Optical Mark Recognition

By:

Danyal Faheem

19I-2014 CS-19

Mehmood Amjad

19I-0472 CS-19

Muhammad

19I-0561 CS-19

Digital Image Processing

Dr. Akhtar Jamil

Table of Contents:

Introduction:

Using the basics of computing which is to automate a tedious and tiring task that was done manually beforehand, we discovered a very common problem scenario. Multiple Choice Question examinations used to be checked manually. Doing so would require manpower to manually look at each and every answer and compare it to an answer key. This process is very time consuming and error prone as human beings are not perfect.

To make this task quicker and less error prone, we realized that we could use computers to check the answer sheets for us, compare it to an answer key and then log the records to a database. 3 steps that had been done manually would now be automated and done by the computer for us.

Literature review:

https://staff.informatics.buu.ac.th/~krisana/cv/paper/spie99.pdf

https://ijeir.org/administrator/components/com_jresearch/files/publications/IJEIR_2338_FINAL.pdf

https://www.researchgate.net/publication/330977246_An_Image_Processing_Oriented_Optical_Mark_Recognition_and_Evaluation_System

Data Set:

Considering we recognized the problem in our university, we used another university's pre-existing dataset of multiple choice questions answer sheets to use as our input images. We would use these images as our sample input and write our program in such a way that it can use these images and produce the desired output.

The sample dataset for some of the pictures can be seen below:

İstanbul Sabahattin Zaim Üniversitesi

CEVAP KARTI

ADI ve SOYADI (Adı ve soyadı arasında bir boşluk bırakınız.)

MERVE ASIKUZUNOĞLU

ÖĞRENCİ NUMARASI

6	0	7	0	8	1	7	2	0

SINAV TÜRÜ

○ QUİZ ○ ARA ● FİNAL ○ BÜTÜNLEME

SORU GRUBU	● B C D E
Öğretim Elemanı Onayı	● B C D E

ÖĞRENCİ BİLGİLERİ

Adı	: Merve
Soyadı	: ASIKUZUNOĞLU
Numarası	: 607081720
Programı	: Yazılım mühendisliği
Dersin Adı	: Makine Öğrenmesi
Tarih	: 18 / 02 / 2020
İmza	

DİKKAT!
Optik form üzerindeki tüm işaretlemelerinizde YALNIZ KURŞUN KALEM KULLANINIZ.

DOĞRU İŞARETLEME ÖRNEĞİ

Başarı Dileklerimizle.

CEVAPLAR

1	B	C	D	31	A	C	D	E	61	A	B	D	E	91	A	B	C	E
2	B	C	D	32	A	B	C	E	62	A	B	C	D	92	A	C	D	E
3	B	D	D	33	A	C	D	E	63	A	C	D	E	93	A	B	C	D	E
4	B	C	D	34	A	B	D	E	64	A	B	C	E	94	B	B	D	E
5	B	C	D	35	B	C	D	E	65	B	C	D	E	95	A	B	C	D	E
6	A	B	C	D	E	36	A	D	E	66	A	C	D	E	96	A	B	C	D	E
7	B	C	D	E	37	A	B	C	D	E	67	B	C	E	97	A	B	D	E
8	B	C	D	E	38	A	C	D	E	68	A	C	D	E	98	B	C	D	E
9	B	D	E	39	A	B	C	D	69	A	B	C	D	99	A	B	C	D	E
10	B	C	D	E	40	A	C	D	E	70	A	C	D	E	100	A	C	D	E
11	B	C	D	E	41	A	B	D	E	71	B	C	D	E	101	A	B	C	E
12	A	B	C	D	E	42	B	C	D	E	72	A	B	C	E	102	B	C	D	E
13	B	D	E	43	A	C	E	73	A	C	D	E	103	A	B	C	E
14	B	C	E	44	A	B	C	D	E	74	A	B	C	E	104	A	B	C	D	E
15	B	D	E	45	A	C	D	E	75	A	C	D	E	105	A	C	D	E
16	B	C	D	E	46	A	B	D	76	A	B	C	D	E	106	A	B	C	D	E
17	B	D	E	47	A	C	D	E	77	B	C	E	107	A	B	D	E
18	A	B	C	D	E	48	A	C	E	78	A	B	C	D	E	108	B	C	D	E
19	B	C	D	49	B	C	D	E	79	A	C	D	E	109	A	B	C	E
20	A	C	E	50	A	B	D	E	80	A	B	C	D	E	110	A	B	C	D	E
21	A	B	D	E	51	A	B	C	E	81	A	B	C	D	E	111	C	X	X
22	A	B	C	D	E	52	B	C	D	E	82	C	X	X	112	B	C	D	E
23	B	E	53	B	C	D	E	83	A	B	C	E	113	X	X	D	E
24	B	C	D	E	54	A	C	E	84	A	B	D	E	114	B	C	D	E
25	B	D	E	55	A	C	D	85	A	X	C	D	E	115	C	X	X
26	B	C	D	E	56	B	C	D	E	86	X	C	116	A	B	C	D	E
27	B	C	D	E	57	A	C	D	E	87	B	C	D	E	117	B	C	E
28	A	B	C	D	E	58	A	B	C	D	88	A	B	D	E	118	A	C	D	E
29	A	B	C	D	E	59	B	C	D	E	89	A	B	C	E	119	A	B	C	E
30	A	B	C	D	E	60	A	B	C	E	90	B	C	D	E	120	B	C	D

Form Kodu: Y

xc/50b/a4

İstanbul Sabahattin Zaim Üniversitesi

CEVAP KARTI

ADI ve SOYADI (Adı ve soyadı arasında bir boşluk bırakınız.)

EDANUR ANLAYAN

ÖĞRENCİ NUMARASI

0 3 0 7 1 6 0 0 4

SINAV TÜRÜ

○ QUİZ ○ ARA ● FİNAL ○ BÜTÜNLEME

SORU GRUBU	● Ⓑ Ⓒ Ⓓ Ⓔ

Öğretim Elemanı Onayı	● Ⓑ Ⓒ Ⓓ Ⓔ

ÖĞRENCİ BİLGİLERİ

Adı	Edanur
Soyadı	ANLAYAN
Numarası	030716004
Programı	Yazılım Mühendisliği
Dersin Adı	Makine Öğrenmesi
Tarih	18 / 02 / 2020
İmza	Edanur

DİKKAT!
Optik form üzerindeki tüm işaretlemelerinizde YALNIZ KURŞUN KALEM KULLANINIZ.

DOĞRU İŞARETLEME ÖRNEĞİ

Başarı Dileklerimizle.

CEVAPLAR

1	●	31	●	61	●	91	●
2	Ⓐ	32	Ⓐ	62	●	92	Ⓐ
3	Ⓐ	33	Ⓐ	63	●	93	Ⓐ
4	Ⓐ	34	Ⓐ	64	Ⓐ	94	●
5	Ⓐ	35	Ⓐ	65	Ⓐ	95	Ⓐ
6	Ⓐ	36	Ⓐ	66	Ⓐ	96	Ⓐ
7	●	37	Ⓐ	67	Ⓐ	97	Ⓐ
8	Ⓐ	38	●	68	●	98	Ⓐ
9	Ⓐ	39	Ⓐ	69	Ⓐ	99	Ⓐ
10	Ⓐ	40	Ⓐ	70	Ⓐ	100	●
11	Ⓐ	41	●	71	Ⓐ	101	Ⓐ
12	Ⓐ	42	Ⓐ	72	Ⓐ	102	Ⓐ
13	Ⓐ	43	Ⓐ	73	Ⓐ	103	Ⓐ
14	Ⓐ	44	Ⓐ	74	●	104	Ⓐ
15	Ⓐ	45	●	75	Ⓐ	105	●
16	Ⓐ	46	Ⓐ	76	Ⓐ	106	Ⓐ
17	Ⓐ	47	Ⓐ	77	Ⓐ	107	Ⓐ
18	Ⓐ	48	Ⓐ	78	●	108	Ⓐ
19	●	49	Ⓐ	79	Ⓐ	109	●
20	Ⓐ	50	Ⓐ	80	Ⓐ	110	Ⓐ
21	●	51	Ⓐ	81	Ⓐ	111	Ⓐ
22	Ⓐ	52	●	82	●	112	Ⓐ
23	Ⓐ	53	Ⓐ	83	Ⓐ	113	Ⓐ
24	●	54	Ⓐ	84	Ⓐ	114	●
25	Ⓐ	55	Ⓐ	85	●	115	Ⓐ
26	Ⓐ	56	●	86	Ⓐ	116	Ⓐ
27	Ⓐ	57	Ⓐ	87	Ⓐ	117	Ⓐ
28	Ⓐ	58	Ⓐ	88	Ⓐ	118	Ⓐ
29	Ⓐ	59	Ⓐ	89	Ⓐ	119	Ⓐ
30	●	60	Ⓐ	90	Ⓐ	120	●

Form Kodu: Y

xc/50b/a4

İstanbul Sabahattin Zaim Üniversitesi

CEVAP KARTI

ÖĞRENCİ NUMARASI

0 3 0 1 1 6 0 0 4

ADI ve SOYADI (Adı ve soyadı arasında bir boşluk bırakınız.)

H A L İ L İ B R A H İ M S E Z E R

SINAV TÜRÜ

○ QUİZ ○ ARA ● FİNAL ○ BÜTÜNLEME

SORU GRUBU	● B C D E
Öğretim Elemanı Onayı	● B C D E

ÖĞRENCİ BİLGİLERİ

Adı : Halil İbrahim

Soyadı : SEZER

Numarası : 030116004

Programı : B.Ü. Sayısal M.H.

Dersin Adı :

Tarih : / /

İmza :

DİKKAT!
Optik form üzerindeki tüm işaretlemelerinizde YALNIZ KURŞUN KALEM KULLANINIZ.

DOĞRU İŞARETLEME ÖRNEĞİ

Başarı Dileklerimizle.

CEVAPLAR

1	●	31	A	61	A	91	A
2	●	32	A	62	A	92	A
3	A	33	A	63	A	93	A
4	A	34	A	64	A	94	A
5	A	35	A	65	A	95	A
6	A	36	A	66	A	96	A
7	A	37	A	67	A	97	A
8	●	38	A	68	A	98	A
9	A	39	A	69	A	99	A
10	A	40	A	70	A	100	A
11	A	41	A	71	A	101	A
12	A	42	A	72	A	102	A
13	A	43	A	73	A	103	A
14	A	44	A	74	A	104	A
15	●	45	A	75	A	105	A
16	A	46	A	76	A	106	A
17	●	47	A	77	A	107	A
18	A	48	A	78	A	108	A
19	●	49	A	79	A	109	A
20	A	50	A	80	A	110	A
21	A	51	A	81	A	111	A
22	A	52	A	82	A	112	A
23	A	53	A	83	A	113	A
24	A	54	A	84	A	114	A
25	A	55	A	85	A	115	A
26	A	56	A	86	A	116	A
27	A	57	A	87	A	117	A
28	A	58	A	88	A	118	A
29	A	59	A	89	A	119	A
30	A	60	A	90	A	120	A

Form Kodu: Y

xc/50b/a4

Methodology:

Our methodology is somewhat simple in this regard. From the dataset, we can see that each part of the image has different segments that are constant. The name of the student is in the top right corner. The roll number is somewhat in the top left corner. The answers for the respective questions are in the bottom right corner.

We can therefore make 3 separate images out of this image that are separated into 3 segments of name, roll number and answers. This will make our work easier.

Taking an example from below image, we can have 3 separate images

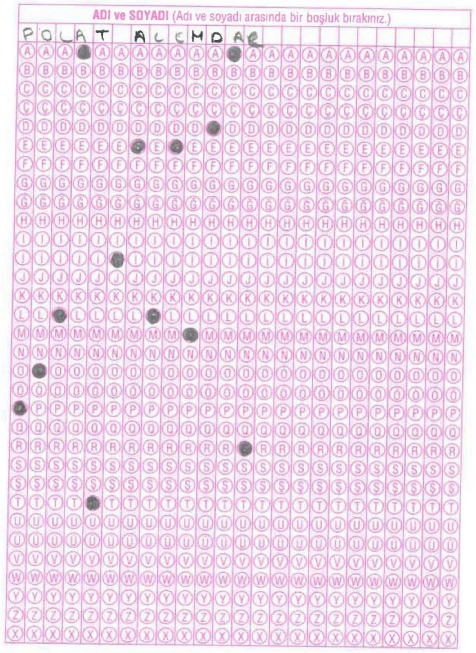


Image of Name bubbles

Image of Roll Number

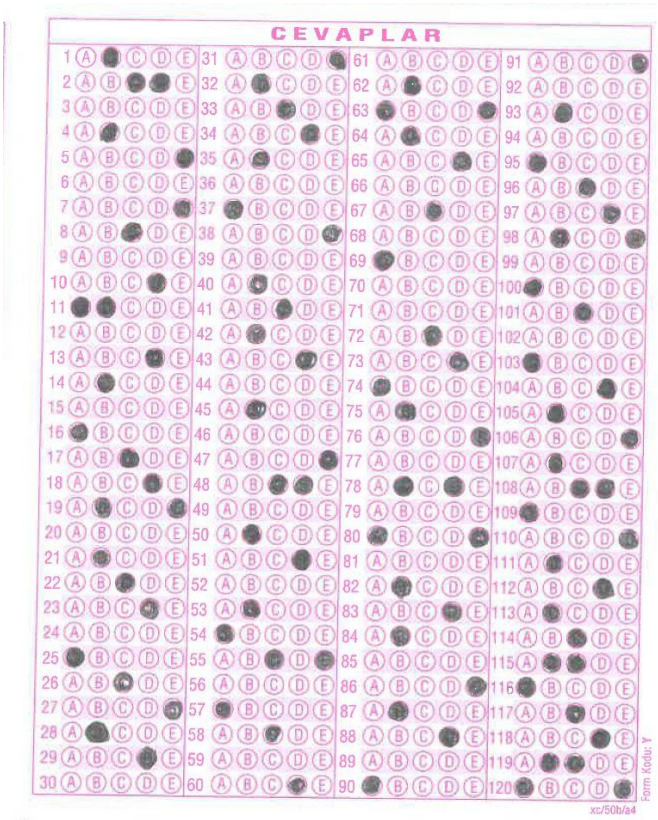


Image of Answer Choices

The next step in the program would be to convert the image to grayscale as this would highlight the spots that have been marked in pencil. Pencil has a generic colour of grey that would be emphasized by converting the image to grayscale.

After this, we can apply thresholding on the resultant images to enhance the marked dots on the sheet. This will help us in clearly distinguishing between the empty circles and the filled circles.

Continuing on, as some circles are smaller than others, we will apply some Morphological Processing to make them stand out more.

Then we will calculate the contours of the image. This will provide us with the locations of each filled circle. Using the contours to find the bounding rectangles we can get the width and height of each filled circle. Calculating the average of this height and width, we can multiply the width by 5 (for 5 options each) to get a window size of row width x row height.

Using this window size, we can break the image into each individual row of 5 options each. This will allow us to get a better look at each option. Using this method, we can get all the separate rows or columns (for roll number and name sheets).

Once we have all the separate we can calculate the number of contours in each row. If it is greater than 3 or less than 2, it means it is an invalid answer and will be awarded a -1 option. For an exact 2 contours, we will calculate the distance between the start till the contour using the row width we calculated first.

Then we will store these results in our database. After this, we can match these values to the answer key already stored in the database and then keep a record of the total marks of the students.

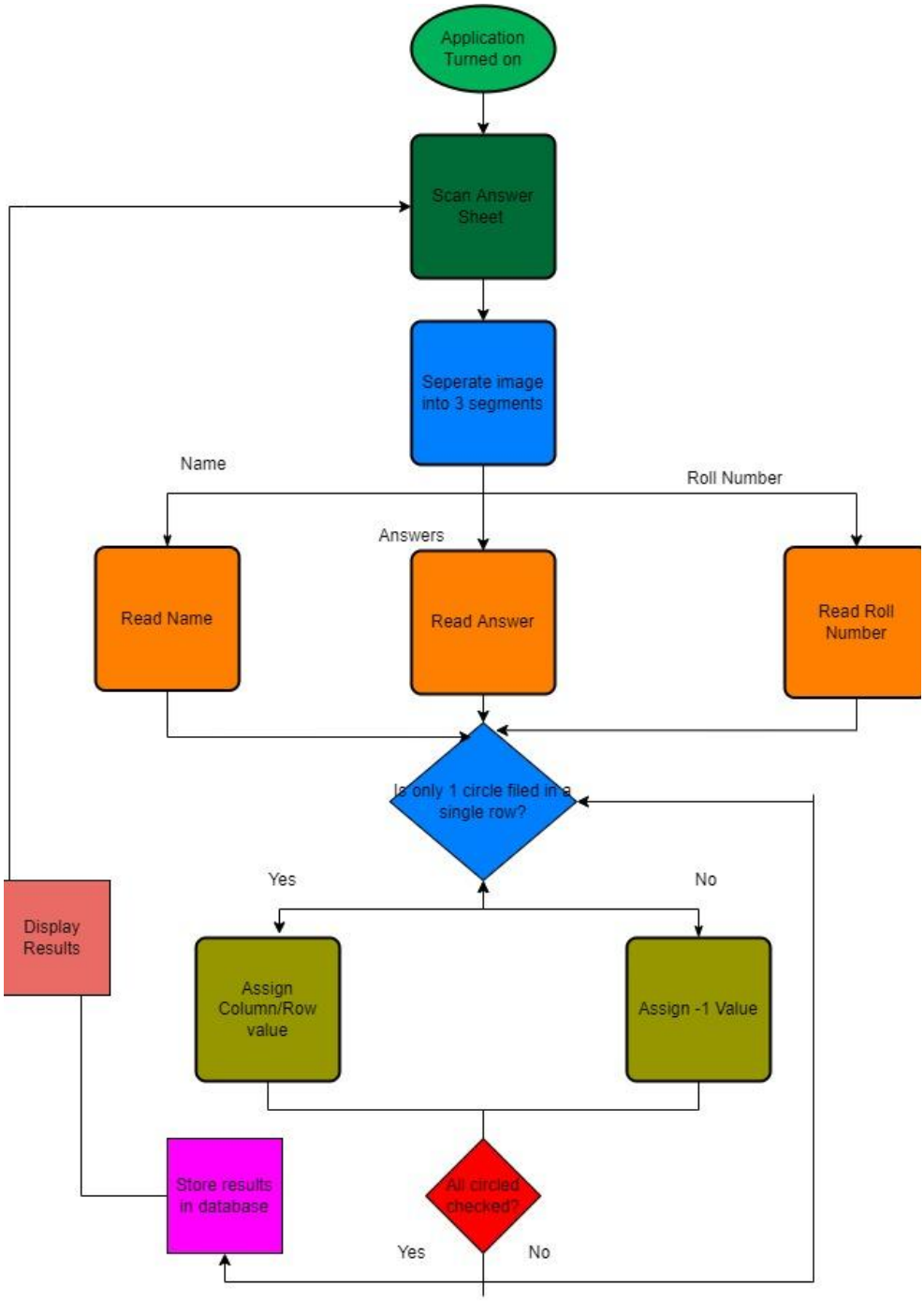
An example of the circles counting and their respective values is as below:

Answers along with Value assigned to each row

As can be seen, the rows with 2 or more circles filled or 0 circles filled, they will be given -1 value whereas 1 circle filled will be given the value of their respective column starting from 0.

Flowchart:

We have also made a flowchart depicting how our program functions according to the input. The link for the flowchart is here for better viewing.



Software Packages:

Name	Tasks
Danyal Faheem	Image Processing, Row/Column Separation, Database Write
Mehmood Amjad	Image Segmentation, Frontend, Database Read
Muhammad	Answer Detection, Answer Value Extraction, Roll Number Value Extraction

Bibliography:

https://staff.informatics.buu.ac.th/~krisana/cv/paper/spie99.pdf

https://ijeir.org/administrator/components/com_jresearch/files/publications/IJEIR_2338_FINAL.pdf

https://www.researchgate.net/publication/330977246_An_Image_Processing_Oriented_Optical_Mark_Recognition_and_Evaluation_System

Thank you!