

МЕТОД РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЗНАКОВ

Задача распознавания текста – широко исследуемая область искусственного интеллекта. По этой теме выпущено достаточное количество публикаций [1,2,3,4], создано множество программных продуктов.

Чаще всего требующий распознавания образ представлен в виде растрового изображения (массива точек).

Существует три традиционных метода распознавания: шаблонное, признаковое и структурное. В шаблонном методе исходное изображение символа преобразуется в набор точек, а затем накладываются на него шаблоны, имеющиеся в базе системы. Шаблон с наименьшим отличием и будет искомым. Эти системы обладают высокой точностью распознавания дефектных (склеенных или разорванных) символов. Недостаток – невозможность распознать шрифт, хоть немного отличающийся от заложенного с систему. Признаковый метод предусматривает для каждого символа вычислять набор признаков и сравнивать эти наборы. Так как набор признаков только в идеальном случае полностью соответствует объекту, то заведомо часть информации о символе теряется. Структурный метод хранит информацию о топологии символа. Этот способ имеет недостатки, если исследуемая буква имеет дефекты, делающие ее похожей на другие буквы. Существует оригинальный метод, в котором знания о букве представлены в виде структурно-пятненного эталона. Изображение представляется в виде пятен, связанных парными отношениями. Технология распознавания с помощью структурно-пятненных эталонов получила название “фонтанное преобразование”. Фонтанное преобразование совмещает в себе достоинства шаблонной и структурной систем и позволяет избежать недостатков, присущих каждой из них по отдельности (Курс лекций по дисциплине "Системы искусственного интеллекта" <http://www.marstu.mari.ru/mmlab/home/AI/index.html>). Однако недостатком всех методов является ограниченность их применения для растрового изображения.

Существует множество случаев, когда исходная информация о знаке представлена в виде массива линий. Например, при использовании карманных компьютеров (Palm), планшетов, при вводе световым пером и т.п. Именно для таких задач применим метод изложенный ниже.

Исследуемый знак представлен в виде одномерного массива линий

$$S = \{L_1, L_2, \dots, L_n\}, \quad (1)$$

где L_1, L_2, \dots, L_n – линии по порядку введенные пользователем

© Ю.П. Юрченко, В.О. Брага, 2002

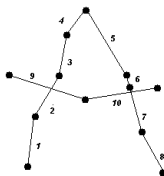


Рис. 1 – Буква “А” представлена в виде множества линий

Например, на рисунке изображена буква “А”, введенная пользователем и представленная в виде множества линий. Формула 1 принимает вид

$$S_A = \{L_1, L_2, \dots, L_{10}\}, \tag{2}$$

где L_1, L_2, \dots, L_{10} – линии образующие введенную букву “А”

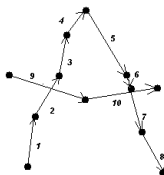


Рис. 2 – Буква “А” представлена в виде множества векторов

Каждая линия характеризуется направлением написания. Исходя из этого, можно представить букву “А” в виде множества векторов. Каждый вектор обладает такими характеристиками: направлением, длиной, координатами точки начала вектора и точки конца вектора. Тогда изображенная на рисунке буква “А” состоит из множества векторов и формула (2) принимает вид:

$$S_A = \{ \vec{L}_1, \vec{L}_2, \dots, \vec{L}_{10} \}, \tag{3}$$

где $\vec{L}_1, \vec{L}_2, \dots, \vec{L}_{10}$ — векторы, образующие букву “А” (рисунок 2).

Направление векторов определяется следующим образом.

Вектор имеет направление вправо (назовем это направление $D1$ – от английского слова direction (направление), если угол между вектором и осью $X > 345^\circ$ и $< 15^\circ$.

Вектор имеет направление вправо и вверх ($D2$), если угол между вектором и осью $X > 15^\circ$ и $< 60^\circ$.

Вектор имеет направление вверх ($D3$), если угол между вектором и осью $X > 60^\circ$ и $< 105^\circ$.

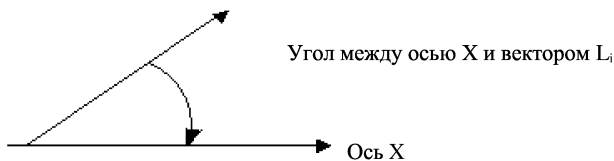


Рис. 3 – Методика определения направления вектора

Вектор имеет направление влево и вверх (*D4*), если угол между вектором и осью $X > 105^\circ$ и $< 165^\circ$.

Вектор имеет направление влево (*D5*), если угол между вектором и осью $X > 165^\circ$ и $< 195^\circ$.

Вектор имеет направление влево и вниз (*D6*), если угол между вектором и осью $X > 195^\circ$ и $< 255^\circ$.

Вектор имеет направление вниз (*D7*), если угол между вектором и осью $X > 255^\circ$ и $< 285^\circ$.

Вектор имеет направление вправо и вниз (*D8*), если угол между вектором и осью $X > 285^\circ$ и $< 345^\circ$.

Длина вектора определяется по формуле

$$Z = \sqrt{V_x^2 + V_y^2}, \tag{4}$$

где Z – длина вектора, V_x – проекция вектора на ось X , V_y – проекция вектора на ось Y (Рисунок 4).

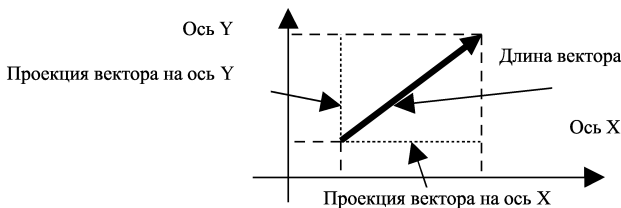


Рис. 4 – Определение длины вектора

Для распознавания знака требуется выделить область знака. Она определяется исходя из координат векторов: $MaxX$ – максимальное значение координаты вектора, образующего знак по оси X , $MinX$ – минимальное значение координаты вектора, образующего знак, по оси X , $MaxY$ – максимальное значение координаты вектора, образующего знак, по оси Y , $MinY$ – минимальное значение координаты вектора, образующего знак, по оси Y . Для определения координат начала и конца вектора разобьем область исследуемого знака на 9 областей, как это показано на рисунке 5.

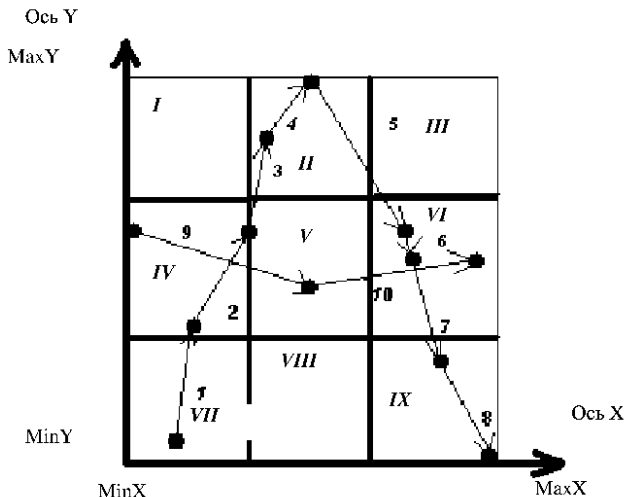


Рис. 5 – Разбиение знака на области

Соответственно, принадлежность определенной области начала или конца вектора определяется по формулам (5 — 13):
 принадлежит области I, если

$$MinX < X < \frac{1}{3} * MaxX \quad \text{и} \quad \frac{2}{3} * MaxY < Y < MaxY, \quad (5)$$

принадлежит области II, если

$$\frac{1}{3} * MaxX < X < \frac{2}{3} * MaxX \quad \text{и} \quad \frac{2}{3} * MaxY < Y < MaxY, \quad (6)$$

принадлежит области III, если

$$\frac{2}{3} * MaxX < X < MaxX \quad \text{и} \quad \frac{2}{3} * MaxY < Y < MaxY, \quad (7)$$

принадлежит области IV, если

$$MinX < X < \frac{1}{3} * MaxX \quad \text{и} \quad \frac{1}{3} * MaxY < Y < \frac{2}{3} * MaxY, \quad (8)$$

принадлежит области V, если

$$\frac{1}{3} * MaxX < X < \frac{2}{3} * MaxX \quad \text{и} \quad \frac{1}{3} * MaxY < Y < \frac{2}{3} * MaxY, \quad (9)$$

принадлежит области VI, если

$$\frac{2}{3} * MaxX < X < MaxX \quad \text{и} \quad \frac{1}{3} * MaxY < Y < \frac{2}{3} * MaxY, \quad (10)$$

принадлежит области VII, если

$$\text{Min}X < X < \frac{1}{3} * \text{Max}X \quad \text{и} \quad \text{Min}Y < Y < \frac{1}{3} * \text{Max}Y, \quad (11)$$

принадлежит области VIII, если

$$\frac{1}{3} * \text{Max}X < X < \frac{2}{3} * \text{Max}X \quad \text{и} \quad \text{Min}Y < Y < \frac{1}{3} * \text{Max}Y, \quad (12)$$

принадлежит области IX, если

$$\frac{2}{3} * \text{Max}X < X < \text{Max}X \quad \text{и} \quad \text{Min}Y < Y < \frac{1}{3} * \text{Max}Y, \quad (13)$$

где X – координата начала или конца вектора по оси X , а Y – координата начала или конца вектора по оси Y .

Учитывая изложенное, получаем множество, описывающее букву “А”

$$S_A = \left\{ \vec{L}_1(Z_1, D3, VII, IV), \vec{L}_2(Z_2, D2, IV, V), \dots, \vec{L}_{10}(Z_{10}, D1, V, VI) \right\}, \quad (14)$$

где L_i – i -й вектор множества буквы, Z_i – длина i -го вектора, D_i – направление i -го вектора, первое число – принадлежность начала вектора одной из 9-ти областей буквы, второе число – принадлежность конца вектора одной из 9-ти областей буквы.

Знак характеризуется не только векторами, но и связью между векторами. Если начало i -го вектора совпадает с концом $i-1$ -го вектора, то такие векторы называются связанными (обозначим *Con* – от английского слова *connected* - связанный). Иначе – несвязанными (обозначим *Uncon* – от английского слова *unconnected* - несвязанный). Например, для буквы “А”, изображенной на рисунке 2, можно сделать следующее описание векторов и связей между ними

$$\text{Uncon}, \vec{L}_1, \text{Con}, \vec{L}_2, \text{Con}, \vec{L}_3, \text{Con}, \vec{L}_4, \dots, \text{Con}, \vec{L}_8, \text{Uncon}, \vec{L}_9, \text{Con}, \vec{L}_{10}. \quad (15)$$

Для распознавания знака необходимо привести его к модели, которая будет анализироваться программой. Для этого необходимо учитывать некоторые особенности написания букв и цифр человеком и приближение метода распознавания от теории к практике.

Ниже перечислены некоторые особенности, которые влияют на построение окончательной модели исследуемого знака:

1. Определяется длина диагонали исследуемого знака, как характеристика величины знака

$$D = \sqrt{(\text{MAX}X - \text{MIN}X)^2 + (\text{MAX}Y - \text{MIN}Y)^2}, \quad (16)$$

где D – искомая величина диагонали знака

2. Исходя из длины исследуемого знака, определяется значимость вектора. Это дополнительное свойство вектора необходимо для отсеечения векторов, которые не характеризуют знак

$$Z(L_i) \geq D * k, \quad (17)$$

где k – коэффициент значимости, определенный практической реализацией метода.

3. Объединяются векторы, которые удовлетворяют следующему условию:

Если вектор L_i и вектор L_{i+1} являются связанными (Con) и имеют одинаковое направление, то эти векторы объединяются в один вектор L_i , начало которого совпадает с началом вектора L_i , а конец – с концом вектора L_{i+1} .

4. Устраняются несвязанные (Uncon) векторы, если конец вектора L_i и начало вектора L_{i+1} удовлетворяют условию

$$|X_2(L_i) - X_1(L_{i+1})| < \Delta \quad \text{и} \quad |Y_2(L_i) - Y_1(L_{i+1})| < \Delta, \quad (18)$$

где $X_2(L_i)$ — конец вектора L_i по оси X , $X_1(L_{i+1})$ — начало вектора L_{i+1} по оси X , $Y_2(L_i)$ — конец вектора L_i по оси Y , $Y_1(L_{i+1})$ — конец вектора L_{i+1} по оси Y , Δ – величина разрыва

Методика построения модели исследуемого знака:

1. формирование множества векторов образующих знак (выражение 3),
2. получение множества (выражение 14),
3. определение размера знака (выражение 16),
4. объединение векторов,
5. удаление незначимых векторов (выражение 17),
6. устранение несвязанных векторов (выражение 18).

Например, для буквы “А” на рисунке 2, модель строиться следующим образом:

1.

$$S = \{ \vec{L}_1, \vec{L}_2, \vec{L}_3, \vec{L}_4, \vec{L}_5, \vec{L}_6, \vec{L}_7, \vec{L}_8, \vec{L}_9, \vec{L}_{10} \} \quad (19)$$

2.

$$S = \{ \vec{L}_1(Z_1, D3, VII, IV), \vec{L}_2(Z_2, D2, IV, V), \vec{L}_3(Z_3, D3, V, II), \vec{L}_4(Z_4, D3, II, II), \vec{L}_5(Z_5, D8, II, VI), \vec{L}_6(Z_6, D8, VI, VI), \vec{L}_7(Z_7, D7, VI, IX), \vec{L}_8(Z_8, D8, IX, IX), \vec{L}_9(Z_9, D1, IV, V), \vec{L}_{10}(Z_{10}, D1, V, VI) \} \quad (20)$$

3.

$$D = \sqrt{(MAXX - MINX)^2 + (MAXY - MINY)^2}, \quad (21)$$

4.

$$S = \{ \vec{L}_1(Z_1, D3, VII, IV), \vec{L}_2(Z_2, D2, IV, V), \vec{L}_{3,4}(Z_{3,4}, D3, V, II), \vec{L}_{5,6}(Z_{5,6}, D8, II, VI), \vec{L}_7(Z_7, D7, VI, IX), \vec{L}_8(Z_8, D8, IX, IX), \vec{L}_{9,10}(Z_{9,10}, D1, IV, VI) \} \quad (22)$$

Выражение 22 можно упростить, поменяв индексацию с учетом объединенных векторов

$$S = \left\{ \vec{L}_1(Z_1, D3, VII, IV), \vec{L}_2(Z_2, D2, IV, V), \vec{L}_3(Z_3, D3, V, II), \right. \\ \left. \vec{L}_4(Z_4, D8, II, VI), \vec{L}_5(Z_5, D7, VI, IX), \vec{L}_6(Z_6, D8, IX, IX), \right. \\ \left. \vec{L}_7(Z_7, D1, IV, VI) \right\} \quad (23)$$

5. Незначимых векторов нет

6. Векторы L_6 и L_7 являются несвязанными, поэтому (23) принимает вид

$$S = \left\{ Uncon, \vec{L}_1(Z_1, D3, VII, IV), \vec{L}_2(Z_2, D2, IV, V), \right. \\ \left. \vec{L}_3(Z_3, D3, V, II), \vec{L}_4(Z_4, D8, II, VI), \vec{L}_5(Z_5, D7, VI, IX), \right. \\ \left. \vec{L}_6(Z_6, D8, IX, IX), Uncon, \vec{L}_7(Z_7, D1, IV, VI) \right\} \quad (24)$$

Таким образом, получаем модель исследуемого знака. Эта модель в процессе распознавания должна сравниваться с шаблонными моделями букв и цифр. Шаблонные модели строятся иначе, чем модель исследуемого знака. Необходимо учесть все возможные варианты написания знака, при этом сделать эту модель универсальной. Исследования показывают, что для построения шаблонных моделей можно использовать простейшую модель буквы или цифры, дополнив ее возможными направлениями написания и расположения точек начала и конца векторов, образующих знак. Например, буква “А” в простейшем случае состоит из трех векторов:

Первый вектор начинается в областях IV, VII, VIII, имеет направление $D2, D3, D4$, заканчивается в областях I, II, III,

Начало второго вектора связано с концом первого вектора, имеет направление $D6, D7$, заканчивается в областях VI, VIII, IX,

Третий вектор начинается в областях I, IV, имеет направление $D1, D2, D8$, заканчивается в областях III, VI, IX.

На рисунке 6 изображены примеры написания буквы “А”, которые могут быть описаны представленной выше шаблонной моделью этой буквы.

Следует отметить, что шаблонные модели характеризуются не только векторами, их положением, но и связями между ними. Соответственно, можно разбить все шаблонные модели на группы по количеству несвязанных векторов. Например, буквы “А”, “В”, “Р” характеризуются отсутствием одной связи. Буквы “Г”, “Н”, “О”, “П”, “С” - являются связанными. Разбиение по такому признаку может значительно ускорить и уточнить процесс распознавания. Распознавание сводится к сравнению модели исследуемого знака и шаблонных моделей. Упрощенный алгоритм распознавания приведен на рисунке 7.

Изложенный выше метод был реализован в программе “ArtInt” (программа находится на веб-сайте http://geocities.com/vova_braha/index.html) и показал точность распознавания до 99%.

Достоинства метода:

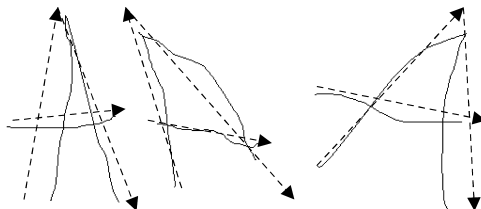


Рис. 6 – Пример написания буквы “А”

1. Универсальность шаблонной модели, так как она описывает все допустимые варианты написания знака
2. Высокая точность распознавания – до 99%
3. Простота программной реализации метода
4. Возможное расширение метода для решения задачи распознавания рукописного слитного текста

Недостатки метода:

1. Сложность создания шаблонной модели знака
2. Ограниченность применения метода только векторным представлением информации о знаке

В заключение хотелось бы отметить, что расширение метода для задачи распознавания рукописного текста усложняется тем, что в этой задаче на распознавание приходится лишь 30%, а остальные 70% лежат в области понимания компьютером смысла документа. Соответственно, дальнейшее развитие метода должно быть тесно связано с развитием методов понимания текста.

Литература

1. Ивахненко А.Г. Самообучающиеся системы распознавания и автоматического регулирования. //К.: “Наукова думка”, 1969—349 стр.
2. Тимофеев А.В. Роботы и искусственный интеллект. // М.: “Наука” 1978-192 стр.
3. Павлидис Т. Алгоритмы машинной графики и обработки изображений. 1986 г.
4. Riedmiller, M., H.Braun. “A direct adaptive method for faster backpropagation learning: the RPROP algorithm”. Proceeding of the IEEE International Conference on Neural Networks, 1993.

Получено: 16.06.2002

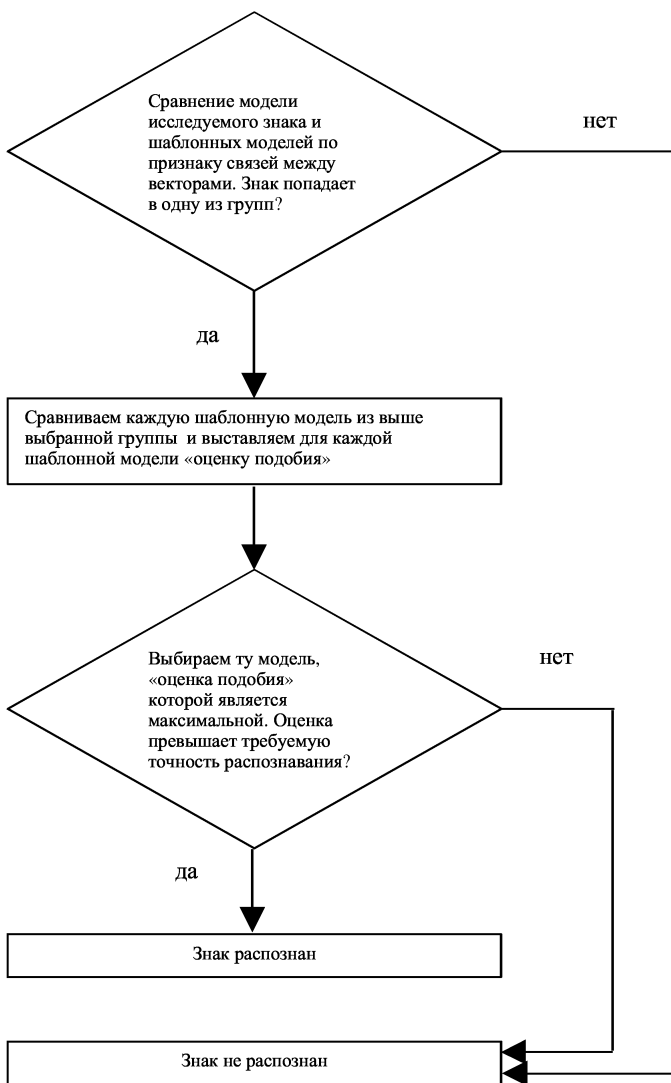


Рис. 7 – Блок-схема алгоритма распознавания текста