

УДК 004.056.5

А. А. Кобозева, С. Н. Григоренко

Одесский национальный политехнический университет
Проспект Шевченко, 1, 65044 Одесса, Украина

Теоретические основы нового подхода к решению задачи выявления результатов клонирования в цифровом изображении, сохраненном в формате с потерями

Предложены теоретические основы нового подхода к решению задачи выявления результатов клонирования в цифровом изображении, сохраненном после изменения в формате с потерями. Подход основан на учете малости изменения объема цилиндрического тела, ограниченно-го сверху графиком интерполирующей функции для матрицы изображения, снизу — плоскостью XOY, в процессе сжатия. Показателем наличия области клонирования/прообраза клона является близость значений двойных интегралов от интерполирующей функции по соответствующим подобластям области определения функции, задающей исходное изображение. Выявлены характерные отличия для сумм элементов матриц абсолютной разности пар непересекающихся блоков матрицы изображения, позволяющие локализовать область клона и его прообраза. Выявленные отличия имеют место независимо от особенностей и характеристик изображения в части прообраза/клона; их последующая количественная оценка даст возможность для разработки метода, решающего поставленную задачу.

Ключевые слова: цифровое изображение, клонирование, формат хранения изображения с потерями, нарушение целостности, матрица цифрового изображения.

Введение

Информатизация и компьютеризация современного общества наряду со значительным положительным эффектом принесла и новые проблемы, связанные с легкостью неавторизованной модификации цифровых информационных контентов, которые в настоящее время являются наиболее часто используемыми в широком спектре областей человеческой деятельности. Широко распространенные графические редакторы, такие как Adobe Photoshop, Coral Draw и другие, их общедоступность, легкость в использовании приводят к необходимости повышения

© А. А. Кобозева, С. Н. Григоренко

эффективности методов пассивной защиты информации, поскольку при использовании любого информационного контента в целях, отличных от развлекательных, необходимым условием является отсутствие его неавторизованных изменений.

Задача проверки целостности цифровых контентов, в частности, изображений (ЦИ), которые и рассматриваются далее в настоящей работе, не является новой. Основы общего подхода к решению данной задачи без привязки к конкретному виду нарушения целостности были предложены одним из авторов в [1], конкретные виды нарушений и методы защиты от них рассмотрены в [2–5]. Однако при всем многообразии существующие методы выявления нарушения целостности ЦИ не являются удовлетворительными, не позволяют говорить о решении рассматриваемой задачи в целом. Они постоянно находятся в состоянии «догоняющего», часто не успевая за предоставляемыми современными графическими редакторами, программными комплексами и возможностями изменения оригинальных цифровых контентов в тех или иных целях.

Наиболее часто используемым способом несанкционированного изменения ЦИ на сегодняшний день является клонирование, при котором одна область изображения (прообраз) копируется и вставляется в другую область (клон) этого же изображения, заменяя собой его оригинальную часть. Упомянутый инструмент реализован практически во всех современных графических редакторах, заставляя уделять значительное внимание разработкам методов и алгоритмов выявления его результатов [5–8]. Одним из наиболее эффективных методов выявления клонирования и локализации области клона в изображении в настоящий момент является метод, предложенный в [5], однако его высокая эффективность имеет место только в случае сохранения измененного ЦИ в формате без потерь, поскольку сохранение с потерями вносит дополнительные к уже имеющимся возмущения в оригинальное ЦИ.

Инвариантность по отношению к формату сохранения измененного изображения декларируется в некоторых существующих методах решения обсуждаемой задачи [6, 7], однако и эти методы не лишены значительных недостатков: они оказываются бессильными в случае, когда область клонирования не содержит контуров, что часто имеет место на практике, когда при помощи клонирования «скрывается» объект, присутствующий на оригинальном ЦИ.

Все вышесказанное говорит об актуальности усовершенствования имеющихся и разработки новых методов и реализующих их алгоритмов для выявления результатов клонирования в ЦИ, которые бы были свободны от недостатков существующих, в частности, эффективность которых не зависела бы от формата сохранения (с потерями, без потерь) измененного ЦИ.

Цель статьи и постановка задачи

Целью работы является разработка общих теоретических основ метода решения задачи выявления (областей) нарушения целостности ЦИ, полученного путем клонирования, эффективного в условиях сохранения измененного ЦИ в формате с потерями, независимо от особенностей и характеристик изображения в части прообраз/клона.

Предполагается, что область клонирования не подвергается никакой постобработке, за исключением последующего сохранения всего полученного измененного ЦИ в формате с потерями, в качестве которого в работе рассмотрен формат Jpeg.

Для достижения поставленной цели необходимо решить следующую задачу: определение качественной характеристики ЦИ, одинаковой/сравнимой для области клона и его прообраза в условиях возможного сохранения измененного изображения в формате с потерями.

Основная часть

Для простоты изложения, не ограничивая общность рассуждений, ниже рассматриваются монохромные ЦИ. В случае цветных изображений (схема RGB) все нижеприведенные рассуждения могут быть применены по отдельности для каждой цветовой компоненты.

Произвольное монохромное ЦИ является результатом дискретизации области определения D и квантования значений некоторой функции двух переменных $f(x, y)$, $(x, y) \in D$ [9], формальным представлением чего служит $(n \times m)$ -матрица F с элементами $f_{ij} \in \{0, 1, \dots, 255\}$, $i = \overline{1, n}$, $j = \overline{1, m}$, или трехмерная поверхность (рис. 1), являющаяся результатом интерполирования элементов F , рассматриваемых в качестве узлов. Обозначим интерполирующую F -функцию $g(x, y)$, $(x, y) \in D$. Заметим, что в общем случае $g(x, y)$ хоть и является приближающей для исходной функции $f(x, y)$, но не является для нее интерполирующей в силу проведенного для значений $f(x, y)$ квантования.

В настоящий момент увеличение объемов хранимой, обрабатываемой, передаваемой информации заставляет прибегать к ее сохранению в форматах с потерями. Пересылка ЦИ, видео, аудио в сети Интернет сегодня практически всегда осуществляется в сжатом виде. Сжатие ЦИ включает в себя предварительное разбиение его матрицы на непересекающиеся (8×8) -блоки с последующим переводом каждого блока в область преобразования (для Jpeg — в область дискретного косинусного преобразования (ДКП)), квантование коэффициентов преобразования с последующим округлением, энтропийным кодированием результата [9]. Присутствующее в процессе сжатия округление делает его необратимым, а сохранение изображения в формате с потерями неизбежно вносит дополнительные возмущения в его матрицу. Эти возмущения тем больше, чем больше коэффициент сжатия или чем меньше коэффициент качества QF , который может принимать значения: $QF \leq 100$.

Для определенности далее рассматриваются оригинальные ЦИ, которые хранятся либо в формате без потерь, либо с потерями при $75 \leq QF \leq 85$ (такой коэффициент качества является наиболее часто используемым на практике, давая, как правило, приемлемое соответствие между объемом сохраненного ЦИ и субъективной оценкой его качества). Предполагается, что после проведения неавторизованных изменений ЦИ злоумышленником, оно будет сохраняться в формате с по-

терями с коэффициентом качества $QF \in \{75,80\}$ (во избежание привлечения дополнительного внимания к измененному (фальсифицированному) изображению).

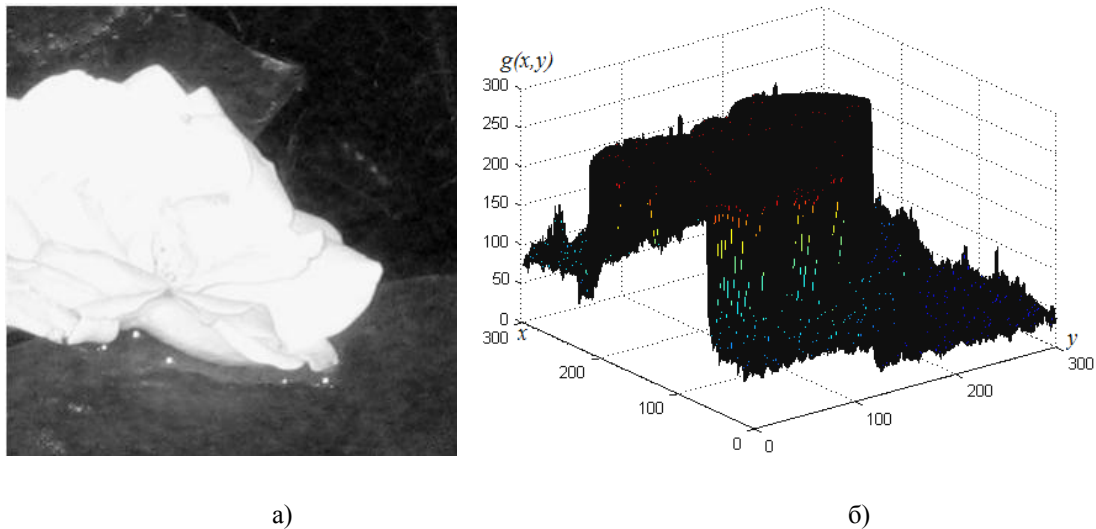


Рис. 1. Иллюстрация представления монохромного ЦИ в виде трехмерной поверхности:
 а) исходное ЦИ; б) соответствующая поверхность, представляющая собой график интерполяционного сплайна 1-й степени для F -функции $g(x, y), (x, y) \in D$

Предположим, что размеры области клонирования таковы, что она содержит в себе целиком хотя бы один (8×8) -квадрат (который не обязательно должен быть одним из блоков, полученных в результате стандартного разбиения матрицы ЦИ при сохранении в Jpeg; здесь речь идет лишь о размерах области).

Клонирование в изображении неизбежно приводит к дублированию части поверхности, являющейся графиком функции $g(x, y), (x, y) \in D$, отвечающей прообразу, в область клона, что приводит к соответствующим изменениям в матрице F : дублированию подматрицы, отвечающей значениям яркости пикселей прообраза, в другое место F , отвечающее после дублирования клону. Обозначим новую функцию $\bar{g}(x, y), (x, y) \in D$, а соответствующую матрицу, результатом интерполирования элементов которой и является $\bar{g}(x, y)$, обозначим \bar{F} . При сохранении измененного изображения в формате без потерь при отсутствии какой-либо обработки области клона части поверхностей, отвечающие клону и его прообразу, являются идентичными, чего в общем случае нельзя утверждать в случае последующего сохранения ЦИ с потерями. При несовпадении сетки разбиения матрицы ЦИ на блоки для области клона и его прообраза (рис. 2 — (K) и (P) соответственно) идентичность соответствующих частей поверхностей будет нарушена в силу отличия (пусть даже незначительного) в значениях яркостей соответствующих пикселей в результате квантования и округления коэффициентов ДКП блоков, оказавшихся несоответствующими для областей прообраза и клона. Соответствующую интерполирующую функцию обозначим $\overline{\bar{g}}(x, y), (x, y) \in D$, а матрицу пикселей $\overline{\bar{F}}$.

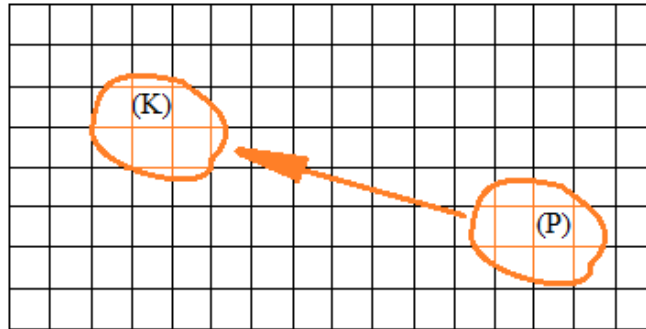


Рис. 2. Иллюстрация возможного несовпадения сеток разбиения области клона и его прообраза при сохранении измененного ЦИ в формате Jpeg

В общем случае: $\bar{g}(x, y) \neq \overline{\bar{g}}(x, y)$, но в силу сделанных выше допущений, части поверхности, отвечающие клону и его прообразу, геометрически будут отличаться друг от друга незначительно (зрительно соответствующие части изображения вообще различаться не будут, что подтверждается представительным вычислительным экспериментом, описание которого не входит в настоящую работу), сохраняя сравнимыми между собой все свои основные характеристики: локальные, глобальные минимумы/максимумы, характер монотонности вдоль одинаковых направлений и т.д. Таким образом, близкими по своим основным геометрическим характеристикам окажутся цилиндрические поверхности с образующими, параллельными оси аппликат, ограниченные графиками функции $\bar{g}(x, y)$, $(x, y) \in K$, и $\overline{\bar{g}}(x, y)$, $(x, y) \in P$, где $K \in D$, $P \in D$ — подобласти области определения D , отвечающие (K) и (P) соответственно, и координатной плоскостью $ХОУ$. Одним из численных показателей этой близости является близость значений объемов тел, ограниченных указанными выше цилиндрическими поверхностями, т.е., с учетом, что $\bar{g}(x, y) \geq 0$, $(x, y) \in D$:

$$\iint_K \bar{g}(x, y) dx dy \approx \iint_P \overline{\bar{g}}(x, y) dx dy. \quad (1)$$

Таким образом, если (K) и (P) — область клона и его прообраза соответственно, то $\iint_K \bar{g}(x, y) dx dy - \iint_P \overline{\bar{g}}(x, y) dx dy \approx 0$.

Обозначим функцию $\overline{\bar{g}}(x, y)$, определенную на области P (области K) через $\overline{\bar{g}}_{(P)}(x, y)$ ($\overline{\bar{g}}_{(K)}(x, y)$). В силу непрерывности интерполирующей функции $\overline{\bar{g}}(x, y)$ на области D и вытекающей отсюда интегрируемости на D , а потому и на P и K , используем формулы численного интегрирования для вычисления интегралов в (1), заменяя их значения на значения соответствующих интегральных сумм с учетом принятых обозначений. Поскольку K и P имеют одинаковую форму, разобьем их естественным образом на частичные подобласти, каждая из которых

отвечает одному пикселю ЦИ. Для простоты изложения, никак не ограничивая общность рассуждений, будем считать площадь каждой такой подобласти единичной. Тогда:

$$\iint_K \bar{g}(x, y) dx dy \approx \sum_{(x, y) \in K} \bar{g}_{(K)}(x, y); \quad \iint_P \bar{g}(x, y) dx dy \approx \sum_{(x, y) \in P} \bar{g}_{(P)}(x, y). \quad (2)$$

Суммы в правых частях формул (2) представляют собой суммы элементов подматриц K_F и P_F матрицы \bar{F} , отвечающих областям клона и его прообраза соответственно. Пусть Z с элементами z_{ij} — матрица разности K_F и P_F : $Z = K_F - P_F$. Из (1) вытекает, что при имеющем место клонировании с последующим сжатием ЦИ ($QF \in \{75, 80\}$) будет иметь место соотношение:

$$\sum_{i, j} z_{ij} \approx 0. \quad (3)$$

Полученное соотношение скорее имеет не количественный, а качественный характер: степень близости $\sum_{i, j} z_{ij}$ к нулю будет зависеть от коэффициента качества QF , использованного при сжатии измененного ЦИ: чем больше значение QF , тем меньше $\sum_{i, j} z_{ij}$ отличается от нуля.

С учетом того, что площади проекций частей поверхности изображения (K) и (P) на плоскость XOY совпадают, и в предположении, что клонирование действительно имеет место (функцию $\bar{g}_{(P)}(x, y)$ можно рассмотреть на $(x, y) \in K$), имеем:

$$\bar{g}_{(P)}(x, y) \approx \bar{g}_{(K)}(x, y), \quad (x, y) \in K, \\ \iint_K \bar{g}(x, y) dx dy - \iint_P \bar{g}(x, y) dx dy = \iint_K \left(\bar{g}_{(K)}(x, y) - \bar{g}_{(P)}(x, y) \right) dx dy \approx 0. \quad (4)$$

Чтобы избежать принципиальной возможности малости значения $\sum z_{ij}$ за счет взаимного уничтожения величин (возможно значительно отличающихся от 0) разных знаков в (3), изменения знака подинтегральной функции в правой части (4) (что принципиально может привести к близости значения интеграла к 0, даже если в некоторых подобластях области интегрирования и наблюдается значительное отличие подинтегральной функции от 0), уточним соотношения (3) и (4), имеющие место для области клона и ее прообраза, следующим образом:

$$\sum_{i, j} |z_{ij}| \approx 0, \\ \iint_K \left| \bar{g}_{(K)}(x, y) - \bar{g}_{(P)}(x, y) \right| dx dy \approx 0, \quad (5)$$

при этом левая часть соотношения (5) представляет объем цилиндрического тела с образующей, параллельной оси аппликат, отличающего цилиндрическое тело, ограниченное сверху $\overline{g}_{(p)}(x, y)$, от тела, ограниченного сверху $\overline{g}_{(k)}(x, y)$ (с образующими, параллельными оси аппликат, и одинаковыми основаниями, лежащими в плоскости XOY), требуя в случае клонирования, чтобы этот объем был незначительным (нулевым).

Далее предполагается, что в измененном изображении области клона и его прообраза не пересекаются. С учетом всего вышесказанного процесс выявления результатов имеющего место клонирования в матрице анализируемого ЦИ возможен путем организации поиска пар всевозможных непересекающихся (8×8) -блоков, для которых сумма элементов матрицы их абсолютной разности будет иметь наименьшее значение. Такие блоки являются подозрительными на то, чтобы принадлежать областям клона и прообраза. Если никаких уточнений для локализации поиска нет, то необходимо проверить всевозможные пары различных блоков матрицы ЦИ (различные блоки генерируются сдвигом на 1 пиксель влево, вправо, вверх, вниз в пределах анализируемого ЦИ).

Один из возможных способов организации поиска пары блоков «клон-прообраз» может быть следующим. Каждому элементу \overline{f}_{ij} $(n \times m)$ -матрицы \overline{F} анализируемого ЦИ поставим в соответствие (8×8) -блок B_{ij} этой матрицы (рис. 3), для которого \overline{f}_{ij} — элемент, находящийся в левом верхнем углу блока, в случае, когда весь блок находится в границах матрицы ($i \leq n - 7, j \leq m - 7$).

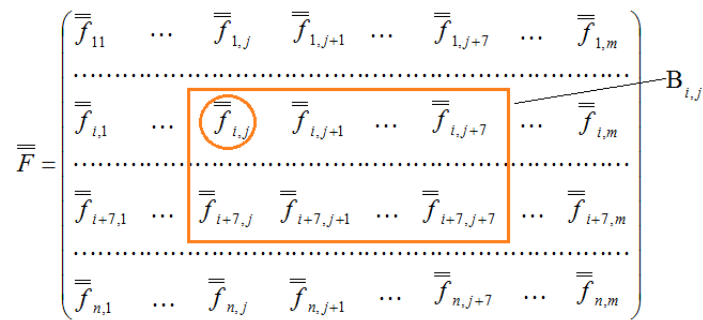


Рис. 3. Иллюстрация соответствия элементам \overline{f}_{ij} матрицы \overline{F} блоков B_{ij}

Каждому из полученных блоков B_{ij} поставим в соответствие матрицу связи $M^{(i,j)}$ с элементами $m^{(i,j)}_{k,l}, k = \overline{1, n-7}, l = \overline{1, m-7}$, элемент $m^{(i,j)}_{k,l}$ которой будут отражать связь блока B_{ij} и блока B_{kl} в следующем смысле: $m^{(i,j)}_{k,l} = \sum_{t,p=1}^8 r_{tp}$, где $r_{tp}, t, p = \overline{1, 8}$, — элементы матрицы R , полученной следующим образом: $R = |B_{ij} - B_{kl}|$.

Элементы матрицы $M^{(i,j)}$ рассмотрим как узлы интерполяции для некоторой функции $h^{(i,j)}(x,y)$ двух переменных. Построим по имеющимся узлам интерполяционный сплайн $H^{(i,j)}(x,y)$ первой степени. Получаемая кусочно-полиномиальная функция $H^{(i,j)}(x,y)$ всегда имеет глобальный минимум, равный нулю, в точке с координатами $k=i, l=j$, поскольку матрица R в этом случае будет точно нулевой, что находит свое отражение на ее графике (рис. 4,б, 5,б). В остальном характеристики поверхности-графика функции $H^{(i,j)}(x,y)$ будут иметь качественные отличия для блока B_{ij} в случае, когда он является частью области клона или прообраза клона, от случая, когда такая ситуация места не имеет. Для первой ситуации поверхность, соответствующая B_{ij} , будет иметь локальный минимум, значение которого хотя и отлично от нуля (в случае сохранения измененного ЦИ в формате с потерями), но значительно меньше других значений матрицы $M^{(i,j)}$ (включая значения матрицы в единичной окрестности глобального минимума). Для второй ситуации, если локальные минимумы и будут иметь место, то их значения будут превосходить элементы из малой окрестности глобального минимума функции $H^{(i,j)}(x,y)$, интерполирующей значения $M^{(i,j)}$. Иллюстративный пример сделанных выше заключений для блока B_{ij} , являющегося блоком клона в ЦИ, сохраненном после операции клонирования в формате с потерями, приведен на рис. 4; для блока, не принадлежащего ни области клона, ни области прообраза — на рис. 5. С учетом упомянутых недостатков методов, предложенных в [6, 7], для иллюстрации сделанных на основе теоретических рассуждений заключений намеренно использованы ЦИ, где клонирование проводится в части, не содержащей контуры и мелкие детали.

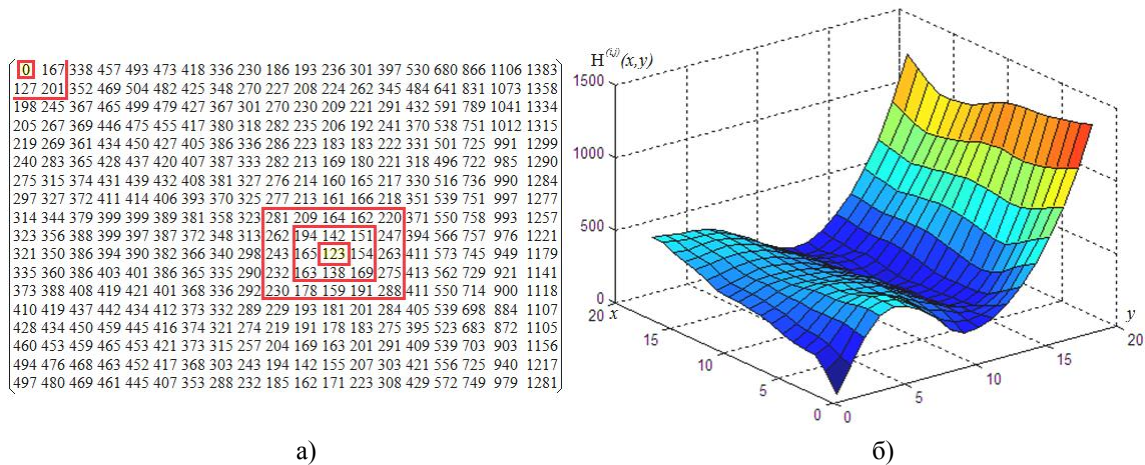


Рис. 4. Блок B_{ij} принадлежит области клона: а) подматрица матрицы $M^{(i,j)}$, содержащая информацию о B_{ij} и блоке, принадлежащем области прообраза клона;
 б) часть поверхности, отвечающей графику функции $H^{(i,j)}(x,y)$, содержащая точки ее глобального и локального минимумов

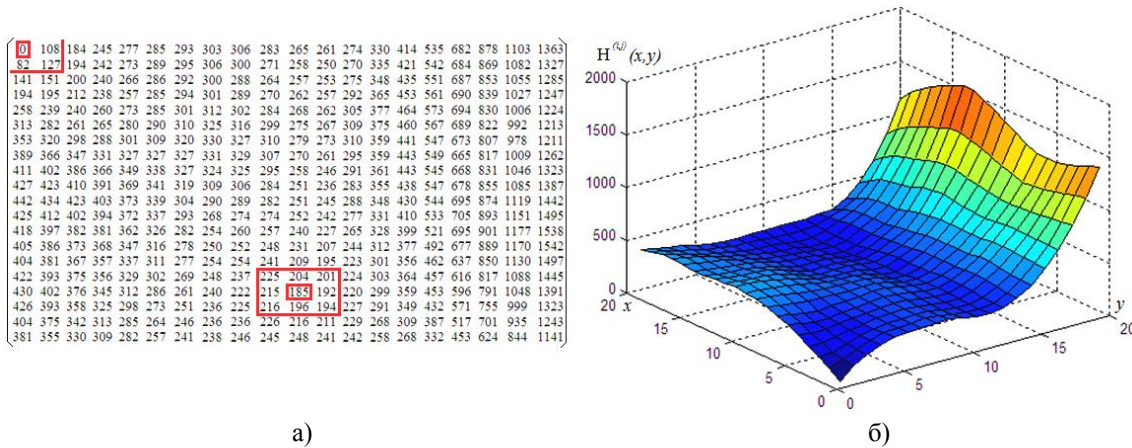


Рис. 5. Блок B_{ij} не принадлежит области клона/прообраза: а) подматрица матрицы $M^{(i,j)}$, содержащая информацию о блоке B_{ij} ; б) часть поверхности, отвечающей графику функции $H^{(i,j)}(x,y)$, содержащая точки ее глобального и локального минимумов

Как видно из приведенных примеров, поверхности, отвечающие $H^{(i,j)}(x,y)$, качественно отличаются в случае принадлежности/непринадлежности блока B_{ij} области клона/прообраза. Очевидно, что выявленные отличия будут иметь место независимо от особенностей и характеристик изображения в части прообраза/клона; будут выражены ярче при сохранении измененного изображения в формате без потерь.

Выводы

В работе предложены основы нового подхода к решению задачи выявления результатов клонирования в ЦИ, сохраненном после изменения в формате с потерями. Качественным показателем проведенного клонирования в ЦИ является выявленная близость значений двойных интегралов по некоторым областям изображения (которые и трактуются как области клона и его прообраза) от функции двух переменных, являющейся интерполирующим сплайном для матрицы изображения, что приводит к появлению локального минимума на поверхности, отвечающей интерполяционному сплайну первой степени для матрицы соответствующего блока из клона/прообраза клона в той части, где находится блок из прообраза/клона. Величина этого локального минимума (в случае отличия от нуля) не превосходит значений в единичной окрестности глобального минимума. Количественная оценка величины локального минимума даст возможность для разработки соответствующего метода, выявляющего результаты клонирования, в том числе, в условиях сжатия ЦИ с потерями.

1. Кобозева, А.А. Общие принципы построения методов выявления нарушения целостности цифрового изображения. Информационные технологии и защита информации в информационно-

коммуникационных системах: монография / под ред. В.С.Пономаренко. — Х.: Вид-во ТОВ «Щедра садиба плюс», 2015. — 486 с.

2. *Rey C.* A survey of watermarking algorithms for image authentication / C. Rey, J.-L. Dugelay // EURASIP J. Appl. Signal Process. — 2002. — N 1. — P. 613–621.

3. *Copy-move* forgery detection and localization by means of robust clustering with J-linkage / I. Amerini, L. Ballan, R. Caldelli [et al.] // Signal Processing. — 2013. — Vol. 28. — N 6. — P. 659–669.

4. *Farid H.* Image Forgery Detection / H. Farid // IEEE Signal Processing Magazine. — 2009. — Vol. 6. — P. 16–25.

5. *Лебедева Е.Ю.* Метод локализации и идентификации оригинальной и клонированной областей изображения / Е.Ю. Лебедева // Інформатика та математичні методи в моделюванні. — 2014. — Т. 4, № 1. — С. 76–84.

6. *Huang H.* Detection of copy-move forgery in digital images using SIFT algorithm / H. Huang, W. Guo, Y. Zhang // IEEE Pacific-Asia Workshop on Computational Intell. and Industrial Application. — 2008. — Vol. 2. — P. 272–276.

7. *Pooja S.* Detecting Region Duplication Forgery in Digital Image using SIFT Features / S. Pooja, S. Kotkar, S. Shriramwar // International Journal of Current Engineering and Technology. — 2014. — Vol. 4, N 3. — P. 1437–1440.

8. *Pan X.* Region duplication detection using image feature matching / X. Pan, S. Lyu // IEEE Transactions on Information Forensics and Security. — 2010. — Vol. 5, N 4. — P. 857–867.

9. *Гонсалес Р.* Цифровая обработка изображений / Р. Гонсалес, Р. Вудс; пер. с англ. П.А. Чочиа. — М.: Техносфера, 2006. — 1070 с.

Поступила в редакцию 10.12.2015