

Supplementary Material

Motion from Structure (MfS): Searching for 3D Objects in Cluttered Point Trajectories

1. Estimating M

Let $\widehat{\mathbf{W}} \in \mathbb{R}^{2F \times 4}$ contains the selected 4 tracks and $\widehat{\mathbf{S}} \in \mathbb{R}^{4 \times 4}$ contains homogeneous coordinates of the selected 4 vertices from the 3D model, where $\widehat{\mathbf{w}}_i$ matches to $\widehat{\mathbf{s}}_i$, $i = 1, \dots, 4$. To estimate $\mathbf{M} \in \mathbb{R}^{2F \times 4}$, we need to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{M}} \quad & \|\widehat{\mathbf{W}} - \mathbf{M}\widehat{\mathbf{S}}\|_F^2 \\ \text{subject to} \quad & \overline{\mathbf{M}}^f \overline{\mathbf{M}}^{f\top} = \mathbf{I}_2, f = 1, \dots, F. \end{aligned}$$

It can be seen that the above problem can be solved independently for each frame f :

$$\begin{aligned} \min_{\mathbf{M}^f} \quad & \|\widehat{\mathbf{W}}^f - \mathbf{M}^f \widehat{\mathbf{S}}\|_F^2 \\ \text{subject to} \quad & \overline{\mathbf{M}}^f \overline{\mathbf{M}}^{f\top} = \mathbf{I}_2, \end{aligned}$$

where $\widehat{\mathbf{W}}^f$ and \mathbf{M}^f contains rows of $\widehat{\mathbf{W}}$ and \mathbf{M} corresponding to frame f . Recall that $\mathbf{M}^f = [\alpha^f \overline{\mathbf{M}}, \mathbf{b}^f]$ with $\alpha^f \in \mathbb{R}$ and $\mathbf{b}^f \in \mathbb{R}^2$ being the scale and translation in frame f .

The above problem may look similar to orthogonal Procrustes analysis, but it is actually a different problem; The matrix $\overline{\mathbf{M}}$ belongs to the Stiefel manifold, not the orthogonal group. Several works have analysed and proposed solutions to the problem [1, 3]. In this work, we solve $\overline{\mathbf{M}}$ using a simple suboptimal approach of finding affine transformation between $\widehat{\mathbf{W}}^f$ and $\widehat{\mathbf{S}}$ then projecting the rotation part to Stiefel manifold. The algorithm is summarized in Alg. 1.

Algorithm 1 Estimating \mathbf{M}^f

Input: $\widehat{\mathbf{W}}^f \in \mathbb{R}^{2 \times 4}, \widehat{\mathbf{S}} \in \mathbb{R}^{4 \times 4}$

Output: $\mathbf{M}^f \in \mathbb{R}^{2 \times 4}$

- 1: Remove the fourth row of $\widehat{\mathbf{S}}$ that comprises of 1s
 - 2: Remove mean from $\widehat{\mathbf{W}}^f$ by $\widetilde{\mathbf{W}}^f := \widehat{\mathbf{W}}^f (\mathbf{I}_4 - \frac{1}{4} \mathbf{1}_4 \mathbf{1}_4^\top)$
 - 3: Remove mean from $\widehat{\mathbf{S}}$ by $\widetilde{\mathbf{S}} := \widehat{\mathbf{S}} (\mathbf{I}_4 - \frac{1}{4} \mathbf{1}_4 \mathbf{1}_4^\top)$
 - 4: Find affine tranformation \mathbf{A} by $\mathbf{A} := \widetilde{\mathbf{W}}^f \widetilde{\mathbf{S}}^\dagger$ where $\widetilde{\mathbf{S}}^\dagger$ is Moore-Penrose pseudoinverse of $\widetilde{\mathbf{S}}$
 - 5: Find thin SVD of \mathbf{A} by $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^\top$
 - 6: Compute $\overline{\mathbf{M}}^f$ by $\overline{\mathbf{M}}^f := \mathbf{U} \mathbf{V}^\top$
 - 7: Compute scale α^f by $\alpha^f := \frac{\text{tr}(\widetilde{\mathbf{S}}^\top \overline{\mathbf{M}}^{f\top} \overline{\mathbf{M}}^f \widetilde{\mathbf{W}}^f)}{\text{tr}(\widetilde{\mathbf{S}}^\top \overline{\mathbf{M}}^{f\top} \overline{\mathbf{M}}^f \widetilde{\mathbf{S}})}$
 - 8: Compute translation \mathbf{b}^f by $\mathbf{b}^f := \frac{1}{4} (\widehat{\mathbf{W}}^f - \overline{\mathbf{M}}^f \widehat{\mathbf{S}}) \mathbf{1}_4$
 - 9: Compose $\mathbf{M}^f := [\alpha^f \overline{\mathbf{M}}^f, \mathbf{b}^f]$
-

2. Additional results

Fig. 1 and 2 show additional results and comparison with baselines. For video with two 3D models, we ran the algorithms two times, each time with different model. For video with one 3D models but two alignments, we ran the algorithms twice with the tracks in $\text{span}(\mathbf{M})$ of the first run removed before the second run. It may be notice that All-R seems to be able to find the correct object (albeit incorrect orientation). This is because the cost function prefers alignment that includes all coherent tracks under a single motion, *i.e.* tracks of “full object”. Hence, in addition to ST selection and guided sampling, the cost function also plays a significant role in the success of our approach. We encourage the readers to view the supplementary video for these results.

In last video in Fig. 2, one alignment of All-R is on top of the other. For STGS-R, the alignment of the left van is flipped, but this can be prevented by incorporating orientation into the algorithm.

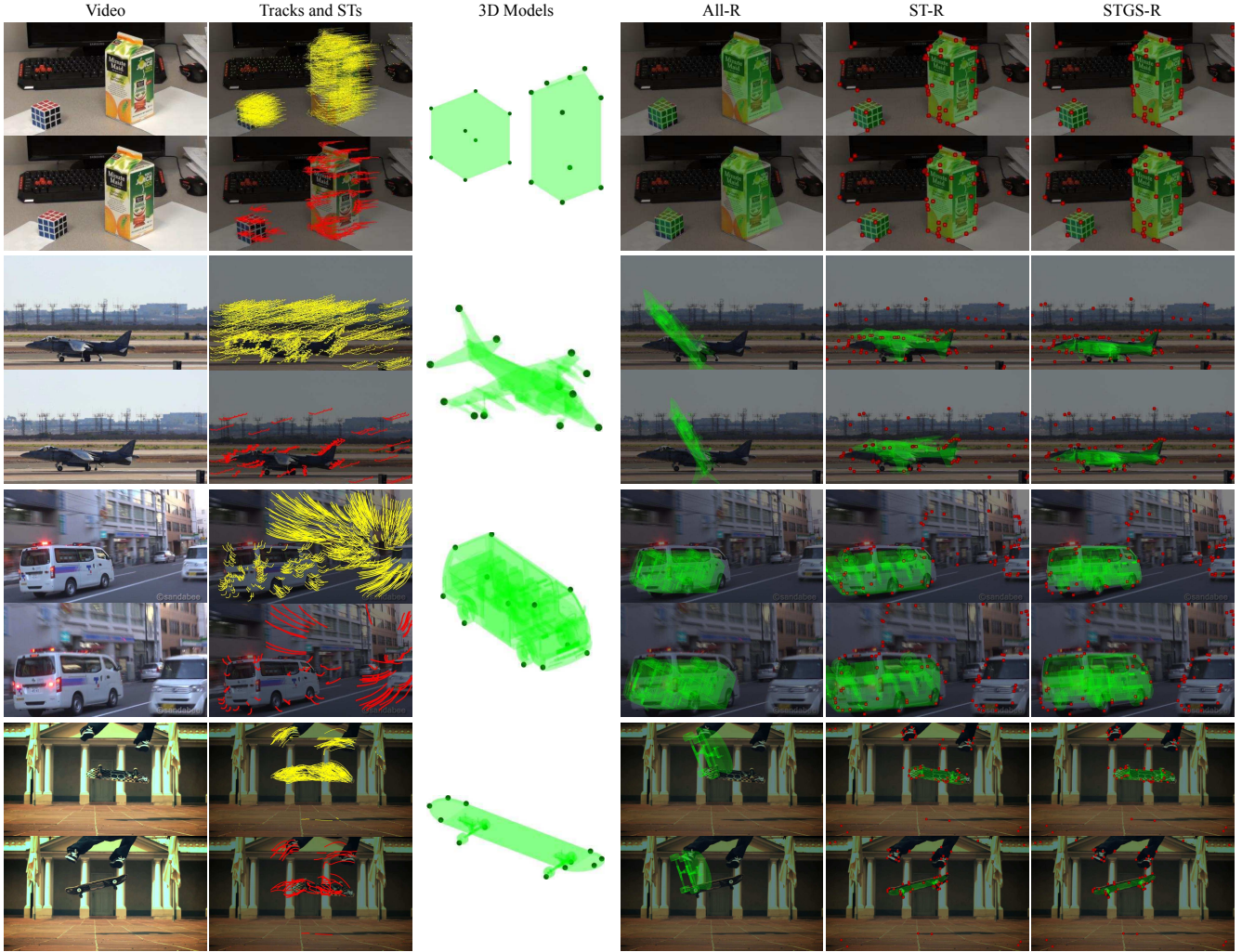


Figure 1. Additional results with comparison with baselines. For each pair of rows, the top shows the first frame while the bottom shows the last frame of each video (except Tracks and STs, and 3D Models).

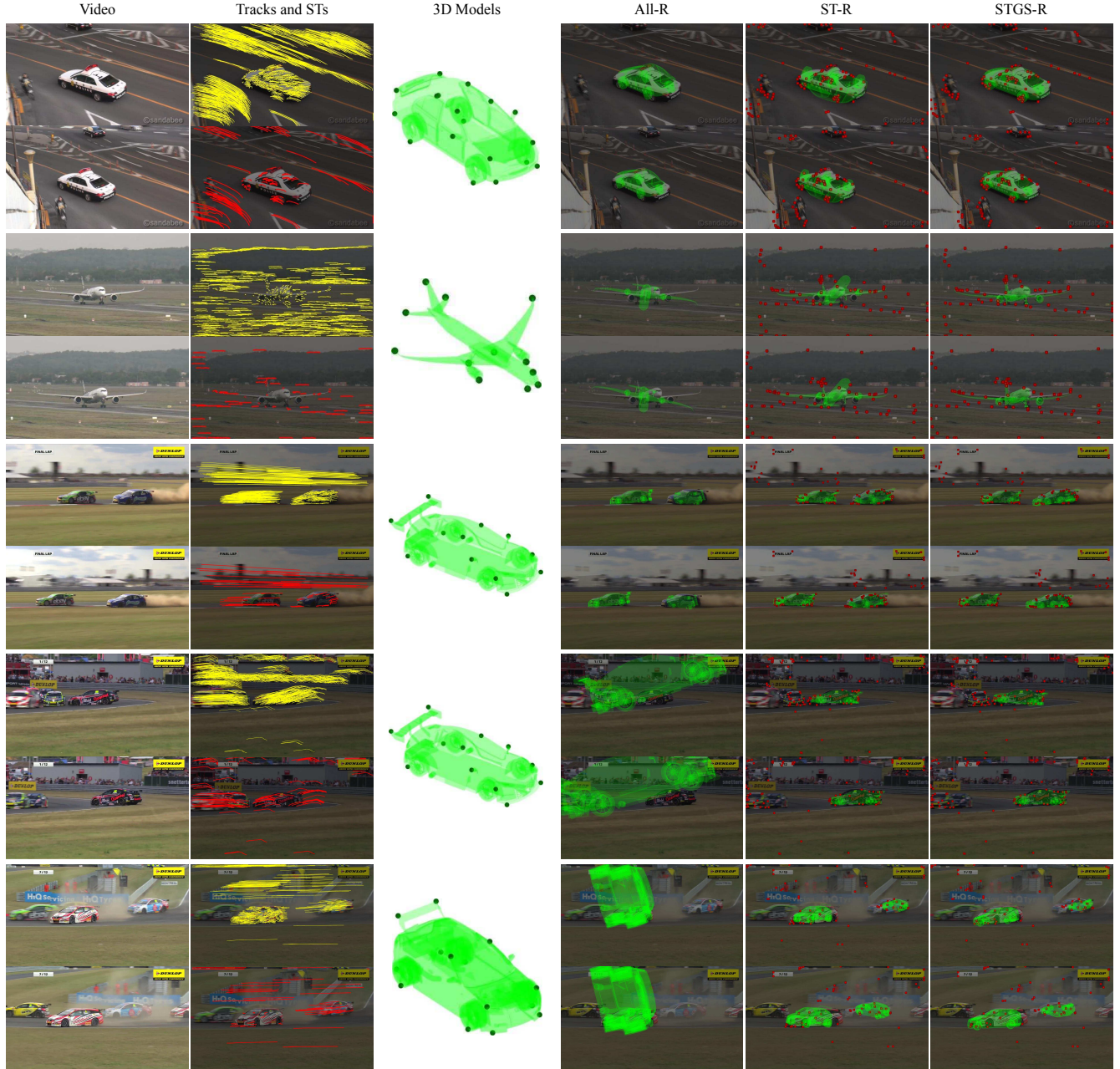


Figure 2. Additional results with comparison with baselines. For each pair of rows, the top shows the first frame while the bottom shows the last frame of each video (except Tracks and STs, and 3D Models).

3. Disadvantages of Using Motion Segmentation and Reconstruction for MfS

In this section, we provide a few examples to show that the pipeline involving motion segmentation, reconstruction, and 3D point cloud registration may fail to solve MfS problem. For this demonstration, we use sparse subspace clustering (SSC) [2] for motion segmentation, and factorization with orthogonal constraint in motion matrix [4] for reconstruction.

Fig. 3 shows an example of incorrect motion segmentation. Here, we show the cases when 3 and 4 are selected as numbers of clusters. It can be seen that the clusters are incorrect; either the van is clustered with other objects (3 clusters) or it is cut in half (4 clusters). This would lead to incorrect reconstruction and registration. On the other hand, even a good segmentation

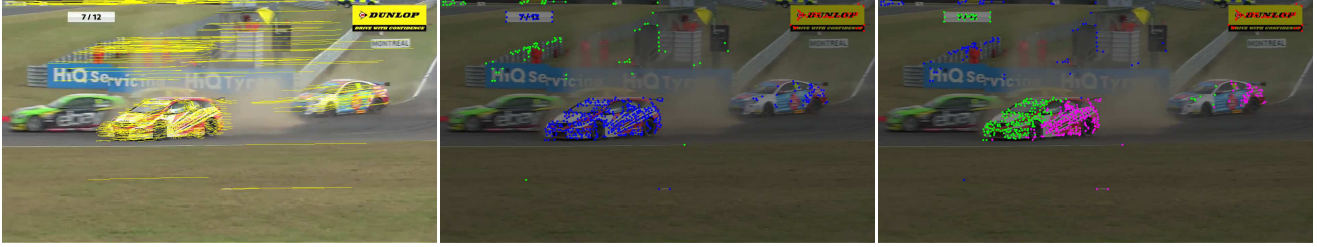


Figure 3. Incorrect segmentation results. (Left) Input tracks. Motion segmentation results with (Middle) 3 clusters and (Right) 4 clusters are shown in different colors.

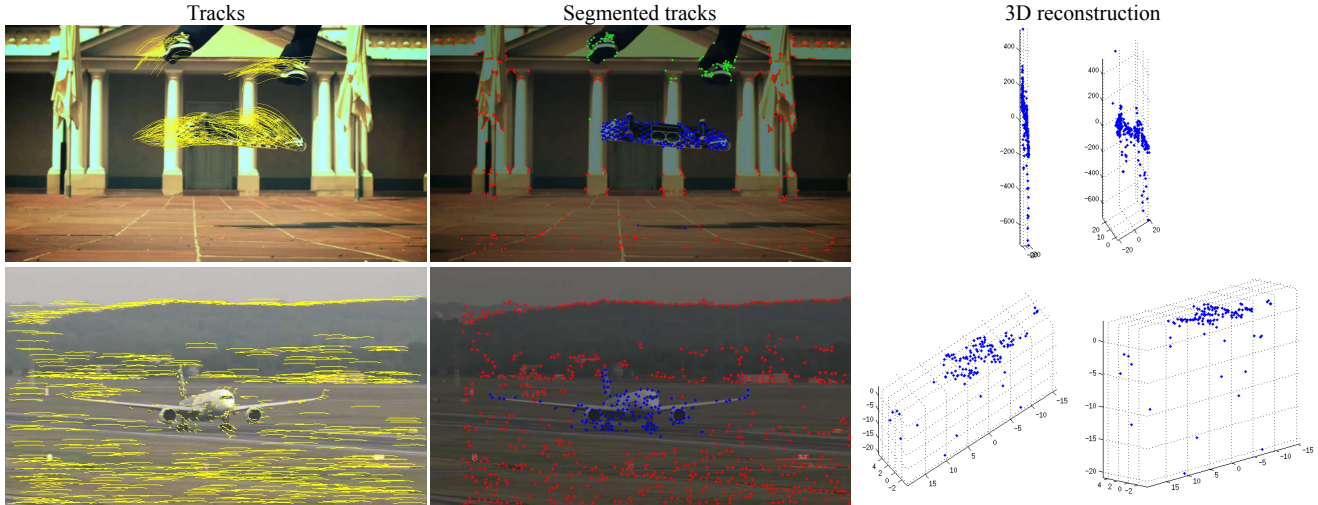


Figure 4. Incorrect reconstruction results.

may not lead to a good reconstruction. Fig. 4 shows two cases where motion segmentation gives a satisfying result, but the target objects fail to reconstruct correctly (notice the scale of the axes) due to outliers and degenerate motion. These reconstructions are not suitable for registering to the 3D models. However, Sec. 2 shows that our approach can align the 3D models to these scenes. Note that even in the case where the reconstruction can be performed well, only partial reconstruction of the visible side of objects will be obtained, making the 3D-3D registration still a complicated task.

References

- [1] M. Dodig, M. Stošić, and J. Xavier. On minimizing a quadratic function on Stiefel manifolds. Technical report, Instituto Superior Técnico, 2009. 1
- [2] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE TPAMI*, 35(11):2765–2781, 2013. 3
- [3] J. Fan and P. Nie. Quadratic programs over the Stiefel manifold. *Operations Research Letters*, 2006. 1
- [4] M. Marques and J. Costeira. Estimating 3D shape from degenerate sequences with missing data. *CVIU*, 113(2):261–272, 2009. 3