Секция 2

МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ И СТЕГАНОГРАФИИ

УДК 004.93'12

О ПОИСКЕ ПОДОБНЫХ ИЗОБРАЖЕНИЙ ПРИ ОБНАРУЖЕНИИ ЦВЗ

Б. Б. Борисенко

В исследуемом методе обнаружения цифровых водяных знаков (ЦВЗ) [1] классификация характеристических векторов производится на основе модифицированной контрольной карты Хотеллинга (МККХ). Достоинством данного метода является необходимость обучения только на выборке пустых контейнеров. При обнаружении ЦВЗ в графических контейнерах для каждого тестируемого контейнера на обучающую выборку накладываются определённые требования. Изображения обучающей выборки должны максимально походить на тестируемое, то есть:

— принадлежать к одному типу изображений (пейзаж, средний план, цветовое поле и т. п.);
— иметь примерно одинаковую гистограмму цветов и одинаковый коэффициент качества).

Суть метода обнаружения ЦВЗ заключается в применении элементов метода контрольных карт, используемого для выявления производственных сбоев, адаптированного под использование задач обнаружения ЦВЗ. При этом:

— на основе обучающей выборки пустых контейнеров вычисляется контрольная граница для классификации. Порог находится из распределения Фишера;
— на основе признаков тестируемого контейнера и обучающей выборки пустых контейнеров вычисляется значение модифицированной статистики Хотеллинга для последующей классификации;
— значение статистики сравнивается с вычисленным порогом и выдается результат о принадлежности к классу пустых контейнеров или стегоконтейнеров.

Результат обнаружения ЦВЗ на основе МККХ существенно зависит от содержимого обучающей выборки. Поэтому важным является использование эффективного алгоритма подбора пустых изображений для обучающей выборки с указанными выше ограничениями и автоматизации поиска таких изображений, так как с увеличением объёма обучающей выборки возрастает вероятность обнаружения [2].

В цифровой обработке изображений поставленная задача известна как задача семантической классификации [3]. Теория распознавания семантики изображения (в зарубежной литературе используется аббревиатура CBIR — Content-based image retrieval, поиск изображений на основе содержимого) позволяет решать, к сожалению, несколько другие задачи. В частности, в базе производится поиск изображений, содержащих конкретную фигуру либо шаблон с определёнными признаками.

В нашем же случае такие подобные изображения будут неприемлемы. Эмпирически было установлено, что метод на основе МККХ показывает лучшие результаты, если

значения соответствующих пикселей изображений обучающей выборки и тестируемого контейнера близки. При этом контекстное содержимое может отличаться.

Однако следует заметить, что направления исследований в обоих случаях совпадают, это:

— выделение признаков изображений;
— многомерное индексирование;
— проектирование систем поиска.

Изначально подобные изображения подбирались субъективно, то есть на основе личного мнения эксперта, составляющего обучающую выборку. Однако такой подход имеет ряд недостатков, среди которых можно упомянуть существенные временные затраты и возможные различия при составлении обучающей выборки разными экспертами. Следовательно, задача автоматизации поиска подобных изображений является актуальной.

Будем считать, что при поиске подобных изображений имеет место определённая относительность. То есть тестируемому изображению I_T более подобно изображение I_1 (I_T более похоже на I_1, чем изображение I_2), если $M(I_T, I_1) < M(I_T, I_2)$, где M — некоторая метрика. Изначально при автоматизации поиска подобных изображений в качестве M предлагалось использовать сумму среднеквадратических ошибок между каждым цветовым слоем пространства RGB изображения I_T и очередного изображения из общей базы, так как изменения цвета в этом пространстве отслеживаются лучше, чем в YUV (даже при анализе изображений формата JPEG). Однако такой подход является достаточно трудоёмким. Поэтому возникла задача вычисления некоторого хэш-образа от изображения для последующего сравнения значений метрики уже не от всего изображения, а от полученного хэш-образа. Поставленная задача была решена применением к изображению вейвлет-преобразования Хаара. При этом количество уровней преобразования определяется следующим образом: последний уровень имеет минимальный номер среди всех уровней, для которых хотя бы один из размеров матрицы коэффициентов аппроксимации не превышает 32.

Таким образом, для тестируемого изображения вычисляется хэш-образ, после чего вычисляются значения метрики для полученного хэш-образа и (уже имеющихся) хэш-образов базы изображений. Значения метрики упорядочиваются по неубыванию и отбирается необходимое количество изображений в обучающую выборку. Результаты стегоанализа на основе МККХ показывают целесообразность требований к изображениям, составляющим обучающую выборку, и эффективность предлагаемого подхода к автоматизации поиска таких изображений.

ЛИТЕРАТУРА

1. *Борисенко Б. Б.* Модификация карты Хотеллинга, нивелирующая влияние тренда, и её применение при обнаружении цифровых водяных знаков // Прикладная дискретная математика. 2010. № 2(8). С. 42–58.
2. *Filler T., Fridrich J., and Ker A. D.* The square root law of steganographic capacity for Markov covers // Proc. SPIE, Electronic Imaging, Security and Forensics of Multimedia XI, San Jose, CA. January 18–21, 2009. P. 212–223.
3. *Smith J. R. and Chang S. F.* Visualseek: A fully automated content based image query system // Proc. ACM Multimedia, Boston, MA. Nov. 1996.