

*Гессен Павел Алексеевич*¹,
магистр, аспирант;
*Сараджишвили Сергей Эрикович*²,
доцент, канд. техн. наук, доцент

СИСТЕМА УЛУЧШЕНИЯ ТОЧНОСТНЫХ ПОКАЗАТЕЛЕЙ ВИДЕООБРАБОТКИ КОМПЛЕКСОВ УМНОГО ДОМА

^{1,2} Россия, Санкт-Петербург, Санкт-Петербургский политехнический
университет Петра Великого;

¹ gessen.pa@edu.spbstu.ru, ² saradzh_se@spbstu.ru, ssaradg@yandex.ru

Аннотация. В целях создания системы улучшения точностных показателей устройств экосистемы умного дома с видеовходом, обеспечивающих выявление объектов в кадре, а также наблюдение за ними в процессе работы с дообучением, произведено исследование мирового опыта в данной области. Основной проблемой являются подходы, которые позволяют с наибольшей точностью, при сохранении высокой вычислительной эффективности, обеспечивать обучение классификаторов на лету без априорного знания о типе объекта наблюдения. Авторами исследования проведена разработка и полунатурное моделирование предлагаемой системы, проведена оценка эффективности ее работы при выполнении задач корректировки положения и выявления объектов с достаточными скоростными показателями с предварительным online-обучением в процессе первичной встречи цели.

Ключевые слова: дообучение, метод опорных векторов, гистограмма направленных градиентов, кластеризация.

*Pavel A. Gessen*¹,
Postgraduate Student (PhD), Master of Science (MSc);
*Sergei E. Saradgishvili*²,
Candidate of Technical Sciences (PhD), Associate Professor

A SYSTEM FOR IMPROVING THE VIDEO PROCESSING ACCURACY INDICATORS OF SMART HOME COMPLEXES

^{1,2} Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russia;

¹ gessen.pa@edu.spbstu.ru; ² saradzh_se@spbstu.ru, ssaradg@yandex.ru

Abstract. In order to create a system for improving the accuracy of smart home ecosystem devices with video input, providing the identification of objects in the frame, their positioning, as well as monitoring them in the process of working with additional training, a study of world experience in this field was carried out. The main problem is the approaches that allow to provide on-the-fly training of classifiers with the greatest accuracy without a priori knowledge of the type of object of observation, while maintaining high computational efficiency. The authors of the study carried out the development and semi-natural modeling of the proposed system, assessed the effectiveness of its work when

performing the tasks of correcting the position and identifying objects with sufficient speed indicators with preliminary online training during the initial meeting of the goal.

Keywords: further training, support vector machine, directional gradients histogram, clustering.

Введение

В настоящее время практически всеобъемлющее использование различных устройств для съемки, хранения и воспроизведения визуальной информации является неотъемлемой частью жизни. Они используются в различных областях, таких как: промышленность [13], медицина [5], транспорт [4], робототехника [11], безопасность [8].

В экосистемах умного дома уже сейчас используются приборы, основывающиеся на данных с видеовхода, и еще больше находятся в процессе реализации. Однако общемировая разработка новых алгоритмов направлена на увеличение точностных характеристик за счет использования больших вычислительных ресурсов [2, 7, 12], что может привести к несовместимости с некоторым аппаратным обеспечением из-за их ограниченных возможностей по производительности, а также экономической нецелесообразности.

Таким образом, актуальной задачей становится разработка независимой системы, которая способствовала бы повышению точности наблюдения за целью и возможности ее повторного выявления. Конструирование такой системы даст возможность адаптировать уже реализованные алгоритмы со скоростным заделом к современным стандартам качества.

1. Постановка задачи

1.1. Цель и задачи работы

Цель работы — увеличить точность наблюдения объекта для существующих алгоритмов наблюдения и добавить механизм выявления цели после потери, если такой отсутствует, или добавить уточнение при его наличии.

Задачи, необходимые для достижения цели:

- а) Исследовать существующие алгоритмы и методы по наблюдению и выявлению целей в видеопотоке и провести их сравнительный анализ.
- б) Предложить метод по улучшению точностных характеристик существующих алгоритмов.
- в) Реализовать предложенный метод.
- г) Показать эффективность применения предложенного метода и его реализации на практике.

2. Исследование предметной области

Рассмотрим отдельно выявление объекта и наблюдение за объектом. В обоих случаях выделим самые популярные существующие подходы.

Данные взяты из обзорных статей и бенчмарков, поэтому стоит учитывать вычислитель, на котором они были запущены, для более адекватной оценки применимой к области исследования. Основными характеристиками для оценивания будут скорость работы алгоритмов и их показатель точности, который будет определен позднее непосредственно для каждого алгоритма.

2.1. Выявление объекта

Для классических алгоритмов мерой оценки будет показатель точности (Accuracy) — средняя вероятность корректно обнаруженных объектов, чем выше, тем лучше. Для нейросетевых алгоритмов используется показатель средней точности (AP — Average precision) — минимальное значение меры IoU для принятия решения об успешном обнаружении. Для алгоритма вычитания фона не получилось найти сведений в представленных метриках, потому приведена мера оценки из оригинальной статьи. Набор данных для нейросетевых алгоритмов — COCO2017. Полученные результаты представлены в таблице 1.

Таблица 1

Характеристики алгоритмов обнаружения

Алгоритм	Вычислитель	Скорость, fps	Оценка (F-measure)
Viola-Jones	700 Mhz Pentium III processor	14.90	90.00
ORB	1GHz ARM chip and 512 MB of RAM	65.30–15.00	78.00–98.00
Template Matching	Xeon 2.4-GHz	250.00–0.05	62.00–98.00
Contour Matching	CPU	6.00–0.90	76.00–92.00
Background subtraction	Raspberry Pi 2	0.066–0.040	0.057–0.084
ConvNeXt V2	GPU	25.60–8.60	51.00–55.70
YOLOv7	GPU	286.00–36.00	38.70–56.80
Mask R-CNN	GPU	6.00–5.00	42.10–42.70
RetinaNet	GPU	14.00–5.00	39.10–40.80
SSD	GPU	17.00–7.00	31.20–33.20

2.2. Наблюдение за объектом

Для классических алгоритмов мерой оценки будет кривая точности [3]. Набор данных для тестирования детерминированных алгоритмов — MIT. Для нейросетевых алгоритмов будет использована мера оценки качества наблюдения за множеством объектов (MOTA). Набор данных

для тестирования нейросетевых алгоритмов — MOT17. Полученные результаты представлены в таблице 2.

Таблица 2

Характеристики алгоритмов сопровождения

Алгоритм	Вычислитель	Скорость, fps	Оценка, %
TLD	CPU	28.00	60.80
KCF	CPU	292.00–172.00	73.20–72.80
DSST	CPU	54.30	80.20
CamShift	2GHz PC	0.70	68.70
CMT	CPU	8.28	67.80
STRUCK	CPU	20.00	65.60
MOSSE	CPU	615.00	43.10
MIL	CPU	38.00	47.50
StrongSORT	GPU	7.50	79.60
FairMOT	GPU	25.90	73.70
BYTETrack	GPU	29.60–11.80	77.30
SMILEtrack	GPU	9.60	81.06

2.3. Выводы

Нейросетевые алгоритмы лучше себя показывают в плане точности, однако сильно проигрывают классическим по скорости, при этом у последних порой есть избыточный задел по производительности, который можно использовать, для улучшения их точностных характеристик.

3. Предлагаемая система

Предлагаемая система работает в совокупности с каким-то существующим алгоритмом наблюдения и состоит из нескольких частей.

Обучение. На первых кадрах происходит обучение SVM-классификатора [10] за счет сбора HOG-дескрипторов [3] с текущего положения объекта (отклик алгоритма слежения принимается за истинное положение объекта).

Корректировка. Уточнение положения объекта в кадре происходит на основе последовательности откликов обученного классификатора. На протяжении всего времени слежения предлагаемая система осуществляет проверку корректности и точности описания истинного положения объекта в кадре. Для этого используются обученный классификатор и метод кластеризации ключевых точек. На каждом кадре после кластеризации находится такая описывающая рамка, которая лучшим образом описывает объект. После этого проводится проверка параметра пересечения рамки слежения и вышеописанного прямоугольника, а результат сохраняется. Если на протяжении нескольких последовательных кадров данный показатель ухудшается, то за рамку сопровождения принимается отклик описывающего прямоугольника кластеризации. При отсутствии

рамки положительного отклика от алгоритма кластеризации и классификатора в предполагаемом месте истинного положения объекта на протяжении некоторого времени происходит пересчет параметров кластеризации, описанный в начале предыдущего подраздела, до тех пор, пока не будет найдена новая комбинация параметров. Данный подход обеспечивает масштабную инвариантность.

Выявление. Выявление объекта после потери во многом схоже с методом корректировки, описанным выше: на каждом кадре собирается множество точек и разбивается на кластеры в соответствии с параметрами, которые были на момент потери. Изображения внутри описывающих рамок полученных кластеров в виде HOG-дескрипторов передаются для проверки классификатору. Данная последовательность действий повторяется на нескольких соседних кадрах, после чего принимается решение о наличии объекта на основе множественных положительных откликов от SVM в одной области. При отсутствии положительных откликов на протяжении нескольких кадров или же положительных откликов, которые не воспроизводятся в схожих местах последовательно, происходит пересчет параметров кластеризации, и вышеописанные действия повторяются для новых значений.

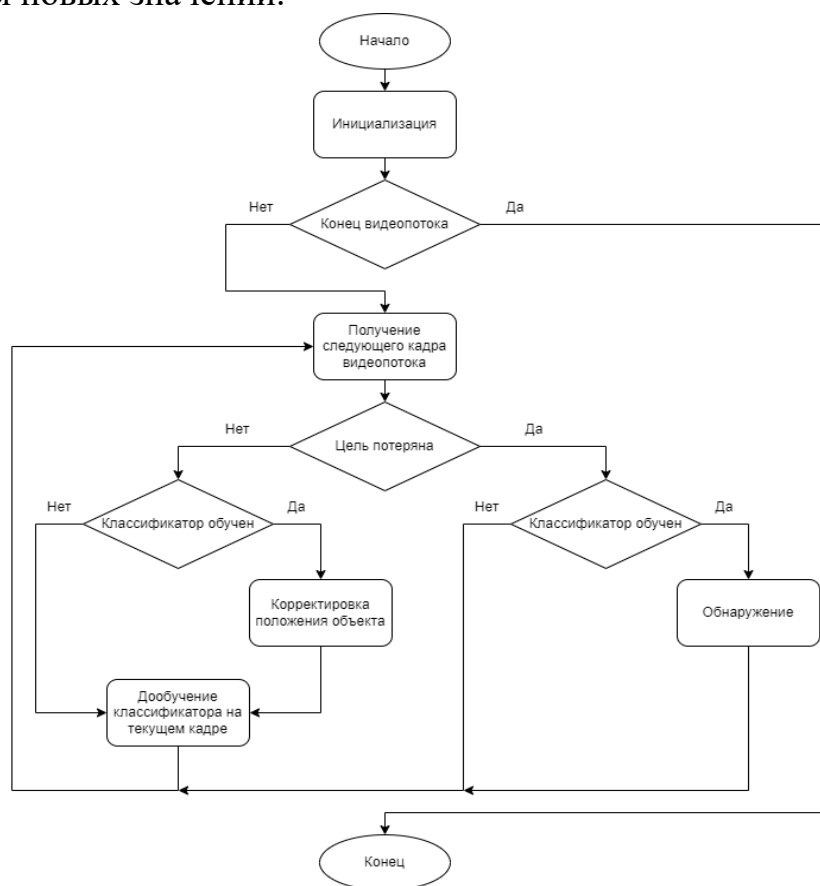


Рис. 1. Схема алгоритма предлагаемой системы

4. Результаты

4.1. Наблюдение

Для правильной и наиболее полной оценки качества, производимых предложенной системой улучшений на протяжении наблюдения и корректировки положения описывающего прямоугольника необходимо и достаточно на основании собранных показателей провести оценку метрики IoU (Intersection over Union) [9]. Подход к проведению тестирования для оценки результатов:

1. Выбор основного алгоритма наблюдения.
2. Использование тестовой среды для оценки его характеристик.
3. Добавление к выбранному алгоритму предлагаемой системы.
4. Повторный запуск полученного комплекса и сбор метрик.
5. Анализ результатов первого и второго запусков.

В качестве улучшаемых алгоритмов сопровождения были выбраны 9 различных реализаций, которые представляют интерес для решения задач в условиях ограничений, а данные по их работе без и с использованием системы представлены на рисунках ниже.



Рис. 2. Графики значений IoU. Синим цветом обозначен оригинальный алгоритм, зеленым — комбинированный с предлагаемой системой

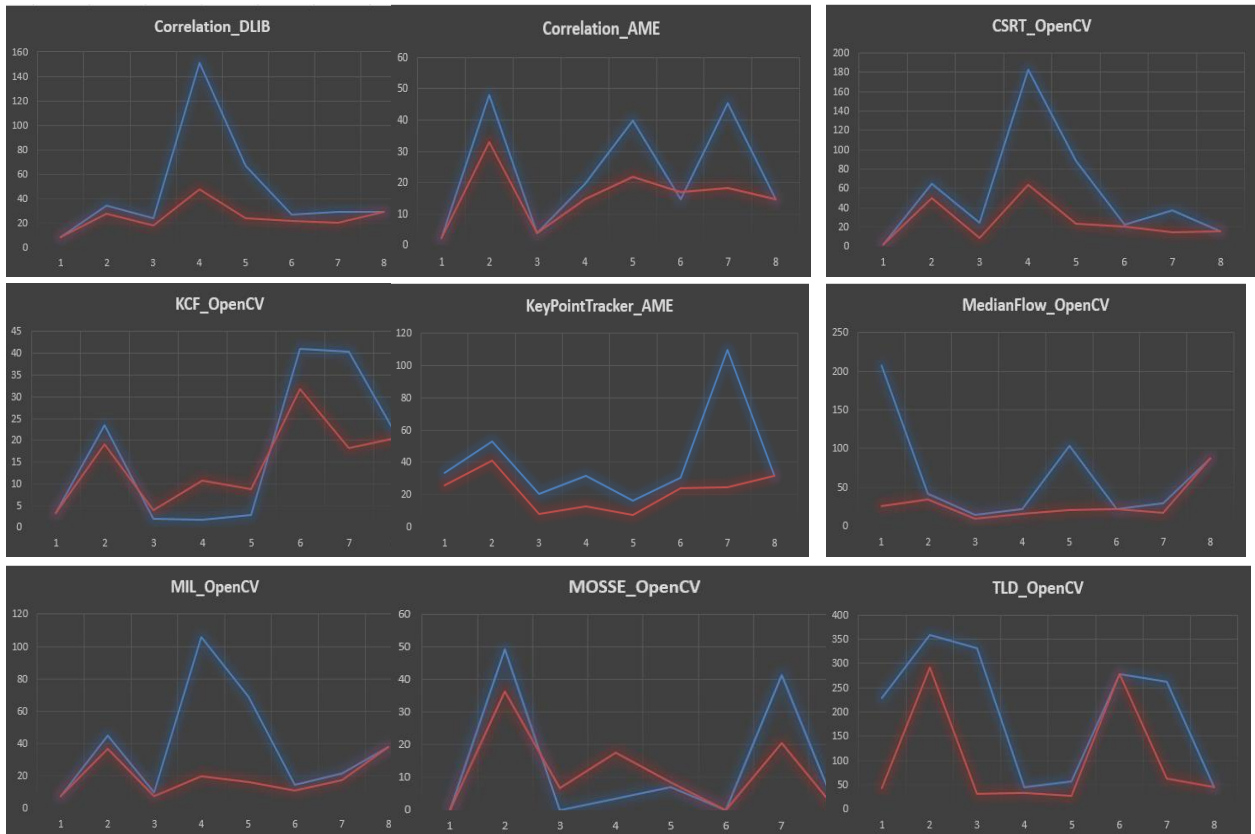


Рис. 3. Графики удаленности центров. Синим цветом обозначен оригинальный алгоритм, красным — комбинированный с предлагаемой системой

4.2. Выявление

Для оценки данного аспекта будет использована параметрическая F-мера (F-score [6]) и многоагентный алгоритм [1].

Подход к проведению тестирования для оценки результатов:

1. Выбор видеозаписи для проверки.
2. Обучение классификатора на протяжении 100-400 кадров при помощи размеченного истинного положения объекта.
3. Имитация срыва наблюдения путем перехода на другой кадр, разделенный с последним кадром обучения не менее, чем на 100 фреймов.
4. Сбор фрагментов кадра двумя способами: метод скользящих окон разного масштаба и предложенный метод кластеризации.
5. Проверка полученных описывающих прямоугольников классификатором.
6. Вычисление F-меры, IoU и разницы дистанции центров.



Рис. 4. Графики IoU для детекции многоагентного алгоритма (синий) и предлагаемой системы (оранжевый)

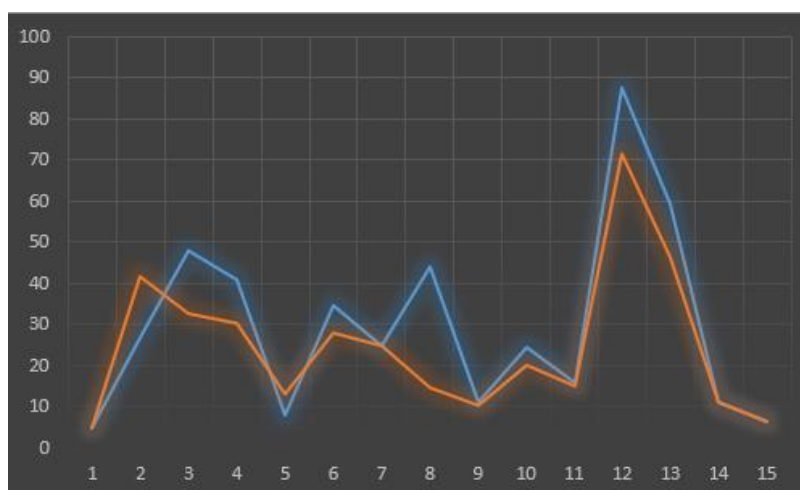


Рис. 5. Графики разницы центров для детекции многоагентного алгоритма (синий) и предлагаемой системы (оранжевый)

4.3. Выводы

Анализируя приведенные результаты наблюдения, можно заметить, что в среднем по всем алгоритмам произошло улучшение показателя IoU на протяжении времени на $\sim 40\%$ (чем выше зеленая линия над синей, тем лучше), а приближение центра к истинному на $\sim 50\%$ (чем ниже красная линия под синей, тем лучше).

Анализируя приведенные результаты по выявлению, получаем среднее значение IoU ~ 0.5 и дистанции до центра ~ 24.6 . Данные показатели свидетельствуют о достаточной точности локализации объекта после его обнаружения (keypoint detector — 0.458 и 29.8, tld — 0.41 и 30.2). Анализ показателей F-меры для кластеризации показывает среднее значение в $\sim 75\%$, что несколько хуже аналогичного для скользящего окна $\sim 83\%$.

Заключение

В работе были изучены существующие подходы и реализации задач обработки видеоданных. Была предложена и реализована система улучшения точностных характеристик устройств умного дома. Проведены оценки качества работы предлагаемой системы.

Список литературы

1. Бондаренко В.А., Ельцова Д.К., Лизин А.И., Павлова В.А., Созинова М.В., Тупиков В.А. Многоагентный алгоритм автоматического обнаружения и сопровождения недетерминированных объектов // Известия ЮФУ. Технические науки. – 2020. – № 1(211). – С. 218–232.
2. Bertinetto L., Valmadre J., Henriques J.F., Vedaldi A., Torr P.H.S. Fully-convolutional siamese networks for object tracking. // ECCV 2016. Lecture Notes in Computer Science. – Vol. 9914. – Cham: Springer, 2016.
3. Dalal N., Triggs B. Histograms of oriented gradients for human detection // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA. – 2005. – Vol. 1. – Pp. 886–893. – DOI: 10.1109/CVPR.2005.177.
4. Dilek E., Dener M. Computer vision applications in intelligent transportation systems: a survey // Sensors. – 2023. – Vol. 23(6). – P. 2938. – DOI: 10.3390/s23062938.
5. Gao Junfeng, Yang Yong, Lin Pan, Park Dong. Computer vision in healthcare applications // Journal of Healthcare Engineering. – 2018. – Pp. 1–4. – DOI: 10.1155/2018/5157020.
6. Goutte C., Gaussier E. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation // European Conference on Information Retrieval. – Springer, 2005. – Pp. 345–359.
7. Li D., Yu Y., Chen X. Object tracking framework with siamese network and re-detection mechanism // J Wireless Com Network, 2019. – P. 261.
8. Rajib Debnath, Mrinal Kanti Bhowmik. A comprehensive survey on computer vision based concepts, methodologies, analysis and applications for automatic gun/knife detection // Journal of Visual Communication and Image Representation. – 2021. – Vol. 78. – Paper id 103165, ISSN 1047-3203. – DOI: 10.1016/j.jvcir.2021.103165.
9. Rezafofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I. Savarese S., Generalized intersection over union: a metric and a loss for bounding box regression // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, 2019. – Pp. 658–666.
10. Shalev-Shwartz S., Singer Y., Srebro N. Pegasos: primal estimated sub-gradient solver for SVM // Math. Program. – 2020. – № 127. – Pp. 3–30.
11. Strat T., Chellappa R., Patel V. Vision and robotics // AI Magazine. – 2020. – Vol. 41, no. 2. – Pp. 49–65. – DOI: 10.1609/aimag.v41i2.5299.
12. Zhang Y., Wang L., Qi J., Wang D., Feng M., Lu H. Structured siamese network for real-time visual tracking // Computer Vision – ECCV 2018. Lecture Notes in Computer Science. Vol 11213. – Cham: Springer, 2018.
13. Zhou L., Zhang L., Konz N. Computer vision techniques in manufacturing // IEEE Transactions on Systems, Man, and Cybernetics: Systems. – 2023. – Vol. 53, no. 1 (Jan 2023). – Pp. 105–117. – DOI: 10.1109/TSMC.2022.3166397.