbb{R}^{n\times n}$, $\mathbf{V}'_2 \in \mathbb{R}^{n\times(m-n)}$, $\mathbf{\Sigma}'_1 = \text{diag}(\sigma_1, \cdots, \sigma_n)$ and $\mathbf{\Sigma}'_2 = \text{diag}(\sigma_{n+1}, \cdots, \sigma_m)$. A similar decomposition exists for $\widehat{\mathbf{A}}$.

Comparing this to (13) if $k = n$, $\mathbf{U}'_1 = \mathbf{U}_1$, $\mathbf{U}'_2 = \mathbf{U}_2$, $\mathbf{\Sigma}'_1 = \mathbf{\Sigma_1}$, $\mathbf{\Sigma}'_2 = \mathbf{0}$, $\mathbf{V}'_1 = \mathbf{V}_1$ and $\mathbf{V}'_2 = \mathbf{I}$. From this we can conclude that in the condition in Wedin's theorem, $\max\sigma(\mathbf{\Sigma}'_2) = 0$ and $\min\sigma(\widehat{\mathbf{\Sigma}}'_1) = \min\sigma(\widehat{\mathbf{A}})$. Thus, the gap only depends on the minimum singular value of the perturbed matrix. This gives our expression for the bound on the distance between $\mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\widehat{\mathbf{A}})$ as:

$$\theta(\mathbf{A}, \widehat{\mathbf{A}}) \leq \frac{\|\mathbf{E}\|}{\min\sigma(\widehat{\mathbf{\Sigma}}'_1) - \max\sigma(\mathbf{\Sigma}'_2)} = \frac{\|\mathbf{E}\|}{\min\sigma(\widehat{\mathbf{A}})}. \tag{18}$$

In order to use the perturbation bound in (18) we must have some knowledge of the nature of the noise matrix $\mathbf{E}$. Since it cannot be measured directly, we will resort to a statistical model. Specifically, we assume that the elements of the noise matrix are Gaussian distributed according to $N(0, \sigma^2)$ and are independent of each other. This is a reasonable characterization of noise that arises from inaccurate measurements – it is most probable that the measurement will be close to the actual value. It was shown in [11] that the mean and variance of the largest eigenvalue λ_1 of the covariance matrix of an $n \times p$ Gaussian random matrix with entries distributed according to $N(0, 1)$ are:

$$\begin{aligned} \mu_{\lambda_1} &= (\sqrt{n-1} + \sqrt{p})^2 \\ \sigma_{\lambda_1} &= (\sqrt{n-1} + \sqrt{p}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}}. \end{aligned} \tag{19}$$

Given that the largest singular value of a matrix is its 2-norm, if we consider a Gaussian random matrix $\mathbf{X}$ with elements distributed as $N(0, \sigma^2)$, then $\mathbf{X} = \sigma\mathbf{X}_N$ where $\mathbf{X}_N$ is a Gaussian random matrix with each element distributed according to $N(0, 1)$. As such, $\|\mathbf{X}\|_2 = \sigma\|\mathbf{X}_N\|_2$. Similarly, the variance can be shown to be scaled by σ^2. Hence for the mean of the 2-norm of our noise matrix $\mathbf{E}$, we have:

$$\mu_{\|\mathbf{E}\|} = (\sqrt{n-1} + \sqrt{p})\sigma. \tag{20}$$

Notice also that from (19), the rate of positive change of the standard deviation becomes increasingly small as the size of the matrix increases, indicating that we can be quite confident of the the mean of the 2-norm for a modestly large n and p.

3.2. Deviation Parameter

From the upper bound on (18) and (20) we will define a deviation parameter $DP(\widehat{\mathbf{A}}_i)$ of submatrix $\widehat{\mathbf{A}}_i \in \mathbb{R}^{P\times 3}$ with x_i rows of missing data as:

$$DP(\widehat{\mathbf{A}}_i) = \frac{(\sqrt{P - x_i - 1} + \sqrt{3})\sigma}{\min(\sigma(\widehat{\mathbf{A}}'_i))}, \tag{21}$$

which provides measure of the sensitivity of the submatrix to perturbation due to noise. A small DP for the submatrix makes it a suitable candidate for inclusion in the final matrix $\widehat{\mathbf{N}}$. The deviation parameter is affected by the noise in measurement matrix, the number of missing elements as well as the singular value spectrum of the submatrix. If the noise is scaled, the numerator of DP will also be scaled. If the number of the non-full rows (x_i) increases, both $(\sqrt{P - x_i - 1} + \sqrt{3})$ and $\min(\sigma(\widehat{\mathbf{A}}'_i))$ will decrease. The fact that the denominator decreases as a result of constructing a reduced $\mathbf{A}' \in \mathbb{R}^{r\times n}$ from r rows of $\mathbf{A} \in \mathbb{R}^{m\times n}$, $r < m$ can be seen by comparing the smallest singular value of the full matrix to that of the submatrix:

$$\begin{aligned}\min(\sigma(\mathbf{A})) &= \min_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \min_{\|\mathbf{x}\|_2=1} \sqrt{\sum_i (\mathbf{a}_i\mathbf{x})^2} \\ &\geq \min_{\|\mathbf{x}\|_2=1} \sqrt{\sum_j (\mathbf{a}_j\mathbf{x})^2} = \min_{\|\mathbf{x}\|_2=1} \|\mathbf{A}'\mathbf{x}\|_2 = \min(\sigma(\mathbf{A}'))\end{aligned} \tag{22}$$

with $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots r$ and $\mathbf{a}_k \in \mathbb{R}^{1\times n}$ is the k^{th} row of $\mathbf{A}$. This shows that we cannot determine if a submatrix provides useful information for shape recovery by simply considering the number of missing elements. Our criterion represents a significant improvement over previous methods in that it also includes a data dependent measure of the sensitivity of the selected submatrix to noise.

3.3. The Deviation Parameter Algorithm

A new algorithm with DP based submatrix selection will be described next. Although the main procedure of the algorithm is the same as Jacobs', we have made several key improvements. First, since the optimal selection is extremely difficult to achieve and the DP criterion cannot guarantee the optimal selection of submatrices, we implement the same algorithm on n different subsets of the matrix, and use the average of all recovered results. Second, we only select the submatrix whose columns in the original matrix are well separated in position. This is because if the motion between two successive frames is small, the localization error on the feature points will seriously degrade the reliability of the data. We set a minimum difference min_diff between the largest and smallest ordinal number of three columns as the ordinal difference condition. Last, we use the most reliable submatrix to recover the 3D structure. This way we can use fewer submatrices to build a smaller final matrix than that used in [6]. The steps in the DP algorithm are:

1. Select a subset of images and use the 2D projective coordinates of all the feature points to build a measurement data matrix.
2. Calculate the deviation parameters of all the possible three-column submatrices which satisfy the ordinal difference condition min_diff and sort all these values from small to large.
3. Choose the submatrices with smallest deviation parameter to construct the final matrix.
4. Use SVD to get the rank 3 linear space by spanning the three singular vectors corresponding to the three smallest singular values, and recover the structure of the scene up to a Euclidean shape.
5. Repeat steps 1 to 4 n times, and set the average as the final solution.

4. Experiment

Our experiments are implemented on a synthetic data set which includes 18 feature points on a 3D face model. We select the mass center of all these points as the origin. This guarantees that there will be no translation in the data. The 3D feature points rotate randomly 30 times, and are mapped onto the xy-plane to build 18×60 2D feature point data matrix $\mathbf{Q}$. The noise matrix $\mathbf{E}$ will be generated on different variances σ^2 from 0.09 to 1. The maximal absolute value in $\mathbf{Q}$ is no more than 100, so the noise level is from 0.3% to 1%. To simulate the situation of missing data, we randomly occlude the data in $\widehat{\mathbf{Q}}$ with the ratio $m\%$ from 10% to 40%. We compare the recovered 3D structure from the DP method to the ground-truth. Comparisons are drawn

	% of submatrices with the smallest DP values					
(σ^2,m %)	0%	10%	20%	50%	80%	99%
(0.25,10%)	2.6851	2.4756	1.6231	4.7435	2101.6	1163.6
(0.25,40%)	1.4052	6.4683	21.879	1.03E6	3.39E6	2.86E6
(1.00,10%)	0.5732	0.3655	0.2680	0.5214	16.1575	319.45
(1.00,40%)	3.7572	2.1422	120.87	13281	1.87E6	2.59E6

Table 1: MSE results for a varying percentage of submatrices ranked by DP.

	(σ^2, m%)			
	(0.09,10%)	(0.09,40%)	(0.25,10%)	(0.25,40%)
J_{max}	2.6557	6616.4	3.2029	1108.6
J_{mean}	2.3325	4959.8	2.5617	176.89
J_{min}	1.6877	2653.6	1.0656	13.439
J_{var}	0.0623	1.02E6	0.2874	82421
DP_{max}	1.3395	3.2419	0.5690	1.5477
DP_{mean}	1.0071	2.1368	0.4718	1.1838
DP_{min}	0.4301	0.9055	0.3325	0.6354
DP_{var}	0.0589	0.4065	0.0039	0.0506
DP_{avg}	0.8655	1.1178	0.3107	0.6891
	(σ^2, m%)			
	(0.49,10%)	(0.49,40%)	(1.00,10%)	(1.00,40%)
J_{max}	1.9762	11253	7.4741	4074.8
J_{mean}	1.5972	6595.8	6.0457	1981.0
J_{min}	1.0910	2009.6	4.1277	354.51
J_{var}	0.0453	7.15E6	0.7951	1.19E6
DP_{max}	1.5251	238.03	4.1117	2.8346
DP_{mean}	1.1468	58.513	2.8940	1.7654
DP_{min}	0.6105	2.0510	1.9444	0.8387
DP_{var}	0.0633	4125.6	0.3167	0.2529
DP_{ave}	0.6576	15.380	1.6346	0.8984

Table 2: MSE results of Jacobs' algorithm and the DP algorithm.

	(σ^2, m%)			
	(0.09, 30%)	(0.09, 40%)	(1.00,30%)	(1.00,40%)
CS_{max}	9.5985	5041.6	519.68	2172.1
CS_{mean}	6.8287	4155.2	354.51	1188.6
CS_{min}	2.6021	2232.1	104.37	101.28
CS_{var}	3.2797	5.10E5	8768.6	4.21E5
DP_{max}	0.4169	3.2419	8.4704	2.8346
DP_{mean}	0.3193	2.1368	5.7242	1.7654
DP_{min}	0.1578	0.9055	2.6806	0.8387
DP_{var}	0.0044	0.4065	2.2694	0.2529

Table 3: MSE results with MUR and DP as the selection criterion.

against Jacobs' algorithm and Chen and Suter's method. All the comparisons are based on the Mean Square Error (MSE) between the ground-truth and the recovered one.

We first verify the effectiveness of DP. Under the conditions of missing data and noise level, we sort all submatrices in a sequence according to their DP values from small to large. In the sorted submatrix sequence, we select the smallest $k\%$ of them to build the final matrix $\widehat{\mathbf{N}}$ and recover the 3D structure. In Table 1, the recovery results in terms of MSE between the ground-truth and the recovered one are listed. We can see that on various noise and missing data conditions, the results recovered from the submatrices with smaller DP are better than those recovered with larger DP. Table 1 also indicates that we should select the submatrix from the set of submatrices with DP's ranked in the smallest 20%.

For comparison, we run Jacobs' algorithm and the DP algorithm 100 times under different noise and missing data conditions. Since both of them cannot guarantee an optimal solution, we compare the average performance based on the first 50 best MSE results. The maximum, mean, minimum and variance among the MSEs of these Jacobs' results are denoted by J_{max}, J_{mean}, J_{min} and J_{var}. DP_{max}, DP_{mean}, DP_{min} and DP_{var} are the corresponding values of the DP algorithm's results. The experimental results are listed in Table 2. In Table 2, the experimental results of the DP algorithm are always better than the results of Jacobs's method. When the data set has a higher noise level or a larger proportion of the data are missing, the DP algorithm's performance is much better than Jacobs'. The reliability of the new method is better because the variance of the recovered results of the DP method is much smaller than that of Jacobs'. Another important thing to notice from the experimental results is that if we average these 50 recovered 3D shapes to get an average shape, the average shape will be a better estimation than any individual one in general. In Table 2, the MSE of the average shape, DP_{avg}, is very close to DP_{min} in most cases or in some cases is smaller. This indicates that the noise effect can be weakened further by averaging multiple recovered results. Figure 1 shows the best results obtained with our algorithm under the condition ($\sigma^2, m\%$)=($1.00, 40\%$), which means large noise and a large proportion of missing data, for the 18 points face model overlayed on the ground-truth, illustrating a very small MSE. By comparison, Figure 2 exhibits a significant deviation from the ground-truth for the best results from Jacobs' algorithm under the same condition ($1.00, 40\%$).

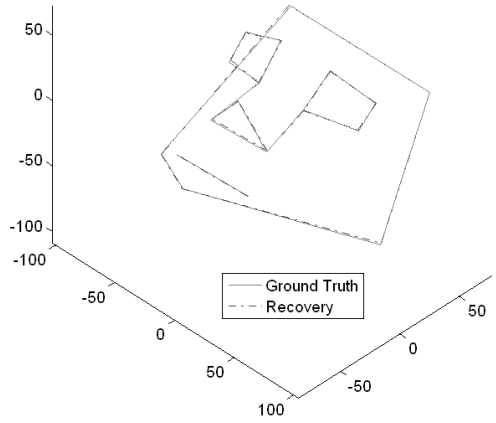

Figure 1: Best results for face model with DP algorithm.

In Chen and Suter's paper, the Minimal Unreliability Ratio (MUR) is proposed as the criterion to select the most reliable submatrix. We compared the performance between two data selection criteria, MUR and DP. The algorithm in [6] is used to recover the 3D structure in both cases. In the case of less missing data, only a few columns are excluded from the most reliable submatrix. For example, if there are about 20% missing data in 18×60 matrix, the number of

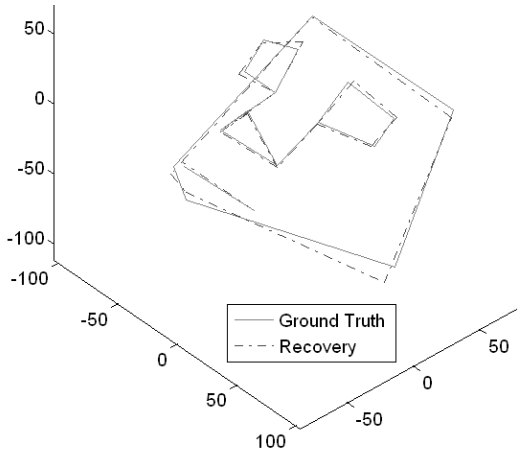

Figure 2: Best results for face model with Jacobs' algorithm.

columns in the most reliable submatrix is 56.91 on average over 100 times. And on average 50.65 columns are left in the most reliable submatrix if the missing data percentage is 40%. We give two sets of comparison results under the different percentage of missing data, 30% and 40% in Table 3. From the results, the performance of the DP algorithm is much better than that of MUR criterion since the MUR cannot reduce that large amount of columns. On the other hand, from Tables 2 and 3, MUR can identify some columns not suitable to be used in recovery, so its results are better than the recovery without any selection. Although we cannot compare the exact performance between Chen and Suter's method and the DP algorithm in this way, the overall effectiveness of our data selection strategy is illustrated.

5. Conclusion

In this paper, a new criterion, the deviation parameter DP, is proposed which can be used to recover the 3D structure more precisely with limited effects from the noise and missing data. This criterion, based on the perturbation estimation of the subspace, takes into account both the effect of noise and the ratio of missing data, as well as the data matrix itself. The DP method performs well under different noise levels and can tolerate large percentages of missing data.

Acknowledgments

This research was partially supported by NIH under grant R01 DC 005241.

References

[1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Int. J. of Comp. Vis.*, 9(2):137-154, 1992.

[2] M. Irani, "Multi-frame optical flow estimation using subspace constraints," *Int. Conf. on Computer Vision (ICCV)*, vol. 1, pp. 626-633, 1999.

[3] B. K. P. Horn and M. J. Brooks (eds.), "Shape from shading," *MIT Press*, Cambridge Massachusetts, 1989.

[4] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Trans. PAMI*, 13(10):992-1007, 1991.

[5] H. Shum, K. Ikeuchi and R. Reddy, "Principal component analysis with missing data and its applications to Polyhedral object modeling," *IEEE Trans. PAMI*, 17(9):854-867, 1995.

[6] D. Jacobs, "Linear fitting with missing data for structure-from-motion," *Computer Vision and Image Understanding*, 82(1):57-81, 2001.

[7] P. Chen and D. Suter, "Recovering the missing components in a large noisy low-rank matrix: application to SFM," *IEEE Trans. PAMI*, 26(8):1051-1063, 2004.

[8] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the EM algorithm," *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 707-714, 2004.

[9] G. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 3^{rd} ed., 1996.

[10] P. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT*, 12:99-111, 1972.

[11] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *The Annals of Statistics*, 29(2):295-327, 2001.