# Анализ CFA-артефактов в задаче обнаружения искажений изображений

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

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086
[2]Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** Одним из часто осуществляемых видов подделки изображений является встраивание в изображение областей, скопированных из другого изображения. 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 CFA-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 выявлять встроенные области с минимальным размером 2×2.

## 1. Введение

Появление большого числа цифровых устройств, с помощью которых можно получать изображения, привело к снижению их стоимости и, как следствие, к их 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, например, 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.

Так, например, 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 свидетельствуют 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.

Данная работа посвящена исследованию 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.

## 2. 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, который представлен на рисунке 1.

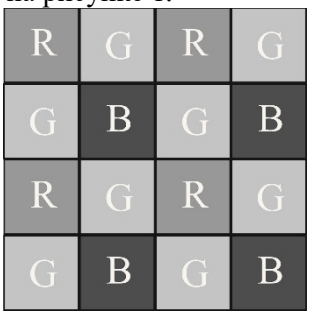



**Рисунок 1.** Фильтр Байера.

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, то 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 рисунке 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*).

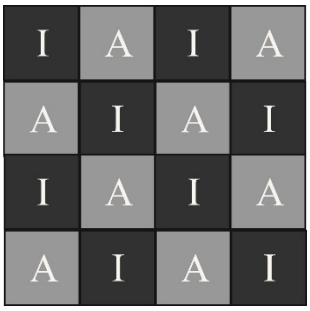



**Рисунок 2.** Зеленый канал изображения ( $A$ – отсчеты, полученные в результате съемки, $I$ – отсчеты, полученные путем интерполяции).

При использовании билинейной интерполяции, значения зеленого канала $s(x,y)$ определяются по формуле (1):

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

где $G_A(x,y)$ – 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.

Ядро интерполяции, используемое в настоящей работе, определено следующим образом:

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

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, было изображение искажено или нет.

## 2. Исследуемый метод обнаружения встраиваний

### *2.1 Теоретическое обоснование работы метода*

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

Пусть $s(x)$ – одномерный зеленый канал изображения, представленный в виде строки, который был получен путем интерполяции с использованием фильтра Байера. Тогда его значения определяются по формуле (3):

$$s(x)=\begin{cases}G_A(x), & x(\mathrm{mod}\,2)=0\\ G_I(x)=\sum_u h(u)G_A(x+u), & x(\mathrm{mod}\,2)\neq 0,\end{cases} \quad (3)$$

где $G_A(x)$ – значения отсчетов зеленого канала изображения, полученные непосредственно в ходе съемки;

$G_I(x)$ – интерполированные значения отсчетов зеленого канала изображения;

$h(u)$ – ядро 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, ошибка предсказания значений зеленого канала может быть определена по формуле (4):

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

где $k(u)$ – ядро предсказателя.

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 формулы (3) в формулу (4), ошибка предсказания может быть приведена к виду:

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

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 $\sum_u k(u)=\sum_u h(u)=1$ 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:

$$e(x)=\begin{cases}G_A(x)-\sum_u k(u)\sum_v h(v)G_A(x+u+v), & x(\mathrm{mod}\,2)=0\\ \sum_u (h(u)-k(u))G_A(x+u), & x(\mathrm{mod}\,2)\neq 0.\end{cases} \quad (5)$$

Полагая, что значения отсчетов, полученных в ходе съемки, независимы и одинаково распределены с математическим ожиданием (МО) $\mu_G$ и дисперсией $\sigma_G^2$, МО ошибки предсказателя может быть вычислено по формуле (6):

$$E[e(x)]=\begin{cases}\mu_G-\mu_G\sum_u k(u)\sum_v h(v), & x(\mathrm{mod}\,2)=0\\ \mu_G\left(\sum_u h(u)-\sum_u k(u)\right)=0, & x(\mathrm{mod}\,2)\neq 0.\end{cases} \quad (6)$$

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

$$Var[e(x)]=\sigma_G^2\left[\left(1-\sum_u k(u)h(-u)\right)^2+\sum_{t\neq 0}\left(\sum_u k(u)h(t-u)\right)^2\right]. \quad (7)$$

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

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

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 полученных отчетах значительно выше дисперсии ошибки, вычисленной для интерполированных отсчетов.

*2.2 Разработка метода обнаружения встраиваний на изображении*

Итак, 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 $\varepsilon$. Следовательно, для того, чтобы выявить наличие/отсутствие артефактов, возникающих после применения интерполяции, нужно вычислить дисперсию ошибки предсказания для отчетов $A$ и $I$.

Пусть $s(x, y)$ – зеленый канал изображения, тогда ошибка предсказания может быть определена по формуле (9):

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

где $k(u, v)$ – двумерное ядро предсказателя.

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 получении 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 $(2K+1) \times (2K+1)$, $c = 1 - \sum_{i=-K}^{K} \sum_{j=-K}^{K} \alpha^2(i, j)$ – 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) $\sigma_W^2 = \frac{K}{2}$. Под Гауссовским окном будем понимать сглаживающий двумерный фильтр, элементы которого распределены в соответствии с нормальным законом распределения. Значение дисперсия элементов $W$ выбрано экспериментально, путем сравнения с другими значениями: $\sigma_W^2 = \left\{ K, \frac{K}{2}, \frac{K}{4}, \frac{K}{8} \right\}$.

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

*2.3 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 $k,l=\overline{0,\left(\dfrac{N}{B}\right)-1}$.

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 «искаженности» 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):

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

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

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

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

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

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 $\varepsilon$, 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 $\varepsilon$ окрестности.

*2.3 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 $\varepsilon$ 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 есть 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 $B\times B$. Тогда функция плотности условного распределения для $M_1$:

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

где $\mu_1 > 0$ – неизвестное математическое ожидание распределения при справедливости гипотезы $M_1$;

$\sigma_1^2$ – неизвестная дисперсия.

Для $M_2$ функция плотности условного распределения:

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

где $\mu_2 = 0$ – математическое ожидание распределения при справедливости гипотезы $M_2$;

$\sigma_2^2$ – неизвестная дисперсия.

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: $\mu_1$, $\sigma_1^2$, $\sigma_2^2$, 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 Е-шаге алгоритма вычисляются вероятности принадлежности каждого отсчета $L(k,l)$ к каждой из моделей. При этом будем считать априорные вероятности каждой из гипотез равными:

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

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):

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

где $P\{L(k,l)|M_1\}$ – вероятность $L(k,l)$ при истинности гипотезы $M_1$;

$P\{L(k,l)|M_2\}$ – вероятность $L(k,l)$ при истинности гипотезы $M_2$.

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: $\mu_1$, $\sigma_1^2$, $\sigma_2^2$ 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, 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.

*2.3 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 по формуле (16), и $R_{FP}$ (false positive rate), характеризующего долю ложных обнаружений по формуле (17) [14].

$$R_{TP} = \frac{N_{m_{R2}}}{N_{R_2}}, \quad (16)$$

где $R_2$ – искаженная область изображения;

$N_{R_2}$ – общее количество блоков в области $R_2$;

$N_{m_{R2}}$ – количество верно обнаруженных искаженных блоков в области $R_2$.

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

где $R_1$ – неискаженная область изображения;

$N_{R_1}$ – общее количество блоков в области $R_1$;

$N_{m_{R1}}$ – количество ложно обнаруженных неискаженных блоков в области $R_1$.

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 коррелированны 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.

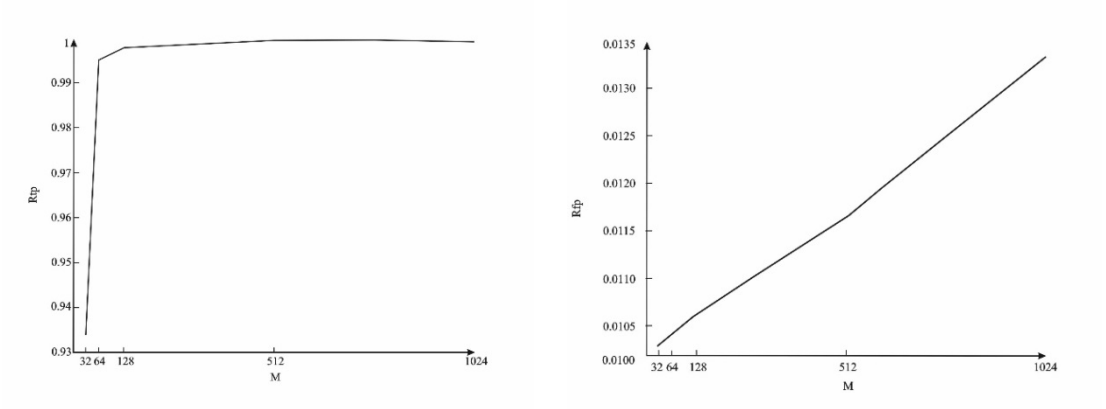
На рисунке 3 представлен пример изображения со вставкой произвольной формы и соответствующая ему карта вероятностей искажения, вычисленная блоками размера 8×8. Встроенная область была получена при помощи другой камеры.

(а) (б) (в)

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

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 рисунке 4.




(а) (б)

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

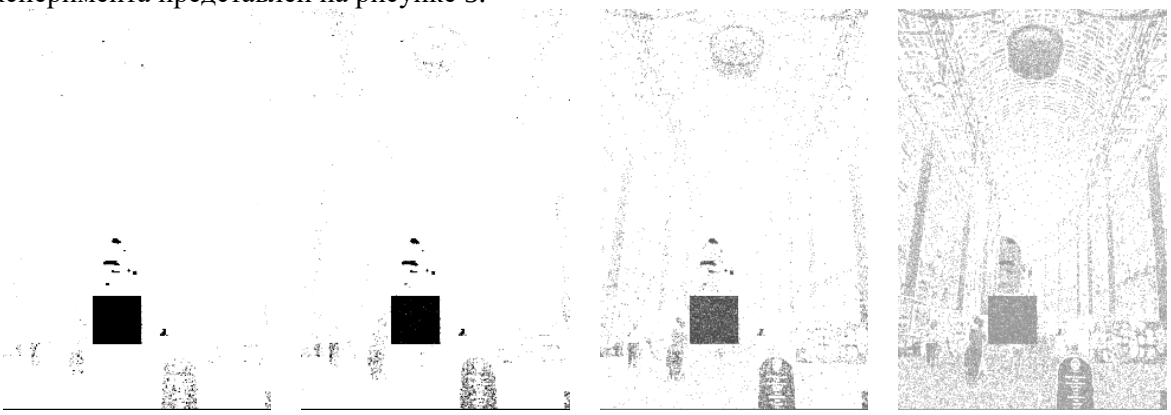
Результаты проведенного эксперимента показали, что с увеличением размера встраиваемой области качество обнаружения улучшается и достигает максимального значения – 1 при размере 512×512, однако и при минимальном исследуемом размере вставки 32×32 $R_{TP} = 0{,}93$,

что характеризует высокое качество обнаружения. При этом число ложно обнаруженных неискаженных блоков изображения незначительно растет и при размере встраиваемой области 1024×1024 составляет $R_{FP} = 0{,}0133$.

*3.1 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.

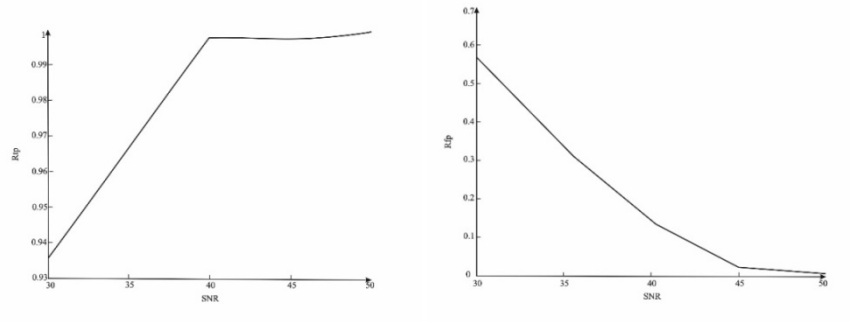
В рамках исследований к каждому из них был добавлен аддитивный гауссовский шум со значением отношения сигнал/шум ($SNR$ (дБ)): 30, 35, 40, 45, 50. Пример проведенного эксперимента представлен на рисунке 5.



(а) (б) (в) (г)

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

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



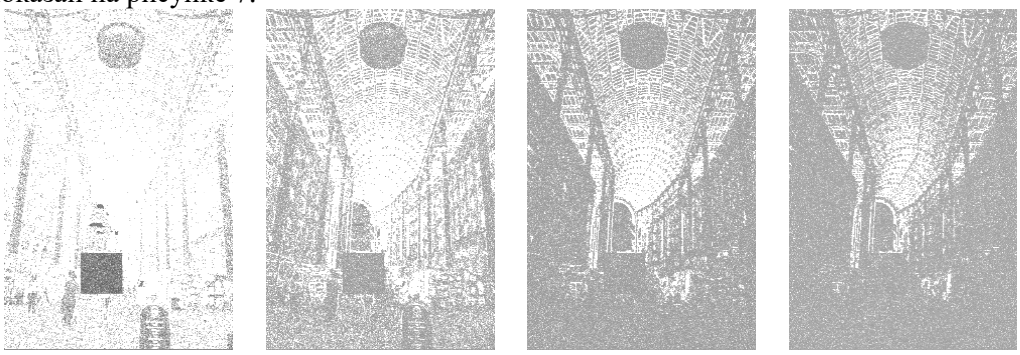

(а) (б)

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

На рисунке 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 различными значениями параметра качества $Q$, изменяемого в

пределах от 0 до 100. Пример работы алгоритма при значениях параметра качества $Q$ = 100, 98, 96, 94 показан на рисунке 7.

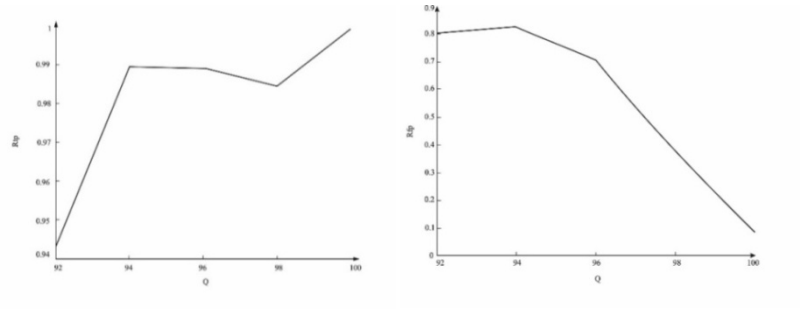


(а) (б) (в) (г)

**Рисунок 7.** Карта искажений изображения со встроенной областью после применения сжатия со значением параметра качества $Q$: а) $Q$ = 100, б) $Q$ = 98, в) $Q$ = 96, г) $Q$ = 94.

На рисунке 8 представлена зависимость характеристик качества обнаружения искажений от значения параметра качества сжатия JPEG $Q$: $R_{TP}(Q)$ 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.

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.



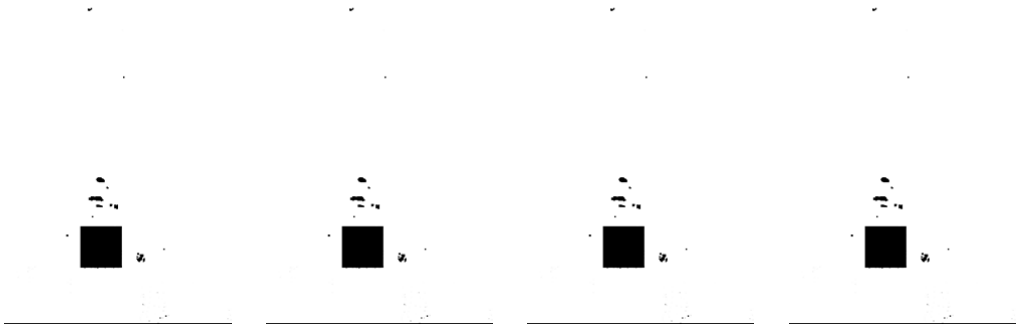
(а) (б)

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

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.

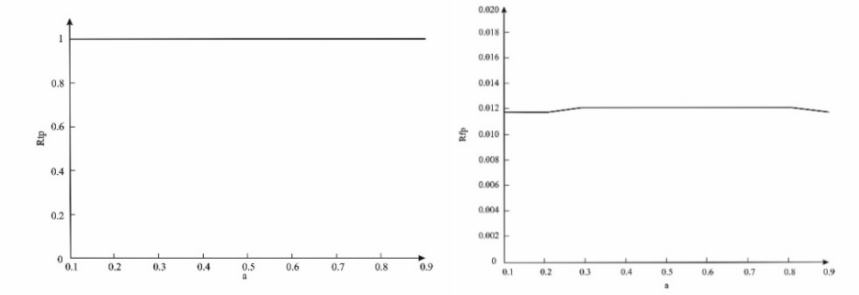
Изображения, вводимые в компьютер, часто являются малоконтрастными, то есть у них вариации функции яркости малы по сравнению с ее средним значением. Реальный динамический диапазон яркостей $[f_{\min}, f_{\max}]$ для таких изображений оказывается намного меньше допустимого диапазона (шкалы яркости). Задача 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.



(а) (б) (в) (г)

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

На рисунке 9 представлен пример работы метода в случае, если к искаженному изображению было применено линейное контрастирование с различными значениями параметров преобразования. На представленном рисунке значение параметра $b$ было зафиксировано: $b$=20.




(а) (б)

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

На рисунке 10 представлена зависимость характеристик качества обнаружения от значения параметра преобразования $a$: $R_{TP}(a)$ и $R_{FP}(a)$.

Аналогичным образом был проведен эксперимент, в котором значение параметра $a$ было зафиксировано, а значения параметра $b$ изменялось.

Результаты показали, что алгоритм устойчив к применению линейного контрастирования и результат обнаружения встроенной области не зависит от значений параметров линейного контрастирования.

## 4. Заключение

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).

**5. Благодарности**

Работа выполнена при поддержке Федерального агентства научных организаций (соглашение № 007-ГЗ/Ч3363/26) и грантов РФФИ 18-07-01312 А "Методы нелинейного снижения размерности гиперспектральных изображений и их применение", 16-29-09494 офи-м "Методы компьютерной обработки мультиспектральных данных дистанционного зондирования Земли для определения ареалов растений в специальных криминалистических экспертизах".

**6. Литература**

[1] Как бороться с подделками фотоотчетов [Электронный ресурс]. – Режим доступа https://club.esetnod32.ru/articles/analitika/kak-borotsya-s-poddelkami-fotootchetov/ (14.08.2017).
[2] Choi, C. Estimation of color modification in digital images by CFA pattern change / C. Choi, H. Lee, H. Lee // Forensic Science International. – 2013. – Vol. 226. – P. 94-105.
[3] Chakraverti, A.K. A Review on Image Forgery & its Detection Procedure / A.K. Chakraverti, V. Dhir // International Journal of Advanced Research in Computer Science. – 2017. – № 4. – Vol. 8. – 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(2). – P. 10-13.
[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(4). – P. 174-178.
[8] Ferrara, P. Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts / P. Ferrara, T. Bianchi, A. Rosa, A. Piva // IEEE Transactions on Information Forensics and Security. – 2012. – Vol. 7(5). – P. 1566-1577.
[9] Popescu, A. Exposing Digital Forgeries in Color Filter Array Interpolated Images / A. Popescu, H. Farid // IEEE Transactions on Signal Processing. – 2005. – Vol. 53(10). – P 3948-3959.
[10] Gallagher, A. Image authentication by detecting traces of demosaicing / A. Gallagher, T. Chen // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. – 2008. – P. 1–8.
[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(2). – P. 351–365.
[12] Bayram, S. Source camera identification based on CFA interpolation / S. Bayram, H. Sencar, N. Memon, I. Avcibas // IEEE Image Processing. – 2005. – Vol. 3. – P. 63-72.
[13] Bishop, C.M. Pattern Recognition and Machine Learning / C.M. Bishop // Springer Verlag. – 2006. – 738 p.
[14] Fawcett T. An introduction to ROC analysis / T. Fawcett // Elsevier. – 2006. – P. 861-874.
[15] The original RAW-Samples Website [Electronic resource]. – Access mode: http://rawsamples.ch (27.08.2017).
[16] Dcraw [Electronic resource]. – Access mode: http://www.centrostudiprogressofotografico.it/en/dcraw/ (27.08.2017).
[17] Photo database [Electronic resource]. – Access mode: http://www.zermatt.ch/ru/Media/Media-corner/Photo-database (30.08.2017).
[18] Columbia University Image Library (COIL-100) [Electronic resource]. – Access mode: http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php (30.08.2017).

# CFA Artifacts Analysis for Image Forgery Detection

**A.A.Varlamova[1], A.V. Kuznetsov[1,2]**

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

**Abstract.** One of the widespread image forgery techniques is image splicing. It represents pasting in an image parts of other images. In this paper, one of the methods of image splicing localization based on analysis of CFA-artifacts that appear on an image during the capturing process is de-scribed. A feature characterizing the presence/absence of CFA artifacts for each image block is measured. The obtained values of the feature define probability of each block to be embed. 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 showed that the suggested method reveals embed regions of different shape, size and nature in images. The method possesses stability against additive Gaussian noise and linear enhancement, but it is not steady against JPEG compression. The advantage of the method is the ability to localize splicing regions even at the smallest 2×2 block level.

**Keywords:** Image Forgery, Color Filter Array, Bayer Filter, Interpolation, Artifact, Tampering Probability Map.