

9.8. ПРЕОБРАЗОВАНИЕ ОРИГИНАЛА ДОКУМЕНТА В ЦИФРОВУЮ ФОРМУ И МЕТОД РАСПОЗНАНИЯ ПОДДЕЛКИ В НЕЙ С ПОМОЩЬЮ ЦИФРОВЫХ ДВОЙНИКОВ ДОКУМЕНТА

Аллаберганов Ахмеджан Атаханович, аспирант. Место учебы: Томский государственный университет систем управления и радиоэлектроники. Подразделение: кафедра АСУ. Фотоника, приборостроение, оптические и биотехнические системы и технологии. Место учебы: Байкальский государственный университет. Подразделение: кафедра криминалистики, судебных экспертиз и юридической психологии. E-mail: nsk-kapital@mail.ru

Аннотация. Как известно, с бумажного носителя (с бумажной версии) невозможно определить признаки монтажа самого печатного текста - воспроизведенного с помощью принтера «знакосинтезирующего устройства». С оригинала бумажного носителя распознается и выявляется любая подделка (фальсификация) в документе – кроме, произведенного монтажа самого печатного текста. Разработан метод распознавания монтажа печатного текста находящегося на бумажном носителе, с помощью цифровых двойников, методом преобразования оригинала документа в цифровую форму, и выделения определённых частей фрагментов текста вплоть до отдельных пикселей, с использованием зон спектра, электромагнитных излучений, с применением Криминалистического многофункционального комплекса (КМК).

Ключевые слова: криминалистическая экспертиза, документ, подделка, монтаж печатного текста, цифровая форма, цифровые двойники носителя.

TRANSFORMATION OF ORIGINAL DOCUMENT TO DIGITAL FORM AND METHOD OF RECOGNITION OF FORGES IN IT WITH DIGITAL DUAL DOCUMENTS

Allaberganov Akhmedzhan Atakhanovich, postgraduate student. Place of study: Tomsk State University of Control Systems and Radioelectronics. Department: Chair of ACS. Photonics, instrumentation, optical and biotechnological systems and technologies. Place of study: Baikal State University. Department: Criminals, forensics and legal psychology chair. E-mail: nsk-kapital@mail.ru

Annotation: As you know, from paper media (from the paper version) it is impossible to determine the signs of the installation of the printed text itself - the "sign-synthesizing device" reproduced using the printer. From the original paper media, any imitation (falsification) in the document is recognized and detected, except for the installation of the printed text itself. A method has been developed for recognizing the installation of printed text on paper using digital doubles, the method of converting the original document into digital form, and highlighting certain parts of text fragments up to individual pixels, using spectral zones, electromagnetic radiation, using the Forensic Multifunctional Complex (FMC)

Keywords: forensic examination, document, fake, installation of printed text, digital form, digital media doubles.

При исследовании реквизитов письма (подпись, расшифровка подписи, оттиск печати) и в том числе печатного текста (текст, линии, бланковая строка) на

бумажном носителе бывают и есть моменты, когда эксперты не могут распознать «монтаж печатного текста», иногда и реквизитов письма, даже и с использованием оптических приборов.

Так как, с использованием современных технологий можно формировать любой текст письма под определенные реквизиты или наоборот и далее, процедура распечатывания при помощи принтера (знакосинтезирующего устройства). После распечатывания (перенос на бумажный носитель), реквизитов письма или текста – происходит сокрытие той информации, то есть, «способ» формирования самой информации - теряется.

Например, сформированный текст и потерянный тот самый «способ» формирования самой информации, продемонстрирован на рис. 1

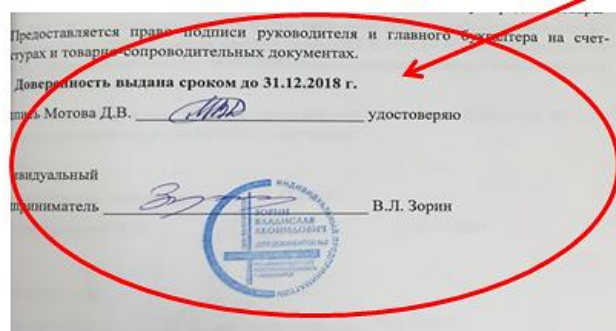


Рис. 1 Формирование информации

Итак, первоначальной формой (цифровой формой), например, является формат «Word» или иной вспомогательный формат на котором формировалась информация, после распечатывания (воспроизведенного с помощью принтера) – информация уже потеряна (скрыта). Документ из «Цифры» превратился в «Аналог», то есть, в бумажный аналог информации.

Проблематика:

Проблемная сторона бумажного носителя: как распознать тот самый «способ» формирования самой информации, который был произведен (сформирован) в «Цифре»?

Автором предлагается: получить цифровые двойники документа (ЦДД), и при помощи ЦДД искать и распознать тот самый «способ» формирования самой информации, который был произведен (сформирован) в «Цифре».

То есть, бумажную аналоговую информацию (которая ранее была в цифре) – превращаем обратно в «Цифру».

Разъяснение автора:

Цифровой двойник документа (ЦДД) – это набор данных, связанный с получением растровых (пиксельных) изображений бумажного варианта текстового документа в различных спектральных диапазонах. Обычное RGB изображение является частью ЦДД, связанного с видимым диапазоном спектра.

Обозначение ЦДД:

Это обозначение, ранее было предложено Автором и д.т.н. профессором Катаевым М.Ю. (Аллаберганов А.А. Катаевым М.Ю. Патент на изобретение МЕТОД ПОЛУЧЕНИЯ ЦИФРОВОГО ДВОЙНИКА БУМАЖНОГО НОСИТЕЛЯ (Заявка № 2019144958 Приоритет изобретения 30.12.2019 Входящий № 086664)).

Для распознавания информации используем собственный разработанный набор методик, получаем необходимую информацию для принятия решений.

Алгоритм распознавания текстовой информации:

1. Текстовая информация содержит 5 (пять) типа элемента: фон бумаги; чернила; рукописный текст; машинописный текст (изготовлен при помощи знакосинтезирующего устройства (принтер)) и оттиск печати.



Рис. 2 Совокупность информации

2. Цифровой двойник (ЦД) – состоит из цифровых образов документа (изображений текстового документа), полученных в различных спектральных диапазонах.

В системе поиска изменений документа (подделки), так как вместо явных физических характеристик бумаги, чернил, возникают их «Цветовые аналоги», т.е. преобразованные в оптической системе цифровой камеры.

Цветовые аналоги имеют следующие характеристики (показатели):

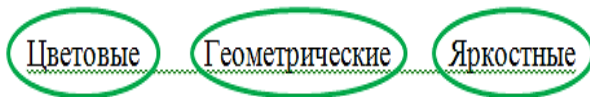


Рис. 3 Совокупность характеристик цветовых аналогов

Для получения технического результата, используем собственные разработанные методы и алгоритмы исследования:

Исследование из цифровой формы (ЦФ)

В данном исследовании - при изменении очередности, изменяя последовательность алгоритма обработки информации, исследование содержания текстовой информации на цифровом носителе, приходим к желаемому результату (определенным показателям).

Алгоритм обработки изображений (цифровых двойников) с целью выделения текстовой информации (ТИ):

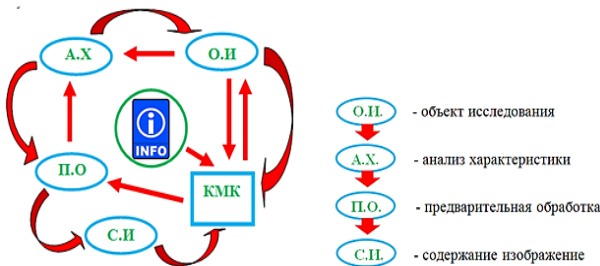


Рис.4 Алгоритм обработки (ЦД) с целью выделения (ТИ)

Алгоритм обработки получения изображений (цифровых двойников) и распознавания текста (РТ):

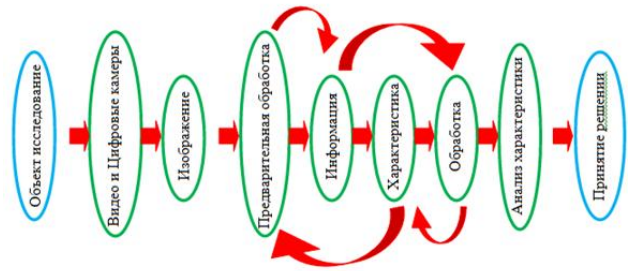


Рис. 5 Алгоритм обработки (ЦД) и распознавания (РТ)

Особенности выделения на изображении текстовую информацию при помощи соответствующих математических алгоритмов. Выделение шрифта, типа чернил связаны с задачи определения нанесения текста (напечатанного), нанесения печати, места печати и др.

Математический алгоритм с целью выделения на изображении (ТИ)

$$V(\text{вольт}) = A+B \cdot \int_{\Delta\lambda} W(\lambda) I(\lambda) d\lambda$$

Площадь
Спектральная кривая каналов R, B, G.

Коэффициенты преобразования излучения в напряжение

Рис. 6 Формула 1. Математический алгоритм

Текстовая информация содержит 5 элементов: фон бумаги; цвет чернила; текст (рукописный или машинописный); тип прибора для машинописного текста (принтер, сканер и др.) и объекты (например, оттиск печати).

При обработке изображений, содержащих текст, получаем информацию в каналы {R, B, G}.

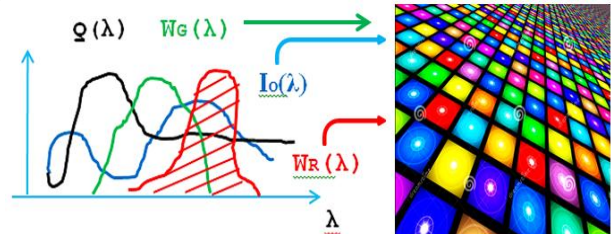
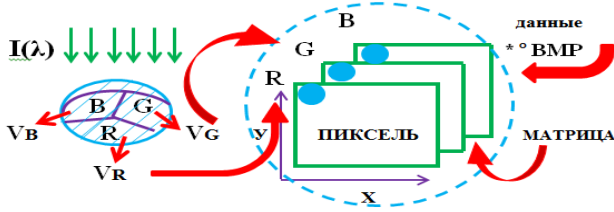


Рис. 7 Информацию по каналам {R, B, G}, полученная при обработке изображений, содержащий текст

Используя стандартные библиотеки обработки изображений можно оценить качество изображения, даваемого оптической системой (цифровой камерой).

Для очистки изображения от шумовой составляющей применяется, далее, цифровая фильтрация изображений. Показано на Рис. 8

Показан схематический рисунок:



где,

- I** – это излучение падающее на объекты;
- λ** – длина волны (спектра);

Рис. 8 Цифровая фильтрация изображений

Обработка и анализ может быть проведена не обязательно в области измеряемых значений, а в области, например, спектрального пространства (например, собственных векторов, Wavelet, преобразований дискретного косинусного преобразования и др.).

Это позволяет определить признаки объектов, которые присутствуют на исследуемом (исходном) изображении - полученную информацию находящиеся на бумажном носителе или в ином носителе.

Особенности нанесения текстовой информации можно подчеркнуть методом увеличения пространственного разрешения при помощи специализированного оборудования КМК (видеомикроскопа).

Схематический рисунок методом увеличения пространственного разрешения при помощи КМК (видеомикроскопа), показан на рисунке 9

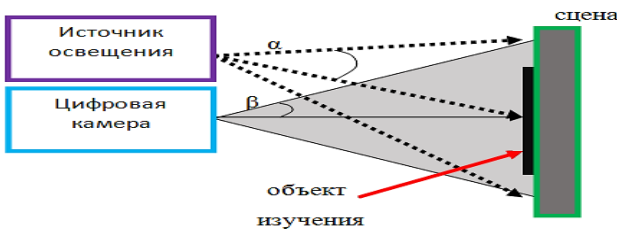


Рис. 9 Исследование, методом увеличения пространственного разрешения

Используя исследовательскую установку (КМК) Криминалистический многофункциональный комплекс, продемонстрирован на Рис. 10



Рис. 10 Криминалистический многофункциональный комплекс (КМК)

Патент на изобретение № 2674495; Заявка № 2018133545; Приоритет изобретения 24 сентября 2018 года; Государственная регистрация в Гос. реестре изобретений РФ 11 декабря 2018 года; Срок действия исключительного права на изобретение 24 сентября 2038 года.

Также, используя набор методик, разработанные автором, включая метод «Метод получения из цифровой формы (изображение) историю документа (изменения в пути), с момента его создания» (Заявка № 2019141187 Приоритет изобретения 12.2019 Входящий 080514) - и распознаем, ту самую историю документа с момента его создания – в «Цифре».

Например, на рис.11 показан формированный текст, под ранее нанесенные реквизиты, тот самый «способ» формирования самой информации:

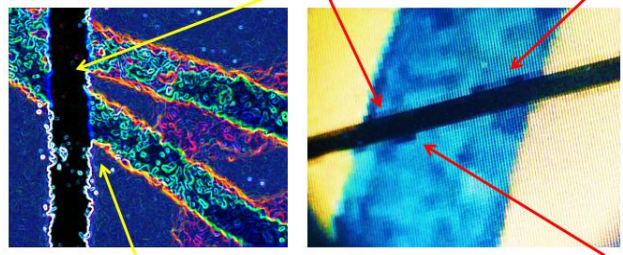


Рис. 11 Формированный текст, под ранее нанесенные реквизиты

Исследование проходило в зонах спектра и при наложении зоны спектров друг на друга, также исследование в негативе, выявлено - очередность нанесения реквизитов, также и присутствия пикселей текстового редактора при монтаже печатного текста на ранее нанесенные реквизиты.

Присутствия пикселей текстового редактора при монтаже печатного текста, продемонстрировано на рисунке 12.

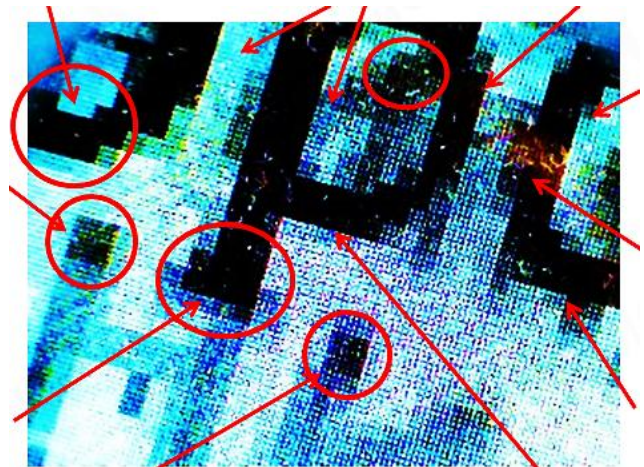


Рис. 12 Присутствия пикселей текстового редактора при монтаже печатного текста

Пояснение автора:

Бумажная аналоговая информация (оригинал) – не «конвертируется» обратно в формат «Word» при исследовании монтажа (формировании) печатного текста, а именно, создается Цифровые двойники (ЦД), в том числе и Цифровые двойники в иных, параллельных цифровых форматах.

Цифровые двойники в иных, параллельных цифровых форматах:

Далее, создаем цифровые двойники в иных, параллельных цифровых форматах.

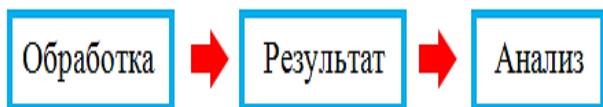
Возникает вопрос,

Вопрос: - Зачем нужен цифровой двойник (ЦД)?

Ответ: Цифровой двойник (ЦД) - это специализированный набор данных, который необходимо обработать набором методов, связанных с данными, для получения заданного количества параметров, позволяющих принять решение об изменении текстового документа.

то есть,

Данные сначала обрабатываются, получаются результаты обработки, которые уже и анализируются.



Вспоминая Методику «Системного анализа»:

«Изложенные выше системные представления являются основой для составления методики системного анализа (алгоритма системной деятельности), в соответствии с которой должен осуществляться переход из состояния проблемной ситуации в состояние желаемой конечной цели – решения проблемы или создания системы.

Этот переход должен происходить системно, упорядоченно, путем последовательного выполнения определенных шагов (этапов)» – (проф. А.М. Кориков «Теория систем и системный анализ»: учебное пособие / А.М Кориков – 2-е изд., доп. и перераб. – Томск: Томск, гос. университет систем управления и радиоэлектроники, 2008. – 264 стр.).

Итак, создаем при помощи исследовательской установки (КМК) разработанное автором, цифровые двойники и двойники в иных, параллельных цифровых форматах и исследование производим при помощи данной установки. Этот процесс продемонстрирован на Рис. 13

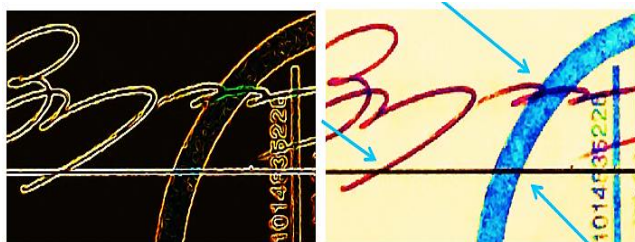


Рис. 13 Получение цифровых двойников и двойников в иных, параллельных цифровых форматах

Итак, получаем Информацию - На пустой лист бумажного носителя (бланк) нанесены Реквизиты (сначала оттиск печати, далее подпись – нарушена очередность нанесения реквизитов), далее под Оттиск печати и Подпись – был сформирован сам печатный текст.

То есть, печатный текст был нанесен при помощи монтажа.

В данном случае, это является – технической подделкой документа.

Напоминание автора: Каждый цифровой двойник представляет собой информацию о документе в различных направлениях исследования изображения:

цветного RGB, инфракрасного, в УФ области спектра, синем, красном цвете и т.д.

Порядок обращения к дополнительной информации определяется величиной оценки качества, которая зависит от условий измерений. Изображение документа может быть разбито на отдельные фрагменты и каждый из них анализируется независимо или последовательно. Для каждого фрагмента изображения формируется своя система оценки качества, из которых, в сумме, формируется общая оценка качества документа.

То есть, Изображение – метод – результат – анализ; Берем, бумажный документ, его основные части составляющего:



Как выше было сказано - Документ из «Цифры» превратился в «Аналог», то есть, в бумажный аналог информации.

И на самом бумажном носителе нет возможности (не представляется) получить ту информацию – которую можно получить из Цифровой формы.

Разъяснение автора: Документ готовился и формировался в цифровой форме, далее, при преобразовании из цифровой формы на бумажный носитель – цифровая информация теряется и не отражается на бумажном носителе.

Также, вместе с информацией и теряются все секреты (тайная информация – самой фальсификации).

Чтобы получить доступ к секретам, то есть, к тайной информации – информация должна стать «Цветным аналогом».

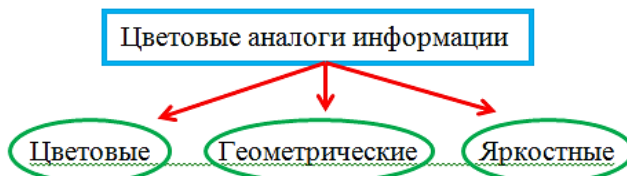


Рис. 14 Совокупность характеристик цветowych аналогов

То есть, «Цветным аналогом» той информации, содержащиеся на бумажном носителе.

Учет этого факта позволяет, что созданный набор данных об отражении документа (текста, бумаги, чернил) в различных участках спектра имеет более широкие возможности анализа, чем только RGB изображение или полученное изображение в другом диапазоне спектра.

Анализ свойств преобразование в цифру и отражения бумаги (носителя) может быть пояснен формулой формирования изображения:

$$I(\lambda) = I_0(\lambda) \cos(\varphi) r(\lambda) W,$$

Здесь $I(\lambda)$ – яркость измеренных пикселей изображения, $I_0(\lambda)$ – яркость источника излучения, $\cos(\varphi)$ – угол освещения, $r(\lambda)$ – спектральный коэффициент отражения и W – шероховатость бумаги.

Заметим, что выражение (1) не учитывает особенности измерительной техники, так как приведено лишь для показа на специфики предлагаемого подхода.

Из дополнительных исследований мы можем получить информацию о свойствах бумаги $rs(\lambda)$ и Ws . Таким образом, имея неизвестный тип бумаги, но, зная характеристики освещения, можно получить $r(\lambda)W$ и сравнить это произведение с известными данными.

Выделение участков документа с текстовой информацией, проводя аналогичные манипуляции, можно получить подобную информацию.

Как было выше указано, чтобы получить доступ к той информации, надо:

- Вернуться обратно, в среду цифровой информации.

Находясь в цифровой среде, можно получить намного больше информации о тех информациях, которые были ранее, до момента преобразования документа в бумажный носитель.

Это указывает на многолетний опыт Автора, которым были произведены более 130 (ста тридцати) судебных экспертиз, исследования документов именно **из цифровой формы (в разных форматах)** по методам, разработанных Автором в рамках гражданских и уголовных судопроизводствах.

Применение данного вида (способа) исследования «Преобразование оригинала документа в цифру и способ распознавания подделки», что намного продвигает сферу экспертной деятельности в рамках производства технико-криминалистического исследования.

Литература:

1. Баринова О.А. Использование современных инструментальных средств для распознавания подлога документов: достоинства и недостатки // Вопросы экспертной практики. 2019. № S1. С. 41-46.

2. Бахтеев Д.В. Компьютерное зрение и распознавание образов в криминалистике // Российское право: образование, практика, наука. 2019. № 3 (111). С. 66-74.

3. Кориков А.М. «Теория систем и системный анализ»: учебное пособие / А.М. Кориков – 2-е изд., доп. и перераб. – Томск: Томск, гос. университет систем управления и радиоэлектроники, 2008. – 264 стр.).

4. Аллаберганов А.А. Катаев М.Ю. / Многофункциональный исследовательский комплекс решения задач анализа текстовой информации. / Международная научно-практическая конференция (МНПК ЭС и СУ – 2018). Часть 1 – 227с.

5. Аллаберганов А.А. Катаев М.Ю. Метод получения цифрового двойника бумажного носителя // Право и практика – 2020. - № 1.

6. Аллаберганов А.А. Криминалистический многофункциональный видеомикроскопический спектральный комплекс. Методика решения задач распознавания и анализа текстовой информации из цифровой формы // Современная наука. Серия "Экономика и право" - № 1. – 2020.

References:

1. Barinova O.A. The use of modern tools for recognizing forgery of documents: advantages and disadvantages // Questions of expert practice. 2019.No S1. S. 41-46.

2. Bahteev D.V. Computer vision and pattern recognition in forensics // Russian law: education, practice, science. 2019.No 3 (111). S. 66-74.

3. Korikov A.M. "Theory of systems and systems analysis": textbook / A.M. Korikov - 2nd ed., Ext. and reslave. - Tomsk: Tomsk, state. University of Control Systems and Radio Electronics, 2008. - 264 p.).

4. Allaberganov A.A. Kataev M.Yu. / A multifunctional research complex for solving problems of textual information analysis. / International Scientific and Practical Conference (MNPК ES and SU - 2018). Part 1 - 227s.

5. Allaberganov A.A. Kataev M.Yu. A method of obtaining a digital double of paper media // Law and Practice - 2020. - No. 1.

6. Allaberganov A.A. Forensic multifunctional video microscopic spectral complex. Methodology for solving problems of recognition and analysis of textual information from a digital form // Modern Science. Series "Economics and Law" - No. 1. - 2020.

**Статья прошла проверку системой «Антиплагиат»;
Оригинальность текста – 94,86%**