ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ и МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 2 Сентябрь 1962/Октябрь № 5

ИССЛЕДОВАНИЕ АЛГОРИТМА РАСПОЗНАВАНИЯ МАШИНОПИСНЫХ ЗНАКОВ

В. К. ЕЛИСЕЕВ, В. А. КОВАЛЕВСКИЙ

(*Киев*)

В настоящей работе описаны исследования одного алгоритма распознавания стандартных знаков с весьма высокой степенью надежности [1]. В основе этого алгоритма лежит предположение, что все знаки одного определенного шрифта, например шрифта пишущей машинки, представляют собой некоторые идеализированные, эталонные, знаки, подвергнувшиеся искажениям: изображение знака может отличаться от соответствующего эталона сдвигом, равномерным потемнением и изменением контрастности. Кроме того, на изображение накладывается шум, т. е. отдельные точки, независимо от остальных, могут быть темнее или светлее соответствующих точек эталона. Предполагается, что все эти искажения описываются простыми статистическими закономерностями.

Алгоритм распознавания заключается в следующем. Вычисляются коэффициенты корреляции неизвестного знака с каждым из эталонных при всевозможных сдвигах в ограниченной области. Неизвестный знак отождествляется с тем из эталонных, которому соответствует наибольший коэффициент корреляции. Для упрощения вместо коэффициента корреляции вычисляется пропорциональное ему скалярное произведение ненормированного вектора, описывающего неизвестный знак, на нормированный вектор эталонного знака. Определение векторов эталонных знаков и их нормировка выполняются заранее. В соответствии с этим алгоритм состоит из двух частей: «обучения», при котором вычисляются нормированные векторы эталонных знаков, и собственно распознавания.

В связи с большой сложностью систем автоматического распознавания образцов, в частности букв и цифр, целесообразно проводить исследования путем моделирования таких систем на универсальной вычислительной машине. Метод моделирования позволяет без затрат материалов и при весьма значительной экономии труда и времени проверять эффективность и надежность любых алгоритмов распознавания.

В Институте кибернетики АН УССР разработана специальная приставка к универсальной вычислительной машине «Киев», представляющая собой универсальный преобразователь изображений [2]. Приставка предназначена для моделирования любых алгоритмов, связанных с обработкой информации, содержащейся в различных рисунках. Приставка работает по следующей схеме.

Координаты центра обследуемого в данный момент небольшого элемен-

та рисунка в виде двоичных кодов поступают из вычислительной машины на преобразователь кода в напряжение. Полученные напряжения воздействуют на отклоняющие пластины электронно-лучевой трубки, экран которой проектируется объективом на исследуемый рисунок. Отраженный от обследуемого элемента световой поток попадает на фотоумножитель. Яркость светового пятна неоднородна и изменяется по мере удаления от центра пятна. Поэтому при получении информации о коэффициенте отражения элемента происходит усреднение локальных коэффициентов отражения с весом. Ток фотоумножителя, пропорциональный усредненному коэффициенту отражения, преобразуется в двоичный код и поступает в вычислительную машину.

Диаметр светового пятна, а следовательно, и размеры элементов разложения можно изменять в некоторых пределах.

Порядок обхода рисунка определяется программой, по которой машина вычисляет координаты точек, являющихся центрами элементов. Рисунок может быть разложен на $256 \times 256 = 65\,536$ элементов. Таким образом, полная информация о рисунке может рассматриваться как массив чисел $\rho[x, y]$, причем индексы x и y могут изменяться от 1 до 256.

Преобразователь позволяет различать 16 уровней почернения. Белому полю соответствует наибольшее значение ρ, равное единице.

Имеется дополнительная, контрольная, электронно-лучевая трубка для визуального наблюдения за движением светового пятна по рисунку.

Ниже рассматривается программа, предназначенная для распознавания 10 машинописных цифр, и приводятся результаты экспериментальных исследований. Программа описана в языке АЛГОЛ-60 [3]. Структура программы достаточно подробно изложена в комментариях.

§1. Описание программы

Обучение

начало действительные массивы ρ [*1* : *256*, *1* : *256*], *M* [*1* : *N*] *и E* [*1* : *n*, *1* : *N*];

процедура *высвечивание B* (*x*, *y*);

комментарий: *Процедура высвечивание обозначает специальную операцию по которой центр светового пятна устанавливается в точку с координатами x, y;*

процедура *растр P* (*x0*, *y0*, *a*, *b*);

комментарий: *Процедура растр заносит в массив M* [*1* : *N*] *значения коэффициентов отражения в точках изображения, расположенных в виде прямоугольной сетки с высотой b, шириной a и расстоянием между точками* Δx *по горизонтали и* Δy *по вертикали. Параметры растра выбираются в соответствии с размерами и средней толщиной линии цифр исследуемого машинописного шрифта. Были приняты следующие размеры растра: число элементов в строке равно 8, в столбце — 16;*

начало

$m := 1$;

для $y := y0$ **шаг** Δy **до** $y0 + b$ **выполнить**
для $x := x0$ **шаг** Δx **до** $x0 + a$ **выполнить**
начало $B\ (x, y)$; $M\ [m] := \rho\ [x, y]$; $m := m + 1$ **конец**
конец P;
процедура *контроль* $K\ (a, b, U, \tau, \pi, \delta)$;
комментарий: *Процедура контроль образует на экране контрольной трубки изображение в соответствии с данными, занесенными ранее в массив* $U\ [1 : N]$. *Число повторных высвечиваний изображения определяется значением параметра* τ. *Изменение яркости отдельных элементов моделируется многократным высвечиванием одной точки, т. е. повторением процедуры* $B\ (x, y)$. *При высвечивании лучше воспринимается негативное изображение, т. е. светлая цифра на темном фоне. Поэтому высвечиваются точки, соответствующие темным элементам занесенного в память изображения. Порог* π *выбирается так, чтобы фон изображения не высвечивался, но даже незначительные отклонения от уровня фона были бы видны на экране. Градации яркости при высвечивании определяются константой* δ: *данная точка высвечивается столько раз подряд, во сколько разность порога* π *и элемента* $U\ [u]$ *превосходит* δ;
начало
для $t := 1$ **шаг** 1 **до** τ **выполнить**
$u := 1$;
для $y := 0$ **шаг** Δy **до** b **выполнить**
для $x := 0$ **шаг** Δx **до** a **выполнить**
начало
$p := \pi$;
высвечивание: **если** $U\ [u] \leqslant p$ **то**
начало $B\ (x, y)$; $p := p - \delta$ **перейти к** *высвечиванию* **конец**
$u := u + 1$
конец
конец K;
процедура *преобразование* $Пр\ (k)$;
комментарий: *Процедура* $Пр\ (k)$ *преобразует числа, характеризующие* k*-й эталонный знак, и заносит полученные данные в двумерный массив* $E\ [1 : n, 1 : N]$, *занимая часть этого массива, соответствующую номеру* k. *Номер* k *задает экспериментатор. Преобразование состоит из вычитания усредненного по данному изображению коэффициента отражения и из нормировки полученного вектора*;
начало
$s := \mu := 0$;
для $m := 1$ **шаг** 1 **до** N **выполнить**
$s := M\ [m] + s$;
$s := s/N$;
для $m := 1$ **шаг** 1 **до** N **выполнить**
$\mu := (M\ [m] - s) \uparrow 2 + \mu$;
$\mu := sqrt(\mu)$;
для $m := 1$ **шаг** 1 **до** N **выполнить**
$E\ [k, m] := (M\ [m] - s) / \mu$;
конец $Пр$;

Начало программы:

растр : P ($x0$, $y0$, $7 \times \Delta x$, $15 \times \Delta y$);

контроль изображения: K ($7 + \Delta x$, $15 \times \Delta y$, M, $\tau 1$, $\pi 1$, $1/16$);

комментарий: *Процедура K с меткой контроль изображения позволяет визуально наблюдать занесенное в массив M [1 : N] изображение*;

преобразование : *Пр* (k);

для $u := 1$ **шаг** 1 **до** N **выполнить**

$U[u] := E[k, u]$;

контроль эталонного знака: K ($7 \times \Delta x$, $15 \times \Delta y$, U, $\tau 2$, $\pi 2$, $\delta 1$);

комментарий: *Процедура K с меткой контроль эталонного знака осуществляет визуальный контроль преобразованного набора чисел, характеризующего k-й эталонный знак*

конец *программы*.

В связи с ограниченным объемом ОЗУ машины «Киев» в одной ячейке памяти хранится по три компоненты массива $E[1 : n, 1 : N]$. Обеспечиваемая при этом точность в 12 двоичных разрядов достаточна для вычисления коэффициентов корреляции.

Программа обучения повторяется для каждого эталонного знака. После того как в преобразователь изображений заложен образец очередного эталонного знака и на пульте управления набран соответствующий номер k, программа выполняется сначала.

Полученная информация об эталонных знаках, хранящаяся в массиве $E[1 : n, 1 : N]$, может быть перенесена на перфоленту или перфокарты. Это избавляет от необходимости повторять процесс обучения при последующих экспериментах.

Распознавание

Алгоритм распознавания не требует точной центровки знака. Поэтому при распознавании используется растр бо́льших размеров, чем при обучении. Число элементов, на которые разлагается эталонный знак, равно 128 (растр 8×16). Если задать точность центровки, по высоте и по ширине равной 4 элементам, то растр при распознавании должен иметь размеры 12×20 (240 элементов).

Алгоритм исследовался при сдвиге с шагом, равным одному элементу (всего 25 сдвигов), и при сдвиге с шагом, равным половине элемента (всего 100 сдвигов). Необходимая величина шага сдвига может быть найдена экспериментально. Этот вопрос рассматривается ниже.

Сдвиги в программе могут реализоваться двумя различными способами. В одном случае сдвигается растр и при каждом сдвиге информация заносится в ячейки памяти заново. В другом — массив исходной информации $M[1 : N1)$ остается неизменным и сдвиги осуществляются путем выборки различных подматриц из матрицы, в качестве которой можно рассматривать массив $M[1 : N1]$. Число строк и столбцов матрицы соответствует аналогичным параметрам растра.

Достоинством второго способа по сравнению с первым является несколько меньшее время счета, потому что в этом случае нет необходимости обращаться к изображению при выполнении каждого сдвига. Однако невоз-

можность реализации при этом сдвигов более мелких, чем один элемент, приводит к необходимости комбинировать оба способа.

В программе 25 сдвигов с шагом, равным одному элементу, осуществляются без повторения растровой развертки. 100 сдвигов при шаге, равном половине элемента, реализуются следующим образом. После каждого цикла распознавания с 25 сдвигами на один элемент происходит изменение одной из координат начала растра ($x0$, $y0$) на величину, равную половине шага растра. Исходная информация снова засылается в массив M [$1 : N1$], и цикл распознавания с 25 сдвигами повторяется.

начало действительные массивы ρ [$1 : 256$, $1 : 256$], M [$1 : N1$] и E [$1 : n$, $1 : N$];
процедура *высвечивание* B (x, y);
процедура *растр* P ($x0$, $y0$, a, b);
процедура *контроль* K (a, b, U, τ, π, δ);
комментарий: *Процедуры высвечивание, растр и контроль определены ранее для обучения*;
процедура *отыскание максимальных скалярных произведений МСП*;
действительный массив r [$1 : n$];
комментарий: *Процедура МСП вычисляет значения скалярного произведения R вектора неизвестного изображения с нормированными векторами эталонных знаков для всех возможных сдвигов с шагом, равным одному элементу. Для каждого номера эталонного знака k отыскивается наибольшее по всем сдвигам значение R, превосходящее величину r [k]. В массиве r [$1 : n$] хранятся максимальные значения R, найденные при предшествующих выполнениях данной процедуры. Число сдвигов по горизонтали обозначено через c по вертикали — через h. Общее число сдвигов равно $c \times h$. Сдвиги в процедуре МСП осуществляются путем выборки подматриц из прямоугольной матрицы неизвестного изображения, в качестве которой рассматривается массив M [$1 : N1$]. Эта матрица имеет размеры, соответствующие размерам растра: число столбцов равно α, число строк β, так что $N1 = \alpha \times \beta$*;
начало
для $k := 1$ **шаг** 1 **до** n **выполнить**
начало
комментарий: *в циклах по m и ml вычисляется скалярное произведение вектора k-го эталонного знака, выбираемого из массива E [$1 : n$, $1 : N$], и вектора сдвинутого изображения, в качестве которого рассматривается подматрица, извлекаемая из массива M [$1 : N1$]. mg — начальный элемент подматрицы при нулевом горизонтальном сдвиге, mp — начальный элемент подматрицы при любом сдвиге, ml — начальный элемент строки подматрицы. Перебор всех возможных сдвигов осуществляется в циклах по mp и mg, перебор эталонных знаков — в цикле по k*;
для $mg := 1$ **шаг** α **до** $\alpha \times (h - 1) + 1$ **выполнить**
для $mp := mg$ **шаг** 1 **до** $mg + c - 1$ **выполнить**
начало
$R := 0$; $e := 1$;
для $ml := mp$ **шаг** α **до** $mp + \alpha \times (\beta - h)$ **выполнить**
для $m := ml$ **шаг** 1 **до** $ml + \alpha - c$ **выполнить**

```
            начало
                R:= M [m] × E [k, e] + R; e:= e + 1;
            конец
        если R ⩾ r [k] то r [k]:= R
    конец для mp и mg
  конец k
конец МСП;
Начало программы:
  растр: P (x0, y0, 11 × Δx, 19 × Δy);
  контроль изображения: K (11 × Δx, 19 × Δy, M, τ3, π3, 1/16);
  для k:=1 шаг 1 до n выполнить
        r [k]:= 0;
  для y:= y0 шаг Δy/2 до y0 + Δy/2 выполнить
  для x:= x0 шаг Δx/2 до x0 + Δx/2 выполнить
  начало
      растр: P (x, y, 11 × Δx, 19 × Δy);
      отыскание максимальных скалярных произведений: МСП
  конец
  комментарий: в циклах по x и по y реализуются сдвиги с шагом, равным половине элемента разложения;
  d:= ϰ, f:= θ;
  для k:= 1 шаг 1 до n выполнить
        если r [k] ⩾ f то начало f:= r [k]; d:= k конец
конец программы
```

Результатом распознавания является значение d. Если максимальное скалярное произведение превосходит порог распознавания θ, то значение d является ответом. Оно равно номеру эталонного знака, для которого получено максимальное скалярное произведение. В противном случае d остается равным $\varkappa$, что является сигналом отказа.

Варьируя величину порога θ, можно изменять в ту или иную сторону отношение вероятности отказа к вероятности неверного ответа.

По окончании процесса распознавания в массиве $r[1 : n]$ содержатся величины, пропорциональные максимальным коэффициентам корреляции распознаваемого знака с каждым из эталонных. Эти величины представляют интерес при исследовании описанного алгоритма и могут быть напечатаны одновременно с результатом распознавания.

Описанные программы обучения и распознавания содержат приблизительно 350 команд.

§ 2. Экспериментальная проверка алгоритма

Исследования производились для шрифта пишущей машинки «Москва».

Цикл распознавания со сдвигом на один элемент длится около 40 сек., со сдвигом на половину элемента — около 2.5 мин.

На фиг. 1 приведены цифры, выбранные в качестве эталонных. В табл. 1 даны результаты распознавания этих же цифр, используемых в качестве неизвестных.

В этой таблице значения корреляционных коэффициентов приведены в относительных единицах. Если обозначить корреляционный коэффициент,

вычисленный для k-го изображения с i-м эталоном, через R_i^k, то содержащиеся в k-й строке таблицы величины равны R_i^k/R_k^k. В каждой строке таблицы приведены данные, полученные при распознавании одного изображения.

При распознавании некоторого неизвестного изображения отыскивается максимум коэффициентов корреляции по всем сдвигам. Поскольку изображение располагается относительно растра случайным образом, при большом шаге сдвига все вычисленные коэффициенты корреляции могут значительно отличаться от действительного максимума. Шаг сдвигов должен быть выбран так, чтобы при самом неудачном расположении изображения максимум коэффициента корреляции с соответствующим данному изображению эталоном был больше максимального из коэффициентов корреляции с другими эталонами.

При распознавании большого числа изображений цифр было установлено, что корреляция с «чужими» эталонами никогда не превышает 80% корреляции со «своим» эталоном. (Корреляция изображения «4» с эталоном «1» достигает 80% от корреляции с эталоном «4», см. табл. 1.) Поэтому шаг сдвигов должен выбираться так, чтобы максимальный из вычисленных коэффициентов корреляции всякой цифры со своим эталоном был не меньше 80% от действительного максимума.

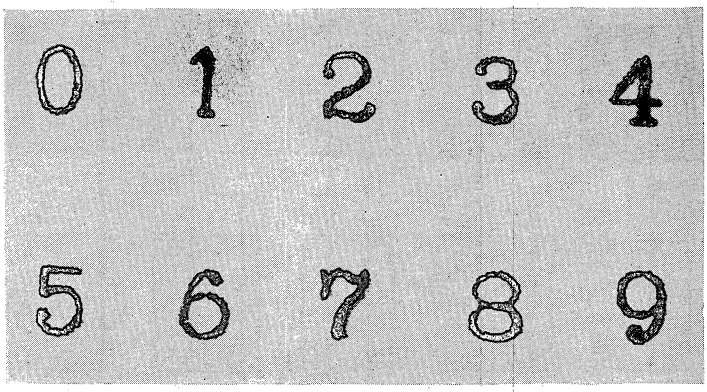
Для оптимального выбора величины сдвига при распознавании были предприняты исследования характера автокорреляционных функций машинописных цифр. При этом должны быть напечатаны все вычисляемые значения корреляционных коэффициентов R_ξ^k для фиксированного номера k. Индекс ξ является номером сдвига и принимает все значения от 1 до w (w — число сдвигов).

При исследованиях корреляционных функций применялась описанная выше программа распознавания со специальной подпрограммой печати корреляционных коэффициентов.

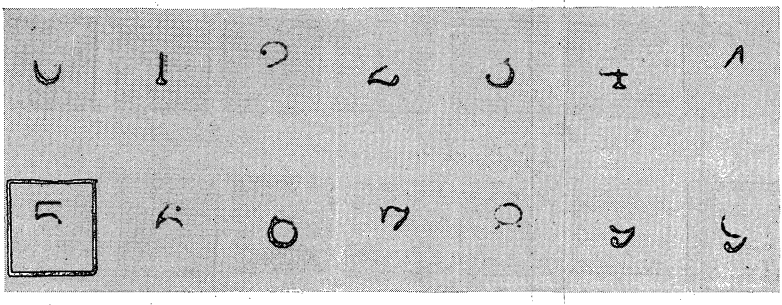
Экспериментально были получены таблицы значений R_ξ^k для 100 сдвигов с шагом, равным половине шага растра. Затем были построены сечения этой таблицы по строкам и столбцам с интерполяцией промежуточных значений R_ξ^k. Кривые равных уровней для цифр «0», «1», «8», построенные по этим сечениям, приведены на фиг. 2*а*, *б*, *в* соответственно. По осям координат отложены величины сдвигов. За единицу принят сдвиг на половину шага растра.

Таблица 1

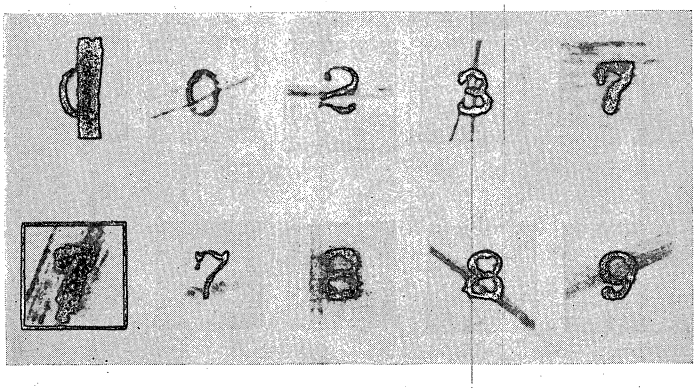
	0	1	2	3	4	5	6	7	8	9
0	1.00	0.56	0.53	0.67	0.44	0.64	0.78	0.39	0.61	0.78
1	0.42	1.00	0.23	0.39	0.74	0.39	0.39	0.42	0.35	0.29
2	0.43	0.32	1.00	0.57	0.34	0.39	0.32	0.39	0.66	0.45
3	0.68	0.50	0.58	1.00	0.42	0.74	0.68	0.37	0.74	0.42
4	0.33	0.80	0.30	0.38	1.00	0.38	0.50	0.40	0.33	0.40
5	0.60	0.39	0.39	0.66	0.32	1.00	0.66	0.32	0.50	0.39
6	0.66	0.47	0.34	0.59	0.47	0.61	1.00	0.32	0.61	0.47
7	0.31	0.48	0.38	0.33	0.45	0.26	0.33	1.00	0.19	0.45
8	0.56	0.54	0.70	0.67	0.35	0.49	0.60	0.23	1.00	0.47
9	0.65	0.37	0.44	0.42	0.49	0.42	0.42	0.49	0.49	1.00



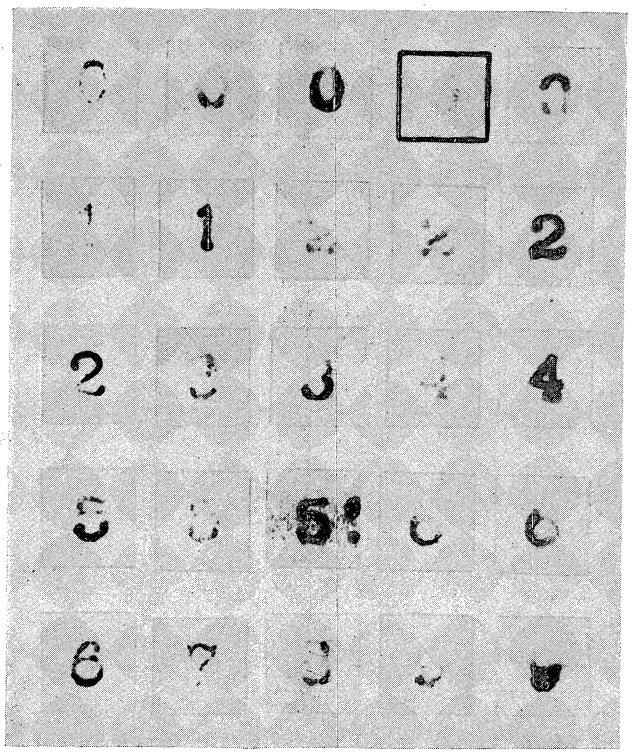
Фиг. 1



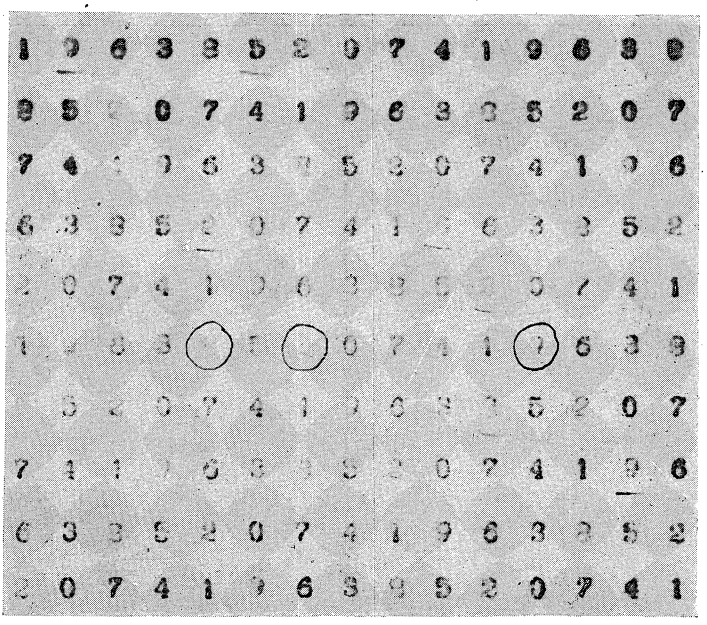
Фиг. 3



Фиг. 5



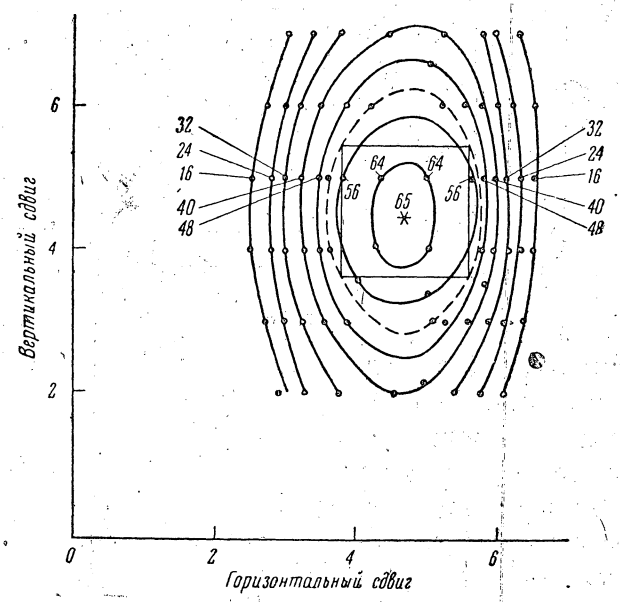
Фиг. 4



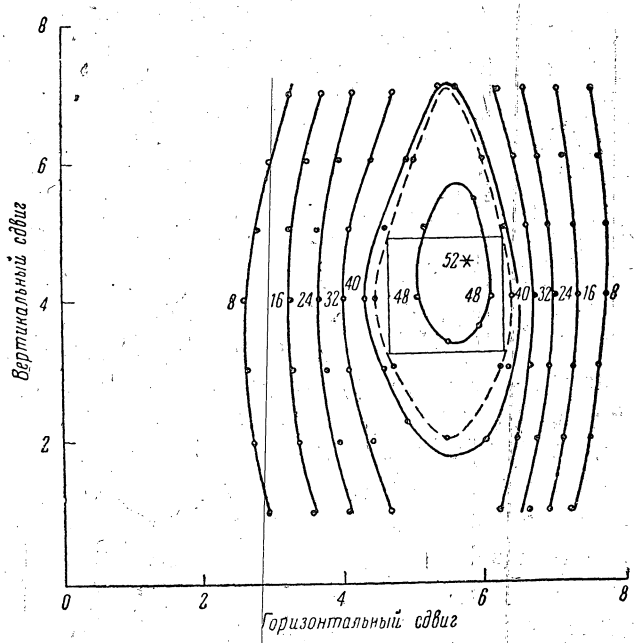
Фиг. 6

На рисунках построены квадраты, вписанные в кривые уровней, соответствующих 80% от максимума. Эти кривые нанесены пунктиром. Очевидно, что при шаге сдвига, превышающем размер стороны квадрата, возможно неправильное распознавание. Как видно из рисунков, сторона наименьшего квадрата составляет 0.83 шага растра.

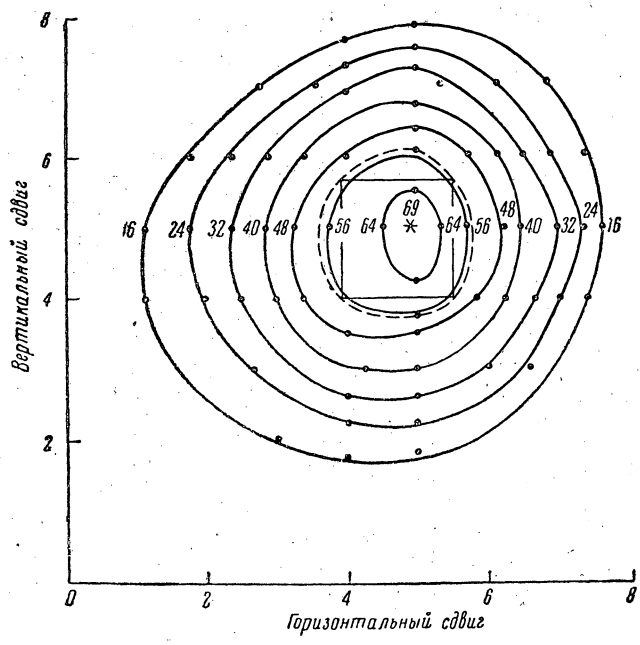
Исходя из приведенных соображений, можно сделать тот вывод, что для надежного распознавания необходимо выбрать шаг сдвигов меньшим, чем 0.83 шага растра. При этом даже в самом неблагоприятном случае $R^k_{\xi_{\max}}$, соответствующий правильному значению индекса k, будет наибольшим.



Фиг. 2*а*



Фиг. 2*б*



Фиг. 2в

Нужно отметить, что исследовались автокорреляционные функции хорошо напечатанных цифр. Автокорреляционные функции цифр, подвергнувшихся искажениям, могут иметь иной вид. Поэтому выполненные исследования позволяют установить только верхний предел для величины шага сдвига. Нами выбран шаг сдвига, равный половине шага растра. Как показали описанные ниже статистические исследования, при таком шаге достигается высокая надежность распознавания.

Таблица 2

Эталоны	5	0	7
0	65 342	53 416	75 774
1	67 166	40 606	73 511
2	33 574	37 630	61 105
3	70 256	41 570	100 744
4	71 152	27 004	125 070
5	102 630	41 472	63 423
6	70 470	47 410	72 502
7	70 376	41 366	135 770
8	61 146	50 252	64 350
9	46 354	45 172	121 402

С целью исследования помехоустойчивости алгоритма была проведена серия экспериментов по распознаванию цифр, подвергнутых искусственным искажениям.

Изучались три типа дефектов цифр: искусственные непропечатки части знака (фиг. 3), очень сильные естественные искажения знака (фиг. 4), искусственные загрязнения (фиг. 5).

В табл. 2 приведены результаты распознавания некоторых из этих цифр. Последние выделены на соответствующих рисунках.

Для каждой из цифр приводятся ответы распознавания и десять наибольших по ξ корреляционных коэффициентов R_{ξ}^{k}, соответствующих эталонам от «0» до «9». Все числа даны в восьмиричной системе счисления.

С помощью описанной выше программы были проведены статистические исследования надежности распознавания на 35 000 машинописных цифр.

При исследовании изучались цифры, напечатанные на бумаге пяти сортов: газетной, белой проклеенной, белой непроклеенной, желтого и зелено-

го цветов. На бумаге каждого сорта цифры печатались в четырех экземплярах (один основной экземпляр и три копии). Процесс получения копий повторялся трижды с помощью одной и той же копировальной бумаги. Таким образом изучалось влияние степени изношенности копировальной бумаги на качество печати цифр в машинописных копиях. Цифры печатались машинисткой в случайных последовательностях, чтобы исключить влияние «обучения»» машинистки на силу удара по клавишам пишущей машинки.

Естественно, что на распознавание каждой из 35 000 цифр было бы затрачено недопустимо много машинного времени. В этом и нет необходимости. В процессе работы с алгоритмом был накоплен достаточно большой опыт по распознаванию машинописных цифр с самыми разнообразными дефектами печати.

Благодаря этому можно было с уверенностью считать, что большинство цифр, напечатанных сравнительно четко, будут правильно распознаны. Достаточно было путем тщательного просмотра всех 35 000 цифр выбрать несколько сот таких, относительно которых нельзя было ручаться за правильность распознавания, и ограничиться исследованием только этих цифр.

На фиг. 6 в качестве примера приведена часть страницы одной из четвертых машинописных копий, использовавшихся при статистических исследованиях надежности распознавания. Все цифры на приведенном рисунке были правильно распознаны.

Величина порога распознавания была принята равной нулю.

В результате исследований были получены на 35 000 цифр 2 неверных ответа и 3 отказа. Отказы объяснялись тем, что изображения были очень бледными и градации коэффициента отражения не превышали шага квантования. Изображение в этом случае воспринимается как белое поле.

Сбои и отказы распределены следующим образом:

первый экземпляр — нет сбоев и отказов;

второй экземпляр — один сбой;

третий экземпляр — один сбой и один отказ;

четвертый экземпляр — один отказ.

Выводы

1. Изучавшийся метод распознавания дает весьма высокую надежность распознавания в условиях очень сильных искажений знаков при печати.

2. Вероятность сбоев при чтении первого экземпляра текста имеет порядок менее 10^{-4}.

3. Надежность при чтении машинописных копий значительно выше, чем в других описанных методах распознавания. Вероятность ошибки имеет порядок 10^{-4}.

Поступила в редакцию
30.03.1962

Цитированная литература

1. В. А. Ковалевский. Корреляционный метод распознавания изображений. Ж. вычисл. матем. и матем. физ., 1962, 2, № 4, 684—694.
2. В. М. Глушков, В. А. Ковалевский, В. И. Рыбак. Универсальная установка для исследования алгоритмов распознавания изображений. В сб. «Принципы построения самообучающихся систем». Киев, Гостехиздат УССР, 1962.
3. Дж. В. Бэкус и др. (под ред. Петера Наура). Сообщение об алгоритмическом языке АЛГОЛ-60. Ж. вычисл. матем. и матем. физ., 1961, 1, № 2, 308—342.