GMSF: Global Matching Scene Flow
- Supplementary Material -

1 Visualization results on FlyingThings3D

In this section, visualization results of the proposed method GMSF and several other state-of-the-art methods on FlyingThings3D [1] (License: Custom Non-commercial) are shown. The results of FLOT [3], 3DFlow [4], GMSF(OURS), and the Ground Truth are shown in the top left, top right, bottom left, and bottom right in each figure, respectively. The blue points represent the target point cloud. The green and orange points represent the warped source points with an EPE_{3D}<0.10 and EPE_{3D}>0.10, respectively.

Figures 1 2 3 4 show that the proposed method works well when the points are dense and there is moderate occlusion. In Figure 1 and Figure 2, the point clouds are dense with very little occlusion. Our proposed method managed to learn robust features and achieves almost 100% accuracy. In Figure 3 and Figure 4, the point clouds are under moderate occlusion, which means parts of the objects might be occluded. The proposed method works very well while the other methods suffer from inaccurate estimation when the objects are partially occluded. This is due to the design choice of the proposed method that the estimated scene flow is refined under the guidance of a self-feature similarity matrix. The estimated scene flow from non-occluded areas is propagated to occluded areas, thus to some extent solving the problem of occlusions.

Figure 5 and Figure 6 visualize some cases when the points are sparse. The proposed method relies very much on learning robust features. Failure cases usually happen at the edge of the objects where the points are sparse since it is difficult to learn robust features.

Figure 7 and Figure 8 visualize some failure cases with severe occlusions. When the objects are completely occluded, none of the methods are able to capture reliable features.

Animations of Figure 3 in the main paper are given in .gif format in the attachment.

2 Visualization results on KITTI Scene Flow

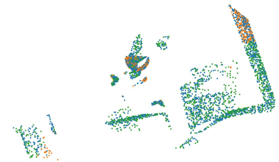
Visualization results of the proposed method GMSF and several other state-of-the-art methods on KITTI Scene Flow [2] (License: CC BY-NC-SA 3.0) are given. The results of FLOT [3], 3DFlow [4], GMSF(OURS), and the Ground Truth are shown in the top left, top right, bottom left, and bottom right in each figure, respectively. The blue points represent the target point cloud. The green and orange points represent the warped source points with an EPE_{3D}<0.10 and EPE_{3D}>0.10, respectively.

Figure 9 and Figure 10 give two overall scenes from KITTI Scene Flow [2] dataset. The results demonstrate that the proposed method can, to some extent, capture background movements e.g. movements on the side of the road. It also works very well under complex scenes with many objects while the other methods fail.

Figure 11 visualizes some zoomed-in examples of vehicles. For autonomous driving, it is very important to capture the movement of the other vehicles. The proposed method works best among the state of the art. All the models are trained on 4 $\times$ NVIDIA A40 GPUs.

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

(a) FLOT

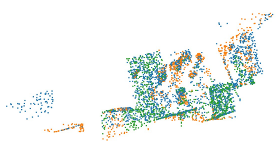


(b) 3DFlow

(c) GMSF

(d) Ground Truth

Figure 1: **Visualization results on FlyingThings3D.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



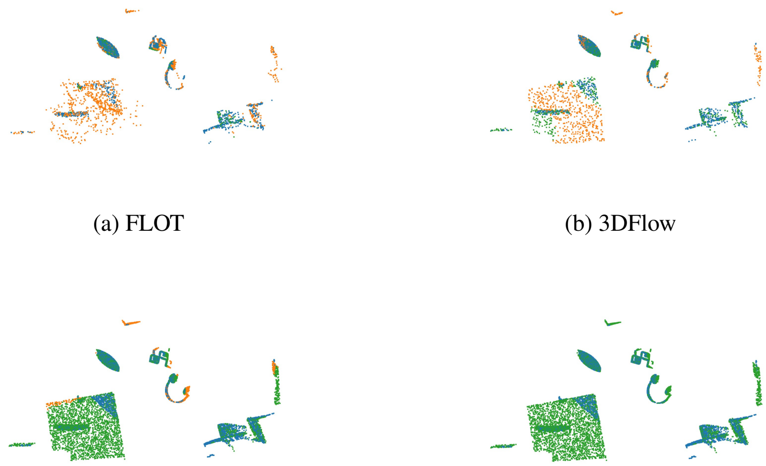
(a) FLOT

(b) 3DFlow

(c) GMSF

(d) Ground Truth

Figure 2: **Visualization results on FlyingThings3D.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



(a) FLOT (b) 3DFlow

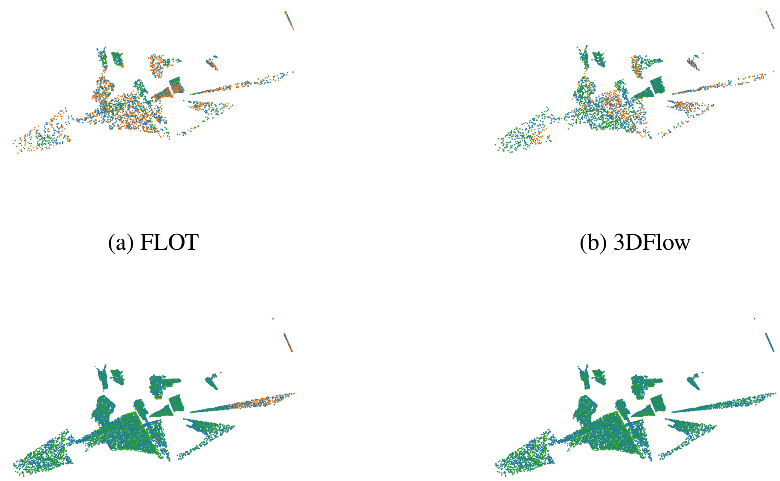
(c) GMSF (d) Ground Truth

Figure 3: **Visualization results on FlyingThings3D.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.

(a) FLOT (b) 3DFlow

(c) GMSF (d) Ground Truth

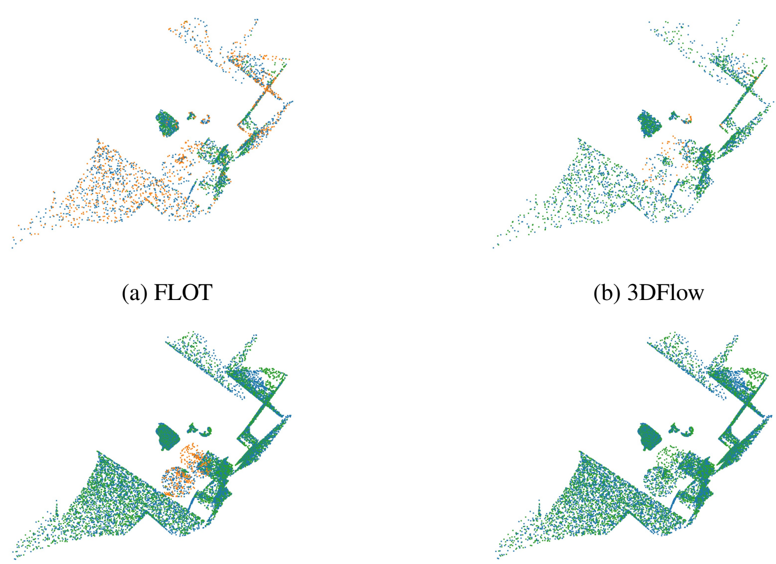
Figure 4: **Visualization results on FlyingThings3D.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



(a) FLOT (b) 3DFlow

(c) GMSF (d) Ground Truth

Figure 5: **Cases when the points are sparse.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



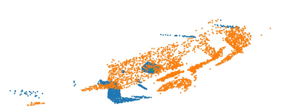
(a) FLOT (b) 3DFlow

(c) GMSF (d) Ground Truth

Figure 6: **Cases when the points are sparse.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.

(a) FLOT

(b) 3DFlow



(c) GMSF

(d) Ground Truth

Figure 7: **Failure cases under occlusions.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.

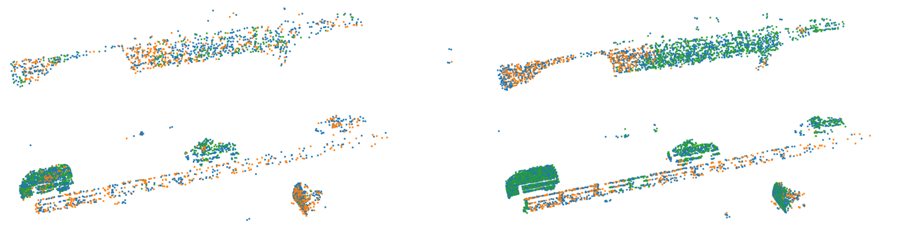
(a) FLOT

(b) 3DFlow

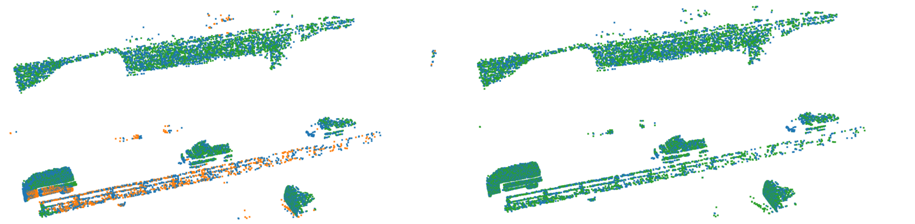
(c) GMSF

(d) Ground Truth

Figure 8: **Failure cases under occlusions.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.

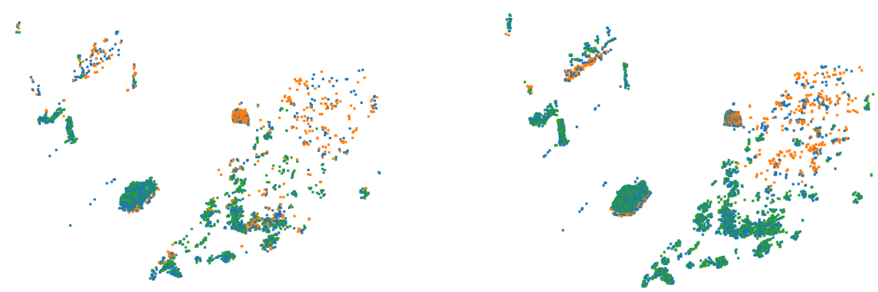


(a) FLOT (b) 3DFlow



(c) GMSF (d) Ground Truth

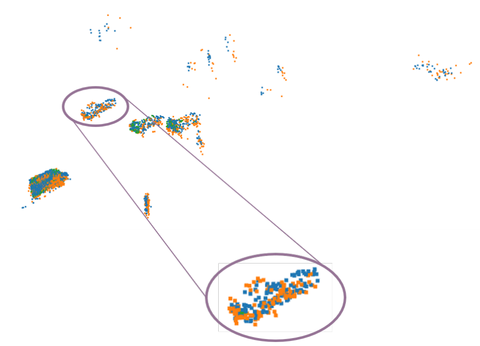
Figure 9: **Visualization results on KITTI Scene Flow.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



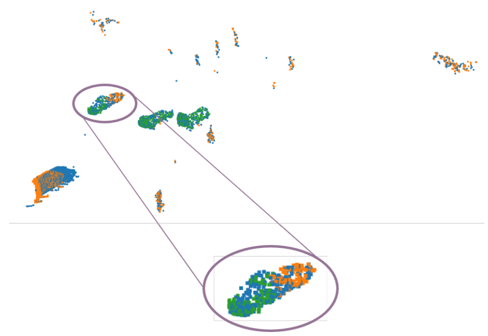
(a) FLOT (b) 3DFlow

(c) GMSF (d) Ground Truth

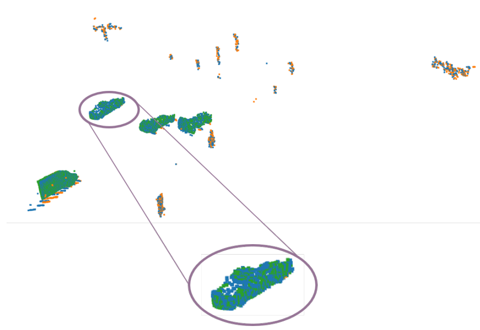
Figure 10: **Visualization results on KITTI Scene Flow.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.



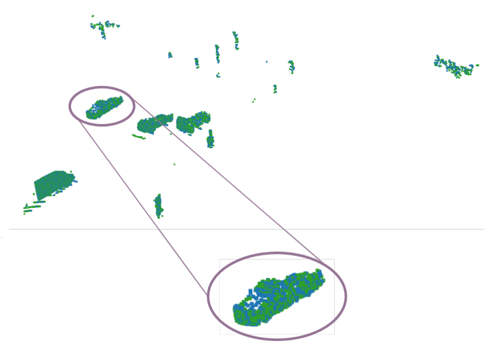
(a) FLOT



(b) 3DFlow



(c) GMSF



(d) Ground Truth

Figure 11: **Zoomed in examples on KITTI Scene Flow.** The top left, top right, bottom left, and bottom right images are the results of FLOT [3], 3DFlow [4], GMSF(OURS), and Ground truth. The blue points represent the target point cloud. The green points represent the warped source points with an EPE_{3D}<0.10. The orange points represent the warped source points with an EPE_{3D}>0.10.

References

[1] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4040–4048, 2016.

[2] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3061–3070, 2015.

[3] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Flot: Scene flow on point clouds guided by optimal transport. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII*, pages 527–544. Springer, 2020.

[4] Guangming Wang, Yunzhe Hu, Zhe Liu, Yiyang Zhou, Masayoshi Tomizuka, Wei Zhan, and Hesheng Wang. What matters for 3d scene flow network. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, pages 38–55. Springer, 2022.