О разработке отечественного специального программного обеспечения для выявления вмешательства в фото- и видеоизображения

Владимир ИВАНОВ, д.т.н., профессор, эксперт
Станислав ЗВЕЖИНСКИЙ, д.т.н., профессор, АО «НПК «Дедал»

Общие положения

Основные особенности процесса и методов обработки с различными целями (в том числе с криминальной) фото- и видеоконтента отражены в предыдущей работе авторов (ТЗ № 1-2021). В целом при исследовании фото- и видеоконтента на предмет выявления фактов подделки или внедрения объектов могут использоваться те же профессиональные средства монтажа фотографий и видеоматериалов, к которым относятся, прежде всего, Adobe Premiere и Adobe Photoshop, а также Pinnacle Studio, WinVCR, iFilmEdit, VirtualDub и др. Из отечественного специального программного обеспечения (СПО) в области анализа изображений следует указать «Vocord Видеоэксперт» (Vocord, Одинцово Московской обл.), «Эскиз-В» (ДиВиЛайн, Томск), «StreamEye Studio» (Elecard, Томск), «Fake Video Detection Service» (VIEN, Москва), «FindFace Security, FindFace SDK» (NtechLab, Москва).

Характеристики распространенных СПО, которые можно отнести к экспертно-криминалистическому классу, показаны в табл.1. Большинство из них использует искусственные нейронные сети (ИНС), реализованные на собственных алгоритмах или из состава общеизвестных библиотек.

Таблица 1. Характеристики СПО для исследования изображений

Продукт / разработчик	Краткое описание
1. Vocord Видеоэксперт/ Vocord (РФ) [1]	Применяется в области защиты информации, криминалистики, цифровой фотографии и телевидении для анализа потокового видео в интернете, DVD-видео, компрессии видеоданных.
2. Эскиз-В/ ДиВиЛайн (РФ) [2]	СПО «Экспертная система криминалистических исследований видеозаписей (ЭСКИЗ-В)» предназначена для повышения результативности видеотехнической экспертизы. Состоит из совокупности программных модулей, позволяющих выполнять исследование видеофайлов, включая анализ метаданных и параметров изображений.
3. StreamEye Studio/ Elecard (РФ) [3]	Elecard StreamEye Studio включает 5 отдельных приложений и 2 инструментальных консольных приложения для анализа видеопотоков. 1. StreamEye – анализ закодированного видео (структура, макроблоки). 2. Stream Analyzer – анализ синтаксиса медиапотока и контейнеров. 3. Video Quality Estimator – анализ качества видео с помощью объективных метрик (PSNR, APSNR, SSIM, DELTA, MSE, MSAD, VQM, NQI, VMAF и др.). 4. YUV Viewer – просмотр последовательности YUV данных в видеофайлах. 5. Quality Gates – сравнение нескольких видеофайлов, закодированных с различными параметрами. Misha Codec Benchmark – инструмент для проверки качества нескольких энкодеров (сравнение параметров закодированных потоков на основе объективных метрик) и качества нескольких потоков, закодированных одним энкодером с разными настройками. Command Line Tools – инструмент для автоматизации анализа большого числа потоков и решения сложных задач с помощью консольных команд.
4. Fake Video Detection Service / VIEN (РФ) [4]	СПО основано на ИНС для покадровой оценки и классификации объектов. Оцениваются характеристики лиц на видео и определяются признаки синтетической генерации образов. В случае обнаружения таковых, лицо на экране выделяется красной рамкой как «фальшивое».
5. FindFace Security, SDK / NtechLab (РФ) [5]	Библиотека на языке СИ, предоставляющая доступ к технологии распознавания лиц на основе ИНС. Решаются 3 задачи по распознаванию лиц: детектирование, извлечение биометрического шаблона и верификация.
6. EnCase Forensic /Guidance Software Inc (США) [6]	СПО для компьютерно-технической экспертизы, проведения расследований и поиска улик правительственными органами, криминалистическими экспертными учреждениями и международными корпорациями.
Microsoft Video Authenticator/Redmond (США) [7]	СПО для «глубокого» обнаружения deepfake, способен анализировать фото или видео на наличие особенностей, которые не видны экспертам – наличие границ между оригинальным изображением и добавленными элементами или небольшие затухания; обнаружение сигнализирует о присутствии манипуляций.
7. Google Alphabet/ Google (США) [8]	БД для исследователей с дипфейк-роликами, в нее входят 300 оригинальных роликов и 3 тысячи фейковых (на базе оригинальных), созданных алгоритмами подмены лиц Deepfakes, Face2Face, FaceSwap и NeuralTextures.
8. Assembler JigSaw / (США) [9]	СПО способно показать, где именно использовался «фотошоп» на фотографии. В него встроено 7 «детекторов» для обнаружения различных типов технологий при редактировании фото. Обнаруживается наложение или соединение разных фотографий, изменение фона, цвета и др.
9. / Institute of Information Technology (США) [10]	ИНС-инструмент фокусируется на движениях лица и головы человека, чтобы определить, был ли подделан материал. Вероятность идентификации подделки видео, созданного ИНС, приближается к 0,96. Для обучения ИНС используется набор данных объемом около 1 000 поддельных видео.
10. AccessData (США) [11]	СПО для компьютерной криминалистики. Сканируется жесткий диск в поисках различной информации, в т.ч. медиафайлы, даже из заблокированных устройств.
11. Forevid / Forevid Forensics [12]	СПО для извлечения данных из изображений: покадровый анализ видео; создание размытия или подсвечивания части видео; экспорт кадров из закладок как картинок (PNG, BMP, TIFF, JPG); экспорт кадров в буфер обмена; добавление закладок; воспроизведение от конца с помощью обратной фильтрации; фильтрация видео и пр.

Математические СПО для исследования фото- и видеоконтента

Для исследования изображений могут использоваться общеизвестные пакеты математического моделирования общего назначения: MathCAD, Matlab, LabVIEV, Wolfram Mathematica, Statistica и др., которые имеют соответствующие инструменты для работы с изображениями, в том числе в реальном времени. Например, в Matlab имеется инструмент Image Processing Toolbox, реализующий более 200 опций для работы с видео и цветовыми матрицами любых стандартов. MATLAB NeuroSolutions содержит 15 типовых моделей ИНС и 5 алгоритмов их обучения. Такое СПО, как Octave, относится к свободно распространяемым и кроссплатформенным, полностью совместимо с программным кодом Matlab. Язык программирования Python имеет в своем составе несколько бесплатных библиотек для работы с матрицами, нейросетями, видео и аудио (библиотеки Keras, Tensorflow, OpenCV). Некоторые пакеты математического моделирования имеют опцию компиляции моделей в исполняемые файлы и создания тематически ориентированных библиотек и программ.

Большинство современных математических СПО общего назначения могут реализовать практически любую модель обработки фото- и видеоматериалов, в т.ч. в реальном или квази-реальном времени. При разработке отечественного СПО для исследования изображений на предмет внутрикадрового монтажа предпочтение следует отдавать инструментам математического моделирования общего назначения, свободно распространяемым, не зависимым от ежегодной оплаты лицензии и генерации ключей, а также от аппаратной привязки ключей и регистрации на сайтах производителей, по следующим причинам:

- наиболее мощные и популярные пакеты математического моделирования (Matlab, MatCAD, Wolfram Mathematica, Statistica, LabVIEV др., в т.ч. облачных сервисов для вычислений) находятся в юрисдикции США и могут быть заблокированы для пользователей РФ (в то время, как отечественного аналога нет);
- разработчики экспертно-криминалистического СПО не раскрывают особенностей извлечения признаков, построения, настроек и параметров математических моделей, используемых в исследовании изображений на предмет внедрения объектов, т.е. СПО работает по принципу «черного ящика»;
- разрабатываемые модели должны быть кроссплатформенными, желательно Linux-ориентированными, поскольку в РФ для использования в важных решениях сертифицировано несколько именно таких ОС – AstraLinux, Синтез-М, Циркон-36 и др.

Вышеназванные пакеты математического моделирования, а также экспертно-криминалистическое СПО реализует методы исследований изображений, которые сведены на рис.1. На практике наиболее востребованными являются методы локализации (обнаружения или распознавания) объектов на цифровых изображениях, классификация которых представлена на рис.2.

Рисунок 1. Возможные методы и результаты обработки изображений

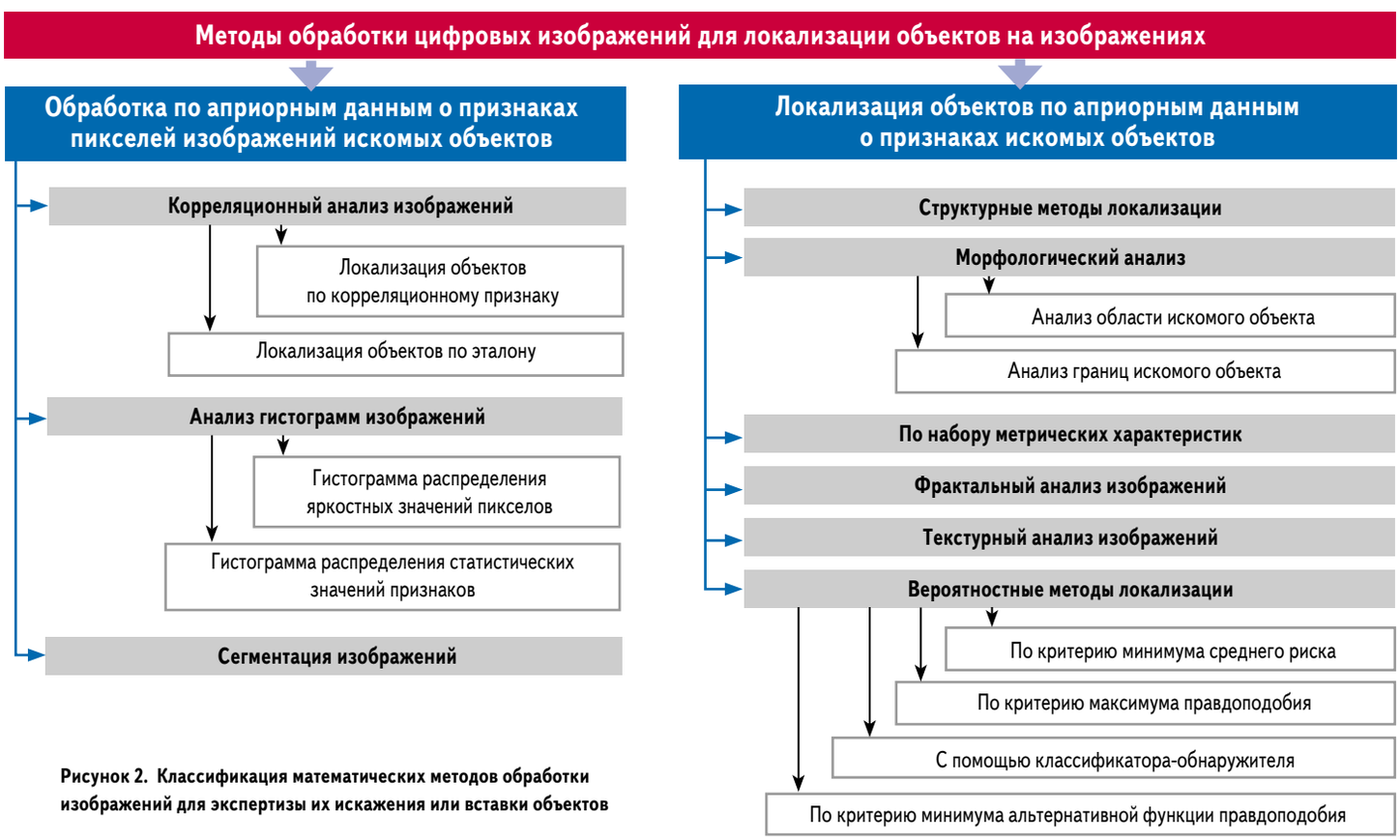

Рисунок 2. Классификация математических методов обработки изображений для экспертизы их искажения или вставки объектов

Возможные прикладные сферы обработки изображений различными математическими методами приведены в табл.2.

Таблица 2. Классификация методов обработки изображений

Метод обработки	Получаемый результат
1. Точечные методы обработки изображений. Гистограммы интенсивности. Преобразования на основе анализа гистограмм интенсивности.	Точечные преобразования - просветление, бинаризация, псевдораскрашивание и др.
2. Пространственные методы обработки изображений. Пространственная частота изображения. Свертка изображения. Усиление края, методы Лапласа, Робертса, Кирша и Собеля, методы сдвига и разности, направленного градиента.	Построение фильтров: НЧ, полосные и ВЧ фильтры. Компенсация смаза, расфокусировки.
3. Геометрические и алгебраические методы обработки изображений. Алгебраические преобразования (сложение, вычитание изображений). Геометрические преобразования (монохромная интерполяция, аффинные и нелинейные преобразования).	Поиск подвижных объектов. Восстановление частично утраченных изображений (в пределах до 15%).
4. Методы межкадровой обработки изображений. Геометрия нескольких проекций.	Стереозрение. Детекция движения объекта.
5. Анализ изображений на основе разложения по базисным функциям. Базисные векторы и матрицы. Разложение Карунена-Лоева. Дискретное преобразование Фурье. Косинусное преобразование. Непрерывное и дискретное вейвлет- преобразование и разложение.	Вейвлетная селекция. Сжатие изображений. Восстановление изображений.
6. Статистические методы анализа текстур. Региональные признаки. Методы измерения текстур на основе статистик 1-го и 2-го порядков.	Поиск неоднородностей в изображениях.
7. Методы анализа формы изображений. Концепции формы. Сегментация, выделение формы и ее представление. Характеристики формы и их измерение. Скелетизация. Преобразование Хафа. Бинарная математическая морфология. Эрозия и дилатация. Морфологические алгоритмы на дискретных бинарных изображениях.	Поиск заданных объектов на изображении.
8. Метрики для измерения сходства изображений. Сравнение спектральных разложений. Классификация методом сравнения с эталоном. Сходство, основанное на поиске оптимального пути. Принцип оптимальности Беллмана и динамическое программирование. «Беспризнаковое» распознавание.	Распознавание объектов на изображениях.
9. Распознавание текстов по изображениям документов. Сегментация документов и текстов. Выравнивание текстов. Распознавание печатных символов. Распознавание рукописных текстов.	Обработка документов.
10. Биометрическая идентификация на основе распознавания изображений. Классификация радужных оболочек глаза методом Даугмана. Классификация силуэтов ладоней методом сравнения гибких объектов. Метод выделения особых точек в папиллярном узоре.	Биометрия.
11. Распознавание динамических сцен. Распознавание жестов. Распознавание мимики. Распознавание поз.	Ситуационный анализ. обстановки

Принципы анализа фото / видео и извлечения признаков подделки

Как показано выше, наиболее точными являются методы анализа структуры самого изображения, к которым относятся:

- RAW-данные с матрицы на выходе АЦП камеры;
- настройки сглаживающих фильтров камеры для оптимизации восприятия изображения человеком (уникальны для разных производителей камер);
- характеристики матрицы («темновые» токи, «горячие», «битые» и «залипшие» пикселы, распределение которых уникально для каждой матрицы) и АЦП (разрядность, время преобразования и др.);
- параметры межпиксельного сглаживания и переходов на границах объектов, баланса белого, средних значений отдельных цветов;
- метаданные изображений (EXIF), хранимые в первичных RAW-файлах и файлах изображений в других форматах после сжатия – JPG, TIF, BMP и др.

Относительно последнего следует заметить, что при конвертации или сжатии изображения метаданные могут частично утрачиваться или модифицироваться. Известны СПО для модификации EXIF-данных в RAW-файлах, и этот признак является ненадежным.

Таким образом, наиболее информативны и точны RAW-данные, но в таком формате на практике фотографии хранятся редко, а видеозаписи вообще не встречаются, поскольку занимают объем, многократно превышающий другие форматы, например, 10-20 раз больше объема сжатых файлов JPG.

RAW-данные подвергаются минимум двойному преобразованию (компрессии), в ходе которого их существенная часть теряется или преобразовывается: при переводе из RAW в JPG или другой формат; при сжатии кодеком смонтированного видеофильма и пр. При этом значительная часть данных, доступных в RAW-интерпретаторе, безвозвратно утрачивается. При повторном захвате видео и перемонтаже с внедрением (или исключением) объектов в кадр, при выводе фильма видеоматериалы вновь подвергается компрессии, зачастую уже с использованием другого кодека. Судя по публикациям, представленным в табл.1 предыдущей статьи авторов, иногда факт многократного перекодирования или сглаживания считается признаком вмешательства в исходный контент. Некоторые производители высококачественной продукции для проверки аутентичности контента включают в изображение специальные метки (цифровые водяные знаки), однако такое решение пока встречается редко (несколько % по рынку).

Тем не менее, можно выдвинуть рабочую гипотезу, что часть RAW-данных все же можно восстановить (или косвенно оценить) по имеющемуся сжатому изображению с частично утраченными EXIF-данными. Также возможна ситуация с идентификацией фото и видео на предмет вмешательства, когда ни один прямой признак не дает однозначного ответа о факте вмешательства в изображение, но по совокупности нескольких слабых косвенных признаков (которые в отдельности не являются доказательствами) можно будет формировать один обобщенный показатель качества изображения. При этом следует использовать математический аппарат big data (большие данные) и data mining (машинное обучение).

В общей классификации методов и программ для обработки фото- и видеоматериалов можно выделить методы, реализуемые на уровне RAW-данных,

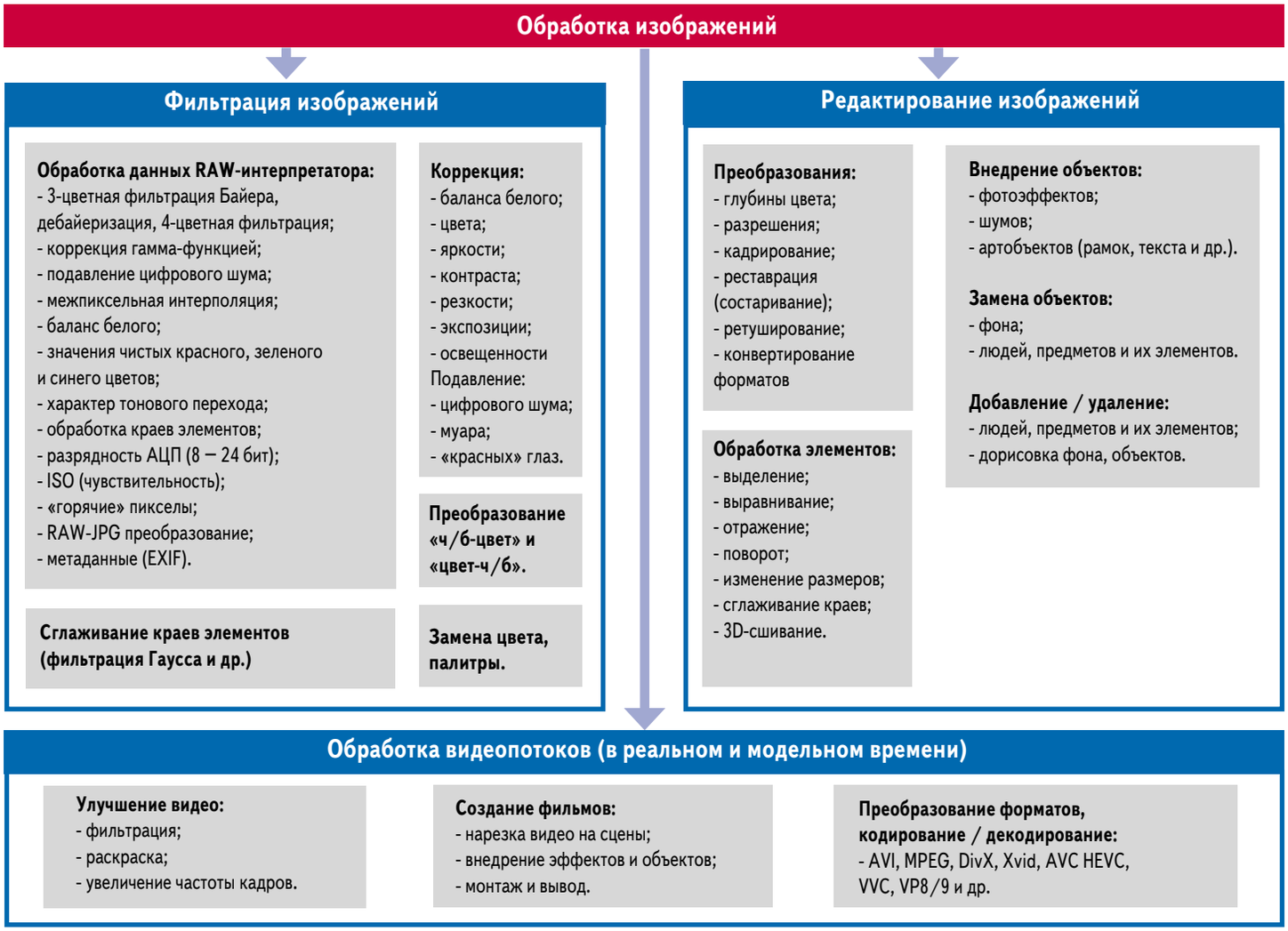

Рисунок 3. Методы извлечения признаков при обработке изображений

а также программные средства, показанные на рис. 3. Файлы сопровождаются метаданными (EXIF), где хранится информация о типе и настройках камеры – ISO, разрядность АЦП, размер снимка, число цветовых компонент, тип кодирования, время создания / редактирования и геометки, диафрагма, выдержка, фокус, время оцифровки, тип баланса белого, яркость, поле зрения и др. Эти данные в совокупности являются уникальными для каждого снимка, поскольку условия съемки непрерывно меняются.

Производители фото- и видеооборудования применяют матрицы от разных производителей и из разных партий, которые, несмотря на постоянное повышение качества, различаются «темновыми» токами, что проявляется в виде уникального расположения «горячих» пикселов на матрице. Каждый производитель использует уникальные фильтры корректировки RAW-данных (с выхода АЦП) для наилучшего восприятия снимка пользователем. К таким фильтрам относится межпиксельное сглаживание (интерполяция), гамма-коррекция, подавление цифрового шума матрицы и др. Таким образом, в совокупности данные о настройке фильтров, характеристик матриц, АЦП, камеры, а также условия сцены делают уникальным каждый снимок.

С точки зрения анализа уникальности снимка и его целостности желательно работать с RAW-данными интерпретатора, но такие файлы имеют большой объем и доступны к использованию только в дорогостоящей технике. Большинство производителей оборудования бюджетного уровня используют сжатие кадра непосредственно в камере (традиционно используется формат JPG). Как показано на рис. 3, фото- и видеоматериалы в процессе окончательной обработки изображений редактируются. При этом вновь могут применяться различные программные фильтры, осуществляться корректировка (цвета, освещенности, баланса белого и т.д.), сглаживание и интерполяция пикселов.

Внедрение объектов из других фото- и видеоматериалов (различные артобъекты, добавление или удаление объектов, в т.ч. человекоподобных) осуществляется программно. Внедряемые объекты, как правило, берутся из других фотографий и видеозаписей, полученных в других условиях съемки, с другими настройками и с помощью камер разных производителей. Таким образом векторы параметров RAW-данных и метаданных внедряемых объектов значительно различается. При внедрении объектов непременно возникают краевые эффекты, которые пытаются сгладить различными фильтрами (например, размытие, сглаживание, гауссова фильтрация). При этом само внедряемое изображение будет характеризоваться аддитивной смесью распределений параметров RAW-данных и настроек фильтров окончательного сглаживания, цветокоррекции и других параметров.

Внедренный в видеофильм объект должен анимироваться в соответствии с сюжетом фильма. Обычно он описывается 3D-моделью, создаваемой на основе его разбиения на элементарные полигоны, – используется метод полигонометрии. Полигоны имеют свой цвет, и для придания реалистичности объекту границы

элементарных полигонов сглаживаются каким-либо фильтром. Таким образом, возможны отличия межпиксельного сглаживания внедренного изображения от характера сглаживания остальной части изображения (сцены).

По окончании монтажа фильма осуществляется его вывод с использованием какого-либо кодека. При этом вновь используется сжатие с частичной потерей качества изображения. Анализ работы видеокодеков показывает, что при сжатии осуществляется разбиение фильма на относительно однородные сцены со статическими объектами, где первый и последний кадры сцены остаются неизменными. Промежуточные кадры, где передвигаются только элементы, также могут различаться для разных объектов. Таким образом, существует возможность из сжатого видеофильма получить оригинальные кадры, которые использовались на этапе монтажа, а также получить информацию о характеристиках промежуточных кадров и элементов.

Таким образом, обнаружение внедренных объектов в фото- и видеоконтент возможно путем:

- извлечения из конечного продукта (фото или видео) остатков RAW-данных для различных областей изображения и оценки их однородности;
- обнаружения краевого эффекта внедренного объекта;
- определения неравномерности характеристик межпиксельного сглаживания на изображении и границах, выделения объектов с аномалиями.

На основании вышеизложенного можно сделать вывод о том, что для выявления перечисленных признаков целесообразно использовать математические методы, основные из которых показаны в табл. 3.

Таблица 3. Основные математические методы и подходы для выявления возможного внедрения объектов в фото- и видеоматериалы

Математический аппарат	Область применения	Выявляемый признак
Стеганография	Выявление в значениях пикселов фотографий зашифрованных текстовых сообщений.	Выявление на изображении скрытых упорядоченных объектов.
Фрактальный анализ	Выявление в хаосе частично организованных структур (с меньшей степенью хаоса).	Область (массив пикселов) на изображении с большей (меньшей) степенью упорядоченности по сравнению с окружающими пикселами.
Двумерный спектральный анализ	Гармонический анализ сигналов и потоков данных.	1. Выявление гармоник высшего порядка на краях возможного внедренного объекта. 2. Выявление искусственных 3D-объектов, созданных с использованием триангуляции. 3. Выявление распределения на матрице «горячих», «битых» и «залипших» пикселов.
Энтропийный анализ	Исследование распределения взаимной энтропии яркости свечения пикселов изображения в некоторой области изображения.	Выявление на изображении областей с разной межпиксельной энтропией: - уменьшение энтропии означает наличие упорядоченности в изображении; - увеличение энтропии означает искусственное «зашумление» изображения.
Корреляционный анализ	Исследование авто- и взаимной корреляции областей изображения с использованием скользящей двумерной «оконной» функции.	Выявление областей с высокой и низкой корреляцией изображения.
Wavelet-анализ	Исследование двумерных вейвлетов изображения.	Обнаружение: 1) краевого эффекта объекта (вейвлеты высших порядков); 2) объектов с «аномальной» структурой; 3) шумов матрицы.
Регрессионный анализ	Исследование регрессии (влияния) одной наблюдаемой случайной величины на другую.	Выявление областей изображения с разными характеристиками сглаживающей (фильтрующей) межпиксельной функции, вычисление функции «предсказания».
Анализ скользящего среднего	Исследование двумерной модели скользящего среднего с «оконной» функцией.	Выявление областей изображения с отклонениями параметров от модели скользящего среднего.
Кластерный анализ	Классификация объектов.	Распознавание характера найденного (локализованного) объекта на изображении.

Выводы

Несомненно, что ввиду высокой потребности (легальной, криминальной) методы искусственного интеллекта для подделки фото- и видеоконтента будут развиваться и далее, что обуславливает необходимость развития «противодействующих» им методов (также основанных на ИИ), математических подходов к выявлению и доказыванию фактов вмешательства в контент. Это определяет безусловную актуальность разработки соответствующих защитных СПО, по крайней мере, на ближайшие двадцать лет.

Литература

1. http://en.vocord.ru/technical-support/demo-version.
2. https://diviline.ru.
3. https://www.elecard.com/ru/products/video-analysis/streameye-studio.
4. https://pt.2035.university/project/project-10.
5. https://findface.pro/findface-sdk.
6. http://www.guidancesoftware.ru/o-kompanii.html.
7. https://hi-tech.news/internet/2519-microsoft-video-authenticator-novyj-instrument-obnaruzhenija-deepfake.html.
8. https://ru.wikipedia.org/wiki/Alphabet.
9. https://ru.wikipedia.org/wiki/Jigsaw_(компания).
10. https://hightech.fm/2019/06/22/deepfake.
11. https://accessdata.com.
12. https://forevid.com.