

ДЕТЕКТОР СМАЗАННОСТИ И РАСФОКУСИРОВКИ НА ОСНОВЕ МОДЕЛИ ТЕКСТА

Костюков М.В.¹

¹ООО «Аби Продакшн», Москва, Россия (109390, г. Москва, ул. 1-я Текстильчиков, д. 12/9, офис 2), e-mail: mikhail.kostyukov@gmail.com

В работе рассматривается подход к детектированию смаза и расфокусировки документов на основе модели текста, посредством анализа второй производной функций яркости, построенных в окрестностях границ объектов по направлениям градиента яркости. Даны описание модели текста, примеры искаженных изображений текста, рекомендации по выбору объективных показателей, на основании которых можно выявить наличие или отсутствие смаза и расфокусировки. Предложен способ, позволяющий перевести выбранные показатели к числовым признакам, инвариантным к различным видам неравномерностей среди локальных характеристик анализируемого изображения. Описаны особенности построения классификатора для данной задачи. Уделено внимание вопросам производительности использованных алгоритмов. Основной анализ и получение характерных признаков объектов реализуется на изображении разницы гауссиан (DoG), что имитирует функцию зрения человека.

Ключевые слова: смаз, расфокусировка, искажение, шум, текстовый документ, обработка текста, OCR, модель текста, лапласиан, DoG, LoG, обработка изображений документов, распознавание символов, инвариантные моменты, классификация, активное обучение.

METHOD FOR BLUR AND DEFOCUS DETECTION BASED ON TEXT MODEL

Kostyukov M.V.¹

¹ABBY PRODUCTION LLC, Moscow, Russia (109390, Moscow, 1-st Tekstilschikov street, 2B, 12/9, office 2), e-mail: mikhail.kostyukov@gmail.com

The paper considers the problem of automatic blur and defocusing detecting on images of documents. The problem is solved based on text model and analysis of second derivative of brightness function along the brightness gradient. The paper gives text model description, examples of text distorted images and some advises on choosing descriptors for blur and defocusing detection and further distortion classification. The way for calculation of invariant number features by local nonuniformities of image from selected descriptors is proposed. The paper describes specificity for the classifier training. Particular attention is paid to the problem for performance of involved algorithms. Principal research is provided on difference of Gaussians image to mimic neural processing of the retina in human eye.

Keywords: blur, defocus, distortion, noise, text document, text processing, OCR, text model, Laplacian, DoG, LoG, document imaging, character recognition, invariant moments, statistical classification, active learning.

Введение

Задача детектирования смаза и расфокусировки изображения не теряет актуальности в течение нескольких десятилетий [6; 7; 9; 12-17; 19; 20]. Были получены определенные успехи в этой области, но работы по детектированию таких искажений на изображениях специального вида – фотографиях документов, отсутствуют или известны лишь в узких кругах [12; 13]. Однако, в связи с возросшей популярностью мобильных устройств и применением в них технологий оптического распознавания текста (OCR), такая задача требует решения.

Обнаружение смаза и расфокусировки востребовано, например, для предоставления информации о параметрах искажения, таких как степень смаза, его направление и другие свойства. На основании этой информации можно принять решение о необходимости переснять изображение или запустить алгоритм устранения искажений в случае, если степень выявленных искажений превышает заданный порог.

В общей постановке задача оказывается довольно сложной. Однако использование априорной информации о свойствах объектов (их модели), которые могут присутствовать на изображении, упрощает ее решение. Использование модели объектов может позволить определить достаточный набор признаков, чтобы идентифицировать тип искажения и его степень.

Таким образом, специализация множества рассматриваемых изображений упрощает задачу.

Решение задачи на практике предъявляет ряд требований:

- быстрое действие;
- достаточная надежность детектирования области с искажением;
- устойчивость к высоким шумам (характерным для мобильных устройств);
- устойчивость к неравномерности освещения, к слишком низкому/высокому освещению;
- устойчивость к гистограммным преобразованиям (например, эквализации [1]);
- возможность получения дополнительной информации о свойствах смаза/расфокусировки.

Сужение класса изображений оставляет задачу слишком сложной при наличии перечисленных требований.

В работах [6; 7; 9; 12-17; 19; 20] сообщается, что о наличии смаза или расфокусировки можно судить по следующим признакам:

- признаки в частотной области;
- особенности в гистограмме градиента;
- отсутствие насыщенных цветов;
- особенности в распределении границ по направлениям;
- локальная корреляция (размытие границ, параллельных движению).

Отдельные признаки или их комбинации с разной степенью успешности используются для нахождения рассматриваемых искажений на фотографиях реального мира. К сожалению, в рамках поставленной задачи подходы, описанные в указанных статьях [6; 7; 9; 12-17; 19; 20], оказываются непригодными по причине высокой вычислительной стоимости или по причине задействованных в подходе моделей объектов и искажений.

Возможность работы с изображением документа, подвергшегося искажениям, зависит как от их степени, так и от относительного размера и контраста информационных элементов. Так, маленькие или слабоконтрастные буквы практически нечитаемы при слабом размытии, хотя большие или сильноконтрастные буквы остаются хорошо различимыми при той же (или даже большей) степени размытости. Эта особенность также должна быть учтена при разработке алгоритма обнаружения любых видов искажений.

Для изложенных целей не нужно определять точную степень искажения. Важно сделать достоверное заключение относительно каждой области изображения, а именно подверглась ли она искажению или нет. И если ответ утвердительный, то с небольшой достоверностью определить какой вид искажения присутствует в изображении: смаз или расфокусировка. Притом, если разрабатываемый алгоритм не всегда сможет установить вид искажения в одной или нескольких областях, это может быть осуществлено при помощи экстраполяции из соседних областей.

Зрительная система человека без преувеличения является одной из самых совершенных систем обработки и анализа изображений. Исследователи в области компьютерного зрения и обработки изображений часто заимствовали биологические принципы функционирования для разработки новых, перспективных алгоритмов. Предлагаемый в статье алгоритм опирается на то, что первичная зрительная система выделяет лапласиан изображения для дальнейшего анализа контуров [1; 2; 15].

В настоящей статье представлен алгоритм детектирования смаза и расфокусировки на фотографиях документов, обладающий приведенными выше свойствами.

Модель текста

Срез идеального изображения документа близок к ступеньке. Если сделать срез через все изображение, то получится функция, близкая к кусочно-постоянной, или ступенчатая функция. Однако из-за несовершенства регистрирующей системы, последствий дискретизации и других процессов ступенчатая функция размывается и накладывается шум. Тем не менее на изображениях с незначительными искажениями границы текста представляют собой хорошо детектируемые (различимые) скачкообразные изменения яркости.

Смаз и расфокусировка заметнее всего влияет на границы отдельных объектов – уменьшается четкость границ (рисунок 1). Наличие объектов с хорошо выделяемыми границами характерно для достаточно обширных классов изображений, при этом границы объектов на изображении встречаются во всех направлениях. Это утверждение можно принять и для фотографий документов.

Построение второй производной перпендикулярно границам

Выводы о смазе или расфокусированности изображения можно делать на основе анализа вторых производных яркости изображения вокруг их точек пересечения нуля. Если в области были объекты с контрастными границами, то точки пересечения нуля детектируются хорошо [1; 2], в противном случае их отсутствие говорит о высоком уровне шума или сильной степени размытости. Расстояние между экстремумами связано со степенью размытости в окрестности точки. На рисунках 3 и 4 приведен пример, каким образом размытие влияет на функцию яркости и ее вторую производную (точнее, ее аппроксимацию, полученную с помощью DoG фильтра [1; 2]).

Шум на исходном изображении вызывает случайный сдвиг точек пересечения нуля второй производной. По этой причине, прежде чем вычислять вторую производную, производят предварительное сглаживание. Вместо явного сглаживания изображения и вычисления вторых производных по направлению в различных точках на изображении, следует использовать LoG или DoG фильтрацию (или любой фильтр, представляющий собой в общем смысле аппроксимацию лапласиана) [1; 2].

Изображение, полученное применением любого из фильтров DoG или LoG, в дальнейшем будем называть изображением лапласиана или просто лапласианом. Применение фильтра может быть эффективно вычислено за линейное время от количества пикселей на изображении [8; 10; 11; 18]. В случае использования лапласиана аппроксимация второй производной по заданному направлению функции яркости получается, как срез на изображении лапласиана, проходящий через пиксель вдоль этого направления. Эту аппроксимацию в окрестности исследуемой точки будем называть профилем точки.

Предлагаемый метод аппроксимации второй производной через использование LoG или DoG фильтрации даёт достаточно точное значение в рамках решения задачи идентификации наличия и степени искажения. Важным, но не решающим преимуществом DoG фильтра перед LoG является свойство радиальной симметрии. Кроме того, нейрофизиологические данные и эксперименты подтверждают, что сигнал, попадающий от сетчатки глаза в мозг, похож на лапласиан [2], что также позволяет надеяться на достаточность (во всех смыслах) лапласиана для решения поставленной задачи.

На рисунке 1 приведен модельный пример: он содержит фрагменты одного и того же изображения. Первый фрагмент иллюстрирует исходное изображение, второй и третий – изображение, размытое гауссовым фильтром. На последнем фрагменте после применения размытия к исходному изображению добавлен гауссов шум. На рисунке 2 приведены эти же изображения после применения DoG-фильтра. Дополнительно динамический диапазон лапласианов был нормирован.

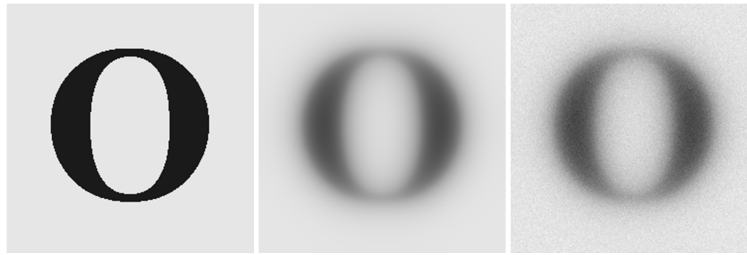


Рисунок 1 - Модельный пример. Слева направо показаны: исходное изображение, расфокусированное, расфокусированное с добавлением шума.



Рисунок 2 - Модельный пример. Слева направо показаны лапласианы для: исходного изображения, расфокусированного, расфокусированного с добавлением шума. Динамический диапазон лапласианов был нормирован.

При выборе параметров для DoG- или LoG-фильтрации следует учитывать особенности класса изображений и особенности устройства камеры, чтобы при известных параметрах шума и характеристиках объектов на изображении наилучшим образом детектировать наличие искажений. Поведение лапласиана в зависимости от этих особенностей проиллюстрировано на рисунках 3 и 4.

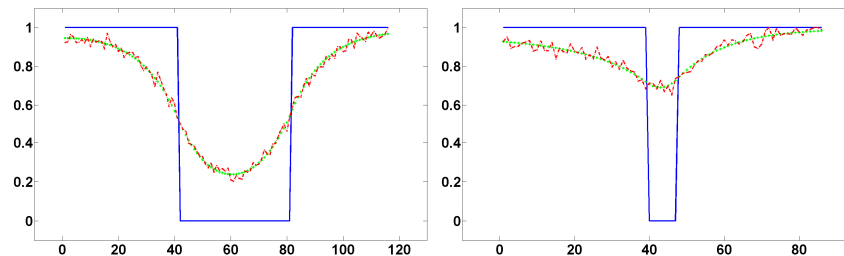


Рисунок 3 - Графики функций яркости (срезы), полученные для изображений, приведенных на рисунке 1. Слева срез для самой правой точки в горизонтальном направлении, справа – срез для самой нижней точки в вертикальном направлении.

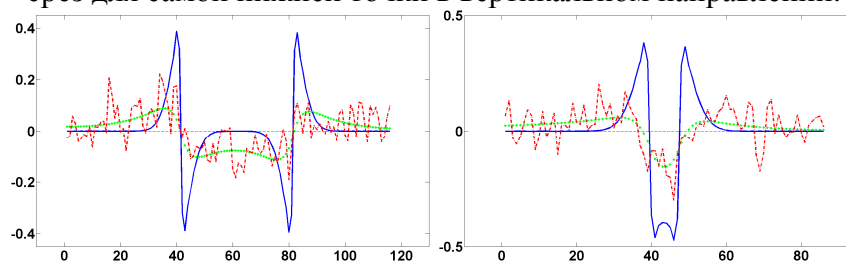


Рисунок 4 - Графики функций яркости (профили), полученные для изображений лапласиана, приведенных на рисунке 2. Слева профили для самой правой точки в горизонтальном направлении, справа – профили для самой нижней точки в вертикальном направлении. Функции, полученные для расфокусированных изображений, были умножены на 5.

Сбор статистики на основании анализа профиля второй производной перпендикулярно границам

После построения изображения лапласиана на нем необходимо найти точки пересечения нулевого уровня [1; 2]. Затем для каждой найденной точки необходимо построить ее профиль (вторую производную функции яркости вдоль направления градиента яркости). Направление профиля (градиента яркости) можно понимать как направление, перпендикулярное границе объекта.

Аналогично алгоритму Canny edge detection [1] можно провести квантование на 8 направлений (через 45 градусов), тем самым предоставляя возможность набрать информацию о свойствах профилей, не прибегая к интерполяции.

В зависимости от искажения, толщины букв и линий, их контраста, степени зашумленности изображения полученные профили будут отличаться – экстремумы будут находиться на разных расстояниях друг относительно друга, разным будет их абсолютное значение, а их величины будут подчиняться разным вероятностным законам распределения. Также следует заметить, что распределение количества детектированных точек пересечения нулевого уровня зависит от силы и направления смаза.

Для анализа искажений следует собирать статистику о свойствах профилей найденных точек пересечения нуля, например о свойствах локальных экстремумов, а именно: их абсолютном значении, расстоянии до соответствующей точки пересечения нулевого уровня и другим. Для каждого экстремума возможно также отмечать степень его надёжности, чтобы исключить ненадёжные экстремумы из подсчёта статистики. Можно полагать, что локальный экстремум является ненадёжным, если нельзя достоверно определить его положение, например, когда он сильно растянут или в окрестности его находятся два или более локальных экстремумов. Большая доля ненадёжных экстремумов может служить признаком присутствия сильных искажений на изображении.

Для различных свойств получаемых профилей вычисляются статистические оценки первого, второго, третьего моментов и их произведения (ковариация). Множество полученных значений предоставляет достаточно информации, чтобы в полной мере судить об исходных величинах, подверженных шумовому воздействию, и оказывается достаточным для рассматриваемой задачи. Это множество значений называется статистикой.

Следует отметить, что статистику нужно собирать для областей, так как в различных областях изображения могут присутствовать различные типы искажений, либо искажения могут отсутствовать вовсе. Помимо этого, сбор статистики в различных областях изображения позволяет нивелировать неравномерность освещенности (конечно, не полностью) и точнее определить параметры искажения. В проведенных экспериментах

статистика была собрана по непересекающимся квадратам. Сбор статистики следует организовывать таким образом, чтобы в каждую область попадало достаточное количество точек пересечения нулевого уровня, иначе статистические данные будут не достоверны.

Обработка набранной статистики

Информацию для статистики предлагается собирать отдельно для каждого выделенного направления градиента, поэтому и статистические оценки (среднее значение, дисперсия, коэффициент асимметрии, ковариационные моменты и т.д.) можно считать отдельно для каждого направления. Таким образом, в случае восьми выделенных направлений каждую оценку в области можно представить в виде 8 векторов, где длина вектора соответствует величине признака, полученного вдоль этого направления.

Собранная для каждого из восьми направлений статистика не является полностью пригодной для последующего анализа. В ней содержатся величины не инвариантные относительно поворотов, количества точек пересечения нуля и т.д. Таким образом, последующий анализ необходимо проводить, учитывая тот факт, что объекты (буквы, линии и др.) на изображениях могут иметь разную ориентацию, контраст. К тому же области документов могут быть по-разному заполнены текстовой информацией, иметь разную среднюю степень освещенности и т.д.

Решением этой проблемы может служить использование различных инвариантов, то есть величин, которые будут независимы от перечисленных выше локальных характеристик. Инварианты легко построить, если рассмотреть полученные векторы как точки на плоскости, имеющие массу, пропорциональную количеству величин, участвующих в формировании соответствующего им вектора. Таким образом, одной из важных возможных характеристик могут стать моменты инерции, точнее собственные значения тензора инерции. Развивая эту идею, также можно использовать хорошо известные инварианты, такие как моменты X_u и их производные [1; 10].

Некоторые статистические оценки собранных данных могут служить в качестве оценок свойств ядра, искажающего преобразования в данной области. Например, вытянутость эллипсоида инерции (эксцентриситет) для векторов расстояний между экстремумами служит индикатором того, смазано изображение или расфокусировано. Упомянутые выше оценки могут быть использованы для построения обратного преобразования.

Параметры фильтра, выделяющего лапласиан, необходимо подбирать, опираясь на размер объектов (букв текста) в пикселях и оценку шумовых характеристик. Оптимальный размер блоков, параметры фильтра для нахождения второй производной и другие необходимые настройки следует подбирать посредством проведения экспериментов.

Статистический анализ экспериментальных данных показал, что классифицировать объекты по одному-трем признакам не представляется возможным. Эта ситуация вызвана тем, что реальный смаз оказывается значительно неоднородным, что может появиться, например, в виде эффекта «двоения» объектов (рисунок 5).

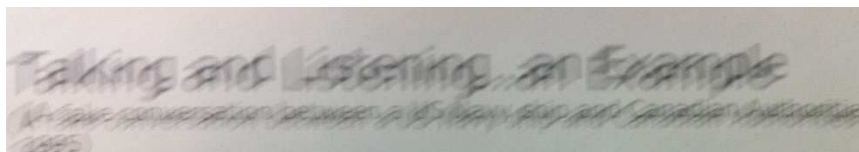


Рисунок 5 - Пример реального изображения.

Обычной практикой в таких случаях стало использование классификаторов [3; 5]. Привлекательность этого подхода очевидна: результат может быть получен автоматически без проведения долгосрочных работ.

Как было указано, весь необходимый анализ может быть реализован на основе алгоритмов, требующих для расчета линейного времени от количества пикселей в изображении. Вычислительно трудоемкими являются этапы построения изображений лапласиана и нахождения точек пересечения нуля на нем. Сложность остальных вычислений мала по сравнению с ними.

Результаты экспериментов

В качестве классификатора в работе было использовано дерево принятия решений. Этот известный и хорошо изученный классификатор имеет свои достоинства и недостатки. К первым свойствам следует отнести возможность содержательно интерпретировать процесс классификации, быстроту обучения и последующей работы. Основной недостаток заключается в том, что результирующий распознаватель сильно зависит от обучающего набора и в конечном итоге может плохо работать на реальных данных [3; 5]. Однако этот недостаток был преодолен при помощи тщательного отбора обучающих примеров, полуавтоматической обрезкой ветвей дерева и введением двух дополнительных классов при классификации, а именно отвечающих случаям, когда искажение возможно и когда ни наличие, ни отсутствие искажения не может быть определено.

Для решения ограничений статистического обучения использовалось активное обучение. Идея активного обучения состоит в том, чтобы постепенно дополнять обучающую выборку сложными для классификации примерами без экспоненциального роста их числа. Так, после первоначального обучения классификатора появляется возможность определить сложные для него примеры – те, которые он проклассифицировал неуверенно. Эти примеры размечаются и добавляются к начальной выборке, и выборка расширяется сложными для обучения и адаптированными под задачу примерами. После нескольких таких итераций

обучения ожидается увидеть классификатор, построенный на относительно малой обучающей выборке и превосходящий по свойствам классификатор, однократно обученный на большой выборке. Похожий метод был предложен в статье Абрамсона и Фройнда [4].

В процессе настройки классификатора было выявлено, что признаков не бывает много: даже те признаки, которым тяжело дать «физическое» объяснение, оказывались задействованы на нижних уровнях дерева решения или в листьях. Можно дать рекомендацию использовать все возможные данные, которые возможно посчитать для исследуемой области, устраняя из анализа лишь те признаки, которые сильно коррелируют.

Для использования классификаторов необходимо подобрать примеры тех классов изображений, которые следует разделять. При сборе обучающей базы обнаружилась проблема, заключающаяся в том, что сложно получить изображения, подвергшиеся целиком только одному виду искажений. Использование дерева решений в комплексе с активным обучением упростило решение задачи классификации, при которой в первую очередь и наиболее достоверно важно выявить наличие искажения в области, и с меньшей точностью нужно определить, является ли выявленное искажение смазом или расфокусировкой. Притом, если классификатор не в состоянии установить с высокой степенью достоверности наличие искажения или его вид в одном или нескольких блоках, то применяется экстраполяция результатов классификации их соседей.

Реализованный алгоритм с использованием классификатора, обученного описанным выше способом, был проверен на базе изображений. Было проконтролировано, что вероятность детектирования искажений хорошо коррелирует с количеством неверно распознанных символов при помощи OCR. Точность определения вида искажения была проверена вручную и признана достаточной. Более того, в тех случаях, где вид искажения не мог определить классификатор, не мог определить вид искажения и человек.

Выводы

В работе рассмотрен новый алгоритм для определения наличия смаза и расфокусировки на изображениях документа. В основу предложенного алгоритма заложена модель текста – предположение о кусочном постоянстве исходного сигнала. Предлагаемый алгоритм обладает несколькими важными свойствами, а именно: высокой достоверностью детектированной области, высоким быстродействием, устойчивостью к высоким шумам, к неравномерности освещения, к слишком слабому или сильному освещению, к гистограммным преобразованиям, дает возможность получить дополнительную информацию о характеристиках смаза/расфокусировки.

Подход может быть применен и для изображений натуральных сцен, поскольку многие натурные изображения также содержат границы типа «ступенька», что дает возможность применить разработанную технику и для них [9].

В работе приведен ряд решений, позволяющих эффективно вычислять промежуточные данные алгоритма. Предложен способ, позволяющий представить данные о смазе по различным направлениям в числовые признаки, инвариантные к повороту, масштабированию, количеству объектов, их контрасту и другим видам неравномерностей среди локальных характеристик анализируемого изображения.

Работы проводятся при финансовой поддержке Министерства образования и науки Российской Федерации в рамках выполнения ГК 07.514.11.4158 по теме: «Разработка алгоритмов поиска и локализации текста на фотографиях и видеокадрах».

Список литературы

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. — М. : Техносфера, 2005, 2006. — 1072 с.
2. Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов / пер. с англ. Н.Г. Гуревич. — М. : Радио и связь, 1987. — 400 с.
3. Мерков А.Б. Распознавание образов. Введение в методы статистического обучения. - М. : Едиториал УРСС, 2011. – 256.
4. Abramson Y. and Freund Y. Active learning for visual object recognition, Technical report, UCSD, 2004
5. Bishop Ch. Pattern recognition and machine learning. – Springer-Verlag, New York, 2006
6. Cannon M. Blind Deconvolution of Spatially Invariant Image Blurs with Phase. IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-24, pp. 58-63, 1976.
7. Chung Y.C., Wang J.M., Bailey R.R., Chen S.W. and Chang S.L. A non-parametric blur measure based on edge analysis for image processing applications. Proc. IEEE Conf. Cybern. Intell. Syst., vol. 1, pp.356 -360 2004
8. Deriche R. Recursively implementing the Gaussian and its derivatives. Proceedings of the 2nd International Conference on Image Processing, Singapore, p. 263–267. 1992
9. Elder J.H., Zucker S.W. Local scale control for edge detection and blur estimation, IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 7, pp.699 -716 1998
10. Flusser J. et al. Moments and moment invariants in pattern recognition. – Wiley, 2009
11. Hale Dave Recursive gaussian filters. CWP-546, 2006.

12. Liu R., Li Z., and Jia J. Image Partial Blur Detection and Classification. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.
13. Liang Luhong, Chen Jianhua, Ma Siwei, Zhao Debin, Gao Wen A no-reference perceptual blur metric using histogram of gradient profile sharpness. Image Processing, IEEE International Conference - ICIP , pp. 4369-4372, 2009
14. Marziliano P., Dufaux F., Winkler S., and Ebrahimi T. A no-reference perceptual blur metric. International Conference on Image Processing, Vol. 3, Rochester, NY, pp. 57–60, 2002.
15. McIlhagga W.H. & May K.A. Optimal edge filters explain human blur detection. Journal of Vision, 12(10):9, 1–13. 2012
16. Rugna J. Da and Konik H. Automatic blur detection for metadata extraction in content-based retrieval context. In SPIE, volume 5304, pages 285–294, 2003.
17. Savakis A.E. and Trussell H.J. Blur identification by residual spectral matching. IEEE Trans. Image Process. 2, 141–151, 1993.
18. Van Vliet L.J., Young I.T., and Verbeek P.W. Recursive Gaussian derivative filters. Proceedings ICPR '98, 509-514. IEEE Computer Society Press, 1998.
19. Yitzhaky Y. and Kopeika N.S. Identification of blur parameters from motion blurred images. CVGIP: graphical models and image processing, 59(5), pp. 321-332, September 1997.
20. Yitzhaky Y., Mor I., Lantzman A., Kopeika N.S. Direct method for restoration of motion-blurred images. J. Opt. Soc. Am. A., 15(6), 1512-1519, June 1998.

Рецензенты:

Кузнецов С.О., д.ф.-м.н., профессор, заведующий кафедрой анализа данных и искусственного интеллекта Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ), г. Москва.

Гриненко М.М., д.ф.-м.н., научный консультант, ООО «Аби ИнфоПоиск», г. Москва.