

Регистрация проходов людей в видеопотоке

Н. Ю. Багров

ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», г. Москва,
Российская Федерация
119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1
mail@bagnikita.com

Аннотация

Проблема подавления ложных обнаружений в видео актуальна для крупных систем видеонаблюдения, поскольку она приводит к увеличению нагрузки на оборудование и операторов в ситуационных центрах. Также из-за особенности алгоритмов распознавания лиц ложные обнаружения часто приводят к ложным срабатываниям алгоритмов идентификации. Помимо лишней нагрузки, при необходимости подсчета уникальных людей в видео, в большинстве систем производится подсчет уникальных лиц через визуальное сопровождение лиц, поэтому важно не учитывать дважды одного человека при разрыве траектории. В работе предлагается подход, который позволяет регистрировать проходы людей при условии перекрытий в сцене, что важно как для решения статистических задач, так и для систем идентификации лиц, где дубликаты лиц приводят к повышенной нагрузке на операторов системы и оборудование. Предложенная модификация алгоритма визуального сопровождения лиц за счет реидентификации незначительно повышает вычислительную нагрузку, поэтому такой алгоритм может применяться на бортовых устройствах с ограниченными вычислительными ресурсами. В ряде случаев, например при использовании нестабильных каналов связи для передачи данных, модификация алгоритма позволяет снизить нагрузку на эти каналы за счет сокращения числа дубликатов лиц. Кроме снижения нагрузки на каналы связи, уменьшение числа дубликатов также снижает общую стоимость решения за счет уменьшения числа оборудования, которое рассчитывается по максимальному потоку людей в пиковые часы. Также это приводит к снижению требований на объем системы хранения данных в системах распознавания лиц. Аналогичный подход можно применять и в других сценариях визуального сопровождения, например, для выявления нарушений правил дорожного движения на транспорте, объединяя траектории по распознанному номеру транспортного средства.

Ключевые слова: детектирование лиц, сопровождение лиц, распознавание лиц, реидентификация, подсчет людей.

Автор заявляет об отсутствии конфликта интересов.

Для цитирования: Багров, Н. Ю. Регистрация проходов людей в видеопотоке / Н. Ю. Багров. – DOI 10.25559/SITITO.16.202004.908-916 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 4. – С. 908-916.

© Багров Н. Ю., 2020



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Person Registration in CCTV

N. Yu. Bagrov

Lomonosov Moscow State University, Moscow, Russian Federation
1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation
mail@bagnikita.com

Abstract

The problem of suppressing false detections in video is relevant for large video surveillance systems, since it leads to an increase in the load on equipment and operators in situational centers. Also in face recognition system false low quality detections often lead to false positives of identification algorithms. In case of people counting applications false detections also lead to phantom detections and counting people twice. This paper proposes an approach to register the passage of people with overlapping in the scene using reidentification algorithm which is embedded in visual tracking. This modification insignificantly increases the computational load, so it can be used on on-board devices with limited computing resources. Another positive side of the algorithm is the reduction of network traffic between face detection appliance and centralized face recognition system. This is important for cellular mobile connections which are often used in transport applications. This approach can also be extended to other tasks, such as identifying traffic violations, by improving optical tracker quality by merging results using license plate recognition. The algorithm was benchmarked in crowded scenes and showed a significant improvement in optical tracking results without using graphics processing units.

Keywords: face detection, face tracking, face recognition, reidentification, people counting.

The author declares no conflicts of interest.

For citation: Bagrov N.Yu. Person Registration in CCTV. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(4):908-916. DOI: <https://doi.org/10.25559/SITITO.16.202004.908-916>



Введение

Фиксацию фактов проходов людей в видеопотоке обычно реализуют с использованием алгоритмов обнаружения лиц и визуального сопровождения. Недостатки данного подхода проявляются в ошибках визуального сопровождения как из-за ошибок самого алгоритма, так и из-за перекрытий лиц людьми другими людьми, дверями на входах, рамками металлодетекторов, колоннами зданий и другими элементами. Алгоритм заключается в применении метода обнаружения лиц на некотором наборе кадров видеопотока через разные интервалы времени, на всех остальных кадрах применяется визуальное сопровождение найденных объектов для ускорения работы. На каждом кадре, где выполнялось обнаружение лиц, решается задача о назначениях, чтобы сопоставить текущие траектории, продолженные методом визуального сопровождения. Результаты работы можно применять для решения различных задач, в работе рассматривается подсчет уникальных лиц и применение в составе системы идентификации лиц. Для последних актуально сокращение числа дубликатов лиц, поскольку они создают повышенную нагрузку на каналы связи, оборудование и операторов в ситуационных центрах. Загрузка каналов связи важна для бортовых и мобильных решений, поскольку в них канал связи обычно реализуют через операторов сотовой связи с ограниченной пропускной способностью и низкой надежностью.

В ряде случаев, например при обнаружении лиц в толпе, алгоритмы основанные только на визуальном сопровождении будут генерировать большое число дубликатов лиц из-за постоянных перекрытий лиц другими людьми. Это искажает статистические оценки (если необходим подсчет) и снижает эффективность работы всей системы, поскольку могут быть обнаружены лица с низким визуальным качеством, которые снижают точность работы алгоритмов распознавания лиц.

Подсчет уникальных лиц

Задача подсчета уникальных лиц актуальна для сбора статистики и посещаемости какого-либо объекта. Часто подобные системы устанавливают на входе в торговые центры и магазины, дополняя результаты работы модулями для классификации пола/возраста и построения отчетов о потоке покупателей. Для подсчета покупателей существуют готовые программно-аппаратные решения, состоящие из блока с камерой и специализированного вычислителя, однако их область применимости ограничена из-за требований на установку камеры вертикального ракурса съемки, также они не позволяют работать с широкими входными группами, что часто бывает в торговых центрах [1,2,3]. Использование обычных камер видеонаблюдения для подсчета уникальных людей позволяет как использовать уже существующие камеры, так и обрабатывать большие открытые пространства.

В случае необходимости подсчета числа людей на улице готовые решения обычно неприменимы из-за невозможности установки камеры вертикального ракурса, либо по причине засветки солнечным светом инфракрасных датчиков в аппаратных счетчиках.

Рассмотрим базой алгоритм подсчета уникальных людей по

камерам видеонаблюдения. Он состоит из детектора лиц, алгоритма визуального сопровождения и алгоритма сопоставления обнаружений. Кадры из видеопотока разделяются на 2 категории кадров:

- кадры, на которых осуществляется детектирование лиц **Dframe** и визуальное сопровождение лиц. Эти кадры распределены через равные промежутки времени;
- кадры, на которых осуществляется только визуальное сопровождение лиц **Tframe**.

Определение: Размеченным лицом будем называть изображение лица человека с координатами ограничивающего прямоугольника лица на одном кадре видеопотока.

Определение: Множество размеченных лиц, относящихся к одному появлению человека в видеопотоке, будем называть треком. Появление считается одним, если повторно этот человек появляется в кадре через длительное время (существенно большее чем длина трека). Допускается пропуск некоторого числа кадров в середине трека из-за перекрытий лиц людей.

Каждый трек характеризуется набором состояний:

- Состояние **ALIVE** назначается тем трекам которые либо были созданы в результате работы детектора лиц, либо для которых алгоритм визуального сопровождения смог продлить траекторию;
- Состояние **LOST** назначается тем трекам, которые либо были не сопоставлены между **Tframe** и **Dframe**, либо прерванные алгоритмом визуального сопровождения. Каждый трек может находиться в состоянии **LOST** не более l_{max} секунд;
- Состояние **DEAD** назначается тем трекам, которые находились в состоянии **LOST** более секунд.

Подсчет количества проходов людей в сцене производится суммированием количества активных и завершенных треков в заданный временной интервал.

В качестве детектора лиц используется Single Shot Multibox Detector, для визуального сопровождения метод, основанный на сумме квадратов попиксельных разностей [4,5]. Существуют более вычислительно сложные и точные алгоритмы визуального сопровождения (Robust online visual tracking with a single convolutional neural network, A prior-less method for multi-face tracking in unconstrained videos), однако скорость их работы низкая и их сложно применять на устройствах с ограниченными ресурсами [6,7]. Сопоставление лиц решается Венгерским алгоритмом с функцией стоимости по формуле 1 [8]. Пусть задано множество треклетов $T=\{t_i\}$ и результаты детектора лиц $D=\{d_j\}$ на k и $k+1$ кадре видеопотока соответственно. $IoU(x,y)$ коэффициент Жаккара. $|\dots|$ — индикаторная функция.

$$|IoU(t_i, d_j) > \phi| * (1 - IoU(t_i, d_j)) + |IoU(t_i, d_j) < \phi| * 2 \quad (1)$$

Алгоритмы визуального сопровождения лиц могут генерировать различные типы ошибок, но для систем идентификации и подсчета людей можно выделить следующие:

- потеря лиц. Поскольку алгоритм сопровождения ищет визуально похожие объекты в некоторой окрестности, то возможна ситуация когда он не сопоставит лицо на новом кадре. В этом случае происходит потеря лица и прекращение трека, что приводит к некорректным результатам подсчета лиц и повышению нагрузки на оборудование. Поскольку при сопоставлении лиц с **Dframe** и **Tframe** учитывается только их пересечение по коэффициенту Жаккара, то алгоритм будет



учитывать проходы людей по несколько раз в случае перекрытия их лиц, например, колоннами или другими людьми в толпе;

- смена идентификатора. В процессе работы алгоритм сопровождения может ошибочно продолжить треки на лица других людей, что приведет как к потери части обнаружений, так и усложнит выбор лучшего кадра для трека.

Для решения этих проблем предлагается модификация функции стоимости по формуле 2 в Венгерском алгоритме за счет использования алгоритма верификации лиц $FR(t_i, d_j) \in [0.0, 1.0]$. В качестве алгоритма для верификации лиц используется специально обученная нейронная сеть на основе измененной архитектуры mobilenet с использованием программной библиотеки Apache MXNet [9,10]. Поскольку данный алгоритм верификации лиц используется для реидентификации на коротком промежутке времени, то в нем используется большая область вокруг лица чем в обычных алгоритмах распознавания лиц чтобы учитывать дополнительные признаки, такие как головной убор и фрагменты верхней одежды человека. Обычные агоритмы распознавания лиц обучаются так, чтобы быть инвариантными к изменению одежды и головного убора человека. Расширение области позволяет неявным образом использовать эти признаки, которые полезны для реидентификации, поскольку за время появления в сцене человек обычно не меняет головной убор и одежду.

Нейронная сеть обучалась на выборке из 40 млн. изображений лиц 400 тыс. различных людей [11]. Данная выборка больше большинства других существующий общедоступных коллекций изображений, таких как Microsoft Celeb 1M и CASIA-WebFace, и была собрана для сценария работы с данными различного визуального качества [12,13]. Размер батча при обучении составил 2500 изображений, что важно для стабильности процесса обучения из-за наличия ошибок разметки в обучающей выборке. Обучение производилось на сервере с 8x Tesla V100 32GB. Для верификации лиц выполняется построение нейросетевого дескриптора $E(x)$ для каждого из двух лиц, где в качестве дескриптора лица используются выходы с одного из слоев нейронной сети (128 вещественных чисел).

Значение $FR(a, b) = M\left(\frac{E(a) \cdot E(b)}{|E(a)| \cdot |E(b)|}\right)$ является результатом сравнения двух изображений лиц по косинусной метрике с применением функции отображения значений M . Функция отображения M является кусочно-линейным преобразованием, которое отображает значения из косинусной метрики по следующим правилам:

- минимальное значение косинусной метрики отображается в 0;
- максимальное значение косинусной метрики отображается в 1;
- производится оценка числа ошибок второго рода алгоритма верификации лиц на размеченной выборке и выполняется построение отображения $FAR \rightarrow Cosine$ на некоторое множество значений FAR (доля ошибок второго рода алгоритма распознавания, т.е. когда разные люди определяются алгоритмом как одинаковые). В работе такое отображение построено для FAR 0.5, 0.1, 0.01, 0.001, 0.0001, 0.00001. Далее значения FAR отображаются на отрезок $[0.0, 1.0]$ по правилу из таблицы

1. Далее применяется обратное преобразование $Cosine \rightarrow FAR$ и происходит интерполяция значения на соответствующем отрезке из таблицы 1. Значение FAR 0.01 означает что на каждые 100 верификаций лиц статистически будет 1 ошибка второго рода.

Т а б л и ц а 1. О т о б р а ж е н и е з н а ч е н и й а л г о р и т м а в е р и ф и к а ц и и л и ц
T a b l e 1. D i s p l a y i n g t h e v a l u e s o f t h e f a c e v e r i f i c a t i o n a l g o r i t h m

FAR	Значение функции верификации лиц
0.5	0.5
0.01	0.6
0.001	0.7
0.0001	0.8
0.00001	0.9

Чем выше значение функции верификации лиц, тем выше сходство лиц. Выборка собрана из различных источников данных (включая источники с низким визуальным качеством и высокой вариативностью выражений и поворотов лиц), поэтому обученная на ней нейронная сеть должна быть устойчива к таким особенностям.

Использование большей области вокруг лица, как показано на рисунке 1, позволяет повысить устойчивость алгоритма к сложным ракурсам, таким как повороты головы и частичные перекрытия, но также приводит к повышению числа ошибок второго рода (когда разных людей алгоритм признает за одинаковых). Поскольку используется небольшое временное окно для реидентификации, то такие ошибки незначительны из-за небольшого объема лиц, используемых для сравнения. Например, в зимнее время большинство людей ходит в шапках, что скрывает часть области лиц, но позволяет алгоритму проводить реидентификацию по особенностям головного убора. Из-за эпидемиологических ограничений лицо может быть частично скрыто медицинской маской, потому увеличение области лица для работы алгоритма также повышает точность работы алгоритма реидентификации.



Р и с. 1. П р и м е р р а с ш и р е н н о г о и б а з о в о г о л и ц а д л я н е й р о с е т и
F i g. 1. A n e x a m p l e o f a n e x t e n d e d a n d b a s i c f a c e f o r a n e u r a l n e t w o r k



Для каждого трека производится выбор лучшего кадра лица с использованием алгоритма оценки визуального качества лиц. Большинство современных алгоритмов оценки качества изображений основаны на нейронных сетях, помимо точности работы в основном они отличаются скоростью работы а набором поддерживаемых характеристик [14,15,16,17,18]. В работе был выбран алгоритм из работы (Lisin N. et al. Improving the Neural Network Algorithm for Assessing the Quality of Facial Images), поскольку он поддерживает выбор лучшего лица в треке по критериям размера, углам поворота, размытия и оценкам перекрытия лица, а также работает достаточно быстро без использования графических ускорителей [14]. Алгоритмы достаточно быстры для их применения в процессе обработки видеопотока на множестве кадров и позволяют исключить сильно перекрытые и размытые лица из реидентификации. Коэффициенты α и β в формуле 2 задают порог минимального

$$\min (|IoU(t_i, d_j)| > \phi * |FR(b_i, d_j)| > \alpha * (1 - IoU(t_i, d_j)), |FR(b_i, d_j)| > \beta * (1 - FR(b_i, d_j))) \quad (2)$$

Данный подход позволяет в том числе объединять траектории людей, которые проходят за колоннами или другими непрозрачными объектами в сцене.

При значениях $\alpha=0$ и $\beta=1$ модифицированная функция стоимости не отличается от базовой. Увеличение параметра α уменьшает вероятность попадание лиц различных людей в один трек, но также увеличивает вероятность разрыва треков. Уменьшение параметра β позволяет объединять треки в случае их разрывов (как в следствие ошибок алгоритма сопровождения, так и из-за перекрытий). Параметры α и β можно подбирать либо аналитически исходя из количества ошибок алгоритма верификации лиц, так и экспериментально используя размеченные видеопоследовательности. В данном эксперименте параметры $\alpha=0,1$ и $\beta=0,7$ подобраны экспериментально.

сходства лиц в одном треке и порог на объединение различных треков в один. Модифицированная функция применяется только для треклетов с состоянием **LOST**, для остальных применяется базовая функция. Существуют аналогичные подходы на реидентификации (*People tracking and re-identification by face recognition for rgb-d camera networks, A real-time and unsupervised face re-identification system for human-robot interaction*), однако они применяются для узкоспециализированного сценария роботов и достаточно ресурсоемкие [19,20]. Заметим, что применения технологии распознавания лиц вносит дополнительные требования на уровень освещения (не менее 100 люкс) и минимальный размер лиц в кадре (не менее 40 пикселей по ширине). На практике это не является существенным ограничением, поскольку уровень освещенности на КПП и объектах транспортной инфраструктуры достаточно высокий и в целевой области лица крупнее^{1,2} [21,22].

Экспериментальная оценка

Экспериментальная оценка базового и модифицированного алгоритма была проведена на видеопоследовательностях, полученных с камер видеонаблюдения. Ракурсы соответствуют типовым ракурсам, применяемым в системах видеонаблюдения на транспорте. Всего было размечено 20 видеопоследовательностей суммарной продолжительностью более 60 минут, на каждом видео было зафиксировано не менее 100 проходов различных людей. Экспертная разметка осуществлялась вручную с использованием программных библиотек Tevian FaceSDK, OpenCV и CVAT^{3,4} [23,24,25].



Р и с. 2. Пример сцены Турникеты
F i g. 2. An example of a scene Turnstiles

¹ Дорман И. Я. Свод правил 120.13330.2012 «Метрополитены» (Актуализированная редакция СНиП 32-02-2003) // Метро и тоннели. 2013. № 1. С. 20-22. URL: <https://elibrary.ru/item.asp?id=24107516&> (дата обращения: 11.08.2020).

² ГОСТ Р ИСО/МЭК 19794-5-2013. Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 5. Данные изображения лица = Information technologies. Biometrics. Biometric data interchange formats. Part 5. Face image data: национальный стандарт Российской Федерации: введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 6 сентября 2013 г. № 987-ст.

³ Tevian FaceSDK [Электронный ресурс] // Единый реестр российских программ для ЭВМ и БД. 2020. URL: https://reestr.digital.gov.ru/reestr/308277/?sphrase_id=364098 (дата обращения: 11.08.2020).

⁴ OpenCV. Computer vision annotation tool (CVAT) [Электронный ресурс]. URL: <https://github.com/openvinotoolkit/cvat> (дата обращения: 11.08.2020).



Характеристики видеопоследовательностей представлены в таблице 2. Сценарий Двери содержит видеопоследовательности, где люди проходят через распашные двери. Этот сценарий содержит большое число перекрытий и разрывов траекторий из-за резких движений людей в кадре (например, когда люди придерживают двери и отворачиваются от камеры). Сценарий Турникеты содержит видеопоследовательности, на которых камера снимает группу турникетов на КПП под углом 45 градусов (пример аналогичной сцены изображен на рисунке 2). В данном сценарии много перекрытий людей. Сценарий Турникет содержит видеопоследовательности где люди проходят через один турникет, камера расположена на уровне лица человека. Сценарий Холл содержит видеопоследовательности где люди проходят через помещение с колоннами. Разрешение всех видеопоследовательностей 1920x1080, частота кадров не менее 25к/с. Уровень освещенности не менее 100 люкс.

Таблица 2. Характеристики видеопоследовательностей
Table 2. Video sequences characteristics

Название	Количество различных людей	Продолжительность (мин)
Двери-1	142	3
Двери 2	139	3
Двери 3	146	3
Двери 4	138	3
Турникеты-1	114	3
Турникеты-2	116	3
Турникеты-3	172	3
Турникеты-4	104	3
Турникеты-5	104	3
Турникеты-6	131	3
Турникеты-7	120	3
Турникет-1	102	5

Название	Количество различных людей	Продолжительность (мин)
Турникет-2	105	5
Турникет-3	113	5
Турникет-4	119	5
Турникет-5	103	5
Холл-1	115	3
Холл-2	120	3
Холл-3	121	3
Холл-4	118	3

Подбор параметров α и β проводился перебором по сетке с оценкой точности работы на 5 независимых видеопоследовательностях суммарной длиной 10 минут. Характеристики видеопоследовательностей аналогичны тем, на которых проводилась экспериментальная оценка.

В таблице 3 представлены результаты работы предложенного алгоритма в сравнении с базовым. Полнотой называется доля правильно построенных треков в видеопоследовательности к общему числу треков в видеопоследовательности. Точностью называется доля правильно сопоставленных треков к общему числу построенных алгоритмом треков. **F-мера** вычисляется как среднее гармоническое между точностью и полнотой. Для сопоставления построенных алгоритмом треков t_i и экспертной разметкой g_i решается задача о назначениях, где в качестве функции стоимости используется покадровая сумма (n - число кадров в видеопоследовательности) коэффициентов Жаккара координат лиц по формуле 3. Если на кадре отсутствует t_i или g_i , то соответствующее значение $IoU(t_i, g_i)$ считается равным нулю.

$$\sum_{i=1}^n IoU(t_i, g_i) \quad (3)$$

Таблица 3. Экспериментальная оценка
Table 3. Experimental evaluation

Название	Полнота	Точность	F-мера	Сравнение алгоритмов		
				Базовый алгоритм	Предложенный алгоритм	F-мера
Двери-1	100%	63.8%	0.778	100%	87.5%	0.933
Двери-2	97.4%	79.6%	0.851	100%	86.4%	0.927
Двери-3	95.6%	76.8%	0.851	95.6%	89.6%	0.925
Двери-4	96.2%	84.6%	0.9	99.2%	96.9%	0.981
Турникеты-1	91.7%	89.8%	0.907	98%	100%	0.99
Турникеты-2	91.8%	86.8%	0.892	98%	95.7%	0.968
Турникеты-3	96.3%	75.7%	0.847	100%	91.2%	0.954
Турникеты-4	92.3%	83.7%	0.878	94.6%	93.3%	0.939
Турникеты-5	97.8%	71.9%	0.828	98.9%	89.8%	0.941
Турникеты-6	100%	84.8%	0.918	100%	87.1%	0.931



Название	Полнота	Точность	F-мера	Полнота	Точность	F-мера
	Базовый алгоритм			Предложенный алгоритм		
Турникеты-7	87.5%	76.3%	0.815	93.8%	88.3%	0.91
Турникет-1	100%	62%	0.765	100%	97.6%	0.988
Турникет-2	100%	63.7%	0.778	100%	96.1%	0.98
Турникет-3	92.3%	67.8%	0.782	92.3%	99.2%	0.956
Турникет-4	100%	70.3%	0.825	100%	97.2%	0.986
Турникет-5	100%	61.9%	0.765	100%	97.7%	0.988
Холл-1	92.9%	92.9%	0.929	92.9%	100%	0.963
Холл-2	90.9%	76.9%	0.833	90%	90%	0.9
Холл-3	100%	100%	1	100%	100%	1
Холл-4	100%	85.7%	0.923	100%	94.7%	0.973

По результатам экспериментальной оценки видно, что предложенный алгоритм повышает полноту на сценариях где часто происходят перекрытия людей (Двери, Турникеты). Обычный алгоритм сопровождения в таких случаях может продолжить сопровождение объекта и допустить переключение на лицо другого человека. Для сценария Турникет базовый алгоритм работает достаточно хорошо из-за специфики установки камеры - в ракурс попадает крупное лицо проходящего человека и смены идентификаторов треков нет. Тестовая видеопоследовательность Холл-3 является простой для работы алгоритмов сопровождения из-за отсутствия перекрытий в области видимости камеры. По результатам сравнения работы алгоритмов видно, что предложенный алгоритм не хуже базового и в большинстве сценариев лучше его сразу по двум критериям. За счет уменьшения числа дубликатов лиц повышается эффективность использования каналов связи, которые могут быть нестибальными и достаточно дорогими в использовании. Специально разработанная нейронная сеть для задач реидентификации незначительно повышает вычислительную нагрузку и в результате обработка одного канала видео 1920x1080 15к/с возможна на одном процессорном ядре с поддержкой набора инструкций AVX2 и частотой не ниже 2.2Ghz.

Заключение

Предложенная модификация алгоритма подсчета людей в видеопоследовательности повышает точность его работы и позволяет экономить сетевой трафик за счет уменьшения числа разрывов треков и отправки дублирующей информации. Из недостатков алгоритма можно отметить более высокие требования к качеству изображений лиц из-за необходимости применять алгоритм верификации лиц, однако на практике при использовании современного оборудования и достаточного уровня освещения на объекте это не является проблемой. В качестве дальнейшего развития алгоритма можно рассмотреть улучшение работы реидентификации лиц в более сложных условиях освещения, исследования по уменьшению кадровой частоты видеопотока при сохранении точности работы.

References

- [1] Del Pizzo L. et al. Counting people by RGB or depth overhead cameras. *Pattern Recognition Letters*. 2016; 81:41-50. (In Eng.) DOI: <https://doi.org/10.1016/j.patrec.2016.05.033>
- [2] Zhou D., He Q. Cascaded Multi-Task Learning of Head Segmentation and Density Regression for RGBD Crowd Counting. *IEEE Access*. 2020; 8:101616-101627. (In Eng.) DOI: <https://doi.org/10.1109/ACCESS.2020.2998678>
- [3] Liciotti D., Cenci A., Frontoni E., Mancini A., Zingaretti P. An Intelligent RGB-D Video System for Bus Passenger Counting. In: Chen W., Hosoda K., Menegatti E., Shimizu M., Wang H. (ed.) *Intelligent Autonomous Systems 14. IAS 2016. Advances in Intelligent Systems and Computing*. 2017; 531:473-484. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-48036-7_34
- [4] Liu W. et al. SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (ed.) *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*. 2016; 9905:21-37. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-46448-0_2
- [5] Shi J., Tomasi Good features to track. In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA; 1994. p. 593-600. (In Eng.) DOI: <https://doi.org/10.1109/CVPR.1994.323794>
- [6] Li H., Li Y., Porikli F. Robust Online Visual Tracking with a Single Convolutional Neural Network. In: Cremers D., Reid I., Saito H., Yang MH. (ed.) *Computer Vision – ACCV 2014. ACCV 2014. Lecture Notes in Computer Science*. 2015; 9007:194-209. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-16814-2_13
- [7] Lin C., Hung Y. A Prior-Less Method for Multi-face Tracking in Unconstrained Videos. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA; 2018. p. 538-547. (In Eng.) DOI: <https://doi.org/10.1109/CVPR.2018.00063>
- [8] Munkres J. Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*. 1957; 5(1):32-38. (In Eng.) DOI: <https://doi.org/10.1137/0105003>



- [9] Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA; 2019. p. 4685-4694. (In Eng.) DOI: <https://doi.org/10.1109/CVPR.2019.00482>
- [10] Chen T. et al. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. *arXiv:1512.01274*, 2015. Available at: <https://arxiv.org/abs/1512.01274> (accessed 11.08.2020). (In Eng.)
- [11] Bagrov N.Y., Konushin A.S., Konushin V.S. A Semi-Automatic Method of Collecting Samples for Learning a Face Identification Algorithm. *Programming and Computer Software*. 2019; 45(3):133-139. (In Eng.) DOI: <https://doi.org/10.1134/S0361768819030022>
- [12] Guo Y., Zhang L., Hu Y., He X., Gao J. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In: Leibe B., Matas J., Sebe N., Welling M. (ed.) *Computer Vision – ECCV 2016*. ECCV 2016. *Lecture Notes in Computer Science*. 2016; 9907:87-102. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-46487-9_6
- [13] Yi D., Lei Z., Liao S., Li S.Z. Learning Face Representation from Scratch. *arXiv:1411.7923*, 2014. Available at: <https://arxiv.org/abs/1411.7923> (accessed 11.08.2020). (In Eng.)
- [14] Lisin N., Gromov A., Konushin V., Konushin A. Improving the Neural Network Algorithm for Assessing the Quality of Facial Images. *CEUR Workshop Proceedings*. 2020; 2744. Article 28. Available at: <http://ceur-ws.org/Vol-2744/paper28.pdf> (accessed 11.08.2020). (In Eng.)
- [15] Nasrollahi K., Moeslund T.B. Face Quality Assessment System in Video Sequences. In: Schouten B., Juul N.C., Drygajlo A., Tistarelli M. (ed.) *Biometrics and Identity Management*. BioID 2008. *Lecture Notes in Computer Science*. 2008; 5372:10-18. Springer, Berlin, Heidelberg. (In Eng.) DOI: https://doi.org/10.1007/978-3-540-89991-4_2
- [16] Wong Y., Chen S., Mau S., Sanderson C., Lovell B.C. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In: *CVPR 2011 WORKSHOPS*. Colorado Springs, CO, USA; 2011. p. 74-81. (In Eng.) DOI: <https://doi.org/10.1109/CVPRW.2011.5981881>
- [17] Nikitin M.Yu., Konushin V.S., Konushin A.S. Neural network model for video-based face recognition with frames quality assessment. *Computer Optics*. 2017; 41(5):732-742. (In Russ., abstract in Eng.) DOI: <https://doi.org/10.18287/2412-6179-2017-41-5-732-742>
- [18] Vignesh S., Priya K. V. S. N. L. M., Channappayya S.S. Face image quality assessment for face selection in surveillance video using convolutional neural networks. In: *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Orlando, FL, USA; 2015. p. 577-581. (In Eng.) DOI: <https://doi.org/10.1109/GlobalSIP.2015.7418261>
- [19] Koide K., Menegatti E., Carraro M., Munaro M., Miura J. People tracking and re-identification by face recognition for RGB-D camera networks. In: *2017 European Conference on Mobile Robots (ECMR)*. Paris, France; 2017. p. 1-7. (In Eng.) DOI: <https://doi.org/10.1109/ECMR.2017.8098689>
- [20] Wang Y., Shen J., Petridis S., Pantic M. A real-time and unsupervised face re-identification system for human-robot interaction. *Pattern Recognition Letters*. 2019; 128:559-568. (In Eng.) DOI: <https://doi.org/10.1016/j.patrec.2018.04.009>
- [21] Kalaiselvi P., Nithya S. Face Recognition System under Varying Lighting Conditions. *IOSR Journal of Computer Engineering*. 2013; 14(3):79-88. (In Eng.) DOI: <https://doi.org/10.9790/0661-1437988>
- [22] Adini Y., Moses Y., Ullman S. Face recognition: the problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997; 19(7):721-732. (In Eng.) DOI: <https://doi.org/10.1109/34.598229>
- [23] Nikitin M.Yu., Konushin V.S., Konushin A.S. Face anti-spoofing with joint spoofing medium detection and eye blinking analysis. *Computer Optics*. 2019; 43(4):618-626. (In Eng.) DOI: <https://doi.org/10.18287/2412-6179-2019-43-4-618-626>
- [24] Bradski G., Kaehler A. *Learning OpenCV*. O'Reilly Media, Inc.; 2008. (In Eng.)
- [25] Yu Q., Cheng H.H., Cheng W.W., Zhou X. Ch OpenCV for interactive open architecture computer vision. *Advances in Engineering Software*. 2004; 35(8-9):527-536. (In Eng.) DOI: <https://doi.org/10.1016/j.advengsoft.2004.05.003>

Поступила 11.08.2020; одобрена после рецензирования 19.11.2020; принята к публикации 05.12.2020.
Submitted 11.08.2020; approved after reviewing 19.11.2020; accepted for publication 05.12.2020.

Об авторе:

Багров Никита Юрьевич, аспирант кафедры интеллектуальных информационных технологий, лаборатория компьютерной графики и мультимедиа, факультет вычислительной математики и кибернетики, ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова» (119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0003-0709-2211>, mail@bagnikita.com

Благодарности: автор выражает особую благодарность кандидату физико-математических наук, доценту, заведующему лабораторией компьютерной графики и мультимедиа факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова Антону Сергеевичу Конушину за полезные рекомендации и ценные советы в подготовке материалов статьи, а также ООО «Технологии Видеоанализа» за оказанную помощь и поддержку при проведении данного исследования.

Автор прочитал и одобрил окончательный вариант рукописи.

About the author:

Nikita Yu. Bagrov, Postgraduate Student of the Department of Intelligent Information Technologies, Computer Graphics and Multimedia Lab, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University (1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation), ORCID: <http://orcid.org/0000-0003-0709-2211>, mail@bagnikita.com



Acknowledgments: The author expresses special gratitude to Anton Sergeevich Konushin, Ph.D. (Phys.-Math.), Associate Professor, Head of the Graphics & Media Laboratory of the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, for useful recommendations and valuable advice in preparing the article materials, as well as LLC "Technologies of Video Analysis" for their help and support in carrying out this study.

The author has read and approved the final manuscript.

