Кластерное согласование RGB-разметки и LiDAR-облаков для семантической навигации беспилотного ровера

Р. Е. Ломовский*, М. И. Кумсков‡

Предложен перенос жёстких меток квантованной RGB-сегментации на LiDAR: они агрегируются по геометрическим кластерам и уточняются MRF с учётом расстояния, нормалей и кривизны. На допустимых проекциях RELLIS-3D mIoU составляет 48,9 %.

Ключевые слова: семантическая сегментация, беспилотный ровер, LiDAR, RGB-изображение, U-Net, EfficientNet, квантование, DBSCAN, марковское случайное поле, alpha-expansion.

1. Задача и метод

Синхронные жёсткие RGB-метки переносятся на LiDAR и согласуются MRF на геометрических кластерах; координаты точки — $\mathbf{x}_i = (x_i, y_i, z_i)$, интенсивность не используется. Это перенос меток с геометрической регуляризацией, не fusion-сеть или 3D-сегментатор: в отличие от PointPainting и Superpoint Graph [7,8], применяется адаптивный DBSCAN с MRF. Метод принимает метки любой 2D-сети, но выигрыш при замене сегментатора не проверялся. Классы RELLIS-3D: грунт, трава, дерево/куст, камень, вода, бревно, препятствие.

U-Net с EfficientNet-B0 [1, 2] выдаёт жёсткие метки; «Прямой перенос» (табл. 1) — её оценка на допустимых проекциях до 3D-коррекции. В RUGD 24 класса сведены к семи: dirt/asphalt/gravel/concrete/mulch — грунт; sand/grass/water/log без изменений; tree/bush — дерево; rock/rock bed — камень; sky/pole/vehicle/container/building/bicycle/person/fence/sign/

bridge/picnic table исключены из loss и mIoU. Train/val/test — 4779/733/1924 кадров без перемешивания соседей. RUGD используется

* *Ломовский Роман Евгеньевич* — студент мех.-мат. факультета МГУ, e-mail: roman.lomovskii@math.msu.ru.

Lomovskii Roman Evgenievich — student, Faculty of Mechanics and Mathematics, Moscow State University.

‡ *Кумсков Михаил Иванович* — доцент мех.-мат. факультета МГУ, e-mail: mikhail.kumskov@math.msu.ru.

Kumskov Mikhail Ivanovich — Associate Professor, Faculty of Mechanics and Mathematics, Moscow State University.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

для предобучения и оценки квантования, но его классы не отождествляются с RELLIS-3D [6].

Для RELLIS-3D: dirt/mud/asphalt/concrete — грунт; grass — трава; tree/bush — дерево/куст; rubble — камень; water/puddle — вода; log — бревно; barrier/vehicle/person/pole/fence/object/building — препятствие; void/sky/прочие исключены. Sequence 00000–00002/00003/00004 — train/val/test (1963 облака), соседние кадры не перемешиваются [5]. U-Net: Dice+cross-entropy, Adam (10^{-4}), batch 8, до 30 эпох; Reduce-on-Plateau (0,5; patience 3), early stopping по val mIoU (5), три seed (42–44), QAT INT8. Компьютер: i7-10700KF, 8 ядер/16 потоков, 32 ГБ; Windows 11 Pro/WSL2 Ubuntu 22.04; Python 3.10.12, PyTorch 2.0.1, OpenVINO 2024.4, NNCF 2.8.1, OpenCV 4.8, NumPy 1.26, scikit-learn 1.2.

Проекция: $q_i = K(R\mathbf{x}_i + t)$, $(u_i, v_i) = (q_{i1}/q_{i3}, q_{i2}/q_{i3})$; точки с $Z_i \le 0$ или вне изображения отбрасываются. При коллизии пикселя видима точка с минимальной глубиной, остальные окклюдированы. Дополнительно в окне 3×3 без центра глубина должна отличаться не более чем на 0,35 м от минимума занятых соседей; без соседей точка видима. Жёсткая метка $\ell_i^0 = S(I)(u_i, v_i)$ присваивается допустимым проекциям.

Адаптивный DBSCAN [3]: $\varepsilon_i = \min\{0{,}55;\ 0{,}16(1 + 0{,}035 r_i)\}$ м, $r_i = \|\mathbf{x}_i\|_2$, MinPts $= 8$; соседство определяется $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \le \min(\varepsilon_i, \varepsilon_j)$. Шум — одноточечные кластеры. По локальным ковариациям находятся n_k, κ_k. Узлы графа — кластеры; ребро проводится при минимальном межкластерном расстоянии $\le \delta = 0{,}65$ м.

Метки кластеров Y_k находятся приближённой минимизацией энергии Поттса:

$$E(Y) = \sum_k U_k(Y_k) + \lambda \sum_{(k,l)\in\mathcal{E}} w_{kl}\, \mathbf{1}[Y_k \ne Y_l], \quad U_k(c) = -\ln \frac{h_k(c) + 1}{N_k^{\mathrm{proj}} + |C|},$$

где $h_k(c)$ — число спроецированных точек класса c, $N_k^{\mathrm{proj}} = \sum_c h_k(c)$; для кластера без проекций $U_k(c) = -\ln \pi_c$, где π_c — частота класса в train. Вес ребра

$$w_{kl} = \exp\!\left(-\frac{\|q_k - q_l\|_2^2}{\sigma_q^2} - \frac{(\kappa_k - \kappa_l)^2}{\sigma_\kappa^2} - \frac{1 - |n_k^T n_l|}{\sigma_n^2}\right)$$

учитывает расстояние между центрами, кривизну и ориентацию нормалей. Коэффициенты DBSCAN, порог видимости 0,35 м и параметры MRF ($\sigma_q = 0{,}55$ м, $\sigma_\kappa = 0{,}08$, $\sigma_n = 0{,}35$, $\lambda = 1{,}4$) фиксировались по sequence 00003 и не менялись на тесте. Alpha-expansion через PyMaxflow [4] останавливается при отсутствии снижения энергии или после 10 проходов. В среднем — около 68 тыс. точек и 1250 кластеров на облако.

2. Результаты

Тест — sequence 00004 (1963 облака), семь классов без void; mIoU и accuracy рассчитаны по общей матрице ошибок с усреднением по точкам. На допустимых проекциях (72,4%) получены 48,9/85,2%; прямой перенос в табл. 1 — базовая оценка 2D-меток. Для полного облака метрики равны 43,5/79,9%, для непроецируемых точек — 29,4/66,1%. Без проекций остаются 12,6% кластеров; распределение их меток по классам в указанном порядке: 34,8; 20,1; 12,4; 7,2; 2,9; 2,0; 20,6%. Табл. 1 содержит среднее ± СКО трёх запусков; LiDAR-время не включает 2D-сеть.

Таблица 1: Абляция на тестовой части RELLIS-3D

Вариант	mIoU, %	Acc., %	LiDAR, мс
Прямой перенос	$39{,}6 \pm 1{,}7$	$79{,}4 \pm 1{,}5$	$6{,}3 \pm 0{,}9$
Голосование без MRF	$44{,}1 \pm 1{,}4$	$82{,}7 \pm 1{,}2$	$9{,}8 \pm 1{,}1$
DBSCAN (пост.) + MRF	$46{,}7 \pm 1{,}2$	$84{,}1 \pm 1{,}0$	$12{,}4 \pm 1{,}4$
DBSCAN (адапт.) + MRF	$48{,}9 \pm 1{,}1$	$85{,}2 \pm 0{,}9$	$12{,}8 \pm 1{,}5$

Классовые IoU в указанном ранее порядке: 68,4; 59,7; 54,6; 41,8; 47,3; 31,2; 39,3%; минимум у бревна, препятствия и камня. Среднее±СКО/медиана/P95: сеть — $12{,}1 \pm 0{,}6$/12,0/13,3 мс; LiDAR — $12{,}8 \pm 1{,}5$/12,5/15,6 мс; полный конвейер — $25{,}7 \pm 2{,}2$/25,2/30,1 мс. Разность 0,8 мс сверх суммы средних 2D/LiDAR — подготовка RGB/тензора, декодирование и передача; чтение диска исключено. Прогревы/измерения: сеть 100/1000, LiDAR и конвейер 50/500.

Список литературы

[1] O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, MICCAI, 2015, 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[2] M. Tan, Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, ICML, 2019, arXiv: 1905.11946.

[3] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, KDD, 1996.

[4] Y. Boykov, O. Veksler, R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**:11 (2001), 1222–1239. DOI: 10.1109/34.969114.

[5] P. Jiang, P. Osteen, M. Wigness, S. Saripalli, “RELLIS-3D Dataset: Data, Benchmarks and Analysis”, ICRA, 2021, 1110–1116. DOI: 10.1109/ICRA48506.2021.9561251.

[6] M. Wigness, S. Eum, J. G. Rogers, D. Han, H. Kwon, “A RUGD Dataset for Autonomous Navigation and Visual Perception in Unstructured Outdoor Environments”, IROS, 2019, 5000–5007. DOI: 10.1109/IROS40897.2019.8968283.

[7] S. Vora, A. H. Lang, B. Helou, O. Beijbom, “PointPainting: Sequential Fusion for 3D Object Detection”, CVPR, 2020, 4603–4611. DOI: 10.1109/CVPR42600.2020.00466.

[8] L. Landrieu, M. Simonovsky, “Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs”, CVPR, 2018, 4558–4567. DOI: 10.1109/CVPR.2018.00479 .

Статья поступила 6 октября 2026 г.

Cluster-Based Alignment of RGB Labels and LiDAR Point Clouds for Semantic Navigation of an Autonomous Rover

R. E. Lomovskii, M. I. Kumskov

We transfer hard labels from quantized RGB segmentation to LiDAR, aggregate them by geometric clusters, and refine them with an MRF using distance, normals, and curvature. RELLIS-3D mIoU on valid projections is 48.9%.

Keywords: semantic segmentation, autonomous rover, LiDAR, RGB image, U-Net, EfficientNet, quantization, DBSCAN, Markov random field, alpha-expansion.

Received on October 6, 2026