

Секция 11

ОПТОЭЛЕКТРОНИКА И ФОТОНИКА

Председатель секции – *Задорин Анатолий Семёнович, д.ф.-м.н., профессор каф. РЗИ*

УДК 004.932.2

А.А. Аллаберганов, М.Ю. Катаев

Многофункциональный исследовательский комплекс решения задач анализа текстовой информации

Приводится описание приборного комплекса, позволяющего получать изображения текста, обрабатывать их и анализировать. Целью работы является автоматизация процессов измерения, анализа и сопоставления графической текстовой информации для решения задач экспертизы в исследовательской криминалистической деятельности. Это позволит обеспечить данное направление новым видом решений и качеством информационных услуг. Представлены описание измерительной установки и основные ее показатели. Показан пример решения криминалистической экспертизы при анализе чековой информации.

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

Входной информацией в системах электронного документооборота, экспертизы, контроля и других приложений могут быть не только документы с печатным текстом, но и рукописные документы. Исторически большое количество рукописных документов создавалось в научной сфере, поэтому ее необходимо перевести в электронный вид, чтобы не потерять знания. Несмотря на то, что задача получения изображений и распознавания текста (РТ) известна с начала 70-х годов, до сих пор имеются как теоретические, так и практические проблемы, связанные с огромным многообразием языков и типов написания символов и текста. На сегодняшний день известно множество методик и алгоритмов, решающих задачу РТ достаточно быстро и эффективно (точность распознавания текста выше 90%). Для многих из известных методик входными данными являются изображения, полученные со сканера, цифрового фотоаппарата, спектрометра или других цифровых устройств. Таким образом, задача получения изображения текста, обработки изображений и распознавания (машинописного или рукописного) текста является актуальной и востребованной в различных сферах практической деятельности.

Предметом исследования в данной работе является криминалистический анализ документов, биологических объектов, предметов археологии или искусства и т.д. Цель исследований связана с повышением точности идентификации изучаемых объектов. Для достижения указанной цели необходима разработка новых методик, способов исследования, повышающих качество анализа.

Текстовая информация имеет важные отличительные особенности, так как состоит из нескольких

частей: текстовой информации (характеристики – шрифт, чернила) и подложки (бумага или другой носитель). Изображение текстовой информации может быть монохромным (бинарным, однобитовым) или цветным (RGB), или полученным в ином участке спектра (например, инфракрасном). Эти особенности позволяют выделять на изображении текстовую информацию при помощи соответствующих математических алгоритмов. Выделение шрифта, типа чернил связано с задачей определения лица, написавшего текст (напечатанного), временем печати, местом печати и др.

Краткое описание установки

Криминалистический многофункциональный комплекс (КМК) представлен на рис. 1.

При этом 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.

Исследовательская установка, разборно-сборная с возможностью переносного действия, содержит силовой каркас с изменениями своего положения порога и установленные на силовом каркасе: 1 – цифровой промышленный видеомикроскоп-камера; 2 – видеоинспекционный промышленный микроскоп-камера; 3 – микроскоп-камера предварительно просмотра с переменным фокусным расстоянием; 4 – микроскоп-эндоскоп-камера; 5 – ультрафиолето-

вый 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 – косопадающие световые фильтры с гаммой разных цветов.

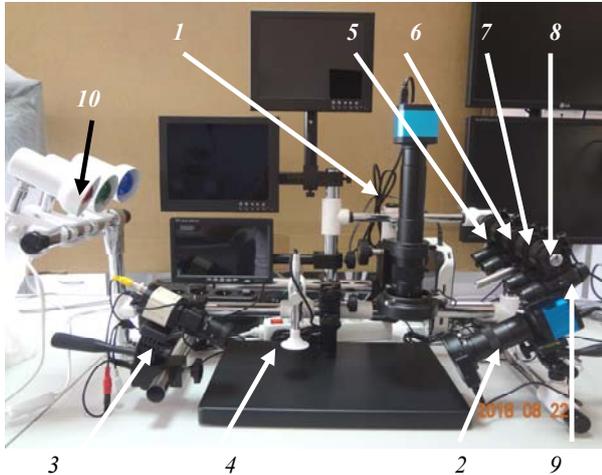


Рис. 1. Криминалистический многофункциональный видеомикроскопический спектральный комплекс

Приборы в комплексе (КМК) работают, как независимо друг от друга, так и в комплексе как единая система. Приборы работают в области видимого спектра, фотографируя и выполняя увеличение, в нужный момент можно при видеозахвате перенести изображение на любой монитор и сделать наложение на другой объект исследования, при этом настраивать прозрачность объектов сравнения. Криминалистический многофункциональный комплекс разбирается полностью в течение 3 мин и собирается в течение 10 мин (масса до 12 кг).

Полученные результаты

В результате примера приведем исследование очередности нанесения реквизитов (рис. 2).

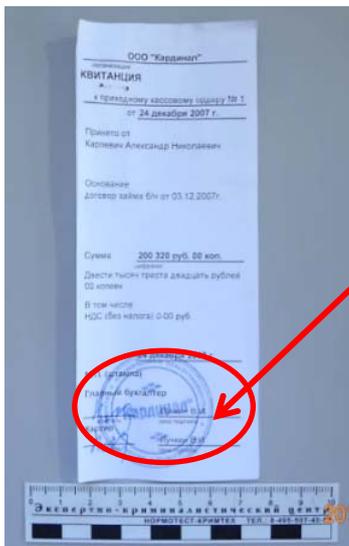


Рис. 2. Пример изображения чековой информации, имеющей плохое качество печати

Для этого выбран чек, на котором есть печать и подписи. Какая из подписей сделана первой и когда нанесена печать, выясняется с помощью обсуждаемого КМК. Одной из форм нанесения информации на бумагу является использование того или иного красящего вещества, и каждое из них обладает свойствами, которые помогут в данном исследовании. Другой формой является нажатие ручкой на бумагу, при написании, что оставляет след (глубина и ширина), который можно зафиксировать.

Заключение

Многофункциональный исследовательский комплекс проводит исследование одновременно в трех и более ракурсах 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.

Литература

1. Фомин Я.А. Распознавание образов: теория и практика. – 3-е изд., доп. – М.: ФАЗИС. – 2014. – 460 с.
2. Журавлев Ю.И. Распознавание. Математические методы. Программная система. Практические применения / Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько. – М.: ФАЗИС, 2006. – 176 с.
3. Местецкий Л.М. Математические методы распознавания образов. – М.: МГУ, ВМиК, 2002. – 85 с.
4. Фаворская М.Н. Модель распознавания изображений рукописного текста / М.Н. Фаворская, А.Н. Горюшкин // Вестник Сиб. гос. аэрокосмического ун-та им. акад. М.Ф. Решетнева. – 2008. – № 2 (19). – С. 52–58.
5. Фан Н.Х. Распознавание печатных текстов на основе применения вейвлет-преобразования и метода главных компонент / Н.Х. Фан, Т.Ч. Буй, В.Г. Спицын // Изв. Том. политехн. ун-та. – 2012. – Т. 321, № 5. – С. 154–158.
6. Андреев С.В., Бондаренко А.В., Горемычкин В.И., Ермаков А.В., Желтов С.Ю. Исследование подходов к построению систем автоматического считывания символьной информации. Препринт ИПМ им. М.В. Келдыша, РАН, № 44. – 2003. – 18 с.
7. Madhvanath S. The Role of Holistic Paradigms in Handwritten Word Recognition. / S. Madhvanath, V. Govindaraju // Trans. on Pattern Analysis and Machine Intelligence. – 2001. – № 23(2). – P. 149–164.

Аллаберганов Ахмеджан Атаханович
Аспирант каф. ССУ ТУСУРа
Эл. почта: nsk-kapital@mail.ru

Катаев Михаил Юрьевич
Д-р техн. наук, профессор каф. АСУ ТУСУРа
Эл. почта: kmy@asu.tusur.ru