

ドキュメントスキャナを用いた半自動採点システムの構築

茂木 良平[†]大枝 真一[‡]木更津工業高等専門学校 情報工学科本科[†]木更津工業高等専門学校 情報工学科[‡]

1. まえがき

試験は、学生の授業に対する理解度を知るために教育現場では必要不可欠なものである。当然、試験を行えば出題者は採点を行わなければならない。その際、解答用紙がマークシート方式ならば機械によって自動で採点されるため効率良く採点する事ができるが、試験ではマークシート方式だけで学生の問題に対する思考過程を詳細に把握することは難しい。そのため自由記述の問題が必要となってくる。しかし自由記述方式の用紙は採点する際に機械で自動採点することができないため、人の手によって一枚一枚めくりながら採点しなければならない。また、点数の再確認や統計データの算出する際に、何度もめくり直すため効率が良いとはいえない。そこで本研究では自由記述方式の解答用紙の採点効率を向上させるための半自動採点システムを構築する。本システムを利用することにより、採点の効率化や統計データも従来の方法より容易に算出することが可能になると予想される。

2. 半自動採点システム

2.1 システムの概要

本研究において作成するシステムは、以下の流れで構成されている。

1. 採点する解答用紙と記述していない解答用紙をドキュメントスキャナを用いてパソコンに取り込む。その際、解答用紙が学籍番号順に並んでいなくてもよい。
2. 取り込んだ用紙の学籍番号の部分の方向線素特徴量を抽出し、識別器を用いて認識し、学籍番号順に自動ソートする。
3. 採点システムソフトウェアに切り抜く対象となる領域を選択するモードを用意し、切り抜く範囲をマウスでドラッグし、始点と終点の座標を得る(図1)。

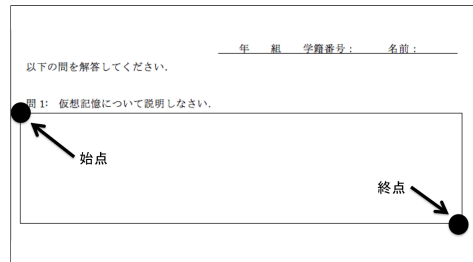


図1 始点と終点を取得

4. 始点と終点を利用して解答欄を抽出する。保存する際には問題毎にフォルダを変えて保存していく。
5. 解答を表示し、採点していく(図2)。

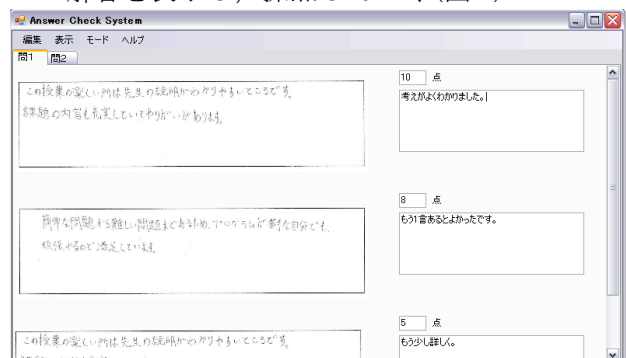


図2 採点システム

本システムでは採点した結果が最終的に CSV ファイルで出力されるため、一般的な表計算ソフトで編集可能となる。

2.2 方向線素特徴量

自動ソートする際に文字認識を行い学籍番号の識別を行う。文字認識を行う際にこの方向線素特徴量が必要となる。方向線素特徴量の算出方法は図3に示す。64×64ドットからなる文字画像領域を8×8ドット単位に分割する。隣接する4単位をまとめて16×16ドットの小領域とし、縦方向と横方向にそれぞれ半分ずつをオーバーラップさせてラスタ走査していく。小領域は全部で7×7の49個得られる。各小領域を図3の下側に示す4つの領域に分割し、各領域の重みを内側から4, 3, 2, 1とする。

各小領域の方向線素特徴量を4次元ベクトル

A Construction of semiautomatic grading system with document scanner

[†]Ryohei Mogi, Dept of Information and Computer Engineering course Kisarazu National College of Technology

[‡]Shinichi Oeda, Dept of Information and Computer Engineering course Kisarazu National College of Technology

(x_1, x_2, x_3, x_4) で定義する。ただし、
 $x_i = 4x_{1i} + 3x_{2i} + 2x_{3i} + 1x_{4i} \quad (i=1, 2, 3, 4)$
 である。

また、添字 i がとる値 1, 2, 3, 4 はそれぞれ、縦、横、+45 度、-45 度の 4 種類の方向線素を意味し、 $x_{1i}, x_{2i}, x_{3i}, x_{4i}$ はそれぞれ中心から外側に向けて 4 つの各部分領域での方向線素 i の個数を表す。

したがって 1 文字の方向線素特徴量は、49 個の小領域の方向線素特徴量を並べたもので、次元数は 196 (49×4) となる [1]。

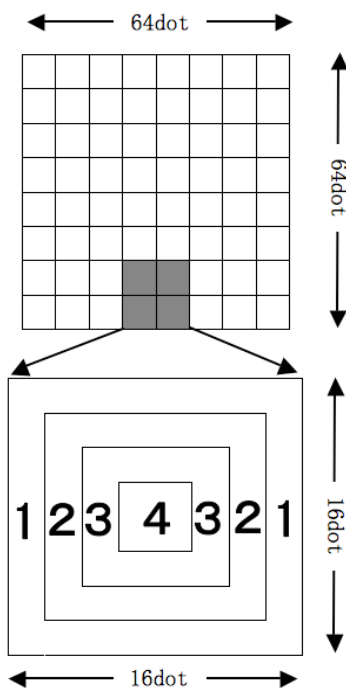


図3 方向線素特徴量

3. 初期実験

3.1 領域抽出

画像を抜き出すためのプログラムを作成した。作成した解答用紙を図 4 に、出力結果を図 5 に示す。抽出された画像は問題毎の出力フォルダに保存する。

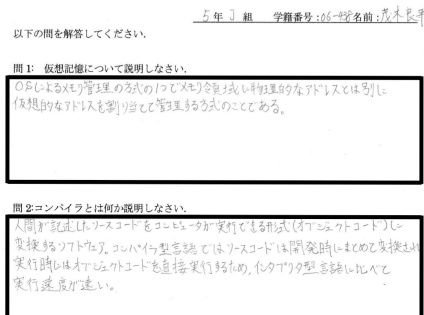


図4 解答用紙

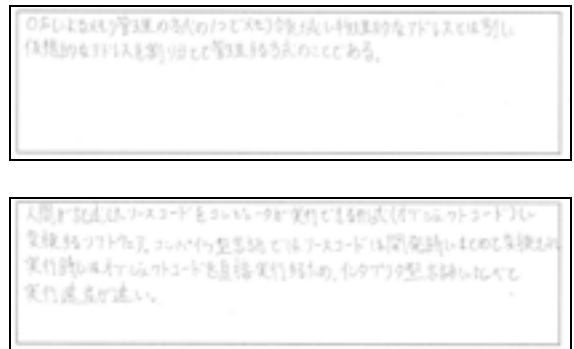


図5 出力結果

3.2 方向線素特徴量抽出

実際に数字の画像に対して方向線素特徴量の抽出を行った。特徴量抽出を行う画像に対して前処理として 2 値化し、さらに細線化を施す。その画像を図 6 に示し、抽出した方向線素特徴量は図 7, 8 に示す。なおグラフの横軸は次元、縦軸は特徴量となっている。

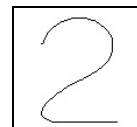


図6 細線化, 2 値化した数字 2

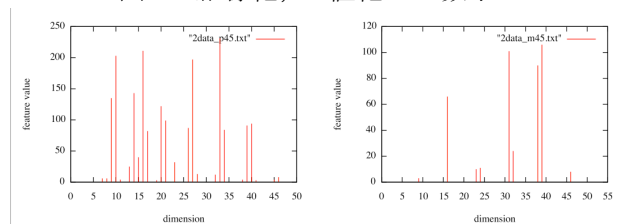


図7 横方向, 縦方向の方向線素特徴量

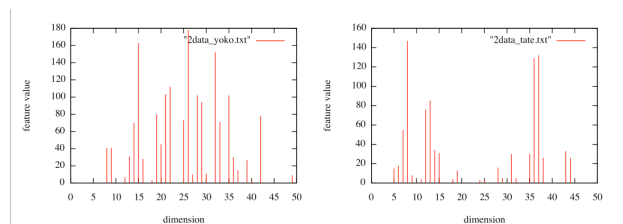


図8 +45 度方向, -45 度方向の方向線素特徴量

4. まとめ

本研究では、自由記述方式の解答用紙に対して、採点の補助を行う半自動採点システムを構築、学籍番号順に自動ソートするために必要な方向線素特徴量の抽出を行った。

参考文献

[1] 山田 奨治, 和泉 勇治, 加藤 寧, 柴山 守, “類似文字検索機能をそなえた電子くずし字辞典の開発”, 情報処理学会研究報告, 2002(52), pp43-50, 2002.