Supplementary Material for the Paper:

Exploiting Symmetry and/or Manhattan Properties for 3D Object Structure Estimation from Single and Multiple Images

Yuan Gao
Tencent AI Lab, Shenzhen, China
ethanygao@tencent.com

Alan L. Yuille
Johns Hopkins University, Baltimore, MD
UCLA, Los Angeles, CA
alan.yuille@jhu.edu

This supplementary material contains more details of the Paper "Exploiting Symmetry and/or Manhattan Properties for 3D Object Structure Estimation from Single and Multiple Images":

1. In Section S1, we show our results on the imperfect annotations (*i.e.* the imperfect symmetric pairs).
2. In Section S2, we discuss how to impose Manhattan constraints as a regularization term for the Symmetric Rigid Structure from Motion (Sym-RSfM) on multiple images.
3. In Section S3, the minimization of the energy function w.r.t. the camera projection matrix R_n under orthogonality constraints is detailed.
4. In Section S4, we provide a way to recover a 2×2 matrix B from BB^T up to a rotation ambiguity.

S1. Experimental Results on The Imperfect Annotations

In this section, we investigate what happens if the keypoints are not perfectly annotated. This is important to check because our method depends on keypoint pairs therefore may be sensitive to errors in keypoint location, which will inevitably arise when we use features detectors, *e.g.* deep nets [1], to detect the keypoints.

To simulate this, we add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to the 2D annotations and re-do the experiments. The standard deviation is set to $\sigma = sd_{max}$, where d_{max} is the longest distance between all the keypoints (*e.g.* for *car*, it is the distance between the left/right front wheel to the right/left back roof top). We have tested for different s by: 0.03, 0.05, 0.07. Other experimental settings are the same as them in the main text, *i.e.* images with more than 5 visible keypoints are used.

The mean rotation errors and the mean shape errors for *car* with $s = 0.03, 0.05, 0.07$ are shown in Tables S1 and S2. Each result value is obtained by averaging 10 repetitions. The results in Tables S1 and S2 show that the performances of all the methods decrease in general with the increase in the noise level. Nonetheless, our methods still outperform our counterparts with the noisy annotations (*i.e.* the imperfectly labeled annotations).

In summary, this section certificates our method is robust to imperfect annotations for practical use.

S2. Imposing Manhattan Constraints to The Symmetric Rigid Structure from Motion

The energy function *w.r.t* R_n, S (when the missing points are fixed) is:

$$\begin{aligned}
\mathcal{Q}(R_n, S) &= -\sum_n \ln P(Y_n, Y_n^\dagger | R_n, S) = -\sum_n \big(\ln P(Y_n|R_n, S) - \ln P(Y_n^\dagger|R_n, S)\big) \\
&= \sum_n ||Y_n - R_n S||_2^2 + \sum_n ||Y_n^\dagger - R_n \mathcal{A} S||_2^2 + Constant \\
&= \sum_n ||\mathbb{Y} - G_n \mathbb{S}||_2^2 + \sum_n ||\mathbb{Y}^\dagger - G_n \mathcal{A}_P \mathbb{S}||_2^2 + Constant
\end{aligned} \tag{S1}$$

	σ = **0.03** d_{max}										σ = **0.05** d_{max}				
	I	II	III	IV	V	VI	VII	VIII	IX	X	I	II	III	IV	V
RSfM	0.57	0.69	0.55	1.09	0.68	1.58	0.67	1.45	0.97	0.37	0.59	0.74	0.53	1.11	0.68
CSF (S)	0.95	1.22	1.10	1.01	1.01	1.03	1.16	**1.02**	1.35	1.04	0.95	1.29	1.05	1.02	1.00
CSF (R)	0.95	1.27	1.07	1.05	1.05	0.98	0.95	1.04	1.05	1.16	0.95	1.24	1.07	1.04	1.07
Sym-RSfM	**0.37**	**0.42**	**0.34**	**0.46**	**0.32**	**0.26**	**0.32**	1.06	**0.25**	**0.12**	**0.39**	**0.45**	**0.31**	**0.48**	**0.31**

	σ = **0.05** d_{max} *(cont.)*					σ = **0.05** d_{max}									
	VI	VII	VIII	IX	X	I	II	III	IV	V	VI	VII	VIII	IX	X
RSfM	1.54	0.69	1.49	0.99	0.40	0.59	0.72	0.56	1.09	0.64	1.54	0.67	1.55	0.97	0.38
CSF (S)	1.04	1.09	**1.03**	1.37	1.02	1.04	1.25	1.05	1.16	0.98	1.03	0.88	1.02	1.20	1.04
CSF (R)	0.97	0.93	**1.03**	0.89	1.16	0.95	1.22	1.06	1.04	0.88	0.96	1.00	1.04	0.91	1.05
Sym-RSfM	**0.28**	**0.33**	1.13	**0.28**	**0.13**	**0.38**	**0.37**	**0.30**	**0.49**	**0.28**	**0.35**	**0.33**	**1.01**	**0.26**	**0.13**

Table S1. The mean *rotation* errors for *car* with imperfect annotations. The noise is Gaussian $\mathcal{N}(0, \sigma^2)$ with $\sigma = sd_{max}$, where we choose $s = 0.03, 0.05, 0.07$ and d_{max} is the longest distance between all the keypoints (*i.e.* the left/right front wheel to the right/left back roof top). The Roman numerals denotes the index of the subtype. Each result value is obtained by averaging 10 repetitions.

	σ = **0.03** d_{max}										σ = **0.05** d_{max}				
	I	II	III	IV	V	VI	VII	VIII	IX	X	I	II	III	IV	V
RSfM	1.48	1.48	1.33	1.37	1.45	1.39	1.21	1.82	1.23	1.07	1.48	1.47	1.34	1.39	1.44
CSF (S)	1.06	2.36	1.14	**0.88**	1.37	1.17	**0.77**	1.13	2.00	0.98	1.10	1.22	1.15	**0.90**	1.34
CSF (R)	1.33	0.99	1.02	1.15	1.18	1.25	0.87	**0.90**	1.42	1.12	1.33	0.98	1.01	1.15	1.16
Sym-RSfM	**1.04**	**0.95**	**0.96**	1.08	**0.90**	**1.12**	0.81	1.80	**0.88**	**0.66**	**1.04**	**0.95**	**0.95**	1.08	**0.90**

	σ = **0.05** d_{max} *(cont.)*					σ = **0.05** d_{max}									
	VI	VII	VIII	IX	X	I	II	III	IV	V	VI	VII	VIII	IX	X
RSfM	1.43	1.20	1.81	1.21	1.08	1.48	1.46	1.33	1.36	1.43	1.48	1.21	1.79	1.22	1.08
CSF (S)	**1.20**	1.04	1.11	1.96	0.97	3.56	1.28	1.15	1.19	1.38	1.22	2.08	1.06	2.47	1.03
CSF (R)	1.25	0.87	**0.88**	1.41	1.11	1.31	**0.92**	0.99	1.15	1.17	1.25	0.87	**0.88**	1.55	1.11
Sym-RSfM	1.25	**0.81**	1.71	**0.88**	**0.67**	**1.05**	0.95	**0.96**	**1.09**	**0.89**	**1.07**	**0.82**	1.80	**0.88**	**0.68**

Table S2. The mean *shape* errors for *car* with imperfect annotations. Other parameters are the same as Table S1.

where $\mathbb{S} \in \mathbb{R}^{3P\times 1}, \mathbb{Y}_n \in \mathbb{R}^{2P\times 1}, \mathbb{Y}_n^\dagger \in \mathbb{R}^{2P\times 1}$ are vectorized $S, Y_n, Y_n^\dagger$, respectively. $\mathcal{A} = \text{diag}[-1, 1, 1]$ is a matrix operator which negates the first row of its right-multiplied matrix. $G_n = I_P \otimes R_n$ and $\mathcal{A}_P = I_P \otimes \mathcal{A}$. $I_P \in \mathbb{R}^{P\times P}$ is an identity matrix. The last equation is obtained by vectorizing the above one by $\text{vec}(AXB^T) = (B \otimes A)\text{vec}(X)$, and $\otimes$ denotes Kronecker product.

If another Manhattan direction is available *e.g.* if we have $S_i = [S_i^x, S_i^{y_i}, S_i^z]^T, S_j = [S_i^x, S_i^{y_j}, S_i^z]^T$ along y-axis, *i.e.* S_i, S_j have the same coordinates on x- and z-axes[1]. Then, we have another constraints: $S_i - S_j = [0, S_i^{y_i} - S_i^{y_j}, 0]^T$. Let the matrix operator $\mathcal{X} = \text{diag}([1,0,1])$ which selects the first and the third row of its right-multiplied matrix, we can rewrite Eq. (S1) by encoding the Manhattan constraints as regularization:

$$\mathcal{Q}(\mathbb{S}) = \sum_n^N \left(||\mathbb{Y} - G_n\mathbb{S}||_2^2 + ||\mathbb{Y}^\dagger - G_n\mathcal{A}_P\mathbb{S}||_2^2\right) + \lambda||\mathcal{X}(S_i - S_j)||_2^2. \tag{S2}$$

We can further rewrite the last term of Eq. (S2) based on $\mathbb{S}$ (instead S_i, S_j) by applying matrix operators on S and then vectorizing it. We define the matrix operator $\mathcal{Y} \in \mathbb{R}^{P\times 2}$ such that its i'th row in first column and j'th row in second column are equal to 1, and otherwise 0. Then, $S\mathcal{Y}$ selects the keypoints S_i, S_j (i'th and j'th columns of S) along the Manhattan direction y. We also use the matrix operator $\mathcal{Z} = [-1, 1]^T \in \mathbb{R}^{2\times 1}$ so that $S\mathcal{Y}\mathcal{Z}$ denotes $S_j - S_i$. Thus, Eq. (S2) can be written by:

$$\begin{aligned}\mathcal{Q}(S) &= \sum_n^N \left(||\mathbb{Y} - G_n\mathbb{S}||_2^2 + ||\mathbb{Y}^\dagger - G_n\mathcal{A}_P\mathbb{S}||_2^2\right) + \lambda||\mathcal{X}S\mathcal{Y}\mathcal{Z}||_2^2.\\ &= \sum_n^N \left(||\mathbb{Y} - G_n\mathbb{S}||_2^2 + ||\mathbb{Y}^\dagger - G_n\mathcal{A}_P\mathbb{S}||_2^2\right) + \lambda||(\mathcal{Z}^T\mathcal{Y}^T \otimes \mathcal{X})\mathbb{S}||_2^2.\end{aligned} \tag{S3}$$

[1] S_i and S_j are not necessary to be symmetric along y-axis.

Taking derivative the energy function in Eq. (S2) *w.r.t* $\mathbb{S}$ and equating it to 0, the update of $\mathbb{S}$ under additional Manhattan constraint becomes:

$$\mathbb{S} = \left(\sum_{n=1}^{N}(G_n^T G_n + \mathcal{A}_P^T G_n^T G_n \mathcal{A}_P^T) + \lambda(\mathcal{Z}^T\mathcal{Y}^T \otimes \mathcal{X})^T(\mathcal{Z}^T\mathcal{Y}^T \otimes \mathcal{X})\right)^{-1}\left(\sum_{n=1}^{N}(G_n^T \mathbb{Y}_n + \mathcal{A}_P^T G_n^T \mathbb{Y}_n^\dagger)\right). \qquad \text{(S4)}$$

Note that it is easy to generalize the matrix operator $\mathcal{Y}, \mathcal{Z}$ if we have M points along y-axis. Specifically, we only need to rewrite $\mathcal{Y} \in \mathbb{R}^{P\times M}$ that selects the M keypoints (M columns of S) along the Manhattan direction y, and $\mathcal{Z} = [-\mathbf{1}_{M-1}, I_{M-1}]^T \in \mathbb{R}^{M\times M-1}$ that makes each of the M keypoints minus the first keypoint.

S3. Update Camera Projection Matrix R_n Under Orthogonality Constraints

In this section, we describe how to do coordinate descent for the camera parameters R_n given the 3D structure S is fixed. In order to minimize Eq. (S1) *w.r.t.* R_n under the nonlinear orthogonality constraints $R_n R_n^T = I$, we follow an alternative approach used in [2] to parameterize R_n as a complete 3×3 rotation matrix Q_n and update the incremental rotation on Q_n instead, *i.e.* $Q_n^{new} = e^{\xi} Q_n$.

Here, the first and second rows of Q_n is the same as R_n, and the third row of Q_n is obtained by the cross product of its first and second rows. The relationship of Q_n and R_n can be revealed by a matrix operator $\mathcal{M}$:

$$R_n = \mathcal{M} Q_n, \qquad \mathcal{M} = \begin{bmatrix} 1,0,0 \\ 0,1,0 \end{bmatrix}. \qquad \text{(S5)}$$

Note that the incremental rotation e^{ξ} can be further approximated by its first order Taylor Series, *i.e.* $e^{\xi} \approx I + \xi$. Finally, we have:

$$R_n^{new}(\xi) = \mathcal{M}(I + \xi)Q_n. \qquad \text{(S6)}$$

Therefore, setting $\partial \mathcal{Q}/\partial R_n = 0$, then replace R_n by Q_n using Eq. (S6) and vectorize it, we have:

$$R_n = \mathcal{M} e^{\xi} Q_n \approx \mathcal{M}(I + \xi)Q_n \qquad \text{and} \qquad \text{vec}(\xi) = \alpha^{+}\beta,$$

$$\alpha = \left(\sum_{p=1}^{P}(S_p S_p^T + \mathcal{A} S_p S_p^T \mathcal{A}^T)^T Q_n^T\right) \otimes \mathcal{M},$$

$$\beta = \text{vec}\left(\sum_{p=1}^{P}(Y_{n,p} S_p^T + Y_{n,p}^\dagger S_p^T \mathcal{A}^T) - Q_n \sum_{p=1}^{P}(S_p S_p^T + \mathcal{A} S_p S_p^T \mathcal{A}^T)\right), \qquad \text{(S7)}$$

where the subscript p means the pth keypoint, α^{+} means the pseudo inverse matrix of α, $\otimes$ denotes Kronecker product.

S4. Recover Matrix B from BB^T

In the following, we describe a method for recovering 2×2 matrix B from BB^T up to a 2D rotation. Let $B = \begin{bmatrix} b_1\cos\theta, & b_1\sin\theta \\ b_3\cos\phi, & b_3\sin\phi \end{bmatrix}$, thus:

$$BB^T = \begin{bmatrix} (b_1)^2, & b_1 b_3\cos(\theta-\phi) \\ b_1 b_3 \cos(\theta - \phi), & (b_3)^2 \end{bmatrix} = \begin{bmatrix} bb_1, & bb_2 \\ bb_2, & bb_3 \end{bmatrix}. \qquad \text{(S8)}$$

If we assume $\phi = 0$ (due to the “fake” rotation ambiguity on yz-plane) and $b_1, b_3 \geq 0$ (due to the “fake” direction ambiguities of $y-$ and $z-$ axes), all the unknown parameters (*i.e.* b_1, b_3, θ) can be calculated by:

$$b_1 = \sqrt{bb_1}, \qquad b_3 = \sqrt{bb_3}, \qquad \theta = \text{arc}\cos(\frac{bb_2}{b_1 b_3}) + \phi, \qquad \phi = 0. \qquad \text{(S9)}$$

References

[1] X. Chen and A. L. Yuille. Articulated pose estimation by a graphical model with image dependent pairwise relations. In *NIPS*, pages 1736–1744, 2014. 1

[2] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:878–892, 2008. 3