

УДК: 004.932.2

DOI: 10.22213/2410-9304-2023-2-27-40

Методика извлечения цифровых представленных текстов по изображению документа и их экспертиза

А. А. Аллаберганов, аспирант, Томский государственный университет систем управления и радиоэлектроники, Томск; Судебная экспертно-криминалистическая лаборатория при УК «НСК-Капитал», Новосибирск, Россия

М. Ю. Катаев, доктор технических наук, профессор, Томский государственный университет систем управления и радиоэлектроники, Томск, Россия

В данной работе решаются проблемы извлечения текстовой информации из изображения документа. Важно отметить, что работа включает в себя комплексное использование специальных знаний по трем направлениям: технико-криминалистическая экспертиза документов; фототехническая и компьютерно-техническая экспертиза. Рассмотрена и детально продемонстрирована информация обнаружения изменений текстовой информации (чернил, тонера и паст пишущих приборов) по цифровым изображениям.

Совершение преступлений в сфере экономики, как правило, связано с фальсификацией (подделкой) документов, и заметное в последние годы увеличение документооборота при регулировании правоотношений привело к значительному росту числа поддельных документов. Это подтверждается статистическими данными о произведенных судебно-почерковедческих и технико-криминалистических экспертизах. Так как документ является основой всей деятельности человечества, оно содержит данные, оформленные в некотором порядке и имеющие правовое значение.

Проблема приобрела множество интересных решений благодаря практическим приложениям, таким как распознавание текстов, перевод, криминалистика и др. Несмотря на то что этой проблемой занимаются долгое время, решение до сих пор отсутствует, т. е. нет проработанной методики с распознаванием изображений документов.

Изображение текстовой информации зависит от типа сцены, цифровой камеры, величины и направления освещенности, подложки и т. д. Здесь важно отметить, что при получении информации (исследуемого документа) в современных условиях можно зафиксировать документ при помощи любого устройства (фотоаппарат, смартфон, сканер и т. д.) и в любых условиях (день/ночь, улица, источник излучения), которые нам предварительно неизвестны. Это является проблемой в современных условиях при исследовании электронно-цифровой информации текстовых документов. В данной работе нами было произведено исследование по решению данной проблемы.

Ключевые слова: документ, цифровые изображения, подлинность, бумага, реквизиты документа, гистограммы яркости, извлечения информации.

Введение

Первым шагом в понимании типа чернил, тонера и паст пишущих приборов является определение характеристик, извлеченных из изображений (цветовые, текстурные), или комбинации из них.

Цель данной работы состоит в том, чтобы обеспечить проверку подлинности чернил, тонера и паст пишущих приборов и их нанесения на оцифрованные документы (изображения).

Данная работа является продолжением исследования предыдущей работы по анализу типов бумаги на подлинность по изображению.

Общая постановка проблемы

Активное внедрение цифровизации во все сферы человеческой деятельности привело к процессу ускорения подачи и приема информации и затронуло сферы науки и практики,

а также криминальную сферу. Цифровая форма документа приводит к ускорению передачи информации, защите и точности назначения передачи, заметному увеличению документооборота между государством, промышленностью и человеком. При этом бумажная форма все еще остается в процессах образования, передачи и преобразования документов и параллельно с этим создаются цифровые формы этих документов в виде изображений или файлов разного типа. Появились новые способы подделки документов, изготовленных путем монтажа с использованием цифровых технологий, в том числе перенесенная цифровая информация (подделка) путем знаковсинтезирования в аналоговую бумажную информацию. Выявление и распознавание данной информации в современных условиях является проблемным. В на-

стоящее время эта область малоизучена, отсутствуют проработанные методики судебной экспертизы документов, позволяющей разрешать данные задачи.

Предметная область ТЭД. Техническая экспертиза документов (ТЭД) относится к традиционным видам криминалистических экспертиз. Предметом ТЭД являются фактические обстоятельства, связанные с изготовлением документов, отождествлением материалов документов и технических средств изготовления документов (то есть изучения свойств документа и технических устройств с помощью специальных приемов и средств), которые могут иметь значение доказательств по уголовным и гражданским делам, устанавливаемым на основе технико-криминалистического исследования документов.

ТЭД включает в себя исследование реквизитов документа и материалов документа.

Актуальность предметной области. Актуальность исследования связана с тем, что одним из компонентов расследования преступлений является изучение документов, связанных с установлением обстоятельств, входящих в предмет доказывания, что невозможно без применения специальных знаний. Необходимо выяснить факт подделки (подлога) документа, а также установить средство, с помощью которого выполнены реквизиты. В ходе исследования документа может быть получена важная информация, которая может быть использована в качестве доказательств в форме производства экспертизы.

Проблемная сторона предметной области. В традиционном ТЭД успешно распознаются документы, содержащие бумагу, текст, подпись, печать. А ее цифровая форма, то есть документы, которые содержат его изображение, не распознается совсем. Так как в современных условиях можно зафиксировать документ при помощи любого устройства (фотоаппарат, смартфон, сканер и т. д.) и в любых условиях (день/ночь, улица, источник излучения), которые нам неизвестны, главным из указанных пунктов является *бумага, текст, подпись, печать*, которые требуют особого внимания. Для глубоко изучения необходимо получить истинное изображение бумаги (типа чернил, тонера и паст пишущих приборов), которое не зависит от внешних факторов (освещения) и от устройств фиксации.

Это и является проблемой в современных условиях при исследовании электронно-цифровой информации текстовых документов (по изображению).

Способ решения задачи. При наличии бумажной версии возможны многочисленные варианты экспертизы по строению бумаги и ее цвету, наличию типов чернил или порошка нанесения текста и т. д. Другое дело, когда имеется лишь цифровое изображение документа, которое не всегда известно, как получено и с помощью какого цифрового аппарата.

Решение задачи состоит в том, что каждое изображение документа может рассматриваться как источник сравнения с изображениями, которые являются шаблонами с известными типами бумаг, типа чернил, тонера и паст пишущих приборов, типами источников изображений. Начальным этапом понимания типа бумаги, типа чернил, тонера и паст является определение характеристик, извлеченных из изображений: цветовые, геометрические, текстурные или их комбинации. Согласно проведенному нами научному исследованию, мы видим отражение бумаги и красящих веществ и можем получить истинное изображение чернил, мастики и порошка (тонера), которое не зависит от внешних факторов (освещения), в том числе от устройства фиксации. Показатели красящих веществ кардинально отличаются друг от друга. Здесь мы можем с точностью анализировать и идентифицировать красящие вещества, нанесенные на бумагу по изображению документа.

Выбранный способ решения задачи. Общая идея подходов к получению изображений документа заключается в том, что источник изображений (сканер или цифровая камера) на выходе оставляет изображение. Это изображение несет в себе всю информацию о том, какой источник освещения был использован, на каком типе бумаги подготовлен документ (и красящие вещества на нем), какая цифровая матрица была в основе формирования изображения. Формирование изображения связано с решением уравнения свертки, где присутствует источник освещения, спектральная функция отклика и спектральный коэффициент отражения, которые могут помочь в экспертизе. Мы предлагаем на основе знаний о формировании изображения использовать заранее заготовленные изображения и показатели от них, чтобы сравнивать с неизвестными, которые подлежат экспертизе.

Полученные результаты. С применением данного метода исследования мы получаем истинное изображение типа чернил, мастики и порошка (тонера) и их характеристики, которые не зависят от внешних факторов (освещения), в том числе и от устройства фиксации. Мы распознаем и идентифицируем по изображению

типа красящего вещества, нанесенного на бумагу, цвету, яркости изображений и условиях получения.

Типы чернил, тонера и паст пишущих приборов

Мастика штемпельная – это штемпельная краска, т. е. краска, предназначенная специально для окрашивания оттисков клише печатей или штампов. Из этого и происходит название «мастичная печать» или «мастичный штамп». Штемпельная мастика имеет различный состав и свойства. По видам применения штемпельная краска делится по видам материалов, для которых она предназначена, и своим характеристикам.

Тонер – мелкий порошок черного цвета. В цветном аппарате (в принтере) применяется фиолетовый, желтый и синий краситель. Порошок состоит из частичек полимеров, покрытых красителем, содержащим магнетит и регулятор заряда. У разных производителей порошок отличается по размеру зерна, намагниченности и дисперсности, рекомендуется производителем использовать те порошки (тонеры), которые предназначены для конкретной модели.

Чернила (пасты для пишущих приборов) представляют собой сложные смеси органических соединений, основными компонентами которых являются красители, синтетические смолы и высококипящие растворители. Кроме того, в большинстве чернил содержатся такие компоненты, как поверхностно-активные вещества, антикоррозионные добавки, электропроводящие соли, консерванты, биоциды, регуляторы pH. То есть красящие вещества окрашенного компонента – метилпроизводных парарозанилина (тетра-, пента- и гексаметилпарарозанилина), метилзамещенных трифенилметановых красителей.

Цифровые источники изображений

Одним из важных источников цифровой формы документов являются сканеры, которые служат связующим звеном между физической, бумажной формой документа и его изображением. Хранение, поиск цифровых документов и обмен ими существенно проще и быстрее при наличии телекоммуникационных технологий пересылки информации между распределенными потребителями. Доверия к изображению до-

кумента меньше, так как обрабатывать цифровую форму возможно с помощью даже простых программных инструментов, несмотря на наличие средств защиты. Таким образом, чтобы узнать неизменность содержания изображения относительно оригинала, необходимо иметь понимание принципов работы планшетного сканера. Это дает понять, как формируется изображение, какого типа возникают шумы в формируемом изображении.

Цифровые камеры могут служить для получения цифровой формы документов, но они плохо приспособлены для сканирования, особенно когда требуется высокое разрешение [1]. На самом деле изображение, полученное камерой, будет затронуто разного рода искажениями, отражениями, тенями и размытием из-за проблемы стабилизации камеры во время экспозиции. Поскольку сканеры используют те же технологии, что и цифровые камеры, но предназначены чисто для сканирования [2] бумажной формы документов, интересно сравнить эти подходы. В обоих подходах используются одни и те же компоненты для получения цветовых значений [3] (RGB) документа, включая объектив, матрицу RGB, аналого-цифровой преобразователь (АЦП). Для работы с изображениями документов применяются одинаковые алгоритмы предварительной обработки изображений для улучшения краев текста, коррекция цвета и т. д. В табл. 1 приведены наиболее заметные различия между сканерами и цифровыми камерами.

Идентификация типа чернил, тонера и паст пишущих приборов по изображению документа

Общая идея подходов к получению изображений документа заключается в том, что источник изображений (сканер или цифровая камера) на выходе оставляет изображение. Это изображение несет в себе всю информацию о том, какой источник освещения был использован, на каком типе бумаги подготовлен документ, какая цифровая матрица была в основе формирования изображения [4]. В действительности, помимо полезного сигнала, матрица фиксирует и некоторый шум, добавляя его к конечному изображению [5], что является одной из задач предварительной обработки изображений.

Таблица 1. Сравнение подходов получения изображения документа сканером и цифровой камерой

Table 1. Comparison of approaches for obtaining a document image by a scanner and a digital camera

Параметр	Сканер	Цифровая камера
Матрица	Трехлинейная	2D
Источник освещения	Люминесцентная лампа, светодиод (LED излучатель)	Естественное освещение, искусственный свет, вспышка
Время сканирования	Зависит от разрешения	Быстрое, но требуется настройка

Формирование изображения связано с решением уравнения свертки, где присутствует источник освещения I_0 , спектральная функция отклика A и спектральный коэффициент отражения ρ в виде

$$I(\lambda(i)) = \int I_0(\lambda') A(\lambda(i) - \lambda') \cdot \rho(\lambda') d\lambda',$$

$$i = R, G, B. \quad (1)$$

Уравнение (1) описывает формирование сигнала в матрице RGB, который далее попадает на АЦП, и выполняется линейное преобразование типа:

$$DN(i) = a + b \cdot I(\lambda(i)), \quad (2)$$

где a, b – коэффициенты преобразования, а значения DN находятся в диапазоне $[0, 255]$.

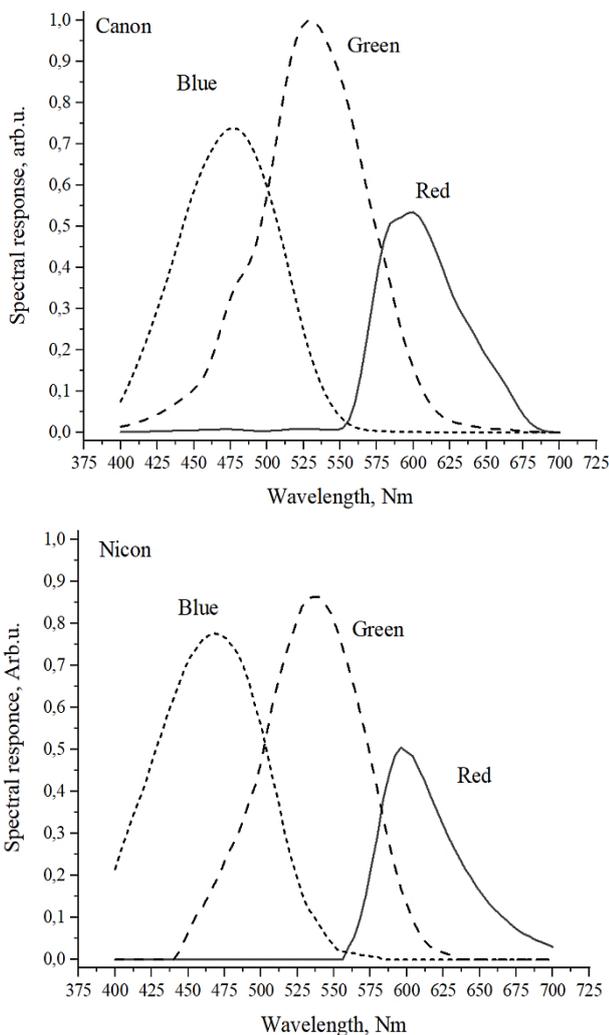


Рис. 1. Показатели «RGB» устройств фиксации (Canon, Nikon)

Fig. 1. "RGB" Values of Fixing Devices (Canon, Nikon)

Поэтому нами делается предположение, что каждое изображение документа может рассматриваться как источник сравнения с изображениями, которые являются шаблонами, с известными типами бумаг, типами источников изображений (рис. 1, 2) [6].

Можно выделить несколько категорий для сравнения: 1) распределение освещенности по пространственной области документа; 2) цветовые особенности; 3) текстурные особенности и 4) пыль и царапины.

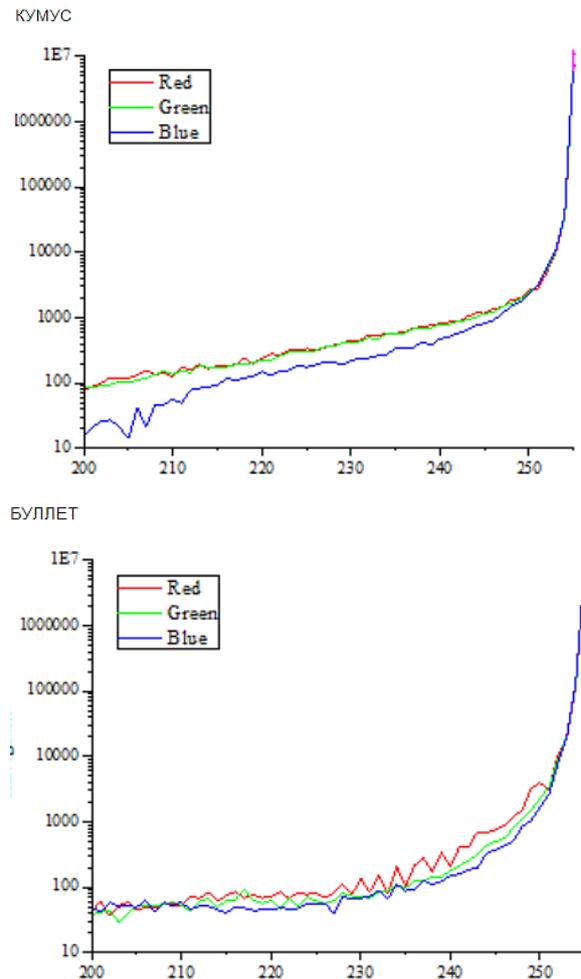


Рис. 2. Показатели «RGB» бумаг («Комус» и «Балет» – тип бумаги)

Fig. 2. "RGB" indicators of papers (Comus and Ballet - type of paper)

Исходя из вышеизложенного нами представлена схема проведения экспериментального исследования – для понимания «методики по исследованию бумажных носителей», которая является частью основной методики «Цифровой двойник изображений документа» [7].

Схема проведения экспериментального исследования продемонстрировано на рис. 3.

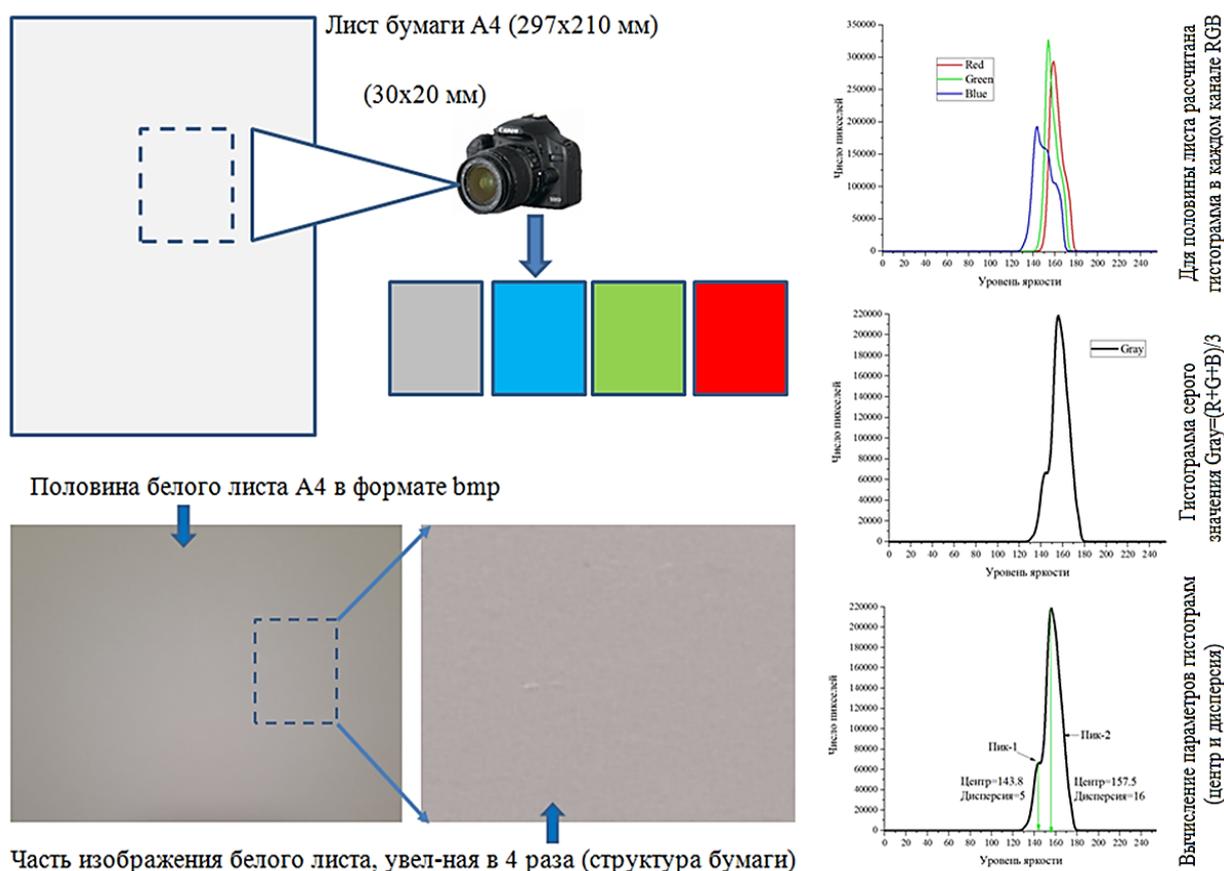


Рис. 3. Схема проведения экспериментального исследования

Fig. 3. Scheme of the experimental study

Получаемые значения обработанного уравнения формирования изображения необходимы для дальнейшего исследования.

Исходное уравнение:

$$I(x,y,\lambda) = [A(\lambda - \lambda') \{I_0(\lambda') * \cos(\varphi) * \rho(\lambda') / R^2\} d\lambda', \quad (3)$$

где a – аппаратные функции каналов R, G, B ; ρ – спектральный коэффициент отражения бумаги; R – расстояние от источника света до листа бумаги.

Перейдя к отражению бумаги, мы можем получить истинное изображение, которое не зависит от внешних факторов (освещения):

$$\langle \rho \rangle = I(x,y,\lambda) / [A(\lambda - \lambda') \{I_0(\lambda') * \cos(\varphi) / R^2\} d\lambda'.$$

Далее будет продемонстрировано, как выглядят аппаратные функции каналов R, G, B для матрицы, условия фиксации и источники излучения (рис. 4). Здесь видно (рис. 4), что для другой матрицы спектры отличаются [8]: амплитуды, каналы, ширины и центры отличаются от матриц у различных производителей [9].

Как выше было сказано: перейдя к отражению бумаги, чернил, тонера и паст, мы можем

получить истинное изображение бумаги, типа чернил, тонера и паст пишущих приборов, которые не зависят от внешних факторов (освещения) и от устройств фиксации. Что мы и демонстрируем и применяем на практике [10].

Исследования типа чернил, тонера и паст пишущих приборов. Для научного экспериментального исследования нами были взяты типы чернил для принтеров и мастики для печатей, типы порошка лазерных принтеров (тонер) и типы чернил для шариковых ручек, нанесенные на бумажные носители, которые распространены для использования документации в РФ.

1. Нанесены красящие вещества на бумажный носитель:

- а) нанесены чернила для струйного принтера на бумажный носитель (бумагу);
- б) нанесена мастика для печатей (оттиски печатей) на бумажный носитель;
- в) нанесен тонер (порошок) для лазерных принтеров на бумажный носитель;
- г) нанесены чернила для шариковых ручек (узлом пишущего прибора) на бумажный носитель (бумагу).

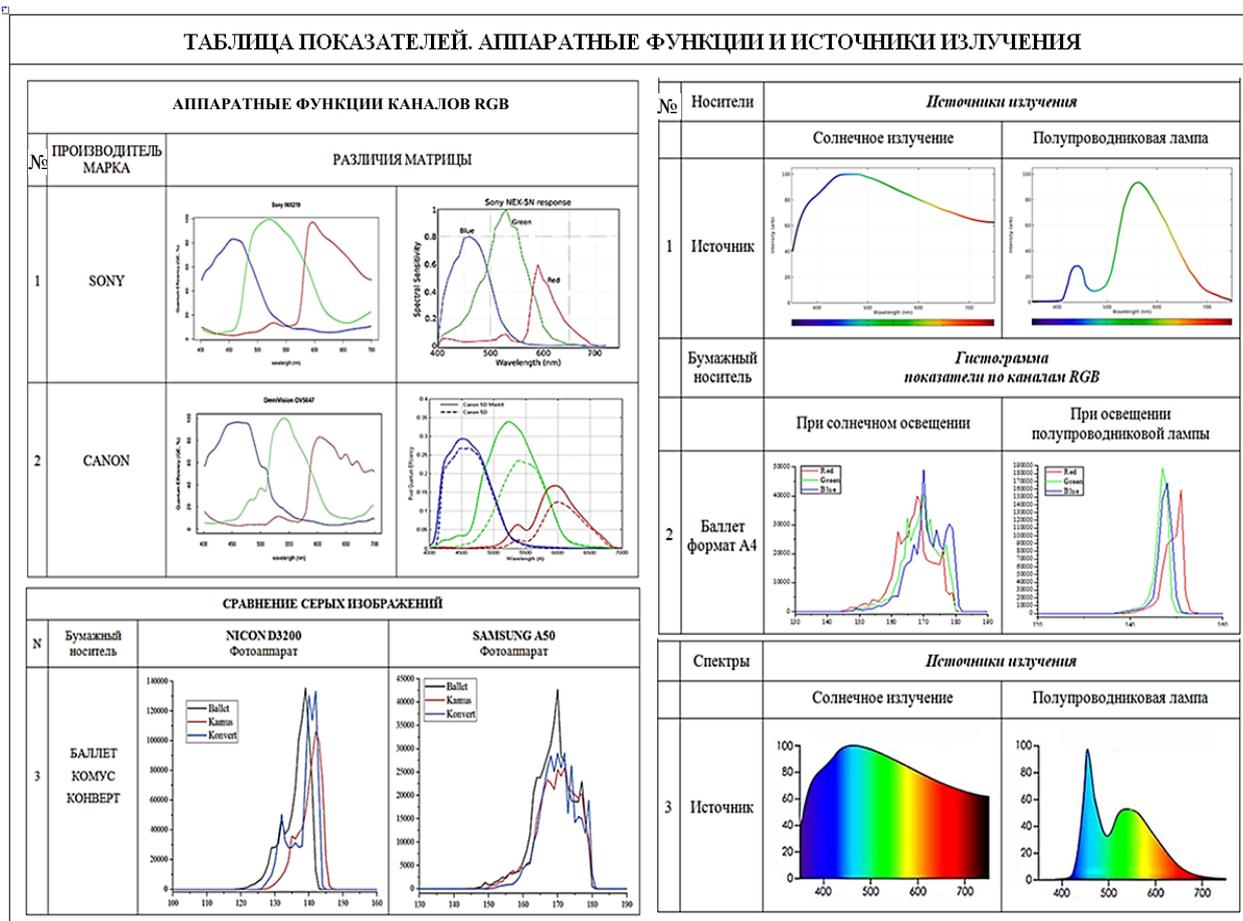


Рис. 4. Аппаратные функции каналов R, G, B разных производителей и условия фиксации и источники излучения

Fig. 4. Hardware functions of R, G, B channels from different manufacturers and fixation conditions and radiation sources

2. Данные бумаги фиксировались (фотофиксация) на различных устройствах, моделях (фотоаппарат, смартфон):

а) устройства фиксации – фотоаппарат Sony, Nikon, смартфон Samsung;

б) условия фиксации – лабораторные условия, домашние условия;

в) источники излучения – солнечный свет (фотон), туман, диодные лампы.

3. С помощью известных программных продуктов произведена обработка изображений и изучения и получена информация о тексте и цвете из изображений:

а) Basic Image Processing with FIJI / ImageJ (<https://www.jmu.edu/microscopy/resources/basic-image-processing-imagej.pdf>, <https://imagej.net/software/fiji/>);

б) приборный комплекс «видеоспектральный компаратор VC-30А» (криминалистический многофункциональный видеокomплекс) и с помощью программного обеспечения – экспертно-програм-

мный комплекс ExpertProf, производство компании «VILDIS» (<https://www.uk-nsk-lab.ru/>);

в) многофункциональный комплекс «КМК», разработанный автором с применением экспертной криминалистической программы «CrimeImage 4.0RUS» (<https://www.uk-nsk-lab.ru/>).

В табл. 2 и 3 продемонстрированы: фиксация чернила для струйного принтера и типы порошка лазерных принтеров (тонер), мастика для печати и типы чернил для шариковых ручек на бумажных носителях и их условия фиксации.

Красящие вещества «чернила и тонер для принтеров». В табл. 2 продемонстрирована фиксация красящих веществ для принтеров и их условия фиксации.

Согласно таблице, мы видим отражение бумаги и красящих веществ и можем получить истинное изображение чернил для струйного принтера и порошка лазерных принтеров (тонер), которое не зависит от внешних факторов (освещения), в том числе от устройства фиксации.

Таблица 2. Фиксация красящих веществ и условия фиксации

Table 2. Fixation of dyes – printing paste and peninks and their fixing conditions

ТАБЛИЦА ИЗМЕРЕНИЙ ПОРОШКА ЛАЗЕРНОГО И ЧЕРНИЛ СТРУЙНОГО ПРИНТЕРА					
№	Устройство фиксации	Чернила (краска) для струйного принтера		Порошок (Тонер) для лазерного принтера	
		Лабораторные условия	Домашние условия	Лабораторные условия	Домашние условия
1	Смартфон SAMSUNG				
2	Фотоаппарат SONY				

Далее, в табл. 3 с помощью известных программных продуктов произведена обработка изображений, изучение и получение информации о красящих веществах и цвете изображений.

Таблица 3. «Basic Image Processing with FIJI/ImageJ»

Table 3. Image processing and study, and obtaining (pulling out) information about dyes and color from images

ПОКАЗАТЕЛИ ИЗМЕРЕНИЙ КРАСЯЩИХ ВЕЩЕСТВ, НАХОДЯЩИХСЯ НА БУМАЖНЫХ НОСИТЕЛЯХ, И ИХ УСЛОВИЯ ПОЛУЧЕНИЯ (ФИКСАЦИЯ С ПОМОЩЬЮ ФОТОАППАРАТА «SONY»)							
№	Наименование красящих веществ на бумаге	Результаты получения измерений					
1	Краска (Струйн. принтер) (Домашние условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	12	10	0	0	9	15
	Max	157	153	178	164	151	178
	Mean	113,576	109,570	13,555	104,492	108,515	127,657
	StDev	31,574	31,532	34,411	32,755	31,918	33,973
1	Краска (Струйн. принтер) (Лабораторные условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	14	14	0	0	0	11
	Max	159	158	178	178	174	158
	Mean	118,616	116,526	118,630	118,630	113,548	116,116
	StDev	30,708	31,343	32,565	32,565	32,841	32,109
2	Тонер (лазерный принтер) (Домашние условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	1	1	0	0	0	3
	Max	169	167	178	164	166	178
	Mean	124.517	121.881	124.474	117.020	121.715	134.688
	StDev	29.667	29.780	30.642	28.562	30.289	30.268
2	Тонер (лазерный принтер) (Лабораторные условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	1	0	0	0	0	0
	Max	180	180	180	180	180	180
	Mean	124.131	123.215	124.161	122.401	122.980	127.103
	StDev	30.066	30.564	30.224	30.435	31.031	28.950

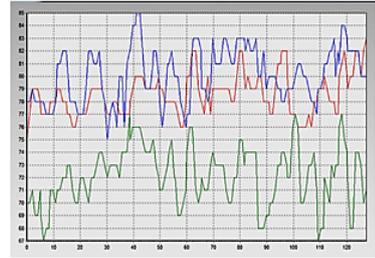
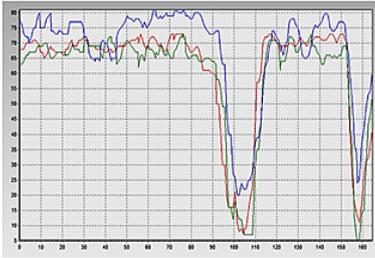
*Примечание: во избежание загруженности данной статьи нами в качестве примера будет представлено исследование одного наименования устройства фиксации (камера).

Согласно табл. 4 показатели красящих веществ кардинально отличаются друг от друга. Здесь мы можем с точностью анализировать

и идентифицировать красящие вещества, нанесенные на бумагу по изображению документа.

Таблица 4. Показатели измерений цветового профиля красящих веществ – чернила для струйного принтера и порошка лазерных принтеров (тонер)

Table 4. Dye stuff Color Profile Measurements - Inkjet Ink and Laser Powder (Toner)

ТАБЛИЦА ИЗМЕРЕНИЙ ЦВЕТОВОГО ПРОФИЛЯ			
Струйный принтер		Лазерный принтер	
Измерения	Показатели	Измерения	Показатели
	X: 242 Y: 310 P: 5,24 R: 77 G: 69 B: 78		X: 155,45 Y: 279,89 P: 36,55 R: 68 G: 69 B: 65

Нами были продемонстрированы показатели измерений красящих веществ – чернила для струйного принтера и порошка лазерных принтеров (тонер), нанесенные на бумажные носители, и их условия получения, фиксация с помощью устройств фиксации: смартфона Samsung и фотоаппарата Sony.

Красящие вещества: мастика штемпельная и чернила для ручки. В табл. 5 продемонстрирована фиксация красящих веществ – мастика для печати и чернила для ручек и их условия фиксации.

Таблица 5. Фиксация красящих веществ и условия фиксации

Table 5. Fixation of dyes and fixation conditions

ТАБЛИЦА ИЗМЕРЕНИЙ МАСТИКИ ШТЕМПЕЛЬНОЙ И ЧЕРНИЛ ШАРИКОВЫХ РУЧЕК					
№	Устройство фиксации	Мастика штемпельная краска для отпечатка печатей		Чернила (паста) для шариковых ручек	
		Лабораторные условия	Домашние условия	Лабораторные условия	Домашние условия
1	Смартфон SAMSUNG				
2	Фотоаппарат SONY				

Согласно таблице мы видим отражение красящих веществ и можем получить истинное изображение мастики штемпельной краски и чернила для ручек, которое не зависит от внешних факторов (освещения), в том числе от устройства фиксации.

Далее, в табл. 6 с помощью известных программных продуктов произведена обработка изображений и изучения и получения информации о красящих веществах и цвете изображений.

Таблица 6. «Basic Image Processing with FIJI/ImageJ»

Table 6. Image processing and study, and obtaining (pulling out) information about dyes and color from images

ПОКАЗАТЕЛИ ИЗМЕРЕНИЙ КРАСЯЩИХ ВЕЩЕСТВ, НАХОДЯЩИХСЯ НА БУМАЖНЫХ НОСИТЕЛЯХ, И ИХ УСЛОВИЯ ПОЛУЧЕНИЯ (ФИКСАЦИЯ С ПОМОЩЬЮ ФОТОАППАРАТА «SONY»)							
№	Наименование красящих веществ на бумаге	Результаты получения измерений					
1	Чернила (Шарик. ручка) (Домашние условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	32	16	6	6	6	45
	Max	168	164	194	157	164	194
	Mean	124,206	119,957	124,199	114,483	118,926	139,188
	StDev	20,368	24,691	24,872	25,978	27,302	9,419
	Чернила (Шарик. ручка) (Лабораторные условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	23	13	2	7	2	30
	Max	188	187	201	187	186	201
	Mean	127,886	125,477	127,906	124,087	124,316	135,314
	StDev	20,821	24,940	23,406	26,204	27,514	10,841
2	Мастика (оттиск печати) (Домашние условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	148	46	14	14	49	91
	Max	156	152	183	146	155	183
	Mean	124.962	121.235	124.926	112.556	121.832	140.390
	StDev	10.339	12.351	17.151	17.606	11.874	5.487
	Мастика (оттиск печати) (Лабораторные условия)	Intensity (unweighted)	Intensity (weighted)	R+G+B	RED	GREEN	BLUE
	Min	55	42	22	22	40	81
	Max	164	160	185	157	161	185
	Mean	127.962	125.509	127.972	120.665	125.621	137.640
	StDev	11.179	13.395	15.141	17.585	13.494	6.623

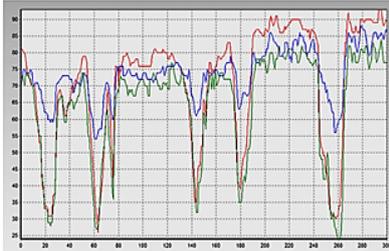
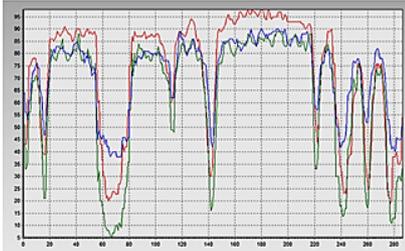
*Примечание: во избежание загруженности данной статьи [11] нами в качестве примера будет представлено исследование одного наименования устройства фиксации.

Согласно табл. 7 показатели красящих веществ кардинально отличаются друг от друга. Здесь мы можем с точностью анализировать

и идентифицировать красящие вещества, нанесенные на бумагу по изображению документа.

Таблица 7. Показатели измерений «Цветового профиля» красящих веществ – чернила для ручки и мастика штемпельная для оттиска печати

Table 7. Indicators of measurements of the "Color profile" of coloring substances - ink for a pen and stamp mastic for a seal imprint

ТАБЛИЦА ИЗМЕРЕНИЙ ЦВЕТОВОГО ПРОФИЛЯ			
Мастика для Оттиска печати		Чернила для Шариковой ручки	
Измерения	Показатели	Измерения	Показатели
	<p>X: 206,1 Y: 491,63 P: 31,95</p> <p>R: 69 G: 66 B: 72</p>		<p>X: 380,89 Y: 349,79 P: 51,79</p> <p>R: 88 G: 73 B: 71</p>

Нами были продемонстрированы показатели измерений красящих веществ – мастика штемпельная для оттиска печати и чернила (паста) для ручки, нанесенные на бумажные носители, и условия получения их фиксации с помощью устройств фиксации: смартфона Samsung и фотоаппарата Sony.

Алгоритм получения информации в разработке метода. Схема проведения экспериментального исследования – для понимания методики по исследованию [12] бумажных носителей, которая является частью основной

методики «Цифровой двойник изображений документа» (ЦДИД).

При изучении нами цифровой (или иной) формы документа возникают результаты измерений в виде изображений, которые получаются при получении изображений в разных спектральных диапазонах. Здесь не надо привлекать технические способы, т. е. привлечение светофильтров, так как это деталь мелкая, а важен принцип получения изображения и спектральный диапазон.

На рис. 5 для понимания продемонстрирован алгоритм получения информации.

АЛГОРИТМ ИНФОРМАЦИИ В РАЗРАБОТКЕ МЕТОДА

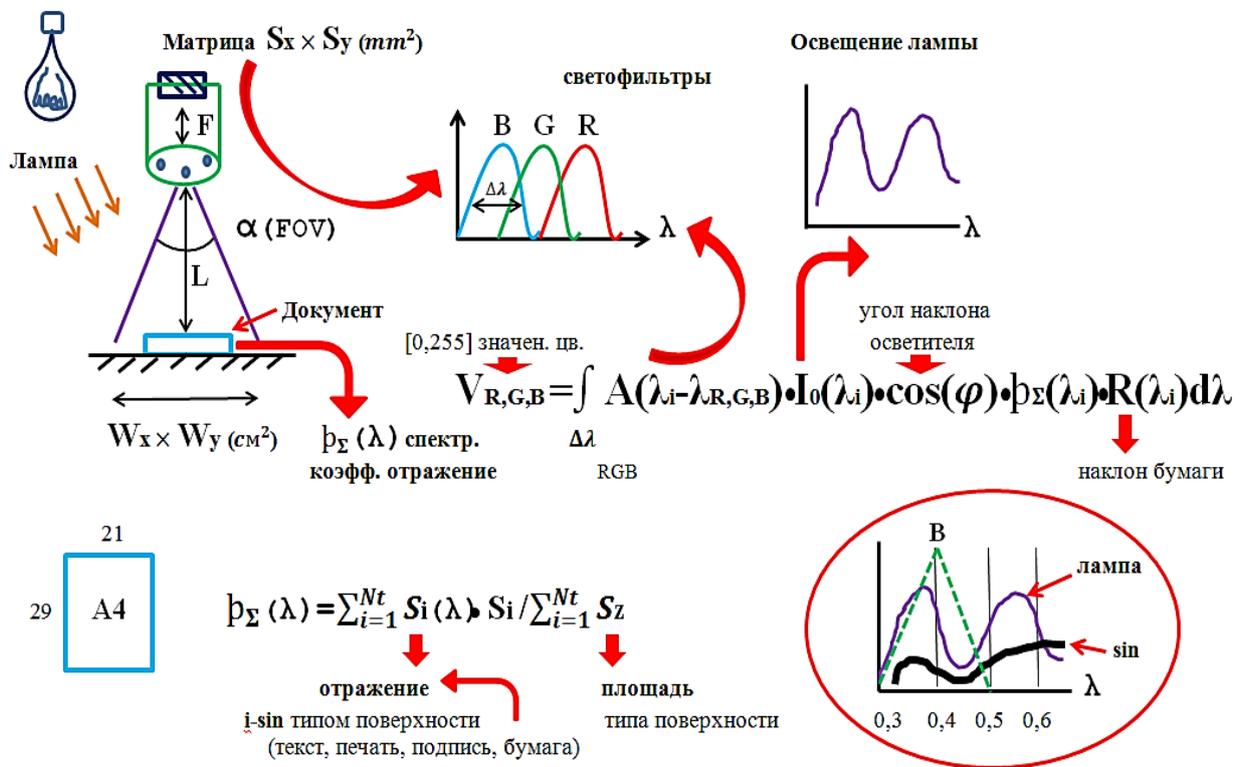


Рис. 5. Алгоритм получения информации в разработке метода

Fig. 5. Algorithm for obtaining information in the development of the method

В рамках исследования цифровой (или иной) формы документа получается набор изображений документа в разных спектральных диапазонах. Итак, на этом пути исследования электронно-цифрового текстового документа (ЦФ – цифровая форма) получаются цифровые изображения, которые нами предлагается называть ЦДД – цифровой двойник документа [13].

Цифровой двойник – это набор данных, связанный с получением растровых (пиксельных) изображений бумажного, цифрового (или иного) варианта текстового документа в различных

спектральных диапазонах. Обычное RGB-изображение является частью ЦДД, связанного с видимым диапазоном спектра.

Алгоритм исследования и анализа (кратко). Как ранее было сказано, каждое изображение документа может рассматриваться как источник сравнения с изображениями, которые являются шаблонами с известными типами бумаг, типами источников изображений.

На рис. 6 представлен алгоритм исследования и анализ.

Допустим, при помощи фиксирующего устройства была произведена фотофиксация бу-

мажного документа. Данный документ подлежит исследованию:

1. Цифровое изображение бумаг передается в цифровой форме в многофункциональное устройство «КМК» (разработанное автором). Идет обработка, анализ и идентификация изображения документа. Далее передается изображение (см. рис. 6):

- а) видеоспектральному компаратору;
- б) Basic Image Processing with FIJI/ImageJ или по очереди, или одновременно пункту «а» и «б».

2. На видеоспектральном компараторе производится анализ, получаем гистограммы и цветовой профиль – показатели измерений цветовой профили красящих веществ (см. рис. 6).

3. С помощью известных программных продуктов «Basic Image Processing with FIJI/ImageJ» производится обработка изображений и изучения [14] и по каналам «RGB» получаем информацию по красящим веществам, также о бумаге и цвете из изображений (см. рис. 6).

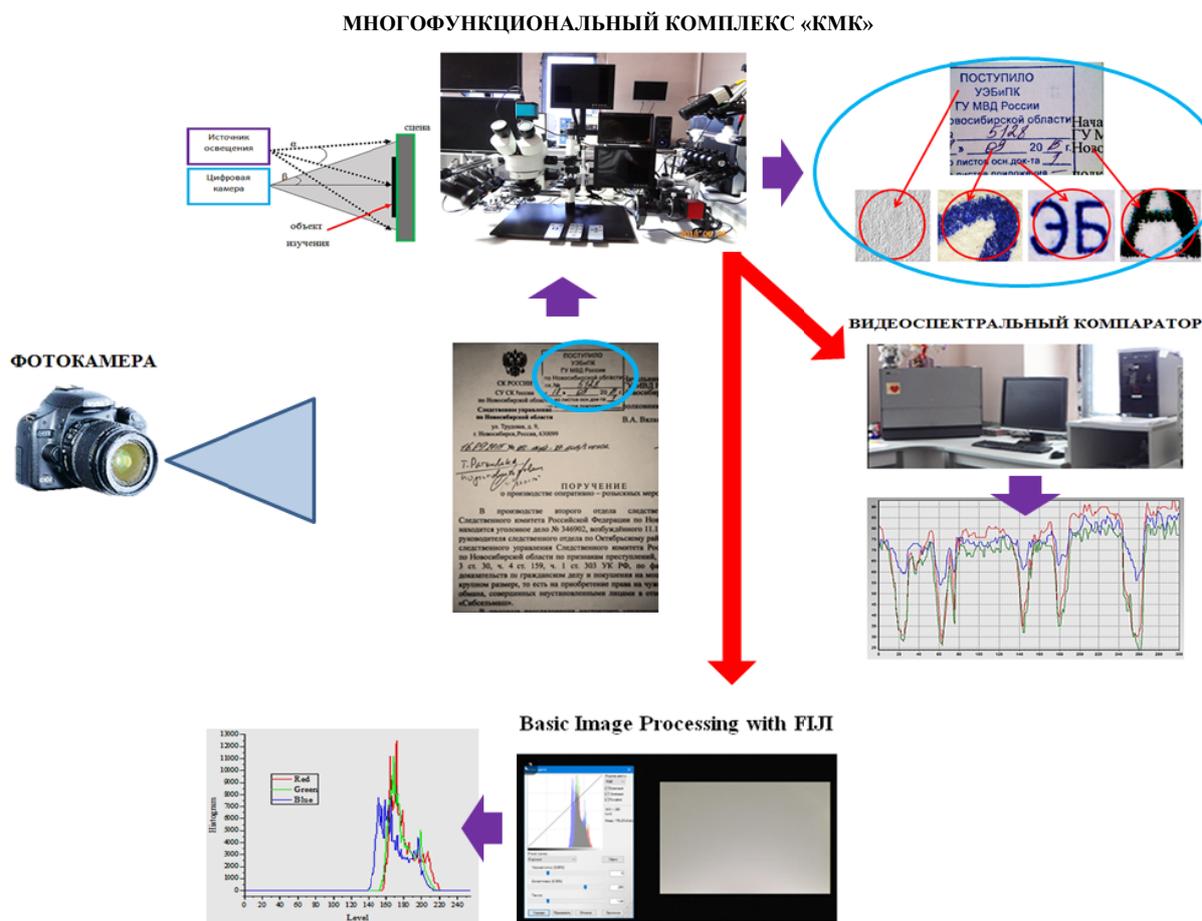


Рис. 6. Алгоритм производства исследования и анализ

Fig. 6. Algorithm production research and analysis

Изображения документа (бумажного носителя) могут быть разделены на несколько зон, каждая из которых исследуется отдельно или по очереди [15], также можно исследовать одновременно на нескольких мониторах (онлайн), включая и в зонах спектра на многофункциональном комплексе «КМК» (см. рис. 6). Каждой разбитой зоне дается своя оценка.

Одна из зон выделена указателем круглой формы синего цвета (см. рис. 6).

Итак, полученные результаты (показателей измерений) при исследовании подтверждают, что красящие вещества (чернила, мастика, тонер и т. д.) отличаются друг от друга показателями гистограмм и цветовой профили (см. табл. 5 и рис. 6).

Выводы по исследованию

Применяя данный метод исследования, мы можем распознать и идентифицировать по изображению красящие вещества, нанесенные на

бумажные носители, цвету изображений и условиям их получения, а также получить истинное изображение красящих веществ (чернил, пасты и т. д.), которое не зависит от внешних факторов (освещения), в том числе от устройства фиксации.

Также есть возможность распознать замену (подделку) любого листа документа из изображения и установить марку бумаги, на которой был изготовлен документ, и на какой бумаге была изготовлена замененная бумага, то есть замененная текстовая информация или реквизиты документа.

Заключение

Вышеизложенный материал показывает необходимость проведения научного исследования, связанного с развитием базы технико-криминалистического обеспечения проведения экспертизы документов, находящихся в цифровой форме (изображении). В соответствии с этим очевидна актуальность темы данной работы.

При этом наблюдается высокая востребованность экспертизы именно цифровой формы документа в виде изображения, что и определяет теоретическое и практическое направление данной работы.

Надо отметить, что на сегодняшний день нет проработанной методики по выявлению и распознаванию подделки документа с электронно-цифровой формы.

Использованные обозначения:

ТЭД – техническая экспертиза документов;
красящие вещества – чернила (для струйных принтеров), мастика (штемпельная краска), тонер (порошок для лазерных принтеров), паста для пишущих приборов (шариковая ручка и т. д.);

ЦФ – цифровая форма;

КМК – криминалистический многофункциональный комплекс;

ЦДД – цифровой двойник документа;

ТИ – текстовая информация;

ЦАЦД – цветовой аналог цифрового двойника;

ЦД – цифровой двойник;

ЦДИД – цифровой двойник изображений документа.

Библиографические ссылки

1. Янг М. Оптика и лазеры, включая волоконную оптику и оптические волноводы. М.: Мир, 2018. 544 с.
2. Фоторефрактивные эффекты в электрооптических кристаллах : учеб. пособие / В. М. Шандаров, А. Е. Мандель, С. М. Шандаров, Н. И. Буримов. Томск : ТУСУР, 2012. 244 с.
3. Шандаров В. М. Волоконно-оптические устройства технологического назначения : учеб. пособие. Томск : ТУСУР, 2012. 198 с.

4. Allaberganov A. A. and Kataev M. Yu. Methodology for obtaining textual information from images and its analysis // Scopus. 2020. P. 1-6.

5. Серебренников Л. Я., Шандаров С. М., Буримов Н. И. Акустоэлектронные приборы и устройства : учеб. пособие. Томск : ТУСУР, 2012. 70 с.

6. Аллаберганов А. А. Комплексные судебные экспертизы документов, изготовленных путем монтажа с использованием цифровых технологий (юридические науки). Использование оптических устройств для распознавания и анализа текстовой информации, представленной в виде изображений (технические науки) : монография. Томск : Томск. гос. ун-т систем упр. и радиоэлектроники ; Москва : Знание-М, 2022. 118 с.

7. Аллаберганов А. А., Катаев М. Ю. Оптоэлектроника как основа метода распознавания и идентификации лица путем декомпозиции и распознавания, идентификация лиц с дальней дистанции в агрессивных условиях (с получением цифрового двойника). Судебная практика : монография. Томск : Томск. гос. ун-т систем упр. и радиоэлектроники ; Москва : Знание-М, 2022. 120 с.

8. Акустооптические лазерные системы формирования телевизионных изображений / Ю. В. Гуляев, М. А. Казарян, Ю. М. Мокрушин, О. Шакин. М. : Физматлит, 2016. 240 с.

9. Ахманов С. А., Никитин С. Ю. Физическая оптика. М. : Издательство МГУ, Наука, 2018. 654 с.

10. Дифракционная компьютерная оптика / под ред. В. А. Сойфера. М. : Наука, 2017. 736 с.

11. Закаэнов Н. П., Киришин С. И., Кузичев В. И. Теория оптических систем. М. : Лань, 2016. 448 с.

12. Пухтин А. Физические основы квантовой электроники и оптоэлектроники. М. : Высш. шк., 2018. 304 с.

13. Аллаберганов А. А., Катаев М. Ю. Метод получения цифрового двойника бумажного носителя // Право и практика. 2020. № 1. С. 80–85.

14. Пухтин А. Н. Квантовая и оптическая электроника : учебник для вузов. М. : Абрис, 2012. 656 с.

15. Шандаров В. М. Волоконно-оптические устройства технологического назначения : учеб. пособие. Томск : ТУСУР, 2012. 198 с.

References

1. Jang M. *Optika i lazery, vključajaja volokonnuju optiku i optičeskie volnovody* [Optics and lasers, including fiber optics and optical waveguides]. Moscow, Mir Publ. 2018. 544 p. (in Russ.).
2. Shandarov V.M., Mandel A.E., Shandarov S.M., Burimov N.I. *Fotorefraktivnye jeffekty v jelektrkoop-tičeskih kristallah* [Photorefractive effects in electro-optical crystals]. Tomsk: TUSUR, 2012. 244 p. (in Russ.).
3. Shandarov V.M. *Volokonno-optičeskie ustrojstva tehnologičeskogo naznachenija* [Fiber-optic devices for technological purposes: Textbook]. Tomsk: TUSUR, 2012. 198 p. (in Russ.).

4. Allaberganov A.A. and Kataev M.Yu. Methodology for obtaining textual information from images and its analysis. Scopus. 2020. P. 1-6.

5. Serebrennikov L.Ya., Shandarov S.M., Burimov N.I. *Akustoelektronnye pribory i ustrojstva* [Acousto-electronic devices and devices: Textbook]. Tomsk: TUSUR, 2012. 70 p. (in Russ.).

6. Allaberganov A. A. *Kompleksnye sudebnye jekspertizy dokumentov, izgotovlennyh putem montazha s ispol'zovaniem cifrovyyh tehnologij (juridicheskie nauki). Ispol'zovanie opticheskikh ustroystv dlja raspoznanija i analiza tekstovoj informacii, predstavlennoj v vide izobrazhenij (tehnicheckie nauki)* [Complex forensic examination of documents made by editing using digital technologies (legal sciences). The use of optical devices for recognition and analysis of textual information presented in the form of images (technical sciences): monograph]. Tomsk: Tomsk. state un-t control systems. and radio electronics; Moscow: Znanie-M, 2022. 118 p. (in Russ.).

7. Allaberganov A.A. Kataev M. Yu. *Optoelektronika kak osnova metoda raspoznanija i identifikacii lica putem dekompozicii i raspoznanija, identifikacija lica s dal'nej distancii v agressivnyh uslovijah (s polucheniem cifrovogo dvojnika). Sudebnaja praktika* [Optoelectronics as the basis of the method of face recognition and identification by decomposition and recognition, identification of faces from a long distance in aggressive conditions (with obtaining a digital twin). Judicial practice: monograph]. Tomsk: Tomsk. state un-t control systems. and radio electronics; Moscow: Znanie-M, 2022. 120 p. (in Russ.).

8. Guljaev Ju.V., Kazarjan M.A., Mokrushin Ju.M., Shakin O.M. *Akustoopticheskie lazernye sistemy formirovanija televizionnyh izobrazhenij* [Acousto-optic laser systems for the formation of television images]. Moscow, Fizmatlit Publ., 2016. 240 p. (in Russ.).

9. Ahmanov S.A., Nikitin S.Ju. *Fizicheskaja optika* [Physical optics]. Moscow : Izdatel'stvo MGU, Nauka, 2018. 654 p. (in Russ.).

10. Difrakcionnaya komp'yuternaya optika [Diffractive computer optics]. (eds. V. A. Soifer). Moscow : Nauka Publ., 2017. 736 p. (in Russ.).

11. Zakaznov N.P., Kiryushin S.I., Kuzichev V.I. *Teoriya opticheskikh sistem* [Theory of optical systems]. Moscow: Lan' Publ., 2016. 448 p. (in Russ.).

12. Pikhtin A. *Fizicheskie osnovy kvantovoi elektroniki i optoelektroniki* [Physical foundations of quantum electronics and optoelectronics]. Moscow : Vysshaya shkola Publ., 2018. 304 p. (in Russ.).

13. Allaberganov A.A. Kataev M.Yu. [Method for obtaining a digital twin of a paper carrier]. *Pravo i praktika*. 2020. No. 1. Pp. 80-85 (in Russ.).

14. A.N. Pikhtin. *Kvantovaya i opticheskaya elektronika : uchebnyk dlya vuzov* [Quantum and optical electronics: a textbook for universities]. Moscow: Abris Publ., 2012. 656 p. (in Russ.).

15. Shandarov V.M. *Volokonno-opticheskie ustrojstva tekhnologicheskogo naznacheniya : ucheb. posobie* [Fiber-optic devices for technological purposes: Textbook]. Tomsk: TUSUR, 2012. 198 p. (in Russ.).

Method of Digitally Presented Texts Extraction from the Image of a Document and their Examination

A. A. Allaberganov, Post-graduate, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia, Judicial Expert Criminalistic Laboratory at the management company "NSK-Capital", Novosibirsk, Russia

M. Yu. Kataev, DSc in Engineering, Professor, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia.

In this work, the problems of extracting textual information from a document image are solved. It is important to note that the work includes complex use of special knowledge in three areas: technical and forensic examination of documents; phototechnical expertise and computer-technical expertise. The article considers and demonstrates in detail the information for detecting changes in textual information (ink, toner and writing instrument pastes) from digital images.

The commission of crimes in the economic sphere, as a rule, is associated with the falsification (forgery) of documents, and the increase in the document flow in the regulation of legal relations, which has been noticeable in recent years, has led to a significant increase in the number of forged documents. This is confirmed by statistical data on forensic handwriting and technical forensic examinations. Since the document is the basis of all human activity, and it contains data drawn up in a certain order and having legal significance.

The problem has got many interesting solutions thanks to practical applications such as text recognition, translation, forensics and etc. Despite the fact that this problem has been dealt with for a long time, it is still missing, i.e. there is no well-developed methodology for image recognition of documents.

The image of text information depends on the type of scene, type of digital camera, amount and direction of illumination, type of substrate and etc. It is important to note here that when obtaining information (from the document under study) in modern conditions, it is possible to fix the document using any device (camera, smartphone, scanner and etc.) and in any conditions (day/night, street, radiation source) that are unknown in advance. Proceeding from this,

and this is a problem in modern conditions, in the study of electronic digital information of text documents. In this work, a study to solve this problem was carried out.

Keywords: document, digital images, authenticity, paper, document details, brightness histograms, information extracting.

Получено: 20.04.23

Образец цитирования

Аллаберганов А. А., Катаев М. Ю. Методика извлечения цифровых представленных текстов по изображению документа и их экспертиза // Интеллектуальные системы в производстве. 2023. Т. 21, № 2. С. 27–40. DOI: 10.22213/2410-9304-2023-2-27-40.

For Citation

Allaberganov A.A., Kataev M.Yu. [Method of digitally presented texts extraction from the image of a document and their examination]. *Intellektual'nye sistemy v proizvodstve*. 2023, vol. 21, no. 2, pp. 27-40 (in Russ.). DOI: 10.22213/2410-9304-2023-2-27-40.