Robust 3D Car Shape Estimation from Landmarks in Monocular Image

Yanan Miao
miaoyn12@mails.tsinghua.edu.cn
Xiaoming Tao
taoxm@mail.tsinghua.edu.cn
Jianhua Lu
lhh-dee@mail.tsinghua.edu.cn

Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Electronic Engineering
Tsinghua University
Beijing, P. R. China

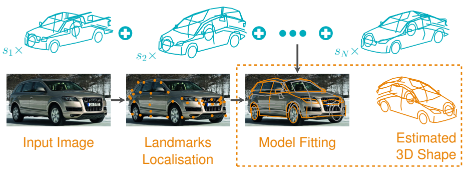

Figure 1: The framework for 3D shape estimation. **Top**: A series of prior 3D shape basis [2]. **Bottom**: The shape estimation procedure for a given input image.

Estimation of the 3D shape of a object from monocular image is an under-determined problem, which becomes harder when the observations are severely contaminated. In this paper, we propose a robust model to estimate 3D shape $\mathbf{X}$ from 2D landmarks $\mathbf{x} \in \mathbb{R}^{2\times p}$ with unknown camera pose M. The 3D shape of the object is assumed as a linear combination of predefined shape basis $\{X_i\}_{i=1}^{N} \in \mathbb{R}^{3\times p}$ weighted by $\mathbf{s} = [s_1, \ldots, s_N]^\mathsf{T} \in \mathbb{R}^N$. To estimate $\mathbf{s}$ and M, we fit the model by minimizing the error between the observations $\mathbf{x}$ and the projected model points $\mathsf{M}\mathbf{X}$ (as shown in Figure 1).

Model. To address the outliers in the observed 2D points, which result from the complex background and illumination conditions, we propose a robust 3D shape estimation model. We explicitly model the outliers with an additional sparse error term $E \in \mathbb{R}^{2\times p}$. Thus, the robust model is then formulated as

$$\min_{\mathbf{s},\mathsf{M}} \frac{1}{2}\|\mathbf{x}-\mathbf{t}-\mathsf{M}\mathbf{X}-E\|_F^2 + \underbrace{\lambda\|\mathbf{s}\|_1 + \eta\|E\|_1}_{\text{non-smooth}}$$
$$s.t. \ \ \overbrace{\mathsf{M}\mathsf{M}^\mathsf{T} = \mathbf{I}_2}^{\text{non-convex}}, \mathbf{X} = \sum_{i=1}^{N} s_i X_i + \mu \quad (1)$$

where $\mathbf{t} = [t_x, t_y]^\mathsf{T} \cdot \mathbf{1}_{1\times p}$ is the translation, and λ, η are the regularization parameters, and μ is the mean shape. The objective function in (1) is non-convex and non-smooth constrained on Stiefel manifold, where the coupling of the unknown shape representation coefficients $\mathbf{s}$ and camera pose M makes it more difficult to be solved.

Method. We propose an efficient numerical algorithm based on Alternative Direction Method of Multipliers (ADMM) [1] to solve this problem. With an auxiliary variable $V \in \mathbb{R}^{2\times 3}$ introduced, the augmented Lagrangian is,

$$\begin{aligned}\mathcal{L}_{\mathsf{M},V,\mathbf{s},E,\mathbf{t},\Lambda} &= \tfrac{1}{2}\|\mathbf{x}-\mathbf{t}-\mathsf{M}\mathbf{X}-E\|_F^2 + \lambda\|\mathbf{s}\|_1 \\ &+\eta\|E\|_1 + <\Lambda, \mathsf{M}-V> + \tfrac{\tau}{2}\|\mathsf{M}-V\|_F^2 \\ s.t. \ \ &\mathsf{M} = V, VV^\mathsf{T} = \mathbf{I}_2, \mathbf{X} = \textstyle\sum_{i=1}^{N} s_i X_i + \mu\end{aligned}$$

where Λ is the multiplier and τ is penalty parameter. We update each block with all the others fixed. Based on some analysis on non-convex optimization of ADMM [3], we set the orthogonality constraints into the smooth sub-problem (V-minimization),

$$\min_V \{\|V - (\mathsf{M}^k + \Lambda^k/\tau^k)\|_F^2 : VV^\mathsf{T} = \mathbf{I}_2\}.$$

The closed-form solution is given by $V^{k+1} = U\mathbf{I}_{2\times 3}W^T$, where U and W satisfy $[U,S,W] = \mathcal{SVD}\left[\mathsf{M}^k + \Lambda^k/\tau^k\right]$. The other sub-problems can be easily solved. Both the optimization of M and $\mathbf{t}$ admit closed-form solutions. The updating of $\mathbf{s}$ is a *Lasso*-problem, and the sparse error pattern E can be efficiently solved by element-wise soft-thresholding. The convergences of ADMM with more than two blocks cannot be always guaranteed [1], and may be influenced by the update ordering. We set a fixed update ordering that can always lead convergence in our experiments.

[1] S. Boyd and etc. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[2] Y. Lin and etc. Jointly optimizing 3d model fitting and fine-grained classification. In *ECCV*, pages 466–480. Springer, 2014.

[3] Y. Zhang. Recent advances in alternating direction methods: Practice and theory. In *IPAM workshop*, 2010.