

(19)



**Евразийское
патентное
ведомство**

(11) **042143**

(13) **B1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2023.01.19

(21) Номер заявки
201700282

(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**

(43) **2018.02.28**

(71)(73) Заявитель и патентовладелец:

**ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ
"ЛАБОРАТОРИЯ ИНФОВОТЧ" (RU)**

(72) Изобретатель:

**Рябов Сергей Сергеевич, Степанов
Виктор Сергеевич (RU)**

(74) Представитель:

Благополучная К.В. (RU)

(56) US-B1-6418430
US-A-5915250
US-A1-20080243842
US-A1-20140195501
US-A1-20130258386
RU-C1-2528140

(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.

B1

042143

042143

B1

Область техники, к которой относится изобретение

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 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, задается 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 высчитываемое по формуле:

$$k[i] = 1 - (L / 2) + i,$$

где $k[i]$ - элемент массива допустимых интенсивностей;

1 - 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 также отбрасывают. В этих случаях размер массива допустимых интенсивностей может быть меньше L.

6) 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.1.6 следующим образом:

а) обходят все пиксели текущего эталона, позиция пикселя в изображении - это индекс второго из-

мерения в массиве из п.1.6;

b) для текущего пикселя обходят все допустимые значения интенсивности, интенсивность - индекс первого измерения в массиве из п.1.6;

с) в массив из п.1.6 по индексу соответствующей интенсивности (п.1.7.b) и по позиции пикселя на изображении (п.1.7.a) 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:

a) 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;

с) во 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;

с) если порог превышен, помещают идентификатор эталонного документа в список найденных эталонов;

d) обнуляют счётчики найденных пикселей;

б) возвращают массив найденных идентификаторов эталонных документов.

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

Способ автоматизированного анализа растровых изображений по изобретению предназначен для осуществления, так называемого, копирайтного анализа (английский аналог - fingerprint detection), задачей которого является установление схожести растровых изображений с изображениями, переданными 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 ложноположительных срабатываний. Параметр R отвечает за степень схожести изображений в рамках алгоритма.

ФОРМУЛА ИЗОБРЕТЕНИЯ

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

- каждому эталонному изображению присваивают уникальный идентификатор;
- преобразуют в заранее заданный формат все электронные файлы эталонных растровых изображений, переводя цвета в градации серого и сжимая до квадрата размером 20 на 20 пикселей, вычисляя интенсивность цвета каждого пикселя;
- создают трехмерный массив, первое измерение которого отражает все возможные значения интенсивности цвета, второе измерение отражает позиции пикселей изображения с заданной в первом измерении интенсивностью, третье измерение содержит идентификаторы эталонов;
- 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 пиксель такой же интенсивности;
- 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.

