Supplementary Material for the Paper:

Symmetric Non-Rigid Structure from Motion for Category-Specific Object Structure Estimation

Yuan Gao[1⋆] and Alan L. Yuille[2,3]

[1] City University of Hong Kong [2] UCLA [3] Johns Hopkins University

This supplementary material contains more details of *Symmetric Non-Rigid Structure from Motion for Category-Specific Object Structure Estimation*:

1. In Sect. 1, we show our results using *median* shape and rotation errors.
2. In Sect. 2, we give the update of each parameters in **M-step** of *Sym-EM-PPCA*.
3. In Sect. 3, the minimization of the energy function of *Sym-PriorFree w.r.t.* the camera projection matrix R_n under orthogonality constrains is detailed.

1 Experimental Results using Median Shape and Rotation Errors

In this section, the **median** *shape* and *rotation* errors are reported. We used the same parameters as them in the main text, *i.e.* we use 3 deformation bases and set λ in Sym-EM-PPCA as 1. The same conclusions can be made by the median errors as them made by the mean errors in the main text.

2 M-Step of Sym-EM-PPCA

This step is to maximize the complete (joint) log-likelihood $P(\mathbb{Y}_n, \mathbb{Y}_n^\dagger, \mathbf{V}|z_n; G_n, \bar{\mathbb{S}}, \mathbf{V}^\dagger, \mathbb{T})$. The complete log-likelihood $Q(\theta)$ is:

$$\begin{aligned}\mathcal{Q}(\theta) =& P(\mathbb{Y}_n, \mathbb{Y}_n^\dagger, \mathbf{V}|z_n; G_n, \bar{\mathbb{S}}, \mathbf{V}^\dagger, \mathbb{T}) \\ =& P(\mathbb{Y}_n|z_n; G_n, \bar{\mathbb{S}}, \mathbf{V}, \mathbb{T}) P(\mathbb{Y}_n^\dagger|z_n; G_n, \bar{\mathbb{S}}, \mathbf{V}^\dagger, \mathbb{T}) P(\mathbf{V}|\mathbf{V}^\dagger) \\ =& 2PN\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_n E_{z_n}||\mathbb{Y}_n - G_n(\bar{\mathbb{S}} + \mathbf{V}z_n) - \mathbb{T}_n||^2 \\ &+ \frac{1}{2\sigma^2}\sum_n E_{z_n}||\mathbb{Y}_n^\dagger - G_n(\mathcal{A}_P\bar{\mathbb{S}} + \mathbf{V}^\dagger z_n) - \mathbb{T}_n||^2 + \lambda||\mathbf{V} - \mathcal{A}_P\mathbf{V}^\dagger||^2 \\ &\text{s. t.} \qquad R_n R_n^T = I, \end{aligned} \tag{1}$$

where $\theta = \{G_n, \bar{\mathbb{S}}, \mathbf{V}, \mathbf{V}^\dagger, \mathbb{T}_n, \sigma^2\}$. $\mathbb{Y}_n \in \mathbb{R}^{2P\times 1}, \bar{\mathbb{S}} \in \mathbb{R}^{3P\times 1}$, and $\mathbb{T}_n \in \mathbb{R}^{2P\times 1}$ are the stacked vectors of 2D keypoints, 3D mean structure and translations. $G_n =$

⋆ This work has been done when Yuan Gao was a visiting student in UCLA.

	aeroplane								bus						
	I	II	III	IV	V	VI	VII	mRE	I	II	III	IV	V	VI	mRE
EP	0.29	0.50	0.39	0.41	0.5	0.46	**0.38**	0.24	0.34	0.28	0.54	0.49	0.92	0.80	0.15
PF	0.94	1.14	1.21	1.16	1.13	1.05	1.17	0.41	1.50	1.32	1.79	1.53	2.36	1.56	0.33
Sym-EP	**0.25**	**0.53**	**0.34**	**0.36**	**0.43**	**0.42**	0.40	**0.23**	**0.25**	**0.24**	**0.32**	**0.27**	**0.65**	**0.38**	**0.12**
Sym-PF	0.47	0.70	0.79	0.52	0.60	0.46	0.65	0.33	1.95	2.07	1.75	1.52	1.46	1.01	1.34

	car											sofa		
	I	II	III	IV	V	VI	VII	VIII	IX	X	mRE	I	II	III
EP	1.05	1.01	1.07	1.01	0.99	1.07	0.94	1.44	0.95	0.84	0.37	1.99	1.86	2.01
PF	1.71	1.66	1.73	1.79	1.60	1.72	1.75	1.48	1.70	1.36	0.81	1.76	1.55	1.32
Sym-EP	**0.93**	**0.88**	**1.02**	**0.99**	**0.87**	**0.99**	**0.83**	**1.40**	**0.91**	**0.67**	**0.32**	**1.12**	**0.81**	**1.07**
Sym-PF	1.73	1.26	1.81	1.43	1.62	1.54	1.32	1.70	1.47	1.16	0.67	1.14	1.08	1.18

	sofa				train					tv				
	IV	V	VI	mRE	I	II	III	IV	mRE	I	II	III	IV	mRE
EP	1.98	2.34	1.81	0.76	1.04	0.45	0.49	0.42	0.71	**0.36**	**0.40**	**0.40**	0.34	**0.26**
PF	2.03	2.51	1.53	1.42	1.81	0.38	0.48	0.30	0.87	0.56	1.01	0.96	0.61	0.66
Sym-EP	1.11	1.79	**0.88**	**0.24**	**0.91**	0.44	**0.41**	**0.25**	**0.55**	0.53	0.52	0.47	0.60	0.28
Sym-PF	**0.87**	**0.90**	0.92	0.76	1.67	**0.32**	0.52	0.47	0.91	0.48	0.91	0.98	**0.13**	0.76

Table 1. The **median** *shape* and *rotation* errors for *aeroplane, bus, car, sofa, train, tv*. Since Pascal3D+ [1] provides one 3D model for each subtype, thus the mean shape errors are reported according to each subtype. While the camera projection matrices are available for all the images, thus we report the mean rotation errors for each category. The Roman numerals indicates the index of subtypes for the mean shape error, and mRE is short for the mean rotation error. EP, PF, Sym-EP, Sym-PF are short for EM-PPCA [2], PriorFree [3,4], Sym-EM-PPCA, Sym-PriorFree, respectively.

$I_P \otimes c_n R_n$, in which c_n is the scale parameter for weak perspective projection, $\mathbf{V} = [\mathbb{V}_1, ..., \mathbb{V}_K] \in \mathbb{R}^{3P \times K}$ is the grouped K deformation bases, $z_n \in \mathbb{R}^{K \times 1}$ is the coefficient of the K bases. $\mathcal{A} = I_P \otimes \mathcal{A}$, and $\mathcal{A} = \text{diag}([-1, 1, 1])$ is a matrix operator which negates the first row of its right matrix.

We first update the shape parameters $\bar{\mathbb{S}}, \mathbf{V}, \mathbf{V}^\dagger$ by maximize the log-likelihood $\mathcal{Q}$. Since these 3 parameters are related to each other in their derivations, thus they should be updated jointly by setting the 3 derivations to 0. According to Eq. (1), we have:

$$\begin{bmatrix} A^*, & (B^*)^T, & C^* \\ B^*, & D^* + 2\lambda\sigma^2 I_{3PK}, & -I_K \otimes 2\lambda\sigma^2 \mathcal{A}_P \\ B^*\mathcal{A}_P, & -I_K \otimes 2\lambda\sigma^2 \mathcal{A}_P^T, & D^* + I_K \otimes 2\lambda\sigma^2 \mathcal{A}_P^T \mathcal{A}_P \end{bmatrix} \begin{bmatrix} \mathbb{S} \\ \text{vec}(\mathbf{V}) \\ \text{vec}(\mathbf{V}^\dagger) \end{bmatrix}$$
$$= \begin{bmatrix} \text{vec}\left(\sum_n G_n^T(\mathbb{Y} - \mathbb{T}_n) + \mathcal{A}_P^T G_n^T(\mathbb{Y}^\dagger - \mathbb{T}_n)\right) \\ \text{vec}\left(\sum_n G_n^T(\mathbb{Y} - \mathbb{T}_n)\mu_n^T\right) \\ \text{vec}\left(\sum_n G_n^T(\mathbb{Y}^\dagger - \mathbb{T}_n)\mu_n^T\right), \end{bmatrix}, \tag{2}$$

where we have:

$$A^* = \sum_n G_n^T G_n + \mathcal{A}_P^T G_n^T G_n \mathcal{A}_P, \qquad B^* = \sum_n \mu_n \otimes G_n^T G_n$$
$$C^* = \sum_n \mu_n^T \otimes \mathcal{A}_P^T G_n^T G_n, \qquad D^* = \sum_n \phi_n^T \otimes G_n^T G_n. \tag{3}$$

The camera parameters t_n, c_n, R_n and the variance of the noise σ^2 can be updated similarly as Bregler's method [2]. We first replace some parameters to make the equation to be homogeneous:

$$\begin{aligned} \tilde{\mathbf{V}} &= [\mathbb{S}, \mathbf{V}], \qquad \tilde{\mathbf{V}}^\dagger = [\mathcal{A}_P\mathbb{S}, \mathbf{V}^\dagger], \\ \tilde{\mu_n} &= [1, \mu_n^T]^T, \qquad \tilde{\phi_n} = \begin{bmatrix} 1 & \mu_n^T \\ \mu_n & \phi_n \end{bmatrix} \end{aligned} \tag{4}$$

Then the estimations of new σ^2, t_n, c_n are:

$$\begin{aligned} \sigma^2 =& \frac{1}{4PN}\sum_n \big(||\mathbb{Y}_n - \mathbb{T}_n||^2 + ||\mathbb{Y}_n^\dagger - \mathbb{T}_n||^2 \\ & -2(\mathbb{Y}_n - \mathbb{T}_n)G_n\tilde{\mathbf{V}}\tilde{\mu_n} - 2(\mathbb{Y}_n^\dagger - \mathbb{T}_n)G_n\tilde{\mathbf{V}}^\dagger\tilde{\mu_n} \\ & +\mathrm{tr}(\tilde{\mathbf{V}}^T G_n^T G_n \tilde{\mathbf{V}}\tilde{\phi_n}) + \mathrm{tr}(\tilde{\mathbf{V}}^{\dagger T} G_n^T G_n \tilde{\mathbf{V}}^\dagger\tilde{\phi_n})\big) \end{aligned} \tag{5}$$

$$t_n = \frac{1}{2P}\sum_{p=1}^{P}(\mathbb{Y}_{n,p} - c_n R_n \tilde{\mathbf{V}}_p\tilde{\mu_n} + \mathbb{Y}_{n,p}^\dagger - c_n R_n \tilde{\mathbf{V}}_p^\dagger\tilde{\mu_n}) \tag{6}$$

$$c_n = \frac{\sum_{p=1}^{P}\left(\tilde{\mu_n}^T\tilde{\mathbf{V}}_p^T R_n^T(\mathbb{Y}_{n,p} - t_n) + \tilde{\mu_n}^T\tilde{\mathbf{V}}_p^{\dagger T} R_n^T(\mathbb{Y}_{n,p}^\dagger - t_n)\right)}{\sum_{p=1}^{P}\mathrm{tr}(\tilde{\mathbf{V}}_p^T R_n^T R_n \tilde{\mathbf{V}}_p\tilde{\phi_n} + \tilde{\mathbf{V}}_p^{\dagger T} R_n^T R_n \tilde{\mathbf{V}}_p^\dagger\tilde{\phi_n})} \tag{7}$$

Since R_n is subject to a nonlinear orthonormality constraint and cannot be updated in closed form, 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 $\mathcal{M}$:

$$R_n = \mathcal{M}Q_n, \qquad \mathcal{M} = \begin{bmatrix} 1,0,0 \\ 0,1,0 \end{bmatrix}. \tag{8}$$

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:

$$R_n^{new}(\xi) = \mathcal{M}(I + \xi)Q_n. \tag{9}$$

Therefore, setting $\partial\mathcal{Q}/\partial R_n = 0$, then replace R_n by Q_n using Eq. (9) and vectorize it, we have:

$$R_n = \mathcal{M}e^{\xi}Q_n \approx \mathcal{M}(I+\xi)Q_n \quad \text{and} \quad \text{vec}(\xi) = \alpha^{+}\beta \tag{10}$$

$$\alpha = \left(c_n^2 \sum_{p=1}^{P} (\tilde{\mathbf{V}}_p^T \tilde{\phi_n} \tilde{\mathbf{V}}_p + \tilde{\mathbf{V}}_p^{\dagger T} \tilde{\phi_n} \tilde{\mathbf{V}}_p^{\dagger})^T Q_n^T\right) \otimes \mathcal{M}; \tag{11}$$

$$\begin{aligned}\beta = \text{vec}\Bigg(c_n \sum_{p=1}^{P} \Big((\mathbb{Y}_{n,p} - t_n)\tilde{\mu_n}^T \tilde{\mathbf{V}}_p^T + (\mathbb{Y}_{n,p}^{\dagger} - t_n)\tilde{\mu_n}^T \tilde{\mathbf{V}}_p^{\dagger T}\Big) \\ - c_n^2 \mathcal{M} Q_n \sum_{p=1}^{P} (\tilde{\mathbf{V}}_p^T \tilde{\phi_n} \tilde{\mathbf{V}}_p + \tilde{\mathbf{V}}_p^{\dagger T} \tilde{\phi_n} \tilde{\mathbf{V}}_p^{\dagger})\Bigg)\end{aligned} \tag{12}$$

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

3 Update Camera Projection Matrix R_n Under Orthogonality Constrains for Sym-PriorFree

In the Sym-PriorFree method, the energy function *w.r.t.* R_n is:

$$Q(R_n) = \sum_n ||Y_n - R_n S_n||_2^2 + \sum_n ||Y_n^{\dagger} - R_n \mathcal{A} S_n||_2^2, \tag{13}$$

where $Y_n, Y_n^{\dagger} \in \mathbb{R}^{2\times P}, S_n \in \mathbb{R}^{3\times P}$ are the 2D symmetric keypoint pairs and the 3D structure of image n. $R_n \in \mathbb{R}^{2\times 3}$ is the camera projection. $\mathcal{A} = \text{diag}([-1, 1, 1])$ is a matrix operator which negates the first row of its right matrix.

We use the same updating procedures as used in the previous section to update R_n for Sym-PriorFree method. Specifically, we firstly parameterize R_n as a Q_n by Eq. (8). Then, Q_n can be updated by $Q_n^{new} = e^{\xi}Q_n \approx (I+\xi)Q_n$.

Therefore, setting the derivation of $\partial\mathcal{Q}/\partial R_n = 0$, then replace R_n by Q_n using Eq. (9) and vectorize it, we have:

$$\begin{aligned} &R_n = \mathcal{M}e^{\xi}Q_n \approx \mathcal{M}(I+\xi)Q_n \quad \text{and} \quad \text{vec}(\xi) = \alpha^{+}\beta, \\ &\alpha = \left(\sum_{p=1}^{P} (S_{n,p}S_{n,p}^T + \mathcal{A}S_{n,p}S_{n,p}^T\mathcal{A}^T)^T Q_n^T\right) \otimes \mathcal{M}, \\ &\beta = \text{vec}\left(\sum_{p=1}^{P} (Y_{n,p}S_{n,p}^T + Y_{n,p}^{\dagger}S_{n,p}^T\mathcal{A}^T) - Q_n \sum_{p=1}^{P} (S_{n,p}S_{n,p}^T + \mathcal{A}S_{n,p}S_{n,p}^T\mathcal{A}^T)\right), \end{aligned} \tag{14}$$

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

References

1. Xiang, Y., Mottaghi, R., Savarese, S.: Beyond pascal: A benchmark for 3d object detection in the wild. In: WACV. (2014)
2. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 878–892
3. Dai, Y., Li, H., He, M.: A simple prior-free method for non-rigid structure-from-motion factorization. In: CVPR. (2012)
4. Dai, Y., Li, H., He, M.: A simple prior-free method for non-rigid structure-from-motion factorization. International Journal of Computer Vision **107** (2014) 101–122