

A
Mini Project
On
“OMR Marksheet Scanning using OpenCV”

Submitted by
Kedar Swami– TY-CSE-B-58
Yashraj Vibhute– TY-CSE-B-65
Siddhant Toggi– TY-CSE-B-61

In partial fulfilment for the award of the degree of
Bachelor of Technology

In
Department of Computer Science & Engineering



Pradnya Niketan Education Society, Pune.NAGESH
KARAJAGI ORCHID COLLEGE OF ENGINEERING &
TECHNOLOGY,SOLAPUR.

2022-2023



Pradnya Niketan Education Society, Pune.

NAGESH KARAJAGI ORCHID COLLEGE OF ENGG. & TECH., SOLAPUR.

Gut No. 16, Solapur-Tuljapur Road, Tale Hipparaga, Solapur – 413 002

Phone: (0217) 2735001/02, Fax. (0217) 2735004

Certificate

This is to certify that the Mini Project entitled “OMR Marksheet Scanning using OpenCV” is completed by the following students of TY. B. Tech CSE Class in satisfactory manner under the guidance of Prof. S.S.

More

Mr. Kedar Sanjaykumar Swami TY-CSE-B-58

Mr. Yashraj Shivanand Vibhute TY-CSE-B-65

Mr. Siddhant Santosh Toggi TY-CSE-B-61

The Mini Project is found to be complete in partial fulfillment for the award of Degree of Bachelor of Technology in Computer Science and Engineering of DBATU, Lonere.

Date of Submission:

(Prof. S.S. More)
Project Guide

(Dr. V. V. Bag)
Head of Department

ACKNOWLEDGMENT

We would like to express our gratitude towards our honorable principal Dr. J. B. DAFEDAR Sir and our honorable Head of Department Dr. V. V. Bag Sir and our guide Prof. S.S. More of N. K. Orchid College of Engineering and Technology, Solapur for their support in accomplishment of our project on **OMR Marksheet Scanning using OpenCV**. We take this opportunity to thank the members of the teaching and non-teaching staff of NKOCET Computer Science Engineering Department for the timely help by them. Lastly thanking our parents for their morale support and encouragement.

Thank You.

Abstract

The main objective of this project is to develop an Image processing based Optical Mark Recognition sheet scanning system. In today's world a whole lot of entrance exams are held as competitive exams which comprise MCQs. The students in such exams have to fill the right boxes or circles for correct answers. Usually, a stencil is given to the examiner to evaluate the correct answers to the questions during the examining phase. But this manual process can lead to a lot of errors such as counting mistakes. To make sure that this is avoided, an OMR system is used in which the OMR answer sheet is to be scanned and this scanned image of the answer sheet is taken in by the software system.

This project report describes the development of an OMR (Optical Mark Recognition) scanning system using OpenCV. OMR forms are used in various fields to gather and analyse data, and the system was designed to be efficient, accurate, and user-friendly. The system consists of an OMR scanner, image processing software using OpenCV, and a data storage and retrieval system. The scanner captures an image of the OMR form, and the OpenCV library is used to detect and read the marks on the form. The data is stored and accessed using the system's user interface. The system was able to accurately detect and read marks on OMR forms, even with different writing utensils or imperfections on the form, and it processed forms quickly, allowing for large volumes of forms to be processed in a short amount of time. Overall, the OMR scanning system with OpenCV provides a useful tool for processing and analysing OMR forms.

Index

Chapter Number	Title	Page Number
1	Introduction	6
2	Literature Survey	8
3	Proposed Work	10
4	Project Design and Technologies	15
5	Future Scope and Conclusion	19

Chapter 1

Introduction:

OMR (Optical Mark Recognition) scanning is a method of capturing data from forms or questionnaires using optical scanning technology. OpenCV is a powerful open-source computer vision library that provides tools for image processing and analysis, making it a popular choice for developing OMR scanning systems. OMR scanning through OpenCV involves capturing an image of the OMR form using a scanner or camera, and then processing the image to detect and read the marks made on the form. OpenCV provides various algorithms for image processing, such as thresholding, contour detection, and template matching, which can be used to detect and read the marks on the form. OMR scanning through OpenCV offers several advantages over traditional OMR systems, including greater accuracy in mark detection, faster processing time, and the ability to handle more complex forms with varying mark types and layouts. Additionally, OpenCV is an open-source library, making it cost-effective and customizable for specific OMR scanning needs.

In India, the multiple-choice questions have become an important part of the educational system. Important tests also use multiple-choice questions to judge the student's academic performance. The Indian Institute of Technology (IIT), the JEE MAINS and All India Pre-Medical Entrance Test (AIPMT) are just some of the many important tests conducted in INDIA to get the students on a common stage. Every year millions of students take these tests and they have to answer various questions asked by highlighting the circles in OMR sheets. Current situations for evaluating these OMR sheets are very overpriced and need dedicated scanners, OMR software and buying customized OMR sheets. So small organizations, institutes, individual teachers and tutors cannot use this easy method of grading without spending a lot of money. They resort to manually grading answer sheets. To check the OMR sheet test responses of a student takes 10 minutes on an average. The main theme of our project comes from the idea that we could build a website that will assist the instructors in auto grading these answer sheets and saving their precious time. Nowadays smartphones are very common. So, there will be no extra cost associated with using this smartphone-based solution. And it is very easy to use without a lot of steps or setup. We have a possibility that our application would do the following activities:

- (i) Automatically checks the answer sheet in the photo,
- (ii) Detect responses for each question and
- (iii) Compare student responses with reference answers.

Overall, OMR scanning through OpenCV provides a powerful and efficient solution for processing and analysing OMR forms, making it a popular choice for various applications, such as educational testing, surveying, and data collection. Optical Mark Recognition (OMR), also called mark sensing, is a technique to sense the presence or absence of marks by recognizing their depth (darkness) on sheet. A mark is a response position on the questionnaire sheet that is filled with pencil or ballpoint pen. The way of marking is simple to everyone and OMR devices can process mark information on sheets rapidly. Thus, OMR has been widely used as a direct input device for data of censuses and surveys and is fit for handling discrete data, whose values fall into a limited number of values. 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 a scanner. The forms contain small circles, referred to as bubbles, or boxes that are filled in by the respondent. Optical Mark Reader (OMR) reads marks written by pencil or ballpoint pen in the predefined positions on the answer sheet. The OMR can judge the existence of written marks by recognizing their darkness on the sheet. In the field of education, OMR technique is often used to process objective questions in the examination, such as the College Boards Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE) in the United States, and the College English Test (CET) in China. However, there are a few distinct drawbacks which limit the application of OMR technology.

Chapter 2

Literature Survey:

A paper regarding this project has proposed a new technique of OMR with ordinary scanners. It says that the given scanned image will be compared to its previously stored template and then on the foundation of the template and given criteria by the user, the answer sheet is marked. This makes the sheet evaluation technique very easy and competent. Different modules required for this system are: Answer Feed Module, Criteria Defining Module, Assessment Module, Result Storage Module. Four types of feasibility studies should be taken into account for this system, which are: Technical Feasibility, Operational Feasibility, Economic Feasibility and Schedule Feasibility

Another research suggested a novel technique for cost effective optical mark reader. Initially in this technique a graphical user interface (GUI) is used to design a custom-made form. Scanning of filled bubbles is done by Regular grayscale scanner. The information of filled bubbles is involuntarily input by processing the scanned images. This system is made up of two independent parts: 1) The interface by which the forms are designed and modified and 2) Reading of filled bubbles from the scanned form that is the recognition part.

Another work done is based on computer vision OMR sheet evaluation with the help of OPENCV. The webcam and its program takes the image when the OMR sheet is kept in front of it. The main objective of this work was to completely eradicate the use of ordinary scanners. OPENCV libraries which are used around the world are used to make a program to further process that image for the extraction of the optical marks. Many steps of image processing are required for this extraction.

1. "A Survey of Optical Mark Recognition Techniques" by R. Sharath Kumar and M. Anandi, in the International Journal of Computer Applications, 2011.

This paper provides an overview of various OMR techniques, including template matching, neural networks, and fuzzy logic. It discusses the advantages and disadvantages of each technique and compares their performance on different types of OMR forms.

2. "Optical Mark Recognition (OMR) System for Automated Examination System" by F. Mohamed and M. Hashim, in the International Journal of Computer Applications, 2010.

This paper describes an OMR system that was developed for use in an automated examination system. The system used a scanner to capture images of the OMR forms and an algorithm to process the images and detect the marks on the forms. The results showed that the system was able to accurately read marks on the forms, even with variations in the marks.

3. "A Novel Method for Optical Mark Recognition Using Convolutional Neural Networks" by H. Xu and L. Xie, in the Proceedings of the 11th International Conference on Natural Computation, 2015.

This paper describes a novel method for OMR using convolutional neural networks (CNNs). The method was able to achieve high accuracy in detecting and reading marks on OMR forms, even in cases where the forms had variations in the marks or layout.

4. "An Intelligent System for Automatic Grading of Multiple Choice Questionnaires Using OMR" by R. Sharma and V. Kumar, in the International Journal of Computer Applications, 2015.

This paper describes an intelligent system for automatic grading of multiple choice questionnaires using OMR

Chapter 3

Proposed Work:

Problem Statement1: While making this system, our first aim was to detect options (bubbles) which are the key concept in our project. To Detect Bubbles, we used OpenCV and the box containing the bubbles is detected using OpenCV algorithm.

3.1.1 System Design

This System is programmed using python and some native libraries. The system can be deployed on any operating system since the whole system will be done in python (platform independent). Aimed at the drawbacks of traditional OMR, our system will have Image tilts correction to improve the recognition precision and Easy Scanning Availability.

As a complex procedure, our system consists of four main parts: Image Acquisition, Tilt correction and mark recognition.

3.1.2 ALGORITHM AND IMPLEMENTATION

This stage will discuss detailed working of system, and will cover basic intuition of how system is actually working.

Step 1: Here take snapshot of answer sheet using webcam for this, it needs VideoCapture and NamedWindow function of OpenCV. Further when the paper is held in front of webcam then press button to take screenshot. This can be achieved using NamedWindow that check whether desired button is clicked by check its event's ascii code if yes then capture the frame of webcam and then write it to desired location.

Step 2: Next step is that there is no need of entire image. Only the answer sheet with little background to it is needed, hence one needs to crop the image. Here one may need opencv and key events. When left button is clicked and its position changes there is need to store all of its coordinates, after the button releases select the area of that coordinates that form rectangle and crop that particular part and while discarding the remaining other parts. This way the input for OMR system is generated.

Step 3: After having the filtered image, the next step will be to convert the image in grayscale format, since it will be helpful for easily detecting marked/ highlighted done by user. Later image will undergo high blur so that we can reduce high frequency noise, that means distortion of image.

Step 4: After converting to grayscale, next step will be to extract page from the given image. For this apply Canny Edge Detection technique it helps to find all edges within image, this algorithm is multi stage algorithm which goes through various steps and produces image which has all edges (horizontal and vertical) within it.

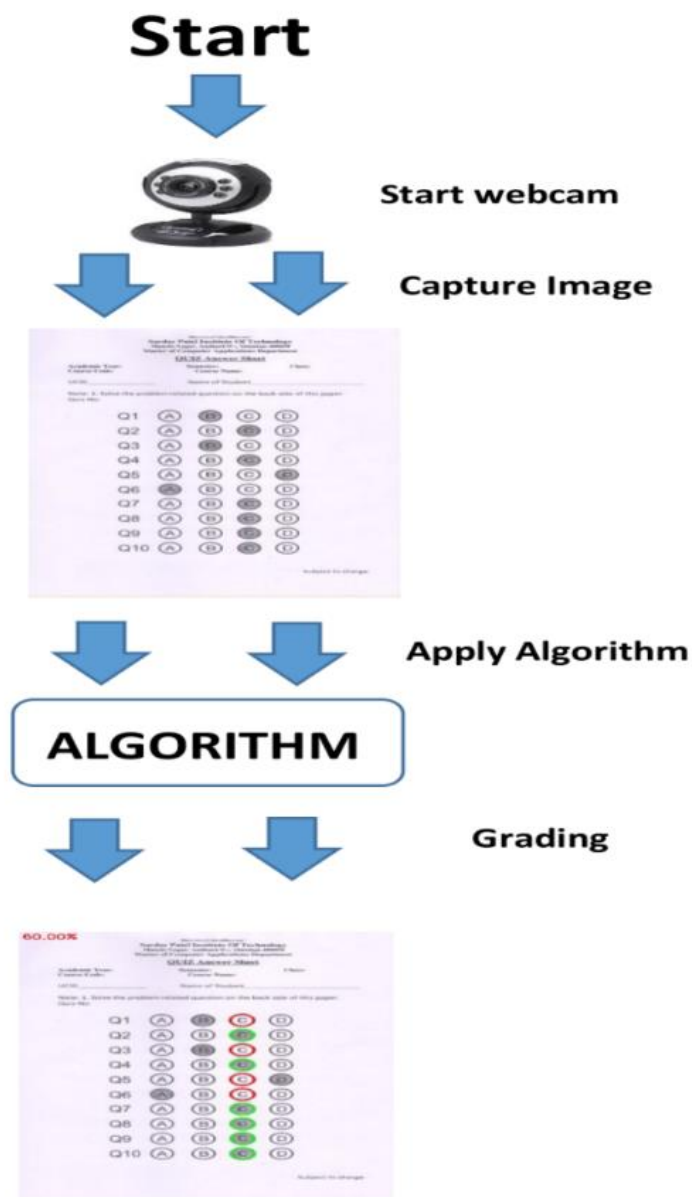


Fig. 1 Application Flow

Step 5: After getting outline, Apply transformation such that it give top-down, bird's eye view of previous image. This is done in order to get four border points of our exam paper. This find out this four points using open cv library “cv.findContours()”. Here it's mandatory to apply Canny Edge Detection before apply this step. As in this stage the aim is to search for white structure from black background and such transformation is only possible using previous step. In order to get only corner border points rather than a continuous line that form's a rectangle (page structure) pass parameter “cv.CHAIN_APPROX_SIMPLE”, It only extract corner point, thus removing redundant points.

Step 6: After getting contours, sort them, so that higher contour are in front of our array and small contour are at last, the basic intuition behind this logic is that the 4 border contour extracted are larger than the contour of object present within the image. Here answer sheet image should be main focal point within the entire image.

Step 7: In order to achieve top-down bird eye view use “four_point_transform()” function. This function is used for geometric transformation to extract only page image from entire image. There may be need to try various approach of getPerspectiveTransform() in order to get best suited result. The reason behind performing perspectiveTransform is to get 90-degree view of our image.

Step 8: Later do binarization of image in short converting image upto 256 grey level to black and white. It is used as pre-processing before any OCR or OMR logic applies. This can be done using Otsu's transformation. As one can see below is generated black and white image, this binarization is done using threshold() of open cv and passing Otsu's parameter to it.

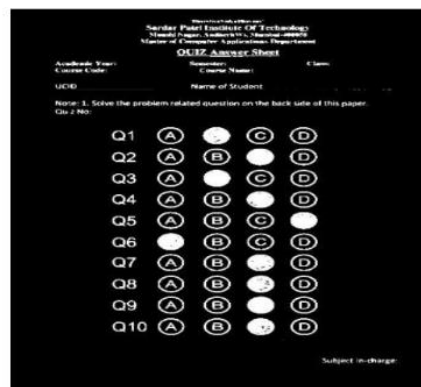


Fig. 2 Gray scale image

Step 9: Next step is to extract bubble options present within the answer page for this we again use perspective transformation and binarization, In order to find bubbles within image there are many ways, Here use HoughCircle() function of opencv which is used to highlight circle area within image. Here loop through contour to find all circles from left-right-top-down approach. This will save it in array where each index of array constitutes five options/contours of bubble that means each row is saved in each index of array, hence getting a nested array.

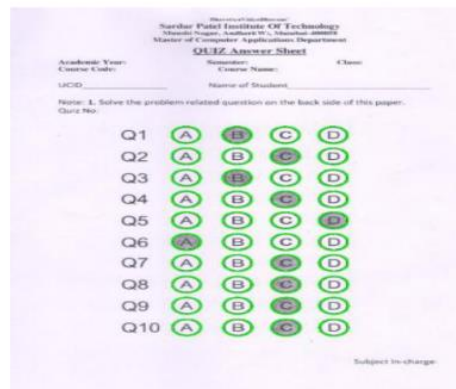


Fig. 3 Finding all bubbles

Step 10: Now next step is to perform function that check whether each answer highlighted is correct or not. Initially one has to provide array of our answer before. Now iterate through the contour array that was achieved earlier and compare each option with our initial answer array to check whether it is correct or not. But the initial step only found out the bubble contour, still one does not know which option is correct within them, after observing image of step 6 one can see the user highlighted or marker bubble option is white where as other options are black with white border. Again, how to find this white bubble? this is done using thresh () of open cv using its parameter one can define whether following bubble is marker or not. If for first row actual answer is 3 then within contour array what needs to be done is select first option arrcontour[0][0] check whether it is white bubble using thresh if not move to arrcontour[0][1], In such way move to next option until one reach end of that row or complete nesting of inner array [0,1,2,3,4] as there are 5 options. After that move to next index of bubble contour arrcontour[1][0] and again repeated previous step, for all index. In case actual and

thresh contour bubble match then increment a counter of correct array. Finally, it has final number of correct options.

Here the correct answer is marked with green border whereas wrong answer is marked with red border this can be done using drawContour() function of open cv.

N K Orchid College of Engineering
and Technology, Solapur

Name Roll No

1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

GRADE

Fig. 4 Final score

Step 11: After reaching last contour of last row, then invoke a function which reads a excel from particular location and the enters marks (score present in correct counter) in it. Then apply basic formula to calculate percentage and display it.

Chapter 4

Project Design and Technology

4.1 Requirement Analysis:

Requirement analysis is a critical first step in developing any software project, including OMR scanning through OpenCV. Here are some key requirements to consider when conducting a requirement analysis for this project:

1.Input: The first requirement is to determine the type of input that will be used for the OMR scanning process. Will the input be paper forms that are scanned. This requirement will help to determine the software needed for the project.

2.Image processing: The image processing requirement involves defining how the scanned images will be processed to extract the marks. This includes image pre-processing, such as noise removal and image enhancement, and feature extraction, which involves identifying the location and type of marks.

3.User interface: The user interface requirement involves designing a user-friendly interface that allows users to scan, process, and analyse the data. This may include features such as batch processing, data visualization, and error correction.

4.Data output: The data output requirement involves determining the format in which the data will be outputted. This may include database storage, spreadsheet export, or custom report generation.

5.Security: The security requirement involves ensuring the security of the data during the OMR scanning process. This may involve data encryption, access control, and secure data storage.

4.2 System Design:

4.2.1 Sequence Diagram

In below given Sequence Diagram, five classes are as mentioned. There are two modes of login 1) login for Faculty and 2) login for students. The Faculty login has access to use the OMR Scanning System and upload the checked marksheet to database, whereas the student login facility will be limited upto only Result Checking.

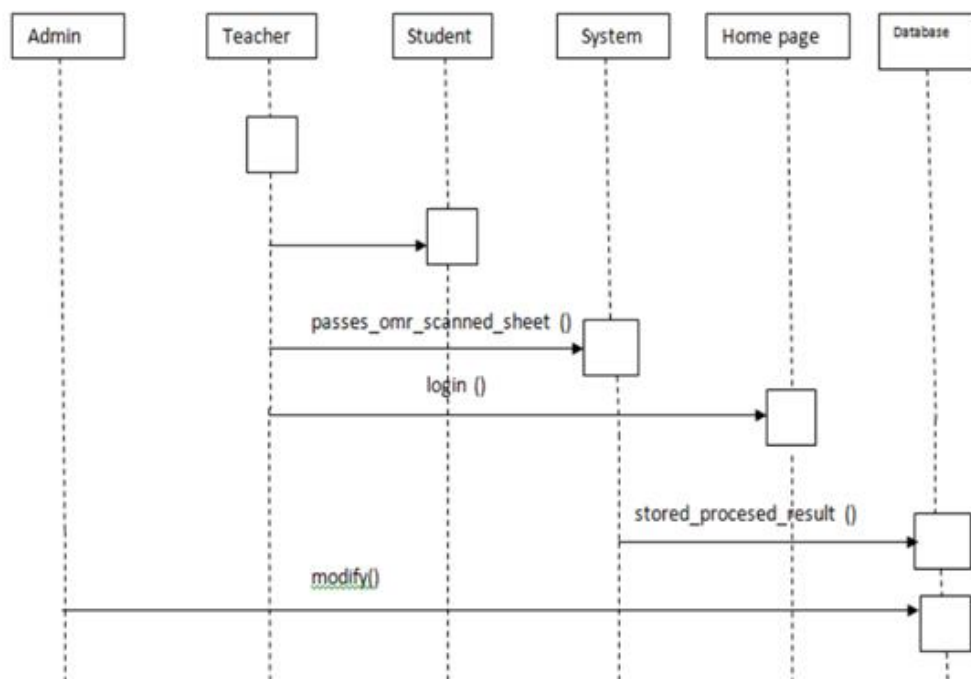
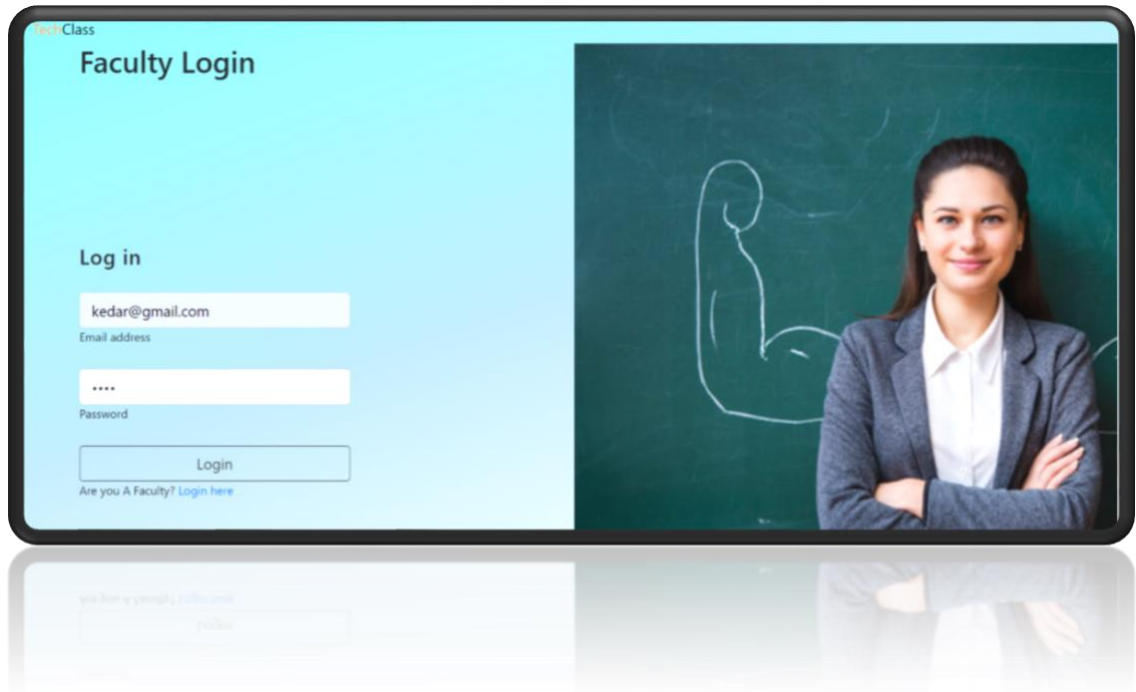


Fig 5 Sequence Diagram

4.2.2 Interface Glimpse:

Our Project GUI looks like the Screenshot given below:



Your Marks Are: 60

Here's Your Answer Paper

N K Orchid College of Engineering
and Technology, Solapur

Name Roll No

1	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

GRADE

Chapter 5

Future Scope and Conclusion:

5.1 Future Scope

Right now, this system is able to work for one semester students marks because of the lack of complex database model. But in future, it might be possible to create such a database with help of skilled developers. For detection of images, tiff images are used which are very memory consuming, but tiff images give better image quality and its detections. So in future, we can make use of less memory images which help to reduce memory wastage. Although we are scanning the images with the help of scanner devices, the hardware is much costlier. In the future we can take snapshots of sheets with the help of our mobile phone, but for this we need to scale the image in a more appropriate way. If in future all the restrictions of this software are able to be removed then it will be possible to implement it to overall Indian as well as world colleges.

5.2 Conclusion

This project is an application that performs OMR technique to extract students' marks using an ordinary scanner. It provides tools to the user to design an OMR sheet based on the layout they want. The design of the sheet will be stored as image format (Tiff). The user can take as many printouts as required, distribute it among others from whom information is desired, and get the lead sheets scanned. The scanned image files will then be provided as input to the software, processing will be done, value of lead fields will be extracted and then the data will be manipulated as instructed by the user and then storage of data will be done in the database. The OMR based marking system is a robust and low-cost OMR technique that can be adopted widely in developing countries. Partially resolves the drawbacks of traditional OMR technique and improves its usability. If this system is implemented successfully, the teachers of colleges will be capable of designing attendance sheets by themselves and the supervisor of colleges can adopt this technique to investigate the students attendance as well as their attention towards studies easily and quickly.

References

- [1] Optical mark reading: A. M. Smith, In Proceedings of the 9th annual ACM SIGUCCS conference on User services, United States, 1981, pp: 257263.
- [2] OCR and OMR Are Hotter Than Ever: Moore B, <http://www.idat.com/a-ocromr.html> (DOV 16 Sep, 2003)
- [3] Optical Mark Recognition: Bergeron B P, Postgraduate Medicine, 104(2): 23-24
- [4] Optical Mark Recognition-FAQ,<http://tev.itc.it/people/modena/kia-ocr-faq.html>.
- [5] ICR/OCR/OMR component for Win32 and .net: <http://www.pegasustools.com>
- [6] [http://en.wikipedia.org/wiki/Optical mark recognition](http://en.wikipedia.org/wiki/Optical_mark_recognition)