Checking multiple choice question exams

Houbakht Attaran
Lecturer in Khavaran high educational institute
Mashhad, Iran
attaran@khi.ac.ir

Fatemeh Sarani Rad
M.sc student of university of Economic science(Tehran)
Mashhad, Iran
f.sarani.rad@gmail.com

Shideh Nessari Ashkezari
M.sc student of Ferdowsi University of Mashhad
International Branch
Mashhad, Iran
Shideh.nessari@gmail.com

Marzieh Rezazadeh
Teacher of technical schools
Mashhad, Iran
m_rezazadeh_3876@yahoo.com

***Abstract*—the aim of this study is to explain an algorithm for identifying checked choices in multiple choices question exam. The target is considered, finding numbers of correct checked tests and recognizing students ID for each answer sheet, base on Black Square. This algorithm uses Morphological operations to eliminate noises and to find distinctive features of image in answer sheet. To identify checked choices, it uses the apparent properties of answer sheet. In final analysis section, system is tested in different situation and it shows that system is resistant to few image rotation and intensity changing, however it is sensitive to state of checking (square must be completely checked in black).**

Keywords-component; Image Rotation; Optical Mark Recognition; checking multiple choice question exams; Morphological Operation

I. INTRODUCTION

Considering when there has been many papers, checking multiple choices exam answer sheets takes many times and also causes mistake in manual way, an automatic system which works carefully is essential. Optical Mark Recognition (OMR) is a procedure of getting information by comparing optical reflex on the specific places on a paper.

One of the most important usages of OMR is checking multiple choices question exam .students choose the answer by blacking square choice on a printed paper. Then after scanning , these papers will be checked by special software automatically.

Because of its low price and easy work, OMR is an optimum and suitable way for checking multiple choice question exams [1]. identifying multiple choice soft wares are being used for along time , although every one has tried to design a better and more careful software , in the past few years, in some cases a collection of optical scanners are used to recognize black choices[2].

Generally there has been two algorithms for identifying checked choices [1, 2, 4, and 5]

The first one is pattern matching. One of its advantages is that this algorithm is not sensitive to noise and dirt, although it will be complicated if there is a few rotation, or fraction and different sizes and also this algorithm has a very complicated calculation.

The second one is combination of edge recognition, morphology and rejection error. This method is more secure and realistic. This article is written based on second algorithm which is included: scanning, preprocess, steps, basic identify, checked steps, and conclusion of different tests.

II. SCANNING (TAKING PICTURE)

At First the picture of answer sheet must be transferred to the computer. Answer sheets must scanned and saved by JPG format. Figure 1 shows an image of answer sheet which is scanned.

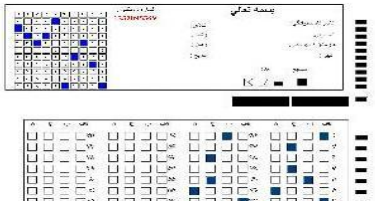

Figure1. Example of a SCANNED ANSWER SHEET

III. PRE PROCESS

The basic duty of this level is to improve image for optimization other process performance.

Output of this step is a binary picture. it include : converting image to gray scale image , normalizing the image size, normalizing intensity and converting image to binary image

A. converting image to gray levels

After scanning some operation must be done to get ready for next level. First step is to convert image to the gray levels. Which means convert picture to black and white with different percents of gray colors. This operation reduces the computation.

B. normalizing image size

Images that are received as input may have different sizes, so after converting, all images must have the same size.

For this purpose, all images size change to 1169*865. Image may become smaller or bigger. After this operation image will have increase or decrease in size. For decreasing image size neighbor pixels must be deleted and for increasing size neighbor pixels must be added.

C. normalizing light

While papers are scanning, images may be captured with different intensity. In order to apply a general algorithm for all received images, they must be normalized .This means that brightness and darkness of all images must be equal.

At first average and variance of image have to calculated Up to specific brightness to determine the current level of image brightness?

Then use formula number 1 for adjusting brightness and darkness to normal level.

$$g(x,y) = A(x,y) * [f(x,y) - m(x,y)] + m(x,y) \quad (1)$$

$$A(x,y) = K\frac{M}{g(x,y)} \quad 0<K<1 \quad (2)$$

In equation 1 “f” is the initial gray level for every pixel, and “σ “and “m” are local average and variance around desired pixel.

In equation 2 “m” is image global average and “k” is a constant factor. ”g” includes initial image normalized pixels [3].

Number 0.989 is considered for “k” in this formula.

D. binary an image

After normalizing image by the written method we can convert an image to a binary image by considering a threshold level as constant ‘T”. This constant is the computed threshold level for an ideal image. This threshold value was calculated based on normalized image. In this pamphlet, the value of ‘T’ is considered as 0.4980. After pre processing steps we got the figure number 2.

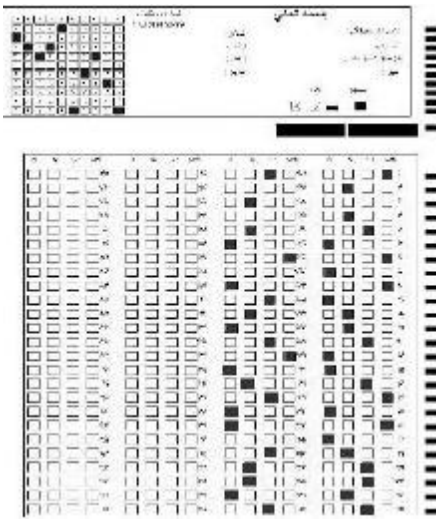

Figure2. IMAGE AFTER USING PREPROCCESS

IV. RECOGNITION OF CHECKED CHOICES

A. eliminate image noise

While paper is scanning, some images may be damaged for some reason such as:

Inappropriate scanning, line corrosion or a dirty answer sheet which makes some extra lines.

Because software needs an image without noise we should omit noises and extra edges .Things that must be processed are marked choices, distinctive feature of image, rectangles which are on the right side of the paper and student ID. This operation is morphological operation.

First erosion operation on the preprocess image is done by using a 4*4 square structural element because choices have square like.

Noises and extra edges are deleted by erosion. After this step, image might become too small. So after erosion, dilation operation must be used the same 4*4 structural element.

Dilation operation will be done again but this time we use a horizontal line structural element .by using this element two rectangles which are on top of the paper will be combined. This combined rectangle is used for distinctive feature of image.

This distinctive feature is considered as origin of coordinate. And other objects places are recognized by this element.

Finally erosion is done again by using the same horizontal item. Now we got an image without any noises or extra edges.

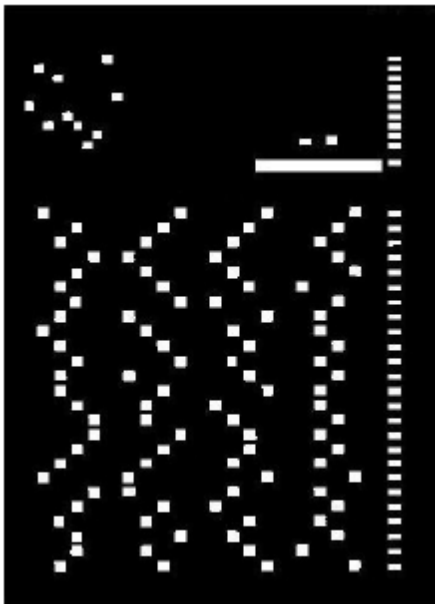

Figure3. IMAGE AFTER REMOVING NOISES

B. *labeling linked component*

For finding objects after deleting noises, labeling will be used.

C. *recognition of distinctive feature of image*

We can use information of the last level (labeling) which are specified by labeling to find distinctive feature of image. The used attributes are:

Coordination of the inscribed rectangle, angle and number of pixel.

Those objects that their length to width ratio is more than constant number D1 and less than constant number D2 will be looked for. Also have bigger pixel range than constant number A.

D1 and D2 are two constant numbers that their values are calculated based on distinctive feature length to width ratio in answer sheet.

'A' is the number of distinctive feature pixel in answer sheet; we set D1, D2 and A equal to 8, 11, 4000 respectively.

D. *Image Rotation*

While scanning Image might rotate, to fix image we must compute the rotation angle, then we rotate image in opposite way.

After finding distinctive feature of image, angle between large diameter of inscribed oval and horizontal axis is calculated .then image rotates in opposite direction of that calculated angle. Because all objects will change their position we need to do labeling level one more time.

E. *Edge recognition*

According to figure 1 there is a small rectangle for each line .these rectangles are used to find question lines. For example the questions (1, 26, 56, 76) are at the first line.

Also at the top of the page there are some rows that each of them shows a line of numbers. For example first rectangle shows zero, second one is 1 and so on. These numbers are used to get student's ID.

Considering that distinctive feature is found, in order to find rectangles corresponding to questions, among image objects the objects whose inscribed rectangle column (X) and row (Y) are bigger than distinctive feature column and row respectively will be searched and these objects are sorted according to their rows (Y) ascending order.

Also to find rectangles corresponding to student's ID among other objects, those object will be searched that their column (X) is bigger than distinctive feature column and their row (Y) is smaller than its row. Then rectangle will be sorted according to their rows (Y) ascending order. Access to each row is easier with this type of sorting.

F. *Calculation of ID*

After reaching ID rectangle, the ID number of answer sheet can be reached as follow:

At the first the number of pixel are calculated in one centimeter (unit of measure)

$$Pixel=Lx/Lx1 \qquad (3)$$

Lx is the length of inscribed rectangle distinctive feature of image and Lx1 is length of distinctive feature on answer sheet.

Then each object that the column of its center (x) range is between X1 and X2 is considered as Ith number in ID according to coordinate of its column (I=1… 10), figure 4.

$$X1 = (Pixel * 17.5)$$

$$X2 = (Pixel * 12.5)$$

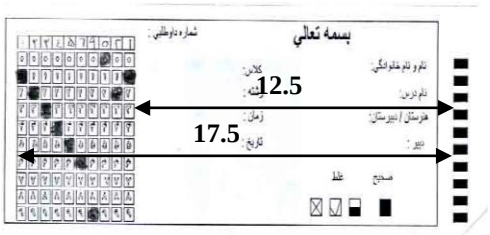

Figure4. SPACE BETWEEN SYMBOLS AND ID

G. *checked choices*

In order to calculate checked choices, those objects are recognized as checked choices whose the column (X) of their center is equal to center column of one of the rectangles of the right side of paper then according to the coordinate of their center row (Y), the row that they are in it , will be found (rows 1-25).

Finally according to their distance to the distinctive feature position in that specific line, it is determined that checked choice is related to what question and what choice in that question (A, B, C, D).

V. TEST AND CONCLUSION

The explained algorithm was tested and run on more than 30 different answer sheets and effective parameters were found.

The effective parameters are:

1. scanning
2. checking method

One of the most effective parameters is scanning method based on different test if scanning did not perform correctly it raises number of errors.

These are some issues while scanning might cause mistakes:

1. Image rotation: if picture rotates with large angle, it damages badly so that even by image rotation procedure it is useless.
2. Wrong resolution for scanning: Low resolution destroys physical feature.

Another parameter is checking choices incorrectly

As it is written on the top of answer sheet, choices must be completely and boldly black.

These parameters raise errors while checking:

1. less filling option
2. incomplete filling option

All conclusion of different situation are mentioned in table 1 to table 3:

Table 1: SYSTEM ERROR PERCENTAGE FOR RIGHT SCANNING

correct scanned paper	Error percentage	Unrecognized option	Recognized option	All options	Number of pages
correct checked paper	0.84	5	595	600	15
correct unchecked paper	10	60	540	600	15

Table 2: SYSTEM ERROR PERCENTAGE FOR WRONG SCANNING BUT CHECKED CORRECTLY

Incorrect scanned paper	Error percentage	Unrecognized option	Recognized option	All options	Number of pages
Pale scanned paper	16.6	100	500	600	15
Rotation angle paper between -10 to 10	33.33	200	400	600	15

Incorrect scanned paper	Error percentage	Unrecognized option	Recognized option	All options	Number of pages
Colorless scanned papers	2	12	588	600	15
rotation angle paper between -10 to 10	25	150	450	600	15

Table 3: SYSTEM ERROR PERCENTAGE FOR INCORRET SCANNED AN CHECKED PAPERS

VI. CONCLUSION

According to the conclusion, it is obvious that the most errors happen when a paper had twirled during scanning, because in this situation some part of paper are eliminate so system can recognize that part. As mentioned before if image has twirl with large angle we cannot fix the picture by image twirl.

At the end we can emphasize that system is resistant and if choices checked correctly, system has high optimum.

Although it is slower than OMR scanners, this system perform correctly as users wish.

REFERENCES

[1] S. Haag and M. Cummings and D. McCubbrey and A. Pinsonnault and R. Donovan, "Management Information Systems for the Information Age" (3rd ed.).Canada: McGraw-Hill Ryerson, 2006.

[2] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Prentice- Hall. Inc, Upper Saddle River. NJ, 2006.

[3] M.B. Ahmad and T.S. Choi , "Local Threshold and Boolean Function Based Edge Detection", IEEE Transactions on Consumer Electronics, Vol. 45, No 3.August 1999.

[4] S. Belongie, J. Malik and J. Puzicha. "Shape matching and object recognition using shape contexts", PAMI, 24(4):509–522, 2002.

[5] Gorgevic Dejan, Grcevski Nikola, Mihajlov Dragan, "A Simple System For Automatic Exam Scoring Using Optical Markup Reader", Second Balkan IFAC International Symposium APPLIED AUTOMATIC SYSTEMS AAS'2000.

[6] Qu Gongyuan, "Directional Morphological Gradient Edge Detector", PhD Thesis, Santa Clara University, 2002.

[7] Csaba Rekeczky, Abraham Schultz, István Szatmári,Tamás Roska , O. Chua. Leon, "IMAGE SEGMENTATION AND EDGE DETECTION VIA CONSTRAINED DIFFUSION AND ADAPTIVE MORPHOLOGY: A CNN APPROACH TO BUBBLE-DEBRIS IMAGE ENHANCEMENT", 1997.

[8] M.A. Oskoei and H. Hu, "A survey on edge detection methods", Technical Report CES-506, University of Essex, Essex, 2010.

[9] Helen tillias Anderson, "EDGE DETECTION FOR OBJECT RECOGNITION IN AERIAL PHOTOGRAPHS", Master thesis, Philadelphia, Pennsylvania, 1986

[10] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.