

МЕТОД ПОВЫШЕНИЯ РАЗРЕШЕНИЯ ДЛЯ ИЗОБРАЖЕНИЙ ДОКУМЕНТОВ С МОБИЛЬНЫХ УСТРОЙСТВ

Загайнов И.Г.¹

¹ООО «Аби ИнфоПоиск», Москва, Россия (127273, г. Москва, ул. Отрадная, 2Б, стр. 6, офис 7-02), e-mail: Ivan_Z@abby.com

В процессе распознавания документов часто встречается ситуация, когда исходное изображение обладает недостаточным разрешением. Это может быть обусловлено ограничением разрешения матрицы в камере мобильного устройства, потерями в оптической системе, алгоритмами кодирования данных в изображении и другими факторами. При этом остается возможность съемки нескольких таких изображений. В данной работе рассматривается задача получения изображений документов более высокого качества и/или разрешения (superresolution) из нескольких изображений, снятых камерой мобильного устройства. Целью является улучшение распознавания исходного документа. На основе допущений о классе изображений предложены модели движения для компенсации межкадрового сдвига. Разработан алгоритм для накопления полезного сигнала из последовательности кадров. Проведен ряд экспериментов по распознаванию получаемых изображений документов, в ходе которых подтверждена эффективность предлагаемого метода.

Ключевые слова: повышение разрешения, интерполяция изображений, обработка изображений документов, распознавание символов, модели движения, оценка движения, компенсация движения.

SUPER-RESOLUTION METHOD FOR DOCUMENT IMAGES FROM MOBILE DEVICES

Zagaynov I.G.¹

¹ООО «АББЫЙ InfoПоиск», Moscow, Russia (127273, Moscow, Otradnaya street, 2B, building 6, office 7-02), e-mail: Ivan_Z@abby.com

It is common situation that resolution of the original document image is not sufficient for optical character recognition (OCR). This may be the result of limited resolution of matrix in the mobile device camera, the losses in optical system, the image encoding algorithms, and other factors. However, there is an option to take a number of images. This paper considers the problem of obtaining higher quality and higher resolution for document images. Several images or video sequence should be available for the purpose. The aim of improvements is better optical character recognition of the original document. The paper gives the assumptions of the class of document images and proposes motion models to compensate inter-frame shifts. The algorithm for signal accumulation from sequence of images is described. The experimental results confirm that the proposed method improves optical character recognition.

Keywords: super-resolution, image interpolation, document imaging, character recognition, motion models, motion estimation, motion compensation.

Введение

Классические методы повышения разрешения с использованием одной или нескольких фотографий сцены подробно описаны в работах [4; 15; 16]. Однако все они предполагают большое разнообразие изображаемых объектов, каковыми могут быть люди, природа, архитектура и т.п. Можно сделать несколько предположений относительно модели получаемых изображений документов, что позволит существенно сократить вычислительные затраты алгоритмов и в то же время получить результаты распознавания (OCR) более высокого качества.

Можно считать, что изображения документов получают одной и той же камерой (устройством), деформация документа и изменения в освещенности сцены незначительны при переходе от одного кадра к другому.

Документ представляет собой текст, напечатанный на бумаге или нанесенный на неё типографским способом. Таким образом, текстовый документ можно представить моделью двумерного кусочно-постоянного сигнала, содержащего значительные по площади участки однородного фона (бумаги) и участки с краской, отличающиеся от фона по яркости (текст).

Можно считать, что движение в сцене обусловлено в основном лишь движением камеры устройства, находящегося в руках человека. Следует заметить, что это движение может быть частично скомпенсировано встроенными алгоритмами компенсации движения.

Для наиболее эффективного повышения качества/разрешения изображений необходимо как можно более точно описать реальное перемещение объектов от одного кадра к другому (задача поиска и компенсации движения рассматривается в [6-9]). В данной работе предлагается использование двух моделей движения, совместное использование которых улучшает результаты последующего распознавания изображений высокого разрешения.

1. Модели движения

Рассмотрим две модели, описывающие движение между двумя кадрами сцены. Первая модель (интерполяционная) предполагает наличие плавно меняющегося поля векторов перемещения с незначительными градиентами компонент векторов, что может быть обусловлено наличием перспективы, небольшими деформациями самого документа, а также искажениями оптической системы. Вторая модель движения (перспективная) предполагает, что все изменения в сцене могут быть описаны лишь перспективным преобразованием одного кадра к другому, близким к тождественному преобразованию.

Обе модели движения используют в качестве исходной информации совокупность векторов перемещения, определенных для блоков изображения, называемую полем. Аналогичное поле векторов перемещения используется во многих алгоритмах кодирования видеоинформации, таких как MPEG-1,-2,-4, H.263, H.264 [6-9]. Для построения этого поля изображение разбивается на непересекающиеся квадратные блоки заданного размера b , затем для каждого из блоков отыскивается вектор перемещения блока от кадра к кадру, обеспечивающий наилучшее совмещение блоков в смысле выбранной метрики подобия [2; 17]. Методы сокращения вычислительных затрат для поиска векторов перемещения могут быть найдены в работах [1; 3; 11; 12; 14].

Оптимальный размер блоков и относительный размер области поиска векторов перемещения выбирались в результате многочисленных экспериментов по оценке движения в сцене. Обнаружена зависимость оптимального размера блока от типа документа и

разрешения камеры: блок по размерам должен приблизительно соответствовать размеру буквы в основном тексте документа.

Получаемое в результате поле векторов перемещения, как правило, содержит большое количество ошибочно определенных векторов, которые можно отбросить, применяя несколько простых правил.

1. Блоки, расположенные на краях изображения, могут не принадлежать документу, для них характерны сильные оптические искажения, поэтому они выбрасываются из рассмотрения. Оптимальный размер области для дальнейшего анализа определялся экспериментально. Ширина отбрасываемой полосы по краю изображения – s блоков.

2. Блоки, содержащие лишь фоновые области документа, не могут характеризовать движение в сцене. Найденные для них вектора перемещения носят случайный характер, поэтому их следует также отбросить. Критерий принадлежности блока к «фоновым» можно определить так:

$$D_b < kD_n,$$

где D_b – оценка дисперсии яркости пикселей в рассматриваемом блоке изображения, D_n – оценка дисперсии шума на изображении, k – коэффициент, определяемый экспериментально для различных типов камер (мобильных устройств).

3. Блоки, вектор перемещения которых отличается от среднего вектора перемещения окружающих блоков на величину, превышающую определенный порог t , отбрасываются (критерий когерентности движения).

Поле векторов перемещения, получаемое в результате применения этих правил, используется в каждой из моделей движения, при этом конечной целью является получение векторов перемещения, наиболее точно описывающих истинное движение каждой точки изображения. В следующем разделе этот процесс будет рассмотрен более подробно для каждой из моделей.

1.1. Обработка в рамках интерполяционной модели движения

В этой модели векторы перемещения отброшенных блоков восстанавливаются из окружения (используются векторы оставшихся блоков) путем итерационного применения свертки с усреднением в окне размером 3×3 блока (9 векторов). Компоненты векторов сглаживаются независимо. В качестве мгновенного значения компоненты вектора перемещения берется среднее значение компоненты из блоков в окне. Процесс повторяется итерационно до тех пор, пока не будут восстановлены значения векторов перемещения у всех блоков изображения.

Получаемое в результате поле векторов перемещения может содержать в себе резкие перепады между обширными зонами, которые возникают при интерполяции или

экстраполяции векторов различных блоков. Для их подавления поле дополнительно сглаживается посредством применения двумерной свертки с прямоугольным окном большого размера, содержащим от 7x7 до 21x21 блоков (от 49 до 441 векторов, соответственно). Поле векторов перемещения для блоков изображения затем необходимо преобразовать в поле векторов перемещения для отдельных пикселей изображения. Предлагается делать это путем применения билинейной интерполяции величин, определенных для блоков. Применение интерполяции по ближайшим соседям, как это сделано во многих стандартах кодирования видео для снижения вычислительных затрат, приводит к появлению заметных блочных артефактов при компенсации движения, и по этой причине ее не следует использовать.

1.2. Обработка в рамках перспективной модели движения

В рамках этой модели предполагается, что движение в сцене хорошо описывается перспективным преобразованием, параметры которого необходимо подобрать. Для векторов перемещения блоков можно написать систему нелинейных уравнений, такую, что ее решением будет вектор параметров перспективного преобразования. Эта система будет переопределенной, для ее решения можно использовать хорошо известный метод Levenberg–Marquardt [10; 13]. В качестве начального вектора параметров задается тождественное преобразование или преобразование, описываемое средним вектором перемещения. По полученным параметрам перспективного преобразования можно определить вектор перемещения для каждой точки изображения.

Вне зависимости от того, какая модель движения используется, каждый пиксель изображения приобретает свой уникальный вектор перемещения, незначительно отличающийся от векторов соседних пикселей. Для обеспечения точной компенсации движения преобразованное изображение получается с использованием бикубической интерполяции [5] яркости пикселей исходного изображения. При интерполяции используются точки исходного изображения, определяемые с учетом их векторов перемещения. После компенсации движения кадры оказываются совмещенными, и можно приступить к процедуре накопления полезного сигнала, которая описана ниже.

2. Повышение качества и разрешения изображений

Алгоритм повышения разрешения состоит в накоплении полезного сигнала в исходном разрешении (повышение качества изображения), но может также совмещать в себе процедуру повышения пространственного разрешения (up-scaling) изображения, чего можно достигнуть несколькими способами:

- 1) геометрическое увеличение исходных изображений. Этот путь заметно увеличивает расход памяти и вычислительные затраты алгоритма, что делает его неприемлемым для мобильных устройств;
- 2) геометрическое увеличение получаемого в результате накопления «улучшенного» изображения. Этот подход не дает существенного прироста в качестве распознавания;
- 3) геометрическое увеличение при по-пиксельной (pixel-wise) компенсации движения с использованием бикубической интерполяции. Процедура бикубической интерполяции в более высокое разрешение производится один раз для каждого кадра перед этапом накопления сигнала. Первый кадр увеличивается без использования движения, последующие – с учетом векторов перемещения. В процессе накопления артефакты, присущие бикубической интерполяции, усредняются и подавляются. Эксперименты подтверждают, что получаемый при этом результат распознается лучше, чем при использовании второго способа.

Процесс накопления сигнала состоит в «добавлении» полезного сигнала из очередного кадра, скомпенсированного с учетом движения, к изображению-аккумулятору. При инициализации процесса, в качестве аккумулятора можно использовать первый кадр из серии, или любой другой, например наименее смазанный/расфокусированный. Для каждого последующего кадра необходимо провести процедуру поиска и компенсации движения к первому кадру или к аккумулятору.

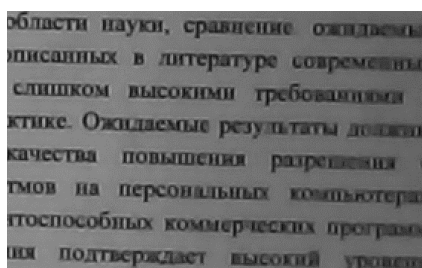
Накопление можно производить простым усреднением сигнала из аккумулятора и добавляемого кадра с различными весами. Если необходимо получить равномерное усреднение между всеми кадрами в серии, вес добавляемого кадра должен вычисляться как:

$$W_n = 1 / n,$$

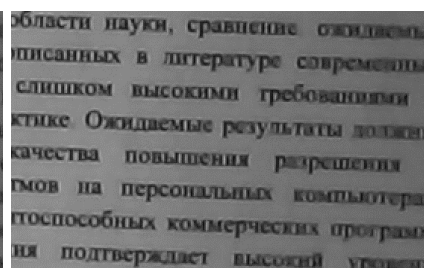
где n – номер добавляемого кадра, а вес аккумулятора как:

$$W_a = 1 - W_n.$$

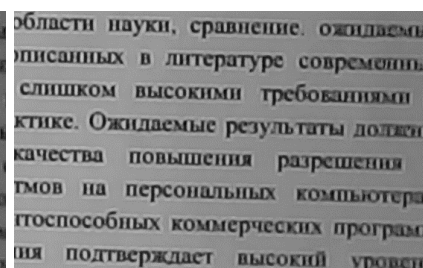
Возможно применение более сложных схем накопления, учитывающих для каждого пикселя добавляемого изображения информацию о его окружении, а также информацию из аккумулятора (такую как дисперсия, локальный контраст и т.п.).



(a)



(б)



(в)

Рис. 1. Один и тот же участок изображения документа из кадра номер 1 (а), из кадра номер 20 (б), и полученного в результате накопления 20 кадров (в), разрешение увеличено в 2 раза.

Результат сложения аккумулятора и очередного кадра помещается в аккумулятор. Далее процедура накопления повторяется для последующих кадров в серии. В результате накопления N кадров получается изображение, имеющее значительно более высокое качество (рисунок 1).

Также для более быстрого и эффективного накопления полезного сигнала можно отсортировать имеющиеся кадры по предварительной интегральной оценке их качества (например, по степени их расфокусировки/смаза) и производить накопление, начиная с наилучших кадров последовательности. В этом случае можно значительно сократить количество используемых кадров и повысить производительность. Кроме того, такой подход гарантирует, что для любого N распознавание накопленного изображения документа окажется не хуже, чем для любого отдельного кадра из этой же серии.

Результаты экспериментов

В ходе экспериментов было выявлено, что наибольшая надежность в определении движения достигается при совместном использовании двух моделей движения, как гипотез с весами, которые могут зависеть от типа документа, характеристик камеры и т.д. В этом случае вектор движения для каждой точки определяется как взвешенная сумма векторов, полученных из двух моделей движения:

$$V = mV_1 + (1 - m)V_2,$$

где V_1 и V_2 – векторы перемещения точек, получаемые при использовании 1-й и 2-й моделей соответственно, m – весовой коэффициент, $m \in [0,1]$.

Оптимальные значения параметров алгоритма (размер блока b , ширина полосы по границам s , коэффициент k , порог t , коэффициент m , коэффициенты бикубической интерполяции) могут зависеть от типа документа и характеристик камеры/устройства. Тип документа может задаваться пользователем или определяться автоматически при предварительном анализе одного из изображений сцены.

Для получения наилучших результатов настройка параметров алгоритма должна производиться в связке с модулем распознавания текста (OCR), который предполагается в дальнейшем использовать. При подборе параметров критерием для оптимизации может выступать количество ошибок в распознанном документе по сравнению с эталоном, различные встроенные в OCR критерии уверенности распознавания или другие метрики. В данной работе для настройки алгоритма использовался ABBYY FineReader.

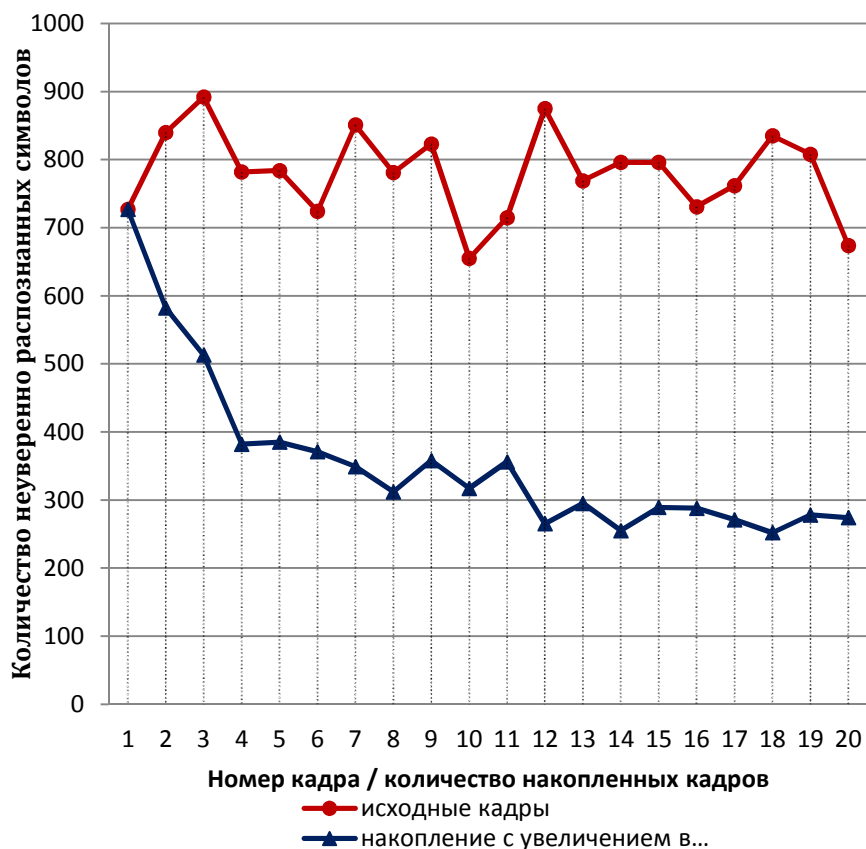


Рис. 2. Результаты работы распознавателя (ABBYY FineReader) на исходных кадрах видеопоследовательности и кадрах, полученных в результате применения метода superresolution с увеличением в 2 раза. Количество неуверенно распознанных символов многократно превосходит количество ошибок распознавания и использовано здесь для наглядности.

На рисунке 2 приведены результаты эксперимента по распознаванию исходных кадров, полученных с мобильного устройства, позволяющего делать фотографии с разрешением 1280x720 пикселей, и кадров, полученных в результате накопления сигнала с увеличением в 2 раза. Использовались обе модели движения, параметры алгоритма: $b=16$, $s=5$, $kD_n=10$, $t=1$, $m=0.4$. Видно, что процесс накопления полезного сигнала носит экспоненциальный характер, наибольший прирост качества распознавания происходит при накоплении информации из первых 5–8 кадров, далее процесс замедляется, после 20 кадра прирост оказывается несущественным, и дальнейший процесс накопления не имеет смысла.

Выводы

Разработанный алгоритм позволяет существенно сократить число ошибок при распознавании документов, получаемых с камер мобильных устройств. Для различных типов документов их количество снижается до 2-3 раз. В то же время число ошибок распознавания типового документа формата А4 может быть изначально слишком велико (для некоторых

типов мобильных устройств оно может измеряться сотнями на страницу), и применение метода, описываемого в статье, по-прежнему не приводит к приемлемому качеству распознавания. Результаты, полученные для изображений с мобильных телефонов iPhone 4S, 5 (с применением описанного метода) соответствуют результатам распознавания изображений документов, получаемых с планшетного сканера. Таким образом, при применении разработанного алгоритма можно использовать эти устройства как альтернативу планшетным сканнерам.

Дальнейшее развитие метода может заключаться в использовании различной информации об окружающих элементах изображения в процессе накопления полезного сигнала и построении более точных и надежных схем для оценки межкадрового движения. Также возможна модификация алгоритмов OCR (настройка их параметров при использовании совместно с методом повышения разрешения) для получения ещё более высоких результатов.

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. Загайнов И.Г. Оптимизация алгоритма компенсации движения в системах кодирования подвижных изображений // Исследовано в России : электронный журнал. – 1999. - 45. - URL: <http://zhurnal.ape.relarn.ru/articles/1999/045.pdf>
2. Boujut, J. Benois-Pineau, O. Hadar, T. Ahmed, P. Bonnet Weighted-MSE based on Saliency map for assessing video quality of H.264 video streams, IS&T // SPIE Electronic Imaging, San Francisco, 2011.
3. Cheung C.K., Po L.M. A New Fast Block Motion Estimation Algorithm Using Normalized Partial Distortion Measure. - International Workshop on Multimedia Data Storage, Retrieval, Integration and Applications, pp. 34-39, January 2000.
4. Daniel Glasner, Shai Bagon, Michal Irani Super-Resolution from a Single Image. - Dept. of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Israel, ICCV 2009.
5. Don P. Mitchell, Arun N. Netravali Reconstruction Filters in Computer Graphics, ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques 22 (4). pp. 221–228, 1988.

6. JTS 1/SC 29, Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – part 2: Video, ISO/IEC 11172-2, 1993.
7. JTS 1/SC 29, Information technology – Generic coding of moving pictures and associated audio information: Video, ISO/IEC 13818-2, 1995.
8. ISO/IEC 14496-10, Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding, 2012.
9. ITU-T SG15, Video Coding For Low Bitrate Communication, DRAFT ITU-T Recommendation H.263, May 1996.
10. Levenberg, K. A method for the solution of certain non-linear problems in least squares, *Quart. Appl. Math.*, 2 (1944), 164–168.
11. Liu L.K., Feig E. A Block-Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 4, pp. 419-422, August 1996.
12. Liu D., Sun W. Block-Based Fast Motion Estimation Algorithms in Video Compression, Stevens Institute of Technology, Technical report 9704, October 1997.
13. Marquardt, Donald W. An algorithm for least-squares estimation of nonlinear parameters, *Journal. Soc. Indust. Applied Math.*, 11, 2 (1963), 431–441.
14. Po L.M., Ma W.C. A Novel Four-Step Search Algorithm for Fast Block Motion Estimation, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 313-317, June 1996.
15. Sean Borman, Robert L. Stevenson Super-Resolution from Image Sequences - A Review, *Midwest Symposium on Circuits and Systems*, 1998.
16. Sina Farsiu, M. Dirk Robinson, Michael Elad, Peyman Milanfar Fast and Robust Multiframe Super Resolution, *IEEE Trans. On Image Processing*, vol. 13, no. 10, October 2004.
17. Stefan Winkler, Praveen Mohandas The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics, *IEEE Trans. Broadcasting* vol. 54, no. 3, September 2008.

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.