



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ,
ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21), (22) Заявка: 2008144384/09, 10.11.2008

(24) Дата начала отсчета срока действия патента:
10.11.2008

(43) Дата публикации заявки: 20.05.2010

(45) Опубликовано: 10.11.2010 Бюл. № 31

(56) Список документов, цитированных в отчете о
поиске: RU 2260208 C2, 10.09.2005. RU 2259592
C2, 27.08.2005. JP 8137990 A, 31.05.1996. GB
1345686 A, 30.01.1974. JP 11353413 A,
24.12.1999. RU 2178916 C2, 27.01.2002. US
2005/0001033 A1, 06.01.2005. US 2008/0166055
A1, 10.07.2008.

Адрес для переписки:

456770, Челябинская обл., г. Снежинск, ул.
Васильева, 13, а/я 245, ФГУП "РФЯЦ-
ВНИИТФ им. академ. Е.И. Забабахина",
ОИС, Г.В. Бакалову

(72) Автор(ы):

Подгорнов Владимир Аминович (RU),
Подгорнов Семен Владимирович (RU),
Щербина Александр Николаевич (RU)

(73) Патентообладатель(и):

Федеральное государственное унитарное
предприятие "Российский Федеральный научно-
исследовательский институт технической
физики имени академика Е.И. Забабахина"
(ФГУП "РФЯЦ-ВНИИТФ им. академ. Е.И.
Забабахина") (RU)

(54) СПОСОБ АВТОМАТИЗИРОВАННОГО РАСПОЗНАВАНИЯ ИДЕНТИФИКАЦИОННОГО
НОМЕРА

(57) Реферат:

Изобретение относится к способам
распознавания знаков с применением
компьютерных средств и может быть
использовано для автоматизированного
определения поврежденных номеров
отработанных тепловыделяющих сборок.
Техническим результатом является повышение
достоверности при распознавании
поврежденных номеров с неизвестным заранее
количеством и шрифтом используемых
цифровых символов по их сохранившимся
фрагментам. Для этого формируют исходное
изображение предполагаемой области
нахождения идентификационного номера,
выделяют прямоугольную область его
размещения, выявляют в ней зоны нахождения
отдельных позиций номера, распознают
символы на выявленных позициях. При этом
предварительно сформированное исходное

изображение очищают с помощью
морфологического фильтра, определяют
координаты приблизительных центров каждой
позиции идентификационного номера в
выявленных зонах их нахождения, определяют
по сохранившимся фрагментам символов
номера координаты каждой позиции и
приблизительный параметр шрифта,
использованного в процессе нанесения
идентификационного номера, формируют
первоначальный набор шаблонов символов с
приблизительным шрифтом распознаваемого
номера. В процессе распознавания символов
осуществляют последовательное наложение
шаблонов с изменяющимися параметрами
шрифта символов на зоны нахождения позиций
номера с циклическим смещением шаблонов
относительно найденных приблизительных
центров позиций и производят сравнительный
анализ, в ходе которого для каждой позиции

выявляют шаблон, имеющий максимальное значение коэффициента совпадения с

фрагментами символа, находящегося на данной позиции. 4 з.п. ф-лы, 3 ил.

R U 2 4 0 3 6 1 7 C 2

R U 2 4 0 3 6 1 7 C 2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY,
PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION(21), (22) Application: **2008144384/09, 10.11.2008**(24) Effective date for property rights:
10.11.2008(43) Application published: **20.05.2010**(45) Date of publication: **10.11.2010 Bull. 31**

Mail address:

**456770, Cheljabinskaja obl., g. Snezhinsk, ul.
Vasil'eva, 13, a/ja 245, FGUP "RFJaTs-VNIITF
im. akadem. E.I. Zababakhina", OIS, G.V.
Bakalovu**

(72) Inventor(s):

**Podgornov Vladimir Aminovich (RU),
Podgornov Semen Vladimirovich (RU),
Shcherbina Aleksandr Nikolaevich (RU)**

(73) Proprietor(s):

**Federal'noe gosudarstvennoe unitarnoe
predpriyatje "Rossijskij Federal'nyj Jadernyj
Tsentr - Vserossijskij nauchno-issledovatel'skij
institut tekhnicheskij fiziki imeni akademika
E.I. Zababakhina" (FGUP "RFJaTs-VNIITF im.
akadem. E.I. Zababakhina") (RU)**

(54) METHOD FOR AUTOMATIC RECOGNITION OF IDENTIFICATION NUMBER

(57) Abstract:

FIELD: information technology.

SUBSTANCE: initial image of the area in which the identification number must be found is formed, a rectangular area is selected, zones for finding separate positions of the number are detected in the area and characters are identified on the detected positions. The pre-formed initial image is cleaned using a morphological filter and coordinates of approximate centres of each position of the identification number in their detection zones are determined, coordinates of each position and the approximate font parametre used in printing the identification number are determined from the stored fragments of characters of the number, an initial set

of templates of characters of the approximate font of the recognised number is formed. During recognition of characters, templates with varying font parametres are superimposed on the zones where positions of the number are found with cyclic shifting of templates relative the found approximate centres of the positions, and comparative analysis is carried out, during which a template is identified for each position, having maximum value of coefficient of coincidence with fragments of a character located at the given position.

EFFECT: more reliable recognition of damaged numbers with an unknown number and font of the digital characters used on stored fragments thereof.

5 cl, 3 dwg

Изобретение относится к способам распознавания знаков с применением компьютерных средств и может быть использовано для автоматизированного определения поврежденных номеров отработанных тепловыделяющих сборок.

Известен способ компьютерного распознавания объектов по патенту РФ № 2191431 от 03.12.99, МПК G06K 09/68, в соответствии с которым предварительно приводят изображение объекта, вводимое в компьютер, к нормальному, стандартному для данного способа виду путем изменения масштаба, поворота в требуемое положение, центрирования, вписания в прямоугольник требуемого размера, на экран монитора выводят изображение распознаваемого объекта, преобразованное в изображение, выполненное в градациях - различных степенях яркости одного цвета, например красного, и на него последовательно, поочередно накладывают изображения хранящихся в памяти компьютера шаблонов, выполненных, например, в градациях зеленого, что позволяет увидеть в зоне перекрытия изображений изображение другого, отличного от первых двух цвета, которое и фиксируют как распознанное в случае тождественных, идентичных, а значит имеющих одинаковый контур изображений распознаваемого объекта и шаблона. В принципе распознаваемым объектом может быть цифровой или буквенный символ.

Недостаток в трудности визуального распознавания объекта, особенно в том случае, когда объект претерпел значительные изменения по отношению к своему первоначальному виду, сохраняемому в памяти как один из шаблонов. Кроме того, в рассматриваемом способе геометрические характеристики рассматриваемого объекта стандартизированы, но в реальности данные параметры (например, вид шрифта и его высота) могут варьироваться.

Известен способ автоматического распознавания электронными средствами символов, таких как буквы и/или цифры, отпечатанных на любом материале, включающем структуры с сильной контрастностью (патент РФ № 2249251 от 20.06.2000, МПК G06K 09/80, 09/62). Этот способ включает в себя следующие этапы: отладка, формирование моделей символов, распознавание, регистрация модели фона вместе с фоном считанного изображения, выделение модели зарегистрированного фона из элементарного изображения фона, комбинирование для каждого положения символа модели букв и/или цифр с элементарным изображением соответствующего фона, создание комбинированных моделей, сопоставление неизвестных символов с комбинированными моделями, распознавание каждого неизвестного символа как соответствующего символа, комбинированная модель которого наилучшим образом накладывается на него в соответствии с технологией "сравнение с шаблоном".

Данный способ хорошо зарекомендовал себя в том случае, когда заранее известны исходные параметры комбинации символов, т.е. вид и размеры шрифта, количество символов, и нашел применение в распознавании номеров ценных бумаг, в частности денег. Однако распознавание становится крайне затруднительным, если речь идет об идентификационных номерах с неизвестными параметрами шрифта и фона.

В качестве прототипа выбран способ распознавания номерных знаков транспортных средств, описанный в статье с тем же названием, авторы: Иванов Г.А., Ларионов А.А., Панин Д.В., адрес в Интернете: [www.http://zntu.edu.ua/cmisis/foto/ivanov/ivan_lar_pan.doc](http://zntu.edu.ua/cmisis/foto/ivanov/ivan_lar_pan.doc).

Способ заключается в формировании исходного изображения предполагаемой области нахождения идентификационного номера, выделении прямоугольной области размещения идентификационного номера (номерной пластины), выявлении в выделенной области зон нахождения отдельных позиций идентификационного номера

и распознавании символов на выявленных позициях номера. При этом распознавание отдельных символов осуществляется с помощью набора нейтронных сетей, организованных в два каскада.

5 Данный способ разработан для распознавания номерных знаков транспортных средств различных стран Европейского союза, которые, в основном, имеют стандартизированное написание с использованием буквенно-цифровых символов единого вида (шрифта).

10 Проблема заключается в том, что при распознавании номера, о котором неизвестно ни количество используемых символов, ни используемый шрифт, который к тому же наносился без особых требований к прямолинейности и межсимвольным расстояниям и претерпел значительные повреждения в агрессивных условиях эксплуатации, данный способ не дает требуемого результата.

15 Задачей данного изобретения является создание способа автоматизированного распознавания идентификационного номера, позволяющего обеспечить максимально возможную достоверность при распознавании поврежденных номеров с неизвестным заранее количеством и шрифтом используемых цифровых символов. Зачастую такие маркировки имеют лишь сохранившиеся фрагменты символов, по которым
20 необходимо восстанавливать исходный символ.

Поставленная задача решается следующим образом.

В способе автоматизированного распознавания идентификационного номера, заключающемся в формировании исходного изображения предполагаемой области нахождения идентификационного номера, в выделении прямоугольной области
25 размещения идентификационного номера, выявлении в выделенной прямоугольной области зон нахождения отдельных позиций идентификационного номера и распознавании символов на выявленных позициях номера, согласно изобретению предварительно сформированное исходное изображение очищают с помощью
30 морфологического фильтра, а после выявления зон нахождения отдельных позиций идентификационного номера определяют по сохранившимся фрагментам символов координаты приблизительных центров каждой позиции и приблизительный параметр шрифта, использованного в процессе нанесения идентификационного номера, исходя из которого формируют первоначальный набор шаблонов символов с
35 приблизительным шрифтом распознаваемого номера. Процесс распознавания символов на выявленных позициях номера производят путем последовательного наложения шаблонов с изменяющимися параметрами шрифта символов на зоны нахождения позиций номера с циклическим смещением шаблонов относительно
40 найденных приблизительных центров позиций и сравнительного анализа, в ходе которого для каждой позиции идентификационного номера выявляют шаблон, имеющий максимальное значение коэффициента совпадения с фрагментами символа, находящегося на данной позиции.

45 Технический результат заключается в оптимизации процесса подгонки шаблонов к фрагментам символов идентификационного номера с неизвестными заранее параметрами. Подгонка осуществляется за счет изменения геометрических параметров шрифта шаблонов и прецизионного поиска оптимального места их расположения в зонах нахождения символов. При этом сам процесс подгонки
50 существенно облегчается за счет предварительной очистки изображения идентификационного номера от фона, благодаря чему фрагменты символов становятся более выделенными. В результате обеспечивается достоверность распознавания посредством выбора тех шаблонов, с которыми происходит

максимально возможное совпадение фрагментов символов по каждой позиции идентификационного номера.

Кроме того, в случае сомнительно распознанных позиций сравнительный анализ может включать эвристическую оценку, проводимую по характерным фрагментам символов, находящихся на данных позициях, что обеспечивает точность распознавания символов, незначительно отличающихся друг от друга.

Кроме того, в качестве приблизительного параметра шрифта идентификационного номера можно использовать характерную ширину символов, соответствующую ширине выявленных позиций. Параметр ширины символов является достаточно универсальным и подходит для всех видов шрифтов.

Кроме того, в процессе распознавания символов на выявленных позициях номера шрифт символов унифицированных шаблонов изменяют посредством уменьшения ширины шрифта, а также уменьшения отношения высоты шрифта к ширине, что увеличивает точность распознавания идентификационного номера, за счет точного совпадения шаблонов с фрагментами символов позиций номера.

Кроме того, при распознавании позиций номера поочередно используют позитивные и негативные шаблоны для каждого символа. Первый тип шаблонов содержит зоны, совпадающие с фрагментами символов, а второй тип шаблона содержит участки фона, не совпадающие с фрагментами символов. Благодаря этому увеличивается вероятность точного распознавания идентификационного номера, так как появляется возможность оценить соотношение между уровнем сигнала (яркости зон, где должны находиться сохранившиеся фрагменты символов) и фона (яркости зон, где не должны находиться фрагменты символов).

На фиг.1 показано исходное изображение предполагаемой области нахождения идентификационного номера.

На фиг.2 изображена прямоугольная область размещения идентификационного номера, выделенная на исходном очищенном изображении.

На фиг.3 представлен первоначальный набор шаблонов символов с приблизительным шрифтом распознаваемого номера.

Способ автоматизированного распознавания идентификационного номера включает в себя следующие этапы.

Вначале формируется исходное изображение предполагаемой области нахождения идентификационного номера (фиг.1). Затем данное изображение очищается морфологическим ориентирно-направленным фильтром, в результате чего частично устраняется искажающее влияние фона.

Из очищенного исходного изображения выделяется прямоугольная область размещения идентификационного номера (фиг.2). Границы области размещения номера могут определяться оператором или находиться либо по характерным реперам (например, краям лыски, на которую нанесен номер), либо по различию в освещенности символов и фона (символы светлее фона и наоборот).

После выявления области размещения номера производится поиск зон нахождения отдельных позиций идентификационного номера и определение по сохранившимся фрагментам символов координат приблизительных центров каждой позиции (фиг.2). С этой целью формируется вектор - строка AM_{row} , каждый j -й элемент которой содержит среднее арифметическое значение элементов j -ого вектор-столбца матрицы I_{filt_cut}

изображения области размещения идентификационного номера на отфильтрованном изображении:

$$AM_{row}(j) = \sum_{i=1}^{N_{row}} I_{filt_cut}(i, j), j = 1, 2, \dots, N_{col},$$

где N_{row} - количество строк в матрице I_{filt_cut} , а N_{col} - количество столбцов в I_{filt_cut} .

5 Из полученной вектор-строки AM_{row} формируется "сглаженная" вектор-строка AM_{row_av} , каждый j -й элемент которой содержит среднее арифметическое диапазона значений $[AM_{row}(j-\Delta), AM_{row}(j+\Delta)]$, при этом параметр Δ определяется заранее, исходя из средней степени поврежденности идентификационного номера. Далее, анализируя
10 вектор AM_{row_av} , определяют координаты $(X(k))$ приблизительных центров каждой позиции номера, например, в результате поиска максимумов функции $f(j) = AM_{row_av}(j)$ (фиг.2). Координата Y может определяться как середина
15 изображения I_{filt_cut} .

По сохранившимся фрагментам символов номера определяется приблизительный параметр шрифта, использованного в процессе нанесения идентификационного номера. В качестве приблизительного параметра шрифта может использоваться
20 характерная ширина символов W_x , соответствующая ширине выявленных позиций (фиг.2). Ширина выявленных позиций определяется в результате анализа положительной части функции $f(j) = AM_{row_av_cut}(j)$:

$$AM_{row_av_cut}(j) = AM_{row_av}(j) - \Delta_{bg},$$

25 где значение Δ_{bg} определяется заранее, исходя из средней степени поврежденности идентификационного номера.

Формируется первоначальный набор бинарных шаблонов символов с приблизительным шрифтом распознаваемого номера. Приблизительный шрифт
30 соответствует найденной характерной ширине W_x символов, а его высота W_y пропорциональна ширине W_x . Кроме того, при распознавании позиций номера могут использоваться позитивные (первый ряд) и негативные (второй ряд) шаблоны для каждого символа, фиг.3. Темным цветом на шаблонах отмечены области, которым присвоены значения 1, а светлым - 0. Первый тип шаблонов содержит ненулевые
35 области, совпадающие с фрагментами символов; а второй тип шаблона содержит ненулевые области, являющиеся участками фона, не совпадающего с фрагментами символов.

Проводится сравнительное распознавание символов на выявленных позициях номера. С этой целью последовательно на каждую позицию номера накладывают
40 поочередно все шаблоны. Наложение происходит посредством поэлементного перемножения матрицы текущего t -ого шаблона ($I_{sam_pos_t}$ - позитивного, $I_{sam_neg_t}$ - негативного) на матрицу (I_{pos_k}) такого же размера, содержащую прямоугольный
45 фрагмент изображения I_{filt_cut} , причем центр данного фрагмента совпадает с координатами $(X(k), Y)$ приблизительного центра текущей распознаваемой k -й позиции номера:

$$I_{res_pos_k_t}(i, j) = I_{sam_pos_t}(i, j) \cdot I_{pos_k}(i, j), \quad \text{при } 1 < i < W_y, 1 < j < W_x;$$

$$50 \quad I_{res_neg_k_t}(i, j) = I_{sam_neg_t}(i, j) \cdot I_{pos_k}(i, j), \quad \text{при } 1 < i < W_y, 1 < j < W_x.$$

Для обеспечения большей точности совпадения шаблонов с фрагментами символов позиций номера матрицы $I_{sam_pos_t}$, $I_{sam_neg_t}$ (т.е. шрифт символов шаблонов) могут

пропорционально изменяться посредством уменьшения ширины W_x , а также отношения высоты W_y к ширине W_x ; кроме того, при наложении шаблонов на позиции номера может осуществляться их незначительная циклическая, практически, прецизионная подвижка относительно друг друга, за счет соответствующего смещения центра матрицы I_{pos_k} относительно текущих координат $X(k)$, Y .

Далее для каждой полученной матрицы $I_{res_pos_k_t}$ $I_{res_neg_k_t}$ вычисляется среднее арифметическое значение ее элементов: для позитивных шаблонов - $am_pos(k, t)$, для негативных шаблонов - $am_neg(k, t)$. После чего, для каждого сопоставления t -ого шаблона с k -й позицией идентификационного номера вычисляется значение коэффициента ($k_ver(k, t)$) совпадения шаблона с фрагментами символа позиции номера. Значение коэффициента $k_ver(k, t)$ может быть вычислено, как соответствующая разность $am_pos(k, t) - am_neg(k, t)$.

В итоге для каждой k -й позиции номера находится t -й шаблон символа, имеющий из соответствующего множества максимальное значение коэффициента k_ver .

Кроме того, для сомнительно распознанных позиций дополнительно проводят эвристическую оценку. Например, для позиций номера, распознанных как 3, по характерным фрагментам производят их сравнение с цифрой 5. Если позиция распознается как цифра 3, то ее сравнивают с цифрой 8. Если позиция распознается как цифра 5, то ее сравнивают с цифрой 6. В другой ситуации для позиций номера, распознанных как 8, по характерным фрагментам производят их сравнение с цифрой 6. Если позиция распознается как цифра 8, то ее сравнивают с цифрой 2 и т.д.

Для описанного алгоритма разработана компьютерная программа. Способ обеспечивает высокую точность дистанционного автоматического распознавания маркировочных номеров отработанных тепловыделяющих сборок и предназначен для использования в системах учета и контроля отработавших ядерных материалов в ядерной энергетике.

Формула изобретения

1. Способ автоматизированного распознавания идентификационного номера, заключающийся в формировании исходного изображения предполагаемой области нахождения идентификационного номера, в выделении прямоугольной области размещения идентификационного номера, в выявлении в выделенной прямоугольной области зон нахождения отдельных позиций идентификационного номера и в распознавании символов на выявленных позициях номера, отличающийся тем, что предварительно сформированное исходное изображение очищают с помощью морфологического фильтра, а после выявления зон нахождения отдельных позиций идентификационного номера определяют по сохранившимся фрагментам символов номера координаты приблизительных центров каждой позиции и приблизительный параметр шрифта, использованного в процессе нанесения идентификационного номера, исходя из которого формируют первоначальный набор шаблонов символов с приблизительным шрифтом распознаваемого номера, при этом процесс распознавания символов на выявленных позициях номера производят путем последовательного наложения шаблонов с изменяющимися параметрами шрифта символов на зоны нахождения позиций номера с циклическим смещением шаблонов относительно найденных приблизительных центров позиций, и сравнительного анализа, в ходе которого для каждой позиции выявляют шаблон, имеющий максимальное значение коэффициента совпадения с фрагментами символа,

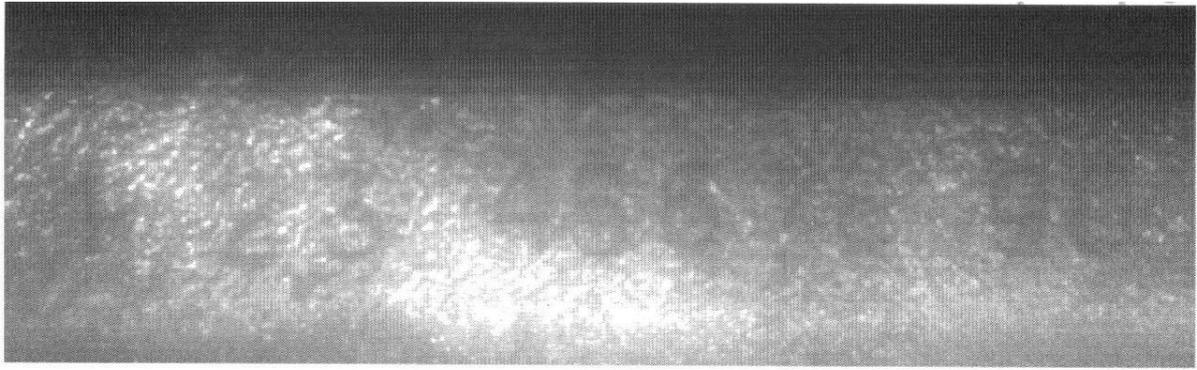
находящегося на данной позиции.

2. Способ автоматизированного распознавания идентификационного номера по п.1, отличающийся тем, что в сравнительный анализ вводят эвристическую оценку сомнительно распознанных позиций, проводимую по характерным фрагментам символов, находящихся на данных позициях.

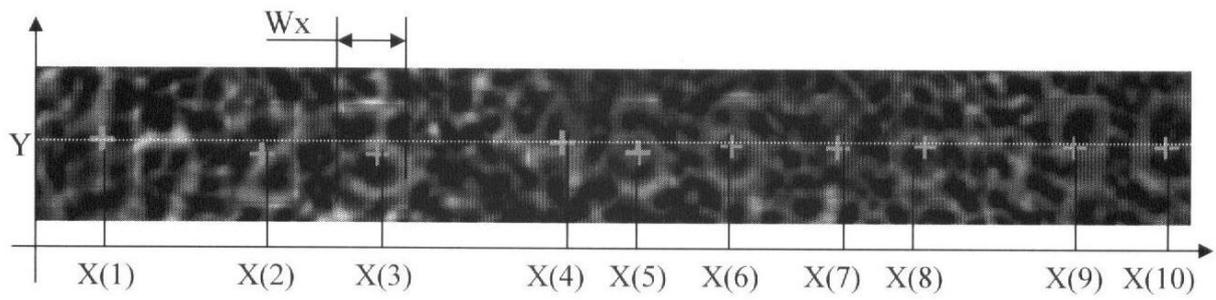
3. Способ автоматизированного распознавания идентификационного номера по п.1, отличающийся тем, что в качестве приблизительного параметра шрифта идентификационного номера используют характерную ширину символов, соответствующую ширине выявленных позиций.

4. Способ автоматизированного распознавания идентификационного номера по п.3, отличающийся тем, что в процессе распознавания символов на выявленных позициях номера шрифт символов шаблонов изменяют посредством уменьшения ширины шрифта и отношения высоты шрифта к ширине.

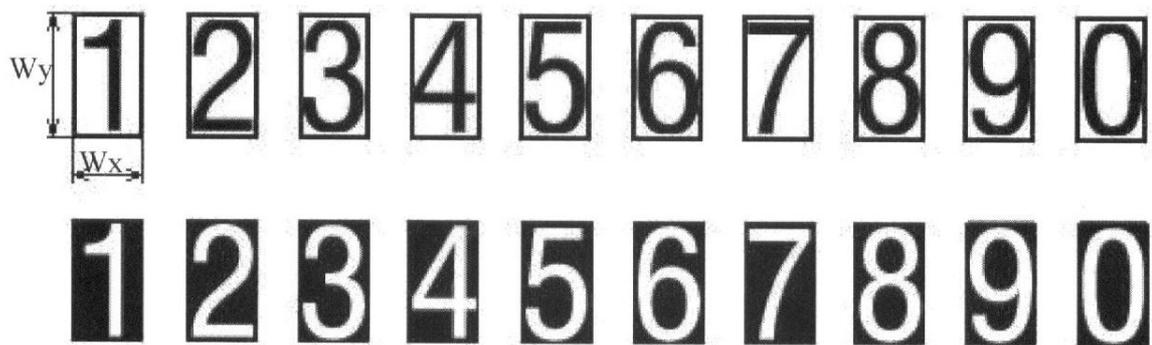
5. Способ автоматизированного распознавания идентификационного номера по п.1, отличающийся тем, что в процессе распознавания символов на выявленных позициях номера поочередно используют позитивные и негативные шаблоны символов.



Фиг.1



Фиг.2



Фиг.3