IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 5, Issue 2, February 2018

Optimized Approach for Parallel OMR sheet Analysis

[1] Varsha Kumari, [2] Dr. Aprna Tripathi
[1][2] Assistant Professor, GLA University

***Abstract* - Recently there has been a considerable increase in the competitive examination and academic examination. Most of these examinations are using Optical Mark Recognition (OMR) sheet for filling the right choice of Objective type questions commonly known as Multiple Choice Questions (MCQs). In this examination process, every question consists of four different choices and the candidate has to mark the right choice. OMR system is used to avoid counting mistake and it reduces manual work of evaluation. In this paper, we are discussing different techniques used in OMR based evaluation. This paper proposes an efficient OMR evaluation technique to reduce the time of evaluation of OMR sheet. The main focus of the paper is on parallel processing in capturing the image and evaluation. The work is divided into two phases. First one is the processing phase in which OMR sheet is captured and data of every question is taken into matrix form. The second phase is the evaluation phase in which the result is calculated on multi-core processor architecture using the proposed algorithm. Also, comparative analysis of sequential and parallel OMR processing has been shown. It is expected to get optimized utilization of processors and processing time.**

***Keywords:* --- Grading system, parallel image processing, MCQ, OMR.**

I. INTRODUCTION

OMR Technology is the technique of gathering information from human by recognizing marks on paper. OMR sheets are still considered a reliable and effective method for collecting opinion or feedback during survey. A lot of competitive examination, colleges uses OMR sheets for tests and assessment. Even though online computer based assessment or surveys are more convenient but it is too expensive to be implemented. Multiple Choice Questions (MCQs) are popularly used as fast and reliable method for entrance examination all over the world. Several applications based on Optical Mark Recognition Technology has been developed to overcome manual grading solutions [3].

The scanning machine, referred to as OMR scanner, is used to read a large numbers of forms automatically. OMR machines uses sensor to detect student's response by determining whether the predefined position is blank or marked. The problem with OMR machine are its price and operating cost due to MCQ scoring papers which are more expensive than plain papers. To reduce the cost of OMR scanner machine, several image based OMR software systems have been developed.

Image-based OMR doesn't require costly hardware devices or special answer sheet. It also allows the user to design his special answer sheet using word processing software. The image based OMR depends on scanning the document. It uses different image processing and pattern recognition techniques to detect the class of the marked bubbles in the answer sheet.

Nowadays, Camera-based document analysis has received considerable attention due to the wide spread of advanced low priced digital camera [3] [4] [9]. The problems faced by camera based document analysis are illumination problems, zooming and focusing problems, low resolution, and perspective distortion, speed and accuracy in evaluation.

Multi-core processors enhance the performance of the system by concurrent execution of the allocated workload on different processors. There are many hardware architectures and software platforms that support parallel processing. The most common hardware architectures are FPGA, GPU, multi-core CPU, and DSP. For the software, many platforms that support developing parallel algorithms are used such as OpenMP, Intel TBB, Intel ArBB, and CUDA [4][10][11].The parallelism can be applied in image processing applications by three main ways: i)Data Parallel, ii) Task Parallel, and iii)Pipeline Parallel.

In this paper an algorithm for camera based OMR is presented with efficient and reliable performance. Parallel image processing has been used to utilize the parallelism of multi-core processors. The performance of proposed algorithm is beneficial in terms of reducing man power and time consumption.

The rest of the paper is organized as follows: section II summarizes the brief review of related work. In section III, the proposed algorithm is discussed. And section IV shows the results and finally ended with the conclusion of the work and future scope.

IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

II. LITERATURE REVIEW

In this section, the literature review related to the existing evaluation systems has been discussed.

Liang Bo, Wand feng and Deng Hui [1] presents a Low Cost OMR solution which uses very low quality printing to decrease the overall cost for using OMR technology. It uses a thresholding process based on the black pixels in the binary image after a set of pre-processing steps.

Sanguansat [2] proposed an automated data entry using Optical Mark Recognition (OMR). They uses regular image scanners in OMR software applications. In this paper, they proposed a solution for preparing questionnaire consisting of only close-ended questions. It also includes proper report of the output obtained in spreadsheet software.

Tien Dzung Nguyen [3] paper presents another technique for grading multiple-choice test which is based on a camera with reliability and efficiency. It demonstrated the use of non-transoptic answer sheet paper with lower cost for achieving accuracy. Accordingly, they have applied skewed adjustment and normalization steps to allocate the answer boxes. And the decision is taken based on the number of black pixels.

A. AL-Marakeby [4] presents a low cost and fast solution for optical mark recognition system working in multi-core processor system. The answer sheet is captured using a digital camera and the image is processed. Initially the borders of the sheet are located then the bubbles are detected. Using bold lines and borders as markers used to detect the location of bubbles solves the problems of rotation, skew, perspective projection and also flipping. The illumination variation problem is solved using the Adaptive binarization technique. Edge following algorithm is a fast method in the detection of lines and borders.

Sumit Tiwari [5] introduces a new method to detect OMR sheet tampering in order to enhance the security of the OMR system, using encrypted QR codes. Examinee details and number of responses for the asked questions they filled are embedded in the encrypted QR code which is printed on the OMR sheet, in avoid any alteration in the OMR sheet details. It also presents an approach for the Detection of OMR Sheet Tampering using ECC algorithm and ZXing library for the encoding and decoding of QR Codes.

D. Chai [6] presented a semi-template-free MCQ test assessment system. The system uses set of mobile applications that are available to grade MCQ tests automatically. Also uses a fixed sheet structure that has specific areas for answer boxes, instructions, and student IDs to detecting the data.

J. A. Fisteus et al. [7] proposed a system called Eyegrade that is based on web-cams. The Eyegrade entails a two-stage thresholding method that recognizes the answer boxes. It recognizes both mark recognition and optical character recognition of handwritten student identification numbers. When compared with similar webcam-based systems, the user interface in Eyegrade has been designed to provide a more efficient and error-free data collection procedure.

Mahmoud Afifi, Khaled F. Hussain [8] proposed another solution by training a classifier to recognize the class of each answer box. It is based on machine learning-based approach, high accurate OMR systems can be developed without any restrictions on the marking of the answer boxes, e.g. crossing out wrong answers. It uses three image classification methods: (1) NBC, (2) BoVW, and (3) CNN, in two different strategies of classification to get the best potential result.

Masakazu Iwamura [9] proposed another method that runs well in real-time even on a laptop PC with a web camera with help of new hashing and voting techniques.
Luis Miguel [10] discussed different parallel programming models (OpenMP, Intel TBB, Intel ArBB, and CUDA) to analyse increase in performance of parallel general-purpose threads (hyper threading, multicore, ccNUMA architectures) or SIMD kernels (CPU vector instructions, GPUs). Harshad B. Prajapat [11] also discusses the parallel and distributed image processing approach.

III. PROPOSED WORK

The algorithm is basically divided into two phases:
(i)The first phase is to capture the image by camera and apply preliminary techniques to enhance the captured image and then convert the data of marked bubble from enhanced image into matrix form.
(ii) The second phase is to compute the score of the candidate by comparing the answer matrix with the output matrix from phase 1.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

Phase 1:
The multiple images can be taken by camera based according to the number of processors available. In particular image of four OMR answer sheets is captured in a single image. Preliminary techniques scaling and rotation is applied on the image to get the enhanced the blurred or tilted image.

The block diagram of the first phase is presented in fig. 1

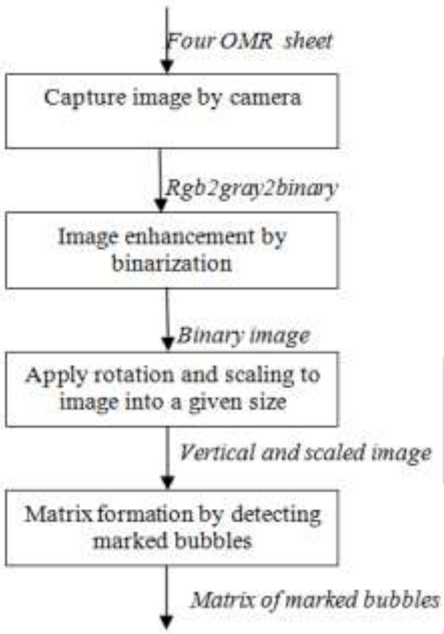

Fig. 1 block diagram of phase 1 of proposed algorithm

Captured image is stored as digital data in colored rgb representation schema. it is converted into gray scale image.

Fig. 2 image of one OMR sheet

Fig 2 shows the gray scale image of one OMR sheet, extracted from the image captured by camera. The image extraction from the digital image of OMR sheet technique require mapping of image in only black and white schema. Therefore this gray image is converted into most suitable binary image format. If the binary image is not placed properly, it may cause inaccurate result. The coordinate positions of the standard template are already stored in database which is then used to identify the responses marked. Find the pixel coordinate value of top left and bottom left position and compare the alignment of binary image and the grid of the previously stored template. Calculates the degree of rotation and translation using eq.1:

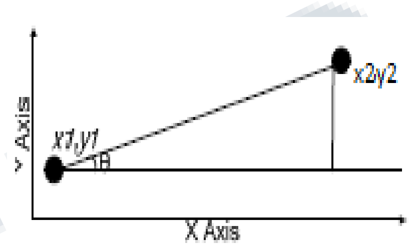

Fig. 3 rotation of OMR sheet

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

If the size of the binary image after rotation does not matches then this image is scaled to convert all coordinate values of scanned image of questionnaire into a uniform system. The question and options region is selected from each OMR picture from the binary image. Fig 3 shows the marked section of each OMR sheets. If 50% or more pixel coordinates for each bubble are colored then the mark is assumed to be “filled”. If the bubble is filled or marked ‘black’ then 1 is stored for the corresponding bubble in matrix otherwise 0 is stored for unfilled bubble. The process is repeated for every question. This output matrix is taken as an input for phase 2.

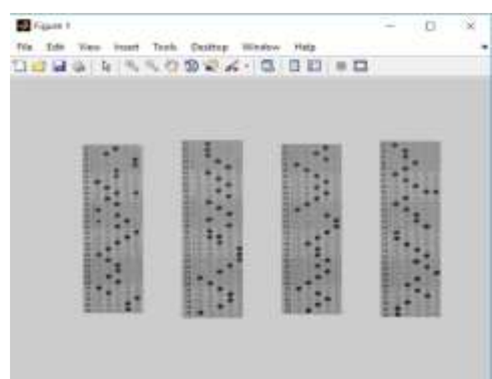

Fig. 4 marked section of each OMR sheet

IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

The following algorithm is applied for the first phase of OMR processing.

Algorithm 1 for phase 1

1: procedure phase1(input image)
2: M<-input image
3: convert image from rgb to grayscale image
4. Applying threshold values, convert the gray image to binary image
5.Apply rotation and scaling to get the image perfectly vertical
6.Select and combine all four marked sections in new image N
7. i<- 1
8.While i<=numberOfQuestions do
9. j<-1
10.While j<=no of bubbles in each row do
11. if the number of gray pixel surrounding center of bubble greater than threshold value then
12. T(I,j)=1
13.else T(I,j)=0
14.Return matrix T

Phase 2:

In this phase the evaluation work is done. The answer matrix is stored in the database for calculation. The output matrix from phase 1 is taken as an input for phase 2. The matrix data is distributed on 4 core for processing using MATLAB pool and single program multiple data schema. Before evaluation, the following three cases are considered:
(i)if no answer marked: it does not changes the score.
(ii)if only one answer marked which matches with the answer matrix data, it is taken into consideration for score evaluation.
(iii)if more than one answer marked or wrong answer marked: negative marking is done in this case.
Considering all three cases, the result is calculated after comparing the answer matrix with all the parallel data matrix.
Parallel evaluations of the score for four forms are done at the same time using following algorithm:

Algorithm 2 for phase 2

Algorithm 2 for phase 2

1.procedure phase2(answer sheet R, T)
2: R <- answer sheet
3.matlabpool (4)
4.Distribute (T)
5.Spmd
6. i <- 1
7: while i <= numberOfQuestions() do
8. j <- 1, c<- 0
9. while: j <= numberOfOptions do
10: if T(I,j)=1
11. A(i) <- j +1
12. c <- c+1
13. End if
14. End while
15. If c>1
16. A(i) =5
17. If c=0
18. A(i)=0
19. End while
20: i <- 1
21: while i <= rows do
22 if A(i)! =0 then
23. If A(i)==R(i) then
24. Total<- Total +1
25. End if
26. If A(i)==5 and A(i)!=R(i)
27. Negative<-negative +1
28. End if
29. End if
30. End while
31. Score <- Total –negative
32. Returns score
33. End
33. matlabpool close

IV. RESULTS AND ANALYSIS

To implement the proposed work , we chosen the Matlab to simulate the parallel processing due to the availability of Matlab parallel Toolbox. The MATLAB has very good support for array and metric processing. Using PCT, the MATLAB can allow solving image processing problems using multi-core. processors, GPUs, and computer clusters[11].

To measure the performance of the system, the algorithm focused on parallel image processing. Since In traditional camera based OMR sheet analysis, the processing of one document is done using single processor. Whereas in proposed algorithm, parallel distribution of image data is considered. Four different OMR sheets are taken as a single image. After applying the discussed preliminary techniques , the enhanced image is converted into matrix data and this matrix data is further distributed on 4 parallel cores containing the questions and options of four different OMR sheet.

TABLE 1: Time analysis of phase1 and Phase2

No of OMR sheets	Phase1 Time(ms)	Phase2 Time(ms)
4	14	11
8	27	17
12	37	28
16	64	42

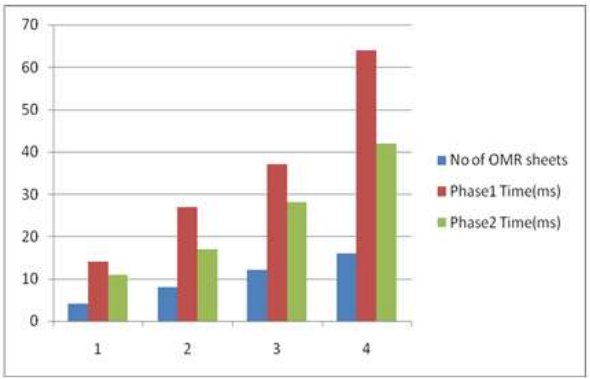

Fig.5 Time computation analysis in phase1 and phase2

TABLE 2 : Time analysis of sequential and parallel processing

No of OMR sheets	Sequential Time	Parallel Time
4	36	24
8	64	48
12	91	78
16	137	112

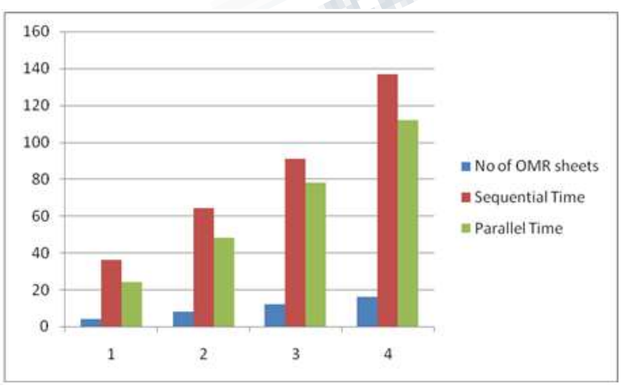

Fig.6 Time computation analysis of sequential and parallel

The time analysis of both sequential and parallel computation of 4 OMR sheet for both the phases of the proposed system is shown in table 2. Fig 5 and 6 show the considerable reduction the time consumption.

The proposed OMR processing algorithm is under implementation and will be analyzed with the data set of 200 OMR sheets. Also comparison will be done with traditional camera based OMR sheet processing through implementation. Hence, it is expected to give better results in time consumption and accuracy.

CONCLUSION

In this paper, OMR analysis algorithm is proposed, the primary objective of which is to reduce the computation time and utilize the computational power of multi-core processor by distributing the image matrix data on parallel cores for processing. The result shows the better performance than the traditional systems that uses sequential algorithm. It is expected to perform better with the data set of 200 OMR sheets. Also it is expected to get good accuracy even by processing blurred images.

REFERENCES

[1] Hui Deng, Feng Wang,Bo Liang, "A Low-cost OMR Solution for Educational Applications" International Symposium on parallel and Distributed Processing with Applications 2008 .

[2] Parinya Sanguansat, "Robust and Low-cost Optical Mark Recognition for Automated Data Entry," IEEE Transactions on, vol. 9, no. 1, pp. 62-66, 2015.

[3] Tien Dzung Nguyen, Quyet Hoang Manh, Phuong Bui Minh, Long Nguyen Thanh, Thang Manh Hoang,"Efficient and Reliable Camera based Multiple-Choice Test Grading System," IEEE International Conference on Advanced Technologies for Communications (ATC), 2011

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

[5] Sumit Tiwari, Sandeep Sahu "A Novel Approach for the Detection of OMR Sheet Tampering Using Encrypted QR Code" , 2014 IEEE International Conference on Computational Intelligence and Computing Research.

[6] D. Chai, “Automated marking of printed multiple choice answer sheets," IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), IEEE International Conference , pp. 145-149, 2016.

[7] J. A. Fisteus, A. Pardo, and N. F. Garc__a, “Grading multiple choice exams with low-cost and portable computer-vision techniques”, Journal of Science Education and Technology, vol. 22, no. 4, pp. 560-571, 2013.

[8] Mahmoud Afifi, Khaled F. Hussain, “The Achievement of Higher Flexibility in Multiple Choice-based Tests Using Image Classification Techniques”, arXiv.org,Cornell University Library , 2017

[9] Masakazu Iwamura, Tomohiko Tsuji, Akira Horimatsu, and Koichi Kise, “Real-Time Camera-Based Recognition of Characters and Pictograms”, 10th International Conference on Document Analysis and Recognition, 2009.

[10] Luis Miguel Sanchez, Javier Fernandez, Rafael Sotomayor, and J. Daniel Garcia, “A Comparative Evaluation of Parallel Programming Models for Shared-Memory Architectures”, IEEE International Symposium on Parallel and Distributed Processing with Applications,2012

[11] Harshad B. Prajapat, Dr. Sanjay K. Vij “Analytical Study of Parallel and Distributed Image” Processing, International Conference on Image Information Processing, (ICIIP),2011