



УДК 004.93

ПРИМЕНЕНИЕ БИНАРНЫХ ДЕСКРИПТОРОВ ДЛЯ ТРЕКИНГА МНОЖЕСТВА ЛИЦ В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ

А.Л. Олейник^а^а Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: andrey_oleynik@niuitmo.ru

Информация о статье

Поступила в редакцию 19.05.16, принята к печати 20.06.16

doi: 10.17586/2226-1494-2016-16-4-670-677

Язык статьи – русский

Ссылка для цитирования: Олейник А.Л. Применение бинарных дескрипторов для трекинга множества лиц в системах видеонаблюдения // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 4. С. 670–677.
doi: 10.17586/2226-1494-2016-16-4-670-677

Аннотация

Предмет исследования. Рассмотрена задача трекинга (отслеживания) множества лиц в видеопотоке. Разработана система трекинга, ориентированная на решение задач автоматизированного видеонаблюдения. При этом специфика условий работы камер видеонаблюдения использована для повышения эффективности системы по сравнению с аналогами общего назначения. **Метод.** Разработанная система состоит из двух подсистем: детектора и трекера. Трекер не зависит от детектора, что позволяет использовать различные методы детектирования лиц. Кроме того, такая структура позволяет обрабатывать детектором только малую часть кадров, существенно повышая скорость работы. Алгоритм трекинга основан на бинарных дескрипторах BRIEF, которые крайне эффективно вычисляются на современных процессорных архитектурах. **Основные результаты.** Система реализована на языке C++, проведены экспериментальные исследования по оценке скорости и качества ее работы. Для измерения качества трекинга использованы метрики МОТА и МОТР. Полученные результаты говорят о четырехкратном увеличении быстродействия по сравнению с базовой реализацией, запускающей детектор на каждом кадре. При этом качество трекинга при сопоставлении с базовой реализацией находится на допустимом уровне. **Практическая значимость.** Разработанная система может быть использована с различными детекторами лиц (в том числе достаточно медленными) для получения полнофункциональной высокоскоростной системы трекинга множества лиц. Алгоритм прост в реализации и оптимизации, поэтому он может быть использован не только в полномасштабных системах видеонаблюдения, но и во встроенных решениях, интегрированных непосредственно в камеры.

Ключевые слова

трекинг множества лиц, бинарные дескрипторы, видеонаблюдение, компьютерное зрение

Благодарности

Работа выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01). Автор выражает искреннюю признательность научному руководителю профессору Г.А. Кухареву, заведующему кафедрой РИС Ю.Н. Матвееву и А.А. Мельникову за критические замечания и советы, которые помогли значительно улучшить качество настоящей статьи.

APPLICATION OF BINARY DESCRIPTORS TO MULTIPLE FACE TRACKING IN VIDEO SURVEILLANCE SYSTEMS

A.L. Oleinik^а^а ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: andrey_oleynik@niuitmo.ru

Article info

Received 19.05.16, accepted 20.06.16

doi: 10.17586/2226-1494-2016-16-4-670-677

Article in Russian

For citation: Oleinik A.L. Application of binary descriptors to multiple face tracking in video surveillance systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 4, pp. 670–677. doi: 10.17586/2226-1494-2016-16-4-670-677

Abstract

Subject of Research. The paper deals with the problem of multiple face tracking in a video stream. The primary application of the implemented tracking system is the automatic video surveillance. The particular operating conditions of surveillance cameras are taken into account in order to increase the efficiency of the system in comparison to existing general-purpose

analog. **Method.** The developed system is comprised of two subsystems: *detector* and *tracker*. The tracking subsystem does not depend on the detector, and thus various face detection methods can be used. Furthermore, only a small portion of frames is processed by the detector in this structure, substantially improving the operation rate. The tracking algorithm is based on BRIEF binary descriptors that are computed very efficiently on modern processor architectures. **Main Results.** The system is implemented in C++ and the experiments on the processing rate and quality evaluation are carried out. MOTA and MOTP metrics are used for tracking quality measurement. The experiments demonstrated the four-fold processing rate gain in comparison to the baseline implementation that processes every video frame with the detector. The tracking quality is on the adequate level when compared to the baseline. **Practical Relevance.** The developed system can be used with various face detectors (including slow ones) to create a fully functional high-speed multiple face tracking solution. The algorithm is easy to implement and optimize, so it may be applied not only in full-scale video surveillance systems, but also in embedded solutions integrated directly into cameras.

Keywords

multiple face tracking, binary descriptors, video surveillance, computer vision

Acknowledgements

This work was partially financially supported by the Government of the Russian Federation, Grant 074-U01. The author expresses his sincere appreciation to Professor Georgy Kukharev, his scientific adviser; Yuri Matveev, Head of SIS Department, and Alexander Melnikov for their critical remarks and advice that significantly improved this paper.

Введение

Трекинг (отслеживание) объектов – это сложная и практически значимая задача компьютерного зрения. Разработка компактных, мощных и относительно недорогих устройств записи и обработки видеoinформации способствует постоянному расширению области применения систем трекинга, которая включает видеонаблюдение, различные мобильные приложения, дополненную реальность, наблюдение за дорожным движением, аудиовизуальные приложения [1], робототехнику. Такое разнообразие возможных областей применения означает, что трекинг объектов представляет собой целый класс задач с различными требованиями к конкретной системе. Таким образом, существование универсального метода трекинга маловероятно.

На сегодняшний день разработано значительное количество алгоритмов трекинга. Подробные обзоры на данную тему представлены в [2, 3], а в [4] проведены экспериментальные исследования по измерению и сравнению характеристик существующих подходов. Кроме того, результаты тестирования множества систем трекинга отражены в итогах конкурса Visual Object Tracking 2015 Challenge [5].

Большая часть существующих методов трекинга лиц основана либо на построении модели внешнего вида отслеживаемого объекта или фона, либо на предсказании траектории движения.

Многие алгоритмы трекинга предполагают построение описания внешнего вида объекта, которое используется для оценки его положения. В зависимости от конкретной области применения, это описание может быть статическим или динамическим. Статическое представление полезно только в том случае, когда объект остается неизменным в процессе трекинга. Динамические модели внешнего вида свободны от этого недостатка, однако подвержены проблеме ускользания трека [6], которая проявляется в постепенном накоплении ошибки трекинга. Для решения проблемы ускользания в условиях изменчивости внешнего вида объекта некоторые авторы используют гибридные представления [7].

Некоторые подходы также применяют описание фона для улучшения качества трекинга. Как правило, в таком случае для отделения отслеживаемого объекта от фона используется бинарный классификатор. Такой подход лежит в основе алгоритмов Ensemble Tracking [8] и Tracking-Learning-Detection (TLD) [9].

Ряд методов трекинга включает в себя модель движения объекта. Данный подход представлен алгоритмами на основе фильтра Калмана [10] и фильтров частиц (Particle Filters) [11, 12].

Большинство существующих систем трекинга в явной или неявной форме используют комбинацию перечисленных подходов. Например, алгоритмы на базе широко известного метода Лукаса–Канаде (Lukas-Kanade) [13, 14] включают как модель внешнего вида объекта, так и модель движения.

В настоящей работе представлена система трекинга множества лиц, предназначенная для решения задач видеонаблюдения. Существует несколько трудностей, характерных для данной области применения. Прежде всего, имеет место упомянутая выше проблема ускользания трека. Кроме того, в условиях сценария трекинга множества лиц возможны ошибки несоответствия, выражающиеся в том, что трекер путает лица с пересекающимися траекториями. И наконец, постоянная необходимость детектирования новых лиц, попадающих в поле зрения камеры, может повлечь существенное ухудшение быстродействия системы трекинга.

С другой стороны, специфические особенности условий работы камер видеонаблюдения могут быть использованы для повышения эффективности систем трекинга. Например, имеет смысл считать лица жесткими объектами и для построения признакового описания лица использовать простые алгоритмы, не требующие значительных вычислительных затрат. Кроме того, детектор лиц может быть использован не только для поиска новых лиц в поле зрения камеры, но и для коррекции ошибок. При этом использование методов трекинга позволяет запускать детектор не на каждом кадре, существенно повышая скорость обработки видеопотока.

Представленная система состоит из детектора лиц и подсистемы трекинга и отличается простотой и эффективностью. Алгоритм трекинга основан на бинарных дескрипторах. На сегодняшний день предложено несколько подходов, основанных на дескрипторах [7, 15–17]. В данной работе применяются дескрипторы BRIEF (Binary Robust Independent Elementary Features) [18], основная особенность которых заключается в высокой вычислительной эффективности. Этот метод, в отличие от представленного в [19], не требует специального оборудования (например, высокопроизводительных графических процессоров) для высокоэффективной реализации.

Метод трекинга с использованием BRIEF, представленный в [7], основан на комбинации статических и динамических словарей. Подход, предложенный в данной работе, не подразумевает применения подобных словарей и использует BRIEF исключительно для оценки межкадрового перемещения лиц.

Постановка задачи

В общем случае задача трекинга множества объектов ставится следующим образом. Для каждого кадра входного видеопотока (как правило, в реальном времени) для каждого отслеживаемого объекта (лица, автомобиля, пешехода и т.д.) указывается его положение (прямоугольник, эллипс, центроид...), идентификатор и, опционально, уровень доверия. Здесь и далее под треком понимается набор положений с одинаковым идентификатором. В зависимости от конкретной задачи присвоение нового трека объекту, который на время скрылся из поля зрения, может считаться как ошибкой, так и нормальным поведением. В рамках предлагаемого подхода такие объекты всегда получают новые идентификаторы (т.е. новые треки). В случае необходимости для объединения треков могут применяться простые методы идентификации по лицу [20].

Рассматриваемый сценарий трекинга множества лиц является частным случаем представленной выше задачи. Таким образом, могут быть введены дополнительные предположения:

- камера установлена стационарно; ее фокусное расстояние и угол обзора неизменны и заранее известны;
- отсутствуют значительные нелинейные искажения кадров видеопотока;
- максимальная скорость движения лиц не превосходит некоторого известного значения;
- отклонения положений лиц от фронтального незначительны.

Эти ограничения позволяют использовать простые методы обработки изображений для трекинга лиц. Например, в данном случае отсутствует необходимость в использовании инвариантных к вращению признаков. Ограничения на максимальную скорость движения лиц позволяют сузить область поиска нового положения лица.

Структура системы

Разработанная система состоит из двух основных подсистем: детектора и трекера (рис. 1). Детектор обрабатывает относительно малую часть кадров видеопотока, на которых создаются новые треки и корректируются существующие. Остальные кадры обрабатываются трекером с использованием алгоритма, не требующего значительных вычислительных затрат.

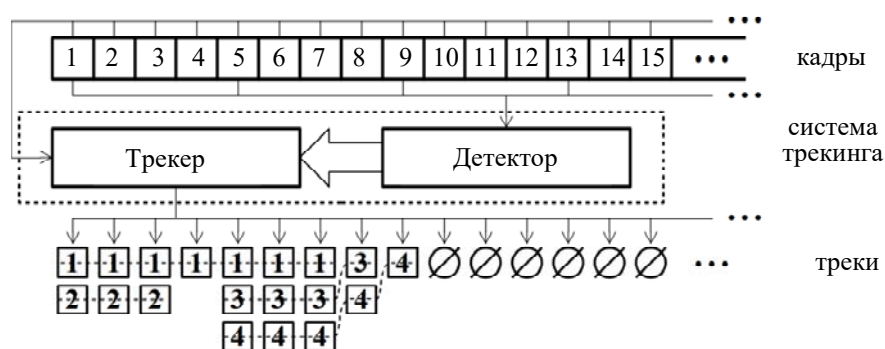


Рис. 1. Структура системы трекинга

Замечание. В настоящей работе термин «трекер» обозначает подсистему, отвечающую исключительно за поддержание треков, в то время как термин «система трекинга» соответствует комбинации детектора и трекера.

Основным преимуществом представленной на рис. 1 структуры является гибкость, так как трекер не зависит от алгоритма детектирования. Кроме того, возможность выбора доли кадров, обрабатываемых детектором, позволяет достигнуть компромисса между скоростью и качеством работы. При этом детектор может быть использован для коррекции треков, что существенно смягчает проблему их ускользания. В представленной системе используется детектор лиц, основанный на каскадных классификаторах [21]. В общем случае может быть использован любой детектор, например, один из представленных в [20].

Трекер выполняет три основные функции:

1. создание и обновление треков;
2. удаление треков;
3. межкадровое поддержание треков.

При этом удаление и межкадровое поддержание треков выполняется на каждом кадре, в то время как их создание и обновление происходит только на тех кадрах, которые были обработаны детектором лиц.

Создание и обновление треков. Прежде всего, трекер оценивает новые положения уже отслеживаемых лиц и удаляет устаревшие треки (см. ниже). Далее обнаруженные детектором лица используются для создания новых треков и обновления существующих. Положения обнаруженных лиц сравниваются с имеющимися треками, в результате чего составляется взвешенный двудольный граф (рис. 2), причем ребра соединяют только те пары лицо–трек, относительная площадь пересечения которых превосходит некоторый заранее заданный порог. Затем строится паросочетание с максимальным количеством ребер и минимальной суммарной стоимостью (рис. 2). Поиск такого паросочетания сводится к решению задачи нахождение потока минимальной стоимости (minimum cost flow) [22]. В большинстве случаев такой подход позволяет корректно разрешать неоднозначные соответствия треков и обнаруженных лиц (например, когда траектории двух лиц пересекаются). Положения треков, которым соответствуют обнаруженные лица, корректируются. Лица, которым не отвечает ни один трек, используются для создания новых треков. Треки, которым не соответствует лицо, не подвергаются модификации.

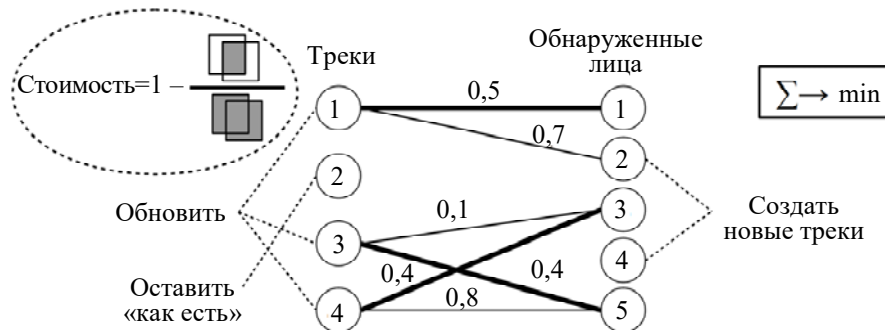


Рис. 2. Сопоставление обнаруженных лиц и существующих треков (пример). Значения стоимости (числа над ребрами) формируются на основе относительной площади пересечения положений лиц, указанных детектором и трекером

Удаление треков. Для предотвращения ускользания треков важно вовремя их удалять. Для того чтобы контролировать накопление ошибки, вводится понятие накопительного уровня доверия C_k .

Допустим, что на k -м кадре известен уровень доверия $c_k \in (0,1]$ (способ его оценки обсуждается ниже). Тогда накопительный уровень доверия C_k определяется как произведение

$$C_k = \prod_{g=0}^G c_{k-g}.$$

Здесь $G \geq 0$ – возраст трека с момента его последнего обновления или создания.

Когда C_k опускается ниже заданного значения, соответствующий трек удаляется. В момент создания или обновления трека значения c_k и C_k устанавливаются в 1.

При постепенном накоплении ошибки трекинга значение C_k будет монотонно убывать и, в конечном счете, опустится ниже порогового уровня. Таким образом, данный подход предотвращает ускользание треков. С другой стороны, в условиях сценария отслеживания множества лиц треки периодически обновляются с помощью детектора лиц. Это предотвращает ложное удаление треков, вызываемое изменениями ориентации и освещения отслеживаемых лиц.

Межкадровое поддержание треков. Оценка положения лица на очередном кадре является важнейшей составляющей алгоритма трекинга. В следующем разделе представлен метод трекинга на основе бинарных дескрипторов.

Метод трекинга лиц на основе бинарных дескрипторов

В разделе рассматривается метод, позволяющий оценить положение лица и уровень доверия c_k на текущем кадре ($\text{№ } k$), основываясь на известном положении этого лица на предыдущем кадре ($\text{№ } k - 1$). Метод использует бинарные дескрипторы – локальные признаки, представленные в форме битовых массивов. Они вычисляются в ключевых точках, которые находятся с помощью детектора ключевых точек. В качестве меры различия бинарных дескрипторов используется расстояние Хемминга. В работе применя-

ется детектор ключевых точек FAST (Features from Accelerated Segment Test) [23] и бинарные дескрипторы BRIEF (Binary Robust Independent Elementary Features) [18].

Ниже приведена вычислительная процедура, реализующая данный подход.

1. В области лица на предыдущем кадре находятся ключевые точки и вычисляются дескрипторы;
2. Исходя из предположений, введенных в разделе «Постановка задачи», область лица с предыдущего кадра расширяется (рис. 3). В этой расширенной области также находятся ключевые точки и вычисляются дескрипторы;
3. Выполняется попарное сравнение дескрипторов из предыдущего и текущего кадров. Пары с расстоянием Хемминга меньше заданного порога T_H сохраняются и используются далее;
4. Вычисляется новое положение лица.

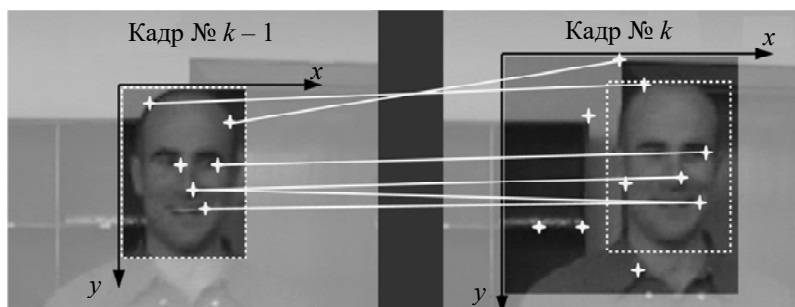


Рис. 3. Поиск нового положения лица

Рассмотрим последний этап более подробно. После сравнения дескрипторов формируется набор сопоставлений: $(p_n, p'_n, w_n); n = 1 \dots N$. Здесь $p_n = (x_n, y_n)$, $p'_n = (x'_n, y'_n)$ – координаты ключевых точек с предыдущего и текущего кадров соответственно. Веса $w_n = [w_{\min}, 1]$ вычисляются путем преобразования соответствующих значений расстояния Хемминга, причем нулевое расстояние соответствует $w_n = 1$, а расстояние T_H – в $w_n = w_{\min}$.

Принимая во внимание введенные выше предположения относительно рассматриваемого сценария трекинга, в модель движения включаются две составляющие: сдвиг $(\delta x, \delta y)$ и *равномерный масштаб* s . Оценка данных параметров сводится к следующей оптимизационной задаче:

$$\theta = (\delta x, \delta y, s) = \arg \min_{\theta} L_w(\theta),$$

$$L_w(\theta) = \sum_{n=1}^N w_n \|f(p_n, \theta) - p'_n\|^2,$$

$$f(p, \theta) = s \cdot p + (\delta x, \delta y).$$

Так как преобразование $f(p, \theta)$ является линейной функцией, искомые параметры θ могут быть получены как решение системы трех линейных уравнений [24]. Полученное решение θ однозначно определяет положение лица на текущем кадре.

Минимальное значение целевой функции можно использовать для формирования оценки уровня доверия c_k :

$$c_k(\theta) = \frac{1}{1 + \alpha \cdot \min_{\theta} L_w(\theta)} \in (0, 1].$$

Здесь α – экспериментально подбираемый параметр.

Так как модель движения обладает всего тремя степенями свободы, вычислительная процедура проста и устойчива. С другой стороны, из-за использования простых признаков (BRIEF) может возникнуть множество некорректных сопоставлений (выбросов). Возможные пути решения данной проблемы включают достаточно низкий (т.е. жесткий) порог T_H , модификацию весов w_n (например, с учетом расстояния соответствующих ключевых точек от центра области лица), а также использование робастных методов оценки параметров, таких как, например, RANSAC (RANdom SAmple Consensus) [24].

Экспериментальные исследования

Целью экспериментальных исследований является оценка скорости работы системы и качества трекинга. Предполагается, что предложенное решение позволяет значительно увеличить скорость обработки видеопотока без существенного снижения качества трекинга.

Представленная система реализована на языке C++ с использованием библиотек OpenCV [25] и LEMON [26]. Каскадный детектор из OpenCV обучен с параметрами, подобранными для высокой скорости обработки и достаточного качества обнаружения.

К сожалению, большинство существующих тестовых баз были разработаны для сценария трекинга пешеходов или автотранспортных средств (например, PETS [27] или VIVID [28]). В связи с этим они не подходят для оценки разработанной системы трекинга лиц.

Для оценки характеристик разработанной системы использованы две видеопоследовательности, имеющие разрешение 352×288 точек (рис. 4). Видео 1 является частью базы SPEVI [12] (рис. 4, а). Видео 2 представляет собой черно-белую запись с камеры видеонаблюдения (рис. 4, б). Для создания разметки использована программа ViPER-GT [29].

Повышение скорости работы, достигаемое за счет использования трекера, зависит от количества кадров, обрабатываемых детектором (рис. 1). Чем больше кадров приходится на один запуск детектора (период запуска детектора, обозначенный символом T_{det}), тем выше скорость обработки кадров всей системой. На рис. 5 показана экспериментальная зависимость между средним временем обработки одного кадра и периодом запуска детектора T_{det} . Экспериментально полученные результаты говорят о значительном (четырёхкратном) повышении скорости работы системы по сравнению с базовой реализацией.



Рис. 4. Используемые для оценки системы видеопоследовательности: запись «motinas_multi_face_frontal» из базы SPEVI (Видео 1) (а) и черно-белая запись с камеры видеонаблюдения (Видео 2) (б)

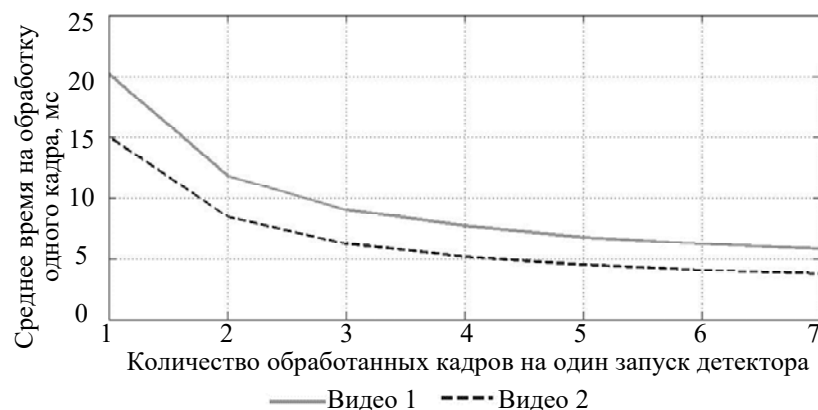


Рис. 5. Экспериментально полученная зависимость среднего времени обработки кадра от периода запуска детектора T_{det}

Для оценки качества трекинга использованы метрики Multiple Object Target Accuracy (MOTA) и Multiple Object Tracking Precision (MOTP) [30]. Метрика MOTA основана на частоте пропусков лиц (*miss*), частоте ложных срабатываний (*fp*) и частоте ошибок назначения идентификатора (*mme*). Метрика MOTP используется для оценки точности позиционирования отслеживаемых лиц. Разработанная система сравнивается с базовой реализацией, которая использует детектор лиц на каждом кадре. Результаты сравнения представлены в таблице.

Чем больше значение T_{det} , тем меньше общее количество отслеживаемых объектов. Иными словами, большие значения периода запуска детектора увеличивают значения *miss* и уменьшают значения *fp* и *mme*. В предельном случае ни один объект не будет отслеживаться (*miss* = 100%, *fp* = *mme* = 0%). Таким образом, для сравнительного анализа систем трекинга следует использовать совокупные метрики MOTA и MOTP.

Несмотря на то, что использование разработанной системы трекинга несколько ухудшает MOTA (таблица), данный показатель находится на приемлемом уровне. Можно видеть, что основная проблема связана с ростом частоты пропусков лиц при увеличении T_{det} . Для того чтобы преодолеть эту проблему, следует использовать надежный детектор: лицо, которое не было найдено детектором, будет пропущено системой трекинга на следующих $T_{det} - 1$ кадрах.

Отметим также, что по показателю MOTP система трекинга превосходит базовую реализацию (в отличие от MOTA, меньшие значения MOTP являются лучшими). Этот парадоксальный результат можно

объяснить следующим образом. Метрика MOTP зависит от множества факторов, таких как качество разметки тестовых видеозаписей и общее количество обнаруженных лиц. Лица, не детектируемые системой трекинга (но обнаруженные базовой реализацией), часто имеют существенное отклонение от фронтального положения, неравномерно освещены или частично скрыты. По этой причине они смещают MOTP базовой реализации в область более высоких значений. Таким образом, можно считать, что все реализации имеют схожие значения показателя MOTP.

Видеозапись	Тестируемая система	<i>miss</i>	<i>fp</i> , %	<i>mme</i> , %	МОТА	MOTP
Видео 1	Базовая реализация	8,40	0,57	0,19	0,91	0,29
	Система трекинга, $T_{det} = 3$	11,00	0,83	0,40	0,88	0,21
	Система трекинга, $T_{det} = 5$	13,00	0,35	0,28	0,86	0,20
Видео 2	Базовая реализация	29,00	1,40	0,46	0,70	0,25
	Система трекинга, $T_{det} = 3$	32,00	2,00	0,64	0,65	0,15
	Система трекинга, $T_{det} = 5$	36,00	1,50	0,40	0,62	0,15

Таблица. Результаты оценки качества трекинга

Заключение

В работе представлена система трекинга для задач видеонаблюдения. Структура системы отличается простотой и гибкостью. Подсистема трекинга основана на эффективных алгоритмах вычисления бинарных дескрипторов и детектирования ключевых точек, что позволяет обрабатывать видеопоток на высоких скоростях.

Тот факт, что система разработана с учетом ряда допущений, не означает, что ее работа будет неминуемо нарушена в случае, если какое-то из этих допущений окажется неверным. Для того чтобы полностью раскрыть потенциал данного подхода, могут быть проведены дополнительные исследования и эксперименты.

Проведенные экспериментальные исследования показали, что предложенный подход позволяет добиться значительного (четырёхкратного) повышения скорости обработки по сравнению с базовой реализацией. При этом качество трекинга осталось на допустимом уровне. Это позволяет интегрировать в систему различные детекторы лиц (в том числе и достаточно медленные) для получения полнофункциональной высокоскоростной системы трекинга множества лиц.

Алгоритм прост в реализации и оптимизации, поэтому он может быть использован не только в полномасштабных системах видеонаблюдения, но и во встроенных решениях, интегрированных непосредственно в камеры видеонаблюдения.

References

- Melnikov A., Akhunzyanov R., Kudashev O., Luckyanets E. Audiovisual liveness detection. *Lecture Notes in Computer Science*, 2015, vol. 9280, pp. 643–652. doi: 10.1007/978-3-319-23234-8_59
- Yilmaz A., Javed O., Shah M. Object tracking: a survey. *ACM Computing Surveys*, 2006, vol. 38, no. 4, pp. 1–45. doi: 10.1145/1177352.1177355
- Yang H., Shao L., Zheng F., Wang L., Song Z. Recent advances and trends in visual tracking: a review. *Neurocomputing*, 2011, vol. 74, no. 18, pp. 3823–3831. doi: 10.1016/j.neucom.2011.07.024
- Smeulders A.W.M., Chu D.M., Cucchiara R., Calderara S., Dehghan A., Shan M. Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, vol. 36, no. 7, pp. 1442–1468. doi: 10.1109/TPAMI.2013.230
- Kristan M., Matas J., et al. The visual object tracking VOT2015 challenge results. *Proc. IEEE Int. Conf. on Computer Vision*. Santiago, Chile, 2015, pp. 564–586.
- Matthews I., Ishikawa T., Baker S. The template update problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, vol. 26, no. 6, pp. 810–815. doi: 10.1109/TPAMI.2004.16
- Minnehan B., Spang H., Savakis A.E. Robust and efficient tracker using dictionary of binary descriptors and locality constraints. *Lecture Notes in Computer Science*, 2014, vol. 8887, pp. 589–598.
- Avidan S. Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no. 2, pp. 261–271. doi: 10.1109/TPAMI.2007.35
- Kalal Z., Mikolajczyk K., Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, vol. 34, no. 7, pp. 1409–1422. doi: 10.1109/TPAMI.2011.239
- Broida T.J., Chellappa R. Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, vol. PAMI-8, no. 1, pp. 90–99.
- Isard M., Blake A. CONDENSATION – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1998, vol. 29, no. 1, pp. 5–28.

12. Maggio E., Piccardo E., regazzoni C., Cavallaro A. Particle PHD filtering for multi-target visual tracking. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, 2007, vol. 1, art. 4217276. doi: 10.1109/ICASSP.2007.366104
13. Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision. *Proc. 7th Int. Joint Conference on Artificial Intelligence, IJCAI*. Vancouver, Canada, 1981, pp. 674–679.
14. Baker S., Matthews I. Lucas-Kanade 20 years on: a unifying framework. *International Journal of Computer Vision*, 2004, vol. 56, no. 3, pp. 221–255. doi: 10.1023/B:VISI.0000011205.11775.fd
15. Ta D.-N., Chen W.-C., Gelfand N., Pilli K. SURFTrac: Efficient tracking and continuous object recognition using local feature descriptors. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. Miami, USA, 2009, pp. 2937–2944. doi: 10.1109/CVPRW.2009.5206831
16. Bilinski P., Bremond F., Kaaniche M.-B. Multiple object tracking with occlusions using HOG descriptors and multi resolution images. *Proc. 3rd Int. Conf. on Crime Detection and Prevention, ICDP*. London, 2009, no. 2. doi: 10.1049/ic.2009.0264
17. Panti B., Monteiro P., Pereira F., Ascenso J. Descriptor-based adaptive tracking-by-detection for visual sensor networks. *IEEE Int. Conf. on Multimedia and Expo Workshops, ICMEW*. Turin, Italy, 2015. doi: 10.1109/ICMEW.2015.7169807
18. Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary robust independent elementary features. *Proc. 11th European Conference on Computer Vision, ECCV*. Heraklion, Crete, Greece, 2010, pp. 778–792. doi: 10.1007/978-3-642-15561-1_56
19. Ishii I., Ichida T., Gu Q., Takaki T. 500-fps face tracking system. *Journal of Real-Time Image Processing*, 2013, vol. 8, no. 4, pp. 379–388. doi: 10.1007/s11554-012-0255-8
20. Kukharev G.A., Kamenskaya E.I., Matveev Y.N., Shchegoleva N.L. *Metody Obrabotki i Raspoznavaniya Izobrazhenii Lits v Zadachakh Biometrii* [Methods for Face Image Processing and Recognition in Biometric Applications] Ed. M.V. Khitrov. St. Petersburg, Politekhnik Publ., 2013, 388 p.
21. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*. Kauai, USA, 2001, vol. 1, pp. 511–518.
22. Ahuja R.K., Magnanti T.L., Orlin J.B. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993, 864 p.
23. Rosten E., Drummond T. Machine learning for high-speed corner detection. *Lecture Notes in Computer Science*, 2006, vol. 3951, pp. 430–443. doi: 10.1007/11744023_34
24. Szeliski R. *Computer Vision: Algorithms and Applications*. London, Springer-Verlag, 2011, 812 p. doi: 10.1007/978-1-84882-935-0
25. OpenCV (open source computer vision). Available at: <http://opencv.org> (accessed 09.06.2015).
26. LEMON Graph Library. Available at: <https://lemon.cs.elte.hu/trac/lemon> (accessed 15.02.2016).
27. Li L., Nawaz T., Ferryman J. PETS 2015: Datasets and challenge. *Proc. 12th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, AVSS*. Karlsruhe, Germany, 2015. doi: 10.1109/AVSS.2015.7301741
28. Collins R., Zhou X., Teh S.K. An open source tracking testbed and evaluation web site. *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, PETS*, 2005, vol. 2, pp. 35.
29. Doermann D., Mihalcik D. Tools and techniques for video performance evaluation. *Proc. 15th Int. Conf. on Pattern Recognition*, 2000, vol. 15, no. 4, pp. 167–170.
30. Bernardin K., Stiefelhagen R. Evaluating multiple object tracking performance: the CLEAR MOT metrics. *EURASIP Journal on Image Video Processing*, 2008, vol. 2008, art. 246309. doi: 10.1155/2008/246309

Олейник Андрей Леонидович

– аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, andrey_oleynik@niuitmo.ru

Andrey L. Oleinik

– postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, andrey_oleynik@niuitmo.ru