ОБНАРУЖЕНИЕ ВСТРАИВАНИЙ НА ИЗОБРАЖЕНИЯХ ПУТЕМ АНАЛИЗА АРТЕФАКТОВ, ОБУСЛОВЛЕННЫХ ПАРАМЕТРАМИ СЕНСОРА РЕГИСТРИРУЮЩЕГО УСТРОЙСТВА

А.А. Варламова [1], А.В. Кузнецов [1, 2]

[1] Самарский национальный исследовательский университет имени академика С.П. Королева, Самара, Россия,

[2] Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Самара, Россия

Аннотация

Встраивание в изображение областей, скопированных из другого изображения, является одним из часто осуществляемых видов подделки изображений. Данная статья посвящена исследованию одного из методов их обнаружения, работа которого основана на анализе артефактов, обусловленных параметрами сенсора регистрирующего устройства, при помощи которого было получено изображение. Для проверки подлинности изображение разбивается на блоки, для каждого из которых вычисляется критерий, определяющий вероятность наличия/отсутствия на нем артефактов и, как следствие, вероятность того, является ли блок встроенным. В экспериментальной части работы проводится анализ точности обнаружения встроенных областей, а также исследование устойчивости метода к различным видам искажений: аддитивному гауссовскому шуму, сжатию JPEG и линейному контрастированию. Результаты экспериментов показали, что метод позволяет обнаруживать встроенные области различной природы, формы и размера, а также обладает устойчивостью к аддитивному гауссовскому шуму и линейному контрастированию для заданного диапазона параметров, но не устойчив к сжатию JPEG. Отличительной особенностью метода является возможность выявления встроенных областей с минимальным размером 2×2.

Ключевые слова: искажение изображения, массив цветных фильтров, фильтр Байера, интерполяция, артефакт, карта вероятностей искажения.

Цитирование: **Варламова, А.А.** Обнаружение встраиваний на изображениях путем анализа артефактов, обусловленных параметрами сенсора регистрирующего устройства / А.А. Варламова, А.В. Кузнецов // Компьютерная оптика. – 2017. – Т. 41, № 6. – С. 920-930. – DOI: 10.18287/2412-6179-2017-41-6-920-930.

Введение

Появление большого числа цифровых устройств, с помощью которых можно получать изображения, привело к снижению их стоимости и, как следствие, широкой доступности для каждого человека. Вместе с этим значительно увеличилось количество программных средств для редактирования изображений. Все это вызвало широкое распространение подделки изображений.

В современном мире любой пользователь может внести изменения в изображение, которые далеко не всегда можно определить невооруженным глазом. Кроме этого, если речь идет о профессиональной подделке, то зачастую и существующие сервисы проверки подлинности изображений не могут ее выявить.

Существует множество примеров из военной и политической сферы, СМИ, из судебных разбирательств, деятельности страховых компаний и многих других областей, когда были выявлены подделки изображений, осуществленных с целью совершения преступлений или сокрытия каких-либо фактов, для нарушения авторских прав или формирования общественного резонанса [1]. В связи с этим остро встал вопрос о защите изображений и о проверке их подлинности.

В зависимости от преследуемой цели подделки изображения могут быть подвергнуты таким искажениям, как ретуширование и встраивание дубликатов (копирование некоторых областей изображения и их вставка в другие области этого же изображения).

В работах [2, 3] рассматриваются методы обнаружения ретуширования на изображениях, а в работах [4, 5] приводятся методы обнаружения дубликатов.

Еще одним из часто используемых методов подделки изображений, обнаружению которого посвящена данная работа, является встраивание областей, скопированных из другого изображения, называемое также фотомонтажом [6].

Для защиты изображений от такого вида подделки можно выполнять встраивание в них цифрового водяного знака [7]. Однако такой метод обладает рядом недостатков. Так, например, его применение ограничено, поскольку проверка подлинности в этом случае может производиться лишь владельцем данных.

Разработаны и другие решения, не требующие встраивания дополнительной информации в изображение. В частности, к ним относятся те методы, которые для выявления подделки используют характеристики сенсора устройства, с помощью которого изображение было получено.

Одной из таких характеристик является массив цветных фильтров (*color filter array*, *CFA*), присутствующий в большинстве современных регистрирующих устройств. Он представляет собой часть светочувствительной матрицы фотоприбора, осуществляющую пространственное цветоделение изображения при помощи фотодатчиков – пикселей матрицы, расположенных за светофильтрами различного цвета.

Присутствие в устройстве CFA-фильтра приводит к тому, что изображения, получаемые с его помощью, содержат артефакты, также называемые в англоязычной литературе CFA-артефактами [8]. Иными словами, CFA-артефакты – это локальные искажения изображения, обусловленные наличием CFA-фильтра в регистрирующем устройстве, при помощи которого было

получено изображение. CFA-артефакты являются уникальными для каждого регистрирующего устройства.

В работе [9] описывается метод выявления CFA-артефактов на изображении. В рамках метода для изображения вычисляется карта вероятностей присутствия CFA-артефактов, затем от нее вычисляется преобразование Фурье (ПФ). Наличие пиков ПФ является признаком периодичности карты и, как следствие, признаком того, что изображение содержит CFA-артефакты. При небольших модификациях метод может быть использован для обнаружения искажений на изображении размера 256×256 и более.

Аналогично, основываясь на том факте, что CFA-артефакты имеют периодическую структуру, в работе [10] представлен алгоритм для определения природы изображения, то есть определения того, было ли оно получено с помощью цифрового устройства или искусственно сгенерировано. В основе работы метода также лежит анализ ПФ. Отсутствие CFA-артефактов на области изображения свидетельствует о том, что область была изменена, является искусственно сгенерированной или в снимающем устройстве CFA-фильтр не используется. Данный метод применим для обнаружения искажений на изображениях размера 64×64 и более, что обусловлено использованием преобразования Фурье.

Данная группа методов не позволяет обнаруживать искажения в случае, если после искажения была повторно применена интерполяция – данная проблема будет решаться в ходе дальнейших исследований. В настоящей работе этот случай не рассматривается, поскольку он является другим видом искажения и выходит за рамки поставленной задачи. Также не рассматривается случай, когда исходное изображение и встроенная в него область были получены с помощью одного и того же регистрирующего устройства – в этом случае CFA-артефакты одинаковы.

В данной работе проводится исследование одного из методов обнаружения встраиваний на изображениях, основанного на анализе CFA-артефактов [8]. С его помощью можно обнаруживать искажения на областях с минимальным размером 2×2. Результатом применения метода является карта вероятностей искажения изображения, которая представляет собой двумерный массив, каждый элемент которого содержит вероятность искажения соответствующей локальной области на изображении.

Работа построена следующим образом. В первом параграфе описываются принципы работы методов обнаружения встраиваний, основанных на анализе CFA-артефактов. Второй параграф посвящен описанию исследуемого метода обнаружения встроенных областей на изображении. Третий параграф содержит в себе результаты экспериментальных исследований качества обнаружения и устойчивости метода на наборе изображений со встроенными областями.

1. Обнаружение встраиваний на изображении путем анализа CFA-артефактов

Несмотря на то, что искажения на изображении часто невозможно обнаружить при помощи визуального анализа, они приводят к изменениям его статистических характеристик. В частности, такие искажения нарушают межпиксельные связи, которые возникают в процессе регистрации RGB-изображения [11].

В большинстве современных камер для получения цветного изображения используется массив цветных фильтров.

Существует ряд разновидностей CFA-фильтров, наиболее распространенным является фильтр Байера (рис. 1). Исследованию влияния характерных особенностей данного фильтра на обнаружение локальных встраиваний (искажений) посвящена данная работа.

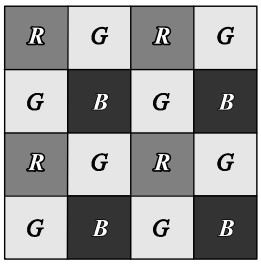

Рис. 1. Фильтр Байера

После прохождения света через CFA-фильтр и сенсор образуется файл формата RAW, в каждом отсчете которого содержится значение только одного цветового канала, то есть в файле RAW присутствует только треть цветовой информации изображения. Помимо этого, в таком файле хранятся EXIF-данные, содержащие информацию о дате и времени создания снимка, устройстве съемки, с помощью которого был сделан снимок, параметры регистрации снимка и т.д.

Ввиду того, что RAW содержит отсчеты изображения, для каждого из которых определено значение только для одного цветового канала из трех, для получения цветного трехканального изображения применяется алгоритм демозаики (интерполяции). Это приводит к возникновению корреляции между отсчетами внутри каждого канала, то есть к возникновению на изображении CFA-артефактов – локальных искажений, обусловленных характеристиками камеры [12].

Задача алгоритма демозаики (*demosaicing*, интерполяция байеровских шаблонов) состоит в получении RGB-изображения по шаблону Байера. Иными словами, с помощью алгоритма демозаики происходит интерполирование каждой из трех цветовых плоскостей в тех отсчетах, где значение соответствующей цветовой компоненты неизвестно.

При получении трехканального изображения с помощью интерполяции в каждом канале вычисляются недостающие значения отсчетов по значениям известных, соседних с ними, отсчетов. Этот процесс можно интерпретировать как процесс фильтрации, при котором ядро интерполяции (маска) периодически применяется к исходному RAW-изображению для получения результирующего трехканального изображения.

Существует множество алгоритмов интерполяции. Наиболее подробно они рассмотрены в работе [9]. При вычислении недостающих значений в зависимости от применяемого алгоритма для расчетов могут

быть задействованы и значения из всех каналов одновременно, что приводит к возникновению межканальных связей.

Далее для простоты будем рассматривать алгоритм без межканальных связей, то есть недостающие значения канала будут рассчитаны на основе известных отсчетов только из того же канала.

Все расчеты, приведенные в работе, выполнены для зеленого канала, для двух других они производятся аналогичным образом.

Наиболее простым алгоритмом интерполяции является билинейная интерполяция. В дальнейшем будет считаться, что при создании изображения применялся билинейный алгоритм интерполяции.

На рис. 2 схематически представлен вид зеленого канала изображения после интерполяции. В позициях *A* расположены отсчеты с известными значениями, то есть те, чьи значения были получены в результате съемки, далее будем называть их отсчетами *A* (*acquired*), а в позициях *I* располагаются отсчеты, значения которых были вычислены по известным значениям посредством интерполяции, далее будем называть их отсчетами *I* (*interpolated*).

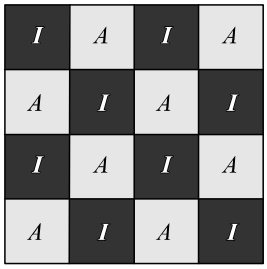

Рис. 2. Зеленый канал изображения (A – отсчеты, полученные в результате съемки, I – отсчеты, полученные путем интерполяции)

Пусть $G_A(x,y)$ – значения отсчетов зеленого канала изображения, полученные непосредственно в ходе съемки, $G_I(x,y)$ – интерполированные значения отсчетов зеленого канала изображения, $h(u,v)$ – ядро интерполяции (в одномерном случае ядро интерполяции будет обозначаться $h(u)$, например, $h(u)=[1\ 2\ 1]/2$), тогда значения зеленого канала $s(x,y)$ при использовании билинейной интерполяции определяются по формуле (1):

$$s(x,y)=\begin{cases} G_A(x,y), & (x,y)\in A \\ G_I(x,y)=\sum_u\sum_v h(u,v)\times \\ \times G_A(x-u)(y-v), & (x,y)\in I \end{cases}. \quad (1)$$

Существуют различные виды двумерных ядер интерполяции, например, ядро билинейной интерполяции $h(u,v)$:

$$h(u,v)=\frac{1}{4}\times\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (2)$$

ядро бикубической интерполяции $h_c(u,v)$:

$$h_c(u,v)=\frac{1}{256}\times\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -9 & 0 & -9 & 0 & 0 \\ 0 & -9 & 0 & 81 & 0 & -9 & 0 \\ 1 & 0 & 81 & 256 & 81 & 0 & 1 \\ 0 & -9 & 0 & 81 & 0 & -9 & 0 \\ 0 & 0 & -9 & 0 & -9 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

В рамках настоящей работы было использовано ядро интерполяции (2), соответствующее наиболее часто применяемому в цифровых камерах фильтру Байера.

При подделке изображения связи между отсчетами, возникающие в результате интерполяции, разрушаются или изменяются.

Поскольку цветовые фильтры (в том числе и фильтр Байера) периодичны, то и корреляция между отсчетами также имеет периодический характер. Таким образом, исследуя корреляцию пикселей локальных областей, возникшую в результате интерполяции, можно определить, была ли искажена та или иная область изображения.

2. Исследуемый метод обнаружения встраиваний

Теоретическое обоснование работы метода

Для простоты будет рассмотрен одномерный случай, поскольку выводы, сделанные по итогам расчетов, справедливы и для двумерного случая.

Пусть $G_A(x)$ – значения отсчетов зеленого канала изображения, полученные непосредственно в ходе съемки, $G_I(x)$ – интерполированные значения отсчетов зеленого канала изображения, $h(u)$ – одномерное ядро интерполяции длины N_h, тогда значения $s(x)$ – одномерного зеленого канала изображения, представленного в виде строки, полученной путем интерполяции с использованием фильтра Байера, определяются по формуле:

$$s(x)=\begin{cases} G_A(x), & x(\mathrm{mod}\,2)=0 \\ G_I(x)=\sum_{u=0}^{N_h-1} h(u)\cdot G_A(x+u), & x(\mathrm{mod}\,2)\neq 0 \end{cases}. \quad (3)$$

В дальнейшем при обозначении применения фильтра предел суммирования указываться не будет для сокращения записи, но предел суммирования будет подразумеваться равным длине ядра интерполяции.

При вычислении интерполированных значений вклад вносят только значения, стоящие в четных позициях, то есть только значения, полученные в ходе съемки, поэтому, как правило, $h(u)=0$ для нечетных значений u. Иначе значение $G_A(x+u)$ равно нулю. Следовательно, если $k(u)$ – ядро предсказателя, то ошибка предсказания значений зеленого канала может быть определена по формуле (4):

$$e(x)=s(x)-\sum_u k(u)\,s(x+u). \quad (4)$$

В случае, когда ядро интерполяции $h(u)$, которое использовалось в регистрирующем устройстве при получении изображения известно, ядро предсказателя совпадает с ядром интерполяции, т.е. $k(u)=h(u)$. В таком случае ошибка предсказания отсутствует. В случае, когда тип фильтра неизвестен, возникает ошибка предсказания.

После подстановки (3) в (4), ошибка предсказания может быть выражена следующим образом:

$$e(x)=\begin{cases} G_A(x)-\sum\limits_u k(u)s(x+u), & x(\bmod 2)=0 \\ \sum\limits_u h(u)G_A(x+u)- & \\ -\sum\limits_u k(u)s(x+u), & x(\bmod 2)\neq 0 \end{cases}.$$

Пусть $k(u)=h(u)$, тогда ошибка предсказания равна нулю при нечетных значениях x и отлична от нуля при четных значениях x. Следовательно, так как значения четных отсчетов были получены непосредственно в результате съемки, а значения нечетных отчетов – в результате интерполяции, то в идеальном случае, когда ядра интерполяции и предсказания совпадают, дисперсия ошибки предсказания равна нулю для интерполированных отчетов и отлична от нуля для гарантированно известных отчетов.

На практике равенство $k(u)=h(u)$ может не выполняться, но, как правило, соотношение $\sum\limits_u k(u)=\sum\limits_u h(u)=1$ справедливо для любых используемых ядер интерполяции.

Ввиду того, что при вычислении ошибки имеют смысл только значения, соответствующие нечетным значениям u, то для оценки ошибки предсказания будем рассматривать именно их. Следовательно, ошибка предсказания может быть выражена следующим образом:

$$e(x)=\begin{cases} G_A(x)-\sum\limits_u k(u)\times & \\ \times\sum\limits_v h(v)G_A(x+u+v), & x(\bmod 2)=0 \\ \sum\limits_u \big(h(u)-k(u)\big)\times & \\ \times G_A(x+u), & x(\bmod 2)\neq 0 \end{cases}. \quad (5)$$

Полагая, что значения отсчетов, полученных в ходе съемки, независимы и одинаково распределены с математическим ожиданием (МО) μ_G и дисперсией σ_G^2, МО ошибки предсказания может быть вычислено по формуле (6):

$$E\big[e(x)\big]=\begin{cases} \mu_G-\mu_G\times & \\ \times\sum\limits_u k(u)\sum\limits_v h(v), & x(\bmod 2)=0 \\ \mu_G\times & \\ \times\left(\sum\limits_u h(u)-\sum\limits_u k(u)\right)=0, & x(\bmod 2)\neq 2 \end{cases}. \quad (6)$$

Дисперсия для четных значений x вычисляется по формуле (7):

$$Var\big[e(x)\big]=\sigma_G^2\left[\begin{array}{l}\left(1-\sum\limits_u k(u)h(-u)\right)^2+\\ +\sum\limits_{t\neq 0}\left(\sum\limits_u k(u)h(t-u)\right)^2\end{array}\right]. \quad (7)$$

Дисперсия для нечетных значений x вычисляется по формуле (8):

$$Var\big[e(x)\big]=\sigma_G^2\sum_u\big(h(u)-k(u)\big)^2. \quad (8)$$

На основании приведенных расчетов можно сделать вывод о том, что дисперсия ошибки предсказания пропорциональна дисперсии отсчетов, полученных в ходе съемки. Однако если ядро предсказателя и ядро интерполяции совпадают, то однозначно дисперсия ошибки в изначально известных отчетах значительно выше дисперсии ошибки, вычисленной для интерполированных отсчетов.

Разработка метода обнаружения встраиваний на изображении

Дисперсия ошибки предсказания выше для отсчетов, полученных с помощью камеры (ранее они были обозначены как отсчеты A), чем для интерполированных отсчетов (отсчетов I). Это утверждение справедливо и для двумерного случая. В случае, если изображение было получено не путем применения алгоритма демозаики или было искажено, что приводит к разрушению артефактов, оставленных после применения алгоритма демозаики, дисперсия ошибки предсказания для обоих типов отсчетов будет иметь близкие значения в пределах некоторого ε. Следовательно, для того чтобы выявить наличие/отсутствие артефактов, возникающих после применения интерполяции, нужно вычислить дисперсию ошибки предсказания для отчетов A и I.

Пусть $s(x,y)$ – канал изображения (как правило, зеленый), значения которого определяются по формуле (1), $k(u,v)$ – двумерное ядро предсказателя, определяемое формулой (2), тогда ошибка предсказания может быть определена по формуле (9):

$$e(x,y)=s(x,y)-\sum_{u\neq 0}\sum_{v\neq 0}k(u,v)s(x+u,y+v). \quad (9)$$

Будем считать, что алгоритм демозаики неизвестен, поэтому $k(u,v)\neq h(u,v)$, где $h(u,v)$ – двумерное ядро интерполяции, которое было использовано при регистрации изображения.

Стоит отметить, что значения отсчетов, полученных в ходе съемки, независимы и одинаково распределены, как правило, не на всем изображении, а лишь локально, поэтому вычисление локальной дисперсии ошибки предсказания как для отсчетов I, так и для отсчетов A будет производиться локально.

Пусть ошибка предсказания стационарна в пределах области размера $(2K+1)\times(2K+1)$, $c=1-\sum\limits_{i=-K}^{K}\sum\limits_{j=-K}^{K}\alpha^2(i,j)$ – масштабирующий множитель,

$\mu_e = \sum_{i=-K}^{K}\sum_{j=-K}^{K} \alpha(i,j)e(x+i,y+j)$ – локально-взвешенное МО ошибки предсказания, $\alpha'(i,j) = W(i,j)$, если $e(x+i,y+j)$ и $e(x,y)$ отчеты одного типа, иначе $\alpha'(i,j) = 0, W$ – Гауссовское окно размера $(2K+1)\times(2K+1)$ со среднеквадратическим отклонением (СКО) $\sigma_W^2 = K/2$. Под Гауссовским окном будем понимать сглаживающий двумерный фильтр, элементы которого распределены в соответствии с нормальным законом распределения. Значение дисперсии элементов W выбрано экспериментально путем сравнения с другими значениями:

$$\sigma_W^2 = \left\{K, \frac{K}{2}, \frac{K}{4}, \frac{K}{8}\right\}$$

(данные исследования в рамках работы не приводятся).

Тогда локально-взвешенная дисперсия ошибки предсказания $\sigma_e^2(x,y)$ определяется по формуле (10):

$$\sigma_e^2(x,y) = \frac{1}{c}\left(\sum_{i=-K}^{K}\sum_{j=-K}^{K} \alpha(i,j)e^2(x+i,y+j) - \mu_e^2\right), \quad (10)$$

где $\alpha(i,j) = \dfrac{\alpha'(i,j)}{\sum_{i=-K}^{K}\sum_{j=-K}^{K}\alpha'(i,j)}$ – весовые коэффициенты.

Вычисление основного критерия метода

После нахождения локально-взвешенной дисперсии ошибки предсказания вычисляется критерий, характеризующий отношение дисперсий ошибки предсказания в исходных и интерполированных отсчетах. По полученным значениям критерия можно определить наличие/отсутствие на изображении CFA-артефактов.

Пусть размер анализируемого изображения $N\times N$, тогда можно вычислить критерий для каждого из непересекающихся блоков изображения размера $B\times B$. Значение блока должно соотноситься с размерами фильтра Байера, при этом минимальный допустимый размер блока – 2×2. Матрица полученных значений дисперсии ошибки предсказателя разбивается на блоки размера $B\times B$. В каждый блок $B_{k,l}$ входят значения дисперсии исходных и интерполированных отсчетов, которые будем обозначать $B_{Ak,l}$ и $B_{Ik,l}$ соответственно, где $k,l = \overline{0,(N/B)-1}$.

Для формирования критерия принадлежности блока изображения к искажённым или неискажённым данным будем применять среднее геометрическое значение локально-взвешенных дисперсий ошибки предсказания в рамках выбранного фрагмента изображения. Стоит отметить, что также можно использовать и любую другую усредняющую меру, например, среднее арифметическое, чтобы получить некоторую характеристику «искажённости» фрагмента.

Пусть $GM_A(k,l)$ – среднее геометрическое значение дисперсии ошибки предсказания для отсчетов A внутри блока $B_{k,l}$ и определяется по формуле (11):

$$GM_A(k,l) = \left[\prod_{i,j\in B_A(k,l)} \sigma_e^2(i,j)\right]^{\frac{1}{|B_{Ak,l}|}}, \quad (11)$$

$GM_I(k,l)$ – среднее геометрическое значение дисперсии ошибки предсказания для отсчетов I внутри блока $B_{k,l}$, и определяется по формуле (12):

$$GM_I(k,l) = \left[\prod_{i,j\in B_I(k,l)} \sigma_e^2(i,j)\right]^{\frac{1}{|B_{Ik,l}|}}, \quad (12)$$

тогда критерий, характеризующий отношение дисперсий ошибок предсказания, может быть рассчитан по формуле (13):

$$L(k,l) = \ln\left[\frac{GM_A(k,l)}{GM_I(k,l)}\right]. \quad (13)$$

Если на блоке изображения $B_{k,l}$ присутствуют CFA-артефакты, то есть он был получен в результате применения алгоритма демозаики, то значение дисперсии будет выше в отсчетах A, таким образом, значение критерия $L(k,l)$ будет положительным. Однако если изображение было получено другим способом, то дисперсии ошибки предсказания для двух типов отсчетов будут иметь близкие значения в пределах некоторого ε, так как значения отсчетов будут одинаково распределены и будут иметь одинаковые статистические характеристики, следовательно, значение $L(k,l)$ будет близко к нулю в рамках выбранного значения ε окрестности.

Вычисление карты вероятностей искажения изображения. Алгоритм Expectation-maximization (EM-алгоритм)

Если в изображение было встроено новое содержимое, то обычно для того чтобы сделать вставку более реалистичной, она сопровождается другими процессами: сглаживанием, компрессией и т.д. Все это нарушает свойства, обусловленные процессом интерполяции, то есть приводит к разрушению CFA-артефактов. Следовательно, значения критерия L в измененном изображении будут неоднородны: в одних его областях его величина будет значительно выше нуля, что является следствием присутствия CFA-артефактов, а в других областях, где CFA-артефакты отсутствуют, значение критерия будет близко к нулю в рамках выбранного значения ε окрестности. Этот факт может быть использован для обнаружения искажений на изображениях путем нахождения по значениям критерия L вероятности присутствия CFA-артефактов в каждом блоке $B_{k,l}$. То есть при помощи полученных значений критерия можно определить карту вероятностей присутствия CFA-артефактов. Для этого используется EM-алгоритм [13].

Пусть существуют две гипотезы:

– M_1 – CFA-артефакты присутствуют;

– M_2 – CFA-артефакты отсутствуют.

Будем считать, что значения критерия $L(k, l)$ распределены по нормальному закону и в случае M_1, и в случае M_2, а также при любом размере блока $B_{k,l}$. Зафиксируем размер блока $B \times B$.

Пусть $\mu_1 > 0$ – неизвестное математическое ожидание распределения при справедливости гипотезы M_1, σ_1^2 – неизвестная дисперсия, тогда функция плотности условного распределения для M_1:

$$P\{L(k,l)|M_1\} \sim N(\mu_1, \sigma_1^2).$$

Пусть $\mu_2 = 0$ – математическое ожидание распределения при справедливости гипотезы M_2, σ_2^2 – неизвестная дисперсия, тогда для M_2 функция плотности условного распределения:

$$P\{L(k,l)|M_2\} \sim N(\mu_2, \sigma_2^2).$$

Будем считать, что параметры распределения в обоих случаях являются постоянными.

Если изображение, полученное с помощью алгоритма демозаики, было изменено, то для каждого отсчета выполняются обе гипотезы, но с разной вероятностью. Это позволяет представить критерий $L(k, l)$ как смесь двух гауссовских распределений с МО $\mu_1 = 0$ в областях, где артефакты присутствуют, то есть область подлинная, и с $\mu_2 = 0$ в областях, где CFA-артефакты отсутствуют, то есть имело место искажение данной области.

Для того чтобы получить параметры смеси двух гауссовских распределений, а именно: μ_1, σ_1^2, σ_2^2, можно воспользоваться EM-алгоритмом. Это итеративный алгоритм, состоящий из двух шагов на каждой итерации, который позволяет разделить смесь нескольких распределений и определить их параметры – МО и дисперсию путем максимизации отношения правдоподобия. Зная параметры для каждого отсчета, можно определить апостериорные вероятности каждой из гипотез $P\{M_1|L(k, l)\}$ и $P\{M_2|L(k, l)\}$.

На E-шаге алгоритма вычисляются вероятности принадлежности каждого отсчета $L(k, l)$ к каждой из моделей. При этом будем считать априорные вероятности каждой из гипотез равными:

$$P\{M_1\} = P\{M_2\} = \frac{1}{2}.$$

Пусть $P\{L(k, l)|M_1\}$ – вероятность $L(k, l)$ при истинности гипотезы M_1, $P\{L(k, l)|M_2\}$ – вероятность $L(k, l)$ при истинности гипотезы M_2, тогда вероятность того, что блок не был изменен и CFA-артефакты на нем присутствуют, то есть вероятность гипотезы M_1, определяется по формуле Байеса (14):

$$P\{M_1|L(k,l)\} = \frac{P\{L(k,l)|M_1\}}{P\{L(k,l)|M_1\} + P\{L(k,l)|M_2\}}. \quad (14)$$

Аналогичным образом по формуле Байеса вычисляется $P\{M_1|L(k, l)\}$ – вероятность того, что блок был изменен, а CFA-артефакты отсутствуют, то есть вероятность гипотезы M_2.

На основании полученных вероятностей оцениваются параметры распределения: μ_1, σ_1^2, σ_2^2, которые далее на М-шаге считаются зафиксированными, что позволяет рассчитать отношение правдоподобия по формуле (15):

$$\Lambda(L(k,l)) = \frac{P\{L(k,l)|M_2\}}{P\{L(k,l)|M_1\}}. \quad (15)$$

Параметры, обеспечивающие максимум отношения правдоподобия (15), и являются искомыми параметрами распределения, а сами вычисленные значения отношения правдоподобия и являются картой искажения изображения, в которой каждый отсчет $\Lambda(L(k, l))$ представляет вероятность наличия в блоке $B_{k,l}$ CFA-артефактов, таким образом, небольшие значения являются признаком того, что блок был искажен.

Критерии качества обнаружения искажений на изображениях

Качество обнаружения алгоритмом искажений может быть определено с помощью критерия R_{TP} (*true positive rate*), характеризующего долю верно обнаруженных искаженных блоков и критерия R_{FP} (*false positive rate*), характеризующего долю ложных обнаружений [14].

Пусть R_2 – искаженная область изображения, N_{R_2} – общее количество блоков в области R_2, $N_{m_{R_2}}$ – количество верно обнаруженных искаженных блоков в области R_2, тогда:

$$R_{TP} = \frac{N_{m_{R_2}}}{N_{R_2}}. \quad (16)$$

Аналогично производится расчет критерия R_{FP}.

Пусть R_1 – неискаженная область изображения, N_{R_1} – общее количество блоков в области R_1, $N_{m_{R1}}$ – количество ложно обнаруженных неискаженных блоков в области R_1, тогда:

$$R_{FP} = \frac{N_{m_{R1}}}{N_{R_1}}. \quad (17)$$

Далее метрики R_{TP} и R_{FP} применяются для оценки качества обнаружения искаженных локальных областей.

3. Экспериментальные исследования

Для проведения экспериментов было использовано четыре файла формата RAW, взятых из базы данных [15]. При этом выбранные файлы были получены с помощью четырех различных камер, использующих фильтр Байера, а именно: Canon EOS 450D, Nikon D50, Nikon D90, Nikon D7000. Тип фильтра, используемого камерой, можно узнать из ее технических характеристик. Для получения трехканального изображения формата TIFF из файла RAW было использовано приложение dcraw [16].

В качестве изображений-вставок для формирования искажений были использованы изображения двух типов: искусственно созданные изображения, то есть

те, у которых отсчеты не коррелированы друг с другом, и изображения, взятые из источников [17, 18], отличающиеся свойствами интерполяции.

Прежде всего, в ходе экспериментов была проверена способность обнаружения алгоритмом искусственно встроенных областей различной природы и формы.

На рис. 3 представлен пример изображения со вставкой произвольной формы и соответствующая ему карта вероятностей искажения, вычисленная блоками размера 8×8. Встроенная область была получена при помощи другой камеры.

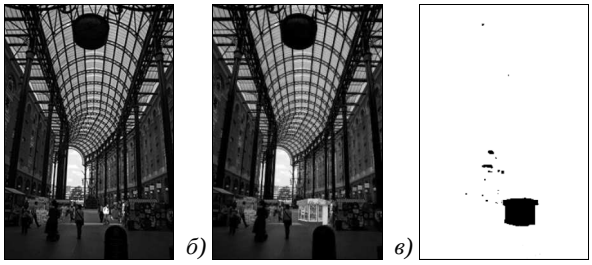

Рис. 3. Примеры работы алгоритма при размере обрабатываемого блока 8×8*: исходное изображение (а), изображение со встроенной областью произвольной формы (б), карта вероятностей искажения изображения (в)*

Стоит отметить, что алгоритм позволяет обнаруживать искажения очень малых размеров, поэтому карта вероятностей искажения может быть рассчитана блоками с минимальным размером 2×2.

Пусть размер встраиваемой в изображение области $M \times M$. Графики зависимости характеристик качества обнаружения от размера встроенной области $R_{TP}(M)$ и $R_{FP}(M)$ представлены на рис. 4.

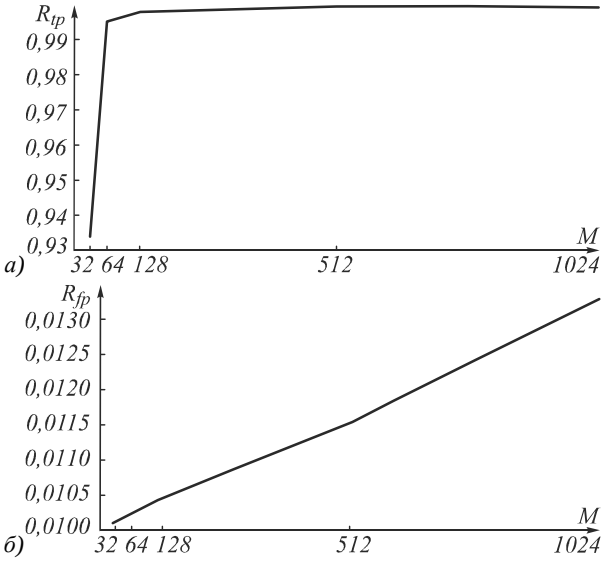

Рис. 4. Зависимость характеристик качества обнаружения от размера встроенной области а) $R_{TP}(M)$*, б)* $R_{FP}(M)$

Результаты проведенного эксперимента показали, что с увеличением размера встраиваемой области качество обнаружения улучшается и достигает максимального значения – 1 при размере 512×512, однако и при минимальном исследуемом размере вставки 32×32, $R_{TP} = 0{,}93$, что характеризует высокое качество обнаружения. При этом число ложно обнаруженных неискаженных блоков изображения незначительно растет и при размере встраиваемой области 1024×1024 составляет $R_{FP} = 0{,}0133$.

Исследование устойчивости метода к различным типам искажений

Для исследования устойчивости алгоритма к различным типам искажений были использованы ранее полученные 40 тестовых изображений с фиксированным размером встроенной области 128×128.

В рамках исследований к каждому из них был добавлен аддитивный гауссовский шум со значением отношения сигнал/шум (*SNR* (дБ)): 30, 35, 40, 45, 50. Пример проведенного эксперимента представлен на рис. 5.

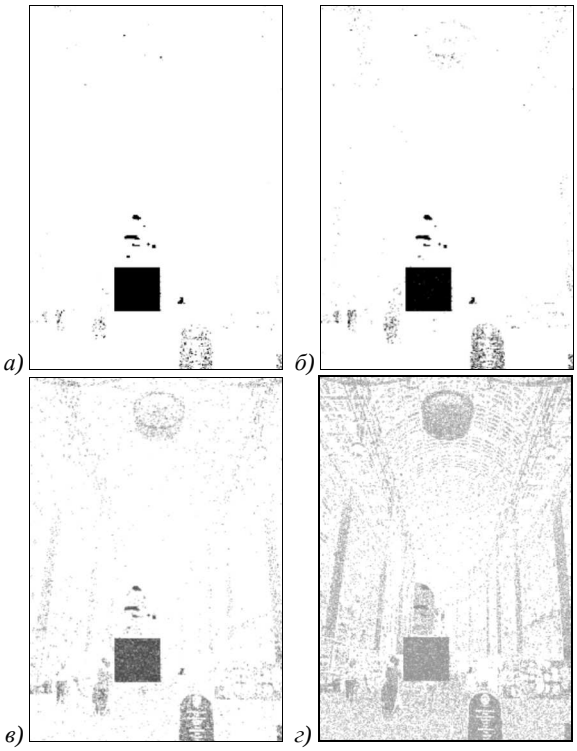

Рис. 5. Карта искажений изображения со встроенной областью после добавления аддитивного гауссовского шума со значением SNR (дБ): SNR = 50 (а), SNR = 45 (б), SNR = 40 (в), SNR = 35 (г)

Для того чтобы улучшить восприятие полученных результатов, ко всем представленным картам вероятностей искажений было применено контрастирование.

На рис. 6 представлена зависимость характеристик качества обнаружения искажений от значения *SNR*: $R_{TP}(SNR)$ и $R_{FP}(SNR)$.

Результаты эксперимента показали, что число верно обнаруженных искаженных блоков на изображении велико для заданного диапазона параметров, однако при $SNR = 35$ дБ и менее число ложных срабатываний алгоритма растет, поэтому можно сказать, что метод работает при значениях $SNR = 35$ дБ и выше.

Далее в рамках экспериментов к тому же набору изображений было применено сжатие JPEG с параметром качества $Q = 100, 98, 96, 94$. Пример работы алгоритма при задании различных значений параметра качества показан на рис. 7.

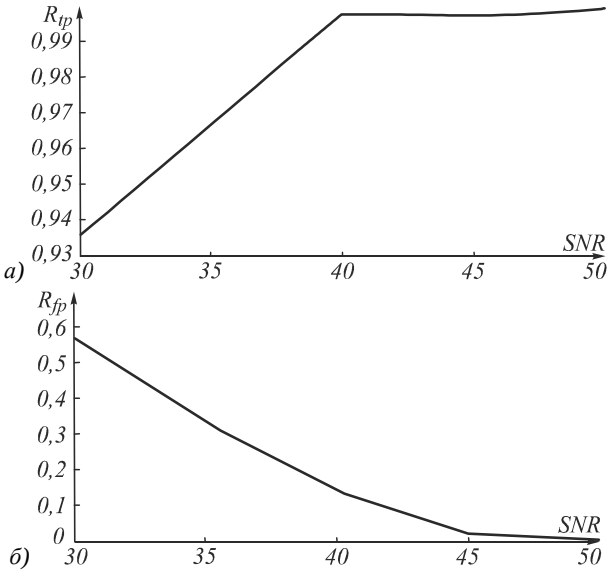

Рис. 6. Зависимость характеристик качества обнаружения от значения SNR: а) $R_{TP}(SNR)$, б) $R_{FP}(SNR)$

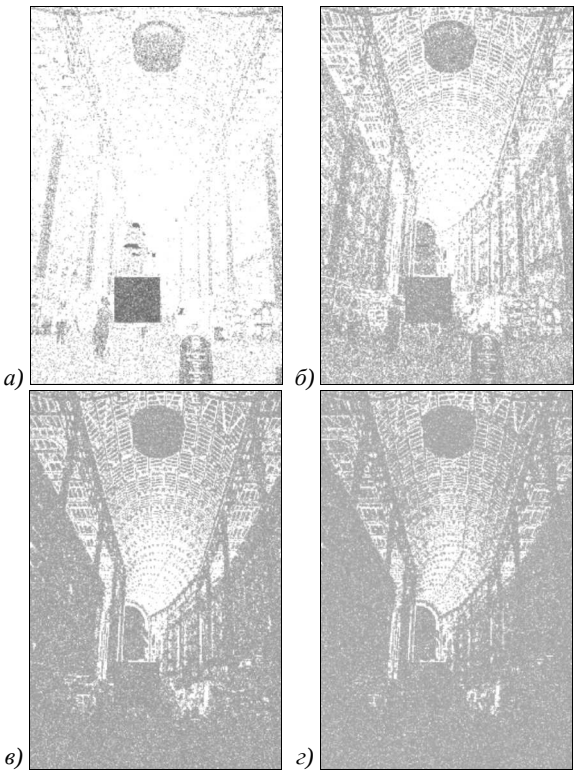

Рис. 7. Карта искажений изображения со встроенной областью после применения сжатия JPEG со значениями параметра качества: Q (дБ): а)Q = 100, б) Q = 98, в) Q = 96, г) Q = 94

На рис. 8 представлена зависимость характеристик качества обнаружения искажений от значения Q: $R_{TP}(Q)$ и $R_{FP}(Q)$. Из полученных результатов видно, что метод не обладает устойчивостью к сжатию JPEG – даже при высоких значениях параметра качества Q число ложных обнаружений велико и уже при $Q = 92$ $R_{FP}(Q) = 0{,}803$. Такой результат можно считать подтверждением очевидных предположений. Применение алгоритма JPEG нарушает интерполяционные свойства на изображении, что и приводит к резкому росту ложно обнаруживаемых фрагментов изображения.

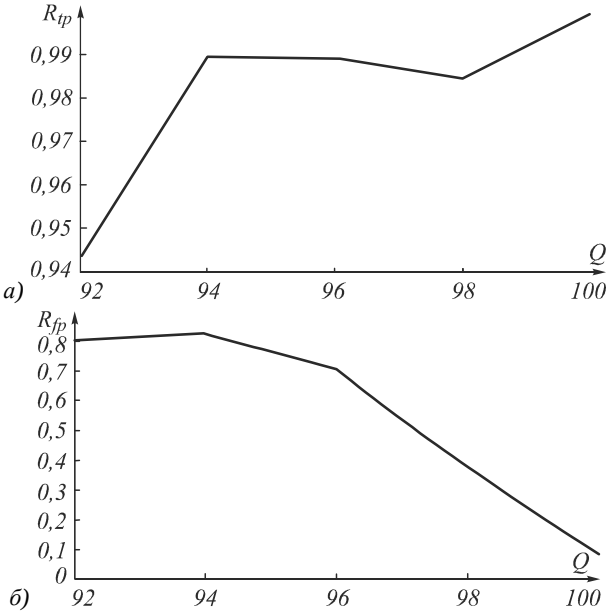

Рис. 8. Зависимость характеристик качества обнаружения от значения Q: а) $R_{TP}(Q)$, б) $R_{FP}(Q)$

В ходе выполнения экспериментальной части работы также было проведено исследование устойчивости метода к сжатию JPEG для случая, если оно применяется только к искаженной области изображения.

Сжатие JPEG, примененное только к искаженной области изображения, не влияет на результат обнаружения, что доказывает тот факт, что алгоритм позволяет обнаруживать встроенные области различной природы.

В рамках экспериментов также было проведено исследование устойчивости метода к линейному контрастированию.

Изображения, вводимые в компьютер, часто являются малоконтрастными, то есть у них вариации функции яркости малы по сравнению с ее средним значением. Реальный динамический диапазон яркостей $[f_{min}, f_{max}]$ для таких изображений оказывается намного меньше допустимого диапазона (шкалы яркости). Задача контрастирования заключается в «растягивании» реального динамического диапазона на всю шкалу. Контрастирование можно осуществить при помощи линейного поэлементного преобразования: $g = af + b$, где a, b – параметры преобразования.

На рис. 9 представлен пример работы метода в случае, если к искаженному изображению было применено линейное контрастирование с различными значениями параметров преобразования, а на рис. 10 представлена зависимость характеристик качества обнаружения от значения параметра преобразования a: $R_{TP}(a)$ и $R_{FP}(a)$. Значение параметра b было зафиксировано: $b = 20$.

Результат эксперимента показал, что алгоритм устойчив к применению линейного контрастирования и результат обнаружения встроенной области не зависит от значений параметров линейного контрастирования.

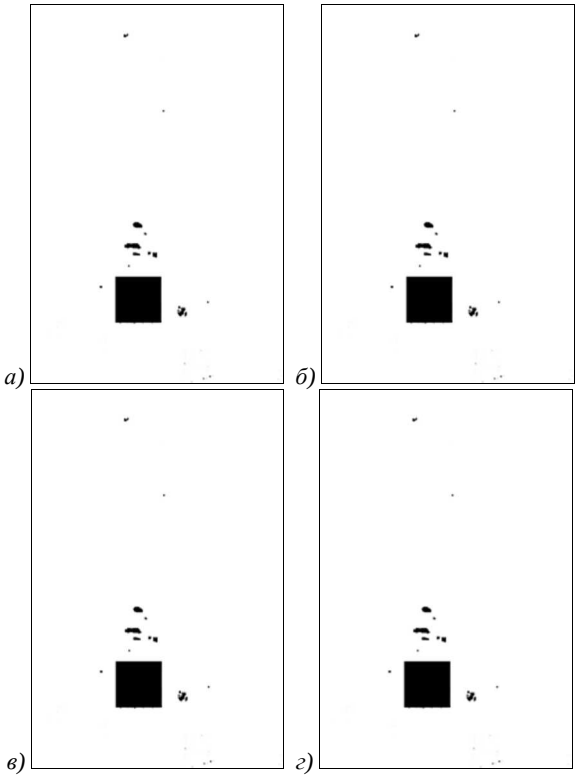

Рис. 9. Карта искажений изображения со встроенной областью после линейного контрастирования с изменяющимся параметром: а (дБ): а)а = 0,2; б) а = 0,4; в) а = 0,6; г) а = 0,8

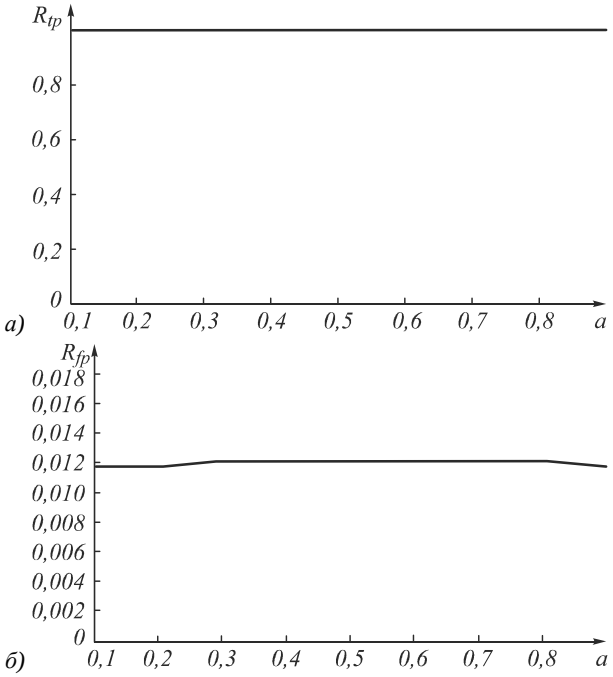

Рис. 10. Зависимость характеристик качества обнаружения от значения а: а) $R_{TP}(a)$, б) $R_{FP}(a)$

Выводы

В работе рассмотрен метод обнаружения встраиваний на изображениях. В ходе проведенных исследований было установлено, что он позволяет обнаруживать на изображениях встроенные области различной природы и формы. С увеличением размера встраиваемой области качество обнаружения растет, однако незначительно увеличивается число ложных срабатываний. Минимальный размер встраиваемой области, при которой она может быть обнаружена, составляет 2×2.

Экспериментальные исследования также показали, что алгоритм устойчив к таким искажениям, как аддитивный белый гауссовский шум при значении выше 35 дБ и линейному контрастированию при любых значениях параметров преобразования. Однако метод оказался неустойчивым к сжатию JPEG. Даже при высоких значениях параметра качества количество ложных срабатываний велико.

Исследуемый метод может быть применен для проверки подлинности изображений. Он позволяет находить встроенные области даже очень малых размеров, однако его применение ограничено (для обнаружения встраиваний на сжатых изображениях он не работает).

Благодарности

Работа выполнена при частичной финансовой поддержке гранта РФФИ № 16-37-00056.

Литература

1. Как бороться с подделками фотоотчетов [Электронный ресурс]. – URL: https://club.esetnod32.ru/articles/analitika/kak-borotsya-s-poddelkami-fotootchetov/ (дата обращения 14.08.2017).
2. **Choi, C.-H.** Estimation of color modification in digital images by CFA pattern change / C.-H. Choi, H.-Y. Lee, H.-K. Lee // Forensic Science International. – 2013. – Vol. 226, Issues 1-3. – P. 94-105. – DOI: 10.1016/j.forsciint.2012.12.014.
3. **Chakraverti, A.K.** A review on image forgery and its detection procedure / A.K. Chakraverti, V. Dhir // International Journal of Advanced Research in Computer Science. – 2017. – Vol. 8, No. 4. – P. 440-443.
4. **Евдокимова, Н.И.** Локальные шаблоны в задаче обнаружения дубликатов / Н.И. Евдокимова, А.В. Кузнецов // Компьютерная оптика. – 2017. – Т. 41, № 1. – С. 79-87. – DOI: 10.18287/2412-6179-2017-41-1-79-87.
5. **Глумов, Н.И.** Поиск дубликатов на цифровых изображениях / Н.И. Глумов, А.В. Кузнецов, В.В. Мясников // Компьютерная оптика. – 2013. – Т. 37, № 3. – С. 360-367.
6. **Burvin, P.S.** Analysis of digital image splicing detection / P.S. Burvin, J.M. Esther // IOSR Journal of Computer Engineering (IOSR-JCE). – 2014. –Vol. 16, Issue 2. – P. 10-13. – DOI: 10.9790/0661-162111013.
7. **Snigdha, K.M.** Image forgery types and their detection / K.M. Snigdha, A.G. Ajay // International Journal of Advanced Research in Computer Science and Software Engineering. – 2015. – Vol. 5, Issue 4. – P. 174-178.
8. **Ferrara, P.** Image forgery localization via fine-grained analysis of CFA artifacts / P. Ferrara, T. Bianchi, A. De Rosa, A. Piva // IEEE Transactions on Information Forensics and Security. – 2012.– Vol. 7, Issue 5. – P. 1566-1577. – DOI: 10.1109/TIFS.2012.2202227.
9. **Popescu, A.C.** Exposing digital forgeries in color filter array interpolated images / A.C. Popescu, H. Farid // IEEE Transactions on Signal Processing. – 2005. – Vol. 53, Issue 10. – P. 3948-3959. – DOI: 10.1109/TSP.2005.855406.
10. **Gallagher, A.C.** Image authentication by detecting traces of demosaicing / A.C. Gallagher, T. Chen // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08). – 2008. – 8 p. – DOI: 10.1109/CVPRW.2008.4562984.

11. **Li, L.** A robust approach to detect digital forgeries by exploring correlation patterns / L. Li, J. Hue, X. Wang, L. Tian // Pattern Analysis and Applications. – 2015. – Vol. 18, Issue 2. – P. 351-365. – DOI: 10.1007/s10044-013-0319-9.
12. **Bayram, S.** Source camera identification based on CFA interpolation / S. Bayram, H. Sencar, N. Memon, I. Avcibas // IEEE International Conference on Image Processing. – 2005. – Vol. 3. – P. 63-72. – DOI: 10.1109/ICIP.2005.1530330.
13. **Bishop, C.M.** Pattern recognition and machine learning / C.M. Bishop. – New York: Springer Verlag, 2006. – 738 p. – ISBN: 978-0-387-31073-2.
14. **Fawcett, T.** An introduction to ROC analysis / T. Fawcett // Pattern Recognition Letters. – 2006. – Vol. 27, Issue 8. – P. 861-874. – DOI: 10.1016/j.patrec.2005.10.010.
15. The original RAW-samples website [Electronical Resources]. – URL: http://rawsamples.ch (date request 27.08.2017).
16. CS. Centro Studi Progresso Fotografico. Dcraw [Electronical Resource]. – URL: http://www.centrostudiprogressofotografico.it/en/dcraw/ (date request 27.08.2017).
17. Photo database [Electronical Resource]. – URL: http://www.zermatt.ch/ru/Media/Media-corner/Photo-database (date request 30.08.2017).
18. Columbia University Image Library (COIL-100) [Electronical Resource]. – URL: http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php (date request 30.08.2017).

Сведения об авторах

Варламова Анна Александровна, родилась в 1995 году. В 2017 году окончила Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет) с отличием по специальности «Информационная безопасность автоматизированных систем». В настоящее время является аспирантом Самарского университета. Основные сферы научных интересов: обработка изображений, обнаружение искажений на цифровых изображениях, защита информации. E-mail: *varlamova.anna.95@mail.ru* .

Кузнецов Андрей Владимирович, родился в 1987 году. В 2010 году окончил СГАУ с отличием по специальности «Прикладная математика и информатика». В 2010 поступил в аспирантуру СГАУ, в 2013 г. защитил диссертацию на соискание степени кандидата технических наук. В настоящее время – доцент кафедры ГИиИБ Самарский университета и научный сотрудник в Институте систем обработки изображений РАН – филиале ФНИЦ «Кристаллография и фотоника». Круг научных интересов включает обработку и анализ изображений, распознавание образов, обнаружение искажений изображений, геоинформатику. Имеет 37 публикаций, в том числе 18 научных статей и 1 монографию.
Страница в интернете: *http://nil97.ssau.ru/employee/detail.php?ID=35*. E-mail: *kuznetsoff.andrey@gmail.com* .

ГРНТИ: 28.23.15.
Поступила в редакцию 20 октября 2017 г. Окончательный вариант – 27 ноября 2017.

IMAGE SPLICING LOCALIZATION BASED ON CFA-ARTIFACTS ANALYSIS

A.A. Varlamova[1], A.V. Kuznetsov[2]
[1]Samara National Research University, Samara, Russia,
[2]Image Processing Systems Institute of RAS – Branch of the FSRC "Crystallography and Photonics" RAS, Samara, Russia

Abstract

Image splicing is a widespread image forgery technique in which fragments from another image are pasted into the image under forgery. In this paper, a method of image splicing localization based on the analysis of CFA-artifacts that appear in the image during the capturing process is described. A feature characterizing the presence/absence of CFA artifacts for each image block is measured. The obtained values of the feature define the probability of each block to be embedded. Analysis of the accuracy of the splicing localization method and its robustness against different types of tampering, such as additive Gaussian noise, JPEG compression, and linear enhancement are presented in the experimental part of the paper. The results show that the suggested method reveals the embedded regions of different shape, size, and nature in images. The method is found to be stable to the additive Gaussian noise and linear enhancement, but not stable to JPEG compression. The advantage of the method is the ability to localize the spliced-in regions as small as a 2×2 block.

Keywords: image forgery, color filter array, Bayer filter, interpolation, artifact, tampering probability map.

Citation: Varlamova AA, Kuznetsov AV. Image splicing localization based on CFA-artifacts analysis. Computer Optics 2017; 41(6): 920-930. DOI: 10.18287/2412-6179-2017-41-6-920-930.

Acknowledgements: The work was partially funded by the RFBR grant #16-37-00056.

References

[1] How to fight against photo report forgeries [In Russian]. Source: ⟨https://club.esetnod32.ru/articles/analitika/kak-borotsya-s-poddelkami-fotootchetov/⟩.
[2] Choi C-H, Lee H-Y, Lee H-K. Estimation of color modification in digital images by CFA pattern change. Forensic Science International 2013; 226(1-3): 94-105. DOI: 10.1016/j.forsciint.2012.12.014.

[3] Chakraverti AK, Dhir V. A review on image forgery and its detection procedure. Journal of Advanced Research in Computer Science 2017; 8(4): 440-443.
[4] Evdokimova NI, Kuznetsov AV. Local patterns in the copy-move detection problem solution [In Russian]. Computer Optics 2017; 41(1): 79-87. DOI: 10.18287/2412-6179-2017-41-1-79-81.
[5] Glumov NI, Kuznetsov AV, Myasnikov VV. The algorithm for copy-move detection on digital images [In Russian]. Computer Optics 2013; 37(3): 360-367.
[6] Burvin PS, Esther JM. Analysis of digital image splicing detection. IOSR Journal of Computer Engineering (IOSR-JCE) 2014; 16(2): 10-13. DOI: 10.9790/0661-162111013.
[7] Snigdha KM, Ajay AG. Image forgery types and their detection. International Journal of Advanced Research in Computer Science and Software Engineering 2015; 5(4): 174-178.
[8] Ferrara P, Bianchi T, De Rosa A, Piva A. Image forgery localization via fine-grained analysis of CFA artifacts. IEEE Transactions on Information Forensics and Security 2012; 7(5): 1566-1577. DOI: 10.1109/TIFS.2012.2202227.
[9] Popescu AC, Farid H. Exposing digital forgeries in color filter array interpolated images. IEEE Transactions on Signal Processing 2005; 53(10): 3948-3959. DOI: 10.1109/TSP.2005.855406.
[10] Gallagher AC, Chen T. Image authentication by detecting traces of demosaicing. CVPRW '08 2008. DOI: 10.1109/CVPRW.2008.4562984.
[11] Li L, Hue J, Wang X, Tian L. A robust approach to detect digital forgeries by exploring correlation patterns. Pattern Analysis and Applications 2015; 18(2): 351-365. DOI: 10.1007/s10044-013-0319-9.
[12] Bayram S, Sencar H, Memon N, Avcibas I. Source camera identification based on CFA interpolation. ICIP 2005; 3: 63-72. DOI: 10.1109/ICIP.2005.1530330.
[13] Bishop CM, Pattern recognition and machine learning. New York: Springer-Verlag; 2006. ISBN: 978-0-387-31073-2.
[14] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27(8): 861-874. DOI: 10.1016/j.patrec.2005.10.010.
[15] The original RAW-samples website. Source: ⟨http://rawsamples.ch⟩.
[16] CS. Centro Studi Progresso Fotografico. Dcraw. Source: ⟨http://www.centrostudiprogressofotografico.it/en/dcraw/⟩.
[17] Photo database. Source: ⟨http://www.zermatt.ch/ru/Media/Media-corner/Photo-database⟩.
[18] Columbia University Image Library (COIL-100). Source: ⟨http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php⟩.

Authors' information

Anna Aleksandrovna Varlamova (b. 1995) graduated with honors (2017) from Samara National Research University, Computer Science faculty. Programme – Information Security of Automated Systems. Nowadays she is postgraduate at Samara National Research University. Main research interests: image processing, digital image forgery detection, information security. E-mail: *varlamova.anna.95@mail.com* .

Andrey Vladimirovich Kuznetsov (b. 1987) graduated with honors (2010) from SSAU, majoring in Applied Mathematics and Informatics. He studied as a post-graduate student at SSAU from 2010 and received his PhD in Technical Sciences in 2013. Nowadays he is a researcher at IPSI RAS. His research interests are currently focused on image processing and analysis, pattern recognition, digital image forgery detection, geoinformatics. He has 37 publications, including 18 scientific papers and 1 monograph. Web-page: *http://nil97.ssau.ru/employee/detail.php?ID=35* . E-mail: *kuznetsoff.andrey@gmail.com* .

Received October 20, 2017. The final version – November 27, 2017.