Supplementary Documentation

Hongjun Jia and Aleix M. Martinez

I. DP Criterion Analysis

We define the Deviation Parameter (DP) of a submatrix $\widehat{\mathbf{B}} \in \mathbb{R}^{m\times r}$ with b rows of missing data ($b = m - p$, where p is the number of full rows) as

$$DP(\widehat{\mathbf{B}}) = \frac{(\sqrt{m-b-1}+\sqrt{r})\sigma}{\min(\sigma(\widehat{\mathbf{A}}))}.$$

If the number of the non-complete rows (b) increases, both $(\sqrt{m-b-1}+\sqrt{r})$ and $\min(\sigma(\widehat{\mathbf{A}}))$ will decrease.

A. Proof of that $\min(\sigma(\widehat{\mathbf{A}}))$ will decrease as b increases

We construct a reduced matrix $\widehat{\mathbf{A}}_1 \in \mathbb{R}^{p_1\times r}$ from p_1 rows of $\widehat{\mathbf{B}} \in \mathbb{R}^{m\times r}$ ($b_1 = m - p_1$), and another matrix $\widehat{\mathbf{A}}_2 \in \mathbb{R}^{p_2\times r}$ from $\widehat{\mathbf{A}}_1$ ($b_2 = m - p_2$). So we have $p_2 < p_1$ and $b_2 > b_1$. We compare the smallest singular values of two matrices,

$$\begin{aligned}\min(\sigma(\widehat{\mathbf{A}}_1)) &= \min_{\|\mathbf{x}\|_2=1} \|\widehat{\mathbf{A}}_1\mathbf{x}\|_2 = \min_{\|\mathbf{x}\|_2=1} \sqrt{\sum_i (\mathbf{a}_i^T\mathbf{x})^2} \\ &\geq \min_{\|\mathbf{x}\|_2=1} \sqrt{\sum_j (\mathbf{a}_j^T\mathbf{x})^2} = \min_{\|\mathbf{x}\|_2=1} \|\widehat{\mathbf{A}}_2\mathbf{x}\|_2 = \min(\sigma(\widehat{\mathbf{A}}_2)),\end{aligned} \tag{14}$$

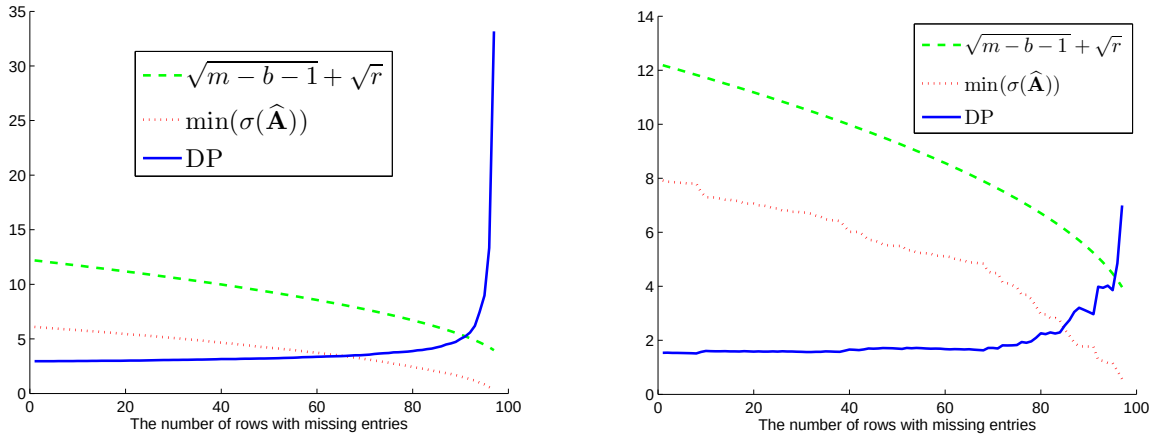
where $\mathbf{a}_k^T \in \mathbb{R}^{1\times r}$ is the k^{th} row of $\widehat{\mathbf{A}}_1$, $i = 1, 2, \cdots, p_1$, $j = l_1, l_2, \cdots, l_{p_2}$, and $\{l_1, l_2, \cdots, l_{p_2}\}$ is a subset with p_2 elements drawn from $\{1, 2, \cdots, p_1\}$.

B. Relationship between DP and the number of missing rows

Fig. S1 shows two examples where the minimum singular value (i.e., the denominator of DP) and the numerator of (12) decrease. In the figure we see that although both the denominator and numerator decrease, the value of DP increases. On average, the more one deletes rows with missing entries, the larger (12) becomes. Thus, in general, our criterion favors matrices with less missing rows, Fig. S1(a). Nevertheless, in each instance, some matrices with more missing rows may have lower DP values and they will be chosen for use in the proposed algorithm, Fig. S1(b).

II. Generation of random matrix of rank r

We generate a random $m \times n$ matrix of rank r with the absolute values of its entries no more than 100 in the following way. First, we generate two random matrices with uniformly distributed entries within $[0, 1]$ with the size $m \times r$ and $r \times n$ respectively. These two matrices multiply to get a rank-r matrix. And the resulting matrix is divided by the entry with the largest absolute value and multiplied by 100 to get a rank-r matrix.



(a) (b)

Fig. S1. (a) The average results over 100 trials. (b) A typical case for the relationship between DP and the number of missing rows. x-axis denotes the number of rows with missing entries.

III. Additional Comparative Results

Several methods proposed in the literature provide a global criterion (e.g., minimization of the reprojection error over all visible data) and solve the problem iteratively as an EM-like approach [2],[11]. The review paper of Buchanan and Fitzgibbon [3] lists several methods of this kind [2],[11],[13]. For example, Brandt [2] proposes an EM-based method to solve the missing-data problem under affine projection. At each iteration of the algorithm (and assuming that the camera motion matrices are given), a closed-form maximum likelihood solution is derived to minimize the reprojection error. Global criteria such as this usually over-fit to the visible data. Buchanan and Fitzgibbon [3] introduce an extra penalty term to alleviate this effect, such as orthogonality. Variations [27],[29],[36] of Wiberg's approach [39] use the alternation strategy to find the closed-form solution of one of the factor while fixing the other.

Finally, Robust Principle Component Analysis (RPCA) (as well as robust SVD) is proposed by De la Torre and Black [5]. These approaches are based on robust statistics and can deal with noise and missing data – the approach will treat them as outliers. This should work reasonably well for low noise and occlusion terms, but will fail for large ones.

We present a comparison of our algorithm to the methods described in [2],[5] and to the variant of RANSAC defined in Section V. All comparisons are obtained using the images of the *box sequence*. The results are in Table IV. Note that RANSAC cannot be applied directly in the problem of SFM with missing data, since the subset selection cannot guarantee the reconstruction of a full submatrix. To avoid this problem, in our implementation of RANSAC, we delete those rows with missing elements. The remaining part of the original data matrix is used to fit the other columns. The overall fitting error is then used to rank the subsets, and the same mechanism used in our approach is used to select the number of columns. Hence, the number of submatrices and the threshold used to discriminate between inliers and outliers are not the classical ones used in RANSAC, but are intrinsically given by the mechanism described above.

The major problem with global approaches is that they are based on the fitting to the non-occluded entries. This may lead to over-fitting. The main reason why our approach outperforms global methods is that it cares about the difference between the approximation matrix ($\mathbf{W}_r$) and the ground-truth ($\mathbf{W}$), i.e. $diff_3$. Over-fitting should not generally happen in our algorithm since our DP criterion is based on the upper-bound of the difference to the ground-truth. In Fig.S2, an example of over-fitting is shown. In this example $diff_3$ becomes diverging while $diff_1$ is still converging, i.e., global methods will increase the value of $diff_3$ for smaller values of $diff_1$.

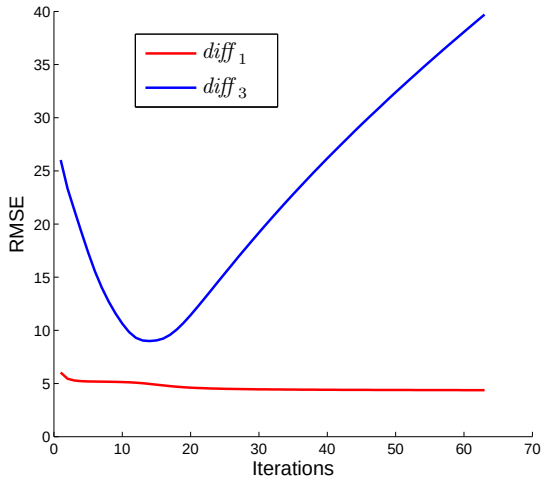


Fig. S2. One example of over-fitting for global methods.

IV. Notation

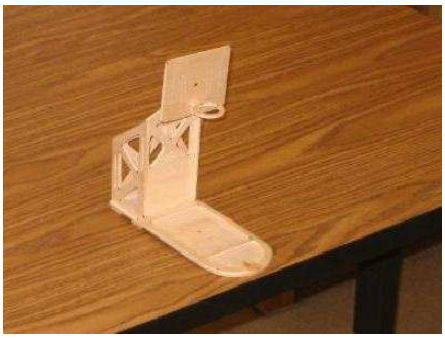
In this paper, matrices are represented with bolded capital letters (e.g., $\mathbf{W}$). Their $(i,j)^{th}$ entry is denoted as w_{ij}. Furthermore, $[(\cdot)_{ij}]$ denotes a matrix with entry $(\cdot)_{ij}$ in the $(i,j)^{th}$ position. This allows us to write $\mathbf{W}$ as $[w_{ij}]$. Noisy matrices, with or without missing entries, are represented with a hat symbol $\widehat{\ }$, e.g., $\widehat{\mathbf{W}}$. The spanning space of the column vectors in a matrix, also known as the range space, is described using the corresponding calligraphic capital letters, e.g. $\mathcal{W}$. The same letter is used to represent the space and the data matrix. $\mathcal{W}$ representing the range space of the matrix $\mathbf{W}$. We can also write $\mathcal{W} = \mathcal{R}(\mathbf{W})$, where $\mathcal{R}$ means range space. The orthogonal space to $\mathcal{G}$ is described using the capitalized letter and an orthogonal symbol, $\mathcal{G}^{\perp}$. A detailed list of symbols and variables is in Table S1.

$\mathbf{W}$	noise-free data matrix (in SFM, it contains the 2D coordinates of the feature points)
$\mathcal{W}$	range space of $\mathbf{W}$
$\widehat{\mathbf{W}}$	measurement matrix with noise and missing entries
$\mathbf{W}_r$	rank-r recovered matrix from $\widehat{\mathbf{W}}$
$\mathbf{E}$	noise matrix
m	number of rows in the matrix $\mathbf{W}$
n	number of columns in the matrix $\mathbf{W}$ (i.e., number of feature points)
q	number of images
r	rank of $\mathbf{W}$
$d\%$	percentage of missing entries in $\widehat{\mathbf{W}}$
σ	variance of the additive Gaussian noise
Γ	visible mask of $\widehat{\mathbf{W}}$
$\mathbf{P}$	camera motion matrix in SFM
$\mathbf{Q}$	shape matrix in SFM
l	number of r-column submatrices of $\widehat{\mathbf{W}}$ or $\mathbf{W}$
$\mathcal{B}_k$	range space of the k^{th} r-column submatrix of $\mathbf{W}$
$\widehat{\mathcal{B}}_k$	range space of the k^{th} r-column submatrix of $\widehat{\mathbf{W}}$
$\mathcal{G}$	intersection space of all $\mathcal{B}_k$
$\mathcal{G}^{\perp}$	orthogonal space to $\mathcal{G}$
$\mathbf{u}$	column vector
$\theta(\mathbf{X}, \mathbf{Y})$	largest principal angle between the spaces defined by the matrices $\mathbf{X}$ and $\mathbf{Y}$
$\theta(\mathcal{X}, \mathcal{Y})$	largest principal angle between the spaces $\mathcal{X}$ and $\mathcal{Y}$
$\widehat{\mathbf{B}}, \widehat{\mathbf{B}}_i$	r-column submatrix from $\widehat{\mathbf{W}}$
$\widehat{\mathbf{A}}, \widehat{\mathbf{A}}_i$	complete part of $\widehat{\mathbf{B}}$
$\mathbf{E}_{\mathbf{A}}$	submatrix of $\mathbf{E}$ defining the noise term on $\widehat{\mathbf{A}}$
$\widehat{\mathbf{M}}_i$	matrix defining the null space of $\widehat{\mathcal{A}}_i$
$\widehat{\mathbf{N}}_i$	matrix extending $\widehat{\mathbf{M}}_i$ with zeros
$\widehat{\mathbf{N}}$	matrix containing all $\widehat{\mathbf{N}}_i$, $\widehat{\mathbf{N}} = [\widehat{\mathbf{N}}_1 \widehat{\mathbf{N}}_2 \cdots \widehat{\mathbf{N}}_l]$
$\sigma(\widehat{\mathbf{A}})$	singular value spectrum of $\widehat{\mathbf{A}}$
b	number of rows in $\widehat{\mathbf{B}}$ with at least one missing entry
$DP(\widehat{\mathbf{B}})$	deviation parameter of $\widehat{\mathbf{B}}$
$\mathbf{p}_{ij}$	2D coordinates of the feature point j in image i, $\mathbf{p}_{ij} = [x_{ij}, y_{ij}]^T$
$\mathbf{q}_{ij}$	2D homogeneous coordinates of the feature point j in image i, $\mathbf{q}_{ij} = [x_{ij}, y_{ij}, 1]^T$
λ_{ij}	projective depth of $\mathbf{q}_{ij}$
$\mathbf{S}$	scaled measurement matrix

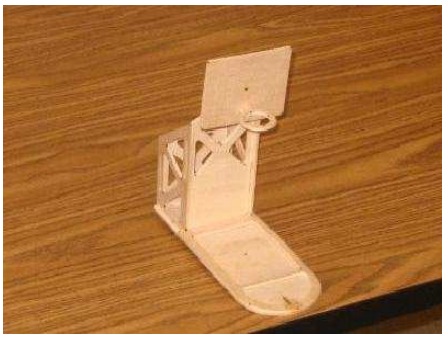
TABLE S1
NOTATION

V. THE WOODEN OBJECT SEQUENCE

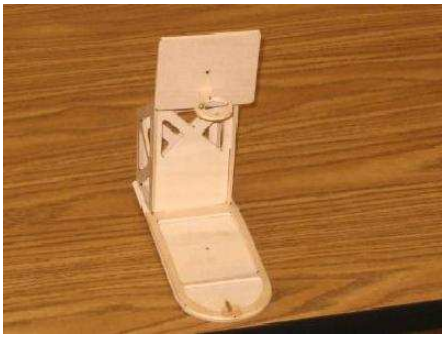
The wooden object sequence used in the paper consists of 7 images. These images are shown in Fig. S3(a-g). In these images, we have selected a total of 32 feature points, which are tracked over the video sequence. Several of the points get occluded when the object rotates, leading to 25% of missing entries in the data matrix. Fig. S4 shows the 2D recovered feature points (small dots) by the proposed DP-based algorithm. These have been overlaid on top of the originally marked positions (specified with large squares). The reconstructed 3D object structure is shown in Fig. S5. To further verify the validity of our reconstruction, we now add three virtual cubes to the 3D structure and project them onto the individual images. The results are in Fig. S6.



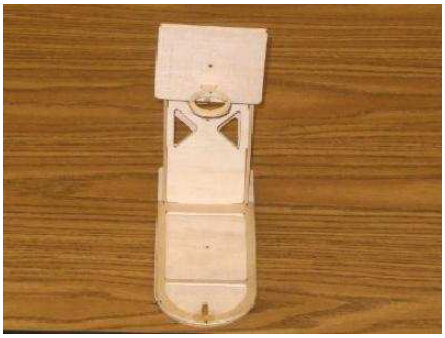
(a)



(b)

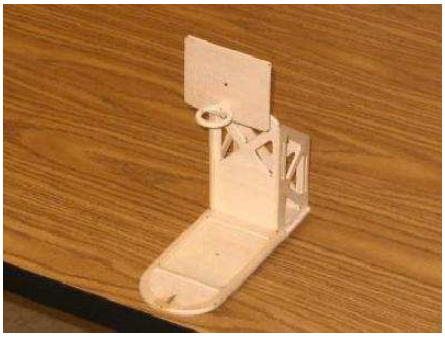


(c)

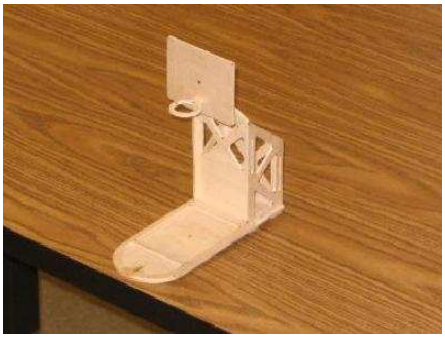


(d)

(e)



(f)



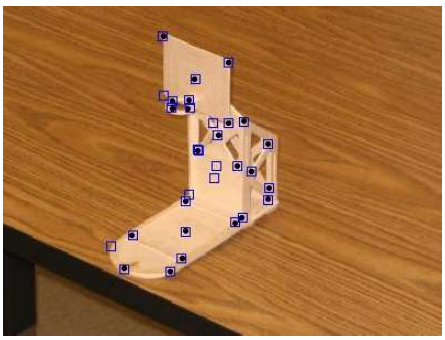
(g)

Fig. S3. The seven images of the wooden object sequence.

(a) (b)

(c) (d)

(e) (f)



(g)

Fig. S4. Marked (dots) and recovered (squares) feature points for the wooden object sequence.

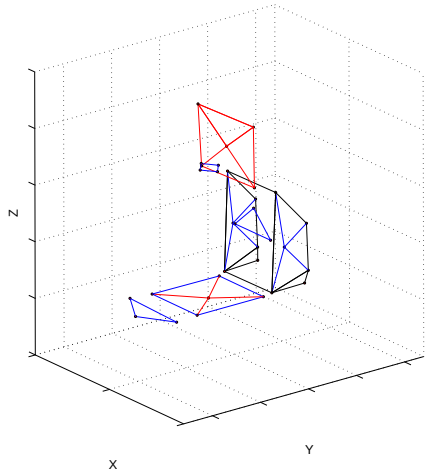
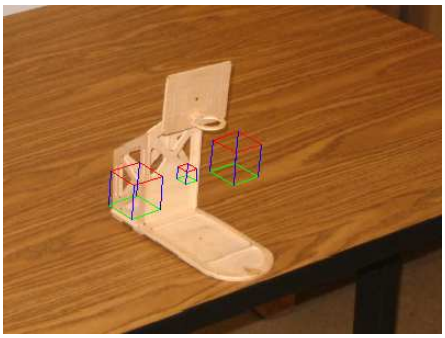
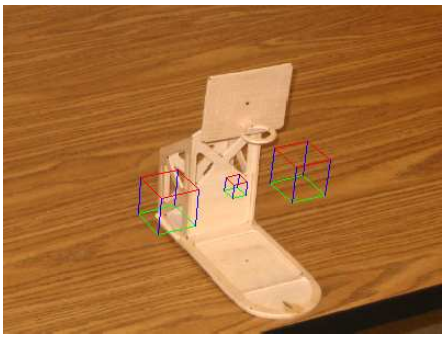


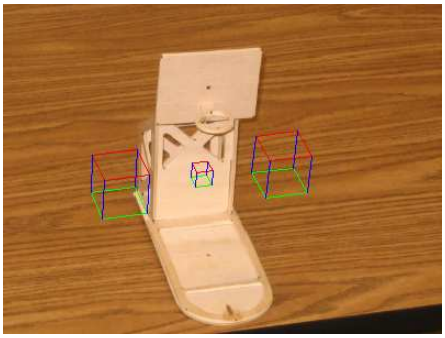
Fig. S5. Recovered 3D structure of the wooden object.



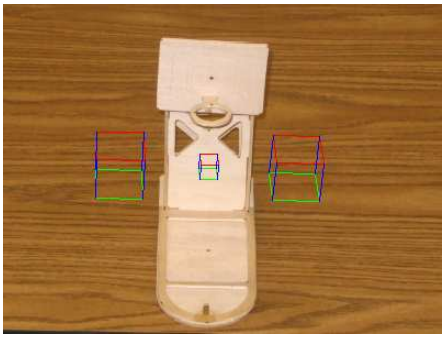
(a)



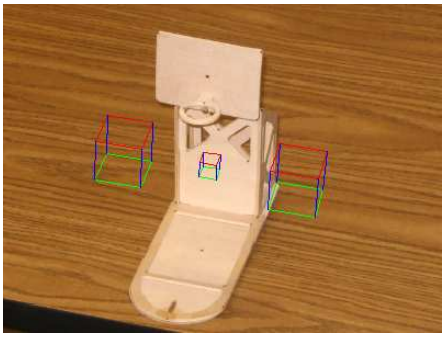
(b)



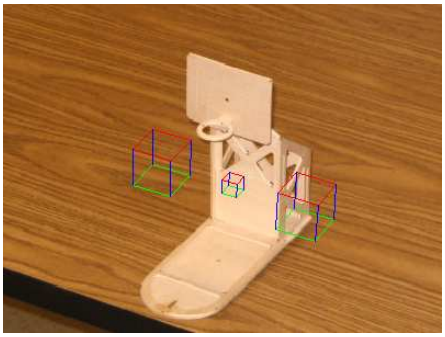
(c)



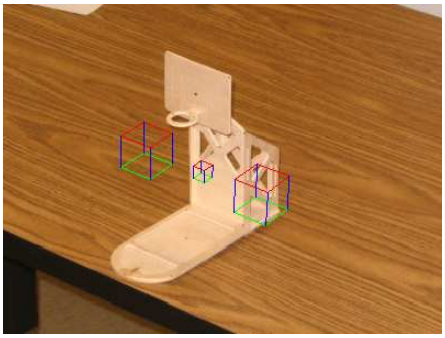
(d)



(e)



(f)



(g)

Fig. S6. Projection of a pair of 3D cubes onto each of the images.