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


Yuan Gao[1], Alan L. Yuille[2,3]

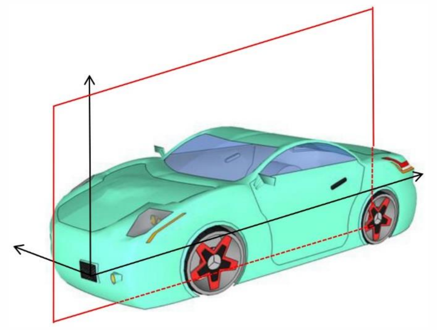
1. Tencent AI Lab 2. UCLA
3. Johns Hopkins University

ethan.y.gao@gmail.com
alan.yuille@jhu.edu


July 21-26 HONOLULU 2017

## Motivation

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



Our goal is to investigate how symmetry and Manhattan structures can improve 3D object structure reconstruction.

## Contributions

**1. Single image 3D reconstruction:**

Manhattan $\longrightarrow$ Camera Matrix
Symmetry
} 3D Structure

However, Manhattan structure may be hard to observe from a single image due to occlusion

**2. Multiple images 3D reconstruction:**

Symmetric Rigid Structure from Motion

This method can deal with occlusion, which are updated iteratively.

By exploring symmetry, Sym-RSfM method achieved state-of-the-art performance on Pascal3D+ dataset.

## Single-image reconstruction

Let 3D KPs $S_1 - S_2 = [x, 0, 0]^T$

$$Y_1 - Y_2 = R(S_1 - S_2) = \begin{bmatrix} r_{11}, & r_{12}, & r_{13} \\ r_{21}, & r_{22}, & r_{23} \end{bmatrix} \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} r_{11}x \\ r_{21}x \end{bmatrix}.$$

Therefore, we have $r_{11}/r_{21}$. Similarly, by exploring the other two Manhattan axes, we have $r_{12}/r_{22}$, $r_{13}/r_{23}$

Imposing Orthogonality Constrains $RR^T = I$

$$\boxed{\text{Known} \begin{bmatrix} 1, & 1, & 1 \\ \mu_1^2, & \mu_2^2, & \mu_3^2 \\ \mu_1, & \mu_2, & \mu_3 \end{bmatrix}} \begin{bmatrix} r_{11}^2 \\ r_{12}^2 \\ r_{13}^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

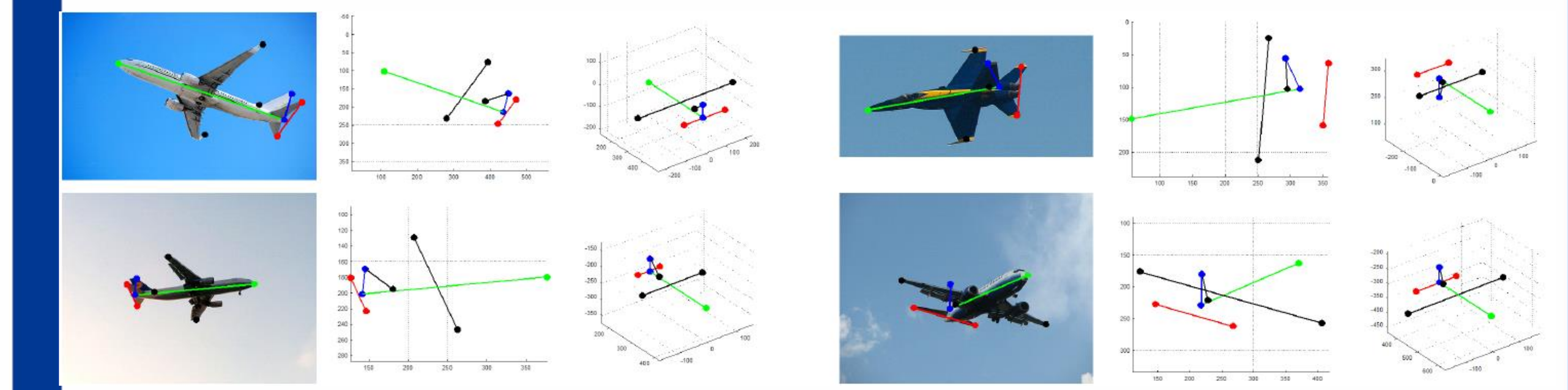
$$\mu_1 = r_{11}/r_{21}, \quad \mu_2 = r_{12}/r_{22}, \quad \mu_3 = r_{13}/r_{23}$$

Then, assume symmetry is along the x-axis:

$$L = \frac{Y - Y^\dagger}{2} = R \begin{bmatrix} x_1, & \dots, & x_P \\ 0, & \dots, & 0 \\ 0, & \dots, & 0 \end{bmatrix} = \begin{bmatrix} r_{11}x_1, & \dots, & r_{11}x_P \\ r_{21}x_1, & \dots, & r_{21}x_P \end{bmatrix},$$

$$M = \frac{Y + Y^\dagger}{2} = R \begin{bmatrix} 0, & \dots, & 0 \\ y_1, & \dots, & y_P \\ z_1, & \dots, & z_P \end{bmatrix} = \begin{bmatrix} r_{12}y_1 + r_{13}z_1, & \dots, & r_{12}y_P + r_{13}z_P \\ r_{22}y_1 + r_{23}z_1, & \dots, & r_{22}y_P + r_{23}z_P \end{bmatrix}.$$

$Y = R\,S \Rightarrow S = R^{-1}\,Y$ (well-defined)



## Multiple-image reconstruction

**Input**: Symmetric 2D keypoint pairs for multiple images for several rigid-deformed objects (e.g. sedan cars of different brand).

**Output**: The Common 3D structure of all the input images $S$, and the camera matrix for each image, $R_n$

Let $Y_n$, $Y_n^\dagger \in \mathbb{R}^{2\times P}$ are symmetric pairs of image $n$:

$$Y_n = R_n S + N_n \qquad Y_n^\dagger = R_n S^\dagger + N_n$$

$$\mathcal{Q}(R_n, S) = \sum_n ||Y_n - R_n S||_2^2 + \sum_n ||Y_n^\dagger - R_n S^\dagger||_2^2. \quad (1)$$

However, Eq. (1) cannot be solve by SVD (as widely used in other SfM methods) on the two terms separately, because:

$$||\underbrace{\mathbf{Y}}_{2N\times P} - \underbrace{\mathbf{R}}_{2N\times 3}\underbrace{S}_{3\times P}||_2^2 + ||\underbrace{\mathbf{Y}^\dagger}_{2N\times P} - \underbrace{\mathbf{R}}_{2N\times 3}\underbrace{S^\dagger}_{3\times P}||_2^2$$

Dependent!!

SVD? ✖

Our solution is decomposing it into two independent terms to enable SVD.

Assume symmetry is along x-axis:

$$\mathbf{L} = \frac{\mathbf{Y} - \mathbf{Y}^\dagger}{2} = \mathbf{R} \begin{bmatrix} x_1, & \dots, & x_P \\ 0, & \dots, & 0 \\ 0, & \dots, & 0 \end{bmatrix} = \underbrace{\mathbf{R}^1}_{N\times 1}\underbrace{S_x}_{1\times P}, \quad \mathbf{M} = \frac{\mathbf{Y} + \mathbf{Y}^\dagger}{2} = \mathbf{R} \begin{bmatrix} 0, & \dots, & 0 \\ y_1, & \dots, & y_P \\ z_1, & \dots, & z_P \end{bmatrix} = \underbrace{\mathbf{R}^2}_{N\times 2}\underbrace{S_{yz}}_{2\times P}.$$

$$\Rightarrow \quad Q(\mathbf{R}, S) = ||\mathbf{L} - \mathbf{R}^1 S_x||_2^2 + ||\mathbf{M} - \mathbf{R}^2 S_{yz}||_2^2, \qquad \mathbf{R} = [\mathbf{R}^1, \mathbf{R}^2]$$

Independent!!

$$\Rightarrow \quad \mathbf{L} = \mathbf{R}^1 S_x = \hat{\mathbf{R}}^1 \underbrace{\lambda\lambda^{-1}}_{\text{Ambiguities!}} \hat{S}_x, \qquad \mathbf{M} = \mathbf{R}^2 S_{yz} = \hat{\mathbf{R}}^2 \underbrace{BB^{-1}}_{\text{Ambiguities!}} \hat{S}_{yz}$$

The ambiguities can be solved by Orthogonality Constrains $\mathbf{RR^T = I}$, where $\mathbf{R} = [\hat{\mathbf{R}}^1\lambda,\ \hat{\mathbf{R}}^2 B]$.

**Final Solution:** $\mathbf{R} = \hat{\mathbf{R}} \begin{bmatrix} \lambda, & \mathbf{0} \\ \mathbf{0}, & B \end{bmatrix}, \qquad S = \begin{bmatrix} \lambda, & \mathbf{0} \\ \mathbf{0}, & B \end{bmatrix}^{-1} \hat{S}.$

Shape (upper table) and Rotation (lower table) errors:

| | Shape error for car | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI | VII | VIII | IX | X |
| RSfM | 0.58 | 0.71 | 0.54 | 1.10 | 0.67 | 1.51 | 0.67 | 1.41 | 0.97 | 0.37 |
| CSF (S) | 0.95 | 1.22 | 1.06 | 1.12 | 1.00 | 1.03 | 1.13 | **1.04** | 1.33 | 1.03 |
| CSF (R) | 0.95 | 1.32 | 1.08 | 1.06 | 1.09 | 0.98 | 1.22 | 1.05 | 1.29 | 1.17 |
| Sym-RSfM | **0.36** | **0.43** | **0.30** | **0.43** | **0.31** | **0.26** | **0.32** | **1.04** | **0.25** | **0.16** |

| | Rotation error for car | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI | VII | VIII | IX | X |
| RSfM | 1.48 | 1.49 | 1.33 | 1.38 | 1.45 | 1.39 | 1.21 | 1.81 | 1.22 | 1.07 |
| CSF (S) | 1.06 | 2.33 | 1.15 | 1.17 | 1.36 | 1.17 | 1.03 | 1.10 | 2.03 | 0.99 |
| CSF (R) | 1.34 | 1.07 | 1.03 | 1.16 | 1.18 | 1.26 | 0.88 | **0.90** | 1.65 | 1.13 |
| Sym-RSfM | **1.03** | **0.96** | **0.95** | **1.07** | **0.89** | **1.00** | **0.81** | 1.66 | **0.88** | **0.71** |

## Reference


1. C. Tomasi and T. Kanade, IJCV, 1992.
2. P. Gotardo and A. Martinez, IEEE T PAMI, 2011.
3. Y. Gao and A. L. Yuille, ECCV, 2016.
4. Y. Xiang, R. Mottaghi, S, Savarese, WACV, 2014.