

Роль алгоритмов распознавания изображений для реставрации исторических документов

В докладе рассматриваются вопросы разработки методики и экспериментального программного обеспечения для оцифровки и реставрации электронных копий реальных бумажных исторических документов. Основным отличием исторического документа является отсутствие цифрового оригинала, то есть исходный документ создавался без участия вычислительной техники. В процессе хранения в документе как носителе информации появляются различные дефекты, затрудняющие его чтение. Кроме того, хрупкость бумаги как носителя приводит к ограничению доступа пользователей к таким документам. Поэтому перед автором доклада встала задача разработки методики восстановления текстовых символов на оцифрованных исторических документах с использованием методов распознавания и сегментации изображений.

В докладе описываются:

1. Метод распознавания по эталонам старопечатных символов на изображениях исторических документов[1];
2. Возможности применения алгоритмов сегментации и кластеризации для целей локализации штрихов текста[2];
3. Методика реставрации потерянных символов в исторических старопечатных текстовых документах.

Методы сегментации применяются в обработке изображений архивных документов для решения следующих основных задач:

- отделение на текстовом (рукописном) документе текста от фона.
- определение областей изображения текстовых документов, полностью или частично испорченных дефектами.

В общем случае кластеризация информативных участков изображения представляет собой последовательное обнаружение таких элементов, как штрихи текста, владельческие, дарственные надписи, элементы графического оформления текстовых документов и другие.

Задача данной разработки – добавить в существующую автоматизированную систему реставрации изображений текстовых документов [2] модуль распознавания текста.

Алгоритмы распознавания позволяют идентифицировать каждую букву текста. Помимо этого автором доклада ведутся работы по применению алгоритмов распознавания для целей восстановления поврежденных старопечатных букв.

Литература

1. Е.Е. Канунова, Е.В. Полякова Особенности распознавания изображений старопечатных текстовых символов// Алгоритмы, методы и системы обработки данных– 2009, №14, С. 46- 50.
2. С.С. Садыков, Е.Е. Канунова, А.Д. Варламов Автоматизированная реставрация изображений архивных текстовых и фотографических документов// Автоматизация. Современные технологии – 2007, №8, С. 10-15