

Метод получения потерянной текстовой информации из цифровой формы (изображений), восстановления и ее анализ

Аллаберганов Ахмеджан Атаханович,

аспирант, кафедра криминалистики, судебных экспертиз и юридической психологии, ФГБОУ ВО «Байкальский государственный университет», аспирант кафедры «АСУ. Фотоника, приборостроение, оптические и биотехнические системы и технологии», ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники»
E-mail: nsk-kapital@mail.ru

В статье будет рассмотрена «потерянная текстовая информация» в цифровой форме. Документ, который был преобразован в электронный файл, формат «PDF». Целью работы является восстановление и получения потерянной текстовой информации из цифровой формы (изображение) распознавание, выявление фальсификации, и ее анализ.

Ключевые слова: изображение, текст, методики анализа, распознавание, выявление и обработка, сравнение, наложение, зоны спектра.

Краткое понятие потерянной текстовой информации в цифровой форме (изображений)

Потеря текстовой информации – это частичное исчезновения или абсолютного исчезновения изображения.

Это потеря контрастности или качества изображения вследствие искажения сигнала или плохого копирования (рис. 1).

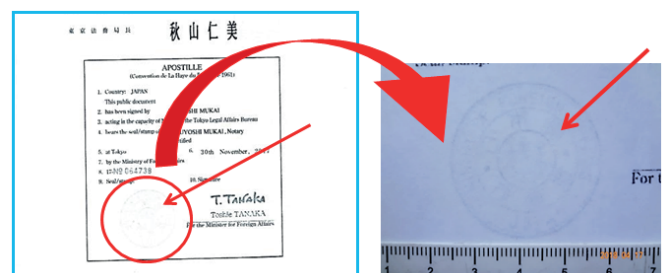


Рис. 1. Потерянная текстовая информация, отпечаток печати на «APOSTILLE» от имени Министерства Иностранных Дел Японии

В картинной плоскости, в полном пространстве или в части областей на изображении, на которых отсутствует какая-либо информация о цвете.

Либо это способ сокрытия исходных данных в недрах фотоизображений, например с помощью операционной системой Windows, также, возможно, используя архиватор Winrar.

Введение

В последнее время, в сфере экономики прослеживается увеличение совершения преступлений с фальсификацией документов. С увеличением товарооборота и расширением экономических зон, привело к увеличению документооборота и при регулировании правоотношений – привело к большому росту числа поддельных документов.

Это подтверждается статистическими данными при производстве судебных экспертиз (с открытых источников (интернет ресурсов) МВД и ЭКЦ МВД РФ).

При производстве технико-криминалистических исследований, прослеживается фальсификация (подделка) документов с применением новых технологий и программных обеспечений.

С применением технических средств и приемов, качество подделки документов выросла, наряду с этим увеличился и объем поддельных документов и правонарушений. Применение новых технологий увеличилось разнообразием подделок и способов направления изготовления подделок.

В последнее время прослеживается – практика предоставления электронных копий документов в судебные инстанции «о правоотношениях». При этом, оригиналы документов каким-то образом исчезают, то есть не предоставляются.

Возникает проблема, исследования документа с электронного формата, то есть из цифровой формы – изображение (текстовая информация).

На сегодняшний день – нет достаточно проработанной методики выделения и распознавания текстовой информации, представленной в виде изображения (цифровой формат).

В результате – объектом исследования в данной работе является документ «Текстовая информация представленный в электронной форме (изображение)», и предметом исследования является «Способы подделок текстовой информации, представленной в виде цифровых изображений», преобразованный формат «PDF».

Научная проблематика. Исследование, по выявлению и распознавания фальсификации (подделки) документов из цифровой формы текстовой информации (изображений).

Текстовая информация, представленная на бумаге (бумажный носитель) часто переносится в цифровой вид, как изображение определенного формата и значит, может быть помещено в электронный файл, например формата «PDF».

Цель исследования. Целью работы является разработка Методики и комплекса приборов для криминалистической экспертизы цифровых изображений текстовой информации документов для выявления способов подделок, обеспечивающей точность и надежность.

Задачи исследования. Задача исследования обнаружение подделки – изменение документа.

1. Исследование изображений и получение характеристик.
2. Подтверждение найденных характеристик с помощью экспериментального комплекса, а также обнаружения новых подтверждающих признаков.

Постановка задачи. Предметом исследования в данной работе является криминалистический анализ документов, представленный в цифровой форме (изображение). Манера исполнения подделок и способ изготовления, применяемые технические средства и приемы при изготовлении подделок в современных условиях.

Цель исследований связана с повышением точности идентификации изучаемых объектов.

Научные результаты исследования и их научная новизна

1. Разработана методика выделения элементов текста и методы исследования по выявлению

и распознаванию фальсификации (подделки) документа с электронного формата.

2. Алгоритм определения и распознавания текстовой информации (текста РТ) изображения, обработка и идентификация объектов исследования.
3. Криминалистический измерительный комплекс (КМК), определяющий подделку документов любой сложности.

Объект исследования. В работе рассматривается документ – «APOSTILLE» от имени Министерство Иностранных Дел и «NOTARIAL CERTIFICATE» от имени Нотариуса г. Токио, Япония (рис. 2).



Рис. 2. Объект исследования «APOSTILLE» и «NOTARIAL CERTIFICATE» – формат «PDF», имеющий плохое качество печати

Документ, из цветного изображения (RGB), преобразовывался в чёрно-белое изображение «Black and White» (B/W), и далее в «изображение в градациях серого» (Grayscale), и помещен в электронный файл, формат «PDF».

Данная текстовая информация содержит 5 (пять) типов элементов: фон бумаги; чернила; рукописный текст; машинописный текст (изготовлен при помощи знаковосинтезирующего устройства (принтер)) и оттиск печати, а также и оттиск личной подписи (печать) «ХАНКО».

Применяемое оборудование. В работе приводится описание приборного комплекса, позволяющего получать изображения текста, обрабатывать их и анализировать.

Краткое описание установки. Криминалистический многофункциональный комплекс (КМК) (рис. 3).

Все упомянутые микроскопы установлены с возможностью изменения своего вертикального и горизонтального положения, поворота на 360° по вертикальной и горизонтальным осям и изменения угла осмотра исследуемого объекта, а световые фильтры, излучатель белого света, инфракрасный излучатель и ультрафиолетовый излучатель установлены с возможностью изменения своего вертикального и горизонтального положения, а также изменения угла освещения исследуемого объекта.

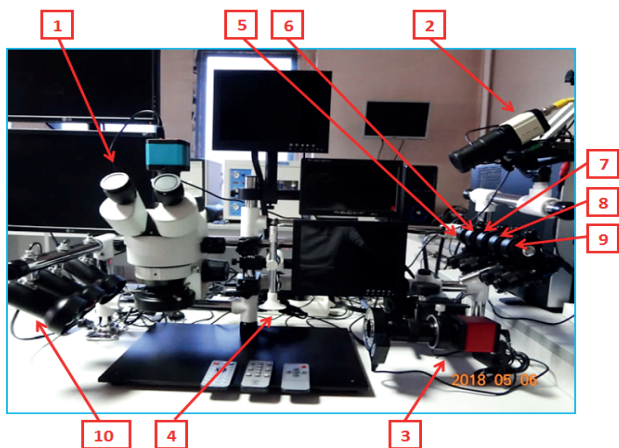


Рис. 3. Многофункциональный биолого-видеомикроскопический спектральный комплекс (данный Комплекс был разработан автором данной статьи. Патент на изобретение № 2674495; Заявка № 2018133545; Приоритет изобретения 24 сентября 2018 года; Государственная регистрация в Гос. реестре изобретений РФ 11 декабря 2018 года; Срок действия исключительного права на изобретение 24 сентября 2038 года)

Исследовательская установка, разборно-сборная с возможностью переносного действия, содержащая силовой каркас с изменениями своего положения порота и установленные на силовом каркасе: 1. Бинокулярный стереоскопический микроскоп камера; 2. Видеоинспекционный промышленный микроскоп камера; 3. Микроскоп камера предварительного просмотра с переменным фокусным расстоянием; 4. Микроскоп-эндоскоп камера; 5. Ультрафиолетовый излучатель «УФ 365 NM»; 6. Ультрафиолетовый излучатель «УФ 395 NM»; 7. Излучатель белого света «Белый А волны (луч)»; 8. Инфракрасный излучатель «ИК в зонах 850 NM»; 9. Инфракрасный излучатель «ИК в зонах 940 NM»; 10. Косопadaющие световые фильтры гаммой разных цветов;

Приборы в Комплексе (КМК) работают, как не зависимо друг от друга, так и в комплексе, как единая система. Приборы работают в области видимого спектра, фотографируя и выполняя увеличение, нужный момент можно при видео захвате перенести изображение на любой монитор и сделать наложение на другой объект исследования, при этом настраивать прозрачность объектов сравнения.

Криминалистический многофункциональный Комплекс – разбирается полностью, в течении 3 (трех) минут и собирается в течении 10 (десяти) минут, (Вес до 12 кг.).

Многофункциональный исследовательский комплекс – проводит исследование Одновременно в трех и более ракурсов с изменением своего положения. Микроскопическому исследованию и в зонах спектра, подлежат большие крупногабаритные предметы (включая, труднодоступные зоны), исследование за пределами зоны Комплекса. Криминалистический комплекс может заменить целую Лабораторию, включающего в себя Оборудование разного направления и профиля.

Техническая проблема, на решение которой направлено данное изобретение, заключается в создании мобильного криминалистического многофункционального исследовательского комплекса, позволяющего осуществлять как контактные, так и дистанционные исследования.

Применяемый алгоритм. В данной работе решаются проблемы восстановления потерянной информации и извлечения текстовой информации из изображения документа.

Алгоритм восстановления текстовой информации и распознавания текста (РТ)

Системы и алгоритмы распознавания, предлагаемые Автором можно использовать в анализе документов и текстовой информации, представленной изображением в криминалистике, медицине, биометрии, промышленности и многих других сферах (рис. 4).

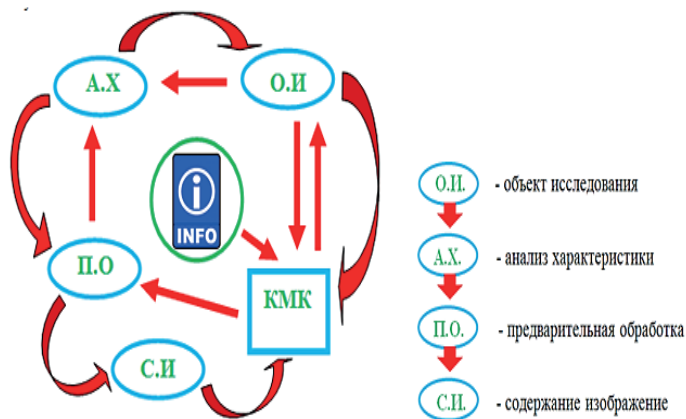


Рис. 4. Схема 1. Алгоритм обработки и распознавания текста (РТ)

Точность, устойчивость и производительность методов распознавания повышаются, за счет расширения области их применения.

Исследование. Исследование проходит «Онлайн» с электронного файла – формата «PDF», при помощи данной исследовательской установки.

Алгоритм обработки, получения изображений и распознавания текста (РТ)

Как известно: цвет – качественный показатель. В данном исследовании – при изменении очередности, изменяя последовательность алгоритма обработки информации, исследование содержания текстовой информации на цифровом носителе, приходим к желаемому результату (определенным показателям).

Процесс исследования. Документ, из цветного изображения (RGB), преобразовывается в чёрно-белое изображение «Black and White» (B/W), и может в «изображение в градациях серого» (Grayscale), также может остаться и в цветном изображении (RGB) и помещается в электронный файл, например формат «PDF».

Целью работы является попытка автоматизации процессов измерения, анализа и сопоставления текстовой информации для решения задач экспертизы в исследовательской криминалистической деятельности.

Так как, это позволит обеспечить данное направление новым видом решений и качеством идентификации объектов исследования (рис. 5).



Рис. 5. Схема 2. Алгоритм обработки изображений с целью выделения текстовой информации

Математический алгоритм

$$I(\lambda) = I_0(\lambda) \times \cos \beta \times R(\alpha) \times P(\lambda)$$

- Где,
 $I(\lambda)$ – внешнее излучение падает в камеру (в линзу);
 $I_0(\lambda)$ – источник излучения;
 $\cos \beta$ – угол падения излучения на поверхность;
 $R(\alpha)$ – функция направления отражения излучения;
 $P(\lambda)$ – коэффициент отражения поверхности;

Рис. 6. Формула 1. Математический алгоритм

Создание математических алгоритмов, позволяющих уточнить область применения, также и обобщить новые методы и методики, кроме того построить и применять качественные методы и эффективно быстро обрабатывающие алгоритмы – для решения определенных криминалистических

и конкретных задач в цифровой форме обработки изображений (рис. 6).

Следует развивать данное направление, позволяющее определять и изучать такие признаки как – части объектов исследования и их частей в цифровой форме в обработке изображений.

Формирование и исследование спектрального и цветового пространства поверхности документа (**объекта** исследования)

Изображение текстовой информации содержит черно-белое изображение, преобразованное из цветного в черно-белое изображение.

Предлагаемые Автором методы и способы позволяют выделять на изображении текстовую информацию и распознать при помощи соответствующих математических алгоритмов. Выделение шрифта, типа чернил связаны с задачи определения лица, написавшего текст (напечатавшего), нанесения печати, способ, места печати и др (рис.7).

$$I(\lambda) = I_0(x) \cdot \cos \beta \cdot R(\alpha) \cdot Q(\lambda)$$

↑ источник излучения ↑ спектр коэф. отражения

структура шершавость бумаги цвет (спектр)

где,
 I – это излучение падающее на объекты;
 λ – длина волны (спектра);
 P – коэффициент отражения;

Рис. 7. Формула 2. Выделение и распознавания на изображении текстовую информацию

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 (рис. 8).

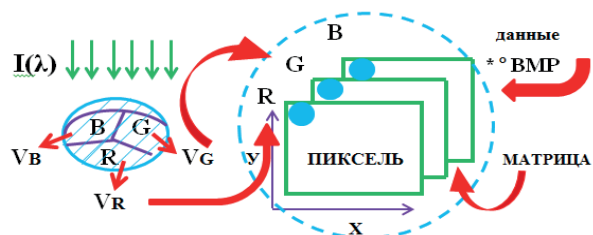


Рис. 8. Схема 3.

Обработка и анализ спектрального пространства позволяет определить признаки объектов, которые присутствуют на исследуемом (исходном) изображении.

При исследовании в видимой зоне спектра, можно выделить наилучший способ (метод) алгоритмической реализации преобразований в зависимости от вида их характеристик объекта и их частей (рис. 9).

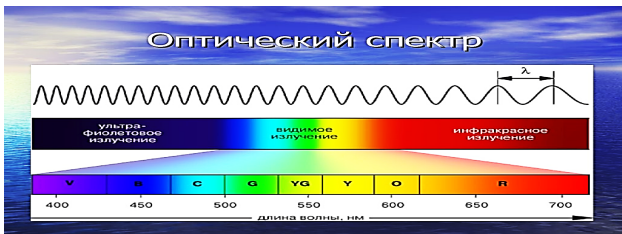


Рис. 9. Видимые зоны спектра

При исследовании и обработке выполняется оценка алгоритмов преобразований, и анализируются зависимости правильного вычисления признаков в объекте от уровня помех.

Данное исследование подтверждается – эффективностью, качеством и быстродействующим алгоритмом.

Полученные результаты. В результате примера, Автор приведет исследование и распознавание фальсификации (подделки) документа в электронном формате «изображение» файле «PDF».

Установление:

- 1) восстановление потерянной текстовой информации в цифровой форме;
- 2) выявление и распознавание фальсификации текстовой информации в электронном формате «изображение» файле «PDF» (рис. 10).

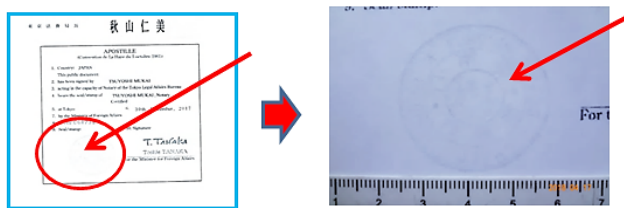


Рис. 10. Оттиск печати от имени МИД Японии преобразованный формат «PDF», с плохим качеством изготовления (для иллюстрации был распечатан)

Обработка и анализ проведена вышеуказанным Автором методом исследования. Это позволяет определить признаки объектов, которые присутствуют на исследуемом (исходном) изображении.

Как известно: Цвет – качественный показатель.

Документы очень плохого качества изготовления, исследование проходило «Онлайн» на Комплексе с электронного файла «PDF», при помощи специализированных видеомикроскопов и специальных ПО, которыми оснащен сам Комплекс (КМК), при исследовании улучшая качество изображения.

Восстановление потерянной текстовой информации (в цифровой форме), определения и распознавания фальсификации

При исследовании и обработки выполняется оценка алгоритмов преобразований, и анализируются зависимости правильного вычисления признаков в объекте от уровня помех.

Данное исследование подтверждается – эффективностью, качеством и быстродействующим алгоритмом.

Получение результата: (выявление подделки документа).

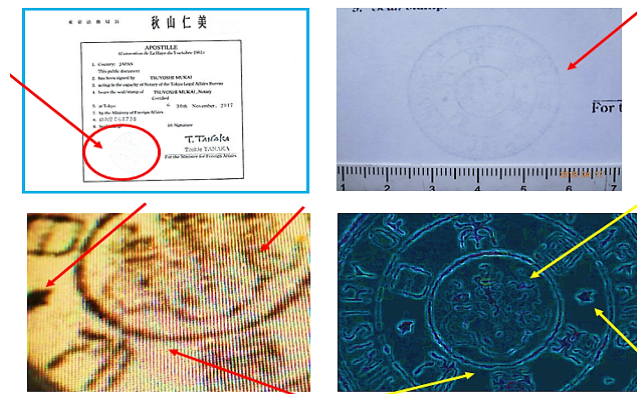
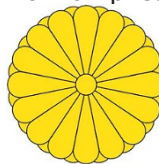


Рис. 11. Исследования в Негативе и в зонах спектра, восстановление и распознавание потерянной Текстовой информации

Выводы

1. Изображение оттиска Круглой печати от имени Министерство Иностраных Дел ЯПОНИИ на «APOSTILLE» – не является изготовленным высокой печатью.
2. В средней части Оттиска печати от имени МИД Японии на «APOSTILLE» – отсутствует изображение Хризантемы.



3. Присутствие «Пятиконечной Звезды» на оттиске печати от имени МИД Японии на «APOSTILLE» – вызывает Большое удивление.
4. Оттиска печати от имени МИД Японии на «APOSTILLE» – является грубой подделкой и изготовлен специалистом очень низкой квалификации и не имеющим познание в символике Японии.

Заключение

Применение данного Метода исследования и Алгоритма обработки текстовой информации в цифровой форме, при получении изображений и распознавания текста (РТ) позволяет применять эффективно быстро обрабатывающие алгоритмы и качественные методы – для решения определенных криминалистических и конкретных задач в цифровой форме обработки изображений.

Литература

1. Фомин Я.А. Распознавание образов: теория и практика. Издание третье, дополненное / Я.А. Фомин // М.: ФАЗИС. – 2014. – 460 с.
2. Журавлев Ю.И. Распознавание. Математические методы. Программная система. Практи-

ческие применения / Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько. М.: ФАЗИС, 2006. – 176 с.

3. Местецкий Л.М. Математические методы распознавания образов / Л. М. Местецкий. – М.: МГУ, ВМиК, 2002. – 85 с.

METHOD OF OBTAINING LOST TEXT INFORMATION FROM DIGITAL FORM (IMAGES), RECOVERY AND ITS ANALYSIS

Allaberganov A.A.

FSBEI of HE «Baikal State University», FSBEI of HE «Tomsk State University of Control Systems and Radioelectronics»

The article will consider the “lost text information” in digital form. The document, which was converted into an electronic file, format “PDF”.

The aim of the work is to recover and receive lost text information from a digital form (image) recognition, identification of fraud, and its analysis.

Keywords: image, text, methods of analysis, recognition, detection and processing, comparison, overlay, spectrum areas.

References

1. Fomin Ya.A. Pattern recognition: theory and practice. Third Edition, supplemented / Ya.A. Fomin // М.: PHASIS. – 2014. – 460 p.
2. Zhuravlev Yu.I. Recognition. Mathematical methods. Software system. Practical applications / Yu.I. Zhuravlev, V.V. Ryzanov, O.V. Senko. М.: FAZIS, 2006. – 176 p.
3. Mestetsky L.M. Mathematical methods of pattern recognition / L. M. Mestetsky. – М.: MSU, VМиК, 2002. – 85 p.