

УДК 004.934

С.Г. Антошук, д-р. техн. наук, проф.
Е.В. Ткаченко, магистр,
Одес. нац. политехн. ун-т

ПОМЕХОУСТОЙЧИВОЕ ВЫДЕЛЕНИЕ ТЕКСТОВОЙ ОБЛАСТИ НА ИЗОБРАЖЕНИИ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-ПРЕБРАЗОВАНИЯ

С.Г. Антошук, О.В. Ткаченко. Перешкодостійке виділення текстової області на зображенні з використанням вейвлет-перетворення. Представлено алгоритм перешкодостійкого виділення текстових областей на півтонових зображеннях з використанням вейвлет-перетворення. Показано можливість виділення текстових областей, що містять шрифти різних розмірів. Проведено порівняльний аналіз з базовим алгоритмом.

С.Г. Антошук, Е.В. Ткаченко. Помехоустойчивое выделение текстовой области на изображении с использованием вейвлет-преобразования. Представлен алгоритм помехоустойчивого выделения текстовых областей на полутоновых изображениях с использованием вейвлет-преобразования. Показана возможность выделения текстовых областей, содержащих шрифты различных размеров. Проведен сравнительный анализ с базовым алгоритмом.

S.G. Antoshchuk, E.V. Tkachenko, Anti-noise localization of text areas on pictures with the use of wavelet transform. The algorithm of anti-noise localization of text areas on gray-scale pictures with the use of wavelet transform is presented. Possibility of selection of the text areas, containing the fonts of different sizes is shown. A comparative analysis with a base algorithm is conducted.

Задача распознавания текстовых областей на полутоновых изображениях весьма актуальна при построении автоматических систем обработки визуальной информации, используемых для учета движущегося транспорта, продукции на конвейере и т.д. Одним из важных этапов решения такой задачи является обнаружение места расположения (выделения) текстовой области. Трудности, возникающие при этом, связаны с наличием помех на изображении, различными возможными размерами и формами надписи, загрязненностью объекта, различного угла обзора и т.д. [1]. Существует ряд задач, в которых нет априорной информации о структуре документа: плакаты, мультимедийные коллекции и т.д., поэтому необходим алгоритм, позволяющий избежать описанных трудностей и не требующий знания структуры документа.

Выделение текстовой области является частным случаем сегментации — одной из базовых процедур обработки изображений — и заключается в представлении последнего в виде совокупности областей, удовлетворяющих заданному критерию однородности. Применение сегментации позволяет существенно сократить объем обрабатываемой информации, т.к. границы сегментов чаще всего являются наиболее информативной частью изображения. Качество выполнения этой процедуры во многом определяет успех решения задачи распознавания и интерпретации объектов [2].

Традиционно существует три подхода к методам сегментации текстовых областей на странице документа: “сверху-вниз”, “снизу-вверх” и смешанный [3, 4].

При первом подходе документ рекурсивно делится на мелкие части. Самые распространенные алгоритмы: проекций [3]; гистограммного анализа; алгоритмы, основанные на правилах [4] и на пространственных преобразованиях (Фурье, Хо) [1, 2, 4]. Для них характерно высокое быстродействие и помехоустойчивость, однако эффективность напрямую зависит от априорных знаний о структуре и классе документа.

Второй подход заключается в уточнении элементов, объединении их в связные области и затем в смысловые структуры в соответствии с алгоритмами: математической морфологии [3],

наращивания областей [3, 4], на основе расчета градиента [5]. Алгоритмы не требуют знания предварительной структуры документа, однако обладают низкой помехоустойчивостью.

Изображение документа, содержащего текстовые области, может быть представлено набором фрагментов изображений, в которых повторяющимися элементами являются символы одного размера (рис. 1, а) при определении ее места нахождения. Надпись может быть рассмотрена как некоторая группа компактно расположенных символов, отличающихся пространственно-частотными свойствами от окружающей ее части изображения и от присутствующего на нем шума.

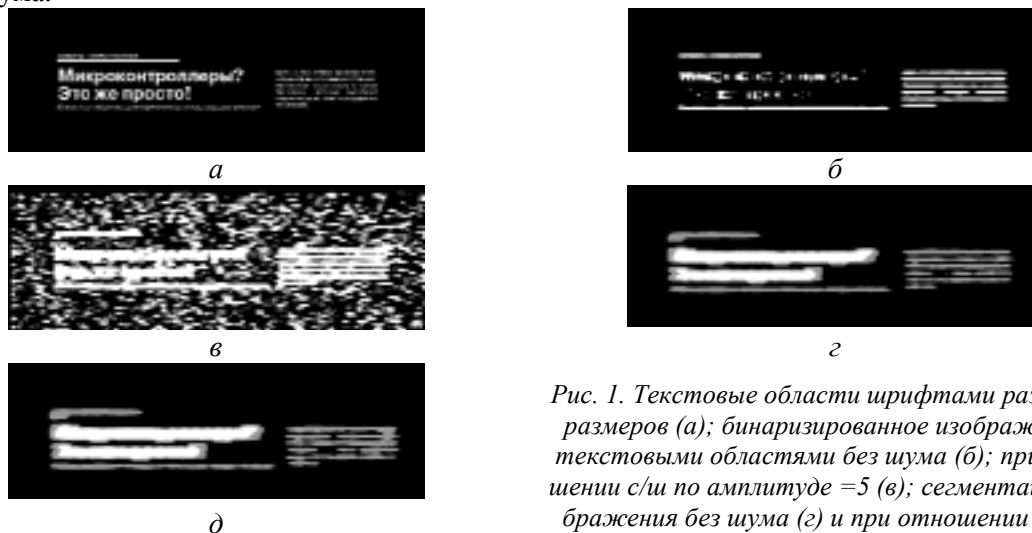


Рис. 1. Текстовые области шрифтами различных размеров (а); бинаризованное изображение с текстовыми областями без шума (б); при отношении с/ш по амплитуде = 5 (в); сегментация изображения без шума (г) и при отношении с/ш по амплитуде = 5 (д) с помощью модифицированного алгоритма

Однородные текстовые области характеризуются определенным уровнем вариаций градиента яркости в горизонтальном направлении, поэтому в качестве критерия сегментации можно использовать значение градиента. В качестве базового для выделения текстовой области выбран алгоритм [5]:

— подчеркиваются перепады яркости с помощью градиента по левой или правой границе с использованием маски $(-1, 1)$ (или $(1, -1)$)

$$\mathbf{G}(i, j) = \begin{cases} 0, & \text{если } \mathbf{I}(i, j+1) - \mathbf{I}(i, j) < 0; \\ \mathbf{I}(i, j+1) - \mathbf{I}(i, j), & \text{иначе,} \end{cases}$$

где в результате матрица $\mathbf{G}(i, j)$ представляет собой изображение, в котором подчеркнуты границы только перехода с фона на символ;

— вычисляется кумулятивная градиентная матрица

$$\mathbf{S}(i, j) = \sum_{k=j-\frac{w}{2}}^{k=j+\frac{w}{2}} \mathbf{G}(i, k), \quad (1)$$

текущее значение которой рассчитывается как сумма абсолютных значений строки матрицы $\mathbf{G}(i, j)$ длиной w , подбираемой эмпирически;

— проводится бинаризация полученного изображения пороговым методом [3]; выбор порога осуществляется на основе анализа гистограммы для матрицы $\mathbf{S}(i, j)$.

Результаты бинаризации показывают, что области разных шрифтов имеют различные неоднородности (рис. 1, б), и для коррекции результатов требуется применение дополнительной морфологической обработки.

При наличии даже незначительных шумов, например, гауссовского нормально распределенного шума с отношением с/ш по амплитуде $=5$) алгоритм становится неработоспособным (рис. 1, в).

Низкая помехоустойчивость алгоритма связана с использованием операции вычисления градиента для подчеркивания перепадов интенсивности.

Предлагается модификация алгоритма выделения текстовых областей изображения документа на основе расчета градиента, что позволит повысить помехоустойчивость выделения и обеспечить выделение текстовых областей шрифтов различных размеров без знания предварительной структуры документа.

Для повышения помехоустойчивости алгоритма сегментации [5] предлагается заменить градиент его оценкой. Для такой оценки могут быть использованы базисные действительные вейвлет-функции, заданные в виде нечетных симметричных функций $\psi(t) = -\psi(-t)$, имеющие компактный или эффективный носитель, в частности, вейвлеты Гаусса (рис. 2, а), гиперболическое вейвлет-преобразование и т.д [6, 7].

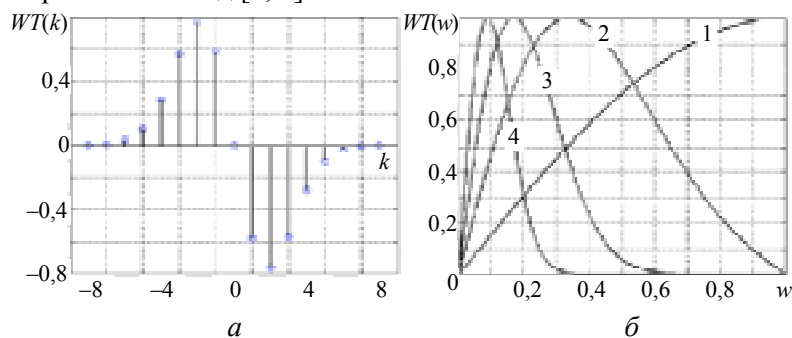


Рис. 2. Импульсная характеристика $WT(k)$ при масштабе $N=17$ (а) и АЧХ $WT(w)$ базисной вейвлет-функции Гаусса вида xe^{-x^2} при масштабах $N_1(1) < N_2(2) < N_3(3) < N_4(4)$ (б)

Использование вейвлет-функций на этапе подчеркивания перепадов яркости позволяет регулировать детализацию изображений объектов [8], мелкие детали подчеркивать с меньшей интенсивностью, крупные — с большей, т.к. вейвлет-преобразование обладает свойством пространственно-частотной локализации. Кроме того, применение вейвлет-преобразования позволяет повысить помехоустойчивость операций обработки изображений [1], которые часто подвержены влиянию мультипликативной помехи, например, условиям освещенности.

Выбранный в качестве вейвлет-функции вейвлет Гаусса $WT(k)$ не содержит колебаний дополнительных лепестков в амплитудно-частотных характеристиках (АЧХ) $WT(w)$ (рис. 2, б) и, таким образом, обладает свойством полосовой фильтрации, что позволяет согласовывать свойства преобразования со спектральными особенностями надписи и соответственно повышает помехоустойчивость алгоритма, а также прост в реализации и обладает высоким быстродействием. Анализ АЧХ базисных вейвлет-функций показал, что их частотные свойства (положение максимума АЧХ) зависят от масштаба N и с его увеличением смещаются в область низких частот [7]. Эффект “полосовой фильтрации” позволяет устранить влияние помех и подчеркнуть элементы символов шрифтов разных размеров с различной интенсивностью.

Для выделения надписи на автомобиле использовалось оконное 2×2 вейвлет-преобразование Хаара [9]. Преобразование производилось дважды. Первый раз обрабатывалось исходное изображение. Второй раз обрабатывалась матрица горизонтальных перепадов интенсивности. Таким образом, результирующая матрица была в 4 раза меньше исходного изображения. Данный алгоритм настроен на определенный размер шрифта и не позволяет выделять шрифты различных размеров.

Поставлена задача сегментирования шрифтов различных размеров и повышения помехоустойчивости при выделении текстовой области. Поэтому для выделения текстовых областей предлагается следующий модифицированный алгоритм:

— производится гомоморфная фильтрация с целью уменьшения мультипликативной помехи, в дальнейшем оценивается только влияние аддитивной помехи;

— производится подчеркивание перепадов яркости изображения вейвлетом Гаусса xe^{-x^2} определенного масштаба k (см. рисунок 2, а). Преобразование изображения осуществляется по строкам, в результате чего получается изображение с подчеркнутыми перепадами

$$\mathbf{IWT}(i, j) = \begin{cases} 0, & \text{если } \sum_{k=j-\frac{N}{2}}^{k=j+\frac{N}{2}} \mathbf{I}(i, k)WT(k) < 0; \\ \sum_{k=j-\frac{N}{2}}^{k=j+\frac{N}{2}} \mathbf{I}(i, k)WT(k), & \text{иначе,} \end{cases}$$

где $WT(k)$ — нормированная базисная вейвлет-функция Гаусса $1 \leq k \leq N$;

N — масштаб функции (количество отсчетов);

— рассчитывается кумулятивная сумма (см. алгоритм [5]):

$$\mathbf{S}(i, j) = \sum_{k=j-\frac{w}{2}}^{k=j+\frac{w}{2}} \mathbf{IWT}(i, k);$$

— проводится сегментация изображения по нескольким уровням интенсивности; для выбора порогов сегментации проводится гистограммный анализ матрицы $\mathbf{S}(i, j)$:

— построена гистограмма интенсивности $L(k)$ матрицы $\mathbf{S}(i, j)$ (рис. 3, а);

— проведена свертка $IW(k)$ значений гистограммы интенсивности $L(k)$ с базисной вейвлет-функцией Гаусса $WT(k)$;

— определены максимумы гистограммы интенсивности $L(k)$ как точки смены знака значений полученной свертки $IW(k)$ с положительного на отрицательный (рис. 3, б, точки 1 и 2);

— определены пороги $porog(k)$ сегментации как середина отрезков (0; 1) и (1; 2) (рисунок 3, б, точки 3 и 4).

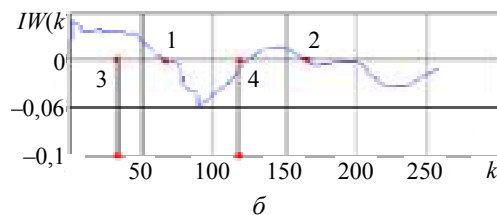
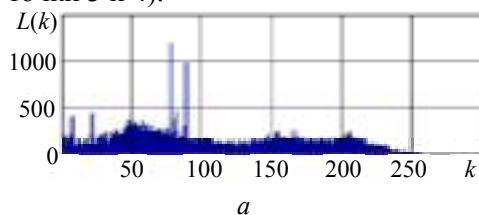


Рис. 3. Гистограмма интенсивности $L(k)$ изображения, полученного после вычисления кумулятивной суммы(2) (а); определение порогов $porog(k)$ сегментации (б)

Сегментация изображения проводилась с учетом яркости соседних пикселей в окне, при этом получено сегментирование изображения с выделенными текстовыми областями различной интенсивности

$$\mathbf{IB}(i, j) = \begin{cases} 0, & \text{если } \sum_{k=j-\frac{w}{2}}^{k=j+\frac{w}{2}} \mathbf{S}(i, j) < \text{porog}(k); \\ L(k), & \text{если } \sum_{k=j-\frac{w}{2}}^{k=j+\frac{w}{2}} \mathbf{S}(i, j) \geq \text{porog}(k), \end{cases}$$

где $L(k)$ — интенсивность;

$\text{porog}(k)$ — порог(и) сегментации, выбранные в результате гистограммного анализа.

Автоматизация выбора порогов сегментации позволила учитывать структуру документа, т.е. расположение надписей шрифтов разных масштабов, без предварительных априорных знаний о структуре документа.

Эксперименты показали, что значения элементов в матрице $\mathbf{S}(i, j)$ прямо пропорциональны размерам шрифта текстовой области.

На сегментации исходного изображения (см. рисунок 1) разному цвету соответствует шрифт разного размера: белому — крупный, серому — мелкий. Кроме того, по сравнению с сегментацией (см. рисунок 1, з), сегменты более однородны.

Проведено исследование на помехоустойчивость модифицированного алгоритма. При добавлении к изображению гауссовской нормальной помехи с соотношением с/ш по амплитуде =5, получено сегментированное изображение, практически аналогичное изображению при отсутствии шумов (см. рисунок 1, д):

Проведена сравнительная оценка помехоустойчивости базового и модифицированного алгоритмов.

Для оценки помехоустойчивости использовался показатель качества как мера близости (схожести) вектора изображения в d -мерном пространстве после сегментации с некоторым идеализированным (эталонным) вектором, который должен быть получен в результате идеального алгоритма. Размерность пространства $d = m \times n$, где m, n — соответственно число строк и столбцов изображения. Для оценки показателя качества предлагается использовать нормированное евклидово расстояние между изображением, полученным обработкой каждым из оцениваемых алгоритмов, и эталонным изображением

$$F^t = 1 - \frac{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (\mathbf{I}(i, j)^t - \mathbf{I}(i, j)^{\text{эт}})^2}}{mn},$$

где $\mathbf{I}(i, j)^t$ — изображение, полученное в результате обработки t -м из оцениваемых методов;

$\mathbf{I}(i, j)^{\text{эт}}$ — эталонное изображение.

Результаты оценки помехоустойчивости (рис. 4) показали, что модифицированный алгоритм работоспособен при отношении с/ш по амплитуде ≤ 2 для гауссовской нормальной помехи, при этом базовый градиентный алгоритм — при отношении с/ш по амплитуде > 10 ; при воздействии мультипликативной помехи модифицированный алгоритм показал приемлемое качество при с/ш по амплитуде ≤ 5 , а базовый при с/ш по амплитуде > 20 .

Данный алгоритм может найти широкое применение в ряде практически важных задач, в которых необходимо выделение символьной информации. Пример работы модифицированного алгоритма при выделении номера автотранспортного средства приведен на рис. 5.

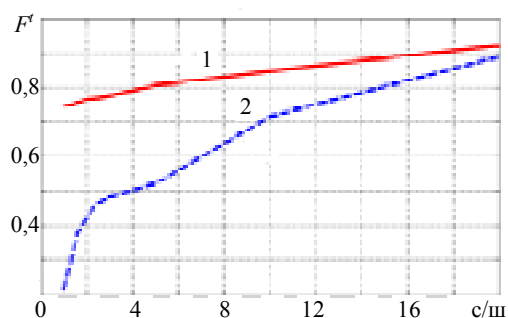


Рис. 4. Показатель качества F^t выделения текстовой области изображения при отношении $c/ш$ по амплитуде < 20 : 1 — модифицированный алгоритм; 2 — базовый алгоритм



Рис. 5. Реальное изображение с текстовой информацией (а), выделенная текстовая область с помощью модифицируемого алгоритма (б)

Модификация алгоритма выделения текстовых областей с использованием вейвлет-преобразований путем замены градиента на вейвлет-преобразование повысила помехоустойчивость процедуры выделения текстовой области. Автоматизация выбора порогов сегментации позволила учитывать размеры шрифтов без априорных знаний о структуре документа.

Алгоритм позволяет сегментировать на изображении текстовые области различных размеров. При отношении $c/ш = 2$ по амплитуде показатель качества F^t процедуры выделения текстовой области модифицированного алгоритма в два раза превысил показатель качества базового. Может быть рекомендован для решения широкого круга прикладных задач, где приходится выделять текстовую область на изображении для дальнейшего распознавания.

Литература

1. Методы компьютерной обработки изображений / Под ред. В.А. Сойфера. — 2-е изд., испр. — М.: Физмалит, 2003. — 784 с.
2. Анисимов, Б.В. Распознавание и цифровая обработка изображений: Учеб. пособие для вузов / Б.В. Анисимов, В.Д. Курганов, В.К. Злобин — М.: Высш. шк., 1983. — 295 с.
3. Syntactic segmentation and labeling of digitalized pages from technical journals / [M. Krishnamoorthy, G. Nagy, S. Seth, M. Viswanathan] // IEEE Computer Vision, Graphics and Image Processing. / — 1993. — Vol. 47. — P. 327 — 352
4. Serra, J. Image Analysis and Mathematical Morphology / Serra J. — New York: Academic Press, 1982. — Vol. 1. — 205 p.
5. Duong, J. Extraction of Text Areas in Printed Document Images pages / J. Duong, H. Emptoz, C. Y. Suen. — Atlanta, GA, USA, 2001. — P. 157 — 164.
6. Иерархический подход при контурном анализе изображений / [Антошук С.Г., Бабилунга О.Ю., Николенко А.А. Ткаченко Е.В.] // Искусств. интеллект. — 2005. — № 34. — С. 605 — 603.
7. Клих, Ю.А. Адаптивные базисные функции вейвлетного преобразования / Ю.А. Клих, С.Г. Антошук, А.А. Николенко // Тр. Одес. политехн. ун-та. — Одесса, 2004. — Вып. 2(22). — С. 121 — 125.
8. Антошук, С.Г. Метод морфологической обработки полутоновых изображений на основе гиперболического вейвлет-преобразования / С.Г. Антошук, О.Ю. Бабилунга., А.А. Николенко // Електромашинобудування та електрообладнання. — 2006. — Вип. 67. — С. 119 — 123.
9. Крылов, В.Н. Автоматизированная система распознавания / В.Н. Крылов, М.В. Максимов, В.О. Давыдов // Искусств. интеллект. — 2002. — № 4. — С. 462 — 469.
10. Оппенгейм, А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер — М.: Мир, 1979. — 416 с.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Крылов В.Н.

Поступила в редакцию 3 октября 2008 г.
