



Столыпинский  
вестник

Научная статья

Original article

УДК 004.89

**ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ЛИЦ ЧЕЛОВЕКА НА  
ИЗОБРАЖЕНИЯХ С КАМЕР ВИДЕОНАБЛЮДЕНИЯ**

**DETECTION AND RECOGNITION OF HUMAN FACES IN IMAGES FROM CCTV  
CAMERAS**

**Малыгин Марк Андреевич**, студент 2 курса магистратуры по направлению подготовки 09.04.01 Информатика и вычислительная техника, программа «Искусственный интеллект», ФГБОУ ВО «Иркутский национальный исследовательский технический университет» (664074 Россия, г. Иркутск, ул. Лермонтова, д. 83)

**Malygin Mark Andreevich**, 2nd year Master's student in the field of study «Informatics and Computer Engineering» (Artificial Intelligence), Federal State Budget Educational Institution of Higher Education «Irkutsk National Research Technical University» (83 Lermontova st., Irkutsk, 664074 Russia)

**Аннотация.** В статье представлена разработка модели искусственного интеллекта для обнаружения и распознавания лица человека на изображениях с камер видеонаблюдения. При использовании классических систем видеонаблюдения необходим человек для постоянного наблюдения. Систему автоматического распознавания лиц на изображениях камеры наблюдения с минимальным вмешательством человека и меньшими затратами можно использовать для

помощи многим организациям, например, таким как правоохранительные органы в выявлении подозреваемых или пропавших без вести, а также лиц въезжающих на запрещенную территорию. Однако в процессе распознавания изображения есть множество трудностей, таких как масштабирование и повороты лица, или изменение интенсивности света. В статье описывается система распознавания лиц на основе изображений, полученных с камер видеонаблюдения с использованием различных методов извлечения признаков и распознавания лиц. Предлагаемая система включает в себя получение изображений с камер видеонаблюдения, предварительную обработку изображений, обнаружение лиц, извлечение из полученных изображений и распознавание. Мы используем два алгоритма извлечения признаков: анализ основных компонентов (РСА) и сверточная нейронная сеть (CNN). Распознавание выполняется путем применения этих методов к набору данных, полученных в реальном времени при различных настройках, таких как уровень освещенности, вращение и масштабирование, для моделирования и оценки производительности. В ходе исследования, система показала конкурентоспособные результаты с точностью более 90%.

**Abstract.** The article presents the development of an artificial intelligence model for detecting and recognizing a person's face in images from CCTV cameras. When using classic video surveillance systems, a person is needed for constant monitoring. With minimal human intervention and lower cost, automatic facial recognition system on surveillance camera images can be used to help many organizations, such as law enforcement agencies, in identifying suspects or missing persons, as well as persons entering a prohibited territory. However, there are many difficulties in the image recognition process, such as scaling and rotating the face, or changing the light intensity. This article aims to develop a face recognition system based on images obtained from CCTV cameras using various methods of feature extraction and face recognition. The proposed system includes acquisition of images from CCTV cameras, image preprocessing, face detection, image extraction and recognition. We use two feature extraction algorithms: Principal Component Analysis (PCA) and Convolutional Neural Network (CNN). Recognition is performed by applying these methods to a set of

real-time data at various settings such as light level, rotation, and scaling to simulate and evaluate performance. During the study, the system showed competitive results with an accuracy of more than 90%.

**Ключевые слова:** сверточная нейронная сеть, детектирование объектов, точность распознавания, системы видеонаблюдения, анализ основных компонентов, распознавание человеческих лиц

**Keywords:** convolutional neural network, object detection, recognition accuracy, CCTV, principal component analysis, recognizing human faces

### **Введение**

Современные организации сталкиваются с серьезными проблемами безопасности; для её обеспечения требуется специально обученные сотрудники, чтобы обеспечить необходимую безопасность. Однако люди совершают ошибки, влияющие на безопасность. Система видеонаблюдения в настоящее время используется для различных целей в быту. Развитие видеонаблюдения превратило простой пассивный мониторинг в комплексную интеллектуальную систему безопасности. Обнаружение лиц и его новые возможности для безопасного контроля доступа, финансовых транзакций и т.д. В последнее время биометрические системы приобрели новое значение. Благодаря достижениям в области микроэлектроники и систем распознавания лиц биометрия стала экономически жизнеспособной. Распознавание лиц является неотъемлемой частью биометрии, в основе которой изображение лица человека соотносится с данными, хранящимися в системе. Извлекаются и реализуются черты лица, используя эффективный алгоритм, и вносятся некоторые изменения для улучшения существующей модели. Распознавание лиц может быть использовано для различных прикладных приложений, таких как идентификация преступников, систем безопасности и аутентификации. Как правило, система распознавания лиц включает в себя этапы обнаружения человека и обработки изображений для облегчения распознавания.

По этой причине в современном мире распознавание лиц стало необходимостью, поскольку с развитием методов распознавания количество идентификаций человека ежедневно увеличивается. За последние два десятилетия распознаванию лиц уделяется большое внимание из-за его различных особенностей, анализа изображений и понимания предметных областей. Распознавание лиц становится важным в других областях, таких как обработка изображений, анимация [1], безопасность [2], человеко-компьютерное взаимодействие [3] и медицина [4]. Распознавание лиц является естественным, неинвазивным и простым в использовании. На сегодняшний день системы распознавания лиц имеют широкий выбор возможностей в области общественной безопасности, развлечений, управления посещаемостью и финансовых платежей. Хотя современные системы распознавания лиц хорошо работают в относительно контролируемых средах, они сталкиваются со значительными проблемами при использовании в существующих системах наблюдения из-за низкого разрешения изображения, фоновых помех, вариаций освещения, а также положения лица и его выражения.

В системах распознавания лиц определяют три этапа работы: предварительная обработка изображения, извлечение признаков и метод классификации для распознавания [5]. Элементы, извлеченные из лица, такие как рот, нос, брови и т. д., являются геометрическими элементами. Обнаруженное и обработанное лицо сравнивается с базой данных лиц, чтобы определить, кто этот человек.

Окклюзии лица, такие как бороды и аксессуары (очки, головные уборы и маски), также включены в алгоритмы распознавания лиц, что делает объект разнообразным и сложным для функционирования в несимуляционной среде. Идеальная система распознавания лиц должна выдерживать изменения освещения, выражений лица, поз и окклюзии, а также захватывать наибольшее количество лиц.

### **Обнаружение и распознавание лиц на изображениях с камер видеонаблюдения**

Метод состоит из четырех основных шагов: получение изображения, улучшение изображения, обнаружение лица и распознавание лица.

Изображения необходимо получать с исходной (обычно аппаратной) камеры, что делает это первым шагом в последовательности рабочего процесса. Система видеонаблюдения постоянно считывает изображения, которые являются предварительно обработанными входными данными.

Для получения изображения используется камера высокого разрешения с интернет-протоколом (IP), которая снимает видео со скоростью 15 кадров в секунду и разрешением 1248x720. Камера делает снимки, которые будут сохранены и доступны для просмотра с помощью программных инструментов.

База данных лиц включает в себя лица тех, кого она распознает. Поскольку распознавание лиц включает алгоритмы классификации, каждое изображение в наборе данных помечено. Изображения лиц каждого человека имеют свои уникальные метки.

После получения изображения предварительная обработка изображения подготавливает его к дальнейшей обработке. Предварительная обработка включает два основных этапа: преобразование изображения в оттенки серого и методы обнаружения границ.

С камеры мы получаем изображение RGB. Пиксель RGB состоит из 1 пикселя красного цвета в сочетании с пикселями синего и зеленого цветов. Изображение RGB усложнило вычисления, поскольку 1 пиксель состоит из 8 бит, поэтому в RGB он стал бы 24-битным. В изображении в градациях серого каждый пиксель является скаляром, поэтому это будет 8-битное изображение.

Фильтр Кэнни определяет края на фотографиях путем обнаружения резких изменений цвета на фотографиях. Мы используем его для улучшения краев изображения. Чем больше улучшений, тем большей точности можно достичь в распознавании выражений лица. Фильтр состоит из фильтров Гаусса и Собеля.

### **Распознавание лиц**

Следующим шагом после получения изображения с камеры является обнаружение лица на изображениях с помощью алгоритма Виолы-Джонса, который различает области лица и не лица.

Алгоритм Виолы-Джонса — первый алгоритм, который обеспечивает конкурентоспособную скорость обнаружения объектов в режиме реального времени. Он обеспечивает надежность с высокой скоростью обнаружения, удобен для приложений реального времени, поскольку может обрабатывать два кадра в секунду. После применения этого для распознавания изображения используются различные методы классификации.

Лицо, обнаруженное методом Виолы-Джонса, извлекается и изменяется в размере как изображение  $40 \times 40$ , а затем используется с помощью различных методов извлечения признаков для поиска особенностей лица.

Мы использовали метод анализа основных компонент (PCA) для извлечения особенностей лица. PCA – это метод, используемый для уменьшения размеров изображений в нашем наборе данных. Он находит характеристики изображений, отличие и отклонение пикселей в одном столбце от другого [6].

### **Валидация и метрики**

В данном случае, средняя точность (mAP) – это оценка общей площади под кривой графика «Recall vs Precision». Для того чтобы получить mAP, нужно выполнить интерполяцию точности на каждом уровне. Затем на основе тестового набора, который был разработан для оценки моделей, обученные модели оценим с помощью тестового набора с использованием средней точности (mAP).

### **Распознавание лиц с использованием сверточной нейронной сети**

Сверточные нейронные сети (CNN) имеют совершенно другую архитектуру, чем простая нейронная сеть. Они имеют входной слой, сверточные слои, слой максимального пула и, в конце, полносвязные нейронные сети.

Для обучения оптимизации весов используется оптимизатор Adam. Оптимизатор Adam – это метод стохастического градиентного спуска, основанный на адаптивной оценке моментов первого и второго порядка. Этот

метод прост в реализации, эффективен в вычислительном отношении, требует мало памяти, инвариантен к диагональному изменению масштаба градиентов и хорошо подходит для задач, которые требуют больших объемов данных и/или параметров. Этот метод также подходит для нестационарных целей и задач с очень шумными и/или редкими градиентами.

### Результаты

В CNN необходимо обучить наш набор данных. Была собрана база данных лиц, в которой содержится около двух тысяч лиц пяти разных человек. Для их распознавания необходимы метки классов, которые пронумерованы от 1 до 5, и каждая метка имеет несколько изображений. Таким образом, метка 1 содержит 400 изображений. На рисунке 1 показаны примеры изображений в наборе данных.



Рисунок 1 – Образцы изображений лиц, используемых для распознавания

Имеющиеся данные были обучены и получена средняя точность предсказания 95,67%. Модель обучили за 4000 эпох. На некоторых этапах обучения точность увеличивалась, а в некоторых точках снижалась, но в конце мы получили максимальную точность 95,67%. Для этого использовались равномерно распределенные данные для обучения и тестового наборов, как показано на рисунке 2.



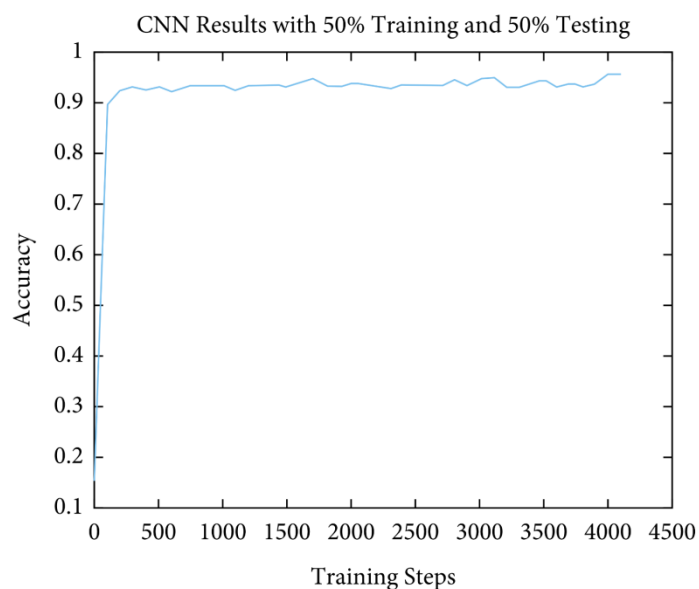


Рисунок 2 – Результаты обучения модели с использованием равномерно распределенных данных

Точность предсказания 95% была получена при использовании большего количества данных для обучения, чем данных для тестирования. Эта точность была получена за 300 эпох и дальше оставалась на одном уровне до самого конца обучения, как показано на рисунке 3.

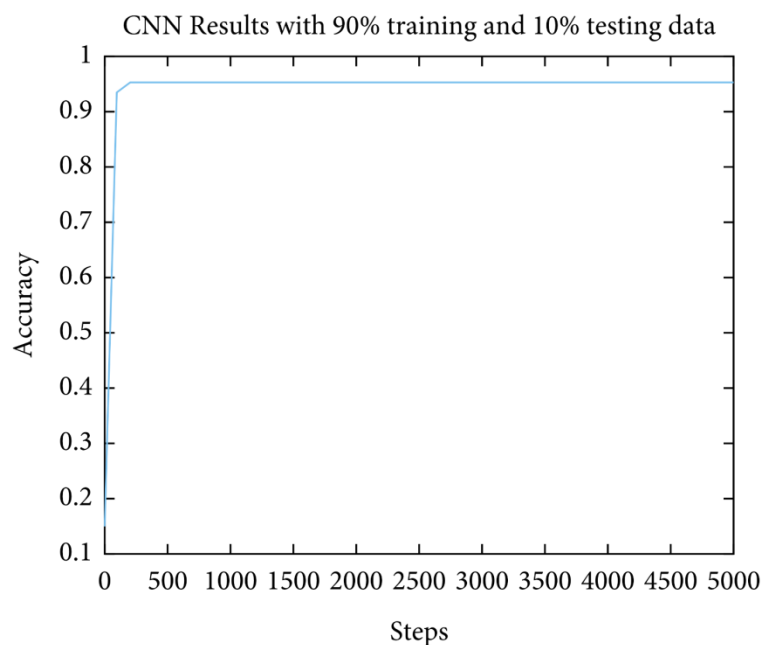


Рисунок 3 – Результаты обучения модели с использованием 90% данных для обучающего и 10% – для тестового наборов



Максимальная точность предсказания равная 97,5% была получена при использовании 80% данных для обучения и 20% данных тестового набора. Модель обучена за 5000 эпох и получена максимальная точность 97,5%. На некоторых этапах обучения точность то увеличивалась, то снижалась, но после достижения 1500 эпох оставалась на одном уровне до конца обучения, как показано на рисунке 4.

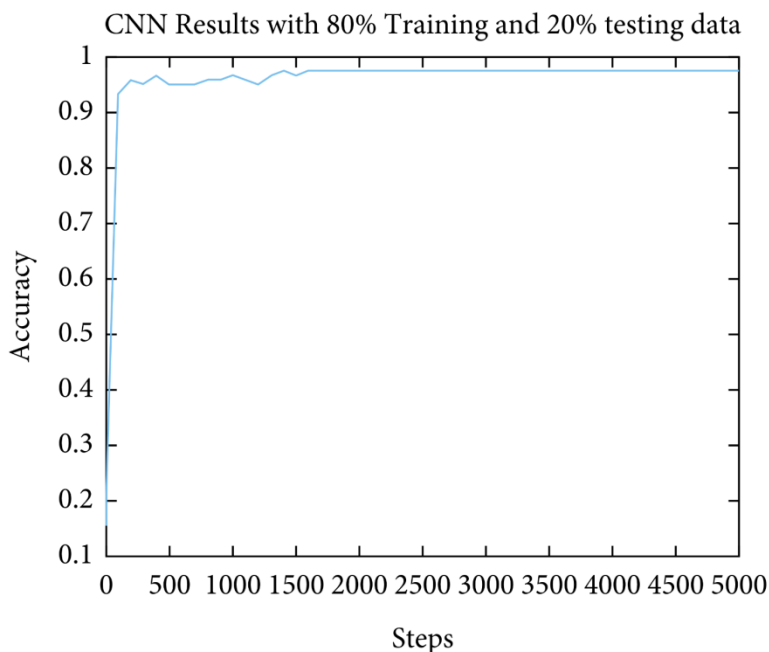


Рисунок 4 – Результаты обучения модели с использованием 80% данных для обучающего и 20% – для тестового наборов

### Выводы

В ходе проведенного исследования было осуществлено обучение модели сверточной нейронной сети для автоматического распознавания лиц на основе изображений с камер видеонаблюдения. Для обучения модели использовался собственный датасет, состоящий из двух тысяч лиц пяти разных людей. Результаты показали, что модель для распознавания лиц может обнаруживать лица на изображении без переобучения.

Модель, обученная на равномерно распределенных данных для обучения и валидационного (тестового) набора показала наиболее высокие показатели точности предсказания и mAP составил 97,5%. При увеличении количества данных для обучения и снижении валидационных данных модель показывает

наименьший mAP, который составляет 95%. В местах большого скопления объектов распознавание лиц требует не только достаточно высокого mAP, но и высокой скорости распознавания. Для задачи, где необходимо наиболее высокое значение точности распознавания, например, в случае задачи идентификации лица человека, данная модель может быть достаточно эффективной.

### Литература

1. A. Deepali, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, Modeling stylized character expressions via deep learning, pp. 136–153, Springer, Taipei, Taiwan, November 2016. [http://dx.doi.org/10.1007/978-3-319-54184-6\\_9](http://dx.doi.org/10.1007/978-3-319-54184-6_9) – (дата обращения: 22.05.2023).
2. S. T. Saste and S. M. Jagdale, “Emotion recognition from speech using MFCC and DWT for security system,” in *Proceedings of the International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 1, pp. 701–704, IEEE, Coimbatore, India, April 2017. <https://doi.org/10.1109/ICECA.2017.8203631> – (дата обращения: 22.05.2023).
3. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis et al., “Emotion recognition in human-computer interaction,” *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001. <https://doi.org/10.1109/79.911197> – (дата обращения: 22.05.2023).
4. J. Edwards, H. J. Jackson, and P. E. Pattison, “Emotion recognition via facial expression and affective prosody in schizophrenia,” *Clinical Psychology Review*, vol. 22, no. 6, pp. 789–832, 2002. [https://doi.org/10.1016/S0272-7358\(02\)00130-7](https://doi.org/10.1016/S0272-7358(02)00130-7) – (дата обращения: 22.05.2023).
5. S. Umer, B. Chandra Dhara, and B. Chanda, “Face recognition using fusion of feature learning techniques,” *Measurement*, vol. 146, 2019. <https://doi.org/10.1016/j.measurement.2019.06.008> – (дата обращения: 22.05.2023).
6. J. Yu, K. Sun, F. Gao, and S. Zhu, “Face biometric quality assessment via light cnn,” pp. 25–32, 2018, *Pattern Recognition Letters*107. <https://doi.org/10.1016/j.patrec.2017.07.015> – (дата обращения: 22.05.2023).

7. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016. <https://doi.org/10.1109/CVPR.2016.90> – (дата обращения: 22.05.2023).
8. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. <https://doi.org/10.1109/5.726791> – (дата обращения: 22.05.2023).
9. P. Kamency, M. Benco, T. Mizdos, and R. Radil, “A new method for face recognition using convolutional neural network,” *Digital Image Processing And Computer Graphics*, vol. 15, pp. 663–672, 2017. <https://doi.org/10.15598/aeec.v15i4.2389> – (дата обращения: 22.05.2023).

### References

1. A. Deepali, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, Modeling stylized character expressions via deep learning, pp. 136–153, Springer, Taipei, Taiwan, November 2016. [http://dx.doi.org/10.1007/978-3-319-54184-6\\_9](http://dx.doi.org/10.1007/978-3-319-54184-6_9) – (date of the application: 22.05.2023).
2. S. T. Saste and S. M. Jagdale, “Emotion recognition from speech using MFCC and DWT for security system,” in *Proceedings of the International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 1, pp. 701–704, IEEE, Coimbatore, India, April 2017. <https://doi.org/10.1109/ICECA.2017.8203631> – (date of the application: 22.05.2023).
3. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis et al., “Emotion recognition in human-computer interaction,” *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001. <https://doi.org/10.1109/79.911197> – (date of the application: 22.05.2023).
4. J. Edwards, H. J. Jackson, and P. E. Pattison, “Emotion recognition via facial expression and affective prosody in schizophrenia,” *Clinical Psychology Review*, vol. 22, no. 6, pp. 789–832, 2002. [https://doi.org/10.1016/S0272-7358\(02\)00130-7](https://doi.org/10.1016/S0272-7358(02)00130-7) – (date of the application: 22.05.2023).

5. S. Umer, B. Chandra Dhara, and B. Chanda, “Face recognition using fusion of feature learning techniques,” *Measurement*, vol. 146, 2019. <https://doi.org/10.1016/j.measurement.2019.06.008> – (date of the application: 22.05.2023).
6. J. Yu, K. Sun, F. Gao, and S. Zhu, “Face biometric quality assessment via light cnn,” pp. 25–32, 2018, *Pattern Recognition Letters* 107. <https://doi.org/10.1016/j.patrec.2017.07.015> – (date of the application: 22.05.2023).
7. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016. <https://doi.org/10.1109/CVPR.2016.90> – (date of the application: 22.05.2023).
8. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. <https://doi.org/10.1109/5.726791> – (date of the application: 22.05.2023).
9. P. Kamency, M. Benco, T. Mizdos, and R. Radil, “A new method for face recognition using convolutional neural network,” *Digital Image Processing And Computer Graphics*, vol. 15, pp. 663–672, 2017. <https://doi.org/10.15598/aeee.v15i4.2389> – (date of the application: 22.05.2023).

© Малыгин М.А., 2023 Научный сетевой журнал «Столыпинский вестник» №5/2023.

**Для цитирования:** Малыгин М.А. ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ЛИЦ ЧЕЛОВЕКА НА ИЗОБРАЖЕНИЯХ С КАМЕР ВИДЕОНАБЛЮДЕНИЯ// Научный сетевой журнал «Столыпинский вестник» №5/2023.