

НЕЙРОСЕТЕВОЙ АЛГОРИТМ ВИЗУАЛЬНОЙ ОДОМЕТРИИ МОБИЛЬНОГО РОБОТА С ОБУЧЕНИЕМ НА СИНТЕТИЧЕСКИХ ДАННЫХ

Диане С.А.К., Лесив Е.А.

Институт проблем управления им. В.А. Трапезникова РАН,

Россия, г. Москва ул. Профсоюзная, д.65

sekoudiane1990@gmail.com, smailsbobs@gmail.com

Аннотация: Предложена архитектура нейронной сети для решения задачи визуальной одометрии. Обучение сети осуществляется на основе данных, собранных в виртуальной среде. На вход системы поступают два последовательных видеокadra. На выходе формируется вектор параметров движения, содержащий информацию о линейных и угловых скоростях робота.

Ключевые слова — генерация обучающих множеств, визуальная одометрия, нейронные сети, автономные роботы, комплексирование сенсорной информации

Введение

За последние несколько десятилетий проведено огромное количество исследований, посвященных вопросам функционирования автономных мобильных роботов в недетерминированных средах. Так, одним из ключевых направлений в области управления автономными роботами является определение их пространственного расположения и скорости движения.

Одновременно с этим активное внимание научного сообщества привлечено к вопросам обучения и применения нейронных сетей для решения задач анализа видеoinформации, в частности для решения задачи визуальной одометрии, когда требуется оценить движение робота в пространстве по последовательности видеокadров с бортовой камеры.

Объединение этих двух направлений исследований может обеспечить высокое качество работы навигационной системы за счет использования способности нейросетевых алгоритмов к обобщению и выделению скрытых зависимостей в изображениях внешней среды, получаемых с бортовой камеры робота.

В отличие от работ [1, 2], где для формирования траектории движения в нейронной сети используется остаточная информация о характере движения за счет встроенных в архитектуру рекуррентных слоев, в предлагаемом подходе на нейронную сеть независимо подается два последовательных кадра для формирования оптического потока.

Для повышения точности бортовой навигационной системы также как и в работе [2] представляется целесообразным комплексировать показания нейронной сети, осуществляющей визуальную одометрию, с показаниями инерциальной навигационной системой робота. Однако в предлагаемом подходе комплексирование данных выполняется не средствами нейронной сети, но осуществляется аналитически на более позднем этапе.

Стоит отметить, что ряд подходов к навигации мобильных роботов основывается на оценке оптического потока методом Лукаса-Канаде [3, 4] с учетом характеристик камеры, таких как угол обзора и разрешение, без применения нейронных сетей. Однако подобные алгоритмы сильно зависят от текстурных особенностей поверхности, по которой перемещается робот, и от качества видеоизображения.

В то же время существующие нейросетевые подходы для решения задачи визуальной одометрии являются более перспективным, однако обладают значительной вычислительной сложностью, что может затруднить их использование в автономных робототехнических системах. Использование нейронных сетей для решения конкретных прикладных задач мобильной робототехники требует не только правильного выбора архитектуры нейронной сети, но и подготовки качественного множества данных для ее обучения.

Большинство работ в данном направлении полагаются на использование обучающих множеств, сформированных вручную, что чревато неполнотой, а зачастую и неточностью в составлении входных и выходных образов нейронной сети, равно как и невозможностью оперативной подстройки сети под решение новых задач. Альтернативный подход, предлагаемый в настоящем исследовании, связан с автоматической генерацией обучающих выборок для решения навигационных задач.

1 Нейросетевая визуальная одометрия

Основными трудностями при разработке алгоритмов визуальной одометрии являются необходимость надежного обнаружения ключевых точек изображения, оценки их смещения с учетом

удаленности отдельных участков изображения от камеры и необходимости отделения движущихся объектов от статического фона. В то же время нейронные сети как универсальные аппроксиматоры функций хорошо зарекомендовали себя при решении подобных слабо формализованных задач, что позволяет рассматривать их как альтернативу аналитическим методам визуальной навигации.

Для решения задачи визуальной одометрии используется нейронная сеть, представляющая собой многослойную вычислительную структуру (рис. 1), на вход которой подаются два последовательных видеокадра X_{t-1}, X_t , а на выходе формируются оценки смещений в трех направлениях и углов поворота по трем координатным осям $\{\Delta x, \Delta y, \Delta z, \Delta \alpha, \Delta \beta, \Delta \gamma\}$.

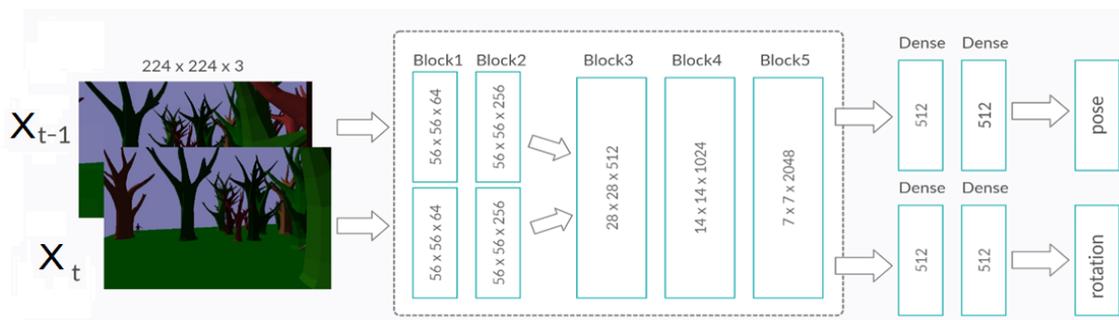


Рис. 1. Архитектура предлагаемой нейронной сети для монокулярной визуальной одометрии

Сверточная часть сети для выделения признаков (Block1 – Block5) строится из структурных блоков ResNet на основе сети ResNet50 [5]. После чего следует два ряда полносвязных слоев для оценки смещения видимой области кадра. При такой архитектуре сеть сначала создает осмысленные представления двух входных изображений по отдельности, а затем объединяет их на более высоком уровне. Объединение представлений осуществляется с помощью корреляционного слоя (Block3), в котором выполняется мультипликативное сравнение фрагментов из двух наборов карт признаков.

В качестве функции потерь для обучения сети применяется функция:

$$Loss(X_{t-1}, X_t) = \|p - \hat{p}\|_2 + \beta \|q - \hat{q}\|_2, \quad (1)$$

где x, \hat{x} - эталонное и оцениваемое сетью изменение положения между (X_{t-1}, X_t) , q, \hat{q} - эталонное и оцениваемое сетью изменение ориентации между (X_{t-1}, X_t) , β - весовой коэффициент, $\|\cdot\|_2$ - евклидова норма.

2 Автоматическое формирование обучающих выборок в виртуальной среде

В рамках научной проблематики визуального анализа изображений можно выделить несколько основных задач, решение каждой из которых допускает применение технологии нейронных сетей, обучаемых на базах аннотированных примеров. Среди них: классификация объектов; сегментация объектов; локализация и оценка геометрических параметров объектов; оценка глубины изображения и трехмерная реконструкция; анализ топологии и лингвистическая интерпретация сцен; оценка состояния внешней среды, а также визуальная навигация.

Подход, объединяющий решение вышеперечисленных задач по настройке визуальных анализаторов, может базироваться на применении технологий виртуальной реальности для формирования синтетических обучающих множеств [6, 7].

Наличие качественного обучающего множества во многом определяет эффективность работы алгоритмов машинного обучения. Следует отметить, что при составлении обучающей выборки следует уделять внимание не только объему данных, но и таким ее аспектам, как сбалансированность и полнота. Целесообразно включать в обучающую выборку изображения, максимально приближенные к условиям дальнейшего использования нейронной сети, а также применять дополнительные технологии аугментации изображений [8].

В настоящем исследовании предлагается технология синтеза обучающих примеров, базирующаяся на применении средств трехмерной графики. В основе программного комплекса для генерации обучающих множеств (рис. 2) лежит модуль синтеза изображений виртуальных сцен. Виртуальная сцена представляет собой совокупность трехмерных объектов различных категорий, снабженных описанием их положения в пространстве, ориентации и цвета.

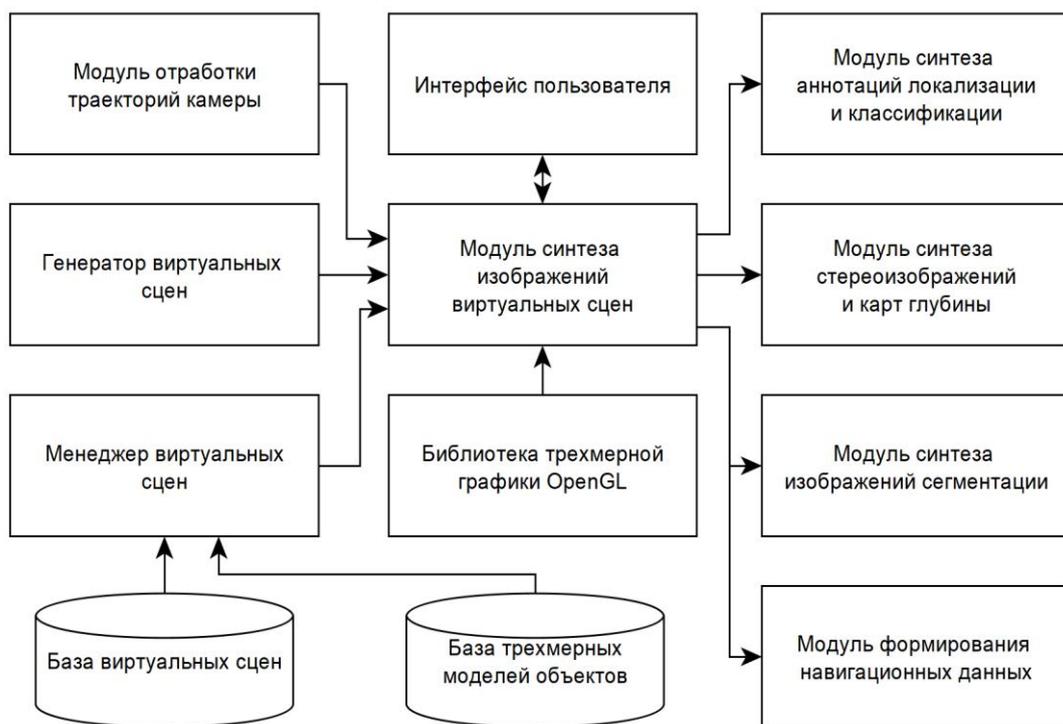


Рис. 2. Структура программного комплекса для генерации обучающих множеств

Таким образом, комплекс позволяет генерировать обучающие выборки для решения задач визуальной классификации, локализации, сегментации, оценки глубины изображений. Кроме того, виртуальная среда предоставляет доступ к точному положению видеокамеры в последовательные моменты времени, что дает возможность синтезировать обучающие выборки для решения задачи визуальной одометрии, а также позволяет моделировать показания подсистемы инерциальной навигации, совмещенной с видеокамерой.

2.1. Формирование обучающих множеств в задаче визуальной одометрии

Для задач настройки нейросетевых алгоритмов на решение конкретных прикладных задач применяется подход, основанный на загрузке заблаговременно подготовленных виртуальных сцен. Разнообразие обучающих примеров при этом достигается изменением ракурса наблюдения в процессе движения камеры по указанной траектории.

На первом этапе решения задачи по подготовке обучающего набора используется программное обеспечение, позволяющее формировать описания виртуальных сцен в виде

$$W = \{o_1, \dots, o_N\}, \quad (2)$$

где o_i - программная структура, характеризующая положение, ориентацию, класс и особенности внешнего облика объекта.

Далее, основываясь на информации о положении объектов в сцене, а также координатах начальной и конечной точек движения, формируются траектории движения камеры. При этом для обеспечения безопасного движения камеры на некотором удалении от препятствий применяется метод потенциальных полей. Различные варианты тестовых траекторий движения объекта в среде представлены на рис. 3.

На втором этапе автоматически сформированные сцены и рассчитанные траектории движения загружаются в трехмерную среду (рис. 4). Производится интерпретация текстовых описаний сцен и формирование соответствующих программных представлений для объектов, перечисленных в загруженном файле.

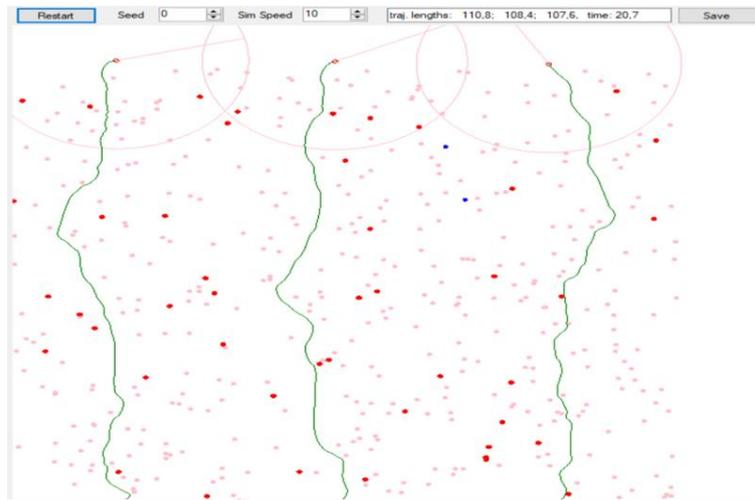


Рис. 3. Тестовые траектории движения в виртуальной среде

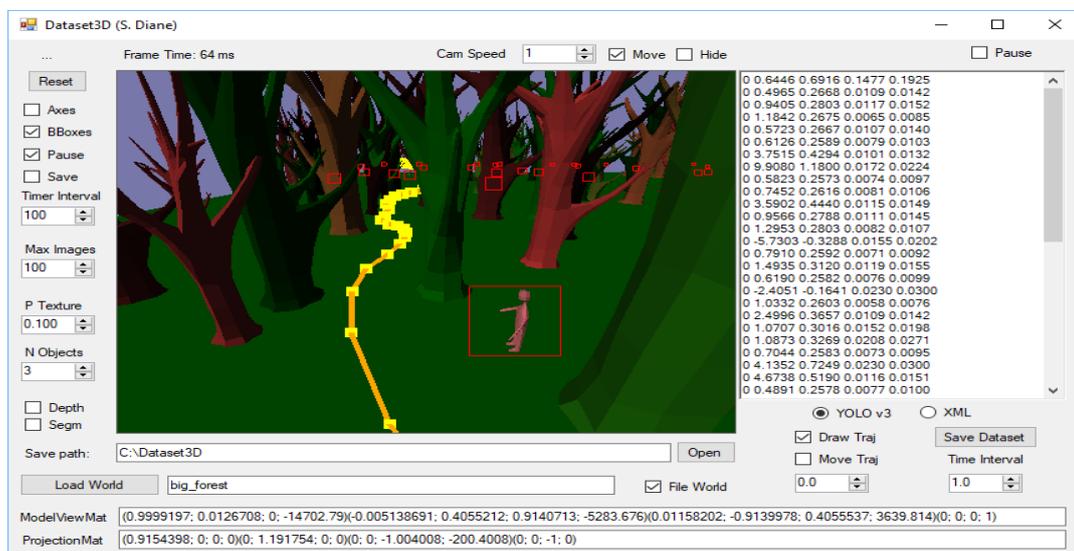


Рис. 4. Результат трехмерной визуализации сцены

Первичные данные для генерации обучающей выборки сохраняются в следующем виде:

$$\{X_1, p_1\}, \dots, \{X_N, p_N\}, \quad (3)$$

где X_i - видеокادر, полученный в виртуальной среде на шаге $i=1..N$, а p_i - вектор, содержащий декартовы координаты и ориентацию камеры в формате углов Крылова в момент времени i .

Обучающие данные, непосредственно используемые для настройки нейронной сети, формируются из первичной выборки по принципу:

$$s_i = \{X_{i-1}, X_i, p_i - p_{i-1}\}, i=2..N, \quad (4)$$

Таким образом, задавая местоположение и ориентацию движущегося объекта в виртуальной среде и программно получая изображения с бортовой видеокмеры робота, можно сформировать необходимое количество обучающих выборок, удовлетворяющих структуре нейронной сети, представленной на рис. 1.

2.2 Уточнение навигационной информации с учетом инерциальных данных

Известно, что повышения точности различных систем обработки информации можно добиться за счет комплексирования данных от нескольких источников. Так, для уточнения координат мобильного робота возможно совместить показания нейросетевой визуальной одометрии с инерциальной навигационной системой робота.

При решении данной задачи необходимо учитывать, тот факт, что погрешность визуальной одометрии минимальна при прямолинейном движении и возрастает в момент поворота камеры. Погрешность же инерциальной навигации в момент поворота камеры, напротив, оказывается меньше.

Таким образом, простейший способ комплексирования может быть реализован следующим образом:

$$\begin{cases} p_i = \alpha \cdot a_i + (1-\alpha) \cdot b_i, \\ \alpha = \max(0, 1 - |\omega|/\omega_0), \end{cases} \quad (5)$$

где a_i – координата робота согласно показаниям визуальной одометрии, b_i – координата робота согласно показаниям инерциальной навигации, α – коэффициент предпочтения данных визуальной одометрии относительно данных инерциальной навигации, ω – текущая угловая скорость, ω_0 – максимальная угловая скорость, допустимая при визуальной оценке движения робота.

Пример использования данного подхода представлен на рис. 5, где совмещение данных инерциальной и визуальной навигации существенно повысило точность оценки траектории робота.

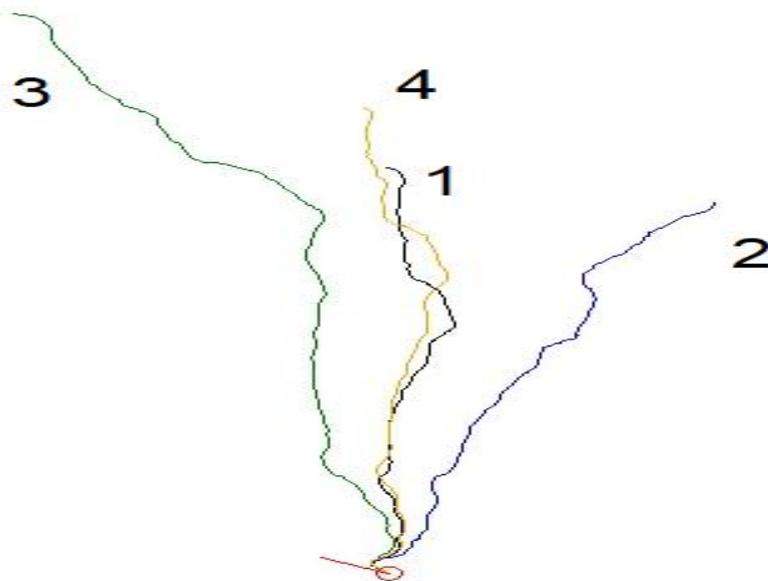


Рис. 5 Результаты комплексирования данных инерциальной и визуальной навигации: 1 – эталонная траектория, 2 – траектория по данным визуальной одометрии, 3 – траектория по данным инерциальной навигации, 4 – траектория, полученная по результатам комплексирования

3 Экспериментальные исследования

Для проведения экспериментальных исследований была разработана виртуальная модель среды, представляющая собой лес. Разработанная трехмерная модель содержит в себе большое количество препятствий и приближена к предполагаемым условиям использования разработанной системы навигации в реальной среде.

В качестве показаний инерциальной системы в данном эксперименте были использованы модели датчиков, рассчитываемых на основе разностных уравнений движения с учетом применения псевдослучайных аддитивных погрешностей. Итоговый сгенерированный набор данных для обучения сети насчитывает в себе 500 обучающих примеров и 50 тестовых. Результат работы навигационной системы показан на рис. 6 (в целях простоты визуализации представлена проекция 6-компонентного выходного вектора нейронной сети на двумерную плоскость XY).

Оценка быстродействия разработанной архитектуры нейронной сети представлена в табл. 1. Измерения проводились с использованием персонального компьютера, оснащенного видеокартой Nvidia GeForce 1070Ti без использования интерфейса TensorRT и на одноплатном компьютере Nvidia Jetson Nano с интерфейсом TensorRT.

Таблица 1. оценка быстродействия предлагаемого подхода

Hardware	FPS
Nvidia GeForce 1070Ti	27
Nvidia Jetson Nano	29

Из результатов проведенных экспериментов можно убедиться в работоспособности предлагаемого подхода, и в возможности его использования для решения прикладных задач мобильной робототехники. Несмотря на то, что в виртуальной модели леса отсутствуют текстуры, и число дескрипторов для оценки оптического потока невелико, нейросеть достойно справилась с задачей визуальной одометрии даже в подобных условиях.

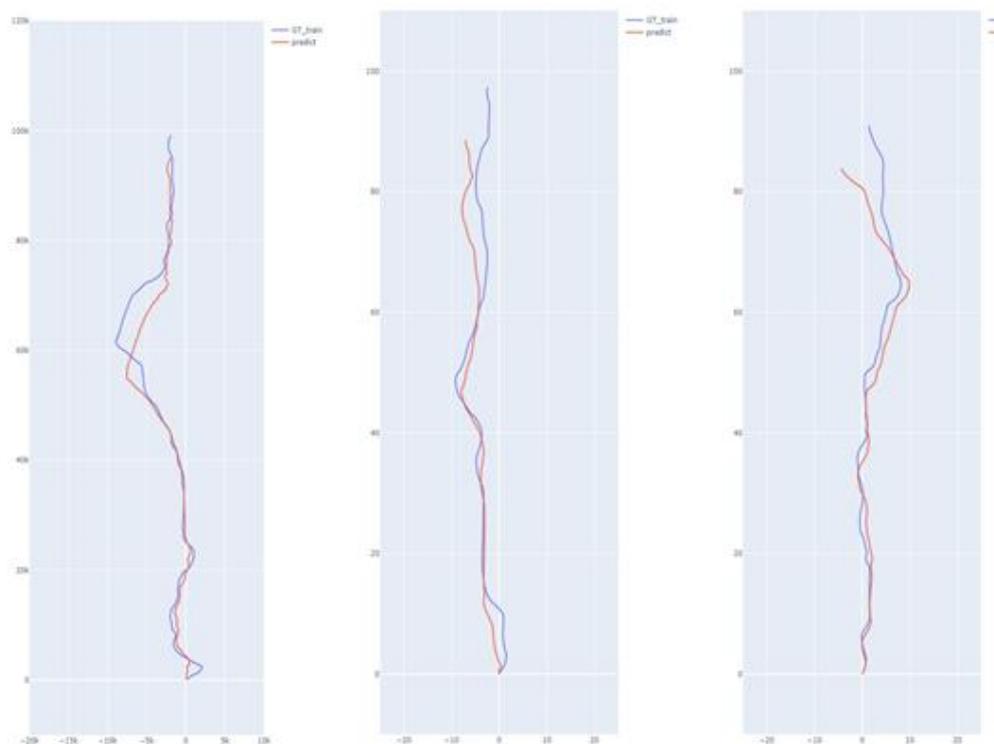


Рис. 6. Результаты визуальной одометрии: эталонные траектории (синий цвет) и восстановленные при помощи нейронной сети (красный цвет)

Заключение

В проведенном исследовании были рассмотрены математические модели локализации робота на основе нейросетевого анализа видеопотока с бортовой видеокамеры в составе комплексной системы навигации.

Полученные результаты подтверждают перспективность развиваемого подхода формирования обучающих множеств в виртуальной среде для решения широкого круга робототехнических задач. Интеграция технологий трехмерной графики и экспертных знаний о предметной области позволяет осуществить эффективную и вычислительно быструю генерацию обучающих множеств для решения задач визуального анализа в необходимом объеме.

В то же время очевидна применимость данного подхода непосредственно для настройки нейросетевых систем визуальной одометрии. Среди дальнейших перспектив данного исследования следует выделить задачу по переносу разработанных технологий на задачи навигации в реальной среде в составе систем навигации колесных роботов и мультикоптеров. Одним из примеров прикладного применения системы может служить задача по поиску людей, потерявшихся в лесу [9].

Возможное научное применение предлагаемого подхода связано с его интеграцией в алгоритмы одновременной локализации и картографирования. Повышение точности навигации повысит качество и скорость построения карты местности как одиночными роботами, так и роботами, действующими в группе.

Литература

1. *Sen Wang, Ronald Clark, Hongkai Wen and Niki Trigoni*, " DeepVO: Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks", in ICRA 2017.
2. *Ronald Clark, Sen Wang, Hongkai Wen, Andrew Markham, Niki Trigoni*, "VINet: Visual-Inertial Odometry as a Sequence-to-Sequence Learning Problem" in AAAI-17.
3. *C. Chuanqi, H. Xiangyang, Z. Zhenjie and Z. Mandan*, "Monocular visual odometry based on optical flow and feature matching," 2017 29th Chinese Control And Decision Conference (CCDC), 2017, pp. 4554-4558.
4. *S. Poddar, R. Kottath, V. Karar*, Evolution of Visual Odometry Techniques. 2018. arXiv:1804.11142.
5. *K. He, X. Zhang, S. Ren and J. Sun*, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
6. *K. Židek, P. Lazorik, J. Pite, A. Hošovský*, An Automated Training of Deep Learning Networks by 3D Virtual Models for Object Recognition. Symmetry 2019, 11(4), 496.
7. *С.А.К. Дуане, Е.А. Лесив*, Методология автоматического синтеза обучающих множеств в задачах нейросетевого управления автономными роботами // Управление развитием крупномасштабных систем (MLSD'2019): материалы 12 междунар. конфер, 1–3 окт. 2019 г., Москва / под общ. ред. С.Н. Васильева, А.Д. Цвиркуна – М.: ИПУ РАН, 2019., стр. 369
8. *Shorten, C., Khoshgoftaar, T.M.* A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60, 2019
9. *Trefilov P.M., Meshcheryakov R.V., Chekhov A.V., Diane S.A.K., Rusakov K.D., Lesiv E.A., Kolodochka M.A.*, A swarm of quadcopters to search for missing people. In Proc. of 19th IFAC Conference on Technology, Culture and International Stability, TECIS 2019, 26 - 28 September, 2019 / Sozopol Bulgaria