



ФЕДЕРАЛЬНАЯ СЛУЖБА
 ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ,
 ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21), (22) Заявка: **2002133900/09**, 17.12.2002

(24) Дата начала действия патента: **17.12.2002**

(43) Дата публикации заявки: **10.07.2004**

(45) Опубликовано: **10.05.2005** Бюл. № 13

(56) Список документов, цитированных в отчете о поиске: **US 5680479 A**, 21.10.1997. **SU 440680 A**, 25.08.1974. **RU 2178203 C1**, 10.01.2002. **EP 1063606 A2**, 27.12.2000. **US 6052481 A**, 18.04.2000. **US 5828783 A**, 27.10.1998. **US 5647022 A**, 08.07.1997. **US 5257323 A**, 26.10.1993. **US 5218538 A**, 08.06.1993.

Адрес для переписки:
129301, Москва, а/я 54, пат.пов. С.А.Платонову

(72) Автор(ы):

**Туганбаев Д.А. (RU),
 Дерягин Д.Г. (RU)**

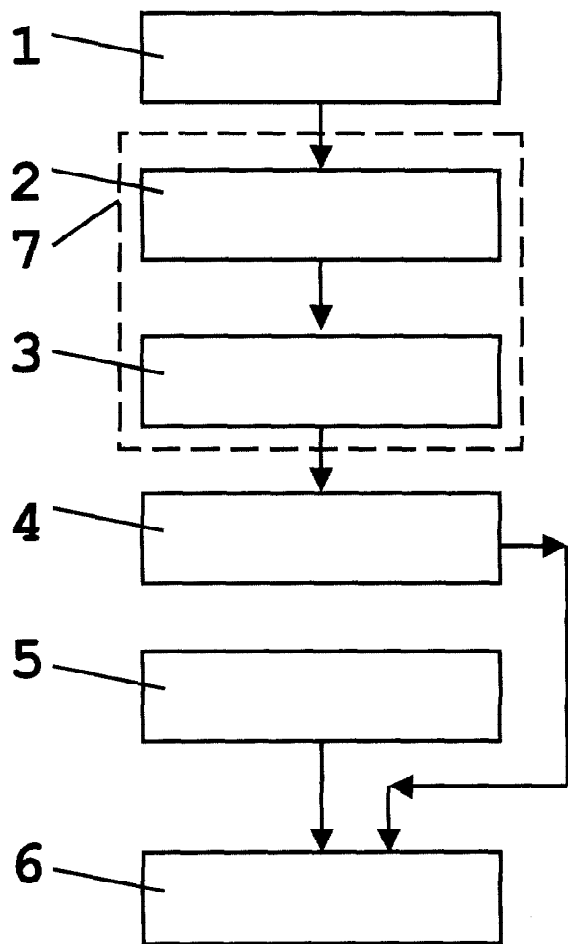
(73) Патентообладатель(ли):
"Аби Софтвер Лтд." (CY)

(54) СПОСОБ ИДЕНТИФИКАЦИИ ЗАЧЕРКНУТЫХ СИМВОЛОВ ПРИ РАСПОЗНАВАНИИ РУКОПИСНОГО ТЕКСТА

(57) Реферат:

Изобретение относится к оптическому распознаванию символов. Его использование при распознавании рукописного текста из растровых изображений обеспечивает технический результат в виде повышения правильности и помехозащищенности распознавания рукописных символов за счет идентификации зачеркнутых символов. Способ характеризуется делением изображения на области, выявлением областей с рукописными символами, применением структурных и признаковых классификаторов для распознавания символов, использованием структурного классификатора как основного

инструмента распознавания, выбором наиболее подходящего из нескольких вариантов символов. Технический результат достигается благодаря тому, что распознавание символа включает в себя распознавание символа по крайней мере одним дополнительным признаковым классификатором символов зачеркивания, осуществляющим одновременное сравнение с символом зачеркивания и по крайней мере одним похожим на него обычным символом, и идентификацию символа как зачеркнутого в случае большего соответствия признакам символов зачеркивания. 1 ил.





FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY,
PATENTS AND TRADEMARKS

(12) **ABSTRACT OF INVENTION**

(21), (22) Application: **2002133900/09, 17.12.2002**

(24) Effective date for property rights: **17.12.2002**

(43) Application published: **10.07.2004**

(45) Date of publication: **10.05.2005 Bull. 13**

Mail address:
129301, Moskva, a/ja 54, pat.pov. S.A.Platonovu

(72) Inventor(s):
**Tuganbaev D.A. (RU),
Derjagin D.G. (RU)**

(73) Proprietor(s):
"Abi Softver Ltd." (CY)

(54) **METHOD FOR IDENTIFICATION OF CROSSED SYMBOLS DURING RECOGNITION OF HAND-WRITTEN TEXT**

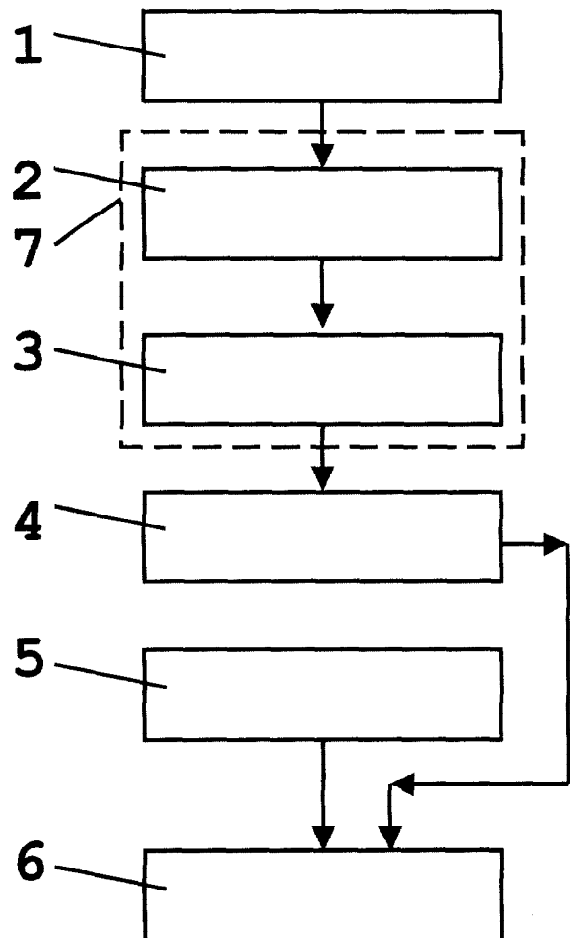
(57) Abstract:

FIELD: optical recognition of symbols.

SUBSTANCE: method includes dividing image on areas, finding areas with hand-written symbols, using structural and sign classifiers for recognition of symbols, use of structure classifier as main recognition tool, selecting best suiting symbol of several variants.. recognition of symbol includes recognition of symbol by at least one additional sign classifier of crossed symbols, performing concurrent comparison to crossed symbol and at least one common symbol like the latter, and identification of symbol as crossed one in case of better compliance to signs of crossed symbols.

EFFECT: higher efficiency.

1 dwg



Изобретение относится к области оптического распознавания символов и, в частности, к способам распознавания рукописного текста из растрового изображения.

Известен способ распознавания текстовой информации, при котором растровое изображение разбивают на области, предположительно содержащие изображения символов текста, с последующим сопоставлением изображения в областях с эталонным изображением, содержащимся в одном или нескольких дополнительных признаковых классификаторах.

Такой способ реализуется в способе распознавания символов по патенту США №5680479, Октябрь 21, 1997.

Недостатком этого способа является то, что в случае распознавания символа зачеркивания результатом является один из символов, подходящий по очертаниям с наибольшей степенью надежности на рисунок линий зачеркивания. Это приводит к ошибке распознавания, поскольку практически всегда этот символ не совпадает с тем, который зачеркнут. Кроме того, известные способы не воспринимают такой результат как ошибку распознавания.

Известный способ не позволяет идентифицировать наличие зачеркнутого символа, что приводит к ошибкам распознавания.

Техническим результатом изобретения является повышение правильности распознавания рукописных символов, повышение помехозащищенности распознавания, обеспечение возможности идентификации зачеркнутых символов.

Это достигается тем, что в системе организуют дополнительный признаковый классификатор, содержащий признаки множества разновидностей символов зачеркивания. При достаточно надежном совпадении с элементом указанного классификатора, символ классифицируют как зачеркнутый и направляют на дополнительную обработку.

Реализация этого способа позволяет существенно повысить качество распознавания рукописных символов, улучшить помехозащищенность распознавания.

Указанные недостатки значительно ограничивают возможности использования известных способов для распознавания текстовой информации.

Известные способы непригодны для достижения заявленного технического результата.

Предлагаемый способ отличается тем, что для распознавания зачеркнутых символов рукописного текста из растрового изображения в системе организуют один обобщенный или несколько частных дополнительных классификаторов, содержащих признаки символов зачеркивания, и используют иную последовательность обработки символа.

Общий классификатор обучают на всем известном наборе признаков зачеркивания. Частные классификаторы обучают на однотипном массиве признаков. Такие классификаторы позволяют выбрать между символом зачеркивания и похожим на него обычным символом. Наиболее типичные символы, сходные с символами зачеркивания, на базе которых создают частные классификаторы, - "3", "З", "8", "Ж" и др.

Распознавание зачеркнутого символа начинают с применения структурных или признаковых классификаторов символов. При появлении многозначного результата, обращаются к одному или более дополнительному признаковому классификатору для уменьшения числа предполагаемых символов. При достижении единственного результата распознанного символа организуют проверку одним общим или несколькими частными дополнительными классификаторами, содержащими признаки различных видов зачеркивания. При достаточно близком совпадении с элементом дополнительного классификатора символ идентифицируют как зачеркнутый и направляют его на дополнительную обработку. Дополнительная обработка может состоять в привлечении оператора для распознавания или в применении полностью автоматизированного способа.

Сущность предложения иллюстрируется на чертеже.

Объектом обработки является графическая область с изображением зачеркнутого символа, выделенная из растрового изображения документа.

Изображение символа поступает в один или последовательно в несколько структурных классификаторов для проведения операции распознавания. Структурные классификаторы

являются в заявляемом способе предпочтительным основным инструментом распознавания символов. В результате распознавания образуются один или более вариантов возможных символов вместе с уровнями доверия каждому варианту. В случае множественного результата распознавания, полученные варианты распознанных символов
5 далее направляются на проверку в один или несколько признаковых классификаторов. Признаковые классификаторы являются в заявляемом способе основным инструментом для уменьшения числа предполагаемых вариантов распознанных символов. Признаковые классификаторы проводят распознавание способами, отличными от структурных классификаторов, что при одновременном использовании классификаторов обоих видов
10 значительно повышает общую надежность распознавания. После проверки в структурных и признаковых классификаторах число возможных вариантов значительно уменьшается и повышается их надежность.

Далее проводят проверку символа в одном общем или нескольких частных дополнительных признаковых классификаторах, предварительно прошедших обучение на
15 символах зачеркивания. В результате работы дополнительного классификатора принимают решение об отнесении символа к зачеркнутым либо принятии варианта, полученного на предыдущих этапах. Символ, считающийся зачеркнутым, направляют на последующую дополнительную обработку, дополнительная обработка может быть полностью автоматизированной или требовать привлечения оператора.

20 Принцип работы следующий. Графическую область 1 с изображениями зачеркнутого символа направляют для анализа в один или более классификатор для распознавания символа (7).

В качестве более предпочтительной реализации используют классификаторы двух типов. Первичный анализ проводят с помощью по крайней мере одного основного
25 структурного классификатора 2. В случае появления множественного результата работы классификатора 2, варианты символов вместе с исходной графической областью 1 направляют для уточнения в один или более признаковых классификатор 3. Далее исходную графическую область 1, вместе с результатами распознавания на предыдущих этапах направляют в дополнительный признаковый классификатор символов зачеркивания
30 4.

При совпадении с достаточно высокой надежностью с элементом дополнительного классификатора 4 символ идентифицируют как зачеркнутый. Далее этот символ направляют на дополнительную обработку зачеркнутых символов 5, после чего распознавание символа завершают 6.

35

Формула изобретения

Способ идентификации зачеркнутых символов при распознавании рукописного текста, характеризующийся делением распознаваемого изображения на области, выявлением
40 областей, содержащих рукописные символы, применением структурных и признаковых классификаторов для распознавания и проверки правильности распознавания символов, использованием по крайней мере одного структурного классификатора как основного инструмента распознавания символов, выбором наиболее подходящего из нескольких вариантов символов, если на предыдущем этапе не удалось распознать символ
45 однозначно, отличающийся тем, что распознавание символа включает по крайней мере выполнение следующей последовательности действий: распознавание символа по крайней мере одним дополнительным признаковым классификатором символов зачеркивания, осуществляющим одновременное сравнение с символом зачеркивания и по крайней мере одним похожим на него обычным символом, идентификацию символа как зачеркнутого в случае большего соответствия признакам символов зачеркивания.

50