

УДК 004.896

МОДУЛЬ ИЗВЛЕЧЕНИЯ РЕЛЯЦИОННОЙ ИНФОРМАЦИИ МЕЖДУ ОБЪЕКТАМИ ДЛЯ ВИЗУАЛЬНОЙ РЕЛОКАЛИЗАЦИИ БПЛА НА КАРТЕ

Домницкий Е.А. (ИТМО)

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

Введение. VSLAM (Visual Simultaneous Localization and Mapping) – визуальная единовременная локализация и картирование - подход, основанный на применении видеокамер (монокулярные, стерео или RGB-D) в качестве основных входных датчиков, фиксирующих визуальную информацию об объектах и средах для локализации и построения локальных карт. Данная область исследований является чрезвычайно актуальной, поскольку получаемые результаты активно используются в интеллектуальных сенсорных и робототехнических системах для решения задач визуальной одометрии, структурного картирования окружения и навигации. В частности, актуальной является задача визуальной релокализации беспилотного агента на заранее известной и размеченной карте в условиях недоступности GPS/GNSS. В настоящем докладе автор предлагает пайплайн модуля визуальной релокализации для VSLAM системы, утилизирующую информацию о взаимоотношении между объектами в зоне видимости, а также некоторую доступную информацию о ландшафте.

Основная часть. Важно уточнить контекст задачи. Задача визуальной релокализации с привязкой к заранее доступной карте рассматривается применительно к беспилотным летательным аппаратам (БПЛА). Под картой подразумевается спутниковый или аэро-снимок области с привязкой к географической системе координат. Основным датчиком будет являться монокулярная камера или некоторая стереосистема камер, способная запечатлеть как можно больший участок среды вокруг БПЛА. Допускается и применение иных датчиков, как, например, гироскоп для определения угла отклонения борта относительно нормали к поверхности, так и радар для измерения высоты средства над точкой среды. В условиях рассматриваемой задачи предполагается, что борт оснащен некоторой сквозной системой визуально-инерциальной одометрии, утилизирующей глубокое обучение (как в работах [1] и [2]). То есть, борт способен извлекать из видеопотока информацию о глубине кадра (приближенные расстояния до точек пространства), а также способен эстимировать эго-движение борта (англ. ego-motion, движение в 3х-мерном пространстве, в т.ч. строить приближенную траекторию). И хотя подобное оснащение само по себе позволяет извлекать информацию о положении борта в пространстве и осуществлять картирование – оно не позволяет привязываться к конкретной заранее известной области, в которой оперирует агент, а также имеет свойство накапливать ошибку, приводящую к искажению информации о положении агента. Так же указанное оснащение не подразумевает извлечения какой-либо семантической информации из кадра (положения объектов, характеристики поверхностей и т.д.). Таким образом, возникает потребность в периодическом уточнении положения агента и его привязке к конкретной области для дальнейшей навигации.

Подобную задачу активно изучают исследователи из области беспилотного транспорта. Так, например, применительно к беспилотному автомобилю, задачу релокализации на заранее доступной карте области решают в своей работе авторы [3]. Авторы утилизируют подход BEV (Bird's Eye View, англ. вид с высоты птичьего полета), при котором сам снимок/карта предварительно размечены (имеется информация в виде масок об объектах, таких как дороги, пешеходные переходы, дорожки и т.п.) и эта информация посредством глубокой нейросети отображается в нейронное BEV представление карты. Суть решения заключается в реконструировании BEV перспективы области вокруг автомобиля, при котором семантически сегментируются объекты дорожной инфраструктуры, определяются высоты этих объектов над поверхностью, полученные данные проецируются на BEV «патч» (плоскость земли), который

посредством глубокой нейросети так же отображается в нейронное BEV представление. Полученный нейронный патч затем сопоставляется с априорной нейронной картой, а на выходе получается вероятностное отображение положения автомобиля на карте. И хотя предложенный подход представляет интерес для обозначенной в настоящем докладе задачи – априорная карта может быть большой, в следствие чего вероятностный вывод может занимать большое время. Более того, авторы искусственно сужают область поиска посредством применения «зашумленной» априорной информации об области положения автомобиля (GPS). Так же рассмотренный подход направлен узко на функционирование в среде дорожно-транспортной инфраструктуры.

Предлагаемый в рамках настоящего доклада пайплайн использует идеи, рассмотренные выше, но расширяет область применения до релокализации на произвольной местности, а так же выдвигает новую идею для сужения области поиска. Для построения нейронной карты, как уже было оговорено, будет применяться спутниковый или аэро-снимок. Он будет предварительно размечаться (либо вручную, либо при помощи моделей для детекции/семантической сегментации) – на нем будут выделяться дороги, деревья, здания (возможно несколько подклассов), водоемы. В добавок к этому, будет использоваться карта высот с привязкой к снимку (при этом возможно извлечение дополнительной информации в виде карты градиентов высот). Семантические метки и высоты будут принимать участие в формировании нейронного BEV представления карты (в результате слияния в слоях нейросети). Однако, семантические метки и высотные ландмарки будут принимать отдельное участие – они будут использоваться для извлечения графов отношения объектов/меток (чем-то похоже на метод, предложенный в [4]), которые затем будут преобразовываться в графовые эмбединги. Предполагается, подобные эмбединги будут извлекаться из разреженных точек по всей карте при помощи GNN (Graph Neural Network), формируя таким образом разреженную таблицу графовых эмбедингов. Во время оперирования на самом БПЛА при помощи моно-камеры или стереосистемы камер из видеопотока будут извлекаться семантические метки окружения (в т.ч. детектироваться объекты), при помощи карты глубины кадра и сенсоров (радара, гироскопа) будут определяться дистанция до и относительные высоты точек пространства. При этом обработка большого числа точек будет очень ресурсозатратной, посему для объектов будет определяться только дистанция (и высота), а высоты считаться для разреженных точек на земле. Семантические метки и информация о высотах/градиентах высот будет проецироваться на локальный BEV «патч» области вокруг БПЛА, который будет впоследствии отображаться на нейронную BEV карту. Отдельно же метки объектов и высотная информация будет применяться для извлечения графового эмбединга, который затем при помощи механизма внимания будет отображаться в таблицу графовых эмбедингов карты, сужая тем самым область поиска совпадений локального BEV «патча».

Выводы. В рамках доклада предложен пайплайн модуля визуальной релокализации для VSLAM системы, утилизирующую информацию о взаимоотношении между объектами в зоне видимости, а также некоторую доступную информацию о ландшафте.

Список использованных источников:

1. Aslan M.F., Durdu A., Yusefi A., Yilmaz A. “HVIOnet: A deep learning based hybrid visual-inertial odometry” // Neural Networks, 2022, 155, 461–474.
2. Almalioglu Y., Turan M., Saputra M.R.U., de Gusmão P.P.B., Markham A., Trigoni N. “SelfVIO: Self-supervised deep monocular visual-inertial odometry and depth estimation” // Neural Networks, 2022, 150, 119–136.
3. A. B. Camiletto, A. Bochicchio, A. Liniger, D. Dai, A. Gawel. “U-BEV: Height-aware Bird's-Eye-View Segmentation and Neural Map-based Relocalization” // arXiv.org, 2023
4. Qia C, Xiang Z., Wang X., Chen S., Fan Y., Zhao X. “Objects matter: Learning object relation graph for robust absolute pose” // Neurocomputing, 2023, 521, 11–26.