

РАСПОЗНАВАНИЕ ТЕКСТОВЫХ ИЗОБРАЖЕНИЙ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

Навойчик А.В., Гурин Н.И.

Белорусский государственный технологический университет, г.Минск, Республика Беларусь, a.navoichyk@hotmail.com

Распознавание текста является одним из направлений распознавания образов. В данном случае, под распознаванием понимается соотнесение изображения объекта, его образа, набора признаков самому объекту. Примерами распознавания текста являются:

1. Оцифровка изображений текста (сканированные книги, статьи, журналы) для последующей работы с его цифровым аналогом;
2. Обработка анкетных бланков;
3. Обработка систем тестирования знаний;
4. Распознавание номеров машин.

Задача распознавания текста остается актуальной на сегодняшний день, так как не существует стопроцентной универсальной системы по распознаванию текста. Система распознавания текста предполагает наличие на входе изображения с текстом (в формате данных графического файла). На выходе системы должен сформироваться текст, выделенный из этого изображения.

Распознавание текста включает в себя следующие этапы:

1. Поступающее на вход системы изображение должно быть очищено от шума и приведено к виду, позволяющему эффективно выделять символы и распознавать их;
2. Система должна разбить изображение на блоки текста, основываясь на особенностях его выравнивания и распределения по нескольким колонкам;
3. Изображение с текстом должно быть разделено на изображения строк, а затем на изображения символов для того, чтобы в дальнейшем обработать каждый символ по отдельности. После данного шага разные системы распознавания работают по своим специфическим алгоритмам;
4. Изображение символа может обрабатываться целиком, для этого оно сравнивается с имеющимися шаблонами, распознается при помощи нейронных сетей или выделяются характеристики изображаемого символа. На выходе четвертого шага появляется возможный вариант буквы. Однако обычно системы на этом не останавливаются и продолжают работу на основе других методов, уточняя полученный результат.

На первом этапе во всех системах оптического распознавания символов выполняются улучшение качества и анализ изображения, поданного на обработку. В том числе применяются специальные фильтры восстановления поврежденных изображений, например, с помощью гипоеллиптической диффузии [1], фильтры, устраняющие смазы [2].

На втором этапе выполняется работа по обнаружению текстов на изображениях. Известны следующие подходы: на основе использования контурной информации (скелетизация [3], выделение краев и выделение углов [4], методы на основе инвариантных моментов [5]); на основе цветовой информации (метод гистограмм [6], анализ главных компонент [7] и различные алгоритмы адаптивной бинаризации — в том числе алгоритмы Niblack, Sauvola, Chistian, Bernsan, Otsu и пр. [8]); на основе анализа текстурной информации (метод опорных векторов, искусственные нейронные сети [9], экспертные системы).

На следующем этапе выполняется сегментация текста. Будем проводить ее в три этапа:

1. Выделение строк - исходное изображение текста необходимо "разрезать" на полосы-строки нужной ширины;
2. Сегментация слов - в изображении текстовой строки выделяем изображения слов;
3. Сегментация символов - в изображении слова проводим границы символов.

Будем рассматривать изображение текста в градациях серого. Исходное изображение можно представить как матрицу яркостей точек B .

$$B = \{b_{ij}\}$$

$$0 \leq b_{ij} \leq b^{\max}$$

$$i = 1 \dots n, j = 1 \dots m$$

где n - ширина картинка, m - высота картинка.

Задача выделения строк сводится к нахождению верхних и нижних граней строк текста, изображённого на исходной картинке. Алгоритм сегментации строк основывается на том, что средняя яркость в изображениях межстрочных промежутков существенно ниже средней яркости в изображениях текстовых строк. Сначала для всех пиксельных строк исходного изображения находим их средние значения яркости.

$$s_j = s_j(B) = \frac{1}{n} \sum_{i=1}^n b_{ij}$$

Затем определяем среднее значение яркости всего изображения.

$$s(B) = \frac{1}{m} \sum_{j=1}^m s_j(B)$$

Средняя яркость в межстрочных промежутках текста должна быть невелика (в идеальном случае она равна нулю). Поэтому яркость верхней границы текстовой строки можно выразить через среднюю яркость изображения.

$$s^t = k^t s(B),$$

$$0 < k^t < 1$$

Яркость нижней границы текстовой строки также может быть выражена через среднюю яркость всего изображения.

$$s^b = k^b s(B),$$

$$0 < k^b < 1$$

Работа алгоритма сегментации строк заключается в последовательном просмотре массива средних значений (s_1, \dots, s_m) и выявлении множества пар индексов (s_i^t, s_i^b) пиксельных строк, соответствующих верхней s_i^t и нижней s_i^b граням изображения строки номер i , удовлетворяющих следующим условиям.

Условия верхней границы текстовой строки:

$$(s_{i-2} < s^t) \wedge (s_{i-1} < s^t) \wedge (s_i > s^b) \wedge (s_{i+1} > s^b) \wedge (s_{i+2} > s^b) \wedge (s_{i+3} > s^b).$$

Условия нижней границы текстовой строки:

$$((s_i > s^t) \wedge (s_{i+1} > s^t)) \vee ((s_{i+1} < s^b) \wedge (s_{i+2} < s^b) \wedge (s_{i+3} < s^b)).$$

В результате формируется множество пар индексов верхних и нижних граней строк. Разность между этими индексами дает высоты текстовых строк. Однако такой алгоритм находит среднюю высоту каждой текстовой строки и срезает символы, выступающие по высоте за эту среднюю высоту. Чтобы избежать этого, необходимо расширить найденные границы.

Для решения задачи сегментации слов понадобится изображение текстовой строки, полученной из исходного изображения документа после применения к нему алгоритма сегментации строк. Для улучшения качества работы алгоритма выделения слов из строки вначале его работы выполняются два преобразования входного изображения:

1. Пороговый фильтр повышения контрастности - такое преобразование, при правильно выбранном пороге, помогает снизить уровень шума, т.е. убрать значительное количество лишних точек;
2. Размазывающий фильтр – для каждой яркой (чёрной) точки исходного изображения закрашиваем соседние точки. В результате такого преобразования близкие точки объединяются в непрерывную область и вместо множества маленьких точек получаем изображение, состоящее из нескольких сплошных пятен с достаточно чёткой границей.

Алгоритм сегментации слов основывается на том, что средняя яркость в межсловных интервалах существенно ниже средней яркости в изображениях слов. Он похож на алгоритм сегментации строк, только просмотр идет по пиксельным столбцам изображения строки.

В большинстве изображений слов символы расположены близко друг к другу и межсимвольные интервалы не так ярко выражены, как в случае межстрочных или межсловных интервалов. Поэтому алгоритм сегментации символов сложнее и не так очевиден, как рассмотренные ранее алгоритмы сегментации строк и слов. Входом для алгоритма сегментации символов служит изображением слова, которое было получено из изображения текстовой строки после применения к нему алгоритма сегментации слов.

Алгоритм сегментации символов основывается на том, что средняя яркость в межсимвольных интервалах ниже средней яркости в изображениях символов. Общая схема состоит из двух основных частей.

1. Нахождение индексов столбцов, соответствующих локальным минимумам средней яркости столбцов c^i ;
2. Выявление и удаление из этого списка индексов ложных границ символов.

Поиск локальных минимумов средней яркости столбцов c^i происходит на смежных интервалах изменения индекса столбца. Размер интервала выбирается исходя из высоты строки. Для большинства шрифтов отношение ширины символа к его высоте не превышает величину 0.3. Поэтому размер интервала выбран:

$$d_j = 0.3m,$$

где m - высота слова в точках.

Поиск минимумов работает следующим образом:

1. Для всех пиксельных столбцов исходного изображения находим их средние значения яркости:

$$c_i = c_i(B) = \frac{1}{m} \sum_{j=1}^m b_{ij},$$

где m - высота слова в точках.

2. Среди значений c_i первый минимум ищем на отрезке $i = 1, \dots, d_j$;
3. Предположим, что он нашелся для индекса i_{\min}^1 ;
4. Следующий минимум ищем на отрезке $i = (i_{\min}^1 + 1), \dots, (i_{\min}^1 + 1 + d_j)$;
5. Процедура поиска повторяется до достижения границы ($i = n$) изображения слова. Все значения индекса i_{\min}^j , соответствующие локальным минимумам, сохраняются в списке W_0 .

Локальный минимум яркости в столбце номер i является кандидатом на принадлежность к межсимвольному интервалу, если значение средней яркости c_i в этом столбце меньше определённой границы яркости c_b и при этом значение средней яркости в столбцах отстоящих от данного локального минимума на 2 пикселя слева или справа больше границы яркости. Границу яркости можно определить через среднюю яркость картинки:

$$c^b = k^b \frac{1}{n} \sum_{i=1}^n c_i(B),$$

где $0 < k^b < 1$ коэффициент, n - ширина изображения слова в точках.

Условие межсимвольных границы выглядит следующим образом:

$$(c_i < c^b) \wedge ((c_{i-2} > c^b) \vee (c_{i+2} > c^b))$$

В результате из списка индексов локальных минимумов W_0 удаляются индексы столбцов, средняя яркость которых не удовлетворяет этому условию, формируется список W_1 индексов межсимвольных границ [10].

На последнем этапе, получив изображения символов, необходимо произвести их распознавание. Для этого будем использовать полносвязную искусственную нейронную сеть (рис. 1). Полносвязная искусственная нейронная сеть состоит из входного, выходного и скрытого слоев. Входной слой состоит из $n * t$ нейронов, где n – высота изображения, а t – ширина. Выходной слой состоит из k нейронов, где k – размер алфавита. Выход нейронов одного слоя является входом для следующего.

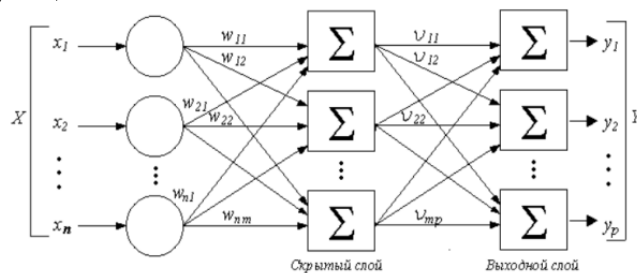


Рисунок 1 – Полносвязная нейронная сеть

Математически нейрон представляет собой взвешенный сумматор, единственный выход которого определяется через его входы и матрицу весов следующим образом:

$$y = f(u), u = \sum_{i=1}^n w_i x_i + w_0 x_0$$

Здесь x_i и w_i - соответственно сигналы на входах нейрона и веса входов, функция u называется индуцированным локальным полем, а $f(u)$ – функция активации. Возможные значения сигналов на входах нейрона считаются заданными в интервале $[0, 1]$. Они могут быть либо дискретными (0 или 1), либо аналоговыми. Дополнительный вход x_0 соответствующий ему вес w_0 используются для инициализации нейрона. Будем использовать сигмоидальную функцию активации. Она является монотонно возрастающей, всюду дифференцируемой, позволяет усилить слабые сигналы и не насыщаться от сильных сигналов.

Для корректной работы нейронную сеть необходимо обучить. Воспользуемся методом обратного распространения ошибок. Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы [11]. Для возможности применения метода обратного распространения ошибки функция активации нейронов должна быть дифференцируема. Метод является модификацией классического метода градиентного спуска. После обучения нейронной сети ее можно использовать для распознавания.

Предложенный алгоритм распознавания текста используется в разрабатываемом приложении для проверки результатов тестирования, проводимого в ходе учебных занятий.

ЛИТЕРАТУРА

1. U. Boscain, J.-P. Gauthier, A. O. Ремизов. Восстановление изображений с помощью гипоэллиптической диффузии, 2013, URL <http://gct.math.nsc.ru/wordpress/wp-content/uploads/2013/10/Remizov.pdf>.
2. X. Li, J. Jiaya. Two-phase kernel estimation for robust motion deblurring // ECCV 1. Lecture Notes in Computer Science, 2010. Vol. 6311, p. 157–170.
3. С. В. Погодин. Выделение и анализ скелетов объектов на цветных снимках // Программные продукты и системы, 2009. Т. 2, с. 42–45.
4. Л. Шапиро, Дж. Стокман. Компьютерное зрение. М.: Бином. Лаборатория знаний, 2006. — 752 с.
5. Н. С. Абрамов, В. М. Хачумов. Распознавание на основе инвариантных моментов // Вестник РУДН. Серия Математика. Информатика. Физика, 2014. Т. 2, с. 142–149.
6. А. Н. Виноградов, Ф. В. Калугин, М. Д. Недев. Выделение и распознавание локальных объектов на аэрокосмических снимках // Авиакосмическое приборостроение, 2007. Т. 9, с. 39–45.
7. Н. В. Завалишин, И. Б. Мучник. Модели зрительного восприятия и алгоритмы анализа изображений. М.: Наука, 1974. — 344 с.
8. Л. Шапиро, Дж. Стокман. Компьютерное зрение. М.: Бином. Лаборатория знаний, 2006. — 752 с.
9. А. А. Талалаев, И. П. Тищенко, М. В. Хачумов. Выделение и кластеризация текстовых и графических элементов на полутоновых снимках // Искусственный интеллект и принятие решений, 2008. Т. 3, с. 72–84.
10. Е. Борисов, Сегментация изображения текста [Электронный ресурс] <http://mechanoid.kiev.ua/cv-text-image-segmentator.html>.
11. Барцев С. И., Гилев С. Е., Охонин В. А., Принцип двойственности в организации адаптивных сетей обработки информации, В кн.: Динамика химических и биологических систем. — Новосибирск: Наука, 1989. — С. 6-55.