

УДК 004.931

**КЛИЕНТ-СЕРВЕРНАЯ СИСТЕМА ДЛЯ ИДЕНТИФИКАЦИИ И МОНИТОРИНГА ПЕРЕМЕЩЕНИЯ ЛЮДЕЙ В ПОМЕЩЕНИЯХ НА ОСНОВЕ АНАЛИЗА ВИДЕОДАННЫХ**

*канд. техн. наук, доц. Р.П. БОГУШ, И.Ю. ЗАХАРОВА  
(Полоцкий государственный университет)*

*Разработаны алгоритмическое и программное обеспечение для идентификации и сопровождения множества людей в помещениях на основе анализа последовательностей изображений, получаемых со стационарных камер видеонаблюдения. Процедуры обнаружения и сопровождения людей используют сверточные нейронные сети. Особенностью алгоритма сопровождения является применение результатов идентификации по лицам для правильного установления соответствия на кадрах между людьми при их идентичных внешних характеристиках. Составной дескриптор изображения каждого человека включает признаки лиц, вычисленные на основе сверточной нейронной сети, и комплекс признаков изображения человека, что позволяет сопровождать людей при невозможности идентификации лиц. Программная реализация применяет библиотеку OpenCV для традиционных операций обработки изображений. Для основных процедур предложенной алгоритмической обработки используется программно-аппаратная архитектура параллельных вычислений, реализованная на GPU Nvidia по технологии CUDA. Такой подход позволяет обеспечить работу в режиме реального времени, если на кадре до пяти человек, при реализации на видеокарте NVIDIA GTX 1060, CPU Intel i7 – 5820k. Выходными данными являются обработанные последовательности изображений с координатами обнаруженного или сопровождаемого человека и его индексом, а также общее количество людей в кадре. Тестирование разработанной клиент-серверной системы выполнено с использованием персонального компьютера и мобильных устройств под управлением операционных систем Android и iOS.*

**Ключевые слова:** *идентификация людей, сопровождение, сверточные нейронные сети, CUDA, система «Умный дом».*

**Введение.** Для видеонаблюдения внутри помещений перспективно использование пространственно-распределенных систем, которые предполагают применение IP-камер. Использование автоматизированного анализа видеоданных в рамках системы «Умный дом» позволит обеспечить быстрый и надежный контроль нежелательного вторжения в жилище посторонних людей, поведения несовершеннолетних, состояния пожилых людей и др.

Обнаружение и сопровождение множества людей в помещении осложняется рядом факторов: низким уровнем освещенности; неоднородным задним фоном, фрагменты которого могут быть схожи по форме, текстуре и цвету с изображениями людей; наличием теней и, соответственно, изменяющимся задним фоном; множественными перекрытиями людей между собой и другими объектами в помещении; высокой схожестью признаков разных сопровождаемых людей; достаточно быстрым движением людей и в ряде случаев изменяющимся ускорением и нелинейной трансформацией траектории движения. Задача сопровождения множества людей заключается в определении местоположения выделенных объектов на текущем кадре видеопоследовательности относительно предыдущего, что позволит построить индивидуальные траектории их движения.

Для сопровождения людей применяются методы на основе сравнения признаков изображений людей и лиц. В первой группе методов в настоящее время наиболее результативным является сопровождение через обнаружение. Этот подход использует ансамбль из детектора объектов и алгоритма для объединения результатов обнаружений на двух кадрах. Эффективное решение проблемы объединения позволяет 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) [1–3], которые достаточно устойчивы к изменениям освещенности, динамическому заднему фону и позволяют осуществлять детектирование даже в случае частичных перекрытий, что также обеспечивает повышение качества сопровождения. Однако, при схожести признаков людей, сложных траекториях движения результативность их работы значительно уменьшается.

Целью работы является создание алгоритмического и программного обеспечения для обнаружения, идентификации и сопровождения множества людей в помещениях с возможностью их практической работы в режиме реального времени на современных вычислительных средствах.

**1. Алгоритмическое обеспечение.** Для обнаружения, идентификации и сопровождения людей разработан алгоритм, который основан на использовании СНС для всех основных задач и состоит из 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;
- определение видимости на кадре;
- выделение рамкой человека при его присутствии в кадре.

Общая схема работы алгоритма представлена на рисунке 1.

Для обеспечения обработки в режиме реального времени на первом этапе (обнаружение людей) необходимо применять быстродействующую СНС, которая также должна отличаться высокой точностью. Среди существующих СНС предъявляемым требованиям удовлетворяет модель YOLO v4 [4], обученная на базе данных MS COCO. Далее для областей кадра, в которых обнаружены люди, формируются признаки при помощи модифицированной модели СНС, состоящей из 29 сверточных слоев и формирующей 128 признаков. После выделения признаков проводится сопоставление информации об уже имеющихся (сопровождаемых) объектах с информацией, полученной от детектора. С этой целью сравниваются признаки для последних  $n$  верных обнаружений объекта, после чего формируется матрица соответствий, которая поступает на вход венгерского алгоритма.

Далее следует поиск лиц в сопровождаемых областях. Выделение области поиска лица выполняется на основе анализа размеров детектированного фрагмента. Если его ширина меньше его высоты более чем в три раза, то анализируется только верхняя часть этого фрагмента, иначе анализируется вся область, описывающая человека. Для обнаружения областей, содержащих лица, применяется мультизадачная трехкаскадная СНС MTCNN [5]. Для данной модели на базе данных Wider Face [6] достигается точность 82% правильных обнаружений и при этом обеспечивается уменьшение вычислительных затрат более чем в два раза по сравнению с моделью детектора лиц RetinaFace [7], которая использует архитектуру ResNet100 для достижения точности 92%.

Полученная на предыдущем шаге область кадра, содержащая лицо, поступает для распознавания. Для этого этапа применяется СНС MobileFaceNet с размером входного слоя  $112 \times 112$  пикселей, которая характеризуется значительно меньшими вычислительными затратами и обеспечивает при этом высокую точность работы (например, на базе данных LFW точность составляет 99,5%, а для СНС LResNet100E-IR – 99,77% [8]). Архитектура MobileFaceNetwork формирует вектор из 128 признаков для лица человека. Затем выполняется идентификация лица путем вычисления расстояний между выделенными признаками на текущем кадре и признаками лиц в базе данных. Среди вычисленных расстояний выбирается минимальное, и если оно больше заданной пороговой величины, то результат распознавания принимается верным. При правильном распознавании лица из базы данных признаки изображения лица, расстояние и имя человека обновляются в составном дескрипторе. В случае пересечения людей в обрабатываемой области может быть обнаружено несколько лиц, которые принадлежат разным сопровождаемым объектам. Поэтому, если значение величины пересечения над объединением (IoU) для лиц больше заданного значения, выполняется оценка схожести между вычисленными СНС-признаками лиц и признаками из составных дескрипторов сопровождаемых людей, максимальная схожесть среди которых определяет соответствие лица сопровождаемому объекту.

Если лицо не распознано с использованием базы данных, то выполняется сравнение признаков обнаруженного лица с соответствующими данными составного дескриптора, который включает:

- координаты центра области человека на кадре при предыдущем его обнаружении;
- ширину и высоту области человека на кадре при предыдущем его обнаружении;
- координаты центра области лица на кадре при предыдущем его обнаружении;
- ширину и высоту области лица на кадре при предыдущем его обнаружении;
- СНС-признаки для последнего верного распознавания лица сопровождаемого объекта ( $f_{rec}^{FCNN}$ );
- расстояние между вычисленными признаками и признаками изображения лица из базы данных;
- число непрерывных результатов обнаружения лица при отсутствии распознавания;
- СНС-признаки всей фигуры человека при последнем правильном обнаружении ( $f_{det}^{FullPCNN}$ );
- СНС-признаки верхней половины фигуры человека при последнем правильном обнаружении ( $f_{det}^{TopPCNN}$ );
- гистограммные признаки человека при последнем правильном обнаружении;
- индекс человека в видеопоследовательности;
- имя человека.

В случаях невозможности обнаружения или распознавания лиц 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 [9].

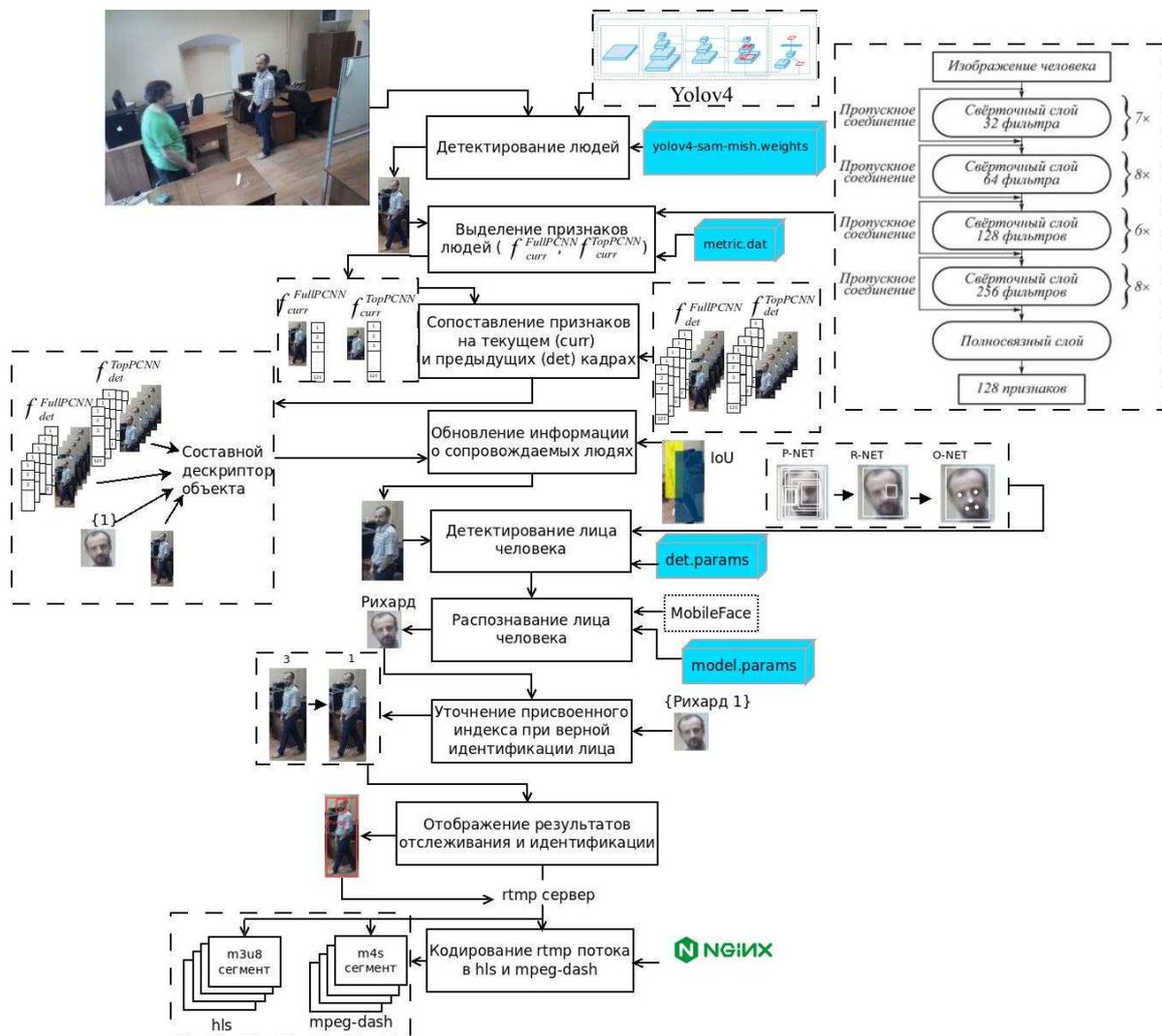


Рисунок 1. – Общая схема работы алгоритма

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.

В результате расчета расстояний для всех сопровождаемых людей на предыдущих кадрах и обнаруженных объектов на входном кадре формируется матрица схожести, к которой применяется венгерский алгоритм, решающий задачу о назначениях, благодаря чему обнаруженному человеку на текущем кадре присваивается имя или индекс сопровождаемого. Для уточнения результатов идентификации формируется словарь, который состоит из имени человека и индекса, под которым это имя впервые было присвоено объекту. Посредством такого уточнения алгоритм сопровождения восстанавливает верный индекс человека, даже если он отсутствовал в кадре продолжительное время, не затрачивая при этом большого количества памяти.

**2. Описание программной реализации модуля.** С учетом специфики использования программного модуля обнаружения, распознавания и сопровождения людей к приложению были сформированы следующие требования:

- работа в режиме реального времени;
- минимизация передаваемого трафика по каналам связи;
- кроссплатформенность;

- обеспечение стабильной работы при сбоях в сети передачи данных;
- возможность подключения большого числа клиентов к видеоданным, формируемым камерой наблюдения.

Для получения видеоданных используется IP видеокamera, обработка выполняется на сервере, к которому может подключиться клиент с использованием мобильного устройства или персонального компьютера, т.е. используется централизованная архитектура системы видеонаблюдения.

Видеокamera формирует поток, для передачи которого используется протокол, обеспечивающий минимальную задержку при передаче данных. Программная реализация алгоритма использует OpenCV для захвата видеопотока, извлечения кадров, изменения его размера, кадрирования, отображения текста и графических фигур. Программно-аппаратная архитектура параллельных вычислений применяется для всех операций на основе СНС, которая позволяет существенно увеличить вычислительную производительность благодаря использованию GPU Nvidia с технологией CUDA. Ее составной частью является ускоренная на GPU библиотека примитивов для сверточных нейронных сетей cuDNN, имеющая свободное распространение и доступная зарегистрированным на NVIDIA разработчикам.

В программной реализации выделены три этапа: обнаружение людей на видеопотоке, отслеживание обнаруженных людей, идентификация отслеживаемых людей по лицам. Для этапа обнаружения используется достаточно быстрый и точный фреймворк машинного обучения, написанный на языке C, Darknet. Данный фреймворк позволяет реализовать СНС Yolov4, которая осуществляет детектирование объектов заданных классов на уровне, сравнимом с GoogleNet-Inceptionv4, затрачивая при этом на свою работу меньше вычислительных мощностей компьютера. При реализации отслеживания также использована библиотека с открытым исходным кодом для алгоритмов машинного обучения Dlib, характеризующаяся возможностью реализации СНС разной структуры и назначения. Для последующей идентификации отслеживаемых людей по лицу была использована библиотека машинного обучения MxNet.

Передача обработанного видеопотока осуществляется через http-сервер nginx, который осуществляет передачу медиаданных значительно быстрее по сравнению с apache, обладает широким набором модулей для упрощения процесса разработки и на основе которого реализованы многие высоконагруженные системы (например, Яндекс, Mail.Ru, ВКонтакте). На мобильные устройства поток передается в формате hls, а на персональный компьютер – в формате mpeg-dash.

Система реализована на основе объектно-ориентированного подхода, каждый модуль приложения является классом. Класс «Детектор» осуществляет обнаружение людей на изображении, использует метод для загрузки структуры СНС, принимающий в качестве входных данных тензор весовых коэффициентов и структуру СНС, и метод для детектирования людей, использующий в качестве входных данных текущий кадр видеопоследовательности, а в качестве выходных – координаты обнаруженных областей с людьми. Для управления составными дескрипторами объектов реализован модуль, главный метод которого в качестве входных данных принимает координаты обнаруженных людей и текущий кадр, а в качестве выходных данных формируются составные дескрипторы для каждого отслеживаемого объекта.

Управляющий класс является надстройкой над несколькими классами, которые предназначены для заполнения и обновления составных дескрипторов при сопровождении объектов. Среди них класс для выделения признаков из областей изображения, в которых содержится человек с методами для загрузки весов СНС и преобразования областей изображений в набор признаков по заданному алгоритму.

Реализован класс для имплементации венгерского алгоритма. Его главный метод осуществляет решение задачи о назначении областей, обнаруженных детектором, к уже существующим отслеживаемым объектам. На вход метода поступает матрица схожести между областями, обнаруженными детектором, и уже существующими отслеживаемыми объектами. В качестве выходных данных формируется вектор, в котором каждому объекту назначен индекс обнаруженного детектором объекта.

Для идентификации людей реализован класс обнаружения и идентификации лиц со следующими методами: для загрузки СНС-моделей; для детектирования лица человека (в качестве выходных данных формирует координаты лица человека и положение особых точек для обнаруженного лица); для построения признаков лица; для формирования базы данных (в качестве выходных данных формируется словарь соответствий идентификации лица, который выдает из базы данных имя максимально похожего лица и степень соответствия).

**3. Результаты тестирования.** Тестирование разработанного модуля выполнено с использованием персонального компьютера и мобильных устройств под управлением ОС Android и iOS. Серверная часть системы реализована на компьютере с видеокартой NVIDIA GTX 1060, CPU Intel i7 – 5820k, RAM 32Gb. Проверка работы программного обеспечения выполнена в браузере Mozilla Firefox на персональном компьютере под управлением ОС Ubuntu (рисунок 2).

Для мобильных устройств под управлением iOS тестирование проведено с использованием iPhone SE в браузере Safari; подключение к серверу с адресом 192.168.43.218, что подтверждается рисунком 3, а (из него видно, что воспроизведение ведется по ссылке <http://192.168.43.218/hls/stream.m3u8>). Браузер Safari

автоматически отображает плеер, в котором имеется возможность воспроизведения видео в полноэкранном режиме. Для мобильных устройств под управлением Android тестирование выполнено с использованием Samsung Galaxy S9 в браузере устройства. Воспроизведение ведется по ссылке <http://192.168.43.218/hls/stream.m3u8> (рисунок 3, б).

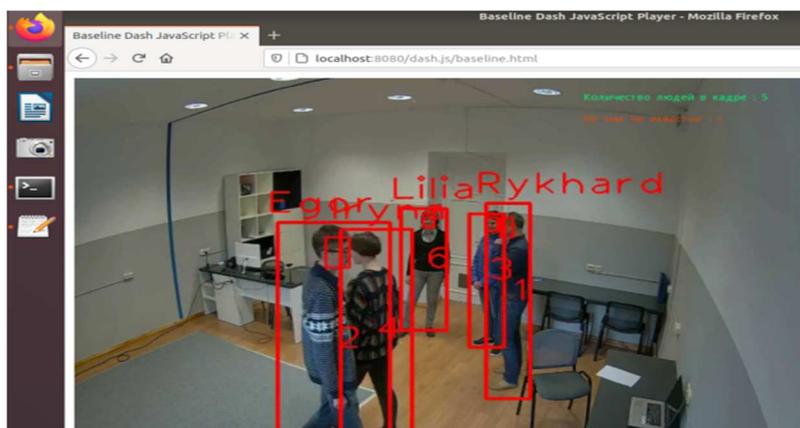
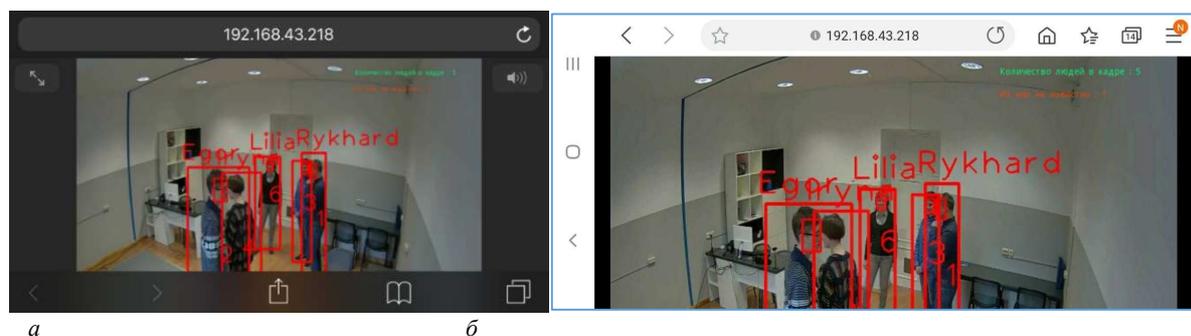


Рисунок 2. – Отображение результатов работы алгоритма на персональном компьютере



а

б

а – iPhone SE; б – Samsung Galaxy S9

Рисунок 3. – Отображение результатов работы на мобильных устройствах

Из рисунков 2, 3 видно, что в кадре одновременно находятся 5 человек, из них один изначально не был зафиксирован в базе данных, и ему соответствует индекс «3» без имени. Остальные объекты, имена и лица которых присутствовали в базе данных, верно идентифицированы и сопровождаются на кадре и в случае без возможности распознавания лица (например, Игуна).

**Заключение.** Для обнаружения, идентификации и сопровождения множества людей в помещениях на основе анализа видеопоследовательностей разработаны алгоритмы на основе сверточных нейронных сетей и выполнена их программная реализация. Библиотека OpenCV применяется для традиционных операций обработки изображений, а для наиболее ресурсоемких процедур алгоритмической обработки используется программно-аппаратная архитектура параллельных вычислений по технологии CUDA. Такой подход позволил обеспечить работу в режиме реального времени, если на кадре до пяти человек, при реализации на видеокарте NVIDIA GTX 1060, CPU Intel i7 – 5820k. Тестирование разработанной клиент-серверной системы выполнено с использованием персонального компьютера и мобильных устройств под управлением операционных систем Android и iOS.

#### ЛИТЕРАТУРА

1. Chahyati, D. Tracking People by Detection Using CNN Features / D. Chahyati, M.I. Fanany, A.M. Arymurthy // In Proceedings of the 4th Information Systems International Conference (ISICO 2017). – 2017. – P. 167–172.
2. Wojke, N. Simple online and realtime tracking with a deep association metric / N. Wojke, A. Bewley, D. Paulus // In Proceedings of the IEEE International Conference on Image Processing (ICIP). – 2017. – P. 3645–3649.
3. Bohush, R. Robust Person Tracking Algorithm Based on Convolutional Neural Network for Indoor Video Surveillance / R. Bohush, I. Zakharava // Communications in Computer and Information Science. – 2019. – Vol. 1055. – P. 289–300.
4. YOLOv4: Optimal Speed and Accuracy of Object Detection [Electronic resource]. – Mode of access: <https://arxiv.org/pdf/2004.10934.pdf>. – Date of access: 29.09.2020.

5. Ma M.H. Multi-View Face Detection and Landmark Localization Based on MTCNN// M.H. Ma, J. Wang // In Proceedings of the 2018 Chinese Automation Congress (CAC). – 2018. – P. 4200–4205. – DOI:10.1109/cac.2018.8623535.
6. WIDER FACE: A Face Detection Benchmark [Electronic resource] // Sh. Yang [et al.] // Computing Research Repository ; 2015 arXiv:1511.06523. – Mode of access: <https://arxiv.org/pdf/1511.06523.pdf>. – (дата обращения 16.06.2019).
7. RetinaFace: Single-stage Dense Face Localisation in the Wild [Electronic resource] // J. Deng [et al.] // Computing Research Repository ; 2019 arXiv:1905.00641v2. – Mode of access: <https://arxiv.org/pdf/1905.00641.pdf>. – Date of access: 16.06.2019.
8. InsightFace [Electronic resource]. – Mode of access: <https://github.com/deepinsight/insightface/wiki/Model-Zoo-MTCNN>. – Date of access: 29.09.2020.
9. Богущ, Р.П. Алгоритм сопровождения людей на видеопоследовательностях с использованием сверточных нейронных сетей для видеонаблюдения внутри помещений / Р.П. Богущ, И.Ю. Захарова // Компьютерная оптика. – 2020. – Т. 44, № 1. – С. 109–116. – DOI: 10.18287/2412-6179-CO-565.

Поступила 28.10.2020

## CLIENT-SERVER SYSTEM FOR PEOPLE IDENTIFICATION AND TRACKING IN INDOOR BASED ON VIDEO DATA ANALYSIS

**R. BOHUSH, I. ZAKHARAVA**

*In this paper we present software that was developed for identification and tracking of multiple people indoors. Proposed approach based on the analysis of image sequences obtained from stationary CCTV cameras. Person detection and tracking procedures use convolutional neural network architectures. Tracking algorithm is characterized by collecting face recognition results for the correct assignment of a person's external features in different frames. An integral descriptor defines each person and contains convolutional neural network based face features and person's image features. His structure allows us to track people when faces can't be recognized properly. The software implementation uses the OpenCV library for basic image processing operations. For the main procedures of our algorithm we use software and hardware architecture for parallel computing that is implemented on an Nvidia GPU and CUDA technology. The proposed approach achieves real-time processing if there are five or less people in a frame simultaneously. The computational experiments are conducted on a PC with GPU NVIDIA GTX 1060 and CPU Intel i7-5820k. The output data is processed image sequence with rectangular areas of people's location and their index. Overall number of people in the current image is showed as well. The developed client-server system was tested using a personal computer and mobile devices running on Android and iOS operating systems.*

**Keywords:** *people tracking and identification, convolutional neural networks, CUDA, "Smart home" system.*