

Распознавание образов

Проблемы и особенности 2D, 3D и 4D-систем распознавания документов, удостоверяющих личность

В.В. АРЛАЗАРОВ^{1,II}

¹ Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

^{II} ООО «Смарт Энджинс Сервис», г. Москва, Россия

Аннотация. В данной работе подробно рассмотрены особенности входных данных (изображений документов, полученных путем сканирования, фотографий и видеопоследовательностей), используемых в современных системах распознавания удостоверяющих документов. Предложена классификация распознающих систем по виду обрабатываемых входных данных. Рассмотрены имеющиеся проблемы и особенности введенных 2D, 3D и 4D-распознаваний. Сформулированы концептуальные модели и подход к построению универсальной промышленной системы распознавания, способной одинаково обрабатывать входные данные, относящиеся к различным классам.

Ключевые слова: распознавание документов, идентификационные документы, искусственный интеллект, OCR, машинное обучение, обработка изображений.

DOI: 10.14357/20790279220301

Введение

Задача автоматического извлечения данных из изображений документов, удостоверяющих личность, стала предметом изучения начиная с 2000-гг, в частности, чтобы обеспечить более эффективный ввод данных и проверку личной информации при регистрации в гостинице, посадке в самолет и т.д. [1]. Хотя изначально основными способами ввода были планшетные и специализированные сканеры. Распознавание документов с помощью камер стало особенно актуально в течение последних 10 лет [2] благодаря широкому распространению портативных камер и мобильных устройств, например смартфонов.

Востребованность систем автоматического анализа и распознавания документов, поддерживающие входные данные в виде фотографий документов [3–5], растет изо дня в день. Дополни-

тельные сложности, возникающие при обработке такого типа входных данных, по сравнению со сканированными изображениями, довольно многочисленны. Во-первых, фон в случае сканированного изображения, как правило, однороден, в то время как фотография документа может быть сделана на произвольном фоне (рис. 1). Разнородный и неконтролируемый фон может оказаться препятствием для точного детектирования и локализации документа, особенно если фон со структурной точки зрения слишком сложен, имеет много высококонтрастных линий или локальных областей, или содержит текст, который используемый алгоритм распознавания может ложно «признать» за часть документа. Вторая существенная трудность распознавания фотографий – это неконтролируемые условия освещения. Изображения, полученные со

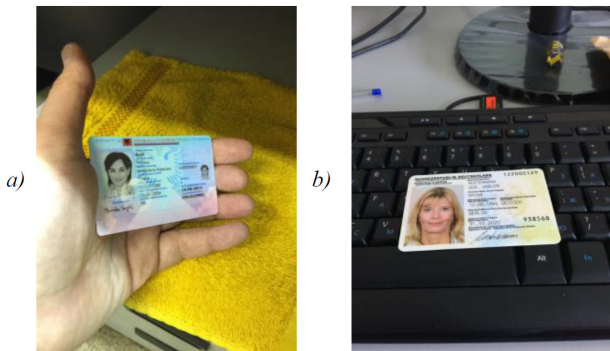


Рис. 1. Примеры фотографий удостоверений личности, сделанных с помощью смартфона

сканеров, всегда равномерно освещены, в то время как фотография может быть получена при слабом и неравномерном освещении, быть пере- или недоэкспонированной (выполненной с недостаточной экспозицией). Неравномерное освещение создает проблемы для детектирования и локализации документа на фотографии, а также для анализа структуры документа, распознавания текста и других компонентов автоматизированной системы анализа документов [6, 7]. Еще одной серьезной проблемой является расфокусировка [8, 9] или наличие размытия [10].

Но самым важным различием между изображениями, полученными с помощью сканеров и камер, является геометрическое положение документа на изображении. На изображении, которое получено с помощью веб-камеры или камеры мобильного устройства, документ может быть повернут на все три угла Эйлера по отношению к оптической системе камеры. Если камера рассматривается в рамках модели камеры-обскуры, то семейство возможных геометрических преобразований документа теперь является подмножеством проективных, что значительно усложняет задачу предварительной локализации документа [11, 12].



Рис. 2. Примеры фотографий внутреннего российского паспорта: а) книжный разворот, б) изогнутая страница

Более того, на одном изображении возможно несколько проективных преобразований для разных частей документа, как в случае захвата обеих страниц, например паспорта (рис. 2,а). Поскольку параметры объектива камеры могут быть неизвестны, изображения документов также могут быть подвержены радиальным искажениям [13]. Наконец, если сам документ не является жестким, он может быть подвержен физической деформации, например, возможен изгиб бумажных страниц паспорта (рис. 2,б).

Помимо фотографий, использование веб-камер и мобильных устройств для получения изображений документов привело к появлению другого вида входных данных – последовательности видеок кадров вместо одной фотографии [14]. Использование видеопоследовательности вместо фотографии позволяет лучше контролировать процесс распознавания документов, а также делает входные данные менее подверженными подделке, поскольку видео сложнее подделать по сравнению с единичным загружаемым изображением.

С точки зрения распознавания и анализа документов, использование нескольких входных изображений одного и того же объекта дает дополнительные преимущества: оказывается возможным применять методы фильтрации и уточнения для повышения точности обнаружения и локализации объекта [15, 16], использовать так называемые методы «сверхразрешения» [17] для получения изображений более высокого качества, а также улучшить результаты распознавания текста путем накопления пок кадровых результатов распознавания и интеграции в один наиболее достоверный [18]. На рис. 3 представлены примеры видеок кадров, снятых для документа, удостоверяющего личность.

Таким образом, видно, что современные системы распознавания удостоверяющих документов получают на вход различные по своей природе данные, что, в частности, позволяет нам интуитивно классифицировать задачу распознавания документов по типам входных данных, приводит нас к рассмотрению задач 2D, 3D и 4D распознавания документов.

В настоящей статье рассмотрено, какие проблемы приносят в промышленные системы распознавания различные входные данные, представлено описание ключевых блоков, возникающих в архитектурах систем распознавания, обрабатывающих соответствующие входные данные, а также предложен способ построения единой системы распознавания удостоверяющих документов, способной обрабатывать входные данные разной природы.

1. Проблемы и особенности 2D, 3D и 4D-распознавание

Перед тем как перейти к рассмотрению ключевых проблем и особенностей, формально введем понятия 3D и 4D-распознаваний.

Определение 1. Будем называть распознавание изображения документа «распознаванием 2D», когда документ расположен в плоскости изображения.

Определение 2. Будем называть распознавание изображения документа «распознаванием 3D», когда документ на изображении может быть повернут на все три угла Эйлера по отношению к оптической системе камеры, которая рассматривается в рамках модели камеры-обскуры.

Определение 3. Будем называть распознавание серии изображений документов «распознаванием 4D», когда исследуемые изображения являются отдельными кадрами видеопотока и могут быть упорядочены с учетом временной шкалы.

2D-системы распознавания удостоверяющих документов традиционно имеют дело с изображениями, полученными с помощью сканеров. В данном случае, помимо обычных для такого рода систем подзадач (включающих идентификацию визуальных элементов документа, анализ структуры, распознавание текстовых полей и т. д.), возникают дополнительные проблемы: произвольное смещение и поворот документа на изображении, а также, в некоторых случаях, неизвестное разрешение изображения (например, когда процесс получения изображения со сканера не контролируем). Тем не менее, все указанные подзадачи, связанные с определением геометрии документа, решаются фактически на плоскости (2D-распознавание).

Если изображение документа получено с помощью веб-камеры или камеры мобильного

устройства (случай 3D-распознавания), то вместо анализа фактически плоского объекта необходимо анализировать трехмерную сцену, где документ должен быть локализован с учетом проективных преобразований, возможных нелинейных искажений и произвольного фона. Помимо геометрических особенностей расположения документа, дальнейшие методы их анализа должны учитывать также возможную расфокусировку, размытие, неравномерное или недостаточное освещение, а также блики на отражающей поверхности документа.

Наконец, если входные данные – это не единичное изображение, а последовательность видеок кадров (случай 4D-изображения), документ необходимо обрабатывать во времени с учетом меняющихся условий съемки. При этом избыточность визуальной информации может быть использована для повышения надежности результатов анализа и распознавания документов, а изменяющиеся условия могут быть использованы для обнаружения и анализа оптических переменных эффектов (голограммы, оптически переменные краски и т. д.), которые широко используются для защиты документов, удостоверяющих личность.

Все перечисленные проблемы и особенности, возникающие при 2D, 3D и 4D распознавании удостоверяющих документов, схематично представлены на рис. 4.

2. Модель универсальной системы распознавания

Несмотря на различные условия получения входных данных (будь то изображения или последовательность видеок кадров), различные модели геометрического положения документа и различные дополнительные искажения (такие как наличие размытия, бликов, неравномерного освещения



Рис. 3. Примеры видеок кадров полученных для удостоверения личности, где сцена изменяется от кадра к кадру

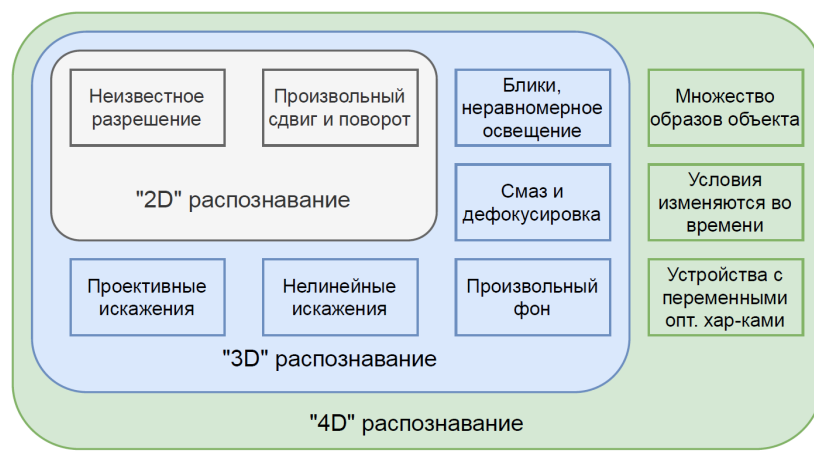


Рис. 4. Особенности 2D, 3D и 4D-систем распознавания документов, удостоверяющих личность

и т.д.), непосредственно объект распознавания – документ, удостоверяющий личность – остается неизменным.

Поэтому разумным является подход построения современной системы распознавания идентификационных документов, учитывающей специфику различных режимов получения изображения. При этом отдельные компоненты при таком подходе и их взаимосвязь должны обеспечивать поддержку множества различных типов идентификационных документов, а архитектура должна поддерживать возможность включения дополнительных методов анализа и обработки документа.

В настоящем разделе представлен список предложений, которых следует придерживаться при построении универсальной распознающей системы (универсальность в данном случае подразумевает поддержку входных данных, соответствующих понятиям 2D, 3D, 4D).

- 1. Использование методов локализации шаблона, устойчивых к проективным искажениям.** Применение таких методов позволяет автоматически обработать проективные преобразования, нелинейные искажения или произвольный фон, которые усложняют проблему распознавания 3D-документов.
- 2. Определение разрешения документа по результатам локализации шаблона.** Информацию о фактическом разрешении следует вычислять после этапа локализации и идентификации документа. Документы, удостоверяющие личность, относятся к классу документов с фиксированной («жесткой») структурой. Поэтому, зная физические размеры идентифицированного документа и определив границы документа на изображении, можно с высокой точностью вычислить фактическое разрешение.
- 3. Независимый анализ качества изображения.**

Анализ проблем расфокусировки, размытости, бликов или неравномерного освещения и прочие проблемы, свойственные случаю распознавания документов в 3D и 4D следует вынести в отдельную подсистему. Выделение данного модуля облегчит анализ в сложных случаях и позволит на ранних этапах отфильтровать входные данные плохого качества.

- 4. Анализ элементов документа на скорректированных и отмасштабированных изображениях.** Использование уже скорректированного (от проективных искажений) изображения шаблона документа позволяет применять одинаковые методы анализа отдельных элементов как в случае распознавания документов в 2D, так и в 3D и 4D.
- 5. Обеспечение доступа к результатам распознавания предыдущих кадров.** В случае 4D-распознавания документов, условия съемки, изменяющиеся во времени, могут быть компенсированы на отдельных этапах процесса благодаря доступу к временному хранилищу цикла распознавания и возможности использовать результаты распознавания из предыдущих видеокadres.
- 6. Использование модулей комбинирования результатов распознавания.** Добавление модулей комбинирования позволяет повысить ожидаемую точность распознавания полей документа за счет использования информации из нескольких кадров, а также проводить анализ оптически изменяемых элементов.
- 7. Фильтрация успешно распознанных полей.** Идея настоящего предложения заключается в реализации модуля, останавливающего распознавание отдельных полей на очередном кадре видеопоследовательности, если такое поле уже успешно распознано. Из-за особенностей удостоверяющих документов зачастую возникает ситуация, когда различные информационные поля

требуют разного количества входных кадров для успешного распознавания. Ранний останов по отдельным полям для случая 4D-распознавания позволяет сократить общее количество распознаваемых элементов, существенно сокращая итоговое время распознавания документа.

Следование указанным предложениям позволяет построить систему распознавания удостоверяющих документов, одинаково хорошо работающую с различными типами входных данных.

Заключение

Распознавание документов, удостоверяющих личность, выходит за рамки классических задач оптического распознавания символов. От автоматизированных систем распознавания документов, удостоверяющих личность, требуется не только выполнить поиск изменяемых полей и данных, подлежащих сбору и структурированию, но и выполнить проверку личности, а также предотвратить мошенничество, причем цена ошибки чрезвычайно высока. Большинство исследований направлено на создание таких систем анализа удостоверений личности, которые ориентированы на определенный тип документов или конкретный способ получения изображения документов. Однако в современном мире растет спрос на решение комплексной проблемы – распознавание документов, удостоверяющих личность, на широком спектре изображений, например, скан-копии, фотографии или видеокadres. При этом условия получения таких изображений могут быть различными, то есть неконтролируемыми.

В настоящей работе представлена классификация распознающих систем по виду обрабатываемых входных данных. Введено формальное определение 2D, 3D и 4D-распознаваний, рассмотрены проблемы и особенности, связанные с каждым типом распознавания. Кроме того, сформулированы концептуальные модели и подход к построению универсальной промышленной системы распознавания, способной одинаково обрабатывать входные данные, относящиеся к различным классам.

Литература

1. *Lladós J., Lumbreras F., Chapaprieta V. and Queralt J.* ICAR: Identity card automatic reader // *Proceedings of Sixth International Conference on Document Analysis and Recognition*. 2001. P. 470–474.
2. *Mollah A.F., Majumder N., Basu S. and Nasipuri M.* Design of an optical character recognition system for camera-based handheld devices // *International Journal of Computer Science Issues*. 2011. Vol. 8. No. 4. P. 283–289.
3. *Ryan M. and Hanafiah N.* An examination of character recognition on id card using template matching approach // *Procedia Computer Science*. 2015. Vol. 59. P. 520–529.
4. *Attivissimo F., Giaquinto N., Scarpetta M. and Spadavecchia M.* An automatic reader of identity documents // *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 2019. P. 3525–3530.
5. *Xu J. and Wu X.* A system to localize and recognize texts in oriented id card images // *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*. 2018. P. 149–153.
6. *Bulatov K., Matalov D. and Arlazarov V.* MIDV-2019: challenges of the modern mobile-based document OCR // *Twelfth International Conference on Machine Vision (ICMV 2019)*. 2020. Vol. 11433. P. 717–722.
7. *Li X., Zhang B., Liao J. and Sander P.V.* Document rectification and illumination correction using a patch-based CNN // *ACM Trans. Graph.* 2019. Vol. 38. No. 6.
8. *Asad F., Ul-Hasan A., Shafait F. and Dengel A.* High performance OCR for camera-captured blurred documents with LSTM networks // *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. 2016. P. 7–12.
9. *Chernov T.S., Razumnuy N.P., Kozharinov A.S., Nikolaev D.P. and Arlazarov V.V.* Image quality assessment for video stream recognition systems // *ICMV 2017*. 2018. Vol. 10696. No. 106961U. P. 1–8.
10. *Nunnagoppula G., Deepak K.S., Harikrishna G., Rai N., Krishna P.R. and Vesdapunt N.* Automatic blur detection in mobile captured document images: Towards quality check in mobile based document imaging applications // *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*. 2013. P. 299–304.
11. *Miao L. and Peng S.* Perspective rectification of document images based on morphology // *2006 International Conference on Computational Intelligence and Security*. 2006. Vol. 2. P. 1805–1808.
12. *Takezawa Y., Hasegawa M. and Tabbone S.* Robust perspective rectification of camera-captured document images // *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 2017. Vol. 06. P. 27–32.
13. *Kunina I., Gladilin S. and Nikolaev D.* Blind radial distortion compensation in a single image using fast Hough transform // *Computer Optics*. 2016. Vol. 40. No. 3. P. 395–403.

14. *Chazalon J., Gomez-Krämer P., Burie J., Coustaty M., Eskenazi S., Luqman M., Nayef N., Rusiñol M., Sidère N. and Ogier J.* SmartDoc 2017 Video Capture: Mobile Document Acquisition in Video Mode // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017. Vol. 04. P. 11–16.
15. *Skoryukina N., Arlazarov V. and Nikolaev D.* Fast method of id documents location and type identification for mobile and server application // 2019 International Conference on Document Analysis and Recognition (ICDAR). 2019. P. 850–857.
16. *Zhukovsky A., Nikolaev D., Arlazarov V., Postnikov V., Polevoy D., Skoryukina N., Chernov T., Shemiakina J., Mukovozov A., Konovalenko I. and Povolotsky M.* Segments graph-based approach for document capture in a smartphone video stream // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017. Vol. 01. P. 337–342.
17. *Haris M., Shakhnarovich G. and Ukita N.* Recurrent back-projection network for video super-resolution // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 3892–3901.
18. *Petrova O., Bulatov K., Arlazarov V.V. and Arlazarov V.L.* Weighted combination of per-frame recognition results for text recognition in a video stream // Computer Optics. 2021. Vol. 45. No. 1. P. 77–89.

Арлазаров Владимир Викторович. Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Заведующий отделом, кандидат технических наук. Количество печатных работ: 130. Область научных интересов: искусственный интеллект, машинное обучение, системы распознавания, информационные технологии. E-mail: vva777@gmail.com

Problems and features of 2D, 3D, and 4D identity document recognition systems

V.V. Arlazarov^{I,II}

^I Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

^{II} Smart Engines Service LLC, Moscow, Russia

Abstract. In this paper, the input data used in modern identification document recognition systems are considered in detail. The classification of recognition systems according to the type of input data being processed is proposed. The existing problems and features of defined 2D, 3D and 4D recognition are considered in detail. Conceptual model and approaches for the construction of a universal industrial recognition system capable of processing input data belonging to different classes in the same way are proposed.

Keywords: *document recognition, identification documents, artificial intelligence, OCR, machine learning, image processing*

DOI: 10.14357/20790279220301

References

1. *Lladós J., Lumbreras F., Chapaprieta V. and Queralt J.* ICAR: Identity card automatic reader // Proceedings of Sixth International Conference on Document Analysis and Recognition. 2001. P. 470–474.
2. *Mollah A.F., Majumder N., Basu S. and Nasipuri M.* Design of an optical character recognition system for camera-based handheld devices // International Journal of Computer Science Issues. 2011. Vol. 8. No. 4. P. 283–289.
3. *Ryan M. and Hanafiah N.* An examination of character recognition on id card using template matching approach // Procedia Computer Science. 2015. Vol. 59. P. 520–529.
4. *Attivissimo F., Giaquinto N., Scarpetta M. and Spadavecchia M.* An automatic reader of identity documents // 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). 2019. P. 3525–3530.
5. *Xu J. and Wu X.* A system to localize and recognize texts in oriented id card images // 2018 IEEE International Conference on Progress in Informatics and Computing (PIC). 2018. P. 149–153.
6. *Bulatov K., Matalov D. and Arlazarov V.* MIDV-2019: challenges of the modern mobile-based document OCR // Twelfth International Conference on Machine Vision (ICMV 2019). 2020. Vol. 11433. P. 717–722.

7. *Li X., Zhang B., Liao J. and Sander P.V.* Document rectification and illumination correction using a patch-based CNN // *ACM Trans. Graph.* 2019. Vol. 38. No. 6.
8. *Asad F., Ul-Hasan A., Shafait F. and Dengel A.* High performance OCR for camera-captured blurred documents with LSTM networks // 2016 12th IAPR Workshop on Document Analysis Systems (DAS). 2016. P. 7–12.
9. *Chernov T.S., Razumnuy N.P., Kozharinov A.S., Nikolaev D.P. and Arlazarov V.V.* Image quality assessment for video stream recognition systems // *ICMV 2017.* 2018. Vol. 10696. No. 106961U. P. 1–8.
10. *Nunnagoppula G., Deepak K.S., Harikrishna G., Rai N., Krishna P.R. and Vesdapunt N.* Automatic blur detection in mobile captured document images: Towards quality check in mobile based document imaging applications // 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013). 2013. P. 299–304.
11. *Miao L. and Peng S.* Perspective rectification of document images based on morphology // 2006 International Conference on Computational Intelligence and Security. 2006. Vol. 2. P. 1805–1808.
12. *Takezawa Y., Hasegawa M. and Tabbone S.* Robust perspective rectification of camera-captured document images // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017. Vol 06. P. 27–32.
13. *Kunina I., Gladilin S. and Nikolaev D.* Blind radial distortion compensation in a single image using fast Hough transform // *Computer Optics.* 2016. Vol. 40. No. 3. P. 395–403.
14. *Chazalon J., Gomez-Krämer P., Burie J., Coustaty M., Eskenazi S., Luqman M., Nayef N., Rusiñol M., Sidère N. and Ogier J.* SmartDoc 2017 Video Capture: Mobile Document Acquisition in Video Mode // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017. Vol. 04. P. 11–16.
15. *Skoryukina N., Arlazarov V. and Nikolaev D.* Fast method of id documents location and type identification for mobile and server application // 2019 International Conference on Document Analysis and Recognition (ICDAR). 2019. P. 850–857.
16. *Zhukovsky A., Nikolaev D., Arlazarov V., Postnikov V., Polevoy D., Skoryukina N., Chernov T., Shemiakina J., Mukovozov A., Konovalenko I. and Povolotsky M.* Segments graph-based approach for document capture in a smartphone video stream // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017. Vol. 01. P. 337–342.
17. *Haris M., Shakhnarovich G. and Ukita N.* Recurrent back-projection network for video super-resolution // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 3892–3901.
18. *Petrova O., Bulatov K., Arlazarov V.V. and Arlazarov V.L.* Weighted combination of per-frame recognition results for text recognition in a video stream // *Computer Optics.* 2021. Vol. 45. No. 1. P. 77–89.

Arlazarov V.V. Head of the Department for the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, Russia. PhD. Number of articles: 130. Research interests: artificial intelligence, machine learning, recognition systems, information technology, queueing theory. E-mail: vva777@gmail.com