

Методы межкадровой интеграции результатов обнаружения документов в видеопотоке мобильного устройства *

А.Е. Жуковский^{1,II}

^I Московский физико-технический институт (Государственный Университет),
Московская область, Россия

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

Аннотация. Работа посвящена задаче обнаружения положения документа в видеопотоке, полученном с мобильного устройства. Особое внимание уделено методам интеграции положений документа, полученных на последовательности кадров. Описывается основанный на фильтре Калмана алгоритм для выбора положений документа по набору предоставленных альтернатив, их интеграции и уточнения в видеопотоке. Приводится анализ работы алгоритма на данных, предоставленных в рамках конкурса ICDAR'15 по обнаружению документов со смартфона.

Ключевые слова: обнаружение документа, видеопоток, интеграция, проективное преобразование, мобильные камеры, фильтр Калмана.

DOI: 10.14357/20790279180502

Введение

В современном мире наблюдается тенденция к преимущественному использованию цифровых версий документов. Они обладают рядом существенных преимуществ перед бумажными: легко поддаются редактированию (в том числе совместному, с удаленным доступом), поддерживают эффективный поиск, проверку отсутствия искажения информации (с помощью цифровой подписи), сохранения истории изменений и резервное копирование. Бумажные документы более традиционны, имеют свои плюсы и часто необходимы. Для получения цифровых копий бумажных документов и их преимуществ встает задача оцифровки [1–3].

Оцифровка зачастую производится при помощи стационарных сканеров. Они обеспечивают фиксированное положение документа и необходимые условия освещения, что закономерно приводит к получению качественных цифровых копий. Альтернативой является получение изображения документа с мобильной камеры, встроенной в смартфон [4]. Повсеместное распространение смартфонов, высокая скорость и легкость получения изображения делают их наиболее привлекательными устройствами для получения цифровой

копии документа. Однако зафиксировать освещение и положение документа в этом случае не представляется возможным, что неизбежно приводит к ухудшению качества изображений.

Для обеспечения возможности использования мобильных камер в качестве устройств захвата документа необходимы алгоритмы для извлечения документа из изображения. В 2015 году в рамках конференции ICDAR впервые были опубликованы данные и организовано соревнование [5], направленное на разработку и оценку качества решений по захвату изображений документов, полученных с мобильных камер. Данные представлены в виде коротких видеоклипов, что учитывает дополнительные возможности мобильных камер.

В настоящей статье рассматривается алгоритм обнаружения документа в видеопотоке. Задача обнаружения в видеопотоке сводится к обнаружению набора возможных положений документа на каждом кадре (при помощи алгоритма, описанного в [6]) и их дальнейшая интеграция в видеопотоке. Интеграцию предлагается осуществлять в жадной итеративной манере, используя для этого фильтр Калмана. Помимо альтернатив, полученных от алгоритма обнаружения документа, также возможно использование дополнительных источников, например, из оценки проективного преобразования между соседними кадрами.

* Выполнено при частичной поддержке РФФИ (гранты 17-29-03170 и 17-29-03236).

Статья построена следующим образом. В части 1 ставится задача обнаружения документа в видеопотоке и описываются сложности, возникающие в реальных условиях работы. В части 2 вкратце описывается алгоритм обнаружения положения документа на отдельном изображении. В части 3 представлен алгоритм интеграции положений документа в видеопотоке. В части 4 рассказывается о возможности получения дополнительных альтернатив из оценки проективного преобразования между кадрами. В части 5 приведены устройство эксперимента, описание используемого набора данных, методология оценки качества и результаты работы алгоритма.

1. Задача обнаружения документа в видеопотоке

Задачу обнаружения документа в видеопотоке можно поставить следующим образом.

Рассматривается трехмерное пространство, в котором расположены камера, документ и источники света. Система координат пространства связана с камерой. Документ представляет собой плоский тонкий объект прямоугольной формы. Излучение источников света неконтролируемо. Положение документа и источников освещения в заданной системе координат может изменяться с течением времени. Оптическая ось камеры направлена на документ, но не обязательно проходит через него (так что на формируемом камерой изображении обязательно находится некоторая часть проекции документа). Камера фиксирует образ документа с некоторой постоянной разрешающей способностью и некоторой не обязательно постоянной частотой, формируя последовательность изображений документа. На каждом изображении из образованной последовательности требуется определить положение проекции документа (например, координаты четырех углов проекции документа). Проективная модель построения изображения изображена на рис. 1.

Помимо данной общей постановки задачи, следует учесть дополнительно сложности, возникающие при работе с документами в реальных условиях:

- Изначально прямоугольный документ на изображении может принять форму произвольного выпуклого четырехугольника, ввиду проективной природы камеры.
- Изображение подвержено искажениям и цифровому шуму, возникающим из-за несовершенства камеры, загрязнений матрицы, объектива, дефектных элементов сенсора и алгоритмов сжатия изображения.

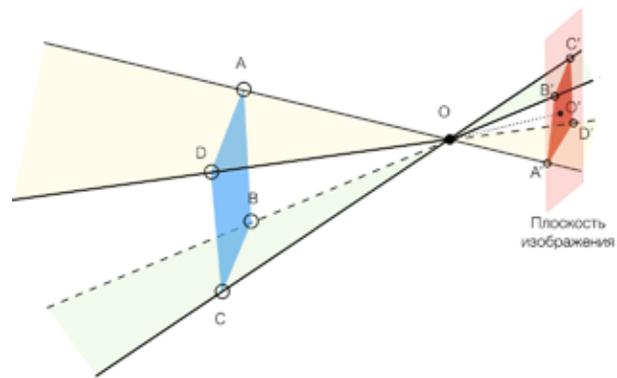


Рис. 1. Проективная модель построения изображения

- В зависимости от характеристик объектива и положения документа относительно плоскости наводки на резкость часть или все изображение может быть размыто.
- Из-за движения и несовершенства камеры могут присутствовать такие явления, как смазывание.
- Ввиду неконтролируемости освещения и относительного движения документа и камеры, освещенность документа может меняться. В захватываемой камерой сцене, в том числе и на документе, могут присутствовать блики и области тени произвольной формы.
- Как объект физического мира, исходный документ может быть подвержен механическим деформациям, так что его края могут изгибаться и не соответствовать модели прямоугольника. Документ может также изначально иметь скругленные, загнутые или обрезанные углы и дефекты прямых краев.
- Сцена, в которой происходит съемка может быть похожей по цвету или текстуре на заполнение документа и неотличимой от документа на изображении.

Описанные моменты существенно усложняют задачу обнаружения положения документа в общем случае, а в некоторых случаях делают ее неразрешимой.

2. Обнаружение документа на кадре

Метод, используемый для обнаружения положения документа на кадре, основывается на обнаружении границ и построении графа пересечений, описанном в [6,7]. Вначале производится первичная обработка изображения, состоящая из масштабирования и морфологического закрытия (для удаления мелких элементов, таких как символы). Далее, на изображении выделяются края, и на

них производится поиск отрезков двумя разными алгоритмами – для увеличения общей надежности метода. После этого по полученным отрезкам строится четырехдольный ориентированный граф пересечений возможных сторон, по которому вычисляются возможные положения документа. Каждому возможному положению назначается свой вес, относительно которого они упорядочиваются. Для того, чтобы получить финальное решение о положении документа, проверяется его соответствие проективной модели прямоугольного документа, и выбирается N альтернатив, удовлетворяющих этой модели и имеющих максимальный вес. В конце работы алгоритма границы документа уточняются на изображении оригинального масштаба.

Стоит отметить, что алгоритм интеграции не зависит от подхода к обнаружению документа и может использоваться с другими подходами.

3. Интеграция результатов обнаружения положения документа

Интеграция результатов в видеопотоке может производиться двумя способами. Первый исходит из предположения, что обнаруженные положения документа на каждом кадре могут быть представлены уже после обработки всего видеоклипа. Это позволяет вначале обнаружить положение документа на каждом кадре, а после использовать метод динамического программирования [8] для уточнения положений по всему набору кадров из видео. Однако, несмотря на то, что такой подход вероятнее приведет к более качественному решению, при реальном применении он повлечет за собой большие накладные расходы в виде времени ожидания ответа и неопределенности времени конца съемки. Альтернативный вариант заключается в использовании жадного подхода. В нем предполагается, что алгоритму не известны последующие кадры, а решение о положении документа на рассматриваемом кадре необходимо принимать, опираясь только на предыдущие положения. При этом также невозможно изменять положения, зафиксированные на предыдущих кадрах. В таком случае информации, доступной для интеграции, намного меньше, чем у первого метода, но алгоритм может использоваться в режиме реального времени, например, для показа текущего положения пользователю.

Предлагаемый алгоритм интеграции положений документа в реальном времени основывается на использовании фильтра Калмана [9]. После получения каждого последующего кадра на нем проводится поиск нескольких альтернатив положения

документа. Также к набору альтернатив добавляются положения, полученные исходя из оценки проективного преобразования между соседними кадрами и из обнаружения углов на изображении.

3.1. Использование фильтра Калмана для межкадровой интеграции

Фильтр Калмана представляет собой итеративный алгоритм фильтрации сигнала. Каждая итерация фильтра Калмана состоит из двух этапов: экстраполяции и коррекции. Во время экстраполяции фильтр строит априорную оценку состояния системы \hat{x}_k^- на текущий k -ый шаг по итоговой оценке с предыдущего шага. На этапе коррекции по априорной оценке \hat{x}_k^- и текущим измерениям z_k строится апостериорная оценка \hat{x}_k , которая и используется в качестве ответа системы на данном шаге.

Используемый фильтр Калмана опирается на положение, скорость и ускорение движения каждого из четырех углов документа в плоскости изображения, всего 24 измерения в векторе x_k состояния системы. Мы используем модель равноускоренного движения для оценки эволюции системы между каждыми двумя кадрами:

$$s_k = s_{k-1} + v_{k-1}t + \frac{a_{k-1}t^2}{2}. \quad (1)$$

В этом случае модель процесса будет выглядеть следующим образом:

$$\hat{x}_k^- = F\hat{x}_{k-1} + w_k, \quad (2)$$

где $w_k \sim N(0, Q_k)$ – нормальный случайный процесс, который описывает случайный характер эволюции процесса, а F – матрица эволюции процесса, имеющая вид:

$$F = \begin{bmatrix} I & I \times \Delta t & I \times \Delta t^2 / 2 \\ 0 & I & I \times \Delta t \\ 0 & 0 & I \end{bmatrix}, \quad (3)$$

где I – единичная матрица 8×8 , 0 – нулевая матрица 8×8 .

На этапе предсказания строятся априорная оценка вектора состояния \hat{x}_k^- и соответствующая ему оценка ковариационной матрицы ошибок P_k^- :

$$\hat{x}_k^- = F\hat{x}_{k-1}; \quad (4)$$

$$P_k^- = FP_{k-1}F^T + Q_k. \quad (5)$$

Они также используются для выбора ближайшего к предсказанию текущего наблюдения z_k из списка альтернатив z_k^i , поступивших от алгоритма обнаружения документа на изображении. Вы-

бираемое наблюдение должно одновременно быть близким к предсказанию и иметь хорошую оценку от алгоритма обнаружения, для этого оно выбирается исходя из максимизации функционала:

$$z_k = \arg \max_{z_k^i} (s_k^i \times \exp(-\frac{d_k^{iT} S^{-1} d_k^i}{2})), \quad (6)$$

где: $d_k^i = z_k^i - \hat{x}_k^-$ – вектор отклонений, $S = HP_k^- H^T$ – ковариационная матрица вектора отклонений, $H = \begin{bmatrix} I & 0 & 0 \end{bmatrix}$ – матрица наблюдений (связывающая истинный вектор состояния и вектор произведенных измерений – координат обнаруженного четырехугольника).

На этапе коррекции априорная оценка вектора состояния дополняется координатами отобранного четырехугольника z_k . Для этого строится матрица коэффициентов усиления K_k , которая минимизирует сумму квадратов математических ожиданий ошибки оценки вектора состояния. Для достижения этого она должна иметь вид:

$$K_k = P_k^- H^T (S + R)^{-1}, \quad (7)$$

где R – ковариационная матрица шума измерений.

Непосредственно скорректированные (апостериорные) значения вектора состояния \hat{x}_k и ковариационная матрица ошибок P_k тогда будут иметь вид:

$$\hat{x}_k = (I - K_k H) \hat{x}_k^- + K_k z_k, \quad (8)$$

$$P_k = (I - K_k H) P_k^-. \quad (9)$$

Полученные из апостериорной оценки вектора состояния \hat{x}_k координаты четырехугольника и будут использоваться как финальный ответ алгоритма интеграции на данном кадре.

4. Оценка проективного преобразования между кадрами

Так как документ представляет собой плоский объект, становится возможным оценить проективное преобразование плоскости [10–12] между двумя кадрами.

Для этого на двух кадрах производится поиск особых точек, для каждой из них вычисляется дескриптор (вектор, описывающий локальную окрестность точки). По ним, исходя из близости дескрипторов, строится начальный набор сопоставлений точек между двумя изображениями. Далее робастно оценивается проективное преобразование и оценивается количество точек ему соответствующих.

В нашей реализации используется алгоритм SURF [13] для выделения 256 точек и дескриптор BRIEF [14] для их описания. Начальный набор со-

поставлений строится методом прямого перебора, а для оценки проективного преобразования и количества сопоставлений, соответствующих этому преобразованию, используется широко распространенный алгоритм RANSAC [15].

Полученная оценка преобразования, связывающего прошлый кадр с текущим, используется для получения проекции найденного на прошлом кадре четырехугольника документа на рассматриваемом кадре. Если качество преобразования, вычисленное по количеству соответствующих ему сопоставлений, достаточно велико, спроецированный четырехугольник добавляется к набору альтернатив, поступающих на вход фильтра Калмана.

5. Эксперимент

5.1. Исходные данные

Для оценки качества работы использовался набор данных, представленный впервые на соревновании SmartDoc, проводимом в рамках конференции ICDAR в 2015 году [5].

Для создания данных использовались шесть типов документов, по пять документов каждого типа. Документы выбирались так, чтобы они покрывали различную структуру и содержание: в них присутствуют как полностью текстовые документы, так и содержащие таблицы или графическое наполнение. Все документы соответствуют формату бумаги А4.

Для каждого из 30 документов при помощи планшета Google Nexus 7 были сняты небольшие видеоклипы (около 10 секунд) в пяти разных фоновых сценариях. Видеоклипы записывались с Full HD разрешением 1920×1080 при переменной частоте кадров.

При съемке планшет двигался относительно неподвижного документа, моделируя реальные условия съемки, поэтому в видеоклипах присутствуют реалистичные искажения, такие как размытие фокуса и движения, перспективные искажения, изменение освещения и даже частичные перекрытия страниц документа. В итоге данные содержат 150 видеоклипов и около 25 тысяч кадров. Примеры кадров из видеоклипов приведены на рис. 2. Для каждого кадра были известны координаты четырех углов проекции документа на изображении.

Отдельно стоит упомянуть пятый фон, существенно отличающийся от остальных. Правый нижний кадр из видеоклипа, представленный на рис. 2, иллюстрирует то множество сложных для алгоритма обнаружения составляющих сцены: наложение прочих элементов сцены (шнур, маркеры) на изображение документа, листы бумаги, лежа-



Рис. 2. Примеры кадров

щие под целевым документом, наполненность сцены прочими объектами. Соответственно, качество работы алгоритмов на этой части данных существенно отличается в худшую сторону.

5.2. Оценка качества работы

Для оценки качества работы итогового алгоритма используется коэффициент Джаккарда [16], который характеризует способность алгоритма к обнаружению точных контуров документа, а также штрафует за пропуск документа на некоторых кадрах.

На каждом кадре вычисляется проективное преобразование координат H , переводящее истинный четырехугольник документа G в прямоугольник известного размера G' . Это же преобразование используется для получения по обнаруженному четырехугольнику S проективно исправленного четырехугольника $S' = H(S)$. Данное преобразование позволяет получить сравнимые оценки на разных кадрах. Для каждого кадра f вычисляется коэффициент Джаккарда, который измеряет сходство проективно исправленных четырехугольников следующим образом:

$$J(f) = \frac{area(G' \cap S')}{area(G' \cup S')}, \tag{10}$$

где $G' \cap S'$ обозначает многоугольник, полученный как пересечение проективно исправленных обнаруженного и идеального четырехугольников, а $G' \cup S'$ – их объединение. Общая оценка работы алгоритма вычисляется как среднее значение для всех кадров во всех видео клипах.

5.3. Результаты

Результаты оценки качества работы алгоритма обнаружения документа в видеопотоке сведены в табл. 1. Алгоритм интеграции позволяет достичь значения коэффициента Джаккарда, равного 0.9548. Разделяя результаты по фону, на котором был расположен документ, можно увидеть, что итоговый алгоритм показывает результат в 0.985 на первых четырех фонах. При этом, если считать качество только по тем кадрам, на которых все углы документа отстоят от соответствующих углов идеальной разметки не более чем на 20 пикселей значение индекса Джаккарда будет со-

Табл. 1

Оценка качества работы алгоритма обнаружения документа в видеопотоке

Место	Метод	Индекс Джаккарда	Доверительный интервал	Фон 1	Фон 2	Фон 3	Фон 4	Фон 5
1	LRDE	0.9716	[0.9710, 0.9721]	0.9869	0.9775	0.9889	0.9837	0.8613
2	ISPL-CVML	0.9658	[0.9649, 0.9667]	0.9870	0.9652	0.9846	0.9766	0.8555
3	SmartEngines	0.9548	[0.9533, 0.9562]	0.9885	0.9833	0.9897	0.9785	0.6884
4	NetEase	0.8820	[0.8790, 0.8850]	0.9624	0.9552	0.9621	0.9511	0.2218
5	A2iA-2	0.8090	[0.8049, 0.8132]	0.9597	0.8063	0.9118	0.8264	0.1892
6	A2iA-1	0.7788	[0.7745, 0.7831]	0.9724	0.8006	0.9117	0.6352	0.1890
7	RPPDI-UPE	0.7408	[0.7359, 0.7456]	0.8274	0.9104	0.9697	0.3649	0.2163
7	SEECs-NUST	0.7393	[0.7353, 0.7432]	0.8875	0.8264	0.7832	0.7811	0.0113

ставлять 0.987. Это позволяет судить о точности и ограничениях использования данной метрики в рамках данного стенда. Наиболее точные результаты на первых фонах позволяют сделать вывод о полезности интеграции для максимально точного наведения.

Заключение

В работе рассмотрена задача обнаружения документа в видеопотоке, полученном с камеры мобильного устройства. Особое внимание уделено задаче интеграции результатов обнаружения в видеопотоке. Предложено использование фильтра Калмана для выбора положения документа из набора альтернатив и интеграции полученных результатов. Проведено исследование предложенного алгоритма на наборе видеоклипов, использованных в конкурсе ICDAR'15 по обнаружению документов со смартфона. Итоговое значение коэффициента Джаккарда, используемого в качестве финальной метрики составило 0.955 на стенде целиком и 0.985 – на первых четырех фонах (90% кадров). За исключением последнего фона, использование межкадровой интеграции позволило достичь наиболее точных результатов среди всех участников соревнования.

Литература

1. Арлазаров В.В., Жуковский А.Е., Кривцов В.Е., Николаев Д.П., Полевой Д.В. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов. // Информационные технологии и вычислительные системы. Том 3. 2014. С. 71-81.
2. Liang J., Doermann D., Li H. Camera-based analysis of text and documents: a survey, Int. J. of Document Analysis and Recognition, vol. 7, Issue 2, 2005, pp. 84-104
3. Doermann D., Liang J., Li H. Progress in Camera-Based Document Image Analysis, IEEE Proc. 7th Int. Conf. on Document Analysis and Recognition, Vol.1, 2003, pp. 606-616
4. Bulatov K., Arlazarov V.V., Chernov T., Slavin O. and Nikolaev D. Smart IDReader: Document Recognition in Video Stream, 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2017, pp. 39-44. doi: 10.1109/ICDAR.2017.347
5. Burie J.C., Chazalon J., Coustaty M., Eskenazi S., Luqman M.M., Mehri M., Nayef N., Ogier J.M., Prum S., Rusinol M. ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc), 13th Int. Conf. on Document Analysis and Recognition. 2015.
6. Арлазаров В.В., Жуковский А.Е., Кривцов В.Е., Постников В.В. Использование графа пересечений в задаче обнаружения документа на изображении, полученном со смартфона // Искусственный интеллект и принятие решений. Том 2. С. 60-69, 2016
7. Zhukovsky A. et al. "Segments Graph-Based Approach for Document Capture in a Smartphone Video Stream," 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2017, pp. 337-342, doi: 10.1109/ICDAR.2017.63
8. Skoryukina N, Shemyakina Y, Arlazarov V. L., Faradjev I. Document localization algorithms based on feature points and straight lines, Proc. SPIE 10696, 10th Int. Conf. on Machine Vision (ICMV 2017), pp. 1-8, 2018, DOI: 10.1117/12.2311478
9. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. 2-е изд., М.: Вильямс, 2005.
10. Kalman R.E. A New Approach to Linear Filtering and Prediction Problems, J. of Basic Engineering 82, 35, 1960.
11. Hartley R., Zisserman A. Multiple view geometry in computer vision, Cambridge University Press, New York, 2003
12. Шемякина Ю.А., Жуковский А.Е., Фараджев И.А. Исследование алгоритмов вычисления проективного преобразования в задаче наведения на планарный объект по особым точкам // Искусственный интеллект и принятие решений. Том 1. 2017. С. 43-49
13. Shemyakina Y, Zhukovsky A., Faradjev I. The Calculation of a Projective Transformation in the Problem of Planar Object Targeting by Feature Points, Proc. SPIE 10341, ICMV 2016, 10341 ed., 9th Int. Conf. on Machine Vision, 2017, vol. 10341, pp. 1-6, 2017, DOI: 10.1117/12.2268590
14. Bay H., Tuytelaars T., Gool L.V. Surf: Speeded up robust features, European Conf. on Computer Vision (ECCV), 2006, pp. 404-417
15. Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary Robust Independent Elementary Features, 11th European Conf. on Computer Vision (ECCV), 2010
16. Fischler M.A., Bolles R.C.. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, 24(6), 1981, pp. 381-395
17. Everingham M., Gool L.V., Williams C., Winn J. and Zisserman A. "The PASCAL visual object classes (VOC) challenge" IJCV, vol. 88, no. 2, 2010, pp. 303–338

Жуковский Александр Евгеньевич. Московский физико-технический институт (Государственный Университет), г. Москва, Россия. Преподаватель кафедры алгоритмов и технологий программирования МФТИ. Smart Engines Ltd., проспект 60-летия Октября, 9, Москва, Россия, 117312. Научный сотрудник. Количество печатных работ: 15. Область научных интересов: компьютерное зрение, распознавание образов, машинное обучение. E-mail: zhukovsky@phystech.edu

Methods for interframe integration of document detection results in a video stream of a mobile device

A.E. Zhukovsky^{I,II}

^I Moscow Institute of Physics and Technology (State University), Moscow region, Russia

^{II} LLC “Smart Engines Service”, Moscow, Russia.

Abstract. The paper is devoted to the task of detecting the position of a document in a video stream received from a mobile device. Particular attention is paid to the methods of integrating the positions of the document obtained on a sequence of frames. The paper describes an algorithm based on the Kalman filter for selecting the document positions for a set of provided alternatives, their integration and refinement in the video stream. The analysis of the performance of the algorithm on the dataset provided in the of the ICDAR’15 competition on detection of documents from the smartphone is given.

Keywords: *document detection, video stream, integration, projective transformation, mobile cameras, Kalman filter.*

DOI: 10.14357/20790279180502

References

1. V.V. Arlazarov, A.E. Zhukovsky, V.E. Krivtsov, D.P. Nikolaev, D.V. Polevoy. Analiz osobennostey ispol'zovaniya statsionarnykh i mobil'nykh malorazmernykh tsifrovyykh video kamer dlya raspoznavaniya dokumentov, [Analysis of specific character of usage fixed and mobile small-size video cameras for document recognition], Informatsionnyye tekhnologii i vychislitel'nyye sistemy [Information Technologies and computing systems], Vol. 3, 2014, pp. 71-81
2. J. Liang, D. Doermann, H.Li. Camera-based analysis of text and documents: a survey, Int. J. of Document Analysis and Recognition, vol. 7, Issue 2, 2005, pp. 84-104
3. D. Doermann, J. Liang, H. Li. Progress in Camera-Based Document Image Analysis, IEEE Proc. 7th Int. Conf. on Document Analysis and Recognition, Vol.1, 2003, pp. 606-616
4. K. Bulatov, V.V. Arlazarov, T. Chernov, O. Slavin and D. Nikolaev. Smart IDReader: Document Recognition in Video Stream, 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2017, pp. 39-44. doi: 10.1109/ICDAR.2017.347
5. Burie J.C., Chazalon J., Coustaty M., Eskenazi S., Luqman M.M., Mehri M., Nayef N., Ogier J.M., Prum S., Rusinol M. ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc), 13th Int. Conf. on Document Analysis and Recognition. 2015
6. V.V. Arlazarov, A.E. Zhukovsky, V.E. Krivtsov, V.V. Postnikov. Ispol'zovaniye grafa peresecheniy v zadache obnaruzheniya dokumenta na izobrazhenii, poluchennom so smartfona [Usage of the intersection graph in the task of camera-based document detection.] Ispol'zovaniye grafa peresecheniy v zadache obnaruzheniya dokumenta na izobrazhenii, poluchennom so smartfona, Iskustvennyy Intellekt i Prinyatiye Resheniy [Artificial Intelligence and Decision Making], vol. 2, pp. 60-69, 2016
7. A. Zhukovsky et al. “Segments Graph-Based Approach for Document Capture in a Smartphone Video Stream,” 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2017, pp. 337-342, doi: 10.1109/ICDAR.2017.63
8. Skoryukina N, Shemyakina Y, Arlazarov V.L., Faradjev I. Document localization algorithms based on feature points and straight lines, Proc. SPIE 10696, 10th Int. Conf. on Machine Vision (ICMV 2017), pp. 1-8, 2018, DOI: 10.1117/12.2311478
9. T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein. Introduction to Algorithms (second ed.). MIT Press and McGraw-Hill. ISBN 978-0-262-53196-2., 2001
10. R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems, J. of Basic Engineering 82, 35, 1960.
11. R. Hartley, A. Zisserman. Multiple view geometry in computer vision, Cambridge University Press, New York, 2003

12. *Y.A. Shemyakina, A.E. Zhukovsky, I.A. Faradjev.* Issledovaniye algoritmov vychisleniya proyektivnogo preobrazovaniya v zadache navedeniya na planarnyy ob"yekt po osobym tochkam [Investigation of algorithms for calculating a projective transformation in the problem of targeting to a planar object from feature points], *Iskusstvennyy Intellekt i Prinyatiye Resheniy* [Artificial Intelligence and Decision Making], vol. 1, 2017, pp. 43-49
13. *Y. Shemyakina, A. Zhukovsky, I. Faradjev.* The Calculation of a Projective Transformation in the Problem of Planar Object Targeting by Feature Points, *Proc. SPIE 10341, ICMV 2016, 10341 ed.*, 9th Int. Conf. on Machine Vision, 2017, vol. 10341, pp. 1-6, 2017, DOI: 10.1117/12.2268590
14. *H. Bay, T. Tuytelaars, L. V. Gool.* Surf: Speeded up robust features, *European Conf. on Computer Vision (ECCV)*, 2006, pp. 404-417
15. *M. Calonder, V. Lepetit, C. Strecha, P. Fua.* BRIEF: Binary Robust Independent Elementary Features, *11th European Conf. on Computer Vision (ECCV)*, 2010
16. *Fischler M.A., Bolles R.C.* Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24(6), 1981, pp. 381-395
17. *M. Everingham, L.V. Gool, C. Williams, J. Winn, and A. Zisserman.* "The PASCAL visual object classes (VOC) challenge" *IJCV*, vol. 88, no. 2, 2010, pp. 303-338

A.E. Zhukovsky. Moscow Institute of Physics and Technology (State University), Institutsky per., 9, Dolgoprudny, Moscow region, 141701, Russia. LLC "Smart Engines Service", Moscow, Russia, prospect 60-letiya Oktyabrya, 9, Moscow, 117321, Russia. E-mail: zhukovsky@phystech.edu