УДК 004.93
DOI: 10.14489/vkit.2020.07.pp.003-014

Р. П. Богуш, канд. техн. наук, **И. Ю. Захарова**
(Полоцкий государственный университет, Полоцк, Республика Беларусь),
С. В. Абламейко, д-р техн. наук (Белорусский государственный университет,
Минск, Республика Беларусь);
e-mail: ablameyko@bsu.by

АЛГОРИТМ СОПРОВОЖДЕНИЯ ЛЮДЕЙ НА ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ ИДЕНТИФИКАЦИИ ПО ЛИЦАМ ДЛЯ НАБЛЮДЕНИЯ ВНУТРИ ПОМЕЩЕНИЙ

Представлен алгоритм сопровождения людей по видеопоследовательностям с использованием результатов идентификации по лицам при сложной траектории их движения внутри помещения. На первом шаге выполняется обнаружение людей с применением сверточной нейронной сети с архитектурой YOLOv3 и их описанием прямоугольной областью. Эксперименты проведены на пяти тестовых видеопоследовательностях с различным количеством людей, снятых в помещениях неподвижной видеокамерой. Получены основные характеристики разработанного алгоритма, которые подтвердили его эффективность.

Ключевые слова: сопровождение людей; распознавание лиц; внутреннее видеонаблюдение; сверточные нейронные сети.

R. P. Bogush, I. Yu. Zakharova (Polotsk State University, Polotsk, Belarus),
S. V. Ablameyko (Belarusian State University, Minsk, Belarus)

ALGORITHM FOR PERSON TRACKING ON VIDEO SEQUENCES USING FACE IDENTIFICATION FOR INDOOR SURVEILLANCE

This paper discusses the algorithmic framework for tracking people on indoor video. To improve tracking accuracy was used face identification algorithm to reduce errorr rate during complicated trajectory of persons in indoor environment. Object detection was performed with CNN Yolov3 that extract rectangular area as a result. Face detection task was resolved eith Cascade CNN MTCNN with following recognition using CNN MobileFaceNetwork. To form person features we used historgrams in HSV colorspave and CNN that includes 29 convolution layers followed by fully connected layer. The Hungarian algorithm was used as decision maker for allignment problem. Experiments were conducted on five videosequences with the variable number of people in it. The main characteristics of the developed algorithm are obtained which confirmed its effectiveness and the possibility of use for indoor video surveillance.

Keywords: Person tracking; Face recognition; Indoor video surveillance; Convolutional neural networks.

Статья поступила в редакцию 08.04.2020 г.

Введение

Задача сопровождения множества объектов на видеопоследовательностях – одна из основных в компьютерном зрении. В настоящее время она имеет различное количество технических применений и в дальнейшем может использоваться для решения следующих практических задач [1, 2]:

– анализ окружающей обстановки в автоматизированных системах вождения транспортными средствами;

– оценка правильности движения в медицине и спорте;

– сопровождение объектов в системах технического зрения на производстве;

– распознавание типа активности человека в системах мониторинга и охраны и т.д.

Рассматриваемая задача характеризуется высокой сложностью реализации и требует точной локализации объектов в кадрах и правильной идентификации на текущем кадре относительно предыдущих. При сопровождении множества людей в помещении мешающими факторами являются:

– неоднородный задний фон, фрагменты которого могут быть схожи по форме, текстуре и цвету с изображениями людей;

– низкий уровень освещенности;

– наличие теней и, соответственно, изменяющийся задний фон;

– множественные перекрытия людей между собой и с другими объектами в помещении;

– высокая схожесть признаков разных сопровождаемых людей;

– достаточно быстрое их движение;

– в ряде случаев изменяющееся ускорение и нелинейная трансформация траектории движения.

Для сопровождения людей применяются методы на основе сравнения признаков изображений людей и лиц. В группе первых методов в настоящее время наиболее результативным является сопровождение через обнаружение [3]. Этот подход использует ансамбль из детектора объектов и алгоритма для объединения результатов обнаружений на двух кадрах. Эффективное решение проблемы объединения позволит корректно соотносить результаты обнаружения различных объектов и формировать устойчивые траектории движения для каждого из них.

В настоящее время широкое развитие и применение для обнаружения объектов получили алгоритмы классификации с применением *сверточных нейронных сетей* (*СНС*) [4 – 6], которые устойчивы к изменениям освещенности, динамическому заднему фону и позволяют осуществлять детектирование даже в случае частичных перекрытий, что повышает качество сопровождения. Однако при схожести признаков людей, сложных траекториях движения эффективность их работы значительно уменьшается.

Сопровождение также осуществляется с применением класса алгоритмов, которые используют выделение ключевых точек человека, и путем формирования на основе этого наборов признаков с учетом расстояний между ними, отношений между расстояниями и др.

В [7] предложена СНС PoseNet для выделения таких ключевых точек. Алгоритм основан на анализе смещения всех точек в пространстве и изменения соотношений расстояний между ними. Подобные алгоритмы характеризуются большей стабильностью при частичных перекрытиях объектов, однако требуют значительных вычислительных затрат.

Для повышения эффективности сопровождения при пересечении траекторий движения людей, долговременном скрытии их за объектами фона, высокой схожести внешних признаков людей перспективным является вычисление признаков лиц и их использование при установлении соответствия людей на кадрах. В статье [8] предложен алгоритм, который использует сопровождение по обнаружению с использованием признаков лица.

На первом этапе используется алгоритм выделения заднего фона, предназначенный для видеокамер с функцией измерения глубины изображения, который позволяет выделять движущиеся области-кандидаты. Далее для них формируются HOG-признаки (Histogram of Oriented Gradients, HOG), которые классифицируются с использованием метода опорных векторов *SVM* (Support Vector Machines) для принятия решения о присутствии человека в данной области. Затем на основе HOG и SVM выполняется обнаружение лица в данной области.

Для области лица рассчитывается вектор признаков, содержащий 128 значений с использованием СНС FaceNet [9]. После нормализации полученного дескриптора выполняется поиск соответствия между лицами в базе данных и

присутствующими на кадре с назначением имен последним.

Для оценки качества алгоритма использовались две видеопоследовательности длиной 162 и 218 кадров соответственно. На каждой из них присутствовали шесть человек. Заметим, что большое количество ошибок возникает в том случае, когда лицо человека скрыто, например, бородой, или долгое время его признаки не могут быть выделены. Кроме того, предложенный алгоритм может быть использован только в системах, оборудованных датчиком глубины.

В работе [10] рассматривается реидентификация (reidentification, Re-ID) людей с применением анализа признаков человека и его лица. Эта задача схожа с задачей сопровождения и направлена на сопоставление изображений людей на разных сценах, полученных с пространственно разнесенных видеокамер, для установления соответствий между ними.

Для выделения признаков человека используется модифицированная СНС AlexNet. Для выделения признаков лица человека используется модифицированная СНС VGG-16. Обе модели продуцируют вектор из 4096 признаков в качестве дескриптора человека и лица соответственно.

В процессе обучения была использована база данных ImageNet, которая не адаптирована для задачи ре-идентификации. Далее положения лица и человека на кадре совместно с признаками, выделенными из детектированных областей, поступают на алгоритм объединения с использованием цепей Маркова. Точность данного алгоритма на базе данных DukeMTMC (Multi-Target, Multi-Camera) оценивается метрикой AUC (Area Under Curve): 92,07 % при использовании признаков лица; 99,99 % при применении признаков человека.

Комбинация признаков в данной работе не оценивалась. Однако в используемой базе данных мало пересечений траекторий движения людей и практически отсутствуют длительные их перекрытия, т.е. наиболее сложных ситуаций мало.

Таким образом, несмотря на наличие большого количества существующих алгоритмических решений, которые могут быть использованы для сопровождения людей, ввиду высокой сложности данная задача не решена в полной мере, в том числе и для сопровождения людей в помещениях.

Цель статьи – разработка алгоритма сопровождения людей в помещениях с улучшенными качественными характеристиками за счет совместного использования признаков людей и лиц.

Структура алгоритма и составного дескриптора сопровождаемого человека

Предлагаемый алгоритм состоит из следующих основных этапов:

– обнаружение людей;

– формирование вектора признаков для каждого человека;

– обнаружение и распознавание лица в детектированной на предыдущем этапе области изображения, идентификация человека по лицу;

– установление соответствия между людьми на кадрах; индексация людей;

– определение их видимости на кадре;

– выделение рамкой человека при его присутствии в кадре.

Структура алгоритма показана на рис. 1.

Для установления соответствия между людьми на входном кадре и сопровождаемыми предлагается формировать составной дескриптор при описании каждого человека P_{ID}, который представляется в виде вектора признаков. При этом в данный дескриптор введены пространственные и СНС-признаки лица человека, по которым его можно идентифицировать и обеспечить высокую результативность сопровождения при возможности распознавания лица. Наличие в дескрипторе пространственных, СНС- и гистограммных признаков изображения людей позволит обеспечить правильное сопровождение человека при невозможности его идентификации по лицу.

Таким образом, предложенный составной дескриптор включает:

– координаты x_{det}^{P}, y_{det}^{P} центра области человека на кадре при предыдущем его обнаружении;

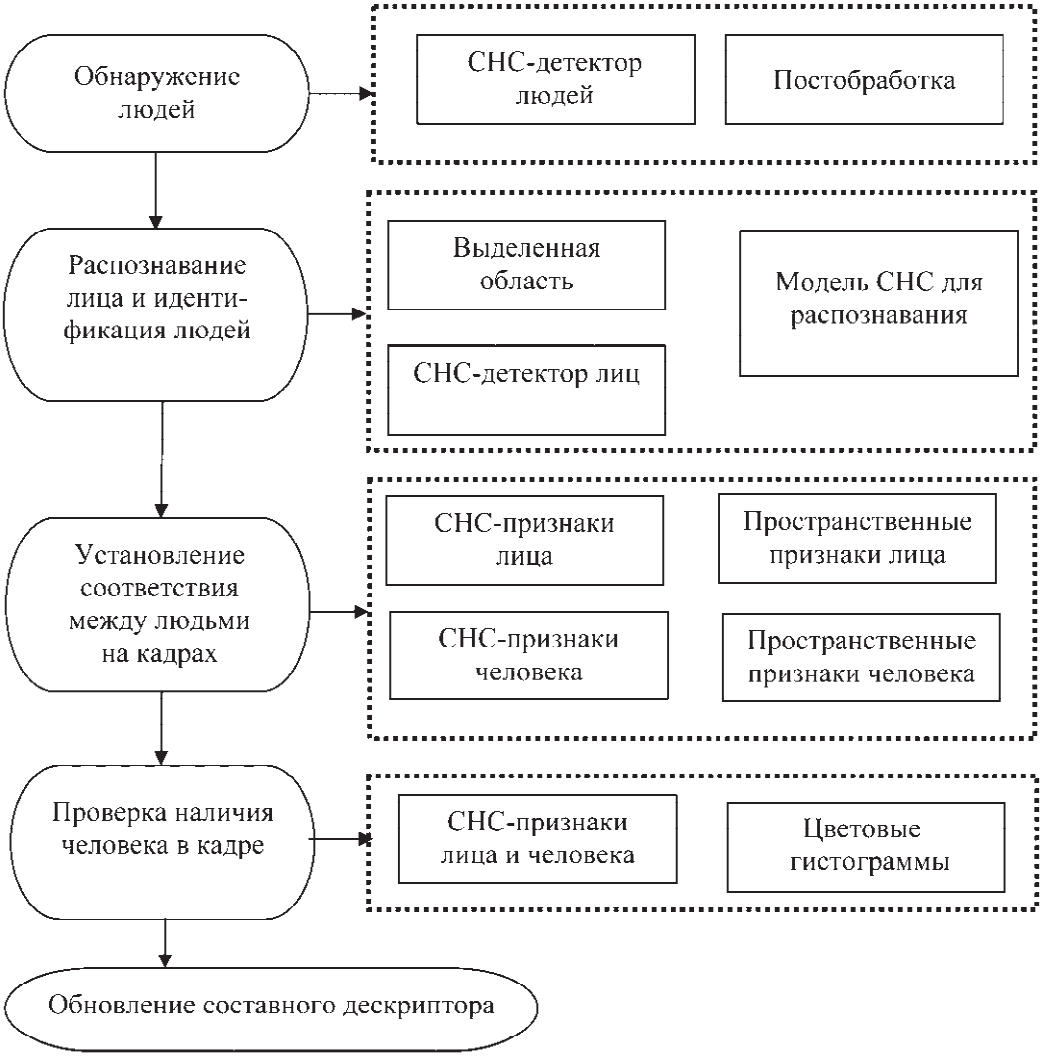

Рис. 1. Структура алгоритма сопровождения людей с идентификацией лиц

– ширину w_{det}^{P} и высоту h_{det}^{P} области человека на кадре при предыдущем его обнаружении;

– координаты центра области лица на кадре при предыдущем его обнаружении x_{det}^{F}, y_{det}^{F};

– ширину w_{det}^{F} и высоту h_{det}^{F} области лица на кадре при предыдущем его обнаружении;

– СНС-признаки f_{rec}^{FCNN} для последнего верного распознавания лица сопровождаемого объекта;

– расстояние d^{F} между вычисленными признаками и признаками изображения лица из базы данных;

– число N_{det}^{F} непрерывных результатов обнаружения лица при отсутствии распознавания;

– СНС-признаки $f_{det}^{FullPCNN}$ всей фигуры человека при последнем правильном обнаружении;

– СНС-признаки $f_{det}^{TopPCNN}$ верхней половины фигуры человека при последнем правильном обнаружении;

– гистограммные признаки N_{det}^{P} человека при последнем правильном обнаружении;

– индекс человека Nr^{P} в видеопоследовательности;

– имя человека Nm^{P}.

Модели сверточных нейронных сетей для обнаружения людей и распознавания лиц

В целях обеспечения режима реального времени комплексной задачи обнаружения и сопровождения людей необходимо для первого этапа применять быстродействующую СНС, которая должна отличаться еще и высокой точностью.

Среди существующих СНС модель YOLO направлена на уменьшение вычислительных затрат при обработке, а ее третья версия [11] использует улучшенную архитектуру для выделения признаков Darknet-53, содержащую 53 слоя и использующую 23 замыкающих соединения (shortcut connection), которые пропускают несколько слоев. При необходимости это позволяет обнулить влияние слоя на результат работы детектора, т.е. дает возможность изменять архитектуру сети так, чтобы конечное количество слоев определялось для конкретной задачи в процессе обучения. По результатам тестирования в метрике топ-5 для данной модели точность составляет 93,8 %.

Таким образом, в качестве модели СНС для детектирования объектов в предлагаемом алгоритме используется YOLOv3, поскольку данная архитектура характеризуется хорошей точностью обнаружения и удовлетворительным временем обработки.

Одним из эффективных подходов для повышения результативности распознавания лиц является использование функции потерь *ArcFace* (*Additive Angular Margin Loss for Deep Face Recognition*) [12], которая применима при обучении практически с любыми архитектурами СНС. На основе представленных в [13] данных видно, что наибольшая точность достигается для модели LResNet100E-IR, которая обучена с использованием ArcFace, однако она потребует значительных вычислительных затрат, что накладывает существенные ограничения при обработке видеопоследовательностей.

Модель MobileFaceNetwork характеризуется значительно меньшими вычислительными затратами, обеспечивая при этом высокую точность работы. Например, на базе данных LFW точность составляет 99,5 %, для сверточной нейронной сети LResNet100E-IR – 99,77 %. С учетом этого для распознавания лиц используется архитектура MobileFaceNetwork, которая формирует вектор из 128 признаков для лица.

Для обнаружения областей, содержащих лица, применяется мультизадачная трехкаскадная СНС MTCNN [14]. Первый ее каскад детектирует области-кандидаты, второй – подавляет ложные, третий – уточняет координаты и детектирует пять ключевых точек лица. Для данной модели на базе данных Wider Face [15] достигается точность 82 % правильных обнаружений. При этом обеспечивается уменьшение вычислительных затрат более чем в два раза по сравнению с моделью детектора лиц RetinaFace [16], которая использует архитектуру ResNet100 для достижения точности 92 %.

Установление соответствия между людьми на соседних кадрах

Основными входными данными для этого этапа являются результаты применения детектора на основе YOLOv3 к кадрам видеопоследовательности. С использованием выбранной СНС обнаруживаются изображения людей различных размеров на кадрах. В связи с тем, что иногда возможны случайные ложные обнаружения, применяется фильтрация таких объектов по размеру.

Идентификация по лицу. Сопровождение на основе идентификации по лицу состоит из следующих этапов:

– определение области поиска и ее масштабирование;

– обнаружение и локализация области лица и его распознавание;

– постобработка при обнаружении и распознавании нескольких лиц в области, которая описывает одного человека;

– постобработка для идентификации, если лицо не распознано на основе сравнения с базой данных.

Область поиска лица выделяется с учетом размеров детектированного фрагмента. Если его ширина более чем в три раза меньше его высоты, то анализируется только верхняя часть этого фрагмента. В противном случае анализируется вся область, описывающая человека. Если человек не скрыт за объектами и стоит или ходит, то осуществляется масштабирование области поиска к размерам входного слоя 300×300.

Для распознавания детектированных лиц применяется сверточная нейронная сеть MobileFaceNet с входным слоем размером [112×112] пикселов, которая обучена с использованием функции потерь ArcFace. Затем проводится идентификация лица путем определения расстояний d_i^F между выделенными признаками f_{curr}^{FCNN} на текущем кадре и признаками f_{db}^{FCNN} лиц в базе данных по формуле

$$d_i^F = \sum_{i=0}^{128} \sqrt{\left(f_{db_i}^{FCNN} - f_{curr}^{FCNN}\right)^2} .$$

Среди вычисленных расстояний выбирается минимальное. Если оно превышает заданную пороговую величину, то результат распознавания считается верным.

При правильном распознавании лица из базы данных признаки изображения лица f_{rec}^{FCNN}, расстояние d^F и имя человека Nm^P обновляются в составном дескрипторе.

В случае пересечения людей в обрабатываемой области могут быть обнаружены несколько лиц, которые принадлежат разным сопровождаемым объектам. Поэтому, если значение величины *пересечения над объединением IoU*(*Intersection over Union*) для лиц больше 0,8, то выполняется оценка схожести между вычисленными СНС-признаками лиц f_{curr}^{FCNN} и признаками f_{rec}^{FCNN} из составных дескрипторов сопровождаемых людей. Максимальная схожесть среди вычисленных определяет соответствие лица сопровождаемому объекту.

Если лицо не распознано с использованием признаков лиц из базы данных, то выполняется сравнение f_{curr}^{FCNN} и f_{rec}^{FCNN}. При превышении порогового значения для полученного результата сравнения распознавание считается положительным и выполняется обновление признака f_{rec}^{FCNN}. В противном случае обновляется только признак f_{curr}^{FCNN}.

Для минимизации случаев ложной идентификации человека P после перекрытия множества людей используется анализ признаков N_{det}^F и IoU по следующему правилу: если $N_{det}^F < 5$; $\mathrm{IoU}\left(P_{curr}^i, P_{curr}^j\right) > 0{,}3$ and $\mathrm{IoU}\left(P_{det}, P_{curr}\right) < 0{,}6$, то человек не идентифицируется.

При отсутствии идентификации людей по лицу. Когда лица не могут быть обнаружены или распознаны, сопровождение выполняется на основе алгоритма, включающего оценку наличия всей фигуры человека, формирование СНС-признаков для всей области и для верхней ее части и их накопление, формирование пространственных признаков и фильтрацию по расстоянию и размерам, вычисление схожести между всеми сопровождаемыми и обнаруженными на текущем кадре объектами и установление соответствия между ними, индексацию и именование людей, определение их видимости на кадре, выделение рамкой человека при его присутствии в кадре. Для формирования признаков человека используется СНС из 29 сверточных и одного полносвязного слоев, которая формирует вектор из 128 значений признаков для входного изображения [6].

При движении в помещении человек может зайти за объект фона, соответственно, признаки будут вычислены для верхней части его фигуры. Поэтому СНС-признаки вычисляются для всей фигуры человека и для ее верхней половины, если ширина выделенного объекта меньше его высоты, иначе принимается решение, что полученные СНС-признаки характеризуют верхнюю часть фигуры.

При $w_{curr}^P < h_{curr}^P$ оценка схожести между сопровождаемыми P_{tr} и обнаруженными на текущем кадре P_{curr} людьми выполняется на основе выражения с учетом СНС-признаков для последних пяти результатов детектора:

$$d(P_{tr}, P_{curr}) =$$
$$= \frac{\alpha}{N_{det}^{F}} \sum_{i=1}^{N_{det}^{F}} \left(\sqrt{\sum_{j=1}^{128} \left(f_{curr}^{FullPCNN} - f_{det}^{FullPCNN} \right)^2} \right) +$$
$$+ \beta \left(\sqrt{\left(x_{curr}^{P} - x_{prev}^{P} \right)^2} + \sqrt{\left(y_{curr}^{P} - y_{prev}^{P} \right)^2} + \right.$$
$$\left. + \sqrt{\left(w_{curr}^{P} - w_{prev}^{P} \right)^2} + \sqrt{\left(h_{curr}^{P} - h_{prev}^{P} \right)^2} \right),$$

где $f_{curr}^{FullPCNN}$, $\left(x_{curr}^{P},\ y_{curr}^{P} \right)$, w_{curr}^{P}, h_{curr}^{P} – соответственно СНС-признаки, координаты центра, высота и ширина объекта на входном кадре;

α, β – корректирующие коэффициенты.

Иначе, оценка схожести рассчитывается для СНС-признаков верхней половины фигуры человека и пересчета координат центра и высоты.

Фильтрация по расстоянию и размерам объектов предусмотрена для исключения ошибок из-за влияния людей, которые могут быть схожи по СНС-признакам, но находятся далеко от сопровождаемого объекта.

В результате расчета расстояний $d(P_{tr}, P_{curr})$ для всех сопровождаемых людей на предыдущих кадрах и обнаруженных объектов на входном кадре формируется матрица схожести, к которой применяется венгерский алгоритм решения задачи о назначениях [17]. В результате обнаруженному человеку на текущем кадре присваивается имя или индекс сопровождаемого. Особенность состоит в необходимости обработки P_{tr} с более ранних кадров, а не только с предыдущего, поскольку возможна кратковременная потеря оптической связи камеры с человеком. Это обусловлено тем, что в помещениях траектории движения людей часто пересекаются, объекты интереса перекрываются относительно камеры видеонаблюдения, например, при разговоре людей или при их совместном движении. Кроме того, человек может иметь несколько точек входа-выхода в кадре, частично или полностью перекрываться другими статическими объектами.

В целях уменьшения вероятности ложного изменения индексации после сложных случаев движения объектов с множественными перекрытиями в составном дескрипторе P_{ID}, на основе которого выполняется непрерывное сопровождение, обновляются только координаты объекта, его ширина и высота.

Проверка наличия человека в кадре

Для методов сопровождения через обнаружения важной является правильная работа детектора. Если человек не обнаруживается на одном или нескольких кадрах из-за ложного пропуска, то признаки в составном дескрипторе не будут обновляться. Это может привести к значительному отличию хранимых признаков от вычисляемых на последующих кадрах и, соответственно, к ошибкам индексации при сопровождении.

Кроме того, необходимо зафиксировать момент перекрытия человека другим объектом или выхода его из кадра для прекращения сопровождения. Актуальна также правильная индексация при возврате человека в кадр через некоторое время.

На первом шаге алгоритма для уточнения наличия человека в кадре используется поиск и распознавание лица для области, в которой был положительный результат детектора на предыдущем кадре. Если лицо найдено, то выполняется сравнение признаков f_{curr}^{FCNN} и f_{rec}^{FCNN}. При превышении полученным значением пороговой величины, человек считается присутствующим в кадре, область выделяется рамкой.

Если лицо не найдено, то анализируются СНС-признаки и признаки цветовых гистограмм областей человека на соседних кадрах. Для уменьшения влияния изменения освещенности изображения преобразуются из цветового пространства RGB (Red, Green, Blue) в цветовую модель HSV (Hue, Saturation, Value), и для оценки схожести используются только данные цветового тона. Человек считается присутствующим в кадре, если выполняется условие

$$d^{PCNN} \geq \varepsilon \quad \text{или} \quad R \leq \eta,$$

где

$$d^{PCNN} = \frac{1}{N_{det}^{F}} \sum_{i=1}^{N_{det}^{F}} \left(\sqrt{\sum_{j=1}^{128} \left(f_{curr}^{FullPCNN} - f_{det}^{FullPCNN} \right)^2} \right);$$

R – мера сходства гистограммных признаков цветового тона для изображений человека (при последнем его правильном обнаружении) и текущего кадра, которая вычисляется на основе евклидова расстояния;

ε, η – пороговые уровни: $\varepsilon = 0{,}3$; $\eta = 0{,}2$ [6].

Рис. 2. Примеры кадров пяти видеопоследовательностей
а – в – для видеоряда 1; *г – е* – для видеоряда 2; *ж – и* – для видеоряда 3; *к – м* – для видеоряда 4; *н – п* – для видеоряда 5

Результаты экспериментов

Для тестирования использовались видеопоследовательности с суммарным количеством кадров, равным 10 250 (рис. 2). Они получены со стационарной видеокамеры в помещениях с различным освещением, нелинейной траекторией движения, полным или частичным перекрытием людей с похожими внешними характеристиками,

выходом людей из помещения с последующим их возвращением в кадр и т.д.

Первая видеопоследовательность (рис. 2, *а – в*) включает 2320 кадров с низким качеством изображения. Наблюдается неравномерность освещения, высокий уровень теней. Число людей в кадре изменяется от одного до трех. Причем два из них имеют идентичную по цветовым характеристикам одежду, а также похожие рост, телосложение и цвет волос. Из рис. 2, *б* видно, что люди на сцене съемки расходятся. Затем они снова пересекаются, двигаясь по сложной траектории (рис. 2, *в*). Расположение лиц по отношению к камере и низкое качество съемки не позволяют идентифицировать людей на многих кадрах.

На рис. 2, *г – е* показаны кадры второго видеоряда из 1350 кадров. Признаки изображений двух человек на них схожи. Данное видео отличается более низким и неоднородным освещением, разным расположением лиц относительно камеры. Таким образом, низкое качество не позволяет идентифицировать людей на многих кадрах, т.е. сопровождение возможно только по признакам людей.

Видеокадры, представленные на рис. 2, *ж – и*, свидетельствуют о том, что на третьем тестовом видеоряде присутствуют два человека, которые передвигаются по помещению по сложной траектории, скрываясь частично за доской и столами, значительно удаляясь от видеокамеры. Данное тестовое видео состоит из 1280 кадров.

Четвертая видеопоследовательность (рис. 2, *к – м*) состоит из 3450 кадров, на которых присутствуют один-три человека. При движении нижняя половина фигуры человека часто скрывается за столом. Имеет место пересечение траекторий, люди многократно входят и покидают помещение.

На рис. 2, *н – п* представлены примеры кадров пятого видео, состоящего из 1850 кадров. На нем присутствуют два или три человека, признаки изображений двух из них схожи. Движение осуществляется по сложной траектории со значительным удалением от видеокамеры и множественным перекрытием, когда два человека скрыты третьим (см. рис. 2, *о*). Поэтому на значительном удалении от видеокамеры из-за низкого качества и большого угла отклонения от профиля лица невозможно идентифицировать людей на многих кадрах.

Для разметки использовался инструмент labelImg [18]. При этом видеопоследовательность разбивается на отдельные кадры, на них выделяются объекты. В качестве классов объектов ставятся индексы, соответствующие порядку появления людей в видео. Затем полученные текстовые файлы объединяются для оценки алгоритма отслеживания.

1. Сравнение характеристик алгоритмов

Параметр	Номер видеоряда (см. рис. 2)	Алгоритм из [5]	Алгоритм из [6]	Предложенный алгоритм
IDF	1	43,8	68,1	75,7
	2	85,1	94	96,7
	3	48,9	88,7	98,9
	4	92,4	92,9	96,4
	5	49,9	90,5	91,2
	1-5	60,1	84,9	**89,6**
FP	1	77	167	127
	2	9	28	12
	3	4	20	7
	4	8	7	10
	5	15	88	27
	1-5	133	310	**183**
FN	1	309	298	220
	2	114	105	63
	3	195	225	19
	4	39	25	38
	5	298	276	238
	1-5	995	929	**578**
MOTA	1	82,6	79,3	84,4
	2	89	88,4	93,5
	3	82,8	79,3	97,8
	4	96	97,7	96
	5	86	84	88,2
	1-5	86,6	84,8	**90,5**
MOTP	1	79,8	78,8	78,1
	2	79	76,8	76,7
	3	82,2	81,3	80,5
	4	83,5	82,2	81,6
	5	81,3	78,1	76,9
	1-5	**81,1**	79,2	78,5

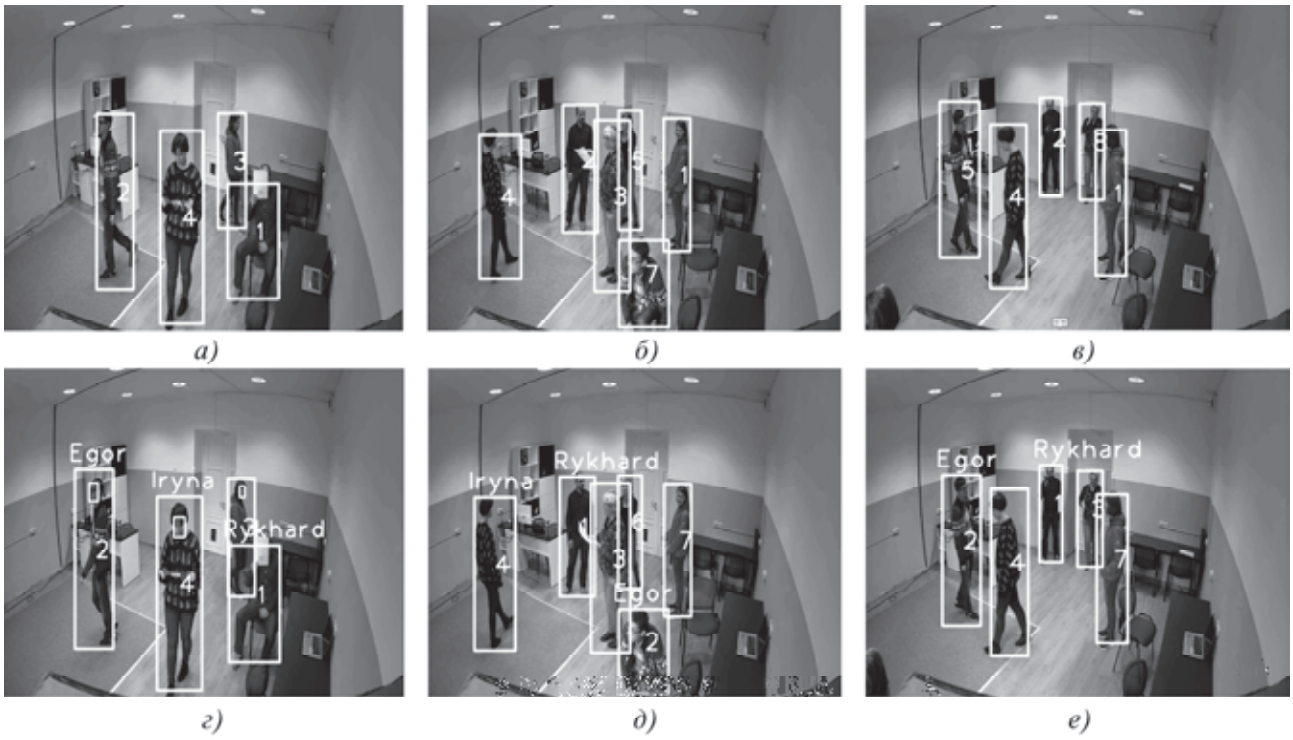

Рис. 3. Примеры сопровождения людей на видео
а – в – без идентификации по лицам; *г – е* – с использованием распознавания и идентификации

Предложенный здесь алгоритм сравнивается с другими (см. таблицу) путем определения следующих основных параметров, широко применяемых для оценки результативности сопровождения объектов [19]:

– IDF, указывающего процент правильной идентификации сопровождаемых объектов;

– MOTA, учитывающего количество ложноположительных (FP), ложноотрицательных (FN) результатов и IDF и характеризующего точность сопровождения объектов во времени с учетом восстановления траектории при кратковременном отсутствии объекта;

– MOTP, показывающего, насколько точно был локализован объект в кадре при сопровождении без учета обнаружения.

Анализ показывает, что процент правильной идентификации сопровождающих объектов значительно улучшен для всех тестовых видеопоследовательностей, параметр MOTA при комплексном тестировании на всех видео улучшен (хотя для четвертой видеопоследовательности не обеспечивается улучшение). Это связано с тем, что люди хорошо различимы по внешним признакам. Точность локализации человека в кадре немного хуже для представленного алгоритма. Скорость работы предложенного алгоритма на используемом для тестирования компьютере при наличии трех человек в кадре составляет 12 кадр./с.

Для сравнения на рис. 3 показаны результаты сопровождения объектов на идентичных фрагментах кадров сложной видеопоследовательности, включающей шесть человек.

На рис. 3, *а*, *г* представлен видеокадр, содержащий четырех человек с правильной индексацией для алгоритма сопровождения без идентификации по лицам (рис. 3, *а*) и с корректной индексацией и идентификацией (рис. 3, *г*). Изображение лица человека с индексом “3” отсутствует в базе данных. При появлении нового объекта в похожей по цвету одежде и пересечении его траектории движения с данным приводит к неправильному присвоению индексов, т.е. к ошибкам в сопровождении (рис. 3, *б*, *в*). При использовании предложенного алгоритма люди, изображения лиц которых присутствуют в базе данных, имеют корректные индексы и имена при отсутствии распознавания (лица не выделены рамками) на входных кадрах (рис. 3, *д*, *е*).

Заключение

Предложен алгоритм сопровождения людей в помещении на основе выполнения следующих основных этапов: обнаружение людей; формирование вектора признаков для каждого; обнаружение и распознавание лица в детектированной на предыдущем этапе области изображения; идентификация человека по лицу; установление соответствия между людьми на кадрах; индексация людей; уточнение сопровождения; выделение рамкой человека при его присутствии в кадре. Для тестирования алгоритма использованы пять видеопоследовательностей с суммарным количеством кадров 10 250. На их основе определены основные характеристики разработанного алгоритма: IDF=89,6; FP=183; FN=578; MOTA=90,5, MOTP=78,5. Алгоритм реализован на языке C++ с применением библиотек компьютерного зрения OpenCV 3.4 и dlib. Все процедуры обработки при обнаружении, сопровождении и идентификации людей на основе сверточной нейронной сети осуществлялись на графическом процессоре с использованием технологии параллельной обработки CUDA (Compute Unified Device Architecture).

Рассмотренный алгоритм способствует не только повышению качественных характеристик сопровождения людей внутри помещений, но и позволяет их идентицифицировать по изображениям лиц, хранимых в базе данных. Кроме того, такой подход может способствовать улучшению результативности сопровождения человека на видеопоследовательностях, формируемых мультикамерными системами внутреннего видеонаблюдения.

Библиографический список

1. **Лукьяница А. А., Шишкин А. Г.** Цифровая обработка видеоизображений. М.: Ай-Эс-Эс Пресс, 2009. 518 с.
2. **Методы** автоматического обнаружения и сопровождения объектов. Обработка изображений и управление / Б. А. Алпатов и др. М.: Радиотехника, 2008. 176 с.
3. **MOTChallenge**: The Multiple Object Tracking Benchmark [Электронный ресурс]. URL: https://motchallenge.net/ (дата обращения: 20.01.2019).
4. **Chahyati D., Fanany M. I., Arymurthy A. M.** Tracking People by Detection Using CNN Features // Proceedings of the 4th Information Systems International Conference (ISICO 2017). Indonesia, Bali, 6 – 8 November 2017. Bali, 2017. P. 167–172.
5. **Wojke N., Bewley A., Paulus D**. Simple Online and Realtime Tracking with a Deep Association Metric // Proceedings of the IEEE International Conference on Image Processing (ICIP). China, Beijing, 17 – 20 September 2017. Beijing, 2017. P. 3645 – 3649.
6. **Bohush R. P., Zakharava I. Yu.** Robust Person Tracking Algorithm Based on Convolutional Neural Network for Indoor Video Surveillance // Communications in Computer and Information Science. 2019. V. 1055. P. 289 – 300.
7. **Iqbal U., Milan A., Gall J.** PoseTrack: Joint Multi-person Pose Estimation and Tracking // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii, Honolulu, 21 – 26 July 2017. Honolulu, 2017. P. 4654 – 4663.
8. **People** Tracking and Re-identification by Face Recognition for RGB-D Camera Networks / K. Koide et al. // Proceedings of the 2017 European Conference on Mobile Robots (ECMR), 6 – 8 September 2017, Paris, France. 2007. P. 1 – 7.
9. **Schroff F., Kalenichenko D., Philbin J.** FaceNet: A Unified Embedding for Face Recognition and Clustering // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA, Boston, MA, 7 – 12 June 2015. Boston, MA, 2015. P. 815 – 823.
10. **Person** Re-identification for Improved Multi-person Multi-camera Tracking by Continuous Entity Association / N. Narayan et al. // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Hawaii, Honolulu, 21 – 26 July 2017. Honolulu, 2017. P. 64 – 70.
11. **YOLOv3:** An Incremental Improvement [Электронный ресурс]. URL: https://arxiv.org/abs/1804.02767 (дата обращения: 10.11.2018).
12. **ArcFace:** Additive Angular Margin Loss for Deep Face Recognition / J. Deng et al. // Computing Research Repository. 2019. arXiv:1801.07698v3 [Электронный ресурс]. URL: https://arxiv.org/abs/1801.07698.pdf (дата обращения: 16.06.2019).
13. **InsightFace Model Zoo** [Электронный ресурс]. URL: https://github.com/deepinsight/insightface/wiki/Model-Zoo MTCNN (дата обращения: 16.06.2019).
14. **Ma M. H., Wang J.** Multi-View Face Detection and Landmark Localization Based on MTCNN // Proceedings of the 2018 Chinese Automation Congress (CAC), 23 – 25 November, 2018, Shannxi Province, China. 2018. P. 4200 – 4205. doi:10.1109/cac.2018.8623535
15. **WIDER FACE:** A Face Detection Benchmark / Sh. Yang et al. // Computing Research Repository. 2015. arXiv:1511.06523 [Электронный ресурс]. URL: https://arxiv.org/pdf/1511.06523.pdf (дата обращения: 16.06.2019).

16. **RetinaFace:** Single-stage Dense Face Localisation in the Wild / J. Deng et al. // Computing Research Repository. 2019. arXiv:1905.00641v2 [Электронный ресурс]. URL: https://arxiv.org/pdf/1905.00641.pdf (дата обращения: 16.06.2019).

17. **Kuhn H. W.** The Hungarian Method for the Assignment Problem // Naval Research Logistics Quarterly. 1955. No. 2. P. 83 – 97.

18. **LabelImg** is a Graphical Image Annotation tool and Label Object Bounding Boxes in Images [Электронный ресурс]. URL: https://github.com/tzutalin/labelImg (дата обращения: 16.06.2019).

19. **Keni B., Stiefelhagen R.** Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics // EURASIP Journal on Image and Video Processing. 2008. V. 1. P. 1 – 10.

References

1. Luk'yanitsa A. A., Shishkin A. G. (2009). *Digital video processing.* Moscow: Ay-Es-Es Press. [in Russian language]

2. Alpatov Yu. A. et al. (2008). *Methods for automatic detection and tracking of objects. Image Processing and Management.* Moscow: Radiotekhnika. [in Russian language]

3. *MOTChallenge: The Multiple Object Tracking Benchmark.* Available at: https://motchallenge.net/ (Accessed: 20.01.2019).

4. Chahyati D., Fanany M. I., Arymurthy A. M. (2017). *Tracking People by Detection Using CNN Features.* Proceedings of the 4th Information Systems International Conference (ISICO 2017), pp. 167 – 172. Bali.

5. Wojke N., Bewley A., Paulus D. (2017). *Simple Online and Realtime Tracking with a Deep Association Metric.* Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 3645 – 3649. Beijing.

6. Bohush R. P., Zakharava I. Yu. (2019). Robust Person Tracking Algorithm Based on Convolutional Neural Network for Indoor Video Surveillance. *Communications in Computer and Information Science, Vol. 1055,* pp. 289 – 300.

7. Iqbal U., Milan A., Gall J. (2017). *PoseTrack: Joint Multi-person Pose Estimation and Tracking.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4654 – 4663. Honolulu.

8. Koide K. et al. (2017). *People Tracking and Re-identification by Face Recognition for RGB-D Camera Networks.* Proceedings of the 2017 European Conference on Mobile Robots (ECMR), pp. 1 – 7. Paris.

9. Schroff F., Kalenichenko D., Philbin J. (2015). *FaceNet: A Unified Embedding for Face Recognition and Clustering.* Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815 – 823. Boston.

10. Narayan N. et al. (2017). *Person Re-identification for Improved Multi-person Multi-camera Tracking by Continuous Entity Association.* Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 64 – 70. Honolulu.

11. *YOLOv3: An Incremental Improvement.* Available at: https://arxiv.org/abs/1804.02767 (Accessed: 10.11.2018).

12. Deng J. et al. (2019). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *Computing Research Repository. arXiv:1801.07698v3.* Available at: https://arxiv.org/abs/1801. 07698.pdf (Accessed: 16.06.2019).

13. *InsightFace Model Zoo.* Available at: https://github.com/deepinsight/insightface/ wiki/Model-Zoo MTCNN (Accessed: 16.06.2019).

14. Ma M. H., Wang J. (2018). *Multi-View Face Detection and Landmark Localization Based on MTCNN.* Proceedings of the 2018 Chinese Automation Congress (CAC), pp. 4200 – 4205. Shannxi Province.doi:10.1109/cac.2018.8623535

15. Yang Sh. et al. (2015). WIDER FACE: A Face Detection Benchmark. *Computing Research Repository. arXiv:1511.06523.* Available at: https://arxiv.org/pdf/1511.06523.pdf (Accessed: 16.06.2019).

16. Deng J. et al. (2019). RetinaFace: Single-stage Dense Face Localisation in the Wild. *Computing Research Repository. arXiv:1905.00641v2.* Available at: https://arxiv.org/pdf/1905.00641.pdf (Accessed: 16.06.2019).

17. Kuhn H. W. (1955). The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly,* (2), pp. 83 – 97.

18. *LabelImg is a Graphical Image Annotation tool and Label Object Bounding Boxes in Images.* Available at: https://github.com/tzutalin/labelImg (Accessed: 16.06.2019).

19. Keni B., Stiefelhagen R. (2008). Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP Journal on Image and Video Processing, Vol. 1,* pp. 1 – 10.

При цитировании использовать:

Богуш Р. П., Захарова И. Ю., Абламейко С. В. Алгоритм сопровождения людей на видеопоследовательности с использованием идентификации по лицам для наблюдения внутри помещений // Вестник компьютерных и информационных технологий. 2020. Т. 17, № 7. С. 3 – 14. doi: 10.14489/vkit.2020.07.pp.003-014

Bogush R. P., Zaharova I. Yu., Ablameyko S. V. (2020). Algorithm for tracking people on video sequences using facial identification for indoor observation. *Vestnik komp'yuternyh i informatsionnyh tekhnologiy, Vol. 17,* (7), pp. 3 – 14. [in Russian language] doi: 10.14489/vkit.2020.07.pp.003-014