

5. Вольмир А. С. Устойчивость деформируемых систем/ А.С. Вольмир. - М.: Наука, 1967. - 984 с.

Summary

Nikitenkov V. L., Koyushev P. I. Stability of a core in the environment with linearly changing rigidity (the decision by means of power series)

In [1],[2] problems of definition of behavior of critical loadings and borders of their change depending on rigidity comprehensive a core of the environment having constant value on all length of a core are considered. In this work rigidity of the environment in which the core is placed, changes on its length linearly. Values of critical loadings and a form of loss of stability (elastic lines) core for various boundary conditions are received at change of coefficient of growth of rigidity of the environment.

Keywords: core, stability, critical loading, rigidity of the environment, elastic line.

Сыктывкарский государственный университет

Поступила 31.05.2013

УДК 004.932

БИНАРИЗАЦИЯ И СЕГМЕНТАЦИЯ ОТСКАНИРОВАННОГО ТЕКСТА

В. Л. Никитенков, А. А. Поберий

В данной статье рассматриваются подходы и методы бинаризации и сегментации применительно к отсканированному тексту. Рассматриваемые методы модифицируются для достижения большей эффективности в распознавании отсканированного текста.

Ключевые слова: обработка отсканированного текста, сегментация текста, бинаризация изображений, сегментация рисунков.

Задачу распознавания любого текста можно разбить на несколько основных этапов: предобработка, сегментация, распознавание, постобработка. Первый этап, как правило, включает в себя обработку изображений с целью уменьшения шумов и бинаризацию — преобразование цветного изображения в чёрно-белое (ЧБ). Сегментация заключается в определении текстовых областей и их последующем дроблении на строки, слова и символы. Распознавание сводится к непосредственному распознаванию символов. В процессе постобработки корректируются ошибочные результаты распознавания. В данной работе в первую очередь рассматриваются этапы предобработки и сегментации отсканированного текста.

1. Бинаризация изображения

Важным этапом предобработки является процесс бинаризации — процесс преобразования цветного изображения в ЧБ. Процесс бинаризации проводится в два этапа: преобразование цветного изображения в серое, серого изображения в ЧБ. Перевод цветного изображения в серое не является проблематичным, значение серого цвета каждого пикселя

можно выразить через три цветовых составляющих соответствующего цветного пикселя: красную, зеленую и синюю (цветовая схема RGB):

$$grey = \frac{(66 \times R + 129 \times G + 25 \times B + 128)}{256} + 16, \quad (1.1)$$

где R , G , B — цветовые составляющие каждого пикселя — красная, зеленая, синяя соответственно.

Второй этап (преобразования серого в ЧБ) является более сложным. Данный этап состоит в выработке такой функции, которая для каждого серого пикселя (принимающего значения от 0 до 255) ставила бы в соответствие либо 0 (белое) либо 1 (черное). Для определения данного значения используется некоторый порог t , при сравнении с которым выбирается выходное значение.

Сложность задачи заключается в том, чтобы подобрать такое значение t , которое корректно отделит фон от текста и графики с наименьшей потерей информации и наименьшим количеством шума. До сих пор нет универсального алгоритма для решения данной задачи, но за последнее время появилось множество различных подходов в решении этой проблемы.

Существующие подходы условно можно разделить на подход глобальный (для всех пикселей один и тот же порог t) и адаптивный (порог t меняется в зависимости от определенных условий). В данной работе для бинаризации изображения будет использоваться адаптивный, а не глобальный порог, изображение будет условно разбиваться на отдельные области и для каждой области будет применяться свой порог. Глобальный порог неприемлем для сканированных изображений, т.к. освещение распределено неравномерно, т.е. некоторые области текста могут быть темнее или светлее, следовательно, значение порога t общее для всех пикселей приводит к потере информации.

Для поставленной задачи можно воспользоваться методом Отсу [2]. Основная идея этого метода заключается в том, что информация на изображении может быть только двух видов: фон или текстовая и графическая информация. Фон представляется светлыми цветами, а информация темной. Составляется частотная гистограмма цветов (от 0 до 255 включительно) с последующим перебором цветов, чтобы отделить фон от текста (см. рис. 1).

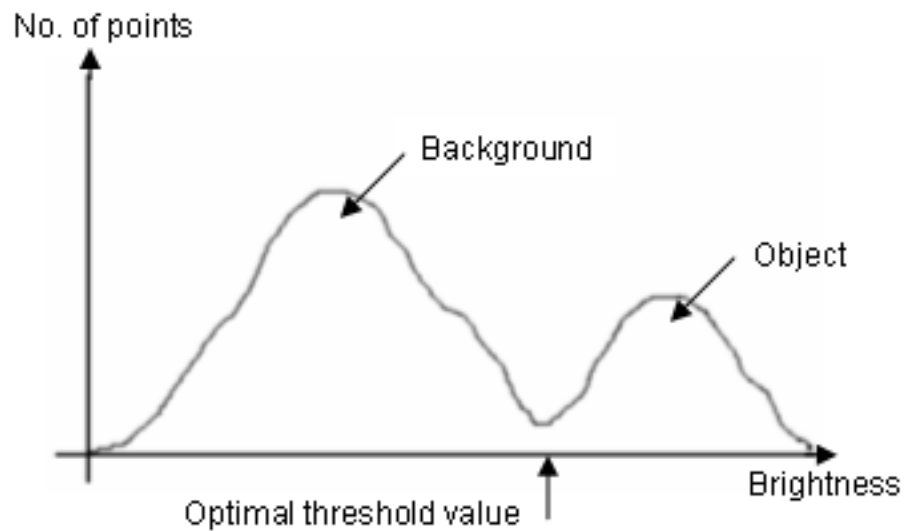


Рис. 1. Гистограмма цветов для нахождения разделяющего порога t .

Метод Отсу обычно используется для глобальной бинаризации. Применение метода Отсу при адаптивной бинаризации может давать шумовые значения для фоновых областей из-за наличия шумов в области изображения (см. рис. 2).



Рис. 2. Шумовые клетки при применении адаптивного метода Отсу.

Для исправления этого недостатка, после определения порога по методу Отсу можно проверить, является ли рассматриваемая область фоновой. Для этого достаточно вычислить дисперсию (сравнивая с некоторым фиксированным порогом); если она небольшая, то это скорее всего фон, в этом случае за порог t можно взять половину максимального значения $256 \times 0,5 = 128$. Описанная корректировка исправила полученные ранее ошибки (см. рис. 3).



Рис. 3. Исправленный адаптивный порог Отсу.

Данную проблему адаптивного метода Отсу также удалось решить разбиением изображения на вертикальные области. Преимущество данного подхода перед адаптивным заключается в том, что разбиение происходит только по одному из измерений (по ширине) и не требуется корректировка в каждой области для исправления описанной ранее ошибки адаптивной бинаризации Отсу (см. рис. 4). Идея данного подхода вытекает из наблюдения, что освещенность сканированного изображения меняется по горизонтали, темнее по краям и на стыке страниц и ярче посередине каждой страницы.



Рис. 4. Порог Отсу при разбиении на вертикальные области.

2. Сегментация

2.1. Нахождение областей. Существует множество методов и подходов к проблеме сегментации. В данной работе рассматривается подход, основанный на разбиении изображения на клетки с последующим анализом взаимной корреляции строк и математического ожидания каждой клетки [3]. Данный метод можно дополнить еще одним параметром — среднеквадратическим отклонением, который также способен характеризовать каждую клетку. Для сегментации серое изображение разбивается на клетки со сторонами, составляющими 1% от ширины изображения. Затем вычисляются статистические характеристики каждой клетки: среднеквадратическое отклонение σ , математическое ожидание μ и средняя взаимная корреляция всех строк клетки $corr$. По полученным характеристикам происходит классификация каждой клетки на фоновую или текстовую.

$$corr < 0,97; 0,03 < \mu \leq 0,85; 3 < \sigma < 9 \implies \text{текст} \quad (2.2)$$

Данные значения подбирались эмпирически [3]. Далее необходимо определить границы областей, для этого обходим изображение, применяя маски для каждого пикселя:

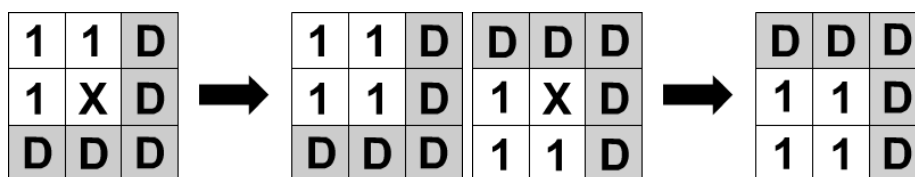


Рис. 5. Маски выделения текстовых областей для пикселей («0» — фон, «1» — текст, «D» — любое значение, «X» — обрабатываемый пиксель).

Обходим изображение в обратную сторону, применяя следующие маски:

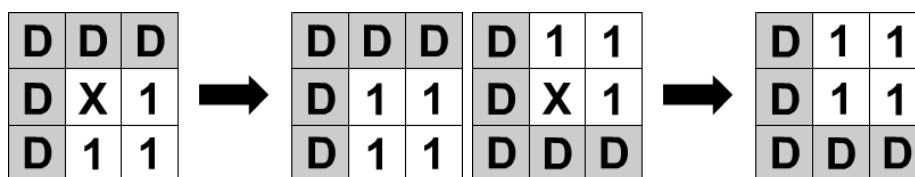


Рис. 6. Маски выделения текстовых областей для пикселей («0» — фон, «1» — текст, «D» — любое значение, «X» — обрабатываемый пиксель).

В результате, текстовые области получают углы, которые можно выделить как начало и конец прямоугольной области:

0	0	0	1	1	0
0	В	1	1	Е	0
0	1	1	0	0	0

Рис. 7. Маски для обозначения начала и конца области («В» — начало, «Е» — конец).

Данный метод помогает быстро находить границы текстовых областей, после чего можно переходить к следующему шагу сегментации — разделению строк. После нахождения текстовой области, ее необходимо разбить на строки, для этого можно воспользоваться статистическим анализом строк изображения [1]. Определяя долю чёрных пикселей каждой строки текстовой области, можно наблюдать колебания: возрастание при входе в текстовую область и убывание при выходе из неё. Опираясь на это наблюдение, можно получить строки. Пример полученных данным образом строк можно увидеть на рис. 8.

еский	интерес	к
ближенных	решен	
Размеры	плоской	

Рис. 8. Выделенные текстовые области строк.

2.2. Разделение слов. Каждую строку необходимо разбивать на слова. Для этого можно оценивать долю чёрных пикселей, но уже каждого столбца. Для разделения слов нужно находить интервалы с наименьшей долей чёрных пикселей, ширина которых соизмерима с шириной одного символа, т.к. интервалы меньшей ширины — это разделители символов, а не слов. При реализации данного метода лучше пробегать столбцы скользящим окном шириной больше одного пикселя, т.к. в одном столбце возможно получить значение меньше порогового из-за разрывов внутри символов, вызванных ошибками бинаризации или низким качеством изображения. Пример выделенных слов на рис. 9.



Рис. 9. Выделенные текстовые области слов.

2.3. Разделение символов. Чтобы отделить символы друг от друга, можно просматривать долю чёрных пикселей в скользящем окне, пробегающем по ширине слова, и если эта доля меньше порога, следовательно, мы попали на разделитель символов (пробел). Но у этих методов есть существенный недостаток — они не инвариантны к повороту изображения, так как все символы будут располагаться под углом и просмотр скользящим окном по ширине может выдать ложные результаты. Чтобы не зависеть от небольших углов наклона изображения, пороговое значение, с которым сравнивается доля черных пикселей, должно вычисляться для каждого случая отдельно, т.к. колебание доли чёрных пикселей при входе и выходе из текстовых областей сохраняется при изменении угла наклона. После сегментации можно переходить к непосредственному распознаванию символов (пример полученных символов см. рис. 10).



Рис. 10. Выделенные текстовые области символов.

2.4. Сегментация изображений. При работе с отсканированным текстом часто приходится иметь дело с графической информацией, т.е. с рисунками, графиками, схемами и т.п. Рисунки, как минимум, требуется убрать из рассмотрения, чтобы не мешать процессу распознавания. Для этой цели можно воспользоваться алгоритмом Блумберга [4], который основывается на мультиразмерном пороговом преобразовании и на методах морфологической обработки изображений (подробнее про алгоритм можно прочитать в [4], а про морфологическую обработку изображений можно подробно прочитать здесь [5]). Но, как упоминается в [6], данный метод подходит исключительно для полутоновых изображений, а для нашего случая желательно распознавать более широкий класс изображения. В упомянутой работе [6] предлагается в алгоритме Блумберга предварительно заливать неполутоновые изображения с помощью морфологической заливки.

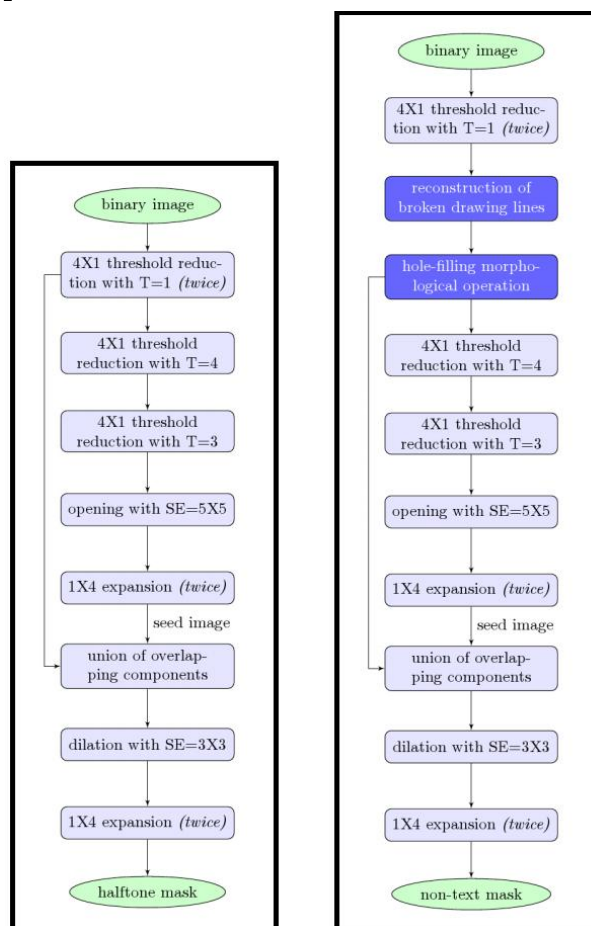


Рис. 11. Слева — алгоритм Блумберга, Справа — модифицированный алгоритм Блумберга с морфологической заливкой.

Морфологическая заливка имеет один существенный недостаток — она работает только для замкнутых областей, что в случае сканированного текста гарантироваться не может, поэтому требуется замкнуть эти области, перед заливкой, что значительно осложняет вычисления.

В [6] предлагается перед заливкой построить прерванные линии, этапы этого модифицированного метода Блумберга можно посмотреть на рис. 8 вместе с этапами самого алгоритма Блумберга.

Описанный выше подход достаточно эффективно справляется с поставленной задачей, но имеет высокую вычислительную емкость. В качестве альтернативы можно немного скорректировать исходный алгоритм Блумберга. Последнюю операцию дилатации («dilation with $SE=3 \times 3$ » см. рис. 8), можно провести раньше, сразу после шага морфологического размыкания («opening with $SE=5 \times 5$ ») и изменить размер структурного элемента (SE) на 5×5 вместо прежнего 3×3 . Эта модификация значительно расширит область, замеченных алгоритмом Блумберга, областей изображения в силу того, что операция расширения применяется для сжатого в 4 раза изображения, т.к. « 4×1 threshold reduction» применяется 4 раза (см. рис. 8.) и структурный элемент операции размерности 5, а не 3, полученный результат представлен на рис. 9.

Данный подход, выполняется немного быстрее самого алгоритма Блумберга (в силу того, что операция дилатация применяется не к конечному изображению, а к уменьшенному в 4 раза, следовательно и будет работать в 4 раза быстрее), но по-прежнему дает ошибки на рисунках с тонкими контурами, поэтому требует доработки, чтобы сравниться в эффективности с подходом морфологической заливки.

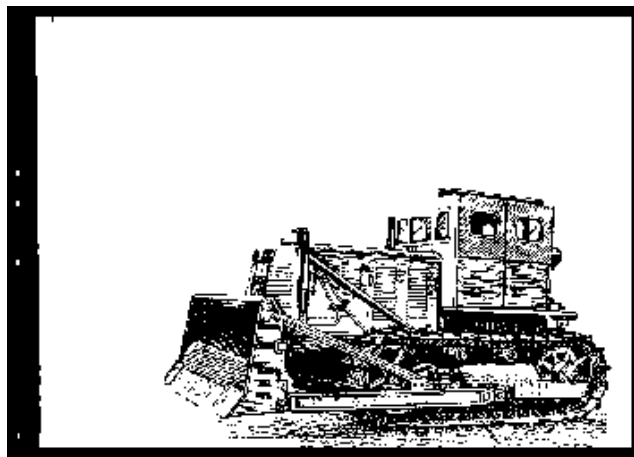


Рис. 12. Применение изменённого алгоритма Блумберга к рассматриваемому изображению.

Литература

1. **Paravassiliou V., Stafylakis T., Katsouros V., Garayannis G.** Handwritten image segmentation into text lines and words // *Pattern recognition*, 2010, №43, 369 - 377
2. **Otsu N.** A threshold selection method from grey-level histograms // *IEEE Trans., Man., Cyber.* 9(1): 62 - 66. *doi:10.1109/TSMC.1979.4310076*
3. **Sauvola J., Piettkainen M.** Page segmentation and classification using fast feature extraction and connectivity analysis, Int. Conf. on Document Analysis and Recognition, 1995.
4. **Bloomberg D. S.** Multiresolution morphological approach to document image analysis, Int. Conf. on Document Analysis and Recognition, 963-971, 1991.
5. **Гонсалес Р., Вудс Р.** Цифровая обработка изображений. -М. Техносфера, 2005.
6. **Bukhari S. S., Shafait F., Bruel T. M.** Improved document image segmentation algorithm using multiresolution morphology. German Research Center for Artificial Intelligence, 2010.

Summary

Nikitenkov V. L., Pobery A. A. Scanned text binarization and segmentation

Methods and approaches for binarization and segmentation for scanned text are considered in this article. Considered methods for scanned text recognition are modified to get more efficiency.

Keywords: scanned text processing, text segmentation, image binarization, picture segmentation.

Сыктывкарский государственный университет

Поступила 11.05.2013