

**РАЗРАБОТКА СИСТЕМЫ ВИЗУАЛЬНОЙ ЛОКАЛИЗАЦИИ НИЗКОЛЕТЯЩЕГО
БПЛА С ПРИМЕНЕНИЕМ ИЗВЛЕЧЕННОЙ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ И
РАСПОЗНАВАНИЕМ ОБЪЕКТОВ-ЛАНДМАРК**

Домницкий Е.А. (ИТМО)

**Научный руководитель – кандидат технических наук, доцент Попов И.Ю.
(ИТМО)**

Введение. Беспилотные летательные аппараты (БПЛА) сталкиваются с необходимостью совершать полет на предельно малых высотах (до 100 м) в таких актуальных приложениях как: сельскохозяйственный мониторинг и точное земледелие, инспекция и обслуживание инфраструктуры, поисково-спасательные операции, стратегическое назначение. Отказ от GPS представляет собой существенное эксплуатационное ограничение, особенно в сельской, сложной пересеченной местности, где сигналы спутниковой навигации могут быть, недоступными, или злонамеренно глушатся. В условиях недоступности GPS/GNSS активно исследуется применение систем визуальной локализации, что в свою очередь является одной из самых сложных задач осуществления автономной навигации. В то время как уже существуют подходы SLAM, строящие разреженную карту и корректирующие ошибку путем замыкания цикла - существенно усложняет задачу необходимость совершать полет на малой высоте на дальность от нескольких до десятков километров. Настоящая работа посвящена предложению и разработке метода визуальной локализации/релокализации, который позволит использовать доступные априори данные о местности для решения поставленной проблемы.

Основная часть. В ходе исследования предметной области, в первую очередь были рассмотрены наиболее популярные методы визуальной локализации и картирования (VSLAM). Среди VSLAM методов, выделяют методы, основанные на извлечении ключевых точек (такие как SOTA ORB-SLAM [1] и его нейросетевые дополнения, подобные [2]), прямые методы (например, LSD-SLAM [3]), а так же полностью нейросетевые пайплайны (подобные Cowan-GGR [4]). Существующие методы так или иначе включают разреженное картирование местности (напр. применяются ключевые точки для построения воксельной карты и замыкание цикла) и/или итеративное определение оптического потока для приближения перемещения в 3х-мерном пространстве с последующим построением траектории (визуальная одометрия). Эти подходы показали свою эффективность, но располагают существенными ограничениями в рамках поставленной проблемы: использование визуальной одометрии неизбежно приводит к накоплению кумулятивной ошибки; даже разреженное картирование местности требует ощутимых ресурсов для хранения “облака” вокселей при движении по протяженному маршруту; разреженная карта местности с определенной частотой требует вычислительных ресурсов на ее очистку от шумовых точек и пакетную корректировку - количество временных и вычислительных ресурсов растет пропорционально размеру карты и пройденному маршруту. Визуальная одометрия и построение траектории движения по прежнему необходимы в условиях высокой однородности среды, поскольку даже с ошибкой будут являться хорошим сигналом для позиционирования. Однако требуется более сильный метод релокализации для устранения кумулятивной ошибки. Таким образом, в поставленной во введении проблеме для осуществления визуальной локализации ключевым требованием является способность системы оперировать в большом географическом масштабе по предварительно известным данным (данным ГИС, спутниковой карте, аэрофото, галерее объектов), отказываясь тем самым от разреженного картирования местности.

Для осуществления локализации по предварительно доступным данным ГИС можно использовать принципы локализации по ортофотоплану и относительно характеристических мест/объектов среды, называемых ландмраками. В задаче локализации при полете на малой высоте точность методов прямого сопоставления перспектив борта и спутника существенно

снижается, если вообще сохраняется. Именно поэтому в условиях поставленной проблемы распознавание ландмарков может сыграть ключевую роль и данный метод локализации считается в рамках работы приоритетным.

При локализации по ландмаркам в рамках поставленной проблематики предложенный метод назван межперспективным мульти-масштабным визуальным распознаванием ландмарков (англ. Multi-scale cross-view landmark recognition). Предлагаемое Ad-hoc решение сочетает возможности таких задач глубокого обучения, как cross view matching и visual place recognition, поскольку домен данных при полете на небольшой высоте содержит признаки, характерные как для наземной съемки, так и для воздушной. Определяющим фактором также является то, что на разной дистанции одни и те же объекты могут иметь разное разрешение - объекты вдали обладают в первую очередь низкочастотными признаками, тогда как ближние объекты изобилуют высокочастотными признаками. В существующих работах касательно названных задач глубокого обучения эта деталь как правило не адресуется, но является критичной в рамках поставленной проблемы, а потому разработка архитектуры и обучение модели в составе требуемого метода будет являться научно новым. Научный интерес будет представлять полученная информация о количестве и свойстве необходимых (и достаточных) исходных данных, предложенном методе region proposal для поиска ландмарков в кадре, равно как и информация о получившемся предельном расстоянии распознавания и чувствительности к условиям видимости.

В ходе работы при изучении существующих наборов данных был сделан вывод о том, что прототипирование и испытание систем визуальной локализации в рамках поставленной проблемы с использованием лишь доступных наборов данных (включая доступные данные ГИС и датасеты для задач визуального распознавания мест) представляется чрезвычайно трудоемким и негибким, если вообще возможным. Поэтому, плагин UnrealCV на UE 5 выбран в качестве инструмента подготовки виртуальной сцены для прототипирования и испытания разрабатываемой системы, сочетая в себе возможности для извлечения тренировочных данных и проведения виртуальных экспериментов. С использованием выбранного инструмента была подготовлена симуляционная сцена протяженностью 10 на 10 км, представляющая из себя пресеченный лесистый ландшафт с водоемами. Сцена содержит дороги, а так же не плотно расположенные человеческие постройки (в т.ч. 2 населенных пункта) - по которым планируется осуществлять локализацию.

Выводы. В ходе настоящей работы, был сделан вывод о непригодности существующих VSLAM алгоритмов визуальной локализации, основанных на частичном картировании местности, для осуществления локализации и корректировки кумулятивной ошибки в больших масштабах. В рамках поставленной проблемы была выявлена целесообразность разработки метода визуальной локализации, основанного на восприятии окружающей среды и ее характеристических элементов в условиях наличия априорных данных о местности. В результате изучения существующих наборов данных в рамках поставленной проблемы, было принято решение о необходимости разработки инструмента для извлечения синтетических симуляторных данных для прототипирования и испытания системы. Предложенный и разрабатываемый в ходе настоящей работы метод может иметь научный и практический интерес с точки зрения получения новой информации о необходимых исходных данных и требованиях к свойствам нейросетевой модели.

Список использованных источников:

1. ACampos C. et al. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multiplanar slam //IEEE Transactions on Robotics. – 2021. – Т. 37. – №. 6. – С. 1874-1890.
2. Ai, Y.; Rui, T.; Lu, M.; Fu, L.; Liu, S.; Wang, S. “DDL-SLAM: A robust RGB-D SLAM in dynamic environments combined with deep learning” // IEEE Access 2020, 8, 162335–162342.
3. Engel J., Schöps T., Cremers D. LSD-SLAM: Large-scale direct monocular SLAM //European conference on computer vision. – Cham : Springer International Publishing, 2014. – С. 834-849.
4. Mumuni F., Mumuni A., Amuzuvi C. K. Deep learning of monocular depth, optical flow and ego-motion with geometric guidance for UAV navigation in dynamic environments //Machine Learning with Applications. – 2022. – Т. 10. – С. 100416.

