

УЛУЧШЕНИЕ ФОТОГРАФИЙ ТЕКСТА: ОЦЕНКА РАЗМЕРА БУКВ И ПОИСК ЛИНИЙ СТРОК

асп. Семейкина Е. В.

Московский Государственный Университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, лаборатория математических методов обработки изображений (МГУ, ВМК)

1. Введение.

Эффективное распознавание рукописного и печатного текста является важной и активно развивающейся задачей компьютерного зрения. Особую сложность представляет распознавание текста при естественной съемке, условия которой не адаптированы для дальнейшего распознавания текста:

- возможно неравномерное освещение (рис. 1а, 1в),
- освещение может быть недостаточным, что приводит к слабому контрасту текста (рис. 1б, 1в),
- фон может быть неоднородным (рис. 1б),
- лист может быть ориентирован не перпендикулярно к оптической оси камеры (рис. 1в),
- лист может быть изогнутым, не представляющим собой плоскость (рис. 1г),
- фотография текста может содержать объекты, не относящиеся к тексту (рис. 1г).

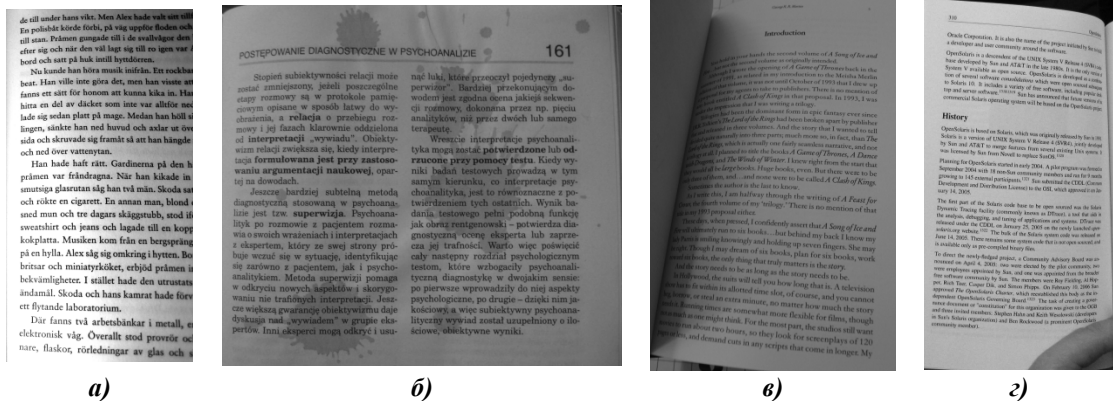


Рисунок 1. Примеры текста при естественной съемке.

В данной работе предложен алгоритм улучшения фотографий текста, являющийся необходимым предварительным этапом обработки в случае естественной съемки. Предложенный алгоритм позволяет:

- определить размер букв (высоту строк) и ширину линий, образующих буквы. Извлечение этой информации необходимо для выравнивания освещенности, локального повышения контраста и непосредственно самого распознавания символов.
- детектировать строки текста, что необходимо для выравнивания текста и дальнейшего распознавания.

2. Многомасштабное детектирование полос и определение размера букв.

2.1. Пространство переменных разрешений

Полосы на изображении могут быть описаны в терминах производных изображения, однако исходное изображение задано на дискретной сетке в плоскости (x, y) . Теория пространства переменных разрешений [1 - 3] предлагает использовать вместо дискретного изображения $I(x, y)$ его версию $L(x, y, t)$, сглаженную с помощью функции Гаусса $G(x, y, t)$:

$$L(x, y, t) = G(x, y, t) * I(x, y), \text{ где } G(x, y, t) = \frac{1}{2\pi t} e^{-\frac{(x^2 + y^2)}{2t}}. \quad (1)$$

Сглаженная функция $L(x, y, t)$ является дифференцируемой и по свойствам свертки её производные могут быть вычислены как свертки исходного дискретного изображения с соответствующими производными функции Гаусса. Таким образом, производные зависят от параметра сглаживания t , который в теории пространства переменных разрешений принято называть *масштабом*. Причем в случае детектирования полос и других особенностей на изображениях не существует одного априори заданного

масштаба, на котором должны вычисляться производные. Особенность должна детектироваться на масштабе, где она наиболее заметна. В случае полосы этот масштаб пропорционален её ширине, которая может быть переменной. Это приводит к рассмотрению однопараметрического семейства изображений $L(x, y, t)$ разной степени размытости. Важно отметить, что так как в качестве ядра размытия используется функция Гаусса, то $L(x, y, t)$ удовлетворяет диффузионному уравнению:

$$L_t = 0,5 \cdot (L_{xx} + L_{yy}). \quad (2)$$

2.2. Детектирование полос на фиксированном масштабе

Яркость изображения в окрестности точки (x, y) можно разложить в ряд Тейлора:

$$dL(x, y) = \vec{g}^T(x, y) \begin{pmatrix} dx \\ dy \end{pmatrix} + \frac{1}{2} \begin{pmatrix} dx \\ dy \end{pmatrix}^T \mathbf{H}(x, y) \begin{pmatrix} dx \\ dy \end{pmatrix} + \dots, \text{ где } \mathbf{H}(x, y) = \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} - \text{матрица Гессе.} \quad (3)$$

В серединной части особенности типа полоса первые производные малы и определяющим становится второе слагаемое. Интерпретируя яркость изображения как поверхность $L(x, y)$, можно сказать, что собственные значения матрицы Гессе λ_1, λ_2 ($|\lambda_1| \geq |\lambda_2|$) характеризуют кривизну этой поверхности в направлениях задаваемых собственными векторами $\vec{v}_1, \vec{v}_2$. Если одно собственное значение близко к нулю, а другое велико по модулю оба значения кривизны в точке имеют один знак и близкие по значению модули, то в этой точке присутствует особенность типа полоса [4].

Таким образом, детектирование *полос* состоит из следующих этапов [5], поясненных ниже:

- Формирование характеристического изображения $\lambda_1(x, y)$.
- Подавление на $\lambda_1(x, y)$ не максимальных значений [6] в направлении собственного вектора $\vec{v}_1(x, y)$

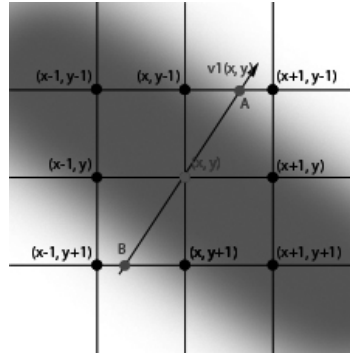


Рисунок 2. Точки A и B , соседствующие с текущей точкой в направлении собственного вектора $\vec{v}_1(x, y)$.

с) *Подавление ложных откликов.* Описанный детектор полос имеет отклик в отдельных точках, не лежащих на полосах. Для их подавления проверяется дополнительное условие: в точках полос градиент должен менять знак. Это условие проверяется в процессе подавления не максимальных значений (b) на изображении $\lambda_1(x, y)$: в точках A и B (рис. 2), соседствующих с текущей точкой в направлении собственного вектора $\vec{v}_1(x, y)$, кроме значений $\lambda_1(x, y)$, сравниваются ещё и значения градиента.

д) *Гистерезис.* Из полученных точек оставляются только те, которые соответствуют достаточно вытянутым особенностям: $0 < \lambda_2 / \lambda_1 \leq \beta$. К концам полос отношение λ_2 / λ_1 возрастает и перестает удовлетворять пороговому ограничению $0 < \lambda_2 / \lambda_1 \leq \beta$. Для того чтобы детектировать полосы полностью, ограничение это дополняется гистерезисным продолжением: сначала выделяются точки, удовлетворяющие $0 < \lambda_2 / \lambda_1 \leq \beta$, затем выделяются точки, примыкающие к ним и удовлетворяющие условию $\lambda_2 / \lambda_1 < 1$. В нашей реализации использовался порог $\beta = 8$.

2.3. Многомасштабное детектирование полос

Для многомасштабного детектирования полос естественным обобщением описанного в разделе 2.2 подхода является использование матрицы Гессе, учитывающей масштаб:

$$\mathbf{H}(x, y, t) = \begin{pmatrix} L_{xx} & L_{xy} & L_{xt} \\ L_{xy} & L_{yy} & L_{yt} \\ L_{xt} & L_{yt} & L_{tt} \end{pmatrix} \xrightarrow{\text{используя уравнение (2)}} = \begin{pmatrix} L_{xx} & L_{xy} & \frac{L_{xxx} + L_{xyy}}{2} \\ L_{xy} & L_{yy} & \frac{L_{xxy} + L_{yyy}}{2} \\ \frac{L_{xxx} + L_{xyy}}{2} & \frac{L_{xxy} + L_{yyy}}{2} & \frac{L_{xxxx} + 2L_{xxyy} + L_{yyyy}}{4} \end{pmatrix} \quad (4)$$

Для детектирования полос необходимо построить набор изображений $\lambda_1(x, y, t)$, используя при вычислении производных набор масштабов $\{t_i\}$, и выполнить процедуру подавления немаксимальных значений одновременно по двум взаимно перпендикулярным направлениям, которые определяются следующим образом. Квадратичная форма (4) задает трехмерный эллипсоид. Его сечение плоскостью перпендикулярной собственному вектору $\vec{v}_2$ матрицы Гессе (1) является эллипсом. Его полуоси выбираются в качестве направлений, по которым производится подавление немаксимальных значений на $\lambda_1(x, y, t)$. Направления выбираются именно таким образом, так как использование двух собственных векторов соответствующих наибольшим собственным значениям (4) может приводить к разрывам полос.

В случае многомасштабного детектирования процедуры немаксимального подавления и подавления отклика на границах производится на каждом масштабе. Затем на аккумулированных со всех масштабов полосах производится проверка условия $\lambda_1/\lambda_2 \geq \beta > 0$ и гистерезисное продолжение. При этом используются значения λ_1 и λ_2 , вычисленные на том же масштабе, на котором произошло детектирование в текущей точке.

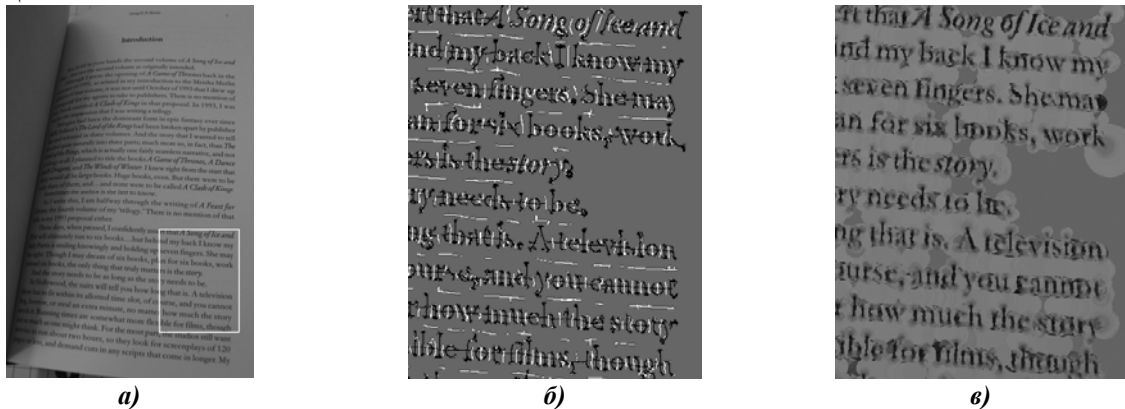


Рисунок 3. а) пример входного изображения, б) центры детектированных полос на увеличенном фрагменте изображения, центры светлых полос на темном фоне ($\lambda_1(x, y, t) < 0$) нарисованы белым, темных ($\lambda_1(x, y, t) > 0$) – черным, в) на однородном фоне в каждой точке, где была детектирована полоса, нарисован круг радиуса равного ширине найденной полосы цветом соответствующим найденной точке на исходном изображении, фрагмент.

На рис. 3 приведен пример многомасштабного детектирования полос. Визуализация на рис. 3в достаточно хорошо воспроизводит структуру текста, сохраняя его читаемость, что говорит о том, что большинство значимых линий детектировано.

3. Определение размера букв и выбор полос, являющихся строками текста.

Как видно на рис. 3 полосами являются не только строки и линии букв, но и светлые промежутки между строками. На темных линиях (с интенсивностью ниже, чем у фона) отклик $\lambda_1(x, y, t)$ детектора полос положителен, на светлых – отрицателен. Таким образом, для детектирования строк выбираются линии с положительным откликом.

На типичной фотографии текста, содержащей небольшое количество посторонних объектов, имеются два преобладающих масштаба: соответствующие размеру букв и ширине линий, образующих буквы. Для определения этих величин строится гистограмма масштабов, на которых были детектированы полосы, и находятся значения масштабов, при которых гистограмма имеет максимумы. Для большей стабильности нахождения преобладающих масштабов гистограмма предварительно сглаживается функцией Гаусса.

Для каждого найденного максимума находится диапазон масштабов, относящихся к этому максимуму, для этого находятся соседние минимумы на гистограмме масштабов (рис. 4б и 4ж). Для каждого преобладающего значения масштаба выбираются точки полос, детектированные на масштабах, лежащих в найденном диапазоне. Группы этих точек анализируются далее. Полосы, соответствующие строкам, имеют близкую ориентацию, в то время как полосы, образующие буквы, могут иметь произвольную ориентацию. Поэтому чтобы найти максимум на гистограмме масштабов, соответствующий строкам, анализируются направления собственных векторов $\vec{v}_1$, в найденных точках. Для выбранных точек строится гистограмма направлений собственных векторов $\vec{v}_1$ и вычисляется отношение m максимума гистограммы направлений к медиане. В случае строк медиана гистограммы близка к нулю, а типичные значения величины $m < 10$; для линий, образующих буквы, $m > 1000$, поэтому по порогу на величину m можно легко различить эти случаи.

Кроме того, используя построенную гистограмму направлений, можно отсеять полосы той же ширины, что строки текста, но с ориентацией далекой от максимума.

Схема изложенного выше алгоритма:

- 1) многомасштабное детектирование полос
- 2) выбор темных полос
- 3) построение гистограммы масштабов, на которых были детектированы полосы
- 4) сглаживание гистограммы масштабов
- 5) поиск максимумов на гистограмме масштабов
- 6) для каждого максимума:
 - a. выбор соответствующего диапазона масштабов между соседними минимумами
 - b. выбор точек, детектированных на масштабах, лежащих в найденном диапазоне
 - c. построение гистограммы направлений полос по выбранным точкам
 - d. вычисление величины m , для определения, соответствует ли рассматриваемый масштаб строкам

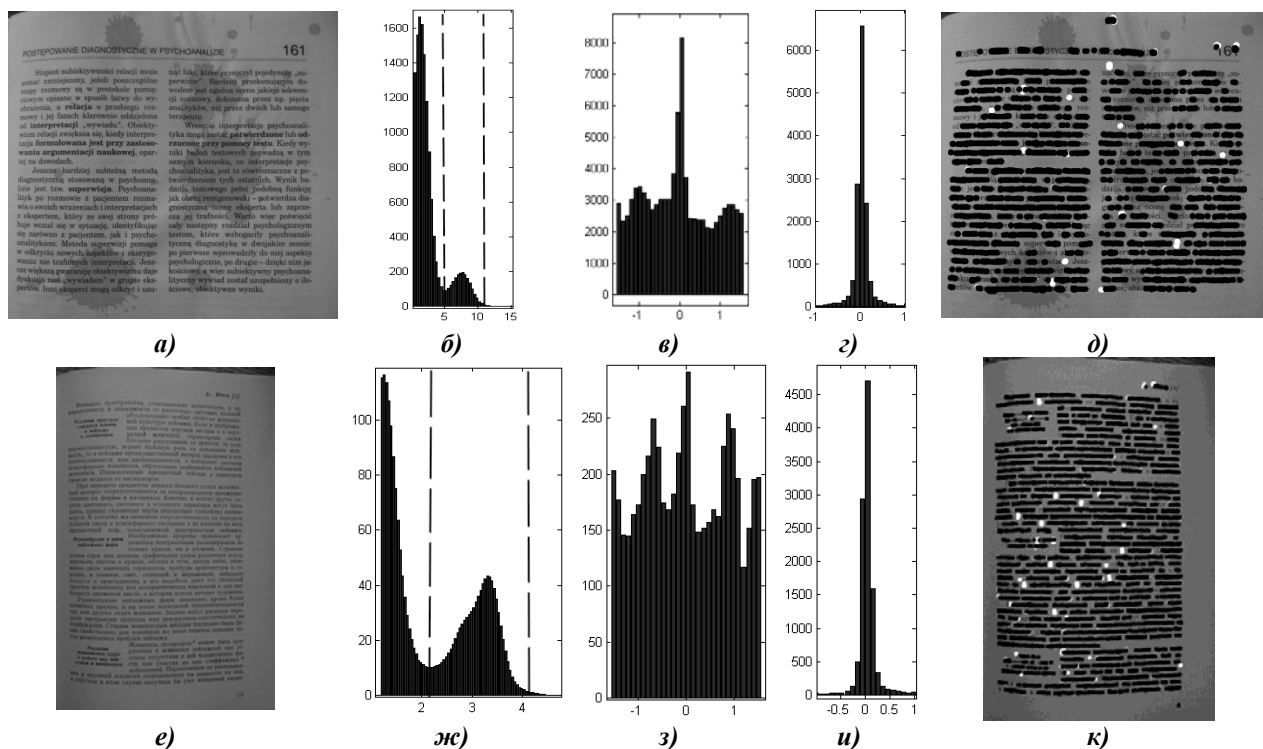


Рисунок 4. а, е) примеры изображений; б, ж) сглаженные гистограммы масштабов и выбранные диапазоны масштабов; в, г, з, и) гистограммы направлений полос, соответствующих максимумам на 4б и 4ж; д, к) полосы с шириной, близкой к найденной ширине строк, черным показаны полосы с направлением близким к преобладающему, белым – полосы с подходящей шириной, но отсеянные по направлению.

4. Благодарности

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы и гранта РФФИ 09-07-92000-ННС_а.

Литература

- [1] A. P. Witkin. *Scale-Space Filtering*. 8th International Joint Conf. on Artificial Intelligence, 1019-1022, 1983.
- [2] J. J. Koenderink, A. J. van Doorn. *The Structure of Images*. Biological Cybernetics, 50: 363-370, 1984.
- [3] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Dordrecht, 1994.
- [4] T. Lindeberg. *Edge Detection and Ridge Detection with Automatic Scale Selection*. CVPR, 30(2): 117-156, 1998.
- [5] N. A. Khanina, E. V. Semeikina, D. V. Yurin. *Color Blob and Line Detection in Scale-Space*. Pattern Recognition and Image Analysis, 21(2): 267-269, Pleiades Publishing, Ltd., 2011.
- [6] J. Canny. *A computational approach to edge detection*. IEEE Trans. PAMI. 8: 34-43, 1986.

Лаборатория математических методов обработки изображений

<http://imaging.cs.msu.ru>

Семейкина Е. В.

Email: esemeikina@graphics.cs.msu.ru

TEXT IMAGES ENHANCEMENT: LETTER SIZE ESTIMATION AND TEXT LINES DETECTION

PhD stud. Semeikina E. V.

Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics

Effective recognition of text is an important task of computer vision. The task is especially complex in the case of images which has not been made specially for text recognition. Such text images can be not uniformly illuminated, weakly illuminated, contain non-text objects, pages can be deformed. In the paper an algorithm for text image enhancement is proposed. The proposed algorithm allows to estimate high of letters and width of lines forming letters. This information can be useful for the further lightening alignment, local contrast enhancement and symbol recognition. Also the algorithm allows to detect text lines which can be used for page deformation compensation. Proposed method is based on multiscale ridge detection, result of which is lines forming letters, text lines and ridges of no interest. Further we select ridges corresponding to text lines via analysis of widths of detected lines and histograms of their orientation.