Supplementary Material:
Partial Sum Minimization of Singular Values in RPCA for Low-Level Vision

Due to space limitation in the main paper, we present additional experimental results in this supplementary material. We show additional results for synthetic dataset, and real-world applications, such as background modeling, High Dynamic Range Imaging, outlier removal for photometric stereo and batch image alignment as well as proof of the main result, Partial Singular Value Thresholding (PSVT) operator. All the parameter not mentioned are same with the parameters used in the main paper.

Index of Contents

• Proof of Partial Singular Value Thresholding

• Synthetic Evaluations
  – Fig. 1: Comparisons of success ratio for other rank conditions (rank 2, 5 and 10).
  – Fig. 2: Comparisons of success ratio for varying row \( m \) (i.e. data dimension).
  – Fig. 3: Comparisons whether the recovered \( \hat{A} \) obtained by RPCA and our method satisfies the target rank constraint.

• Background Modeling and High Dynamic Range (HDR) Imaging
  – Fig. 4: Comparisons of low-rank matrix and sparse error between RPCA and our approach on Sculpture Garden sequence.
  – Fig. 5: HDR results of Sculpture Garden sequence obtained by RPCA and our approach.

• Outlier Removal for Photometric Stereo
  – Fig. 6: Outlier rejection results for Photometric Stereo on Horse dataset.
  – Fig. 7: Comparison of the photometric normal among the least square method (LS), Wu et al. [4] and ours on Horse dataset.
  – Table 1: Singular values of inputs and refined data by Wu et al. [4] and ours on Horse dataset.

• Batch Image Alignment
  – Fig. 8: The batch image alignment with similarity model from RASL [3] and ours for \( n = 10 \).
  – Fig. 9: The batch image alignment with similarity model from RASL [3] and ours for \( n = 100 \).
  – Fig. 10: The batch image alignment with affine model from RASL [3] and ours for \( n = 4 \).
  – Fig. 11: Detail comparisons of Fig. 10.
  – Fig. 12: The batch image alignment with affine model from RASL [3] and ours for \( n = 16 \).
  – Fig. 13: Detail comparisons of Fig. 12.

• References
Proof of the Partial Singular Value Thresholding (PSVT) operator

We present the proof of the proposed PSVT (see Sec. 4 of the main paper). We first introduce the von Neumann’s lemma (see the details in de S\'a et al. [8]).

Lemma (von Neumann)  For any matrices \( B, Z \in \mathbb{R}^{m \times n} \) and \( \sigma(\cdot) \) is a vector of the singular values, the following equality holds:

\[
\max \{ \langle U Z V^T, B \rangle | U \in \mathcal{U}_m, V \in \mathcal{U}_n \} = \langle \sigma(Z), \sigma(B) \rangle, \tag{1}
\]

where \( \mathcal{U}_n \) denotes the set of \( n \times n \) unitary matrices, \( < A, B > = \text{Tr}(A^T B) \) and, for any matrix \( A \in \mathbb{R}^{m \times n} \), hence

\[
\langle A, B \rangle \leq \langle \sigma(A), \sigma(B) \rangle. \tag{2}
\]

Moreover, equality holds in Eq. (2) if and only if there exists a simultaneous singular value decomposition (SVD) \( U \) and \( V^T \) of \( A \) and \( B \) as the following form:

\[
A = U \cdot \text{diag}(\sigma(A)) \cdot V^T, \quad \text{and} \quad B = U \cdot \text{diag}(\sigma(B)) \cdot V^T. \tag{3}
\]

von Neumann’s lemma shows that the value of \( \langle A, B \rangle \) is always bounded by the inner product of \( \sigma(A) \) and \( \sigma(B) \). Notice that the maximum value of \( \langle A, B \rangle \) can be only achieved when \( A \) has the same singular vector \( U \) and \( V \) of \( B \). This fact is useful to derive the PSVT.

Problem [PSVT]  Let \( \tau > 0 \), \( l = \min(m,n) \), and \( X, Y \in \mathbb{R}^{m \times n} \) which can be decomposed by SVD and \( Y \) can be considered as the sum of two matrices, \( Y = Y_1 + Y_2 = U_{Y_1} D_{Y_1} V_{Y_1}^T + U_{Y_2} D_{Y_2} V_{Y_2}^T \), where \( U_{Y_1}, V_{Y_1} \) are the singular vector matrices corresponding to the largest singular values from the first to the \( N \)-th, and \( U_{Y_2}, V_{Y_2} \) are the singular vector matrices corresponding to the largest singular values from \( N+1 \)-th to the last from SVD. Let’s consider the following minimization problem

\[
\arg \min_X \frac{1}{2} \| X - Y \|_F^2 + \tau \sum_{i=N+1}^l \sigma_i(X), \tag{4}
\]

Its solution is given by the following PSVT operator:

\[
P_{N,\tau}(Y) = U_Y (D_{Y_1} + S_\tau[D_{Y_2}]) V_{Y_1}^T = Y_1 + U_{Y_2} S_\tau[D_{Y_2}] V_{Y_2}^T,
\]

where \( D_{Y_1} = \text{diag}(\sigma_1, \ldots, \sigma_N, 0, \ldots, 0) \),

\[
D_{Y_2} = \text{diag}(0, \ldots, 0, \sigma_{N+1}, \ldots, \sigma_l),
\]

where \( S_\tau[\cdot] \) represents the soft-thresholding (or shrinkage) operator [1].

Proof of PSVT  Let’s consider \( X = U_X D_X V_X^T = \sum_{i=1}^l \sigma_i(X) u_i v_i^T \) where \( U_X = [u_1, \ldots, u_m] \in \mathcal{U}_m \), \( V_X = [v_1, \ldots, v_n] \in \mathcal{U}_n \) and \( D_X = \text{diag}(\sigma(X)) \), where the singular values \( \sigma(\cdot) = [\sigma_1(\cdot), \ldots, \sigma_l(\cdot)] \geq 0 \) are sorted in a non-increasing order. Eq. (4) can be derived as follows:

\[
\frac{1}{2} \| X - Y \|_F^2 + \tau \sum_{i=N+1}^l \sigma_i(X) = \frac{1}{2} \left( \| Y \|_F^2 - 2 \langle X, Y \rangle + \| X \|_F^2 \right) + \tau \sum_{i=N+1}^l \sigma_i(X)
\]

\[
= \frac{1}{2} \left( \| Y \|_F^2 - 2 \sum_{i=1}^l \sigma_i(X) u_i^T Y v_i + \sum_{i=1}^l \sigma_i(X)^2 \right) + \tau \sum_{i=N+1}^l \sigma_i(X)
\]

\[
= \frac{1}{2} \| Y \|_F^2 + \frac{l}{2} \sum_{i=1}^l ( -2 \sigma_i(X) u_i^T Y v_i + \sigma_i(X)^2 ) + \tau \sum_{i=N+1}^l \sigma_i(X). \tag{6}
\]

For more detail representation, we change the parameterization of \( X \) to \( (U_X, V_X, D_X) \) and minimize the function:
Figure 1. Success ratio for synthetic data with varying the number of columns \( n \). Comparison between RPCA (Top row) and ours (Bottom row) for rank-2, 5, 10 cases. X-axis represents the column size, and Y-axis represents the corruption ratio \( r \in [0, 0.4] \). We fixed \( m = 10000 \) and varied \( n \) and \( r \). Color magnitude represents success ratio \([0,1]\). The white dotted lines are provided as a guide for easier comparison.

\[
J(U_X, V_X, D_X) = \frac{1}{2} \sum_{i=1}^{l} (-2\sigma_i(X)u_i^TYv_i + \sigma_i(X)^2) + \tau \sum_{i=N+1}^{l} \sigma_i(X). \tag{7}
\]

From von Neumann’s lemma, the upper bound of \( u_i^TYv_i \) is given as \( \sigma_i(Y) = \max \{ u_i^TYv_i \} \) for all \( i \) when \( U_X = U_Y \) and \( V_X = V_Y \). The lower bound envelope of \( J(U_X, V_X, D_X) \) is obtained at \( U_X = U_Y \) and \( V_X = V_Y \). Then Eq. (7) becomes a function only depending on \( D_X \) as follows:

\[
J(U_Y, V_Y, D_X) = \frac{1}{2} \sum_{i=1}^{l} (-2\sigma_i(X)\sigma_i(Y) + \sigma_i(X)^2) + \tau \sum_{i=N+1}^{l} \sigma_i(X)
= \frac{1}{2} \left( \sum_{i=1}^{N} (-2\sigma_i(X)\sigma_i(Y) + \sigma_i(X)^2) + \sum_{i=N+1}^{l} (-2\sigma_i(X)\sigma_i(Y) + \sigma_i(X)^2 + 2\tau \sigma_i(X)) \right). \tag{8}
\]

Since Eq. (8) consists of simple quadratic equations for each \( \sigma_i(X) \) independently, it is trivial to show that the minimum of Eq. (8) is obtained at \( \hat{D}_X = \text{diag}(\hat{\sigma}(X)) \) by derivative in feasible domain as the first-order optimality condition, where \( \hat{\sigma}_i(X) \) is defined as

\[
\hat{\sigma}_i(X) = \begin{cases} 
\sigma_i(Y), & \text{if } i < N+1, \\
\max (\sigma_i(Y) - \tau, 0), & \text{otherwise}.
\end{cases} \tag{9}
\]

Consequently, the solution of Eq. (4) is \( X^* = U_Y \hat{D}_X V_Y^T \). This result is identical to the PSVT operator. Now, we know that a feasible solution \( X^* = U_Y(D_{Y1} + S_r[D_{Y2}])V_Y^T \) exists, which is a non-increasing operator for the objective function.
Figure 2. Success ratio for synthetic data with varying numbers of columns rows $m$ (a-d). Comparison between RPCA and ours for the rank-1 case (a,b), and for the rank-3 case (c,d). The Y-axis represents the corruption ratio $r \in [0, 0.4]$. The X-axis represents the log scale row size $\log_{10} m \in [\log_{10} 100, \log_{10} 12800]$ in (a-d). We fixed $n = 16$, and varied $m$ and $r$. The color magnitude represents success ratio $[0,1]$. Our method can successfully recover $A$ and $E$ up to 15% of severe corruption for the rank-1 case (b), and leads to more robust results than RPCA despite 5% higher corruption for the rank-3 case (d).

Figure 3. Comparison for the Rank Deficiency of $\hat{A}$. (a,b) shows average ratio $\frac{\sigma_3}{\sigma_1}$ (similar to the inverse value of the condition number) of the estimated low–rank matrix $\hat{A}$ for the rank-3 case (a typical example of photometric stereo), obtained by RPCA (a) and our approach (b). If the ratio is lower than 0.01, we consider that the recovered matrix has a rank lower than $N$. The red regions mean that the rank of the recovered matrix is lower than the constraint rank. Our method shows that the rank constraint is satisfied for almost all of the regions, while RPCA has some regions whose rank is lower than the target rank. We measure the average ratio from 50 trials.
Figure 4. Sculpture Garden sequence [7]. Comparison results of low-rank matrix and sparse error between RPCA and our approach. (a) Samples out of 5 input multi-exposed images. (b,d) Recovered low-rank $A$ obtained by RPCA and the proposed approach, respectively. (c,e) Recovered sparse error $E$ obtained by RPCA and the proposed approach, respectively. The degenerated low-rank matrix of (b) yields dense non-zero entries in (c) that should be originally sparse. On the other hand, our proposed method shows well modeled background scene and successfully detects outlier regions in (d).

Figure 5. HDR results of Sculpture Garden sequence [7] obtained by RPCA (a) and our approach (b). (b) shows the visibly better result than (a) due to better-modeled background scene as shown in Fig. 4-(d).
Figure 6. Outlier rejection results for Photometric Stereo. (a) Three sampled input images out of five. (b-c) low-rank and sparse image from Wu et al. [4]. (d-e) low-rank and sparse image from ours. We considered 5 input images and corrupted 2 of them by painted artifacts to mimic outliers.

Table 1. Singular values of photometric stereo input for \( n = 5 \) in Fig. 6 and Fig. 7-(b,c).

<table>
<thead>
<tr>
<th></th>
<th>( \sigma_1 )</th>
<th>( \sigma_2 )</th>
<th>( \sigma_3 )</th>
<th>( \sigma_4 )</th>
<th>( \sigma_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_i(O_R) )</td>
<td>137.3020</td>
<td>23.5079</td>
<td>19.5820</td>
<td>15.6315</td>
<td>12.8782</td>
</tr>
<tr>
<td>( \sigma_i(O_G) )</td>
<td>137.8808</td>
<td>23.5852</td>
<td>19.5483</td>
<td>15.5554</td>
<td>12.9935</td>
</tr>
<tr>
<td>( \sigma_i(O_B) )</td>
<td>138.6918</td>
<td>23.3590</td>
<td>19.7637</td>
<td>15.5297</td>
<td>12.9103</td>
</tr>
<tr>
<td>RPCA ( \sigma_i(\hat{A}_R) ) in Fig. 7-(b)</td>
<td>124.5052</td>
<td>7.4036</td>
<td>\textbf{0.0001}</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>RPCA ( \sigma_i(\hat{A}_G) ) in Fig. 7-(b)</td>
<td>125.7143</td>
<td>7.2594</td>
<td>\textbf{0.0001}</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>RPCA ( \sigma_i(\hat{A}_B) ) in Fig. 7-(b)</td>
<td>126.7097</td>
<td>7.3432</td>
<td>\textbf{0.0157}</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Ours ( \sigma_i(\hat{A}_R) ) in Fig. 7-(c)</td>
<td>137.9819</td>
<td>22.9652</td>
<td>16.3428</td>
<td>1.5540</td>
<td>0.1773</td>
</tr>
<tr>
<td>Ours ( \sigma_i(\hat{A}_G) ) in Fig. 7-(c)</td>
<td>139.1555</td>
<td>23.0132</td>
<td>16.3316</td>
<td>1.5898</td>
<td>0.1549</td>
</tr>
<tr>
<td>Ours ( \sigma_i(\hat{A}_B) ) in Fig. 7-(c)</td>
<td>140.0027</td>
<td>22.6999</td>
<td>16.2750</td>
<td>1.7833</td>
<td>0.2152</td>
</tr>
</tbody>
</table>
Figure 7. Results of the photometric normal estimation and depth by the least square method (LSQ), Wu et al. [4] and ours. Each top row represents the normal estimation and each bottom row shows the reconstructed depth from the estimated normals. Corrupted inputs by texture are included in 2 out of $n = 5$ inputs, and 4 out of $n = 10$ inputs. (b,c) show the estimated normal from the outlier-removed images by RPCA and our method. Wu et al. [4] returns planar surface normal when the rank of input matrix is lower than 3 due to the lack of observations (refer to Table 1). When more input images are available, RPCA begins to return detail preserved results, as shown in Fig. 7-(e). On the other hand, our method consistently provides robust results for both limited and sufficient observations, as shown in Fig. 7-(c,f).
Figure 8. Results of the batch image alignment from RASL [3] and ours for $n = 10$. Similarity transformation is used as geometric transformation model $g$.

Figure 9. Results of the batch image alignment from RASL [3] and ours for $n = 100$. Similarity transformation is also used as geometric transformation model $g$. Fig. 8 and 9 show that our method performs identically to conventional RPCA with many samples. It is consistent results with synthetic result in Sec. 4.1 of the main paper.
Figure 10. Aligning planar surfaces despite occlusions by RASL [3] and ours with \( n = 4 \) images. Affine transformation is used as geometric transformation model \( g \). Our method can correctly detect the outliers as well as robustly align the images even if the geometric model has more degree of freedom.

Figure 11. Detail comparisons of the averages of the aligned results \( O \circ g \) (Middle) and the average of the recovered low-rank component \( A \) (Right) between RASL [3] and ours in Fig. 10.
Figure 12. Aligning planar surfaces despite occlusions by RASL [3] (top row) and ours (bottom row) for $n = 16$. (a) Original images from 16 views. (b) Aligned result $O \circ g$. (c) Recovered low-rank component $A$. (c) Occlusions $E$. Affine transformation is used as geometric transformation model $g$. Our method can correctly detect the outliers as well as robustly align the images even if the geometric model has more degree of freedom.

Figure 13. Detail comparisons of the averages of the recovered low-rank component $A$ between RASL [3] and Ours in Fig. 12. The result of our method shows more clean image than RASL. It means that our method decomposes outliers better than RASL.
References


