

ЭКСПЕРИМЕНТ "ЦИФРОВОЕ ПИСЬМО"

А.С.Нудельман

1. Проведение эксперимента. Цель эксперимента "цифровое письмо" — проверить эффективность применения понятий и идей, разработанных в [1,2], для построения "читающих" автоматов.

В ходе эксперимента был построен автомат $P-WF$ (в виде АЛГОЛ-программы для ЭВМ "Минск-22"), "знающий" все десять арабских цифр. Объектом чтения для этого автомата может являться произвольное целое (без знака) число, написанное от руки. Чтобы чтение происходило безошибочно, написание чисел должно удовлетворить трем требованиям:

1. Высота каждой цифры должна быть не меньше трети высоты цифрового слова (рис. 1,а).

2. Различные цифры не должны иметь общих точек, принадлежащих основным линиям этих цифр (рис. 1,б,в). Соединения цифр возможны только за счет "хвостов" и соединительных линий.

3. Цифры должны быть написаны так, чтобы восприятия этих цифр различными людьми совпадали.

В проведении эксперимента приняло участие 20 человек, которые представили 150 образцов написания чисел (некоторые из этих образцов показаны в приложении). Для распознавания брались случайные числа вплоть до шестизначных (хотя автомат $P-WF$ рассчитан на "восприятие" чисел произвольной значности, объем оперативной памяти ЭВМ "Минск-22" вынудил ограничиться шестизначными числами). Все числа, предъявленные автомату $P-WF$, были "прочитаны" им правильно.

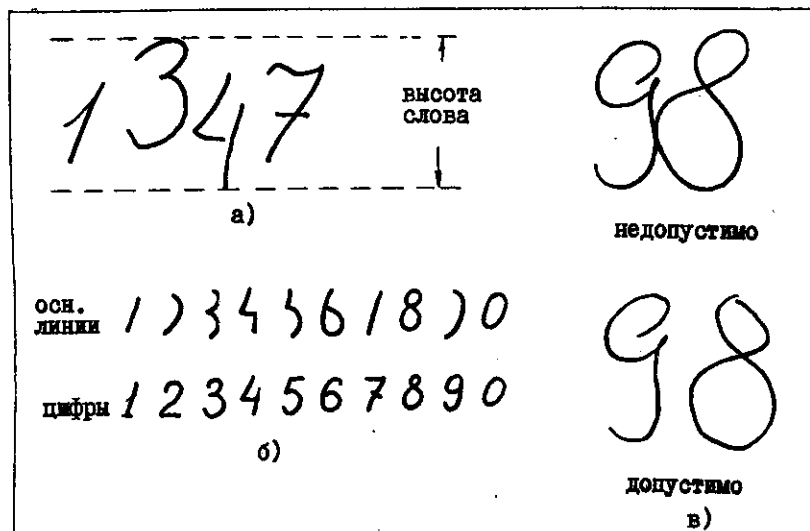


Рис. 1

II. Принцип работы автомата $P-WF$. Входной информацией для автомата $P-WF$ является конструктивная фигура [1, стр. 75], которая задается двумя множествами: множеством MT нумерованных точек, куда занесены координаты [1, стр. 38] точек, и множеством MC связей, элементы которого - пары точек (номеров точек). Конструктивная фигура есть, по сути, кусочно-линейная аппроксимация предъявляемого линейчатого изображения, поскольку пара точек, принадлежащая MC , представляет собой аппроксимирующий отрезок прямой, концевыми точками которого являются точки этой пары, а множество MC характеризует всю совокупность аппроксимирующих отрезков. Множество MT содержит все (и только их) точки, встречающиеся в парах из MC .

Будем говорить, что отрезок прямой $[t_1, t_2]$ хорошо аппроксимирует участок $L(t_1, t_2)$ линии фигуры, если δ (см. рис. 2) не превышает 2% высоты цифрового слова. Число точек в MT и расположение этих точек в фигуре выбираются так, чтобы аппроксимация исходного изображения была хорошей, т.е. каждый отрезок прямой должен иметь и хорошо аппроксимировать соответствующий

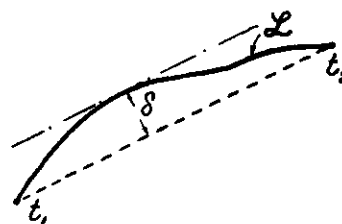


Рис. 2

участок линии фигуры и, кроме того, любая точка фигуры должна принадлежать некоторому участку линии, имеющему отрезок, аппроксимирующий этот участок. Нумерация точек в MT и единицы длины при измерении координат точек могут быть произвольными.

В соответствии с моделью восприятия [2, стр. 79, 83] алгоритм, распознающий образы (в данном случае - цифры), включает

в себя, прежде всего, механизм установления значения истинности на воспринимаемом изображении предложений, являющихся описаниями этих образов. Отсюда ясно, что алгоритм распознавания характеризуется, прежде всего, заложенными в основу его работы описаниями образов. Если реализованные в автомате $P-WF$ предложения $S_0, S_1, \dots, S_9$ (описания цифр 0, 1, ..., 9) формулировать так, как это сделано для предложения S_F [1, стр. 56-59], то изложение будет слишком громоздким. Поэтому здесь применен другой способ формулировки предложений, который назовем графическим. Конечно, подобное представление предложений не так полно, как формальное, поскольку при графическом представлении невозможно указать все детали соотношений между метрическими характеристиками объектов описания (линий, точек, направлений и углов). Кроме того, графическое описание может отразить только чисто конъюнктивную схему предложения [1, стр. 53], т.е. только конъюнкцию высказываний, входящих в графическое описание. Однако графическое представление предложений способно выразить общий характер предложений и является наглядным.

Прежде чем формулировать правила графического описания, введем несколько определений. Будем говорить, что направление $\vec{E}$ равно числу α ($\vec{E} = \alpha, 0 \leq \alpha < 2\pi$), если и только если при повороте этого направления против часовой стрелки на угол α оно становится вертикальным, идущим снизу вверх. Выражение " $\vec{E} < \alpha$ " будет означать, что $\vec{E} = \beta$ и $\beta < \alpha$. Будем говорить, что направление $\vec{E}$ принадлежит сектору $[\alpha, \beta]$, $0 \leq \alpha, \beta < 2\pi$ (обозначать: $\vec{E} \in [\alpha, \beta]$), если

$$\vec{E} = \gamma \& (\alpha < \beta \Rightarrow \alpha \leq \gamma \leq \beta) \& (\beta \leq \alpha \Rightarrow \gamma \geq \alpha \vee \gamma \leq \beta).$$

Пусть $\mathcal{L}(t_1, t_2, \dots, t_n)$ - конструктивная линия, т.е. линия, "составленная" из прямолинейных отрезков $[t_i, t_{i+1}]$, $i = 1, 2, \dots, n-1$. Линию $\mathcal{L}$ будем называть правосторонней, если

$$\forall i (1 \leq i \leq n \Rightarrow \vec{E}(t_i, t_{i+1}) \in [\alpha, \alpha + \pi (\text{mod } 2\pi)]),$$

и левосторонней, если

$$\forall i (1 \leq i \leq n \Rightarrow \vec{E}(t_i, t_{i+1}) \in [\beta, \beta + \pi (\text{mod } 2\pi)]).$$

Здесь $\vec{E}(t_i, t_{i+1})$ - направление от точки t_i к точке t_{i+1} , а α и β таковы, что $\vec{E}(t_{i-1}, t_i) = \alpha$ и $\vec{E}(t_i, t_{i+1}) = \beta$. Линия, не являющаяся ни правосторонней, ни левосторонней, будем называть свободной.

Теперь сформулируем правила графического описания. Всего таких правил - три.

1. Высказывание "существует линия $\mathcal{L}_i(t_j, t_k)$ такая, что в любой точке этой линии направление линии [1, стр. 41] (идущее по касательной в сторону продолжения линии) принадлежит сектору $[\alpha, \beta]$ " выражается линиями типа а, б и в (рис.3). При этом показывается номер линии, номера концевых точек и сектор $[\alpha, \beta]$ с вершиной в начальной точке линии. Всегда $k = j+1$.

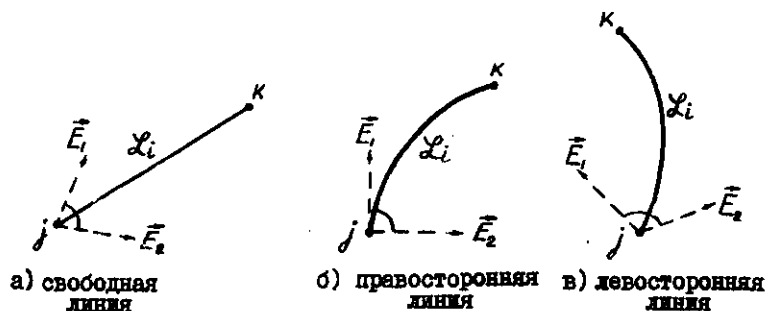


Рис. 3

На рис.3 имеются обозначения направлений $\vec{E}_1$ и $\vec{E}_2$ - крайних направлений сектора, которые не принадлежат графическому описанию. Они даны здесь только для того, чтобы сказать, что $\vec{E}_1 = \alpha$, а $\vec{E}_2 = \beta$.

2. Высказывание "существует точка t_n " выражается точкой, отмеченной крестиком и обозначенной номером n .

3. Если совокупность линий графического описания предложения S принять за изображение F (точнее, за фигуру [1, стр.39]), то S должно быть истинным на F , причем эта истинность обеспечивается и теми линиями и точками, которые указаны в графическом описании. (Этот пункт, в частности, говорит о том, что все метрические соотношения на графическом описании удовлетворяют предложению S . Но всегда надо иметь в виду, что предложение S требует выполнения любого метрического отношения с некоторой "погрешностью" (например, i - равенство длин, направлений и углов [1, стр. 46-49].)

Графические описания арабских цифр представлены на рис.4. Ниже указаны дополнительные особенности описаний $S_0, S_1, \dots, S_9$, не представленные в графической форме. Предполагая, что читателю известны геометрические свойства нормально написанной цифры (эти свойства в описаниях присутствуют), в дополнениях мы приводим лишь те детали описаний, которые специфичны именно для данного автомата.

Общим для всех описаний является то, что некоторые линии могут быть пустыми (не содержать ни одной точки). Кроме того, здесь используется интенсивность перестройки отдельно взятой линии, которая рассматривается при этом как самостоятельная фигура [2, стр. 82-83]. Интенсивность перестройки линии $\mathcal{L}$ обозначается через $\mathcal{I}(\mathcal{L})$.

S_0 : Число нарушений левосторонности линий - не более двух на всю фигуру (используется идея условной истинности [2, стр. 81]).

$$y(t_4) - y(t_2) \geq 0,875 (y(t_0) - y(t_2)).$$

$$(y(t_i) - y - \text{координата точки } t_i)$$

S_1 : Если $\mathcal{L}_2 \neq 0$, то

$$m[t_2, t_1] \geq 0,75 m[t_0, t_1] \& \vec{E}(t_2, t_3) / \vec{E}(t_0, t_1) < 0,79.$$

($m[t_i, t_j]$ - расстояние между точками t_i и t_j , а $\vec{E}(t_2, t_3) / \vec{E}(t_0, t_1)$ - направление от точки t_2 к точке t_3 в изображении, повернутом так, что направление от t_0 к t_1 является вертикальным, идущим снизу вверх).

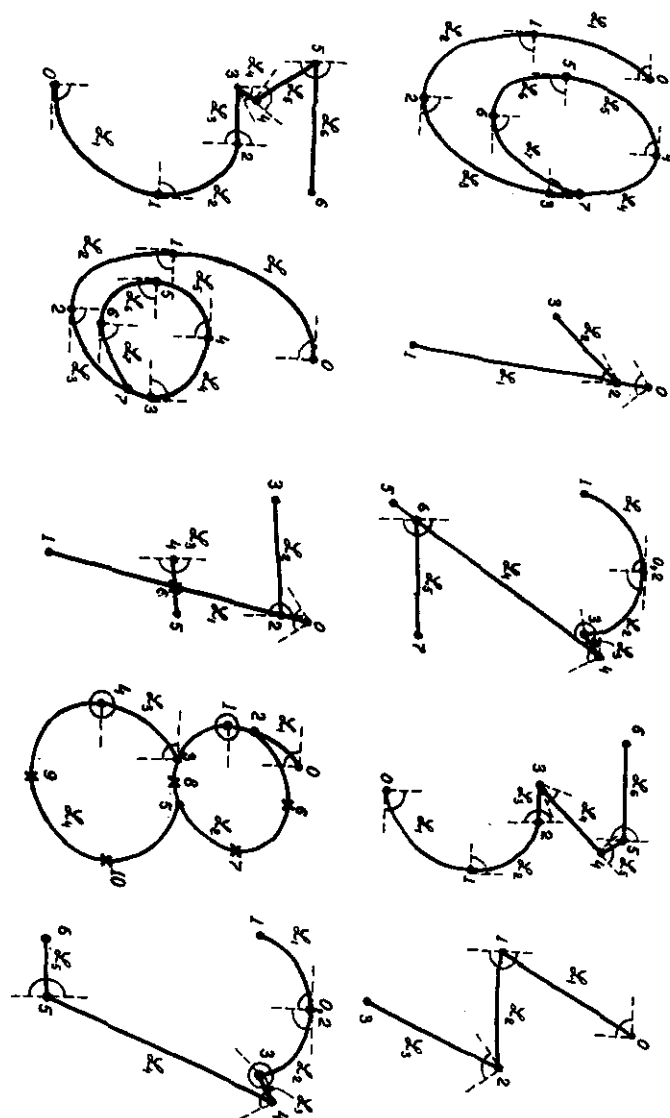


Рис. 4

S_2 : Если $L_i \neq \phi$, то $J(L_i) \leq 0,2$.

$$m[t_6, t_5] < 0,5 m[t_3, t_5]$$

S_3 : В начале линии L_6 возможно одно нарушение направления этой линии.

$$J(L_1 \cup L_2 \cup L_3) \leq 0,1.$$

S_5 : То же, что и для S_3 .

S_6 : Может быть только одно нарушение левосторонности - в конце последней линии.

$$y(t_4) - y(t_2) < 0,875(y(t_0) - y(t_2)).$$

S_7 : $L_2 \neq \phi$ и $m[t_2, t_1] \geq 0,75 m[t_0, t_1]$. К тому же, если L_2 не свободна, то

$$\tilde{E}(t_2, t_3) / \tilde{E}(t_0, t_1) \geq 0,79.$$

S_8 : Число нарушений левосторонности линий - не более двух на всю фигуру.

$$J(L_1 \cup L_2) \leq 0,2.$$

S_9 : Если $L_i \neq \phi$, то $J(L_i) \leq 0,2$.

Заметим, что предложения $S_0, S_1, \dots, S_9$ могут принадлежать языку описания [1, стр. 55] со следующими параметрами: $q_2 = 4$, $q_3 = 4$, $q_4 = 0$. Иными словами, в описаниях цифр сравнение длин линий осуществляется не точнее $100 \cdot 2^{-4}\%$, измерение направлений происходит с погрешностью $\pi \cdot 2^{-4}$, а углы не используются.

В соответствии с идеей активного восприятия исходное изображение достраивается допустимыми отрезками [2, стр. 80]. Следовательно, собственно распознавание цифр происходит не в исходной фигуре F (множества MT и MC), а в дополненной фигуре F (множества MT и MCA). Множество MCA всех связей содержит пары точек (связи, отрезки) из MC и пары точек, выражающие допустимые в F отрезки. Для любого отрезка из MCA указано, является ли этот отрезок реально существующим в F (принадлежит ли он MC) или это - достроенный отрезок.

Теперь ясно, что для установления истинности высказывания (из описания образа) "существует линия $L(t_1, t_2)$ " в фигуре F "ищется" конструктивная линия $L(t_1, t_2, \dots, t_n)$ такая, что

$$t_1 = t_{i_1}, t_2 = t_{i_n} \text{ и } \forall k (1 \leq k < n \Rightarrow \langle t_{i_k}, t_{i_{k+1}} \rangle \in MCA).$$

Здесь $i_k, k = 1, 2, \dots, n$, — номера точек, установленные множеством MT (но не описаниями!).

Требование однозначности написания цифр исключило необходимость применения в автомате $P-WF$ целевой достройки фигуры [3]. По этой же причине здесь не используется понятие условной истинности [2, стр. 81] в полном объеме.

Работа автомата $P-WF$ состоит в следующем. В исходном цифровом слове (фигуре F) определяется крайняя слева цифра — F_1 . После этого из фигуры F "вычеркиваются" все точки, принадлежащие основным линиям (рис. 1) цифры F_1 , и в оставшейся фигуре снова ищется крайняя слева цифра. Если цифровое слово содержит n цифр, то вышеописанная процедура повторяется n раз. После того, как определена (и "вычеркнута") последняя цифра, дальнейший поиск цифр не дает результата и автомат заканчивает свою работу.

Пусть F — изображение n -значного числа и пусть $F_1, F_2, \dots, F_{i-1}$ — уже найденные в F цифры ($1 \leq i < n+1$). На i -м шаге общего цикла работы "воспринимается" фигура $\bar{F}_i$ (множества MT_i и MCA_i) такая, что $\bar{F}_i = F \setminus (F_1 \cup F_2 \cup \dots \cup F_{i-1})$. В процессе выполнения i -го шага параллельно формируются два ряда фигур:

$$R_1: \bar{F}_i^1 = \phi, \bar{F}_i^2, \dots, \bar{F}_i^m$$

и

$$R_2: F_i^1 = \phi, F_i^2, \dots, F_i^m.$$

Ряд R_1 таков, что

- а) $\forall k (1 \leq k \leq m \Rightarrow \bar{F}_i^k \subseteq \bar{F}_i)$,
- б) $\forall k (1 \leq k \leq m \Rightarrow \bar{F}_i^k \text{ есть цифра})$,
- в) если $F' \subseteq \bar{F}_i$ и F' — цифра, то существует такое k , что $\bar{F}_i^k = F'$.

Поскольку в формировании ряда R_1 используются только основные утверждения описаний $S_0, S_1, \dots, S_g$ (что сокращает время работы автомата), то вполне возможна подфигура $\bar{F}_i^k$, которая в R_1 "читается" по-разному, т.е. для различных k и ℓ : $\bar{F}_i^k = \bar{F}_i^\ell$. Для сокращения времени формирования ряда R_1 фигуры $\bar{F}_i^k$ не "выбрасываются" из $\bar{F}_i$ с целью дальнейшей проверки

истинности какого-либо предложения на этой фигуре, а "строятся" согласно предложениям $S_0, S_1, \dots, S_g$ из отрезков, входящих в MCA_i .

Фигуры ряда R_2 являются промежуточными решениями. Фигура (цифра) $\bar{F}_i^m$ есть решение автомата $P-WF$ на i -м шаге общего цикла работы, т.е. — цифра F_i . При $k > 1$ цифра $\bar{F}_i^k$ равна либо цифре $\bar{F}_i^k$, либо цифре $\bar{F}_i^{k-1}$. Конкретный выбор определяется взаимным расположением цифр $\bar{F}_i^k$ и $\bar{F}_i^{k-1}$, их размерами, а также теми деталями описаний $S_0, S_1, \dots, S_g$, которые не использовались ранее (при построении ряда R_1).

3. Заключение. В ходе эксперимента "цифровое письмо" построен автомат, "читающий" n -значные десятичные числа, написанные от руки различными людьми, и сформулированы ограничения на написание чисел, практически сохраняющие естественность письма. Имеется уверенность (не опровергнутая проведенным экспериментом), что автомат $P-WF$ способен в принципе правильно распознавать любое число, написанное с учетом указанных ограничений любым человеком.

Результат эксперимента показал эффективность применения понятий и идей, изложенных в [1, 2], для построения читающих автоматов. Алгоритм работы автомата $P-WF$, как считает автор, может служить логической основой автоматов, предназначенных для чтения числовых документов.

Л и т е р а т у р а

1. НУДЕЛЬМАН А.С. Об одном формализованном подходе к изучению зрительного восприятия. — В кн.: Вычислительные системы. Вып. 55, Новосибирск, 1973, с. 36–78.
2. НУДЕЛЬМАН А.С. Активное восприятие зрительной информации. — Там же, с. 79–93.
3. НУДЕЛЬМАН А.С. Принцип целевой достройки в активном восприятии зрительной информации. — Настоящий сборник, с. 69–72.

Поступила в ред.-изд.отд.
8 мая 1974 года

Приложение

84398	5753	961	21
0206	518	477505	42813
68	57	9849	2320
69785	5294	481	21
323	619	554172	57875
88	20	6036	7753
34020	4391	765	28
069	858	566429	50581

Некоторые образцы написания чисел