IJARESM
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ENGINEERING, SCIENCE & MANAGEMENT

EFFICIENT SYSTEM FOR EVALUATION OF OMR SHEET-A SURVEY

Divya Patel[1], Shaikhji Zaid[2]
M.E. Students[1], Department Of EC, SNPIT & RC, Umrakh, Gujarat, India[1]
Assistant Professor[2], Department Of EC, SNPIT & RC, Umrakh, Gujarat, India[2]

Abstract: ***Optical Mark Recognition (also called Optical Mark Reading and OMR) is the process of capturing human marked data from document forms such as surveys and tests. This technique is widely used in various applications like exam evaluation, automated attendance marking, voting and community surveys etc. Pencil or pen marks made in predefined positions on paper forms as responses to questions or tick list prompts can be read by the reader. These marks are digitally entered into a computer for further analysis. OMR is very useful when data is to be collected from a large number of sources simultaneously and a large volume of data must be collected and processed in a short period of time. Optical Mark Recognition is the technique used to scan a marked paper to detect the presence or absence of the mark in a predetermined position. Problem with the OMR technique lies is that, it cannot process thin papers and low printing precision answer sheets. This paper analysis a new technique that overcomes the limitation of OMR process. It scans the hard copies of papers which are designed and marked in a specific format and saves that scanned copy in JPEG format, then identifies the responses marked from the JPEG image and stores the result in a database. Errors and their solutions which could creep in due to scaling, rotation, translation of the scanned copy are also presented.***
Keywords: Optical Mark Recognition, Image processing, Adaptive Threshold, Scanner.

I. INTRODUCTION

OMR technology has changed much in recent years. Now a day in schools, colleges and classes OMR technology is used. Exams are conducted using OMR answer sheet checking system because by using this technology the conduction of exam is getting much easier, powerful, and cheap[6,12]. optical markup recognition (also called optical mark reading) is the process of capturing human-marked data from document forms such as surveys and tests.

OMR is a technology that detects the absence or presence of a mark, but not the shape of the mark. OMR software interprets the output from the scanner, and translates it into the desired ASCII output. Forms are scanned through an OMR scanner. The forms contain small circles, referred to as bubbles, or boxes that are filled in by the respondent. Optical Mark Reader reads marks written by pencil or ballpoint pen in the pre-defined positions on the questionnaire sheet. The OMR can judge the existence of written marks by recognizing their depth (darkness) on the sheet[9].

II. OPTICAL MARK RECOGNITION USES AND ACTIVITIES

The most common use of optical mark recognition is to process student responses to a multiple choice exam, or responses to a questionnaire or feedback form. Typically the questions are provided on paper, and students mark their responses onto special pre-printed forms. These forms are then read automatically[4]. OMR solution that uses scanners and supports plain sheets and low printing quality.

We have a possibility that our application would do the following activities[7]:

- checks the answer sheet scanner,
- detect responses for each question.
- Compare student responses with reference answers which would be pre stored in database.

OMR based evaluation is preferred over the manual methods when-

- A large volume of data is to be collected and processed in short period of time.
- Data is to be collected from large number of sources simultaneously.
- Questionnaires consists of multiple choice questions or selections of categories.
- Very high accuracy is required.
- Survey collectors are limited.

OMR machines with specialized and dedicated Infra-Red OMR scanners have been recently popular mainly because of their high execution speed and appreciable accuracy. Infra-Red OMR scanners work only for specific color and thickness of the form hence cannot be printed on a general purpose printer. Another problem with dedicated machine is their cost and maintenance. Also these scanners expect the form to be in perfect condition; even a slight damaged, crush or fold would be rejected. Although these machines address the problem in hand, their usage is limited due to high cost, strict specifications and inflexibility in format of the form[14].

The proposed solution to this problem is very simple and cost-effective, which could easily replace the heavy and expensive present day dedicated OMR machines. Unlike scanners and computers, OMR machines are not commonplace hardware resource. The proposed system uses commonly available scanner and computer. The filled-in forms are scanned by a normal scanner, and scanned images are provided as input to an ordinary computer. An image processing application is developed that gives the information of filled and unfilled bubbles in the form. The system is made to run in multiple threads, thus making parallel computation and evaluating thousands of forms in few minutes. The computed results are stored in a spread sheet, thus making it easy to understand and interpret the information. The system offers flexibility to the users, which allows designing and printing of forms on simple ordinary sheets, without the aid of any special pre-processing and color constraints. The OMR forms evaluated by dedicated scanners generally range in the quality of 90-110 gsm, which are much more expensive than the common plain papers (60 – 70 gsm), thus making the system to function on low cost. The substitution of special or expensive dedicated high computational machines by ordinary scanner and computer with no stringent requirements on paper quality makes the proposed system cost effective[14].

Basically designed software is to check the answer sheet and display the results. Firstly the OMR answer sheet will be scanned and the scanned image of the answer sheet will be given as input to the software system. Using Image processing we will find the answers marked to each of the 50 questions by finding the region of interest. Summation of the marks & displaying of total marks will be also implemented[12].

A. Research Methodology

In digital image processing there are various types of images used and in order to find the target object using template matching technique the first need to be converted into binary image.

1) Grayscale Image: Gray scale image uses 1byte space for storing each pixel and each pixel intensity value can varies between 0 to 255, where 0 refers to Black means Highest intensity and 255 refers to White means lowest intensity.

2) Color Image: It uses 3 bytes per pixel as it has RGB components. The intensity of each pixel is 0-255.

3) Binary Image: A binary representation of image is called true digital image format because it map image only in two color value either 0(for black pixel) or 1 (for white pixel). Binary Image uses only 1 bit space for mapping each pixel.

III. VARIOUS TECHNIQUES FOR ASSESSMENT

In [5] At present most of desktops, laptops, tablets, and even smart phones are shipped with multi-core processors. The efficient utilization of multi-core processors computation power can't be achieved by developing traditional applications with sequential algorithms. Parallel algorithms utilize the capabilities of these processors. They are very well suited for parallel processing. This work represents a low cost and fast solution for optical mark recognition system working in multi-core processor system. In this system a solution for camera based OMR is presented. This system turns on special designs of the answer sheet to add some marks which speed up the detection of bubbles. The system is insensitive to rotation scaling and illumination variations. In addition to that the flipped images can be processed and recognized without correction. The solution keep out of the way of heavy computational algorithms such as skew correction, circle detection and Hough transform, to increase the speed of the system.

There are various components of the camera based OMR system:

A. Preprocessing and Bubble Detection

Image should be thresholded then the bubble location is finding in the answer sheet. Each pixel in the image is separated as an object or a background. The borders are move out giving us information about the skew and perspective distortion. There are external borders and another internal separation lines between each column. Underline markers are used to change over the bubble detection process from circle detection into a fast line tracking process**.**

Adaptive Binarizarion: One of the main problems in camera based document analysis is the binarization process. Various types of degradations such as uneven illumination, shadows, low contrast, smears and heavy noise densities often make thresholding of the document images a difficult job [11].

The simple fixed threshold level used for binarization is not suitable for the lighting variation as shown in Fig.1. the adaptive binarization is used where the image is divided vertically into

columns with its different thresholding values. Also inside the single column each row has its different thresholding value depending on contrast analysis.

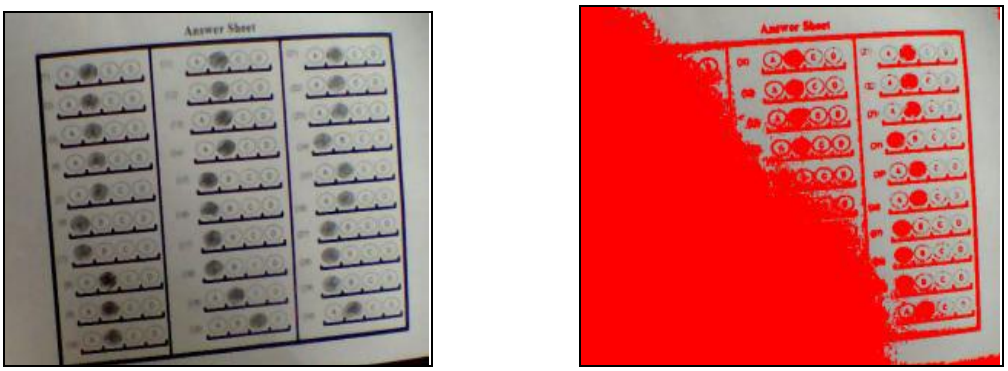


Figure 1: Binarization with constant threshold level

1) Borders Extraction: The design of the template with a thick line borders works out many problems and increase the reliability and speed. These lines are insensitive to noise and help us to work out the rotation and perspective distortion problems. There are many line detection algorithms with different complexities and robustness. Hough transform is robust for noise and occlusion, but the calculation and/or memory costs are very high.[8] Projection is the quickest way in finding horizontal and vertical lines in an assigned image, because such lines will produce peaks in projection profiles[16]. Without using structural processing, the thick line is chased using edge tracking. If the line is fall apart a connection algorithm is used to connect both segments of the line. There are no dependencies between different lines detection and the processed are parallelized easily.

***2) Bubble Detection*:** The edge tracking algorithm with the added heuristics has accomplished a fast and robust bubble detection results. In addition to that it works out the problems of skew, rotation, and perspective distortion. The parallelization in the bubble detection process is accomplished by assigning a specific number of rows to each core.

B. Feature Extraction and Classification

The simple solution to separate the bubble is the brightness difference between marked and unmarked bubbles. This simple solution has some problems. The small errors and deviations in finding the location of the bubble cause categorization errors. Also the different lighting conditions make problems in finding the threshold between the two classes brightness. To move toward with these problems different features are extracted. To achieve the best categorization results training process is applied.

1) Feature Extraction: The gray level difference between Marked and unmarked bubble is the main feature in the classification process. To trim back the effect of noise and illumination variation, the difference in gray level of the current bubble and the background has been used. The gradient features are used in the case of light variations.

2) Classification: The classification process can be simple task, while the number of classes are only two. But practically the errors and deviations in finding the location of the bubbles in addition to the different lighting conditions make overlap between the two classes and add some complexity to the classifier. Many classifiers have been used to test the accuracy and performance of the system. Naive Bayes, QDF, MQDF, and Neural Networks classifiers are used in that system.

IV. TECHNIQUE FOR LOW COST SOLUTION OF OMR PROCESS

In [10] Proposed technique is a low cost solution of OMR process. It neither requires high cost computational machine (reader) for scanning nor expensive high quality paper. Also this is an image based technique which can be used in small scale industries, institutes and schools. There are four basic steps in this proposed method.

A. Template Design

In this phase, a layout of the questionnaire, a template is designed using Microsoft word. Microsoft word is an application which can be learnt easily and also be used by non technical persons. A sample layout of template is shown in Fig.2.

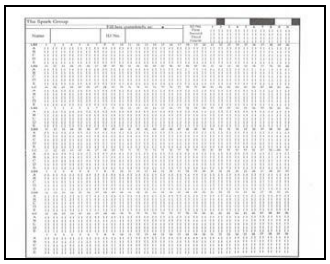


Figure 2: Layout of questionnaire

The designed template layout has a grid which helps in checking the alignment and scaling. All coordinates of template including top left corner, top right corner, bottom left corner and the bottom right corner of grid are saved in database. Also the aspect ratios of all the option fields are consistent.

B. Image Capturing

To capture the image of the questionnaire, flatbed scanners are used. These scanners allow users to work with prints, paper documents, and even three dimensional objects. Normally, the flatbed scanners take 10-20 seconds to scan an 8 x 10 inch image. This image is stored in JPEG format.

C. Perform Transformation and Scaling

During the scanning process, if the form is not put properly on the scanner, or if the resolution of the scanner is not set to normal, the scanned image may be tilted or translated with respect to the original template referring Fig.3.

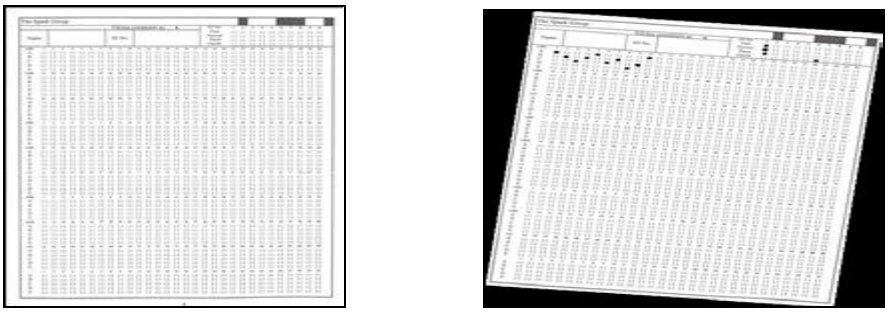


Figure 3: Scaling and Rotation Issues

This might also happen while photocopying the form (to get multiple copies of blank form).

1) Transformation: If the scanned image of the form is tilted, the rotation needs to be done. It makes use of the grid present on the form and calculates the degree of rotation and translation, then corrects it to match the standard template before any processing gets started. The coordinate positions of the standard template are already stored in database which is then used to identify the responses marked.

For resolving rotation, the steps to be followed are:

- Find the pixel coordinate value of top left and bottom left position of the scanned image of questionnaire form.
- Now compare the grid of questionnaire to the grid of template. The grid of scanned image may be inclined by an angle Θ to the standard template. The grid of scanned image will be rotated by Θ to match with the grid of standard template.
- The rotation angle Θ (shown in Fig.4.) of the questionnaire with respect to template can be calculated by the Eq. 1 :

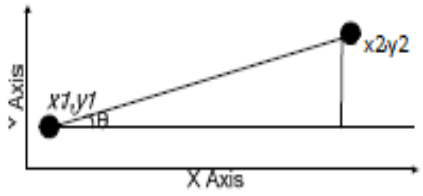


Figure 4: Rotation angle of questionnaire to template

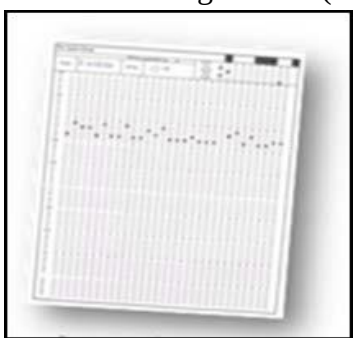
Rotation angel $\Theta = \tan^{-1}[(y2 - y1) / (x2 - x1)]$ ……(1)

2) Scaling: If the aspect ratio of printed form is differ from the standard template, scaling is need to be done. For resolving scaling, the steps to be followed are:

- Pixel values of all the coordinates of the scanned image are stored in the database.
- With the help of the coordinates of top left corner, top right corner and bottom left corner the height and breadth of standard layout and scanned image are to be calculated.
- Now the height and breadth of scanned image are compared to the standard template and scaled to convert all coordinate values of scanned image of questionnaire into a uniform system.

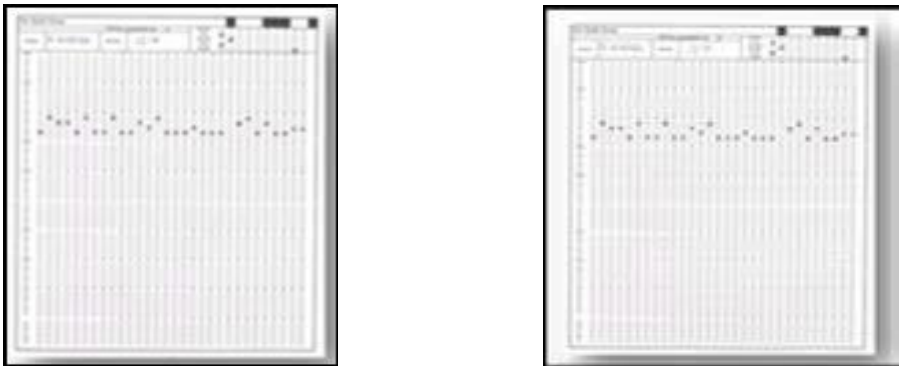
D. Finding marks on questionnaire

As all the coordinates of the template layout and scanned image of questionnaire are matched, marks recognition process becomes easy. All the marks on the questionnaire image would be 'filled' or 'blank'. More than 70% random pixels of single mark are checked. If 50% or more pixel coordinates are colored then the mark is assumed to be "filled". In single choice question if two or more than two marks are colored than the mark with higher percentage of colored pixel is considered to be 'filled'. For image processing mainly two steps are followed i.e. checking inclination and then checking scaling. After alignment(rotation and scaling) of the questionnaire image, mark recognition is to be done and results are generated(referring Fig.5).



Scanned Copy

RESULTS



Aligned Copy

Figure 5: All steps of Image Processing

CONCLUSION

Evaluation of OMR (with simple scanner) is accurate, time effective and cost effective. The scanning can be done by simple scanner. The system efficiency depends on the speed of the scanner. We have given an initial contribution to evaluation of methods for performance analysis of various techniques of assessment of OMR sheet using ordinary 2D scanner. By means of such evaluations we support system designers in choosing performance analysis method that is most suitable for their particular requirements.

REFERENCES

[01] Infotronicx. (2010). OMR system.

[02] Hui Deng, Feng Wang, Bo Liang, "A Low Cost OMR Solution For Educational Applications", IEEE 978-0-7695-3471-8/08 DOI 10.1109/ISPA, 2008.

[03] K. Toida, "An Overview of the OMR technology: based on the experiences in Japan", Workshop on Application of new information technology to population: Paper based data collection and capture, Thailand, 1999.

[04] Bhavana B.Nerkar, "Optical Markup Recognition for Exam system," International Journal of Emerging Technology and Advanced engineering Volume 5, Issue 2, February 2015.

[05] A. AL-Marakeby, "Multi-Core Processors for Camera based OMR", International Journal of Computer Applications (0975 – 8887) Volume 68– No.13, April 2013.

[06] Rusul Hussein Hasan, Emad I Abdul Kareem "An Image Processing Oriented Optical Mark Reader Based on Modify Multi-Connect Architecture MMCA" International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 02, Issue 07, [July– 2015].

[07] Soumitra Patole, Aniket Pawar, Avanish Patel, Akshay Panchal ,Prof. Ravindra Joshi, " AUTOMATIC SYSTEM FOR GRADING MULTIPLE CHOICE QUESTIONS AND FEEDBACK ANALYSIS", International Journal of Technical Research and Applications e-ISSN: 2320-8163.

[08] Kazuhito Murakami and Tadashi N aruse "High Speed Line Detection by Hough Transform in Local Area" 15th International conference on pattern recognition ,2000.

[09] Krishna G, Rana H R, Madan I, Sahu R, " Implementation Of OMR Technology With The Help Of Ordinary Scanner", international journal of advanced research in computer science and software engineering, vol-3, issue- 4, April 2013.

[10] Astha Gupta, Sandhya Avasthi, "Image based low cost method to the OMR process for surveys and research", International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-7,July 2016.

[11] Landu Jiang1 Kai Chen1 Shibo Yan1 Yi Zhou2 Haibing Guan3 ,"Adaptive Binarization for Degraded Document Images",Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on.

[12] Ms.Sumitra B. Gaikwad, "Image Processing Based OMR Sheet Scanning" International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 3, March 2015.

[13] TienDzung Nguyen, Quyet Hoang Manh, "Efficient and reliable camera based multiple-choice test grading system" 2011 International Conference on Advanced Technologies for Communications (ATC 2011).

[14] Rakesh S, KailashAtal, Ashish Arora, "Cost Effective Optical Mark Reader" International Journal of Computer Science and Artificial Intelligence Jun. 2013, Vol. 3 Iss. 2, PP. 44-49.

[15] NutchanatSattayakawee, "Test Scoring for Non-Optical Grid Answer Sheet Based on Projection Profile Method" International Journal of Information and Education Technology, Vol. 3, No. 2, April 2013.

[16] ZHI-HONG ZHAO, XUE-DONG TIAN, BAO-LAN GUO , "A STUDY ON PRINTED FORM PROCESSING AND RECONSTRUCTION" ,Proceedings of the First International Conference on Machine Learning and Cybernetics 2002.

