

Геометрическая коррекция фотографий документов

Ramiz Zeynalov

Department of Computational Mathematics and Cybernetics

Moscow State University, Moscow, Russia

ramiz_zeynalov@mail.ru

Аннотация

На сегодняшний день получили широкое распространение сотовые телефоны и КПК со встроенными камерами, которые часто используются для фотографирования разворотов книг. При этом текст на фотографии претерпевает значительные геометрические искажения, возникающие из-за того, что поверхность книги имеет неплоскую форму. Существующие алгоритмы распознавания текста не могут справиться с такими входными данными. В данной работе предлагается новый алгоритм геометрической коррекции фотографий документов, существенно повышающий точность распознавания. Алгоритм опирается на оценку формы поверхности с текстом.

Ключевые слова: строки текста, геометрическая коррекция, неравномерное освещение, распознавание текста

1. ВВЕДЕНИЕ

В настоящее время актуальна задача предобработки фотографий книг перед их распознаванием. При этом требуется произвести фотометрическую и геометрическую коррекцию изображения текста.

Постановка задачи:

Дана фотография неплоской поверхности книги - страница или разворот с текстом. Требуется подготовить фотографию к распознаванию, то есть, формально, увеличить процент правильно распознанных символов.

Изображения документов с фотоаппарата зачастую содержат следующие дефекты:

1. Неравномерное освещение (блики)
2. Шум
3. Непрямые строки текста (поверхность с текстом неплоская)

Первые две проблемы из списка встречаются во многих задачах обработки изображений. Наибольшая сложность заключается именно в выпрямлении строк текста.

Существующие методы решения этой задачи можно разделить на два принципиально разных подхода – форма строк оценивается методом SFS ([1], [7]), форма строк оценивается путём анализа самих строк на изображении ([2], [3], [4], [8]). Первый подход более универсальный и поэтому хуже подходит для решения более узкого класса задач. В некоторых методах из второй группы не используется никакой пространственной информации ([2], [3], [8]).

Для решения этой задачи предлагается метод, основанный на оценке формы поверхности документа. Достаточно распространённый пример – фотография разворота книги.

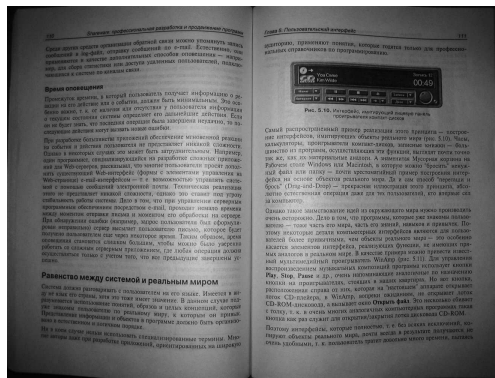


Рис. 1. Пример фотографии разворота книги

При решении этой задачи разумно ввести некоторые предположения, которые смогут упростить решение задачи без существенного ограничения области применимости метода.

Используемые предположения

1. Строки текста книги параллельны друг другу
2. Гладкость поверхности книги
3. На изображении страница ориентирована вертикально

2. ОПИСАНИЕ МЕТОДА

Модель искажения

Полное искажение рассматривается как последовательность следующих искажений:

1. Радиальная дисторсия
2. Проективное преобразование
3. Неплоская форма поверхности страницы

Влияние первого искажения в исходных данных решаемой задачи, как правило, достаточно мало. Поэтому в предложенном методе влияние этого искажения не учитывается. В отличие от первого искажения, вторые два можно увидеть – вследствие влияния проективного преобразования и неплоской формы поверхности страницы строки текста сверху и снизу страницы загибаются с разной силой. Поэтому при решении поставленной задачи учитывается композиция проективного преобразования и неплоская геометрия поверхности страницы.

Метод состоит из следующих 8 шагов

1. Предобработка изображения
2. Нахождение строк текста на изображении
3. Анализ формы строк текста
4. Построение грубой поверхности страницы на некотором множестве точек
5. Уточнение поверхности страницы
6. Аппроксимация поверхности (полиномом 3-й степени)

7. Коррекция геометрии
8. Выпрямление проблемных областей, используя преобразования на плоскости

Предобработка заключается в коррекции освещения страницы и удалении бликов возникающих, при искусственном освещении страницы от лампы. Коррекция освещения осуществляется стандартным методом попиксельного деления изображения на его сильно размытую копию. Этот метод является частным случаем метода Retinex [6]. Размытие изображения производится фильтром Гаусса с большими значениями радиуса размытия и стандартного отклонения σ .

Блики на подобных фотографиях присутствуют в областях, в которых нормаль к поверхности направлена в оптический центр камеры, в результате чего освещённость в таких точках достигает максимального значения. Удаление бликов производится путём применения гамма-коррекции.

Поиск строк текста на изображении выполняется путем поиска связанных компонент. Для этого к изображению сначала применяется операция пороговой бинаризации, а затем морфологическая операция расширения двумя различными структурными элементами: диском и горизонтальной линией. Применение таких структурных элементов позволяет объединить символы и слова одной строки в единую связную область. Результат применения морфологических операций продемонстрирован на рис. 2.

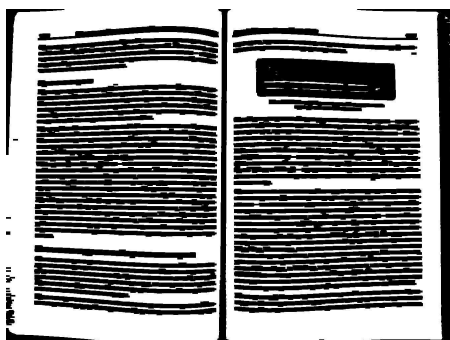


Рис. 2 Результат применения последовательности морфологических операций

Затем выполняется поиск связанных областей, соответствующих строкам. Часть найденных областей отбрасывается по следующим критериям:

- площадь меньше порога P_1 (шум);
- площадь больше порога P_2 (рисунки в тексте);
- длина по горизонтали меньше порога L_{\max}

Для найденных строк производится аппроксимация полиномами 3-й степени, используя МНК [5].

Построение грубой поверхности страницы на некотором множестве точек основывается на предположении, что в результате выполнения повторного проецирования строки должны быть прямыми. С учётом этого предположения вычисляется значение Z-координаты соответствующей точки на поверхности страницы. Каждая строка в результате должна оказаться прямой. С учётом этого рассчитываются смещения точек вдоль оси OZ. Значения Z-координаты вычисляются в точках, принадлежащих найденным образам областей, соответствующих строкам текста.

Пусть $(x_i; y_i)$ – точка строки, $p(x)$ – полином, аппроксимирующий форму этой строки, $\bar{y} = \sum y_i$ – среднее значение y по точкам строки, y_i' – значение ординаты данной точки на изображении после коррекции, h – высота изображения, тогда значение глубины поверхности находится следующим образом:

$$\Delta Z = \frac{y_i - y_i'}{h} \quad (1)$$

$$y_i' = \bar{y} + y_i - p(x_i) \quad (2)$$

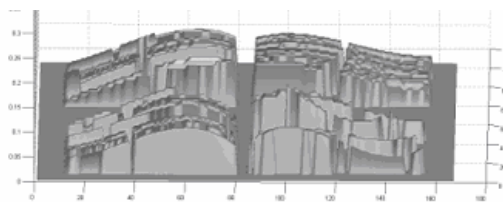


Рис. 3 Оценка поверхности в некоторых точках

В предположении, что глубина поверхности изменяется только в направлении оси OX, вычисление значения глубины поверхности в остальных точках производится путём усреднения значения координаты Z для каждого значения координаты Y по столбцам. В результате получается поверхность, определённая уже во всех точках, но не достаточно гладкая, что может привести к плохим результатам на стадии обратного проецирования поверхности (рис. 4).

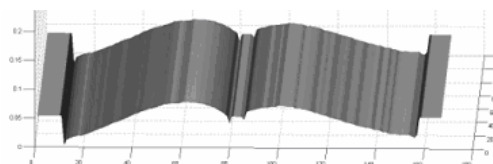


Рис. 4 Значения поверхности доопределены

Аппроксимация поверхности производится для каждой страницы отдельно. При этом предполагается, что Z-координата поверхности меняется только в направлении OX. Каждая страница аппроксимируется полиномом 3-й степени от переменной X. В результате получается гладкая поверхность (рис. 5).

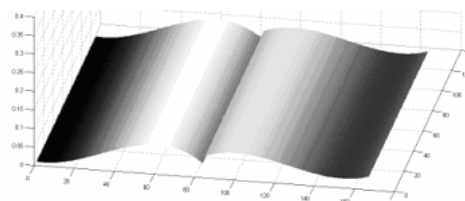


Рис. 5 Гладкая поверхность

Коррекция геометрии использует результат предыдущей стадии – полученную гладкую поверхность. На данном этапе не требуется откалиброванной камеры – сначала производится реконструкция поверхности книги, затем

рассчитывается повторная проекция объекта на экранную плоскость камеры с учётом полученной поверхности.

Введём систему координат с начальной точкой в центре книги. Оси OX и OY направим параллельно границам книги. Ось OZ направим так же, как направлена средняя нормаль к поверхности книги (перпендикулярно плоскости книги). Камера расположена в точке (0; 0; d). Через точку (0; 0; d-f) будет проходить экранная плоскость камеры. Здесь f – фокусное расстояние камеры. Константы d и f неизвестны.

Рассмотрим некоторую точку на экранной плоскости, принадлежащую области строки. Пусть её ордината равна y_1 . Она соответствует некоторой точке (Y; Z) на поверхности (рис. 6).

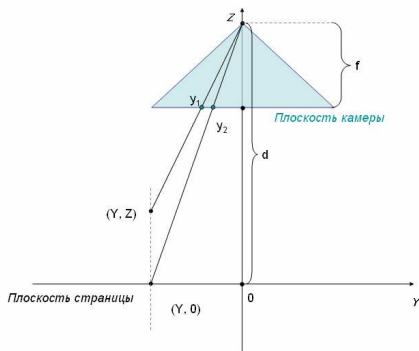


Рис. 6 Вычисление координат образа точки

В результате геометрической коррекции эта точка должна оказаться в точке (Y; 0). Тогда она будет соответствовать некоторой точке с ординатой y_2 на экранной плоскости.

Зависимость Z от y_1 и y_2 :

$$Z = d \cdot \left(\frac{y_1 - y_2}{y_1} \right) \quad (3)$$

Здесь мы видим, что фокусное расстояние f сократилось, но при этом осталась зависимость Z от d, которое не известно. Далее в алгоритме производится аппроксимация функции Z(X, Y) функцией $Z_c(X, Y)$. Очевидно, что при умножении Z на некоторую константу C получаем некоторую функцию Z' , в результате аппроксимации которой получается некоторая функция Z'_c , которая будет связана с функцией Z_c следующим образом:

$$Z' = C \cdot Z \quad (4); \quad Z'_c = C \cdot Z_c \quad (5);$$

То есть, множитель C сохраняется.



Рис. 7 Вычисление координат образа точки

Проведём усреднение значения координаты y по всей строке:

$$\bar{y} = \sum_{i=1}^n y_i, \text{ здесь } y_i - \text{координата } i\text{-ой точки из множества}$$

точек строки, n – количество точек в строке. $y'_1 = \bar{y} \cdot y'_1$ соответствует Z_c . Выразим y'_1 через y_2 :

$$y'_1 = \frac{y_2 \cdot d}{d - Z_c} = \frac{y_2}{1 - \frac{Z_c}{d}} \quad (6)$$

Из формул (3) и (6) и того факта, что при аппроксимации не теряется множитель d, получаем, что неизвестная константа d в предпоследней формуле сокращается. Таким образом, при вычислениях её можно не учитывать, т. е. положить равной единице. Наконец, выразим y_2 через известные величины Z и y_1 (рис. 7):

$$y_2 = y_1 \cdot (1 - Z) \quad (7)$$

своей статье "Уютный интерфейс" ("Компьютерра", № 42, 1998, <http://www.computerra.ru/online/1998/270/1778/>): "Красота программы, как и самолета, — не в лишних украшениях. Самолет, увешанный бантиками, с художественной ледкой и резными крыльями красного дерева не пролетит ста метров. Красота состоит в целесообразности всей конструкции самолета, в еще большей степени это относится к программам".

Рис. 8 Результат коррекции геометрии

На рис. 8 демонстрируются дефекты, сохраняющиеся после коррекции геометрии:

1. Строки на концах недостаточно прямые
2. Расстояние между строками непостоянно

На стадии постобработки производится окончательное выпрямление тех фрагментов, которые не удалось выпрямить на предыдущих этапах. При этом производится преобразование в плоскости изображения без использования какой-либо перспективной информации. Проблема, которую помогает решить это преобразование - это дефекты между страницами: резко загибающиеся строки, которые были подвержены искажению ещё и в направлении оси OX, а не только в вертикальном направлении.

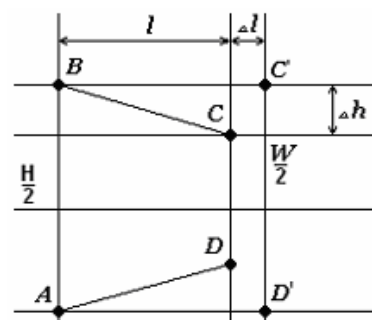


Рис. 9а Постобработка – линейное растяжение

Как показано на рис. 9, на этой стадии производится отображение трапеции ABCD (окрестности линии границы между страницами) на прямоугольник ABC'D'. При этом параметры преобразования (l, dl, dh) сейчас требуется задавать пользователю.

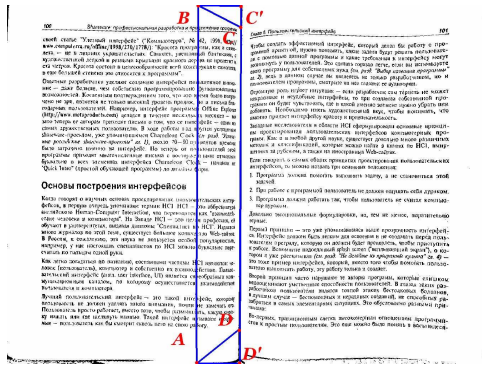


Рис. 9b Постобработка – линейное растяжение

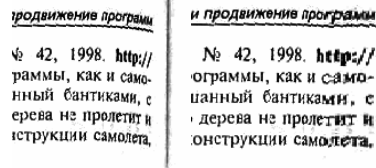


Рис. 10 Постобработка – фрагмент входного изображения и соответствующий фрагмент результата

На рис. 10 можно увидеть, что это преобразование помогает решить проблему загибающихся строк.

3. ЗАКЛЮЧЕНИЕ

Алгоритм был реализован на C++. Обработка одного изображения размерами порядка 1400x1000 занимает порядка 3 секунд (Athlon-64 3200+, 1GB RAM), качество распознавания заметно повышается после применения предложенного метода, причём результат оказывается немного лучше, чем результат метода, используемого в Fine Reader 9.0 ([9]). Тестирование производилось на 10 изображениях. В результате использования предложенного метода удалось правильно распознать 51% строк, в результате применения метода, встроенного в Fine Reader 9.0, 20%.

В настоящее время продолжается усовершенствование алгоритма по нескольким направлениям. Во-первых, планируется сделать метод более универсальным, а именно, отказаться от некоторых предположений, которые сужают область применимости метода (произвольные углы поворота страниц, аппроксимация более сложными функциями и другие). Во-вторых, усовершенствовать алгоритм постобработки, а именно, коррекцию тех искажений, которые остались после применения предыдущих шагов. В-третьих, метод имеет большое количество параметров (P_1 , P_2 , L_{max} , ...), значения которых зависят от разрешения изображения. Необходимо автоматическим образом определять значения таких параметров. И, наконец, требуется интегрировать реализацию метода в существующие программные средства.

4. ЛИТЕРАТУРА

[1] Владимир Вежневцев. “Задача восстановления 3D формы по закраске”, <http://cgm.graphicon.ru/content/view/59/62/>, 2004
 [2] Zheng Zhang, Chew Lim Tan. “Correcting Document Image Warping Based on Regression of Curved Text Lines”, *In Proc. ICDAR-2003*, pp.589-563, 2003

[3] Adrian Ulges, Christoph H. Lampert, Thomas M. Breuel. “Document Image Dewarping using Robust Estimation of Curled Text Lines”, *In. Proc. ICDAR-2005*, pp.1001-1005, 2005
 [4] Adrian Ulges, Christoph H. Lampert, Thomas M. Breuel. “Unwarping Images of Curved Documents Using Global Shape Optimization”, *In Proc. International workshop on Camera-based Document Analysis and Recognition*, pp.25-29, 2005
 [5] Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков “Численные Методы”, *Бином, Москва, 2003*
 [6] Zhixi Bian, Yan Zhang, “Retinex Image Enhancement Techniques”, *Course notes*, <http://library.graphicon.ru/paper/633>, 2000
 [7] Li Zhang, Chew-Lim Tan, “Warped Document Image Restoration Using Shape-From-Shading and Physically-based Modeling” *In Proc. WACV*, pp.29-32, 2007
 [8] Антон Масалович, Леонид Местецкий, “Использование патча Безье для аппроксимации искажения изображений текстовых документов”, *In Proc. Graphicon-2007*, pp.171-174, 2007
 [9] Fine Reader homepage, <http://www.abbyy.ru/finereader/>

5. ОБ АВТОРАХ

Зейналов Рамиз Шакирович – студент факультета вычислительной математики и кибернетики Московского Государственного Университета им. М. В. Ломоносова

E-Mail: ramiz_zeynalov@mail.ru

Document images geometric correction

Abstract

By reason of PDA and Smartphones with integrated digital cameras became widespread, people have started to use it to capture images from broadside of a book. Naturally the text on this-way captured images is suffered by both geometrical and perspective distortions because of non-planar geometry shape of the book broadside. The OCR methods couldn't get text from such images with satisfactory accuracy. In this article a new method of geometric correction of such images is described. Using this method the recognition accuracy is improved. This method is based on the surface shape estimation.

About the authors

Ramiz Zeynalov is student at Moscow State University, Faculty of Computational Mathematics and Cybernetics. Contact E-Mail: ramiz_zeynalov@mail.ru