

# Метод совмещения изображений на основе максимизации резкости\*

Д.В. ТРОПИН<sup>I,IV</sup>, Д.П. НИКОЛАЕВ<sup>II,IV</sup>, Д.Г. СЛУГИН<sup>III,IV</sup>

<sup>I</sup> Московский физико-технический институт, г. Долгопрудный, Россия

<sup>II</sup> Институт проблем передачи информации им. А.А. Харкевича РАН, г. Москва, Россия

<sup>III</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия

<sup>IV</sup> ООО «Смарт Энджинс Сервис», г. Москва, Россия

**Аннотация.** В статье предложен метод быстрого совмещения изображений на основе максимизации резкости усредненного изображения. Рассмотрен алгоритм совмещения для глобально-сдвиговой модели, показана возможность его эффективного вычисления с помощью БПФ. Для проективной модели изложен подход к совмещению изображений путем сопоставления отдельных частей и использования RANSAC для вычисления финального преобразования. Приведены результаты работы алгоритма на примере задачи восстановления изображения документа в видеопотоке, демонстрирующие улучшение качества финального изображения.

**Ключевые слова:** анализ изображений, совмещение изображений, оптический поток, БПФ, RANSAC.

**DOI:** 10.14357/20790279180515

## Введение

Совмещение изображений – важная часть многих систем и алгоритмов обработки изображений, таких как панорамирование, стабилизация съемки видео, привязка и калибровка спутников по земной поверхности, поиск объектов в видеопотоке и многих других. С активным развитием мобильных устройств значительно расширился круг решаемых ими задач, в частности в области поиска и распознавания документов в видеопотоке [1]. Все актуальнее становится разработка быстрых алгоритмов обработки изображений, способных работать на устройствах ограниченной производительности в режиме реального времени, с учетом возможных внешних факторов в неконтролируемых условиях съемки, таких как изменение освещенности, дефокус, шум, блики и другие виды искажений.

В общем случае под задачей совмещения изображений понимается нахождение параметров преобразования, переводящего одно изображение в другое. Существуют разные подходы к решению данной задачи, такие как классический метод оценки оптического потока [2, 3], а также алгоритмы, использующие детекторы и дескрипторы особых точек [4] и их комбинации с нейронными сетя-

ми [5]. Однако они не всегда применимы к задачам обработки изображений в видеопотоке из-за большой вычислительной сложности и/или неустойчивости к неконтролируемым условиям съемки. В работе предлагается метод быстрого совмещения изображений для нахождения соответствия между кадрами видеопотока, содержащих плоский псевдоригодный объект – документ.

## 1. Постановка задачи

Пусть имеется набор изображений, представляющих собой кадры видеопотока. Установление соответствия между кадрами можно свести к задаче совмещения пары соседних кадров (назовем их опорный и локальный) и оценке параметров модели оптического потока, переводящего локальное изображение в опорное в соответствии с критерием совмещения. При совмещении изображений с помощью алгоритма Лукаса-Канаде [2] используется критерий качества, основанный на минимизации невязки между опорным и преобразованным локальным изображением в каждой точке изображения:

$$\operatorname{argmin}_p \|I_0 - \hat{T}_p I_1\|_2, \quad (1)$$

где  $\hat{T}_p$  – преобразование в соответствии с векторным полем  $P$ ,  $I_0$  – опорное изображение,  $I_1$  – локальное

\* Работа выполнена при финансовой поддержке РФФИ (проекты № 17-29-03263 и № 17-29-03236).

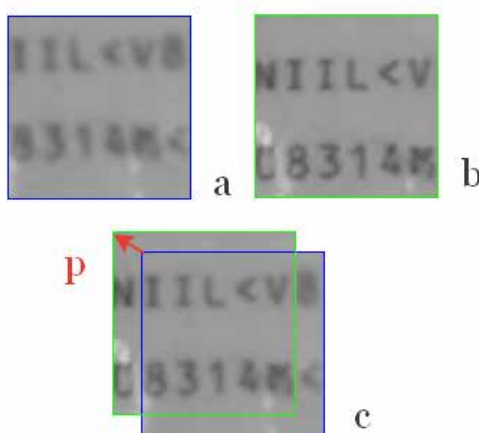
изображение. В общем случае векторное поле строится для всех точек изображений, так что задача совмещения весьма сложна и трудоемка. В работах [6, 7] предложены более эффективные методы совмещения изображений в соответствии с критерием (1) с использованием фазовой корреляции преобразований из расширенной группы движений.

В задаче совмещения кадров видеопотока, содержащих документ, основным объектом на изображении является плоский псевдоригидный объект. Таким образом, можно перейти от общей модели оптического потока к глобально-аналитической, где преобразование двумерной плоскости будет определяться взаимным расположением и движением объекта сцены и камеры. В частности, при съемке документа с камеры мобильного устройства это будет проективное преобразование, а при работе с отсканированными данными его частный случай – сдвиг [8].

При работе с данными, полученными с мобильного устройства в реальных условиях съемки, почти всегда не выполняется требование о яркостной константности, ключевое для критерия качества (1). Чтобы сделать алгоритм устойчивым к локальным изменениям интенсивности освещения предлагается модифицированный критерий качества совмещения, а именно – оценка точности совмещения не исходных изображений, а границ объектов на них. В качестве такого критерия будем использовать условие максимальности функционала резкости усредненного изображения:

$$\operatorname{argmax}_p \|\vec{\nabla}(I_0 + \hat{T}_p I_1)\|_2, \quad (2)$$

где  $\hat{T}_p$  – проективное преобразование с параметрами  $P$ ,  $I_0$  – опорное изображение,  $I_1$  – локальное изображение.



**Рис. 1.** Опорное (а), локальное (б) и совмещенное (с) изображения; d – значения функционал резкости для различных сдвигов p

## 2. Глобально-сдвиговая модель

Рассмотрим глобально-сдвиговую модель оптического потока, при которой каждая точка опорного изображения представляет собой точку локального изображения, сдвинутого на некоторый одинаковый вектор. Например, такая модель возникает при совмещении двух изображений, полученных после нескольких сканирований одного документа.

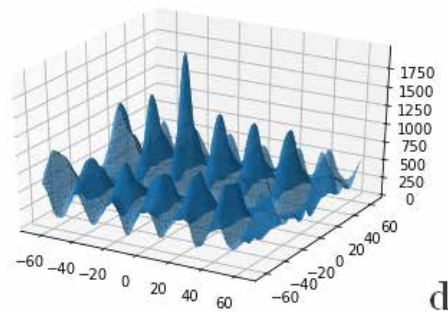
Пусть опорное и локальное изображения имеют размеры  $W \times H$ . Нахождение параметров вектора сдвига сводится к перебору всех его возможных

параметров  $p \in \left[-\frac{W}{2}; \frac{W}{2}\right] \times \left[-\frac{H}{2}; \frac{H}{2}\right]$ , вычисления

функционала резкости и нахождения его максимума (рис. 1). Оценка трудоемкости в данном случае составляет  $O(W^2 H^2)$  операций, что не позволяет эффективно производить совмещения изображений больших размеров.

Если расписать подробнее функционал резкости, то можно заметить, что задача нахождения вектора сдвига  $p$  соответствует поиску точки максимума свертки двух изображений (3):

$$\begin{aligned} & \operatorname{argmax}_p \|\vec{\nabla}(I_0 + \hat{T}_p I_1)\|_2 \\ &= \operatorname{argmax}_p \left[ \sum_v \left( \vec{\nabla} I_0(v) + \hat{T}_p \vec{\nabla} I_1(v) \right)^2 \right] = \\ &= \operatorname{argmax}_p \left[ \sum_v \left( \vec{\nabla} I_0(v) \right)^2 \right. \\ & \quad \left. + 2 \sum_v \vec{\nabla} I_0(v) \cdot \hat{T}_p \vec{\nabla} I_1(v) + \sum_v \left( \hat{T}_p \vec{\nabla} I_1(v) \right)^2 \right] = \\ &= \operatorname{argmax}_p \sum_v \vec{\nabla} I_0(v) \cdot \hat{T}_p \vec{\nabla} I_1(v). \quad (3) \end{aligned}$$



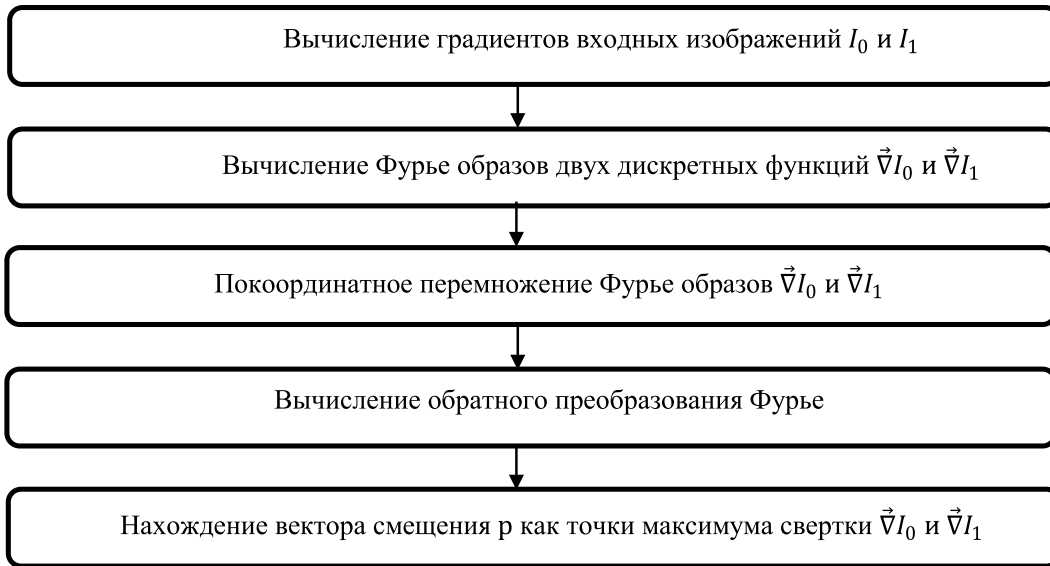


Рис. 2. Нахождение вектора смещения в случае глобально-сдвиговой модели

Свертку функций  $\vec{V}I_0$  и  $\vec{V}I_1$  можно эффективно вычислить с помощью Быстрого Преобразования Фурье (БПФ) по теореме о свертке, где  $\mathfrak{Z}[f]$  обозначает преобразование Фурье функции  $f$ :

$$\mathfrak{Z}[f * g] = \mathfrak{Z}[f] \cdot \mathfrak{Z}[g]. \quad (4)$$

В результате схема алгоритма принимает следующий вид (рис. 2).

Сложность алгоритма имеет асимптотику БПФ дискретных функций и составляет

$O(WN * \log_2(WN))$ , что значительно лучше первоначального переборного варианта.

### 3. Глобально-проективная модель

В случае совмещения кадров видеопотока, содержащего плоские псевдоригидные объекты, переход от одного кадра к другому осуществляется с помощью проективного преобразования. Для соседних кадров с небольшой разницей по времени проективное преобразо-



Рис. 3. опорное (а), локальное (b) и совмещенное (d) изображения; с – векторное поле сдвигов соответствующих фрагментов

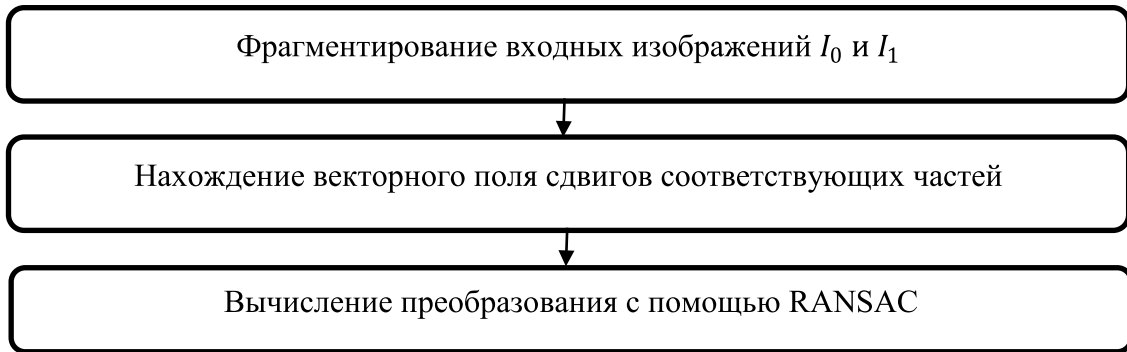


Рис. 4. Схема глобально-проективной модели

вание почти в каждой небольшой окрестности точек документа мало отличается от сдвига, вследствие чего оно может быть представлено кусочно-сдвиговой аппроксимацией. Производя разбиение опорного и локального изображения на небольшие части (например, прямоугольники), для каждой пары частей, используя алгоритм предыдущей главы, находим вектор сдвига, в результате получаем общее векторное поле сдвигов (рис. 3). Для нахождения проективного преобразования и оценки его параметров используем метод RANSAC [9]. На вход методу подается набор центров прямоугольников и соответствующие вектора сдвигов. Выходом являются оценка и параметры наиболее оптимального проективного преобразования. Схема алгоритм представлена на рис. 4.

Используя найденные проективные преобразования для соседних кадров можно определить преобразование между любыми двумя кадрами, используя их композицию. Стоит отметить, что алгоритм также устойчив к аффинным преобразованиям, которые являются частным случаем проективных, а именно к поворотам и изменением масштаба в некоторых границах (рис. 5).



Рис. 5. Пример векторного поля в случае уменьшения масштаба

Также был рассмотрен вопрос о применимости алгоритма в случаях значительных проективных преобразований документа (например, при съемке с низким FPS, когда разница во времени между соседними кадрами является достаточно большой). Предложена идея о возможности использования итерационной схемы нахождения проективного преобразования, когда на каждом шаге вычисляется совмещение опорного изображения с преобразованным локальным от предыдущего шага. Критериями остановки процесса могут являться прекращение повышения резкости усредненного изображения, уменьшение модуля векторного поля до определенного порога, близость матрицы проективного преобразования к тождественной и т.д. На рис. 6 показан пример сходимости такого итерационного процесса для трех итераций.

#### 4. Результаты

Разработанный алгоритм был применен в рамках решения задачи восстановления документа в видеопотоке на конкурсе "Smartdoc2017" международной конференции ICDAR-2017 [10]. Задача конкурса состояла в следующем: заданы идеальное прямоугольное изображение документа  $I$  и видеоролик  $V$ , снятый на камеру мобильного устройства. На кадрах  $\{F_1 \dots F_N\}$  ролика  $V$  присутствует документ  $I$ , причем документ в кадре проективно искажен, может присутствовать частично, иметь засветы, неравномерное освещение, быть частично закрыт другими предметами, присутствовать дефокус всего документа или его части, а также содержать другие искажения. Требуется, используя особенности идеального изображения, найти на кадрах ролика документ и собрать наиболее качественное выходное изображение. Подход к поиску и восстановлению изображения предлагался следующий:

1. Разбираем ролик  $V$  на кадры  $\{F_1 \dots F_N\}$ .
2. На идеальном изображении  $I$  выделяем особые точки  $\{P(I)\}$ .



**Рис. 6.** Опорное (a) и локальное (b) изображения; векторные поля первой (c), второй (d) и третьей (e) итерации; f – совмещение опорного и финального изображений

3. На каждом следующем кадре  $F_i$  выделяем особые точки  $\{P(F_i)\}$ .
4. Сопоставляя особые точки на кадре  $\{P(F_i)\}$  с точками идеального изображения  $\{P(I)\}$ , находим положение документа в кадре и проективное преобразование  $T$ , приводящее его в прямоугольную форму  $T(F_i)$ , с помощью RANSAC [11]. Если документ не найден или найден с низкой оценкой уверенности, тогда переходим на шаг 3.
5. Если это первый найденный документ на кадрах ролика, используем найденное изображение документа  $T(F_i)$  в качестве основы  $V$  для дальнейшей процедуры восстановления изображения и продолжаем с шага 3.
6. Если уже есть текущее восстановленное изображение  $V$ , то производим разбиение  $V$  и

$T(F_i)$  на прямоугольные части  $\{B_1 \dots B_M\}$  и  $\{T(F_i)_1 \dots T(F_i)_M\}$ . Далее для каждой части производится оценка четкости путем подсчета суммы модулей градиентов. Если на какой-либо части  $T(F_i)_k$  четкость выше, чем на  $V_k$ , тогда  $T(F_i)_k$  вставляем на место  $V_k$  изображения  $V$ . Далее продолжаем с шага 3.

Алгоритм отработал на исходных данных задачи, но качество выходного изображения оказалось не очень приемлемым. Проблема связана с тем, что поиск документа в кадре не обладает необходимой точностью наведения границ документа, из-за чего при склейке выходного изображения наблюдались искажения частей изображения  $V$  (рис. 7).

В результате было принято решение использовать предложенный алгоритм совмещения изображения для коррекции изображения  $T(F_i)$



Рис. 7. Примеры восстановленных изображений

так, чтобы найти преобразование  $Q$ , наиболее точно совмещающее  $T(F_i)$  с  $V$ . Тогда в итоговом алгоритме появляется дополнительный шаг: находим преобразование  $Q$  такое, что  $Q(T(F_i))$  наилучшим образом совмещается с  $V$ . Далее вместо  $T(F_i)$  используем  $Q(T(F_i))$  – скорректированное изображение.

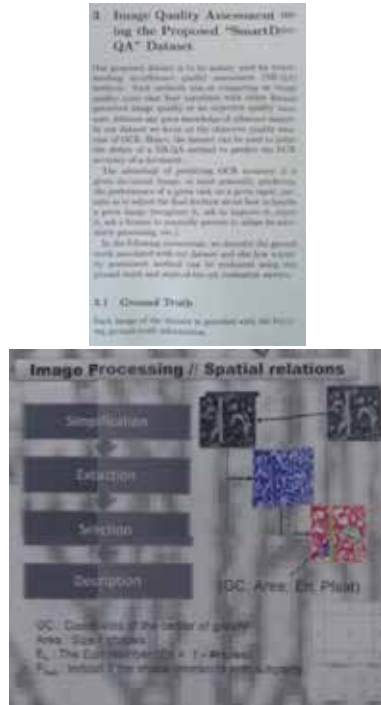


Рис. 9. Результаты работы алгоритма после добавления совмещения

Окончательная схема алгоритма показана на рис. 8.

В итоге качество выходных изображений алгоритма значительно улучшилось (рис. 9). Данный результат показывает корректность и применимость предложенного алгоритма для решения задач совмещения изображений, в частности документов.

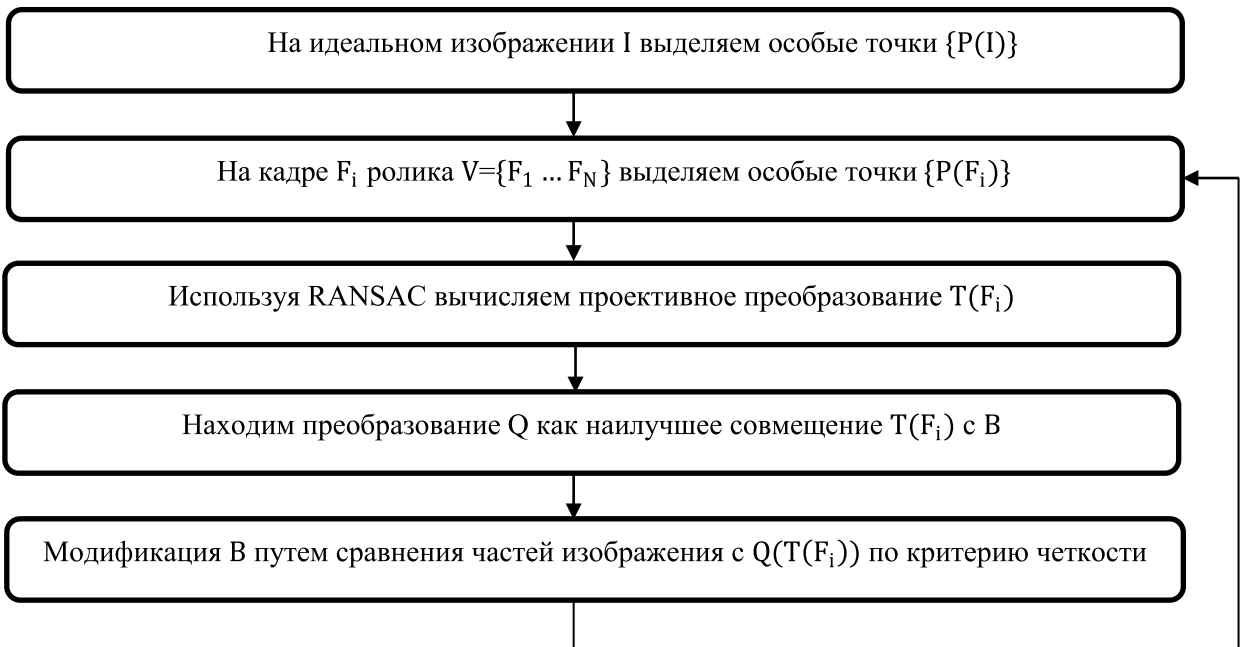


Рис. 8. Схема алгоритма восстановления изображения документа в видеопотоке

### Заключение

В работе предложен метод совмещения изображений на основе максимизации резкости усредненного изображения. Обоснована корректность такого подхода для глобально-сдвиговой модели, разработан алгоритм нахождения совмещения в случае проективных преобразований. Рассмотрена итерационная модель в случае невозможности нахождения совмещения за один шаг. Работа алгоритма проверена на реальной задаче восстановления документа в видеопотоке. Дальнейшие исследования будут включать в себя вопросы возможности применимости алгоритма и его устойчивости в зависимости от фрагментирования исходных изображений, площади документа в кадре, границ проективного искажения документа и сходимости итерационного подхода.

### Литература

1. Bulatov K., Arlazarov V.V., Chernov T., Slavin O. and Nikolaev D. "Smart IDReader: Document Recognition in Video Stream," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 39-44. doi: 10.1109/ICDAR.2017.347
2. Lucas B.D., Kanade T. An iterative image registration technique with an application in stereo vision // Seventh International Joint Conference on Artificial Intelligence (IJCAI-81). – Vancouver, 1981. – P. 674-679.
3. Horn B.K.P., Schunck B.G. Determining optical flow. Artificial Intelligence // Artificial Intelligence. – Cambridge, 1981. – P. 185–203.
4. Lowe D.G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. – 2004. – № 56.
5. Nanne van Noord, Eric Postma. Learning scale-variant and scale-invariant features for deep image classification. Pattern Recognition. Volume 61, January 2017, Pages 583-592
6. Kuglin C.D. and Hines D.C. The phase correlation image alignment method. International Conference of Cybernetic Society. Proc IEEE 1975. New York. PP 163-165.
7. Szeliski Richard. Image Alignment and Stitching: A Tutorial. Foundations and Trends in Computer Graphics and Vision Vol. 2, No 1 (2006) 1–104
8. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов / Арлазаров В.В., Николаев Д.П., Полевой Д.В и др. // Информационные технологии и вычислительные системы. – 2016. – №2, – С. 71-81.
9. 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. – 1981. – №24. – P. 381-395.
10. Smartphone Document Capture Competition <http://smartdoc.univ-lr.fr/task/>
11. Natalya Skoryukina, Julia Shemiakina, Vladimir L. Arlazarov, Igor Faradjev. Document localization algorithms based on feature points and straight lines. Proceedings Volume 10696, Tenth International Conference on Machine Vision (ICMV 2017). doi: 10.1117/12.2311478

**Тропин Даниил Вячеславович.** Московский физико-технический институт, г. Долгопрудный, студент. Количество печатных работ: 1. Область научных интересов: обработка и анализ изображений. E-mail: tropin.dv@phystech.edu

**Николаев Дмитрий Петрович.** Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, г. Москва, заместитель директора по научной работе. Количество печатных работ: более 180. Область научных интересов: машинное зрение, быстрые алгоритмы обработки изображений, распознавание образов. E-mail: dimonstr@itp.ru

**Слугин Дмитрий Геннадьевич.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва. Ведущий программист. Количество печатных работ: 9. Область научных интересов: системный анализ, алгоритмы обработки изображений, распознавание образов. E-mail: slugin@isa.ru

## The method of image alignment based on sharpness maximization

D.V. Tropin<sup>I,IV</sup>, D.P. Nikolaev<sup>II,IV</sup>, D.G. Slugin<sup>III,IV</sup>

<sup>I</sup> Moscow Institute of Physics and Technology (State University), Moscow, Russia

<sup>II</sup> Institute for Information Transmission Problems RAS (Kharkevich Institute), Moscow, Russ

<sup>III</sup> Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

<sup>IV</sup> Smart Engines, Moscow, Russia

**Abstract.** This paper describes an image alignment method based on sharpness maximization of average image. The proposed algorithm is substantiated for global-shift model of optical flow using an efficient way of calculation with Fast Fourier Transformation. For projective model is proposed an approach of image alignment by comparison separate fragments and using RANSAC to obtain the final transform. The experimental results of solution the problem of restoring document’s image in a video stream showing increase in quality of output images are presented.

**Keywords:** *image analysis, image alignment, optical flow, FFT, RANSAC.*

**DOI:** 10.14357/20790279180515

### References

1. Bulatov K., Arlazarov V.V., Chernov T., Slavin O. and Nikolaev D. “Smart IDReader: Document Recognition in Video Stream,” 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 39-44. doi: 10.1109/ICDAR.2017.347
2. Lucas B.D., Kanade, T. An iterative image registration technique with an application in stereo vision // Seventh International Joint Conference on Artificial Intelligence (IJCAI-81). – Vancouver, 1981. – P. 674-679.
3. Horn B.K.P., Schunck B.G. Determining optical flow. Artificial Intelligence // Artificial Intelligence. – Cambridge, 1981. – P. 185–203.
4. Lowe D.G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. – 2004. – № 56.
5. Nanne van Noord, Eric Postma. Learning scale-variant and scale-invariant features for deep image classification Pattern Recognition. Pattern Recognition. Volume 61, January 2017, Pages 583-592
6. Kuglin C.D. and Hines D.C. The phase correlation image alignment method. International Conference of Cybernetic Society. Proc IEEE 1975. New York. PP 163-165.
7. Szeliski Richard. Image Alignment and Stitching: A Tutorial. Foundations and Trends in Computer Graphics and Vision Vol. 2, No 1 (2006) 1–104
8. Arlasarov V.V., Zhukovsky A.E., Krivtsov V.E., Nikolaev D.P., Polevoy D.V. Analysis of features of the use of fixed and mobile small-sized digital video camera for OCR. // Information Technologies and Computing Systems – 2016. – №2, –P. 71-81.
9. 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. – 1981. – №24. – P. 381-395.
10. Smartphone Document Capture Competition <http://smartdoc.univ-lr.fr/task/>
11. Natalya Skoryukina, Julia Shemiakina, Vladimir L. Arlazarov, Igor Faradjev. Document localization algorithms based on feature points and straight lines. Proceedings Volume 10696, Tenth International Conference on Machine Vision (ICMV 2017); doi: 10.1117/12.2311478

**D.V. Tropin.** Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia, master’s student. E-mail: tropin.dv@phystech.edu

**D.P. Nikolaev.** Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia. Deputy director for scientific issues. E-mail: dimonstr@iitp.ru

**D.G. Slugin.** Institute for Systems Analysis, Federal Research Center «Computer Science and Control» of Russian Academy of Sciences, Moscow, Russia. Lead programmer. E-mail: slugin@isa.ru