

Разработка системы анализа данных бланков тестирования

Ковалева И.Л., Савчук Н.А.

Белорусский национальный технический университет

Одним из методов контроля успеваемости и качества образования абитуриентов и студентов в Белорусском национальном техническом университете является диагностическое тестирование. Участники тестирования раз в квартал выполняют предложенное им задание, результаты которого заносят в специально разработанные бланки тестирования. Проверка результатов осуществляется в ходе автоматической обработки бланков. Однако качество автоматического анализа бланков при использовании базовой системы не всегда соответствует желаемому уровню. Поэтому было принято решение о выполнении параллельного анализа бланков и разработке для этого дополнительной системы, реализующей иные алгоритмы предварительной обработки изображений бланков тестирования и их распознавания.

На этапе предварительной обработки в системе анализа данных бланков тестирования выполняются фильтрация сканированного изображения бланка, повышение его контрастности, бинаризация, поворот и сегментация характерных зон. Затем в каждой из сегментированных зон выделяются отдельные символы. Положение символов на бланке строго зонировано и заранее известно. Каждая зона однозначно определяет возможный набор расположенных в ней символов: характерные символы (крестики), цифры, буквы, комбинация цифр и букв.

Затем выполняется распознавание сегментированных символов. Для повышения качества распознавания в системе анализа данных бланков тестирования предусмотрено использование двух систем распознавания с учителем. Набор информативных признаков для этих систем распознавания был сформирован на основании того факта, что все символы заносятся в бланк ответов в соответствии с определенными правилами, а их написание максимально приближено к их печатному представлению. При этом на каждом бланке приведен пример образцового написания соответствующего символа. Все это позволило назвать подобные рукописные символы «рукопечатными» и использовать для их распознавания информативные признаки, полученные с помощью двух алгоритмов. В первом алгоритме определяются и затем анализируются характеристические точки утоньшенного изображения распознаваемого символа, во втором алгоритме используются зонды. При обучении систем распознавания использовались как образцовые символы, так и реальные символы, полученные с бланков тестирования.