ОПРЕДЕЛЕНИЕ ПОДЛИННОСТИ ЦИФРОВЫХ ФОТОГРАФИЙ НА ОСНОВЕ ХАРАКТЕРИСТИК ЦИФРОВЫХ УСТРОЙСТВ И ГРАФИЧЕСКИХ РЕДАКТОРОВ

Кадан А. М., Каспер П. С., Лазарь А. И., Радевич Н. А., Сенько Д. Ю., Шишкин Е. А.
Кафедра системного программирования и компьютерной безопасности,
Гродненский государственный университет имени Янки Купалы
Гродно, Республика Беларусь
E-mail: {kadan, kasper_ps_14, lazar_ai_14, radevich_na_14, senko_dj_13, shishkin_en_14}@mf.grsu.by

Представлен метод подтверждения подлинности цифровых фотографий, основанный на учете особенностей реализации формата JPEG, определяемых программно-технической реализацией цифровых фотокамер различных брендов и моделей. Работа метода основана на применении алгоритмов машинного обучения и позволяет определить бренд и модель цифровой фотокамеры, которой была сделана цифровая фотография, или определить графический редактор, которым в изображение были внесены изменения.

Введение

Цифровые фотокамеры и полученные с их помощью фотографии зачастую являются носителями криминалистически значимой информации. В то же время, количество различных способов мошенничества в сфере ИТ постоянно растет. Поэтому актуальной, и не только для специалистов в области компьютерной технической экспертизы (КТЭ), является задача обеспечения возможности подтверждения подлинности цифровых изображений. Подобные задачи актуальны, к примеру, в финансовых сферах, страховом деле, информационной безопасности, защите информации и криминалистике.

Под «подтверждением подлинности» будем понимать возможность определения бренда цифровой камеры и ее модели либо определение класса программного средства, с помощью которого было изменено оригинальное изображение.

I. Методы подтверждения подлинности JPEG-изображений

Традиционно в качестве атрибутов JPEG-изображений выделяются:

1. метаданные EXIF;
2. таблица Хаффмана, заранее записанная в устройство. Определяется производителем на основании характеристик устройства;
3. таблицы квантования, которые также записаны в устройство заранее и определяются производителем;
4. размеры эскиза характерные для устройств определенной марки и модели. При этом эскиз также имеет характерные для используемого устройства эскизные таблицы Хаффмана и квантования;
5. размер итогового изображения.

Если нужно решить, является ли цифровая фотография подлинной, проводится исследование предоставленных цифровых фотографий. Традиционно, эксперты применяют такие методы как [1]:

- визуальный метод;
- анализ EXIF-данных;
- эффект двойного квантования (double quantization effect);
- анализ уровня ошибки (error level analysis);
- копирование-перемещение фона (copy-move forgery);
- определение сетки артефакта блока (block artifact grid detection);
- определение качества изображения (image quality detection);
- интерполяция матрицы фильтров цветов (color filter array interpolation);
- базовый метод [2].

II. Различия в структуре файлов JPEG-изображений

Алгоритм JPEG в наибольшей степени пригоден для сжатия фотографий и изображений, содержащих реалистичные сцены с плавными переходами яркости и цвета. Необходимость аппаратной реализации алгоритма JPEG в цифровых фотокамерах привели к тому, что практически все его реализации имеют различия у различных производителей цифровых фотокамер.

Согласно требованиям стандарта ISO/IEC 10918-1 [3], файл JPEG содержит последовательность маркеров, каждый из которых начинается с байта 0xFF, свидетельствующего о начале маркера, и байта — идентификатора. Некоторые маркеры состоят только из этой пары байтов, другие же содержат дополнительные данные, состоящие из двухбайтового поля с длиной информационной части маркера (включая длину этого поля, но за вычетом двух байтов начала маркера т.е. 0xFF и идентификатора) и собственно данных.

Среди маркеров выделяют группу основных маркеров, которые присутствуют в реали-

зации JPEG-формата практически у всех разработчиков, и целый ряд дополнительных (не обязательных) маркеров.

Также в структуре JPEG-формата у различных производителей отличается типы используемых маркеров, количество вхождений маркеров одного типа, длина блока кода, связанного с маркером, порядок появления маркеров в файле и другие характеристики, также связанные с маркерами.

Различия в использовании маркеров наблюдаются не только в файлах различных брэндов, но также и для различных моделей одного и того же производителя, а также и для одних и тех же моделей при выборе различных режимов (настроек) фотосъемки. Использование графического редактора также вносит изменения в структуру исходного JPEG-файла, что, возможно, позволит идентифицировать программное средство, с помощью которого была нарушена подлинность исходного цифрового изображения.

III. Подготовка обучающей выборки

База цифровых изображений была получена путем выгрузки изображений с сайта [4], на котором представлены различные бренды и модели фотоаппаратов вместе с примерами оригинальных фотографий.

Полученная с сайта база включала более 1500 различных примеров бренд-модель и более 25000 оригинальных цифровых фотографий. Для ее получения были разработаны специальные программные модули на языке Python, позволившие автоматизировать работу. Также были созданы таблицы брендов, моделей бренда, фотографий, связанных с моделью бренда.

На основании структуры JPEG-файлов этих цифровых фотографий были сформированы два множества признаков (data set'ов) для применения методов машинного обучения.

В первый data set на основании анализа каждой из фотографий были отобраны 12 признаков: (1-2) Название камеры (бренд), название модели; (3-4) количество всех маркеров, количество маркеров начала 0xFFD8; (5-6) длина и количество таблиц квантования 0xFFDB; (7-8) длина и количество начала кадра, базового метода 0xFFC0; (9-10) длина и количество таблиц кодов Хаффмана 0xFFC4; (11-12) длина и количество начала закодированного изображения 0xFFDA.

Во второй data set были включены данные о 54 признаках на основе основных и дополнительных маркеров.

IV. Обучение моделей

Для построения автоматизированной системы подтверждения подлинности изображений на основе построенных обучающих выборок (решения задачи классификации) были обучены модели методов машинного обучения, построенные на основе алгоритмов наивного байесовского классификатора; решающих деревьев; k-ближайших соседей; случайного леса.

Качестово обучения было протестировано методом кросс-валидации. Сами алгоритмы машинного обучения были заимствованы из библиотеки sklearn компании Google.

В результате было установлено, что наилучший уровень предсказания имеет модель, основанная на алгоритме случайного леса – 87%.

Такой результат позволяет найти практическое применение в области информационной безопасности и криминалистике. Однако исследования в этой области имеют дальнейшие пути развития. Например, пополнение базы данных оригинальных фотографий, расширение количества классов марок и моделей фотоаппаратов, поиск новых характеристик для классификации и дополнение базы изображениями, измененными в различных редакторах.

V. Сервис проверки подлинности изображений

Разработан и функционирует как клиент-серверное приложение прототип сервиса проверки подлинности цифровых изображений. Сервис является локальным на компьютере эксперта, так как передавать в локальную сеть, а тем более в Интернет, криминалистически значимую информацию недопустимо.

База признаков, используемых для прогнозирования результата, – для определения подлинности цифрового изображения путем подтверждения (или отрицанияния) наличия у него характеристик, присущих фотографиям, сделанных фотокамерой конкретных бренда и модели, – может пополняться экспертом самостоятельно. Особое внимание уделено фотографиям, сделанным камерами цифровых мобильных устройств, с учетом их очень широкого распространения и весьма большого разнообразия.

1. Как проверить фотографию на подлинность [Электронный ресурс]. – Режим доступа: http://www.smtdp.com/ru/kak-proverit- fotografiyu-na- podlinnost/. – Дата доступа: 21.08.2017.
2. Метаданные в цифровой фотографии [Электронный ресурс]. – Режим доступа: http://www.ixbt.com/digimage/metadxph.shtml. – Дата доступа: 20.02.2017.
3. ISO/IEC 10918-1: 1993(E). Information technology – digital compression and coding of continuous-tone still images – requirements and guidelines [Электронный ресурс]. – Метод доступа: https://www.w3.org/Graphics/JPEG/itu-t81.pdf. – Дата доступа: 22.08.2017.
4. Digital Camera Reviews, Canon Cameras, Nikon Cameras, DSLR and SLR Cameras - Steves Digicams [Электронный ресурс]. – Метод доступа: www.steves-dificams.com/camera-reviews/. – Дата доступа: 22.08.2017.