УДК 003.26.09:004.056.55

Автоматизация внедрения в изображения неизвлекаемых цифровых меток для защиты авторских прав с использованием стеганографии

Гросс Яна Александровна yana.gross.05@mail.ru
Волосатова Тамара Михайловна tamarav@bmstu.ru
Князева Светлана Вадимовна knyazevasv@bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Россия

Исследование посвящено разработке устойчивого метода цифровых водяных знаков для защиты авторских прав на изображения. Анализируется уязвимость классических методов к статистическому стегоноанализу и обработке. Предлагается двухуровневая система: скрытое «ядро», внедряемое в изображение стеганографическим методом, и механизм визуализации, который динамически накладывает видимый водяной знак при обнаружении «ядра» детектором на платформе.

***Ключевые слова**: цифровые водяные знаки, стеганография, защита авторских прав, LSB-метод, DCT-преобразование, стегоанализ*

Введение. В условиях цифровой эпохи, характеризующейся лавинообразным ростом объемов контента, защита интеллектуальной собственности приобретает критическую важность. Простота несанкционированного копирования и распространения цифровых произведений требует разработки надежных механизмов для подтверждения авторства и противодействия незаконному использованию. Одним из ключевых инструментов для решения этой задачи являются цифровые водяные знаки (ЦВЗ), обеспечивающие скрытое внедрение идентификационной информации в сам медиафайл.

Современные стеганографические методы, включая LSB-замещение, уязвимы к стегоноанализу. Атаки, такие как метод хи-квадрат или RS-анализ, обнаруживают внедрение сообщений через статистические аномалии в младших битах изображений, что ставит под угрозу скрытность цифровых водяных знаков.

Ключевой целью данного исследования является разработка и анализ метода внедрения цифрового водяного знака, который был бы неотличим от естественного шума изображения и, как следствие, устойчив к обнаружению и удалению. Пользователь получает изображение, в котором надежно и незаметно для глаза зашита информация (например, фамилия и инициалы автора), удалить эту информацию, не испортив изображение, невозможно, потому что неизвестно, что именно нужно изменить. В то же время важно, чтобы этот скрытый идентификатор можно было легко и быстро прочесть. Для этого необходим детектор — программа или скрипт. Детектор знает тот самый секретный ключ, который использовался для псевдослучайного распределения

битов, поэтому он может точно определить, в каких именно пикселях или коэффициентах искать информацию.

Реализация видимой вотермарки при публикации решается на уровне платформы, куда загружается изображение:

– сервис/платформа (например, социальная сеть или фотохостинг) после загрузки изображения автоматически запускает детектор;

– детектор ищет в изображении «ядро» ЦВЗ с помощью известного ему ключа;

– если водяной знак обнаружен и успешно прочтен, система динамически накладывает поверх изображения видимую графическую вотермарку (логотип, имя автора, знак копирайта и так далее). Для пользователя это выглядит так: он загрузил «чистую» картинку, а на сайте она отобразилась уже с водяным знаком. При этом скрытая информация все еще содержится в рисунке и не позволяет удалить видимый водяной знак.

Обзор методов. Метод замены наименьшего значащего бита (LSB) — самый простой и интуитивный. Он основан на том, что младшие биты цветовых каналов минимально влияют на восприятие изображения. Метод предполагает замену этих битов на биты скрытого сообщения. Это позволяет спрятать сообщение размером примерно равным $(\text{Ширина} \times \text{Высота} \times 3)/8$ байт. Например, для изображения 1000×1000 пикселей это около 375 КБ. Однако существенным недостатком остается низкая устойчивость к любым преобразованиям изображения, включая сжатие с потерями, изменение размеров или цветокоррекцию. Кроме того, факт внедрения легко обнаруживается статистическими методами, такими как критерий хи-квадрат или RS-анализ, которые выявляют аномалии в распределении значений соседних цветов [1].

Более устойчивыми считаются методы, работающие в частотной области, такие как алгоритм Коха на основе дискретного косинусного преобразования. Здесь изображение разбивается на блоки, для каждого из которых вычисляются спектральные коэффициенты. Скрываемые данные внедряются в коэффициенты средних частот, манипулируя их соотношениями. Этот подход обеспечивает значительно более высокую устойчивость к сжатию и геометрическим искажениям, поскольку основные преобразования затрагивают, прежде всего, высокочастотные компоненты, оставляя средние частоты относительно сохранными. Однако такие методы требуют больших вычислительных ресурсов и являются более сложными в реализации [2].

Отдельное направление в стеганографии представляют блочные методы, основанные на принципах распределения информации и контроля четности. Их фундаментальное отличие от последовательных алгоритмов заключается в отказе от поточечного внедрения данных в пользу структурного подхода, оперирующего группами пикселей как едиными объектами. Основной принцип метода заключается в разбиении изображения-контейнера на блоки фиксированного размера, например, 3×3 или 4×4 пикселя. Каждый такой блок рассматривается как независимая ячейка, способная нести один бит информации.

Внедрение осуществляется путем вычисления бита четности для выбранного блока, например, как суммы всех значений пикселей по модулю два. Если бит четности не совпадает с внедряемым битом, значение одного случайно выбранного пикселя инвертируется. Это гарантирует совпадение бита четности с передаваемым битом. Случайный выбор пикселя обеспечивает равномерное распределение искажений и затрудняет обнаружение внедрения. Такой подход повышает устойчивость к локальным искажениям (обрезка, шум, цветокоррекция). Однако блочные методы характеризуются существенно меньшей вместимостью по сравнению с классическими подходами. Емкость контейнера определяется не общим количеством пикселей, а количеством блоков, на которые разбито изображение. Например, для изображения размером 1000×1000 пикселей, разбитого на блоки 4×4, максимальный объем скрываемых данных составит примерно 62,5 КБ, что значительно меньше потенциальной емкости LSB-метода. Кроме того, возникает проблема надежного извлечения сообщения, требующая точного знания размера и схемы разбиения на блоки на стороне получателя [1].

На сегодняшний день цифровые водяные знаки, особенно на основе спектральных методов, активно используются в различных отраслях для защиты авторских прав и отслеживания контента. Практическое внедрение происходит преимущественно в системах видеохостинга, где каждый зритель получает уникальную версию контента с невидимыми идентификаторами, встроенными непосредственно в видеопоток. Эти идентификаторы привязываются к конкретному пользователю, сессии или устройству, что позволяет точно установить источник утечки в случае пиратского распространения [3]. Эффективность технологии подтверждается ее применением крупными стриминговыми платформами, образовательными сервисами и корпоративными системами для защиты конфиденциальных материалов. Развитие методов основано на повышении устойчивости идентификаторов, уменьшении вычислительной сложности и обеспечении совместимости с различными устройствами и форматами доставки контента [4].

Для защиты авторских прав и создания цифровых водяных знаков, требующих устойчивости к преобразованиям, в дальнейшем будут рассмотрены преимущественно методы частотной области. Этот приоритет обусловлен их принципиально иным уровнем устойчивости к деструктивным воздействиям по сравнению с пространственными методами. Если LSB-внедрение уязвимо даже к базовым операциям, таким как конвертация формата или изменение яркости, то модификация средне- и низкочастотных коэффициентов спектрального представления сигнала (например, в DCT или DWT-области) сохраняет целостность внедренной информации при сжатии, масштабировании и ряде геометрических искажений.

Практическая часть. Исходное изображение не содержит видимой вотермарки; ее добавляет система-детектор. Если злоумышленник скачает картинку с сайта, он получит файл без видимого знака, но со скрытым ядром. При попытке загрузить ее на другой сайт с детектором, ядро будет обнаружено, и на изображение будет наложена видимая вотермарка.

Для реализации изученных методов на практике необходима комбинация двух слоев: невидимый, устойчивый стеганографический слой (ядро). Его нельзя убрать, так как он неотличим от самого изображения; механизм визуализации, который активируется специальным детектором и проявляет скрытую информацию, создавая видимую вотермарку.

В ходе исследования были последовательно проведены два эксперимента по сокрытию информации в изображениях формата WebP, направленные на сравнительный анализ методов стеганографии.

В ходе первого эксперимента, основанного на методе LSB-замещения, в качестве инструментальной базы использовалась среда разработки Visual Studio и библиотека libwebp — официальная библиотека для работы с изображениями формата WebP от Google, обеспечивающая точное декодирование и кодирование изображений без потерь. На подготовительном этапе исходное изображение в формате JPEG было преобразовано в WebP с сохранением максимального качества. Далее на языке C++ были разработаны две программы: encode.cpp для внедрения текстового сообщения в младшие биты пикселей и decode.cpp для его последующего извлечения. Сборка программ выполнялась с помощью системы CMake, которая автоматизировала процесс компиляции и управления зависимостями через сценарий CMakeLists.txt. На этапе внедрения сообщения программа encode.exe модифицировала младшие биты цветовых каналов изображения, создавая файл stego_image.webp. Визуальная проверка подтвердила полную идентичность исходного и модифицированного изображений, что объясняется незначительностью вносимых изменений: поскольку каждый цветовой канал кодируется 8 битами, замена младшего бита приводит к изменению значения цвета не более чем на 0,4 %, поскольку цветовые каналы в RGB принимают значения от 0 до 255, следовательно, максимальное относительное изменение составляет 1/255, то есть $(1/255)\times 100\,\% \approx$ $\approx 0{,}392\,\%$, что находится ниже порога восприятия человеческого глаза [5]. Статистический анализ изображений выявил существенные изменения в структуре изображения, представленные в табл. 1, в частности, существенное увеличение объема файла и умеренный прирост количества градаций серого, указывающий на изменение цветового распределения.

Декодирование зашифрованного текста с помощью decode.exe прошло успешно. Однако дальнейшие испытания показали, что даже небольшие изменения насыщенности или резкости приводят к потере скрытой информации и выводу случайных символов. Кроме того, любое сохранение с потерями уничтожает LSB-биты. Если такое изображение загрузить в соцсеть, оно будет пережато и сообщение исчезнет, поэтому для реальной защиты информации LSB-метод не подходит.

Второй эксперимент был направлен на исследование более устойчивого подхода — спектрального метода на основе дискретного косинусного преобразования (DCT). В рамках реализации спектрального метода применялось блочное дискретное косинусное преобразование с последующей модификацией избранных коэффициентов в частотной области. Особое внимание уделялось

селекции оптимального диапазона частот, обеспечивающего компромисс между визуальной незаметностью и устойчивостью к типичным преобразованиям изображения. Модификации подвергались исключительно среднечастотные компоненты, что обеспечивало сохранение визуального качества контейнера при одновременном повышении устойчивости к операциям сжатия с потерями. Обратное преобразование и последующее кодирование выполнялись с применением оптимизированных процедур, минимизирующих накопление артефактов.

Таблица 1

Сравнительный анализ статистических характеристик изображения до и после внедрения сообщения с использованием метода LSB-замещения

	До	После	Изменение
Размер файла	208,650 байт	245,258 байт	+17,5 %
Уникальных цветов	30,450	31,974	+5 %
Градаций серого	902 (0,07 %)	972 (0,08 %)	+7,2 %
Изменения в палитре цветов			
#caafa3	2,098	1,672	–20,3 %
#c9ada2	2,036	1,657	–18,6 %
#c3a89d	1,790	1,711	–4,4 %
#c8aca1	1,722	1,589	–7,7 %
#c2a69b	1,708	1,643	–3,8 %
#cdb1a6	1,699	1,452	–14,5 %
#c5a99e	1,660	1,544	–7,0 %
#c6aa9f	1,645	1,541	–6,3 %
#ccb0a5	1,528	1,598	+4,6 %
#c7aba0	1,500	1,537	+2,5 %

Данный подход продемонстрировал существенное превосходство над пространственными методами в тестах на устойчивость [5]. В качестве программной основы использовалась библиотека OpenCV, предоставляющая оптимизированные реализации алгоритмов компьютерного зрения. Аналогично первому эксперименту, были разработаны программы dct_encode.exe и dct_decode.exe для внедрения и извлечения сообщения соответственно. Процесс внедрения заключался в преобразовании изображения в частотную область с последующей модификацией коэффициентов средних частот, наименее заметных для зрительного восприятия. Визуальный анализ результатов также показал полное отсутствие различий между исходным и модифицированным изображениями, что свидетельствует о сохранении высокого качества контейнера. Статистический анализ изображений, результаты которого представлены в табл. 2, зафиксировал незначительное увеличение размера файла — всего на 1,8 %, что свидетельствует о высокой степени совместимости внесенных модификаций с алгоритмом сжатия. Данный факт

позволяет предположить, что примененный метод стеганографического внедрения не приводит к возникновению статистических аномалий, интерпретируемых как цифровой шум. Кроме того, цветовая статистика изображения сохранила высокую стабильность: прирост количества уникальных цветов и градаций серого составил менее одного процента, что подтверждает минимальное визуальное и структурное воздействие на исходное изображение.

Таблица 2

Сравнительный анализ статистических характеристик изображения до и после внедрения сообщения с использованием DCT-метода

	До	После	Изменение
Размер файла	208,650 байт	212,400 байт	+1,8 %
Уникальных цветов	30,450	30,480	+0,1 %
Градаций серого	902 (0,07 %)	918 (0,07 %)	+1,7 %
Изменения в палитре цветов			
#caafa3	2,098	2,095	–0,1 %
#c9ada2	2,036	2,032	–0,2 %
#c3a89d	1,790	1,785	–0,3 %
#c8aca1	1,722	1,718	–0,2 %
#c2a69b	1,708	1,705	–0,2 %
#cdb1a6	1,699	1,695	–0,2 %
#c5a99e	1,660	1,655	–0,3 %
#c6aa9f	1,645	1,640	–0,3 %
#ccb0a5	1,528	1,525	–0,2 %
#c7aba0	1,500	1,495	–0,3 %

Человеческое зрение более чувствительно к резким перепадам, низким частотам и изменениям яркости, чем к плавным изменениям, высоким частотам и изменениям цветности. Метод DCT использует эти особенности, встраивая сообщение в менее заметные компоненты, распределяя изменения по всему боку и при этом сохраняя важную визуальную информацию. Для Y-канала (компоненты яркости) мы сознательно не вносим изменения, что дает 0 % искажений в самой важной компоненте изображения. Для U-канала (компоненты цветности) применяется избирательное модифицирование среднечастотных коэффициентов с амплитудой возмущения приблизительно 0,007 единиц. Учитывая динамический диапазон значений цветоносных компонент от –127 до +128, относительное искажение составляет: (0,007 / 128) × × 100 % ≈ 0,0055 %. Столь незначительная модификация обеспечивает уровень изменений, не обнаруживаемый при визуальном анализе. Для V-канала (второй цветоносной компоненты) изменения не применяются, что позволяет дополнительно снизить совокупный уровень искажений. Данное решение основано на принципе избирательного воздействия, при котором минимизация

количества модифицируемых каналов способствует сохранению визуальных характеристик исходного изображения.

Вопрос о предельном объеме внедряемого сообщения зависит от количества блоков, на которые сегментируется изображение. Существует компромисс между объемом данных и визуальной заметностью изменений: увеличение размера водяного знака повышает уровень искажений. На рис. 1 представлена полученная зависимость пикового отношения сигнала к шуму (PSNR) от объема вводимой информации, полученная экспериментальным путем (табл. 3):

– высокое визуальное качество (PSNR > 70 дБ) сохраняется при объеме сообщения примерно до 80 символов;

– хорошее визуальное качество (PSNR > 60 дБ) достигается при объеме до 1000 символов.

Для скрытной стеганографии рекомендуется использовать не более 20–30 % емкости, что составляет 400–600 символов, достаточно для ЦВЗ.

При использовании технических средств возможно выявление сокрытия данных, даже если изменения визуально незаметны. Стеганография в DCT-области, включая YUV подход, устойчива к обнаружению, но уязвима для методов, сравнивающих гистограммы DCT-коэффициентов исходного и пересжатого изображения. Современные методы обнаружения используют машинное обучение, достигая 70–80 % эффективности для простых DCT-реализаций [5]. Внедрение продвинутых DCT-методов с адаптивным квантованием и случайным выбором позиций снижают обнаруживаемость до 40–60 %. Ключевой принцип — обеспечение «естественности» изменений, чтобы статистические и визуальные характеристики оставались неотличимыми от исходного носителя. пояснениями и расшифровкой использованных обозначений.

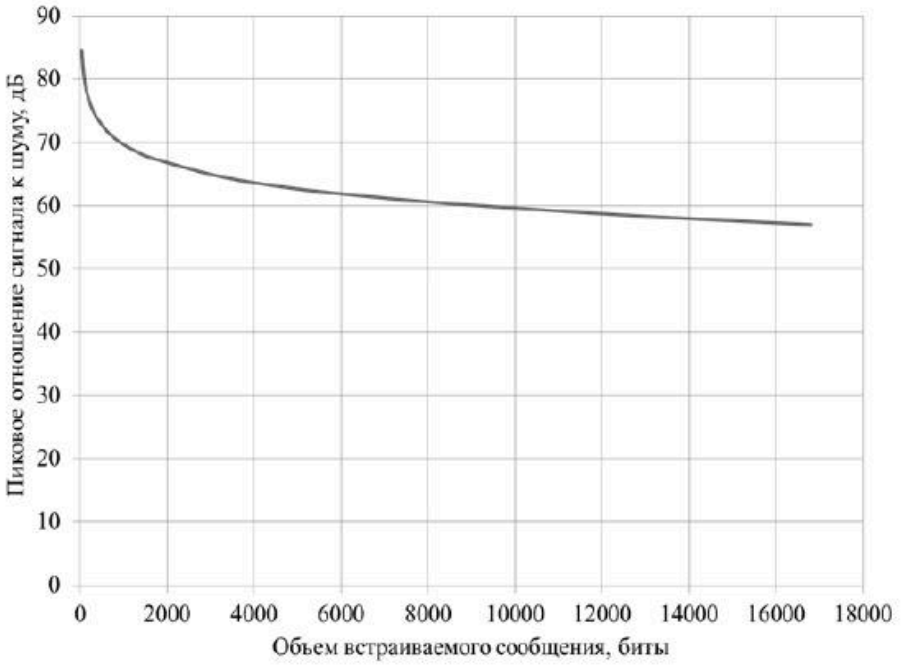

Рис. 1. График зависимости пикового отношения сигнала к шуму от объема встраиваемого сообщения

Таблица 3

Результаты эксперимента: значения PSNR для различных объемов встраиваемого сообщения

Размер сообщения в символах	Размер сообщения в битах	PSNR, дБ
4	32	84,58
8	64	81,57
12	96	79,80
16	128	78,56
20	160	77,60
40	320	74,59
80	640	71,58
120	960	69,81
160	1280	68,57
200	1600	67,61
400	3200	64,60
800	6400	61,59
1000	8000	60,63
1200	9600	59,81
1600	12800	58,43
2100	16800	57,00

Заключение. На текущем этапе цель исследования достигнута частично. Практическая реализация и сравнение двух подходов — простого LSB-замещения и спектрального метода на основе DCT — дали четкие и ожидаемые результаты. Если LSB-метод подтвердил свою принципиальную уязвимость даже к базовой обработке, то применение DCT-преобразования позволило создать скрытое «ядро», которое успешно выдерживает сжатие с потерями, незначительные геометрические искажения и цветокоррекцию, оставаясь неотличимым от естественного шума изображения. Полученные результаты открывают путь для практического внедрения двухуровневой системы защиты в реальных цифровых платформах. Данный метод может быть интегрирован в системы видеохостинга, фотобанки и социальные сети для автоматического отслеживания распространения контента. Динамическое наложение видимого водяного знака только при отображении изображения на платформе решает проблему порчи эстетики оригинала для правообладателя, одновременно создавая для нарушителя непреодолимый барьер в виде неудаляемого скрытого идентификатора. Дальнейшие исследования могут быть направлены на адаптацию метода для видеоформатов и повышение его устойчивости к целенаправленным атакам с использованием машинного обучения.

Литература

[1] Скрывать не скрывая. Еще раз о LSB-стеганографии, хи-квадрате и… сингулярности? https://habr.com/ru/articles/422593/ (дата обращения 11.10.2025).
[2] Цифровые водяные знаки. Стеганография. URL: https://habr.com/ru/sandbox/16609/ (дата обращения 11.10.2025).
[3] Защищает ли Netflix свой контент? URL: https://habr.com/ru/companies/ruvds/articles/565432/ (дата обращения 11.10.2025).
[4] Forensic Watermarking: 4 Types, Implementation & Challenges Explained. URL: https://www.vdocipher.com/blog/forensic-watermarking/ (accessed 11.10.2025).
[5] Грибунин В.Г. Цифровая стеганография. Москва, СОЛОН-Пресс, 2023, 262 с.

Automating the embedding of non-removable digital watermarks into images for copyright protection using steganography

Gross Yana Alexandrovna yana.gross.05@mail.ru
Volosatova Tamara Mikhailovna tamarav@bmstu.ru
Knyazeva Svetlana Vadimovna knyazevasv@bmstu.ru

BMSTU, Moscow, Russia

The study is devoted to the development of a robust digital watermarking method for protecting image copyrights. The vulnerability of classical methods to statistical steganalysis and processing is analyzed. A two-level system is proposed: a hidden core, embedded into the image using a steganographic method, and a visualization mechanism that dynamically overlays a visible watermark when the core is detected by a detector on the platform.

Keywords: *digital watermarking, steganography, copyright protection, LSB method, DCT transform, steganalysis*