ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ НАСТРОЙКИ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ НА ОСНОВЕ СИНТЕЗА ОБУЧАЮЩИХ ПРИМЕРОВ В ВИРТУАЛЬНОЙ СРЕДЕ

Диане С.А.К., Лесив Е.А., Зинченко И.В.
Институт проблем управления им. В.А. Трапезникова РАН
diane1990@yandex.ru, smailsbobs@gmail.com, ziv97@mail.ru

Аннотация. Описана технология автоматического синтеза обучающих множеств для настройки нейросетевых анализаторов изображений на обучающих множествах, синтезированных в виртуальной среде с использованием средств трехмерной графики, а также в реальной среде с применением технологий визуального трекинга. Предложены принципы формирования обучающих выборок для задач классификации объектов и локальной навигации автономных роботов. Исследована возможность применения нейросетевых классификаторов, обученных на виртуальных множествах, при решении реальных прикладных задач.

Ключевые слова: классификация визуальных образов, визуальная одометрия, синтез обучающих множеств, трехмерная графика, нейронные сети

Введение

За последние несколько десятилетий проведено огромное количество исследований, посвященных вопросам картографирования среды функционирования автономных роботов. Одновременно с этим активное внимание научного сообщества привлечено к вопросам обучения нейронных сетей для решения задач визуальной классификации и анализа геометрических параметров объектов. С одной стороны, очевидно, что совмещение двух этих направлений обеспечит повышение информативности формируемых карт, за счет добавления в них семантической информации об объектах, расположенных в исследуемой зоне.

С другой стороны, задача визуальной классификации является комплексной и на сегодняшний день решена не полностью. С развитием вычислительной техники стало очевидно, что наилучшие результаты при распознавании визуальных образов дают сверточные нейронные сети. Однако использование данной технологии для решения конкретных прикладных задач требует не только правильного выбора архитектуры нейронной сети, но и подготовки качественного множества данных для ее обучения.

Большинство работ в данном отношении полагаются на использование обучающих множеств, сформированных вручную [1, 2], 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 задач визуального анализа изображений.

1 Задачи визуального анализа изображений

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:

1. визуальная классификация одиночных объектов [1];
2. локализация и оценка геометрических параметров объектов [2, 5];
3. визуальная навигация и оценка состояния внешней среды [6];
4. визуальная сегментация объектов на изображении [7];
5. оценка глубины изображения и 3D-реконструкция [8];
6. анализ топологии и лингвистическая интерпретация сцен [9].

Подход, объединяющий решение вышеперечисленных задач по настройке визуальных анализаторов, может базироваться на применении технологий виртуальной реальности [10, 11].

2 Автоматическое формирование обучающих выборок в виртуальной среде

Наличие качественного обучающего множества во многом определяет работу алгоритмов машинного обучения. Следует отметить, что при составлении обучающей выборки следует уделять внимание не только объему данных, но и таким моментам, как сбалансированность классов и порядок их следования. Данные должны содержать сопоставимый объем экземпляров для каждого класса и должны быть перемешаны. Целесообразно включение в обучающую выборку данных максимально приближенных к условиям дальнейшего использования нейронной сети.

В настоящем исследовании предлагается технология синтеза обучающих множеств, базирующаяся на применении средств трехмерной графики (библиотека OpenGL [12]) и разработанного на их основе программно-алгоритмического комплекса (рис. 1).

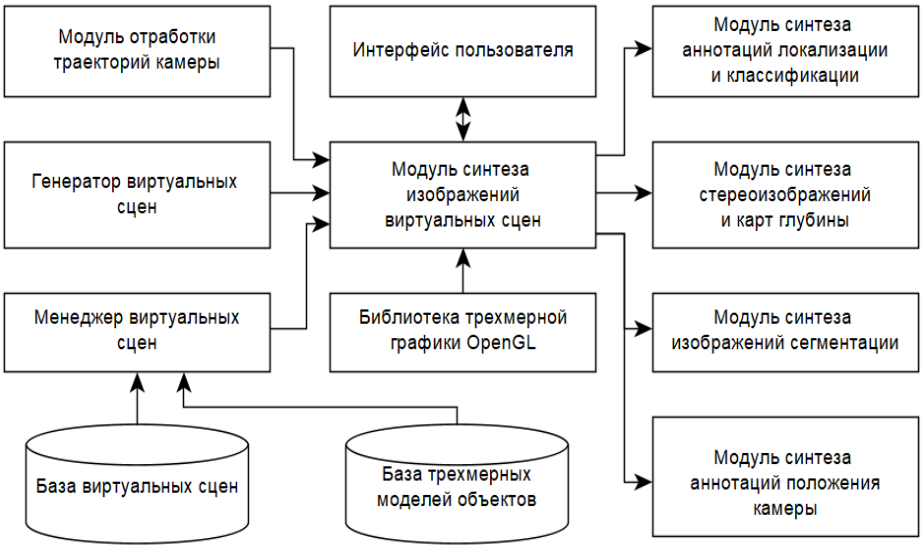

Рис. 1. Структура программного комплекса для генерации обучающих множеств

В основе программного комплекса для синтеза обучающих множеств (КСОМ) лежит модуль синтеза изображений виртуальных сцен. Виртуальная сцена представляет из себя совокупность трехмерных объектов различных категорий, снабженных описанием положения в пространстве, ориентации и цветовых характеристик.

2.1 Формирование обучающих множеств в задачах визуальной классификации

В соответствии с вышеперечисленными задачами визуального анализа изображений КСОМ позволяет генерировать обучающие выборки для решения задач визуальной классификации, локализации, сегментации, оценки глубины изображений. Кроме того, виртуальная среда предоставляет доступ к точному положению камеры в последовательные моменты времени, что дает возможность синтеза обучающих выборок и для решения задачи визуальной одометрии.

Формирование виртуальных сцен на этапе, предшествующем отрисовке, может выполняться двумя способами.

Для задач грубой настройки нейросетевых классификаторов, когда взаимное положение различных объектов непринципиально и, напротив, требуется как можно большее разнообразие перемещений объектов по сцене, применяется подход, суть которого в следующем. Задается или случайным образом выбирается число N объектов, одновременно наблюдаемых в сцене. Формируется множество случайных положений для данных объектов

$$P = \{p_1, ..., p_N\}, \quad (1)$$

Производится устранение ситуаций взаимопроникновения объектов на основе метода потенциальных полей:

$$p_i' = p_i + \min(d_{\max}, \sum_{j=1, j\neq i}^{N} \eta / (p_j - p_i)^2),$$

где $p_i' = \{x', y', z'\}$ – обновленное положение объекта; d_{max} – максимальное смещение объектов; η- коэффициент силы отталкивания.

Для задач настройки нейросетевых классификаторов на решение конкретных прикладных задач применяется подход, основанный на загрузке заблаговременно подготовленных виртуальных сцен. Разнообразие обучающих примеров при этом достигается уже не вариацией положения предметов в сцене, а изменением ракурса наблюдения в процессе движения камеры по указанной траектории.

На первом этапе решения конкретной прикладной задачи по настройке визуального классификатора используется программное обеспечение, позволяющее формировать описания виртуальных сцен в виде

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

где o_i - программная структура, характеризующая положение, ориентацию, класс и особенности внешнего облика объекта. Обеспечивается совместимость форматов хранения описаний сцен с КСОМ для корректной загрузки виртуальной среды.

На втором этапе автоматически сформированные сцены загружаются в КСОМ: производится интерпретация текстовых описаний сцен и формирование соответствующих программных представлений для объектов, перечисленных в файле сцены. Далее производится определение достоверных результатов визуального анализа на основе прямого доступа к свойствам загруженных программных структур. Аннотации, содержащие желаемые результаты анализа типа и положения объектов, сохраняются совместно с изображениями в каталог на диске для дальнейшей настройки нейронных сетей.

Отметим, что первый этап может выполняться как с применением готовых трехмерных моделей в качестве элементов множества W, так и с применением моделей, синтезируемых процедурно. Последний подход предпочтителен в силу неограниченных возможностей по вариации параметров и, как следствие, внешнего облика данных объектов.

2.2 Процедурная генерация моделей

Процедурно генерируемые объекты задаются в виде набора параметров, что, с одной стороны, является более компактным по сравнению с явным перечислением образующих геометрических примитивов в файле трехмерной модели, а с другой – позволяет разнообразить внешний вид объекта для формирования обучающих множеств высокого качества.

В общем случае генерируемый объект задается формулой

$$o = \{l, p_1, ..., p_K\}, \quad (3)$$

Где l – класс объекта, p_i – параметры, определяемые экспертом и влияющие на геометрическую форму объекта через соотношения, заложенные в соответствующую аналитическую модель.

На рис. 2 представлены примеры процедурно синтезированных объектов для моделирования как помещений, так и открытых участков местности.

Так, например, для генерации объекта типа «Дерево» множество (3) принимает вид:

$$o = \{"Дерево", h, k, n_b, n_l\}, \quad (4)$$

Где h - высота ствола дерева, k - коэффициент толщины ветвей, n_b - число ответвлений на каждом уровне формирования дерева, n_l - число уровней рекурсивного ветвления.

Каждая ветвь формируется в виде типовой четырехгранной пирамиды, смещение и поворот которой задаются случайным образом в пределах, допускаемых родительским объектом (стволом или ветвью), а длина выбирается пропорционально расстоянию от точки позиционирования ветви до конечной вершины родительского объекта. Подобный подход позволяет генерировать широкий класс объектов, схожих с реальными деревьями различного типа. То же можно сказать и об остальных объектах, представленных на рис. 2.

Описанная возможность процедурной генерации объектов может быть использована не только для формирования обучающих множеств в задачах классификации визуальных образов, но и при настройке систем технического зрения, решающих задачи уклонения от препятствий. Наличие исчерпывающей информации о геометрии синтезированного объекта позволяет осуществить оптимальное планирование траектории вблизи него и сопоставить полученный план движения с частичной визуальной информацией, доступной на борту автономного мобильного робота.

Дополнительное увеличение числа обучающих примеров может быть получено с применением общеизвестных методов аугментации: сдвиг, поворот, отражение изображений, зашумление, размытие, а также сравнительно нового метода нейросетевой аугментации данных [13].

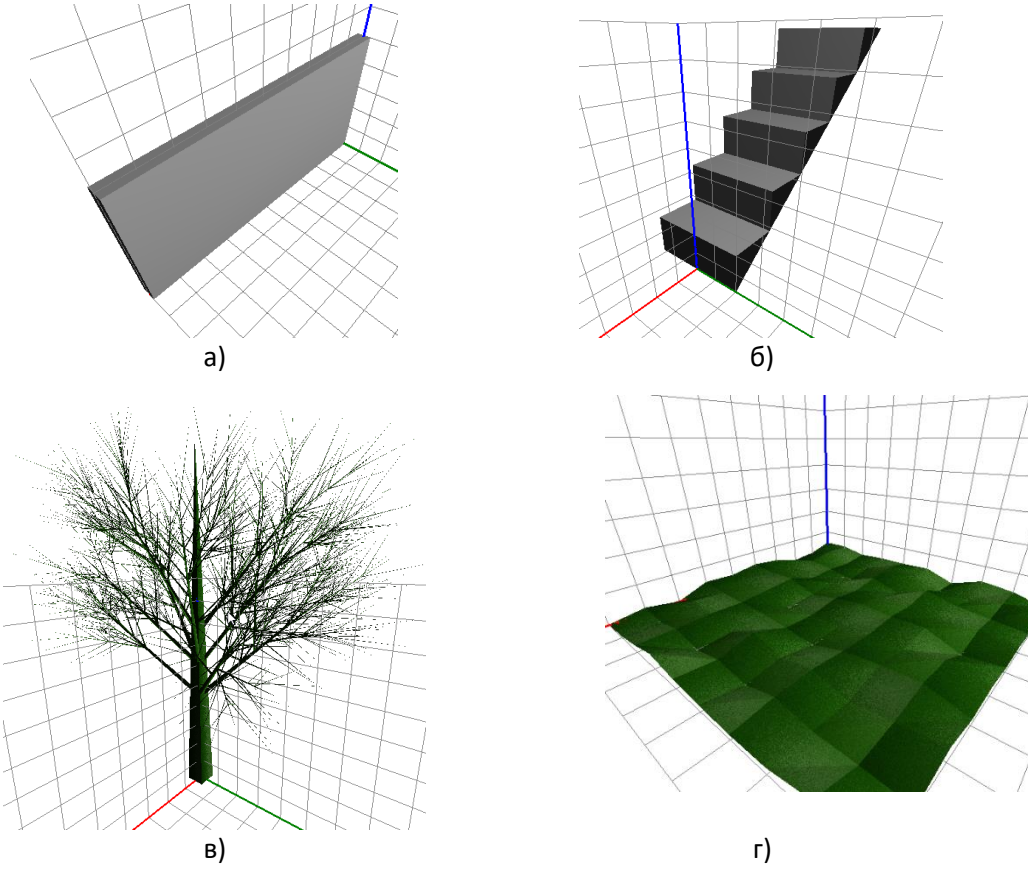

Рис. 2. Примеры автоматически сгенерированных объектов: а) стена; б) лестница; в) дерево; г) фрагмент ландшафта

2.3 Формирование обучающих множеств в задачах визуальной одометрии

Задача визуальной одометрии играет важную роль в робототехнике, когда требуется по последовательности видеокадров с бортовой камеры робота определить его перемещение в пространстве. Решение данной задачи базируется на оценке оптического потока по видеоизображению и пересчету его в параметры движения с учетом характеристик камеры, таких как угол обзора и разрешающая способность [15].

Основные сложности разработки алгоритмов визуальной одометрии заключаются в необходимости надежного детектирования ключевых точек на изображении, учета удаленности отдельных участков изображения от камеры, отделения подвижных объектов от статичного фона. В то же время нейронные сети как универсальные аппроксиматоры функций хорошо

зарекомендовали себя при решении слабо формализованных задач, что позволяет рассматривать их в качестве альтернативы аналитическим методам визуальной навигации.

На вход систем визуальной одометрии подаются два последовательных видеокадра X_{t-1}, X_t, а на выходе в общем случае формируются оценки смещений в трех направлениях и углов поворота по трем координатным осям $\{\Delta x, \Delta y, \Delta z, \Delta\alpha, \Delta\beta, \Delta\gamma\}$. Таким образом, очевидна структура нейронной сети, способной воспроизводить решение данной задачи (рис. 3).

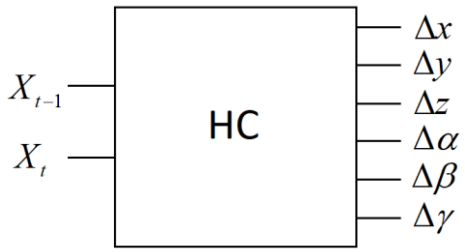

Рис. 3. Структура нейронной сети для оценки параметров визуальной навигации

Задавая местоположение и ориентацию подвижного объекта в виртуальной среде и получая программным способом изображения с бортовой видеокамеры, можно сформировать обучающую выборку требуемого размера, удовлетворяющую структуре, примеры в которой удовлетворяют структуре, представленной на рис. 3.

3 Автоматическое формирование обучающих выборок в реальной среде

Задача визуальной классификации допускает автоматическое формирование обучающих множеств не только в виртуальной среде, но и на базе реального видеопотока. Данное утверждение относится к случаю, когда известно изображение с одного из ракурсов анализируемого объекта и требуется дополнить обучающее множество визуальными образами с других ракурсов. При допущении о том, что перемещение целевого объекта относительно камеры происходит плавно, можно ожидать незначительное изменение визуального образа на соседних кадрах и возможность копирования его в обучающее множество на протяжении нескольких временных отсчетов.

Вследствие движения целевого объекта возникает неопределенность в его положении на видеокадре, что осложняет извлечение соответствующего изображения в качестве обучающего примера. Решение данной проблемы возможно с применением технологий визуального трекинга, которые позволяют с небольшими вычислительными затратами определять положение области, выбранной в поле зрения камеры ее смещении или иных трансформациях [14].

В простейшем случае мера схожести областей изображения на текущем и предыдущем кадрах может быть задана в следующем виде:

$$E(i_0, j_0) = \sum_{i=i_0}^{i_0+n-1} \sum_{j=j_0}^{j_0+m-1} (x_t(i, j) - x_{t-1}(i, j))^2, \quad (5)$$

где x_t – участок растра в момент времени t, n – ширина участка изображения, m – высота участка изображения, i_0, j_0 – начальная точка области изображения.

Цель отслеживания состоит в минимизации функционала (5) по координатам i_0, j_0.

Пополнение базы изображений может выполняться при выполнении условий:

$$\varepsilon_1 < E < \varepsilon_2, \quad (6)$$

где ε_1 – минимальный порог различия визуальных образов для предотвращения добавления в базу идентичных примеров, а ε_2 – максимальный порог различия визуальных образов для снижения вероятности добавления ложных обучающих примеров.

Наглядной иллюстрацией прикладного применения данного подхода является задача преследования цели (рис. 4). Первоначально цель идентифицируется экспертом с одного лишь ракурса, однако в процессе маневрирования изображение цели меняется, но отслеживается алгоритмом трекинга и добавляется в базу примеров с учетом условий (6). Переобучение нейросетевого классификатора на дополненную базу позволяет детектировать целевой объект даже после потери его основным алгоритмом слежения.

а) б)

Рис. 4. Отслеживание визуального образа в задаче преследования цели: а) ведомый и ведущий роботы; б) видеокадр с бортовой камеры ведомого робота

4 Экспериментальная оценка применимости технологии

Вне зависимости от выбранного способа формирования обучающего множества для решения задач визуального анализа, перечисленных в п. 1, наиболее целесообразным является применение сверточных нейронных сетей.

Важнейшим вопросом в оценке применимости развиваемой технологии является проверка способности нейронных сетей, обученных на виртуальных примерах, осуществлять визуальный анализ изображений, полученных в реальной среде функционирования автономных роботов.

При поиске ответа на данный вопрос в качестве примера была рассмотрена задача поиска человека в лесу. По результатам обучения нейронной сети YOLO v2 [1] на грубо проработанных виртуальных сценах с лесными массивами настроить нейросетевой классификатор на распознавание человека в реальных условиях леса (рис. 5).

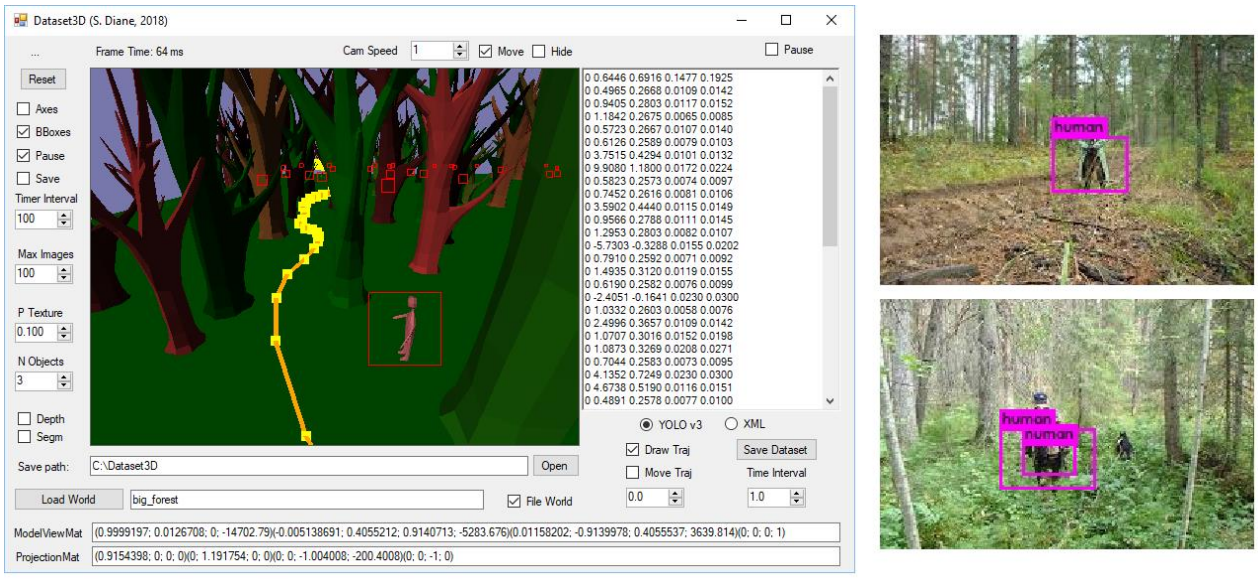

Рис. 5. Результаты моделирования виртуальных сцен (слева) и распознавания схожих объектов в реальной среде (справа)

В робототехнических приложениях, когда наряду с визуальной информацией доступны навигационные данные, результаты классификации могут быть нанесены на семантическую карту [2 -4] или же переданы напрямую оператору, контролирующему процесс функционирования робота.

Заключение

Полученные результаты подтверждают перспективность развиваемого подхода. 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.

Резюмируя вышеизложенное, можно выделить три основных способа автомтического формирования обучающих множеств: генерация примеров в виртуальной среде, аугментация выборки с применением методов обработки изображений, пополнение базы примеров в процессе слежения за целевым объектом в реальной среде.

Следует понимать, что обучение на виртуальных множествах не дает стопроцентной точности в настройке нейронных сетей под конкретную задачу. Тем не менее, дальнейшее повышение качества функционирования нейросетевых анализаторов изображений возможно путем комбинации трех перечисленных методов.

Работа выполнена при поддержке программы президиума РАН №30 "Теория и технологии многоуровневого децентрализованного группового управления в условиях конфликта и кооперации".

Литература

1. *J. Redmon, S. Divvala, R. Girshick, and A. Farhadi*, "You Only Look Once: Unified, Real-Time Object De-tection", in CVPR 2016.
2. *S. Diane, E. Lesiv, I. Pesheva, A. Neschetnaya*, Multi-Aspect Environment Mapping with a Group of Mobile Robots. 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), pp. 474-478.
3. *C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka,* Multi-hierarchical semantic maps for mobile robotics, in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems.
4. *Legovich Yu.S., Diane S.A.K., Rusakov K.D.* Integration of modern technologies for solving territory patroling problems with the use of heterogeneous autonomous robotic systems / Proceedings of the 11th International Conference "Management of Large-Scale System Development" (MLSD). Moscow: IEEE, 2018.
5. *Joseph Redmon, Anelia Angelova,* Real-Time Grasp Detection Using Convolutional Neural Networks, 2014, arXiv:1412.3128.
6. *Loquercio, Antonio & Maqueda, Ana & R. Del Blanco, Carlos & Scaramuzza, Davide. (2018).* DroNet: Learning to Fly by Driving. IEEE Robotics and Automation Letters. PP. 1-1. 10.1109/LRA.2018.2795643.
7. *Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia*, Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
8. *D. Eigen, C. Puhrsch, R. Fergus*, Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, arXiv:1406.2283, 2014
9. *R. Socher, Cliff C.-Y. Lin, A. Ng, and C. Manning.* Parsing Natural Scenes and Natural Language with Re-cursive Neural Networks. In Proc. of the 26th International Conference on Machine Learning (ICML), 2011
10. *Toshev, A. Makadia, K. Daniilidis*, Shape-based Object Recognition in Videos Using 3D Synthetic Object Models. In 2009 IEEE Conference on Computer Vision and Pattern Recognition.
11. *K. Židek, P. Lazorík, J. Pite, A. Hošovský,* An Automated Training of Deep Learning Networks by 3D Virtual Models for Object Recognition. Symmetry 2019, 11(4), 496.
12. *Дональд Херн, М. Паулин Бейкер*. Компьютерная графика и стандарт OpenGL = Computer Graphics with OpenGL. 3-е изд. - М.: Вильяме, 2005.- 1168 с.
13. *Xinyue Zhu, Yifan Liu, Zengchang Qin, Jiahong Li,* Data Augmentation in Emotion Classification Using Generative Adversarial Networks. 2017. arXiv:1711.00648
14. *L. Cehovin, A. Leonardis and M. Kristan,* Visual object tracking performance measures revisited. 2016. arXiv:1502.05803v3.
15. *S. Poddar, R. Kottath, V. Karar,* Evolution of Visual Odometry Techniques. 2018. arXiv:1804.11142.