УДК 343
ББК 67
DOI 10.24411/2073-3313-2020-10197

МЕТОД РАСПОЗНАВАНИЯ ТЕКСТОВОЙ ИНФОРМАЦИИ ИЗ ЦИФРОВОЙ ФОРМЫ (ИЗОБРАЖЕНИЕ), ОПРЕДЕЛЕНИЕ ОЧЕРЕДНОСТИ НАНЕСЕНИЯ РЕКВИЗИТОВ

Ахмеджан Атаханович АЛЛАБЕРГАНОВ, аспирант кафедры криминалистики, судебных экспертиз и юридической психологии Байкальского государственного университета, аспирант кафедры «АСУ. Фотоника, приборостроение, оптические и биотехнические системы и технологии» Томского государственного университета систем управления и радиоэлектроники
E-mail: nsk-kapital@mail.ru

Научная специальность: 12.00.12 — криминалистика; судебно-экспертная деятельность; оперативно-розыскная деятельность

Аннотация. Разработан метод обработки и анализа спектрального пространства, получения изображений и распознавания текста (текстовой информации) видеоспектральным и видеомикроскопическим методом исследования с применением Криминалистического многофункционального комплекса (КМК) при помощи криминалистического программного обеспечения, позволяющий определить признаки объектов, присутствующих на исследуемом (исходном) изображении.

При этом в видимой зоне спектра может быть выделен оптимальный способ (метод) алгоритмической реализации преобразований в зависимости от вида характеристик объекта и их частей. Предложенный метод (способ) исследования — распознавание очередности нанесения реквизитов из цифровой формы, выделение элементов текста и распознавание объектов и их частей, выявление фальсификации (подделки) документа, текстовых знаков в электронном формате (цифровой форме) — способен существенно повысить эффективность экспертной деятельности в рамках производства технико-криминалистического исследования.

Ключевые слова: криминалистическая экспертиза, документ, подделка, очередность нанесения реквизитов, цифровая форма.

Annotation. A method has been developed for processing and analyzing the spectral space, obtaining images and recognizing text (textual information) by the video spectral and video microscopic research method, using the Forensic Multifunctional Complex (CMC) with the help of forensic software, which allows to determine the features of objects that are present on the investigated (source) image. At the same time, in the visible region of the spectrum, the optimal method (method) of the algorithmic implementation of transformations can be selected depending on the type of their characteristics of the object and their parts.

The proposed research method (method), recognition of the sequence of drawing details from a digital form, highlighting text elements and recognizing objects and their parts, identifying falsification (fake) of a document, text characters in electronic format (digital form) can significantly increase the effectiveness of expert activities in the production forensic research.

Keywords: forensic examination, document, fake, sequence of application details, digital form.

Текстовая информация, представленная на бумаге (бумажном носителе) часто переносится в цифровой вид, такой, как изображение определенного формата, и, значит, может быть помещена в электронный файл, например, формата «PDF» (см. рис. 1).

Рис. 1. Текстовая информация, помещенная в цифровой формат

На сегодняшний день нет достаточно проработанной методики выделения и распознавания текстовой информации, представленной изображением.

Целью работы является распознавание фальсификации из цифровой формы, определение очередности нанесения реквизитов, изменение и формирование самих объектов исследования и их частей при изготовлении реквизитов и части письма — при изготовлении разными способами текстовой информации. Применение новых способов и методов исследования, а также иных приборов и комплексов с техническими характеристиками оптических измерительных устройств дает возможность выделения элементов текста и распознавания объектов и их частей из цифровой формы (см. рис. 2).

Рис. 2. Исследование при помощи специального видеомикроскопа

Применяемое оборудование

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

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

Размещение конструктивных элементов установки на силовом каркасе и обеспечение их подвижности по осям позволяет проводить исследования во всех плоскостях, при этом все упомянутые микроскопы установлены с возможностью изменения своего вертикального и горизонтального положения, поворота на 360є по вертикальной и горизонтальным осям и изменения угла осмотра исследуемого объекта, осуществлять разворот как отдельных элементов, так и всей установки на 360є. Также за счет возможности монтажа / демонтажа установка получается переносной и быстро сборно-разборной.

Оснащение установки упомянутыми микроскопами, съемно установленными с возможностью вращения, позволяет проводить исследования за пределами установки, например, крупногабаритные объекты можно исследовать на расстоянии до 2,5 метров от установки, а также в труднодоступных местах.

Наличие инфракрасных излучателей, ультрафиолетовых излучателей, излучателя белого света и световых фильтров позволяет проводить исследования одновременно в разных зонах спектра и накладывать зоны спектра друг на друга.

Исследование проходит онлайн с электронного файла формата «PDF» при помощи установки — «Многофункционального биолого-видеомикроскопического спектрального комплекса» (см. рис. 3).

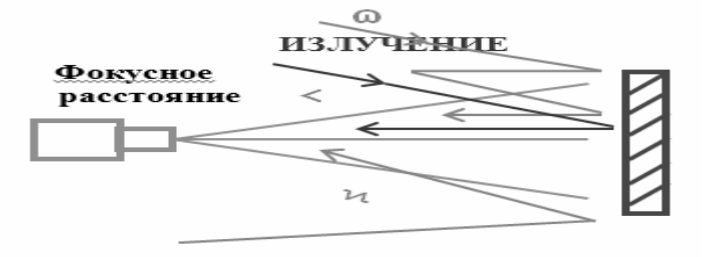


Рис. 3. Исследование при помощи специального видеомикроскопа

Алгоритм обработки изображений и распознавания текстовой информации (см. рис. 4).

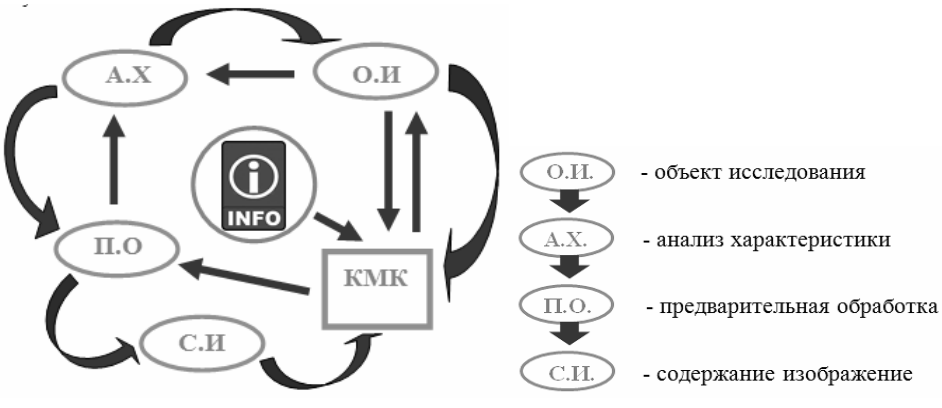


Рис. 4. Алгоритм обработки и распознавания текста (РТ)

Постановка задачи

Предметом исследования в данной работе является криминалистический анализ документов. Задача получения изображений и распознавания текста (РТ), т.е. точности идентификации изучаемых объектов.

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

Алгоритм обработки, получения изображений и распознавания текста (РТ) (см. рис. 5).

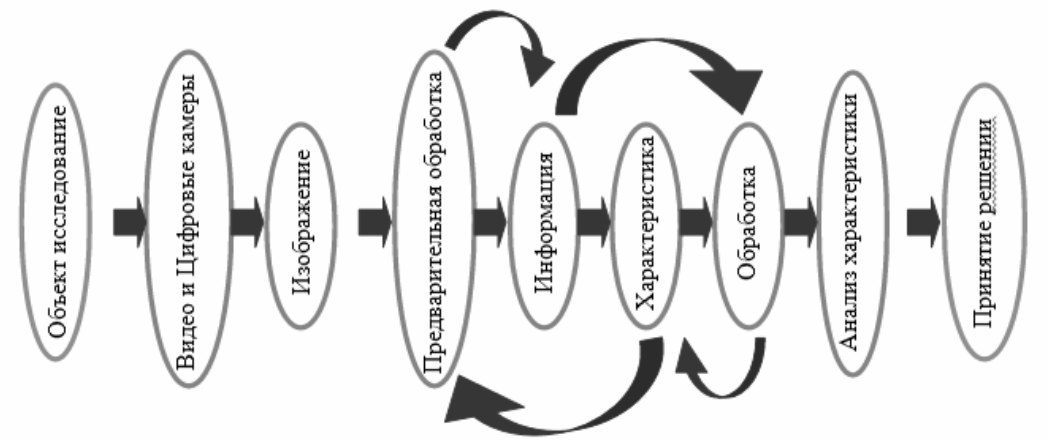


Рис. 5. Алгоритм обработки изображений с целью выделения текстовой информации

Процесс исследования

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

Это позволит обеспечить данное направление новым видом решений и качеством идентификации объектов исследования.

Математический алгоритм:

$$I(\lambda) = I_{\circ}(\lambda) \times \cos\beta \times R(Q) \times \text{Þ}(\lambda)$$

Где,

$I(\lambda)$ – **внешнее излучение падает в камеру (в линзу);**

$I_{\circ}(\lambda)$ – **источник излучения;**

$\cos\beta$ — **угол падения излучения на поверхность;**

$R(Q)$ – **функция направления отражении излучения;**

$\text{Þ}(\lambda)$ – **коэффициент отражения поверхности;**

Создание математических алгоритмов, позволяющих уточнить область применения, а также обобщить новые методы и методики, и, кроме того, построить и применять качественные методы и эффективно быстро обрабатывающие алгоритмы необходимо для решения определенных криминалистических и конкретных задач в цифровой форме обработки изображений.

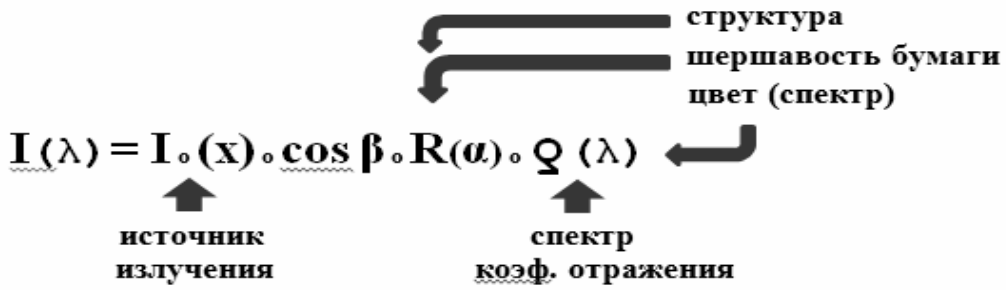
Следует развивать данное направление, позволяющее определять и изучать такие признаки, как части объектов исследования и их части в цифровой форме в обработке изображений.

Формирование спектрального и цветового пространства поверхности документа (объекта исследования)

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

Предлагаемые методы и способы позволяют выделять на изображении текстовую информацию и распознавать ее при помощи соответствующих математических алгоритмов. Выделение шрифта, типа чернил, оттиска печати связано с задачей определения очередности нанесения реквизитов.

Выделение и распознавание на изображении текстовой информации:



где,

I – это излучение падающее на объекты;

λ – длина волны (спектра);

β – коэффициент отражения;

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 (см. рис. 6).

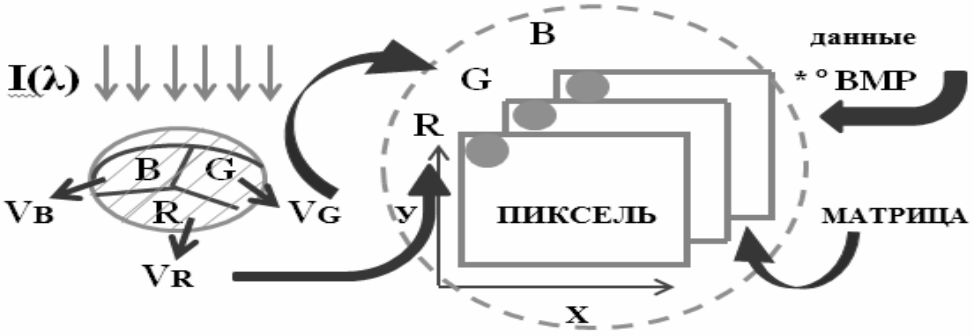


Рис. 6. Цифровая фильтрация изображений

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

При исследовании в видимой зоне спектра (см. рис. 7) можно выделить наилучший способ (метод) алгоритмической реализации преобразований в зависимости от вида характеристик объекта и их частей.

Рис. 7. 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.

Алгоритм обработки изображений, распознавания очередности нанесения реквизитов в цифрофой форме (в различных форматах) (см. рис. 8).

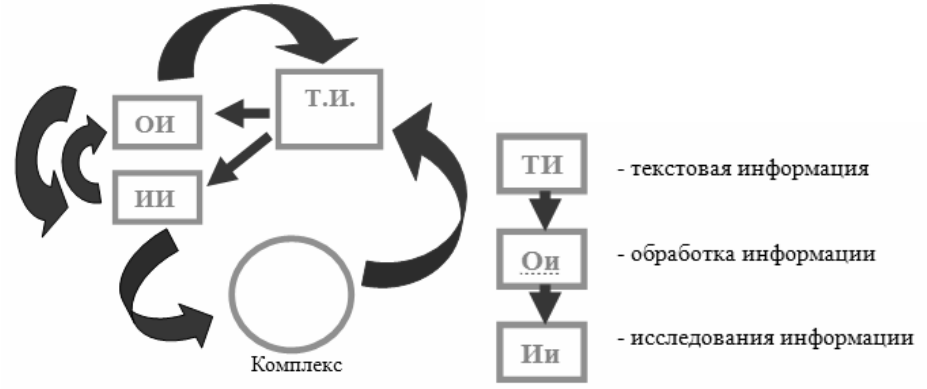


Рис. 8. Алгоритм обработки изображений, распознавания очередности нанесение реквизитов в цифровой форме

Получение результата:

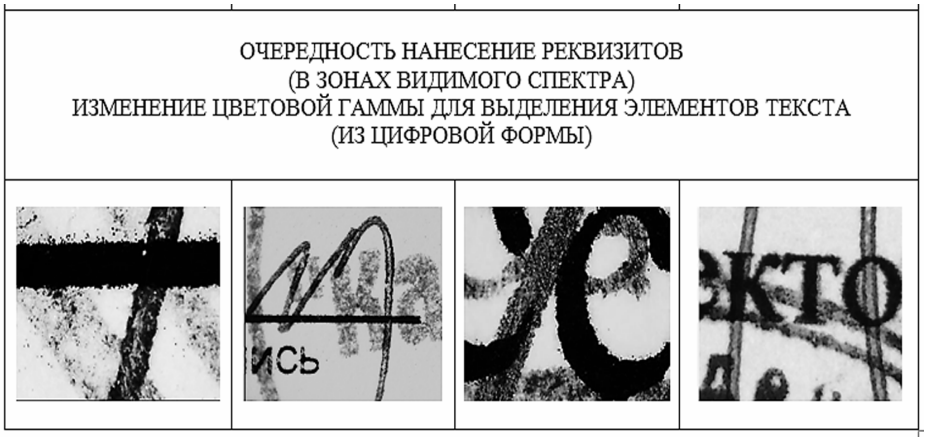


Рис. 9. Очередность нанесения реквизитов (в зонах видимого спектра), изменение цифровой гаммы для выделения элементов текста (из цифровой формы)

На рис. 9 показаны (на основе проведенных автором исследований (судебных экспертиз), с разных цифровых форматов) изменение (с целью подделки) и формирование самих объектов исследования и их частей при изготовлении реквизитов и части письма — при изготовлении разными способами текстовой информации.

Применение данного метода (способа) исследования — распознавание очередности нанесения реквизитов из цифровой формы, выделение элементов текста и распознание объектов и их частей, выявление фальсификации (подделки) документа, текстовых знаков в электронном формате (цифровой форме) — намного продвинет сферу экспертной деятельности в рамках производства технико-криминалистического исследования.

Библиографический список

1. *Аллаберганов А.А., Катаев М.Ю.* Многофункциональный исследовательский комплекс решения задач анализа текстовой информации. / Междунар. науч.-практ. конф. (МНПК ЭС и СУ — 2018). Часть 1. 227 с.

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

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

4. *Булдакова Т.И., Славин О.А., Путинцев Д.Н.* Алгоритмы интеграции результатов распознавания в видеопоследовательностях полей документов, удостоверяющих личность // Междунар. журнал прикладных и фундаментальных исследований. 2017. № 7-2. С. 172—175.

5. *Усилин С.А., Арлазаров В.В., Алиев М.А., Маталов Д.П.* Программа распознавания признаков подлинности «Smart Document Forensics». Свидетельство о регистрации программы для ЭВМ RU 2018615343. 07.05.2018 г. Заявка № 2018612851 от 23.03.2018 г.