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]John Hopkins University

Motivation

Many objects, especially that made by human, have intrinsic symmetry property, e.g. cars, aeroplanes, etc.

Our goal is to is to investigate how symmetry can improve NRSfM.

Note that we assume the deformation between different objects is non-rigid and symmetric, e.g. from sedan to hatchback cars. While the object itself is rigid (also symmetric).

Contributions

By exploring symmetry, we proposed two NRSfM methods which gain significant improvement over their baseline counterparts.

Both of our methods can deal with occlusions, which are updated iteratively.

Sym-EM-PPCA: this is an extension of EM-PPCA [1]. It imposes symmetric constraints on both 3D structure and deformation bases. Sym-Rigid-SfM [5] is used to initialize Sym-EM-PPCA with hard symmetric constraints on the 3D structure.

Sym-PriorFree: it extends the matrix factorization methods of [2, 3] by symmetry, to initialize a coordinate descent algorithm for occlusions.

Ambiguities in NRSfM

The key idea of non-rigid SfM is to represent the non-rigid deformations of objects in terms of **a linear combination of bases**:

$$\mathbf{Y} = \mathbf{RS} \quad \text{and} \quad \mathbf{S} = \mathbf{Vz}, \quad \mathbf{RR}^T = I \quad \Rightarrow \quad \mathbf{Y} = \mathbf{RVz} \quad s.t. \quad \mathbf{RR}^T = I$$

Ambiguities 1: (between **R** and **S**, will be solved by $\mathbf{RR}^{\mathrm{T}} = I$)

$$\mathbf{R} \leftarrow \mathbf{RA}_1 \text{ and } \mathbf{S} \leftarrow \mathbf{A}_1^{-1}\mathbf{S} \;\Rightarrow\; \mathbf{Y} = \mathbf{RS} = \mathbf{RA}_1\mathbf{A}_1^{-1}\mathbf{S}$$

Ambiguities 2: (between **RV** and **z**, essentially "gauge freedom" as in [4])

$$\mathbf{z} \leftarrow \mathbf{A}_2\mathbf{z} \text{ and } \mathbf{V} \leftarrow \mathbf{VA}_2^{-1} \;\Rightarrow\; \mathbf{Y} = \mathbf{RVz} = \mathbf{RVA}_2\mathbf{A}_2^{-1}\mathbf{z}$$
$$\mathbf{z} \leftarrow \mathbf{z} + \alpha\mathbf{w} \text{ and } \mathbf{RV} \perp \mathbf{w} \;\Rightarrow\; \mathbf{Y} = \mathbf{RVz} = \mathbf{RV}(\mathbf{z} + \alpha\mathbf{w})$$

Different ways to deal with Ambiguities 2 leads to original EM-PPCA [1] and PriorFree [2,3] methods.

Sym-EM-PPCA method

It uses a conjugate (Gaussian) prior on z to deal with Ambiguities 2.

A mean shape is used to model the 3D structure of *n*'th image:

$$S_n = \bar{S} + \mathbf{V}z_n$$

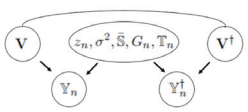
Let superscript † denote symmetry, and $\mathbb{Y}_n, \bar{\mathbb{S}}, \mathbb{T}_n$ denote the stacked vectors of 2D keypoints, 3D mean structure, and translations. $G_n = I_P \otimes c_n R_n$, in which c_n is the scale parameter, R_n is the camera projection. **V** is the deformation bases, z_n is the coefficient. After exploring symmetry, we have:

$$P(\underset{2P\times 1}{\mathbb{Y}_n}|\underset{K\times 1}{z_n}, \underset{2P\times 3P}{G_n}, \underset{3P\times 1}{\bar{\mathbb{S}}}, \underset{3P\times K}{\mathbf{V}}, \underset{2P\times 1}{\mathbb{T}}) = \mathcal{N}(G_n(\bar{\mathbb{S}} + \mathbf{V}z_n) + \mathbb{T}_n, \sigma^2 I)$$
$$P(\mathbb{Y}_n^\dagger|z_n, G_n, \bar{\mathbb{S}}, \mathbf{V}^\dagger, \mathbb{T}) = \mathcal{N}(G_n(\mathcal{A}_P\bar{\mathbb{S}} + \mathbf{V}^\dagger z_n) + \mathbb{T}_n, \sigma^2 I)$$

Then, we maximize marginal probability $P(\mathbb{Y}_n, \mathbb{Y}_n^\dagger|G_n, \bar{\mathbb{S}}, \mathbf{V}, \mathbf{V}^\dagger, \mathbb{T})$ by EM, where z_n is the latent variable with conjugate (Gaussian) prior.

E-step: maximize the posterior probability of z_n:

$$\begin{aligned}
&P(z_n|\mathbb{Y}_n, \mathbb{Y}_n^\dagger; \sigma^2, \bar{\mathbb{S}}, \mathbf{V}, \mathbf{V}^\dagger, G_n, \mathbb{T}_n)\\
\sim &P(z_n, \mathbb{Y}_n, \mathbb{Y}_n^\dagger|\sigma^2, \bar{\mathbb{S}}, \mathbf{V}, \mathbf{V}^\dagger, G_n, \mathbb{T}_n)\\
= &P(\mathbb{Y}_n|z_n; \sigma^2, \bar{\mathbb{S}}, \mathbf{V}, G_n, \mathbb{T}_n)P(\mathbb{Y}_n^\dagger|z_n; \sigma^2, \bar{\mathbb{S}}, \mathbf{V}^\dagger, G_n, \mathbb{T}_n)P(z_n)\\
= &\mathcal{N}(z_n|\mu_n, \Sigma_n)
\end{aligned}$$



M-step: maximize the joint likelihood with fixed z_n:

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

Initialization: R_n, $\bar{S}$ and the occluded points $Y_{n,p}, Y_{n,p}^\dagger$ can be initialized by [5], c_n is initialized as 1. **V** and $\mathbf{V}^\dagger$ are initialized by the PCA on the 2D KPs

Sym-PriorFree method

As revealed in [4], Ambiguities 2 are in fact "gauge freedom", which facilitates a prior free matrix factorization method [2,3].

The original prior free SfM model is [2,3]:

$$\underset{2N\times P}{\mathbf{Y}} = \underset{2N\times 3N}{\mathbf{R}}\ \underset{3N\times P}{\mathbf{S}} = \begin{bmatrix} R_1S_1 \\ \vdots \\ R_NS_N \end{bmatrix} = \begin{bmatrix} z_{11}R_1, & ..., & z_{1K}R_1 \\ \vdots & \ddots & \vdots \\ z_{N1}R_N, & ..., & z_{NK}R_N \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_K \end{bmatrix} = \underset{2N\times 3K}{\mathbf{\Pi}}\ \underset{3K\times P}{\mathbf{V}}$$

where **Y, R, S** are stacked 2D keypoints, projection matrix and 3D structures. z_{nk} is the coefficient associated with *k*'th basis of image *n*. **V** is the deformation bases, and $\mathbf{\Pi} = \mathbf{R}(\mathbf{z} \otimes I_3)$.

The matrix factorization is conducted on **Y** *w.r.t* **Π** and **V**:

$$\min ||\mathbf{Y} - \mathbf{\Pi V}||_2^2 \qquad \text{(because rank(}\mathbf{Y}\text{)} = \min(2N, 3K, P)\text{.)}$$

After imposing symmetry, we have: (superscript † denotes symmetry)

$$\mathcal{Q}(\mathbf{R}, \mathbf{S}) = ||\mathbf{Y} - \mathbf{RS}||_2^2 + ||\mathbf{Y}^\dagger - \mathbf{RS}^\dagger||_2^2 = ||\mathbf{Y} - \mathbf{\Pi V}||_2^2 + ||\mathbf{Y}^\dagger - \mathbf{\Pi V}^\dagger||_2^2$$

Note that we cannot do SVD on **Y** and $\mathbf{Y}^\dagger$ separately, because the two energies are dependent. In other words, doing SVD separately cannot ensure to provide the same **Π** and the symmetric **V**, $\mathbf{V}^\dagger$.

Therefore, we further explore symmetry to decouple the dependency. Assume the object is symmetric along *x*-axis in world coordinate system (we can always rotate the world coordinate system to ensure that), we have:

$$\mathbf{L} = \frac{\mathbf{Y} - \mathbf{Y}^\dagger}{2} = \mathbf{R}\begin{bmatrix} x_1^1, & ..., & x_1^P \\ 0, & ..., & 0 \\ 0, & ..., & 0 \\ \vdots & ..., & \vdots \end{bmatrix} = \underset{2N\times K}{\hat{\mathbf{\Pi}}^1}\ \underset{K\times P}{\hat{\mathbf{V}}_x}, \qquad \mathbf{M} = \frac{\mathbf{Y} + \mathbf{Y}^\dagger}{2} = \mathbf{R}\begin{bmatrix} 0, & ..., & 0 \\ y_1^1, & ..., & y_1^P \\ z_1^1, & ..., & z_1^P \\ \vdots & ..., & \vdots \end{bmatrix} = \underset{2N\times 2K}{\hat{\mathbf{\Pi}}^2}\ \underset{2K\times P}{\hat{\mathbf{V}}_{yz}}$$

Therefore, the decoupled energies are:

$$\mathcal{Q}(\mathbf{\Pi}, \mathbf{V}) = ||\mathbf{L} - \hat{\mathbf{\Pi}}^1\hat{\mathbf{V}}_x||_2^2 + ||\mathbf{M} - \hat{\mathbf{\Pi}}^2\hat{\mathbf{V}}_{yz}||_2^2$$

The SVD (initial) on **L** and **M** gives us $\hat{\mathbf{\Pi}}^1, \hat{\mathbf{V}}_x, \hat{\mathbf{\Pi}}^2, \hat{\mathbf{V}}_{yz}$, which is up to ambiguities H^1, H^2 to the true $\mathbf{\Pi}^1, \mathbf{\Pi}^2, \mathbf{V}_x, \mathbf{V}_{yz}$:

$$\mathbf{L} = \mathbf{\Pi}^1\mathbf{V}_x = \hat{\mathbf{\Pi}}^1H^1(H^1)^{-1}\hat{\mathbf{V}}_x, \qquad \mathbf{M} = \mathbf{\Pi}^2\mathbf{V}_{yz} = \hat{\mathbf{\Pi}}^2H^2(H^2)^{-1}\hat{\mathbf{V}}_{yz}$$

These ambiguities can be solved by applying orthonormality constraints $\mathbf{RR}^{\mathrm{T}} = I$, on $\hat{\mathbf{\Pi}}$, where $\hat{\mathbf{\Pi}} = [\hat{\mathbf{\Pi}}^1, \hat{\mathbf{\Pi}}^2]$. The (under-determined) solutions lie in the intersection of three subspaces [2,3], i.e. any solution that satisfies these three subspaces gives the same 3D reconstruction results, only up to different rotations and scales [4]. Please refer to our paper for the details for solving H^1 and H^2.

After solving H^1 and H^2, we can recover **R** by normalizing each row of **Π** to have unit L2 norm, then **S** can be estimated accordingly by minimizing the reconstruction error with low-rank prior on **S**.

Results

Please see our paper for more results! (The last two rows are ours)

Pascal3D+ dataset [6] is used. The Roman numerals are the mean shape error for each subtype (e.g. sedan car, because we only have groundtruth shape for subtypes), mRE is the mean rotation error for each category.

	car											sofa			
	I	II	III	IV	V	VI	VII	VIII	IX	X	mRE	I	II	III
EP	1.10	1.01	1.09	1.05	1.03	1.07	0.99	1.46	1.00	0.85	0.39	2.00	1.87	2.03
PF	1.76	1.67	1.76	1.77	1.65	1.79	1.67	1.57	1.70	1.42	0.86	1.71	1.41	1.46
Sym-EP	**0.99**	**0.89**	**1.05**	**1.02**	**0.92**	**1.00**	**0.89**	**1.39**	**0.95**	**0.68**	**0.34**	**1.18**	**0.81**	**1.08**
Sym-PF	1.74	1.41	1.70	1.48	1.69	1.58	1.43	1.69	1.52	1.30	0.79	1.33	1.15	1.36

We also investigate the noise annotations, as it affect the symmetry. Experiments on *aeroplane* with different Gaussian noise σ show that our methods is robust to noise..

	$\sigma = 0.03\ d_{max}$								$\sigma = 0.05\ d_{max}$			
	I	II	III	IV	V	VI	VII	mRE	I	II	III	IV
EP	0.34	0.59	0.49	0.45	0.54	0.55	**0.45**	0.33	0.37	0.58	0.51	0.47
PF	0.92	1.01	1.05	1.06	1.13	1.03	1.06	0.52	0.93	1.04	1.05	1.08
Sym-EP	**0.34**	**0.54**	**0.47**	**0.44**	**0.52**	**0.55**	0.46	**0.32**	**0.35**	**0.54**	**0.47**	**0.43**
Sym-PF	0.79	0.93	1.01	0.93	0.91	0.79	0.94	0.60	0.83	0.99	1.09	0.98

Reference

1. Torresani, L., Hertzmann, A., Bregler,. IEEE T PAMI, 2008.
2. Dai, Y., Li, H., He, M.. In: CVPR, 2012.
3. Dai, Y., Li, H., He, M.. IJCV, 2014.
4. Akhter, I., Sheikh, Y., Khan, S. In: CVPR, 2009.
5. Gao, Y., Yuille, A.L. arXiv preprint arXiv:1607.07129, 2016.
6. Xiang, Y., Mottaghi, R., Savarese. In: WACV, 2014.