

ОЦЕНКА 6D-ПОЗИЦИИ НЕИЗВЕСТНЫХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ ОДНОРАКУРСНОЙ ГЕНЕРАТИВНОЙ 3D-РЕКОНСТРУКЦИИ

Лоу Цзячэн¹

Научный руководитель – канд. техн. наук, доцент Шаветов С. В.¹

¹Университет ИТМО

jiacheng.lou@itmo.ru

Введение

Определение положения и ориентации объекта в пространстве по шести степеням свободы — одна из базовых задач компьютерного зрения и робототехники [1]. Существующие методы, в том числе PVNet [2], PVN3D [3], GDR-Net [4] и SAM-6D [5], как правило, требуют готовой CAD-модели объекта. Без неё эти системы работать не могут, что сильно сужает круг задач, где их можно применить на практике.

Перспективным направлением служат диффузионные генеративные модели (SAM3D [6]), восстанавливающие 3D-геометрию из единственного RGB-снимка, однако они страдают от неоднозначности масштаба: синтезируемые облака точек определены в нормализованном пространстве без привязки к реальным физическим размерам объекта. Аналогичное ограничение характерно для методов нейронного рендеринга — NeRF [7] и 3D Gaussian Splatting [8]. Устранение этого разрыва составляет научную проблему настоящего исследования.

Основная часть

Предложен замкнутый алгоритм оценки 6D-позиции без использования предварительных CAD-моделей, включающий три последовательных этапа.

На первом этапе базовая модель SAM3 [9] извлекает бинарную маску целевого объекта, после чего сеть SAM3D реконструирует его полную трёхмерную форму, включая скрытые части, по единственному ракурсу.

На втором этапе устраняется неопределённость масштаба. Карта глубины обратно проецируется в трёхмерное пространство через матрицу внутренних параметров камеры, после чего облако точек очищается медианной статистической фильтрацией. Далее вводится оригинальный алгоритм TP-ICP (Template-Prior Iterative Closest Point) на основе классического ICP [10]: вращательная компонента инициализируется и фиксируется через механизм сопоставления шаблонов SAM-6D, а итеративной оптимизации подлежат только коэффициент метрического масштаба и вектор трансляции. Это обеспечивает глобальную сходимость алгоритма и устойчивость к окклюзиям и шуму сенсора.

На третьем этапе откалиброванная метрическая модель служит опорным шаблоном в алгоритме SAM-6D для оценки финальной 6D-позиции объекта.

Выводы

Верификация на наборе данных LINEMOD [11] подтвердила работоспособность предложенного подхода: восстановленный метрический масштаб демонстрирует высокую точность, а спрогнозированные ограничивающие рамки визуально близки к результатам эталонной CAD-модели. Метод открывает практическую возможность внедрения роботизированных систем захвата в неструктурированных средах без предварительного создания базы CAD-моделей. Дальнейшее направление работы — расширение верификации на более сложные наборы данных и интеграция с системами планирования траекторий манипулятора для натуральных испытаний.

Литература

1. Thalhammer S., Bauer D., Hönig P. et al. Challenges for monocular 6-D object pose estimation in robotics // *IEEE Transactions on Robotics*. 2024. Vol. 40. P. 4065–4084.
2. Peng S., Liu Y., Huang Q. et al. PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019. P. 4561–4570.
3. He Y., Sun W., Huang H. et al. PVN3D: A deep point-wise 3D keypoints voting network for 6DoF pose estimation // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020. P. 11632–11641.
4. Wang G., Manhardt F., Tombari F. et al. GDR-Net: Geometry-guided direct regression network for monocular 6D object pose estimation // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. P. 16611–16621.
5. Lin J., Liu L., Lu D. et al. SAM-6D: Segment Anything Model Meets Zero-Shot 6D Object Pose Estimation // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024. P. 27906–27916.
6. Chen X., Chu F. J., Gleize P. et al. SAM 3D: 3DFy Anything in Images [Электронный ресурс] // *arXiv*. 2025. URL: <https://arxiv.org/abs/2511.16624>
7. Mildenhall B., Srinivasan P. P., Tancik M. et al. NeRF: Representing scenes as neural radiance fields for view synthesis // *Communications of the ACM*. 2021. Vol. 65. No. 1. P. 99–106.
8. Kerbl B., Kopanas G., Leimkühler T. et al. 3D Gaussian splatting for real-time radiance field rendering // *ACM Transactions on Graphics*. 2023. Vol. 42. No. 4. P. 139:1–139:14.
9. Carion N., Gustafson L., Hu Y. T. et al. SAM 3: Segment Anything with Concepts [Электронный ресурс] // *arXiv*. 2025. URL: <https://arxiv.org/abs/2511.16719>
10. Besl P. J., McKay N. D. Method for registration of 3-D shapes // *Sensor Fusion IV: Control Paradigms and Data Structures*. 1992. Vol. 1611. P. 586–606.
11. Hinterstoisser S., Lepetit V., Ilic S. et al. Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes // *Asian Conference on Computer Vision*. Berlin; Heidelberg: Springer, 2012. P. 548–562.