Test Scoring for Non-Optical Grid Answer Sheet Based on Projection Profile Method

Nutchanat Sattayakawee

***Abstract*—OMR systems for test scoring were widely developed as commercial solutions and research topics. The filling area of the answer sheets, found so far, is in form of bubble or lozenge. However, most schools in Thailand use grid answer sheets. This paper proposes the algorithm of test scoring for grid answer sheets. The method used is based on projection profile and thresholding techniques. The proposed method is also able to detect more than one or no selected choice. Among 300 test samples with 3 types of grid answer sheets and total 16500 questions, the average accuracy result was 99.909%, which was considered satisfactory. The error was caused by some pale marks from input rather than the algorithm itself.**

***Index Terms*—Grid, multiple-choice test, OMR, projection profile, test grading, test scoring, threshold.**

I. Introduction

Multiple-choice test scoring is a repetitive task, achieved by matching the answers and the keys, for each student, question by question. The instructors can be lightened the work load from this handy task by the emergence of Optical Mark Recognition (OMR) technology, providing automated test scoring. OMR for test grading has been used since 1960s [1] and existing in form of hardware and software. Hardware-based OMR consists of dedicated scanner which makes it more expensive and harder to maintain than software solution [2]. Consequently, software-based OMR is increasingly chosen, especially in small to medium institutes. However, it is surprising that this type of system is not widely used in educational institutions in Thailand in spite of the fact that multiple-choice tests are employed in almost every school, in particular primary and secondary schools. The drawback of any OMR solutions, making them not interesting enough to get adopted in schools, is their fixed cost, e.g. scanner device, software and maintenance costs, but importantly continually added cost of expensive optical pre-printed answer sheets in form of whether bubble or lozenge [3]–[4] shown in Fig. 1. These types of answer sheets require the students to waste their time blackening the whole area of a bubble or a lozenge, not allowing them to cross or tick on those areas.

From the survey carried out in this research, 56 schools out of 74 interrogated secondary schools from all regions of the country use traditional grid answer sheet in the form presented in Fig. 2. The majority of school-level instructors still grade the students manually using that mentioned type of answer sheets due to their very low costs. Importantly, it is noticeable that no OMR solutions exist for such type of answer sheets. A key factor of this effect is probably scoring reliability and accuracy issues. As a result, multiple-choice test scoring for traditional grid answer sheet is an attractive challenge to step over.

(a) (b)

Fig. 1. Mark in (a) bubble and (b) lozenge forms of optical answer sheet.

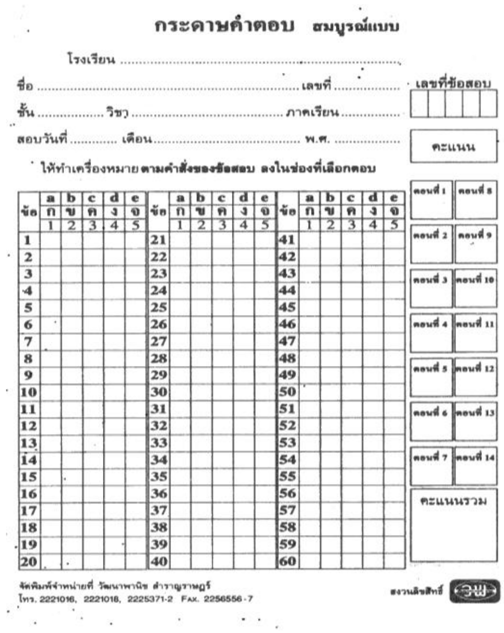


Fig. 2. Example of traditional grid answer sheet in Thailand.

The following sections are organized as follows. Exploration of test scoring OMR methods is in section II. Section III includes some assumptions for the proposed algorithm which is detailed in section IV. The results and conclusion are presented in sections V and VI respectively.

II. Exploration of Test Scoring OMR Methods

The general flowchart of OMR for test scoring found in almost all reviewed papers [5]–[8] are shown in Fig. 3.

Answer sheet image acquisition is photoelectric conversion consisting of lens and sensor as presented in [5]. In general, it comes in form of scanners and cameras. For fast acquisition, auto-document feeder is also employed. There exist auto-feed scanners, whereas camera solution proposed in [6] needs to afford this device. Step 2 in Fig. 3 consists of optimizing the input image to be ready for scoring in the next step. General pre-processes, as found in [6], [8], [9], involves

Manuscript received October 25, 2012; revised January 15, 2012.

N. Sattayakawee is with the Department of Computer Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Kamphaen Saen, Naknornpathom, 73140, Thailand (e-mail: fengncn@ku.ac.th).

DOI: 10.7763/IJIET.2013.V3.279

image binarization, skew correction and offset adjustment. For scoring the answers, it is no doubt that the most common method has always been counting the number of black pixels in the given segments as chosen in [5]–[10].

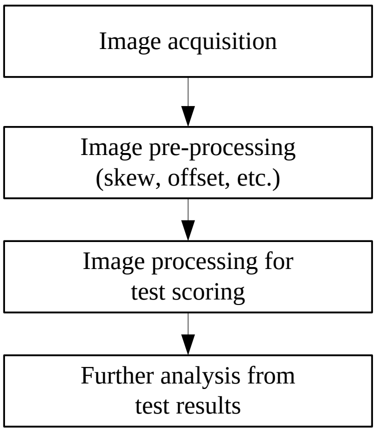


Fig. 3. Flowchart of OMR for test scoring.

The segments that must be detected in test scoring are segments for each question and segments for each choice in a given question. The segmentation and test grading are done by projecting the image vertically and horizontally, which is called image projection. The algorithms in [5]–[7] used the markers at the border of the answer sheet to help determining question and choice coordinates in the sheet. By comparing ON or black pixels among all choices in the same question, the choice with the maximum ON pixels is determined as the response of that question as proposed in [7], [9]. That is the reason why general OMRs cannot detect two-response answer or even reject the input image when no choice is selected for a given question. This does not make sense for the assessors who are the end users since the score for such a question in those cases should be zero if single correct choice is allowed. The performance measured is certainly accuracy and some also give importance to processing time. Item analysis, done in [3] and [11], completes full-path assessment. This last step is important for helping the instructors evaluate their students and the quality of the tests. However, the researches focusing on OMR usually ignore this as in [5]-[7], [9].

III. Assumptions

Since this paper focuses on the algorithm of test scoring for non-optical grid answer sheet, the following assumptions are made:

1) The scanned image is good enough, not skewed and well oriented, because there already exist many well-defined techniques to apply to achieve pre-processing steps.
2) The students can mark a cross with any types of pen or pencil whose color is well-seen from background color of the answer sheets, e.g. black, blue, green.
3) If the students make some mistakes on any choice marks, they must be erased cleanly; no cross-out is allowed.
4) The students should mark a cross in the same manner, especially when crossing more than one choice in the same question so that the algorithm can detect multiple responses. If not so, the algorithm may count the choice with maximum ON pixels as response. An example of non-uniform marks is shown in Fig. 4.

IV. Test Scoring with Traditional Grid Answer Sheet

An overall architecture of test scoring follows the pattern in Fig. 3. Image acquisition is done by a scanner. After the answer sheet image is pre-processed, the test grading is carried out using principally projection profile and thresholding. Step 3 in Fig. 3 can be detailed in Fig. 5.

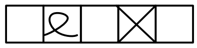


Fig. 4. Multiple non-uniform marks in a question.

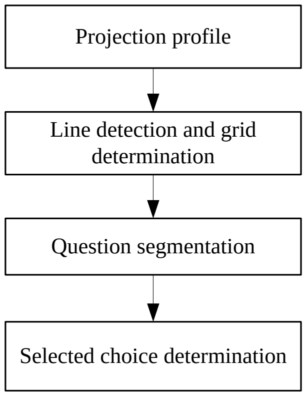


Fig. 5. Flowchart of test scoring with traditional grid answer sheet.

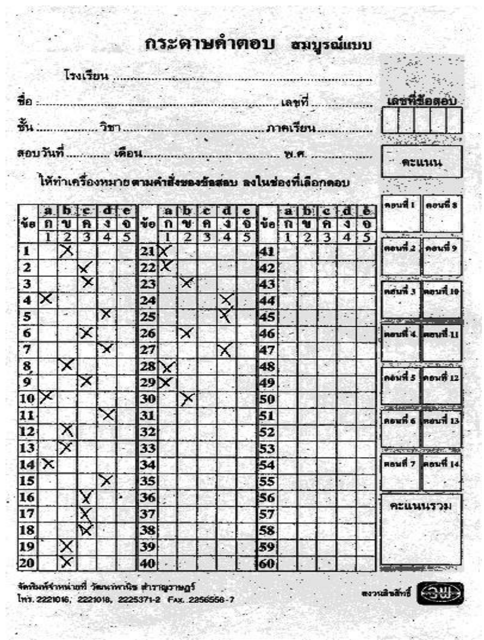


Fig. 6. Grey scaled image of the answer sheet.

A. *Answer Sheet Pre-Processing*

In this proposed method, the answer sheet image is converted from colored image to grey scaled one and is resized proportionally to the width of 400 pixels. After this simple pre-processing, the answer sheet to be processed later looks like the one in Fig. 6.

B. *Test Scoring Processing*

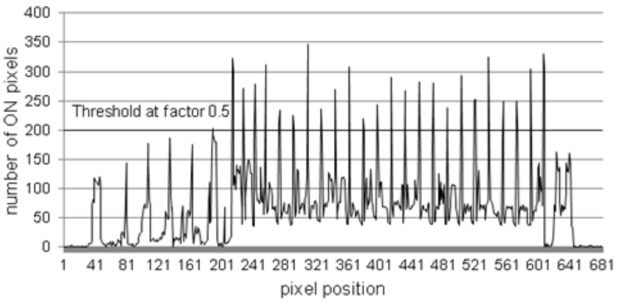
From Fig. 5, there are 4 steps in grading the test, which are detailed as follows.

1) Projection profile

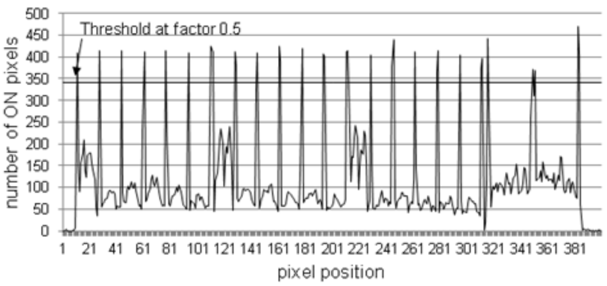
The image is projected horizontally and vertically. An example of answer sheet projection profile can be shown in Fig. 7. From Fig. 7, we observed that grid zone will have high frequency of ON pixels and each line of the grid is almost equally separated. This remark was used in next step.

2) Line detection and grid determination

The purpose of this step is to determine the image positions considered as lines of the grid in the answer zone. Line detection seems to be easily achieved by selecting the pixel with high frequency, thus thresholding is chosen.



(a)



(b)

Fig. 7. Projection profile of the answer sheet (a) horizontal projection (b) vertical projection.

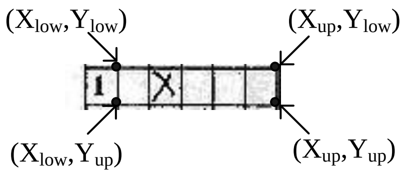


Fig. 8. Question zone represented by the quadruple (Xup, Xlow, Yup, Ylow).

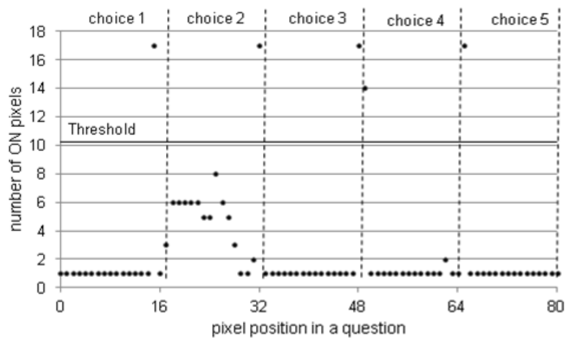


Fig. 9. Projection profile of answer zone of a question.

What must be pre-determined are the threshold and the numbers of vertical lines and horizontal lines in the grid noted as N_{row} and N_{col}, e.g. from Fig. 6, $N_{row} = 21$ and $N_{col} = 19$. The method then consists of finding the N_{row} horizontal lines and N_{col} vertical lines at the right positions in the image. In the algorithm, the threshold can be configurable, but from the observation, as demonstrated in Fig. 7, factor 0.5 was suitable to find N_{row} and N_{col} lines. Detecting horizontal lines are done using horizontal projection and vertical lines using vertical projection. The algorithm iterates all the pixels of the width and the height of the image to find those lines. From Fig. 7, there can be more than one pixel representing a line. In this case, the first position whose frequency is above the threshold is selected and consecutive positions above the threshold are ignored. After that, the number of detected lines, noted as $N_{detected}$, as well as their positions, is identified. $N_{detected}$ must be compared with N_{row} or N_{col} for different projections, noted as $N_{desired}$. There are 3 cases to get solved:

1) $N_{detected} = N_{desired}$: the evaluation of whether all detected lines are roughly equally separated is processed by using the standard deviation; if this is the case, the process is finished, if not the threshold is declined by factor 0.03;
2) $N_{detected} > N_{desired}$: by evaluating the gaps between each line with the average gap width, some lines which are not well separated are removed out until $N_{detected}$ is equal to $N_{desired}$, after that the evaluation of the previous case is carried out;
3) $N_{detected} < N_{desired}$: the threshold is declined by factor 0.03 and the whole process recommences.

Emerged from the remark figured out in step *1) Projection Profile*, the evaluation of the gap width between each line with standard deviation can solve false positive and false negative lines.

TABLE I: ACCURACY RESULTS OF TEST SCORING

Answer Sheet type	Number of questions	Accuracy Percentage
I	50	100%
I	100	99.7%
II	30	100%
II	60	100%
III	30	100%
III	60	100%

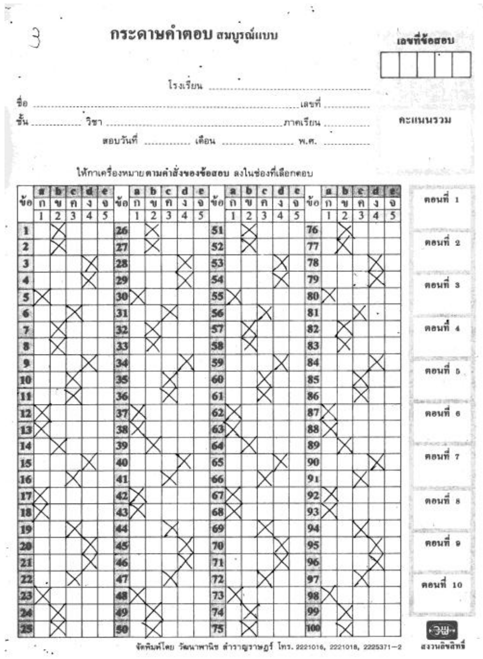
3) Question segmentation

After determining the correct line positions in the grid, each question zone must be determined in order to score the response of that question. Question number is excluded from question segment. This step takes the template of the answer sheet and line positions from the previous step to separate one question from the others. Each question comprises 4 positions (X_{up},X_{low},Y_{up},Y_{low}) which are lower and upper bounds of the question answer zone as shown in Fig. 8.

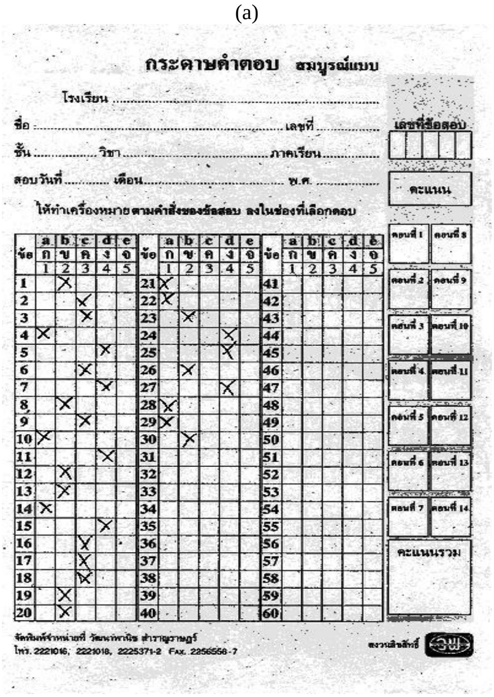
4) Selected choice determination

This step aims at determining the choice selected in the answer sheet for each question by using the results from 2 previous steps. The number of choices in each question must be known in prior, noted as N_{choice}. Choice segmentation is done by simply calculating the average choice width from $(X_{up} – X_{low})/N_{choice}$. For each choice, local vertical projection profile is used to determine selected choice. In each question zone, lines separated between each choice still exist, thus thresholding is used to separate cross mark on a choice from those lines. The frequency of ON pixels above the threshold factor of 0.6 won’t count in selected choice determination. Fig. 9 shows the local projection profile which helps analyze the answer of a student for each question. The choices with high rate of ON pixels are candidates to be selected. The algorithm needs to check for more than 1 choice selected or no choice selected. This achievement is obtained by checking for 2 choices, noted as C1 and C2, where C1 represents the choice with highest frequency of all choices, noted as F_{C1} and C2 the choice with second highest frequency noted as F_{C2}. Two choices will be considered for 2 selected choices from the students when $F_{C2} \geq 0.7*F_{C1}$. However, this can be false positive, meaning that C2 is not actually selected, but detected as is. To double check this situation, mode is used to determine the cross mark or unselected choice by observing the projection graph for each choice as shown in Fig. 9. If the mode is above a half of total number of the sample, the choice will be considered as unselected choice. From local projection profile example in Fig. 9, which resulted from Fig. 8, choice 2 is selected.

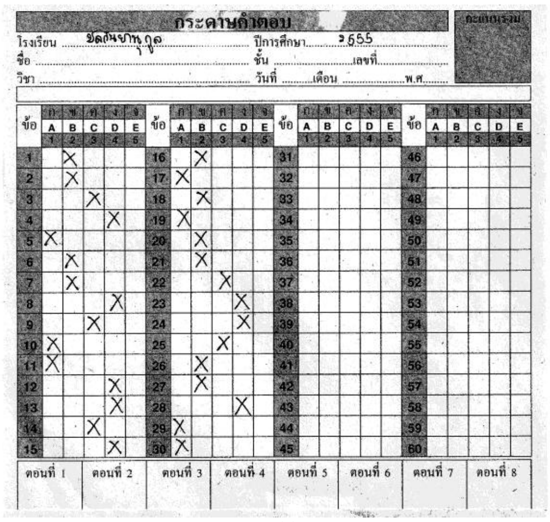
By counting the selected choice mapping to the key for each question, the score of the test is obtained.



(a)



(b)



(c)

Fig. 10. Tested types of answer sheets with (a) type I: 100 questions of 4 columns and 15 rows (b) type II: 60 questions of 3 columns and 20 rows and (c) type III: 60 questions of 4 columns and 15 rows.

V. Results

In the experimentation, percentage of correctness was measured. Three types of answer sheets were tested as illustrated in Fig. 10: type I containing 100 answers, 25 answers in each column; type II containing 60 answers, 20 answers in each column; and type III containing 60 answers, 15 answers in each column. Each type of answer sheets was experimented in 2 tests with different number of questions. Each test was defined by 50 samples, each sample marked the same answers, but 1% of the questions contained 2-choice, no selected choice or cross-out choice answers in the samples randomly. The questions with cross-out choice were expected to be detected as 2-choice answers. The samples of 6 tests, as shown in Table I, came from 6 different classes of Chonkanyanukul School, Thailand. Table I presents the accuracy result of each test. Total 16500 questions were tested in three types of answer sheets, but 15 answers were detected incorrectly in answer sheet type III. This error came from the fact that the ink of a student’s pen was pale from questions 21 to 77. However, the algorithm was still able to detect 42 answers correctly. Consequently, the average accuracy result was 99.909%.

VI. Conclusion and Future Work

This paper proposes an algorithm for test scoring using non-optical traditional grid answer sheets of Thailand. The algorithm is based on projection profile and thresholding methods. The main distinctions from other OMRs for test grading are the answer sheet forms which are much harder to process and the detection of 2 or no selected choice. The average accuracy was 99.909%. The actual error came from an abnormal input, rather than the algorithm itself. This algorithm can also be applied in other different types of answer sheets in grid form.

For future work, the algorithm can be improved in many angles, especially some repetitive task that can be learned from the same pattern of answer sheet. For example, line and grid determination should be pre-determined before scoring operation or should be learned from first answer sheets in order to accelerate the whole operation. The algorithm should be able to recognize the cross mark so that the cross-out of the wrong choice is allowed. For full traditional answer sheet OMR, the algorithm must be able to identify the student by OCR (Optical Character Recognition) techniques.

ACKNOWLEDGMENT

I am very thankful to Mr. Winai Witayalai, a teacher of Chonkanyanukul school, Thailand, for providing all the answer sheets samples and Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Thailand, for all facilities that let me accomplish the work.

REFERENCES

[1] A. M. Smith, "Optical mark reading - making it easy for users," in *Proc. 9th annual ACM SIGUCCS conf. User services*, USA, 1981, pp. 257-263.

[2] Software Farmer. (October 10, 2012). Optical Mark Recognition. Software Farmer Inc.. [Online]. Available: http://www.softwarefarmer.com/content/view/14/1

[3] Gravic. (October 10, 2012). Remark Classis OMR Software for test scoring and survey scanning. Gravic Inc., [Online]. Available: http://www.gravic.com/remark/classicomr/

[4] Scantron. (October 10, 2012). Scantron Genuine Forms. Scantron Corp. [Online]. Available: http://www.scantron.com/forms/

[5] S. Hussmann and P. W. Deng, "A high-speed optical mark reader hardware implementation at low cost using programmable logic," *Real-Time Imaging*, vol. 11, no. 1, pp. 19-30, February 2005.

[6] T. D. Nguyen, Q. H. Manh, P. B. Minh, L. N. Thanh, and T. H. Hoang, "Efficient and reliable camera based multiple-choice test grading system," in *Proc. 2011 ATC international conf.*, 2011, pp. 268-271.

[7] H. Deng, F. Wang, and B. Liang, "Low-Cost OMR Solution for Educational Applications," in *Proc. 2008 ISPA*, 2008, pp. 967-970.

[8] C. Saengtongsrikamon, P. Meesad, and S. Sodsee, "Scanner-Based Optical Mark Recognition," *Information Technology Journal*, vol. 9, pp. 69-73, June 2009.

[9] K. Chinnasarn and Y. Rangsanseri, "An image-processing oriented optical mark reader," in *Proc. the International society for optical engineering*, USA, 1999, pp. 702-708.

[10] D. Ivetic and D. Dragan, "Projections Based OMR Algorithms," in *Proc. 8th INES conf.*, 2004, pp. 226-232.

[11] A. Rane, A. Kumar, H. Saini, and M. Sasikumar, "Extending Moodle to Support Offline Assessments," in *Proc. ELELTECH*, 2009, pp. 31-39.

Nutchanat Sattayakawee was born in Phatthalung, Thailand in 1984. She received the diplôme d'ingénieur (Master's Degree) specialized in computer engineering from National Institute of Applied Sciences of Lyon (INSA de Lyon), Lyon, France, in 2008. She is currently a lecturer at the Department of Computer Engineering, Faculty of Engineering at Kamphaeng Saen, KasetsartThailand. Her research interests include image processing, web engineering, software engineering and programming methodology.