

РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ СИМВОЛОВ НА ОСНОВЕ ЛИНЕЙНОГО ОПИСАНИЯ СТРУКТУРНЫХ ХАРАКТЕРИСТИК

Предложен метод построения линейных описаний объекта на изображении. Описание строится на основе интерпретации структурных связей элементов объекта и обладает устойчивостью к изменениям масштаба и смещений объекта. Исследованы процедуры сравнения предложенных описаний, компьютерные эксперименты подтвердили эффективность подхода.

Ключевые слова: символ, блок, линейное описание, состояние, сравнение описаний.

Введение

Проблема распознавания текста остается одной из наиболее актуальных в области систем компьютерного зрения ввиду ее необычайной сложности. Существующие OCR-системы дают неплохие результаты распознавания текста, однако главным условием их успешного функционирования остается качество входного изображения. Наиболее качественные методы распознавания символов связаны обычно с использованием нейронных сетей при значительных временных затратах на их обучение. Использование же структурных и численных признаков, которые характеризуются простотой построения в реальных задачах остается недостаточно исследованным.

Статья посвящена разработке метода построения линейного описания изображения символа, которое является инвариантно-устойчивой характеристикой его структурных свойств. Описание создается с помощью операции сканирования значимых линий изображения и их последующей специальной обработки. Идея метода заключается в реализации «человекоподобного» распознавания, а именно – распознавания по форме.

Целью данной работы является разработка метода первичного распознавания на основе линейного инвариантного описания, который позволяет сократить множество эталонных изображений с последующим вторичным распознаванием. При этом необходимо достигнуть сохранения высокого качества первичной обработки за приемлемое время.

Линейное описание

Основу предлагаемого подхода составляет построение некоторого представления символа как цепочки состояний образующих его структурных элементов [1, 2]. Особенности построения таких

описаний и их сравнения должны обеспечить его устойчивость к изменениям масштаба и начертаний одного класса.

Рассмотрим изображение $I(x, y)$ как совокупность горизонтальных линий $L = \{l_1, l_2, \dots, l_n\}$ с некоторым шагом Δy . Блоком назовем последовательность пикселей одинаковой яркости, чередование которых образует линию. В рамках каждой из линий может быть выделено m блоков

$l = \bigcup_{i=1}^m b_i$. В дальнейшем под «блоком» будем

подразумевать лишь те из них, которые несут в рамках решаемой задачи смысловую нагрузку, а именно – блоки черных пикселей. Например, рис. 1 иллюстрирует линию изображения, которая состоит из четырех блоков.



Рис. 1 Определение блока

Значение m формируется исходя из практических соображений после анализа множества эталонных изображений. Например, для символов латинского алфавита количество подобных блоков в линиях колеблется от 1 до 4. Необходимо учитывать, что m может быть различным даже для одного и того же набора символов различных начертаний [3, 4].

Пусть $S^i = \{s_1, s_2, \dots, s_{k_i}\}$, $i = \overline{1, m}$ – множество состояний, которые могут образовывать комбинации блоков в каждой линии, k_i – количество возможных состояний во множестве S^i . Формирование блоками определенного состояния может быть определено как на основании анализа только текущего блока, так и с учетом предыдущих.

Например, единственный блок в линии может указывать на одно из состояний из множества $S^1 = \{ \text{"vertical line"}, \text{"horizontal line"}, \text{"diagonal line"} \}$. Возможный набор состояний может быть сгенерирован после автоматического анализа всех эталонных изображений и их группировки.

Основная сложность оценки подобных состояний для каждого из множеств S^i связана с тем, что необходимо обеспечить полный охват всех присутствующих состояний, а также взаимно-однозначное соответствие их последовательностей эталонному множеству.

Таким образом, каждая из линий l изображения отображается в одно из всех

возможных состояний: $l \Rightarrow s$, соответственно, множество всех линий формирует набор состояний

$L \Rightarrow C = \bigcup s$, который и образует его линейное описание.

Определим набор правил, которые будут формировать состояния. Блоки и линии характеризуются параметрами, определяющими их поведение и состояние линии в целом. В качестве таких параметров были выбраны их центры тяжести c и ширина w : $b = \{c, w\}$.

Собственно правила отображения совокупности блоков внутри конкретной линии в конкретное состояние представляют собой набор case-ситуаций, сформированных, исходя из анализа эталонного множества. Например, следующая ситуация для S^1 отображает линию l , состоящую из единственного блока b_1 в одно из состояний:

$$[w_1 > \delta W, b_1 = \{c_1, w_1\}, l = \{b_1\}] \Rightarrow l \rightarrow s_2,$$

где $\delta \in [0; 1]$ – некоторое пороговое значение,

W – ширина изображения, $s_2 \in S^1$.

Применение множества правил к каждой из линий L дает возможность сформировать линейное описание эталонного изображения в виде строки

набора состояний $L \Rightarrow C = \bigcup s$.

Построение профиля

Одновременно с линейным описанием для каждого из объектов в поле зрения строится его профиль, который представляет собой более общий аналог описания, а именно отображение

$L \Rightarrow P = \bigcup S$. Подобные профили не обладают свойством однозначного соответствия одному из эталонных классов, а могут соответствовать сразу нескольким классам. Такое отображение может быть использовано для вспомогательных целей,

например, для того, чтобы сделать процесс распознавания более «строгим».

Для обеспечения устойчивости профиля и линейного представления к искажениям масштаба выполняется сокращение повторяющихся значений.

На рис. 2 приведены примеры эталонных и тестовых изображений.

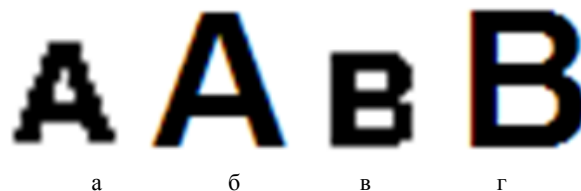


Рис. 2 Примеры эталонных изображений

Линейные описания для изображений на рис. 2, а и рис. 2, б соответственно имеют вид $C_1 = \{beae0bcba\}$, $C_2 = \{beae0bcefabc\}$, где символ

«а» соответствует реализации состояния «горизонтальная линия», «б» – «вертикальная линия», «с» – «наклонная линия», «е» – «единственная линия раздваивается». Значения до символа «0» характеризуют состояния при сканировании изображения горизонтальными линиями с продвижением сверху вниз, после символа «0» – при сканировании вертикальными линиями с продвижением слева направо. Отличия в вертикальном профиле связаны со значительными изменениями структуры символа по краям. Сокращенные комбинированные профили изображений совпадают и равны $P_1 = P_2 = \{1212121\}$. Изображения на рис. 2, в и рис. 2, г также имеют различия по линейному представлению $C_1 = \{agaga0aje\}$,

$C_2 = \{aegfaega0aje\}$ и одинаковые профили $P_1 = P_2 = \{121213212\}$.

Сравнение линейных представлений

Сравнение линейных представлений должно быть выполнено с учетом нескольких факторов. Основной проблемой в сравнении является переменная длина описаний вследствие различий по качеству для двух разных изображений одного и того же класса. Одним из наиболее простых методов сравнения двух описаний различной длины является расстояние Левенштейна [5].

Рассмотрим несколько других моделей сравнения предложенных линейных описаний.

Простое сравнение

Наиболее простым с точки зрения объема вычислений методом подсчета меры соответствия является манхэттэнское расстояние между описаниями согласно следующему выражению:

$$\text{distance}(C_1, C_2) = \sum_{i=1}^N \sum_{j=1}^M \rho(s_i, s_j), \quad s_i \in C_1, \quad s_j \in C_2,$$

$$\rho(s_i, s_j) = \begin{cases} 1/N, & s_i = s_j, \\ 0, & s_i \neq s_j. \end{cases}$$

где C_1, C_2 – описания, N, M – длины описаний соответственно, $N \geq M$. Существенным нюансом является отбрасывание всей предыдущей части каждого из фрагментов описания при нахождении совпадающих элементов.

Сравнение с подтверждением

Недостатком простого сравнения является чувствительность к существенному искажению структуры описания. В этом случае предложено использование подтверждающей процедуры, которое заключается в сравнении линейных описаний и их профилей. Описания при этом считаются соответствующими только при условии отображения профиля в правильный эталонный класс:

$$\text{profile}(P_1, P_2) = \sum_{i=1}^N \sum_{j=1}^M \rho(S^i, S^j),$$

$$\rho(S^i, S^j) = \begin{cases} 1/N, & S^i = S^j, \\ 0, & S^i \neq S^j. \end{cases}$$

Суммарное сравнение

Этот метод сравнения заключается в учете метрики профиля и линейного описания одновременно. Степень доверия профилю регулируется параметром ξ , исходя из его структуры и неоднозначности, логичным видится установка этого коэффициента с сохранением условия $\xi < 1$, конкретно при проведении экспериментов было выбрано значение $\xi = 0.5$:

$$\text{distance}(C_1, C_2) = \sum_{i=1}^N \sum_{j=1}^M \rho(s_i, s_j) + \xi \times \text{profile}(P_1, P_2), \quad s_i \in C_1, \quad s_j \in C_2,$$

$$\rho(s_i, s_j) = \begin{cases} 1/N, & s_i = s_j, \\ 0, & s_i \neq s_j. \end{cases}$$

Сравнение с голосованием

Возможность накопления измерений во многих практических задачах позволяет принять более уверенное решение, обоснованное применением статистических агрегатных функций.

В текущей модели голосование может быть записано в виде:

$$\max(\text{distance}(C_1, C_2)) \rightarrow \{e_1^d, e_2^d, \dots, e_{N_1}^d\},$$

$$\max(\text{profile}(P_1, P_2)) \rightarrow \{e_1^p, e_2^p, \dots, e_{N_2}^p\},$$

$$\text{voting}(C_1, C_2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \rho(e_i^p, e_j^d),$$

$$\rho(e_i^p, e_j^d) = \begin{cases} 1, & e_i^p = e_j^d, \\ 0, & e_i^p \neq e_j^d, \end{cases}$$

где e^d, e^p – метки эталонного класса как результат распознавания с использованием линейного описания и профиля соответственно, N_1, N_2 – количество наиболее вероятных эталонных классов при распознавании по линейному описанию и профилю соответственно.

Экспериментальные исследования

Эталонное множество при экспериментальных исследованиях включало 442 изображения шрифта Agial в верхнем и нижнем регистрах полужирного начертания разных размеров, а также цифры, формируя таким образом 62 эталонных класса.

Тестовое множество включало 1590 изображений символов различного размера шрифта Agial, а также нескольких других шрифтов без засечек, аналогичных Agial. Символы были получены из синтетических и реальных изображений текста, в том числе и после сканирования текста с разным качеством.

Распознавание было построено на основе сокращенного варианта линейного описания символа. Кроме того, была выполнена ручная коррекция результатов для тех ситуаций, когда невозможно определить символ без контекста его использования, например, для букв «l» и «I», которые для шрифта Agial идентичны.

Правильным первичным распознаванием назовем такое распознавание, при котором правильный эталонный образ находится в α результатах выдачи, которые соответствуют входному изображению с наибольшей вероятностью.

Таблица 1 иллюстрирует вероятности правильного первичного распознавания при разных значениях α при сравнении по расстоянию Левенштейна. Можно говорить о том, что предложенный метод с использованием описаний с вероятностью около 0,95 выдал номер правильного эталонного класса среди первых 10 результатов.

Таблица 1

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$	$\alpha = 10$
Успешно распознано	0,78	0,86	0,88	0,92	0,93	0,95

В табл. 2 сведены результаты первичного распознавания для всех остальных методов, рассмотренных выше.

Таблица 2

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$	$\alpha = 10$
Голосование	0,55	0,62	0,70	0,79	0,84	0,87
Суммарное сравнение	0,59	0,66	0,70	0,78	0,82	0,86
Суммарное голосование	0,51	0,62	0,70	0,77	0,81	0,86

Время вычисления линейных описаний и профилей для всех эталонных изображений составило около 3 с. (E8200, 2x 2.66 ГГц, 2Гб ОЗУ). На рис. 3 показано время первичного распознавания отдельного символа в зависимости от размера изображения.

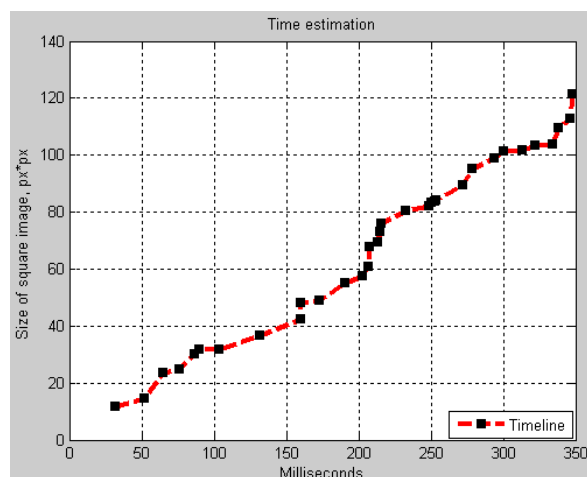


Рис. 3 Зависимость времени первичного распознавания от размера изображения символа

Представление символа изображения в виде строкового описания, которое отображает структуру символа, дает возможность быстро выполнить первичное распознавание и выделить подмножество классов в эталонном множестве с целью последующего распознавания.

Выводы

Исследования посвящены разработке метода, который позволяет выполнить первичное распознавание изображения по его линейному описанию. Отображение изображения в линейное представление достигается путем структурного анализа символа изображения с помощью сканирования линиями. Предложенное представление сочетает простоту построения обычных количественных характеристик с анализом структурных особенностей изображения и

инвариантностью к искажениям масштаба и смещений.

Научной новизной работы является метод построения линейного описания и профиля изображения, а также процедур их сравнения, который дает возможность сократить пространство эталонных изображений без существенных вычислительных затрат, осуществив тем самым первичное распознавание.

В качестве дальнейших исследований можно рассматривать пути улучшения метода построения описания, например, с помощью анализа состояний не только текущей линии, но и нескольких предыдущих с их корректировкой.

Список литературы

1. *Imaged document text retrieval without OCR* / Chew Lim Tan, Weihua Huang, Zhaohui Yu, Yi Xu // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jun 2002.– Vol. 24, Issue 6. – P. 838 – 844.– ISSN: 0162-8828.
2. *Chen, F.R. Detection and Location of Multi-Character Sequences in Lines of Imaged Text* / F.R. Chen, D.S. Bloomberg, L.D. Wilcox // *Journal of Electronic Imaging*.– 1996.– Vol. 5.– P. 37-49.
3. *Different types of fonts* [Электронный ресурс] / Режим доступа: <http://www.newbiewebdesign.com/different-types-of-fonts>.– 15.01.2012.
4. *The Classification of Fonts* [Электронный ресурс] / Режим доступа: <http://speckyboy.com/2010/11/02/the-classification-of-fonts/>
5. Левенштейн В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов / В. И. Левенштейн. – Доклады Академии Наук СССР, 1965. – № 163.4. – С. 845-848.

Рецензент: д-р техн. наук, проф. Пуятин Е.П., Харьковский национальный университет радиотехники

Авторы:

Гороховатский Алексей Владимирович

Харьковский национальный экономический университет, Харьков, кандидат технических наук, старший преподаватель кафедры информатики и компьютерной техники.

Тел. 758-77-10 (4-38), E-mail: gorohovatsky@rambler.ru

Передрий Елена Олеговна

Харьковский национальный университет радиотехники, Харьков, преподаватель кафедры информатики и компьютерной техники.

Тел. 758-77-10 (4-38), E-mail: elena_peredriy@ukr.net

Розпізнавання зображень символів на основі лінійного опису структурних характеристик

О.В. Гороховатський, О.О. Передрій

Запропоновано метод побудови лінійних описів об'єкту на зображенні. Опис будується на основі інтерпретації структурних зв'язків елементів об'єкту та є стійким до перетворень масштабування та зміщення об'єкту. Досліджено процедури порівнянь описів, комп'ютерні експерименти підтверджують ефективність запропонованого підходу.

Ключові слова: *символ, блок, лінійний опис, стан, порівняння описів.*

Recognition of images of symbols based on linear description of structural properties

O.V. Gorokhovatskyi, O.O. Peredrii

Method for the construction of linear descriptions of an object on image is proposed. Description is based on an object's structural elements interpretation and has resistance to scale and move changes. Comparison procedures for descriptions are investigated, computer experiments confirmed the effectiveness of the proposed approach.

Keywords: *symbol, block, linear representation, state, descriptors comparison.*