

(19)



**Евразийское
патентное
ведомство**

(21) **201700282** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки
2018.02.28

(22) Дата подачи заявки
2017.06.30

(51) Int. Cl. **G06K 9/64** (2006.01)
G06T 7/136 (2017.01)
G06F 17/30 (2006.01)

(54) **СПОСОБ АВТОМАТИЗИРОВАННОГО АНАЛИЗА РАСТРОВЫХ ИЗОБРАЖЕНИЙ**

(31) **2016134827**

(32) **2016.08.26**

(33) **RU**

(71) Заявитель:
**ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ
"ЛАБОРАТОРИЯ ИНФОВОТЧ" (RU)**

(72) Изобретатель:
**Рябов Сергей Сергеевич, Степанов
Виктор Сергеевич (RU)**

(74) Представитель:
Благополучная К.В. (RU)

(57) Изобретение относится к автоматизированному анализу растровых изображений. Техническим результатом является расширение арсенала технических средств за счет создания сравнительно быстрого и универсального способа, который позволяет выявлять в потоке данных/документах схожие с эталоном растровые изображения. Способ автоматизированного анализа растровых изображений заключается в том, что преобразуют в заранее заданный формат все электронные файлы эталонных растровых изображений, переводя цвета в градации серого и сжимая до квадрата малой площади порядка 20×20 пикселей. Сохраняют преобразованные электронные файлы эталонных растровых изображений в специализированной базе данных. Преобразуют каждый электронный файл анализируемого растрового изображения в заранее заданный формат, идентичный эталонному. Осуществляют поиск каждого пикселя изображения в специализированной базе данных и извлекают из неё информацию о том, в каких эталонных изображениях на данном месте встречается схожий пиксель. Для каждого найденного эталонного растрового изображения подсчитывается количество совпавших пикселей. Количество совпавших пикселей сравнивают с заранее заданным для каждого эталонного изображения порогом. Все эталоны растровых изображений, пороги которых превышены, считаются найденными.

A1

201700282

201700282

A1

Способ автоматизированного анализа растровых изображений

Область техники, к которой относится изобретение.

Настоящее изобретение относится к автоматизированному анализу растровых изображений и может быть использовано при разработке новых и совершенствовании существующих систем проверки растровых изображений на совпадение с эталонными изображениями.

Уровень техники.

В настоящее время весьма остро стоит проблема так называемого перехвата данных. Такая проблема может встретиться в случае отслеживания различных документов, в т.ч. содержащих изображения, проходящих по сети компании, на предмет наличия в них конфиденциальной информации.

В настоящее время известно несколько систем или способов, позволяющих решить эту проблему.

Например, в патенте России №2420800 (приоритет от 30.06.2009, опубликовано 10.06.2011) раскрыт способ поиска похожих по смысловому содержанию электронных документов, в котором задают правила формирования уникальных слов, взвешиваются уникальные слова и связи между ними, строят на основе этого семантическую сеть и сравнивают семантические сети документов. Этот способ достаточно трудоемок и пригоден лишь в ограниченной области.

Известен способ индексации и поиска цифровых изображений (патент на изобретение РФ №2510935, приоритет от 23.09.2011, опубликовано 10.04.2014). Способ, описанный в данном патенте, подходит для поиска изображений определённой категории, в том числе цветовой, но при этом не подходит для задачи поиска конкретного изображения в обширной базе данных, где может присутствовать много изображений с одинаковыми преобладающими цветами. Так как автор патента ставил задачу поиска изображений, подходящих под некоторое описание, он сам указал в качестве недостатка некоторых других методов тот факт, что они могут классифицировать изображения с похожими преобладающими цветами как совершенно различные. В то же время, заявленное изобретение решает задачу поиска конкретного изображения с поправкой на искажения, возникающие при его сохранении в разных форматах и разных масштабах. При такой постановке задачи

недостатком становится именно невозможность различить изображения с похожими характеристиками.

Раскрытие изобретения.

Таким образом, существует потребность в расширении арсенала технических средств за счет создания сравнительно быстрого и универсального способа, который позволил бы выявлять в потоке данных растровые изображения, схожие с эталонными и который бы преодолевал недостатки известных решений.

Для решения этой задачи и получения указанного технического результата в настоящем изобретении предложен способ автоматизированного анализа растровых изображений, заключающийся в том, что:

1. Создают специализированную базу данных эталонных изображений, специализация которой заключается в том, что данные из эталонных изображений хранятся специальным образом. Каждое эталонное изображение описывается тремя параметрами: растр (набор пикселей), порог срабатывания (R , задаётся в процентах от 1 до 100), размер диапазона допустимых интенсивностей пикселя (L). Первый параметр используется как эталонные данные, второй и третий влияют на качество поиска.

Для создания такой базы данных выполняют следующие шаги:

- 1) каждому эталонному изображению присваивают уникальный идентификатор;
- 2) все электронные файлы эталонных растровых изображений преобразуют в заранее заданный формат (bmp);
- 3) каждое эталонное изображение переводят в градации серого, где 1 пиксель кодируется 1 байтом, диапазон интенсивностей от 0 до 255;
- 4) каждое эталонное изображение ужимают до размера 20 на 20 пикселей;
- 5) для каждого пикселя строится допустимый диапазон интенсивностей, для этого строится массив допустимых интенсивностей размера L , каждый элемент которого представляет собой значение интенсивности вычисляемое по формуле:

$$k[i] = 1 - (L / 2) + i, \text{ где}$$

$k[i]$ — элемент массива допустимых интенсивностей,
 i — интенсивность в текущем пикселе,

$L/2$ — половина размера диапазона допустимых интенсивностей,
 i — номер текущего элемента массива допустимых интенсивностей, принимает значения от 0 до L . В граничных ситуациях, когда, например, разность интенсивности и $L/2$ меньше 0, отрицательные значения интенсивности отбрасывают. Так же поступают с верхней границей - значения интенсивности больше 255 так же отбрасывают. В этих случаях размер массива допустимых интенсивностей может быть меньше L .

- б) создают трехмерный массив первое измерение которого отражает все возможные значения интенсивности цвета (256 элементов), второе измерение отражает все пиксели изображения (400 элементов, при целевом размере изображения 20 на 20 пикселей), третье измерение будет содержать идентификаторы эталонов (размер заранее неизвестен, зависит от количества эталонных изображений).
 - 7) каждое подготовленное изображение раскладывают по массиву из п.6 следующим образом:
 - а) обходят все пиксели текущего эталона, позиция пикселя в изображении — это индекс второго измерения в массиве из п.1.6;
 - б) для текущего пикселя обходят все допустимые значения интенсивности, интенсивность — индекс первого измерения в массиве из п.1.6;
 - с) в массив из п.6 по индексу соответствующей интенсивности (п.1.7.б) и по позиции пикселя на изображении (п.1.7.а) добавляется идентификатор эталонного документа.
 - 8) создают вспомогательный массив, в котором ключом является уникальный идентификатор эталонного изображения, а значениями — счётчик найденных пикселей (инициализируется 0) и порог срабатывания эталона R .
2. Когда на анализ поступает файл, содержащий растровое изображение, то производят следующие действия:
- 1) изображение преобразуют в заранее заданный формат (bmp);
 - 2) изображение переводят в градации серого, где 1 пиксель кодируется 1 байтом, диапазон интенсивностей от 0 до 255;
 - 3) изображение ужимают до размера 20 на 20 пикселей;
 - 4) для каждого пикселя анализируемого изображения:
 - а) по интенсивности, которая является индексом первого измерения массива из п. 1.6 получают массив позиций, в которые входит пиксель с

- такой интенсивностью;
- b) по позиции пикселя в анализируемом изображении, которая является индексом в массиве полученном в п. 2.4.a, получают список идентификаторов эталонов.
 - c) в вспомогательном массиве из п.1.8 увеличивают счётчики найденных пикселей для эталонов полученных в п. 2.4.b.
- 5) обходят вспомогательный массив:
- a) счётчики найденных пикселей делят на размеры изображения (400) и умножают на 100, таким образом получают процент соответствия анализируемого изображения данному эталону;
 - b) сравнивают процент соответствия с соответствующими порогами;
 - c) если порог превышен помещают идентификатор эталонного документа в список найденных эталонов;
 - d) обнуляют счётчики найденных пикселей;
- б) возвращают массив найденных идентификаторов эталонных документов.

Настоящее изобретение может быть реализовано в любой вычислительной системе, например в персональном компьютере, на сервере и т.п.

Способ автоматизированного анализа растровых изображений по настоящему изобретению предназначен для осуществления так называемого копирайтного анализа (английский аналог - fingerprint detection), задачей которого является установление схожести растровых изображений с изображениями, переданным ранее в базу данных (библиотеку) в качестве эталонных.

Особенностью данного алгоритма является то, что параметрами R и L можно влиять на показатели полноты и качества алгоритма. Например, увеличивая значение L, можно повышать устойчивость к шуму и артефактам ценой повышенного риска ложно-положительных срабатываний. Параметр R отвечает за степень схожести изображений в рамках алгоритма.

Формула

1. Способ автоматизированного анализа растровых изображений, заключающийся в том, что:

- преобразуют в заранее заданный формат все электронные файлы эталонных растровых изображений, переводя цвета в градации серого и сжимая до квадрата малой площади порядка 20x20 пикселей;

- сохраняют преобразованные электронные файлы эталонных растровых изображений в специализированной базе данных;

- преобразуют каждый электронный файл анализируемого растрового изображения в заранее заданный формат, идентичный эталонному;

- осуществляют поиск каждого пикселя изображения в специализированной базе данных и извлекают из неё информацию о том, в каких эталонных изображениях на данном месте встречается схожий пиксель;

- для каждого найденного эталонного растрового изображения подсчитывается количество совпавших пикселей;

- количество совпавших пикселей сравнивают с заранее заданным для каждого эталонного изображения порогом;

- все эталоны растровых изображений, пороги которых превышены, считаются найденными.

2. Способ по п.1, в котором параметрами R и L влияют на показатели полноты и качества алгоритма, увеличивая значение L повышают устойчивость к шуму и артефактам, увеличивая риск ложно-положительных срабатываний.

ЕВРАЗИЙСКОЕ ПАТЕНТНОЕ ВЕДОМСТВО

ОТЧЕТ О ПАТЕНТНОМ
ПОИСКЕ(статья 15(3) ЕАПК и правило 42
Патентной инструкции к ЕАПК)

Номер евразийской заявки:

201700282

Дата подачи: 30 июня 2017 (30.06.2017) | Дата испрашиваемого приоритета: 26 августа 2016 (26.08.2016)

Название изобретения: Способ автоматизированного анализа растровых изображений

Заявитель: ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ЛАБОРАТОРИЯ ИНФОВОТЧ"

 Некоторые пункты формулы не подлежат поиску (см. раздел I дополнительного листа) Единство изобретения не соблюдено (см. раздел II дополнительного листа)

А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ: G06K 9/64 (2006.01)

G06T 7/136 (2017.01)

G06F 17/30 (2006.01)

Согласно Международной патентной классификации (МПК) или национальной классификации и МПК

Б. ОБЛАСТЬ ПОИСКА:

Минимум просмотренной документации (система классификации и индексы МПК)

G06F 1/00-17/50, G06T 1/00-17/30, G06K 9/00-9/82, G06Q 10/00-50/34, H04N 9/00-9/38, 13/00-13/04, 19/00-19/98, H03M 13/00-13/53M H04L 9/00-9/38, 12/00-12/955, H04H 60/00-60/98, G09G 5/00-5/42, G11B 7/00-7/30, 13/00-13/08, H04W 4/00-12/12, 40/00-48/20, 60/00-60/06, 68/00-68/12, 74/00-84/22, 92/00-99/00

Другая проверенная документация в той мере, в какой она включена в область поиска:

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
A	US 6418430 B1 (ORACLE INTERNATIONAL CORPORATION) 09.07.2002	1-2
A	US 5915250 A (VIRAGE, INC) 22.06.1999	1-2
A	US 2008/0243842 A1 (XEROX CORPORATION) 02.10.2008	1-2
A	US 2014/0195501 A1 (INTERACTIVE INTELLIGENCE, INC. et al.) 10.07.2014	1-2
A	US 2013/0258386 A1 (WAGNER DAVID et al.) 03.10.2013	1-2
A	RU 2528140C1 (ОАО "НПО КАРАТ") 10.09.2014	1-2

 последующие документы указаны в продолжении графы В данные о патентах-аналогах указаны в приложении

* Особые категории ссылочных документов:

"А" документ, определяющий общий уровень техники

"Е" более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее

"О" документ, относящийся к устному раскрытию, экспонированию и т.д.

"Р" документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета

"D" документ, приведенный в евразийской заявке

"Т" более поздний документ, опубликованный после даты

приоритета и приведенный для понимания изобретения

"Х" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности

"У" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории

"&" документ, являющийся патентом-аналогом

"L" документ, приведенный в других целях

Дата действительного завершения патентного поиска: 14 августа 2017 (14.08.2017)

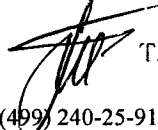
Наименование и адрес Международного поискового органа:

Федеральный институт

промышленной собственности

РФ, 125993, Москва, Г-59, ГСП-3, Бережковская наб., д. 30-1. Факс: (499) 243-3337, телетайп: 114818 ПОДАЧА

Уполномоченное лицо:



Т. М. Иванова

Телефон № (499) 240-25-91