CHAPTER

5

GENERIC OPTICAL ANSWER EXTRACTION FOR MULTIPLE-CHOICE FORMS

Aliyu Muhammad and Musa Mohd Mokji

5.1 INTRODUCTION

The concept of automatic Multiple-Choice (MC) answer extraction started with a machine; a product called Optical Mark Reader (OMR). It uses a dedicated scanner which allows the machine to scan through the specially designed form (the sheet trans-optic in nature) by illuminating light beams onto its surface. The surface's reflectivity determines the position of a tick mark or label in the form of digital signals. It was costly and uneconomical to utilize, not to mention huge. Nowadays, MC forms have become dominant, reliable, and one of the easiest ways of assessing and acquiring information used in academia, survey, elections, and form applications, to mention a few. The extraction of answers from the MC forms has evolved from traditional manual methods to automated answer extraction systems/techniques. However, many of those techniques solely depend on fixed form designs or pre-defined templates (this can be a set of rules, guidelines, or a map of the MC form.), which are a considerable burden since unknown forms made from another technique cannot be processed, while others rely on manual user input to operate due to lack of automatic template

generation. This research offers a dynamic MC test answer extraction technique that automatically generates templates for unknown MC forms. The novelty of the proposed technique is in its ability to generate a working template automatically and, in turn, extracts answers from unknown forms regardless of the design, formatting and origin. Detailed steps and algorithms will be discussed later in the chapter. Six hundred samples of twelve different MC form designs commonly in use have been implemented in this study to test the performance of the technique algorithms. The experimental results show that the proposed MC answer extraction technique has successfully generated the template with no error according to the study objectives and scope. Based on the generated template, 100% extraction accuracy has been achieved.

5.2   MULTIPLE CHOICE ANSWER EXTRACTION TECHNIQUES

From the context of this research, the term 'Multiple-Choice (MC)' form can be defined as a form of assessment in which the respondents are asked to select the best possible answer or response out of listed choices or options (Parul & Kaur, 2012). The form can also be classified into two categories; the ones with pre-stored answers, which can either of answer/response extraction or grading (pass or fail) and the ones without pre-stored answers, mainly for non-grading techniques such as the ones used in survey and election. It should also be noted that a pre-defined template refers to an existing or fixed MC form template that has already been defined and thus, should not be confused with pre-stored answers, which is a selection of stored MC form answers used for grading. The automatic MC answer extraction is a process of detecting the presence or absence of a mark in terms of its position, intensity, or depth on a form. A mark is a label or an indicated position on the form filled in with a marker such as a pencil. This process relates to that of the OMR. However, from the context of this study, the OMR refers to the process (the technology) of MC automatic answer extraction, while the optical mark read refers to the machine.

Therefore, an MC answer extraction technique can be regarded as an automatic assessment technique whereby the selections in each form are recognized and extracted following the pre-defined template.

The early traditional (manual) procedure for extracting responses from the MC scripts is generally a manual data entry process which has many disadvantages in terms of speed, accuracy (such as mistakes and typos) and cost, not to mention tiredness and fatigue due to long hours of the manual data entry. The MC answer extraction techniques provide greater feasibility, incorporating flexibility on the chosen materials, simplicity in handling, and effectiveness in delivering accurate results. Just as in the earliest machine, most of the current automatic MC form answer extraction techniques require the forms to be printed on the trans-optic sheet and marked/shaded with special markers such as number two pencils, while others require the forms to have special design formats and features such as identification marks or grid lines. This occurs mainly due to a lack of an automatic template generation algorithm that will automatically generate a working template for any Alien form. Hence, this study's primary purpose is to design and develop an automatic template generation algorithm to extract answers from unknown MC forms. Presently, there are different techniques and methods for MC answer extraction, categorized as photoelectric, online, and image processing based.

For the Photoelectric based technique, the OpScan® machine from Scantron, as an example, offered simultaneous data transfer between a scanner and a computer with a document throughput rate of 1,500 to 10,000 forms per hour (Opscan8 Datasheet, 2012). The advent of advanced programming languages such as MATLAB and C-programming languages enable answer extraction in real-time on Computer Systems (PCs) using an on-screen Graphic User Interface (GUI) and, in some cases, over the internet. Then, extracting selected responses from MC forms using dedicated machines and Image Processing Based (IPB) have been around since the late 70s. The Remark Office OMR (Remark Office OMR, 2012) and Form Return (OMR User Guide, 2012) are among the few. The technique employs