Journal of Scientific Research & Reports

26(3): 53-59, 2020; Article no.JSRR.55980
ISSN: 2320-0227

Answer Sheet Recognition System Design

Guanyang Gao[1*] and Yaoyao Wang[2]

[1]School of Mechanical Engineering, North China University of Water Resources and Electric Power, Zhengzhou, 450045, China.
[2]School of Hydraulic Engineering, Changsha University of Science and Technology, Changsha, 410114, China.

Authors' contributions

This work was carried out in collaboration between both authors. Author GG designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author GG managed the analyses of the study. Author YW managed the literature searches. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2020/v26i330235
Editor(s):
(1) Dr. Kleopatra Nikolopoulou, University of Athens, Greece.
(2) Dr. Ana Paula da Silveira Simões Pedro, University of Aveiro, Portugal.
Reviewers:
(1) Raidah Mahirah, STAI Al-Gazali Bulukumba, Indonesia.
(2) Pasupuleti Venkata Siva Kumar, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/55980

Short Research Article

Received 09 February 2020
Accepted 15 April 2020
Published 21 April 2020

ABSTRACT

Examination is an indispensable way for students to verify their learning achievements and select talents. With the continuous development of computer technology, especially the emergence of answer sheets, computers have become the mainstream instead of manual examination. The main research content of this design is to use MATLAB software to identify and obtain the information in the answer sheet, import the completed answer sheet into the system, and use MATLAB software to preprocess the image to obtain an image that can be identified later. Through the system to identify the information in the answer card and compare the correct answers, finally realize the identification of the candidate's student number, identification of subjects, reminders of missing information, and calculation of scores and other functions. This article designed the GUI interface, which provided convenience for human-computer interaction, making the system operation more convenient and the results clear.

Keywords: MATLAB; GUI; answer sheet recognition; image processing.

**Corresponding author: E-mail: 774781041@qq.com;*

1. INTRODUCTION

With the rapid development of modern education, examinations have become the most equitable selection method in modern education, and also the decisive means of teaching and learning quality inspection. Regardless of whether you are looking for a job or a promotion within an organization, the exam is one of the important references for selection. Manual examinations have the disadvantages of large workload, heavy tasks, and inevitable errors. The appearance of answer sheets has attracted the attention of teachers and students. One of the results of the current response is the computer automatic examination. Its reliability directly determines the fairness, impartiality and accuracy of the selection results, which increases the reliability of the automatic examination and increases the credibility of the selection. Through the recognition of the student's answer sheet through the computer, you can also store the student's results in the computer directly used by the school, and manage it more efficiently through data processing, which is also convenient for long-term storage and future data analysis. Answer sheet image recognition technology can easily get the answer card score, which can then be applied to different teaching and research systems for data sharing and can also be connected to many information technologies, which is also a very important function [1].

With the variety of test questions, especially the English test, selective test, psychological test, etc. In the test papers, the choice of questions accounted for a large proportion. Single-choice and multiple-choice questions are more objective, so many people begin to use computer to mark papers instead of manual markup to analyze the test papers, making the papers more efficient and accurate. It also laid a solid foundation for future computer marking [2].

2. ANSWER SHEET IMAGE PROCESSING

First of all, digital images need to be pre-processed, mainly including image transformation, image filtering, image correction and other measures that can greatly enhance the effect of pictures. For different needs, we can pre-process the pictures purposefully. In a word, image preprocessing plays an important role in digital image processing [3-4].

2.1 Image Gray Processing

The images people usually see are composed of the three primary colors RGB. In MATLAB software, it is composed of a $\mathrm{N} \times \mathrm{M} \times 3$ data matrix. Each element is replaced by pixels with blue, red, and green color values in the array. This can represent a total of $2^{24} = 16777216$ colors according to various combinations of red, green, and blue, which produces what people call twenty-four true color images. Linear grayscale conversion can be understood as assuming that the original image $\mathrm{f}(\mathrm{x},\mathrm{y})$ has a grayscale range $[a, b]$ and the grayscale of the original image is increased by the transformed. We can express it by formula 2.1.

$$\mathrm{g}(x,y) = \frac{d-c}{b-a}[f(x,y) - a] + c \quad (1)$$

After linear transformation through formula 1, the black areas of the image are exposed as much as possible, and the less bright white areas are whitened, so the contrast of the image is increased. If $\mathrm{c} = 0$ and $\mathrm{d} = 255$ in the assignment formula, then it can be reduced to:

$$\mathrm{g}(x,y) = \frac{255}{b-a}[f(x,y) - a] + c \quad (2)$$

Once the grayscale image belongs to the 0 ~ M category, the grayscale levels of most pixels are distributed in the interval [a,b], but only a few pixels have grayscale levels not in this interval. There is also a tool for implementing grayscale transformation in MATLAB: imadjust function, which can be used to achieve grayscale processing of images. This function can be used to obtain the style of the grayscale image that you want to see to enhance the contrast. The function expression format is shown in Equation 3.

$$\mathrm{J} = \mathrm{imadiust}(I, [L, H], [B, T], bite) \quad (3)$$

The grayscale processing of the answer sheet image is shown in Fig. 1.

2.2 Image Binarization

Binary image refers to an image that has been binarized. The processed image has only two colors, black and white, so it is called a black and white image. The grayscale value of the image has nothing to do with the middle grayscale value, and they are independent of each other. In MATLAB software, all binary images are conveyed by a matrix. This matrix has only two

values, 0 and 1, which correspond to black and white [5]. In MATLAB software, the function to turn a grayscale image into a binary image is as shown in Equation 2.4:

$$\mathrm{I} = \mathrm{im2bw}(J, N) \tag{4}$$

In this function, the image returned by image J after binary transformation is represented by I. The grayscale image to be transformed is represented by J. The threshold value of grayscale image is represented by N. If you change the size of the threshold, the corresponding binary image will also be transformed. The binarized image is shown in Fig. 2.

2.3 Image Smoothing

The function of image smoothing filter is to remove the noise in the image. Each image has its own different noise. The noise source comes from many different situations, interference from outside the system (such as electromagnetic interference), and some noise generated inside the system (Such as camera thermal noise). Low-pass filtering is generally a popular method to reduce noise in spatial domain or frequency domain. Smoothing filtering generally includes average, median, gaussian filtering and other methods.

Median filtering can determine the gray value of each pixel according to the median of pixels in adjacent areas. It is a non-linear smoothing technique that is used to eliminate single noise around. Median filtering can break through the obstacles of blurred image details caused by linear filters, such as the minimum mean square or average filter, which has the most significant effect on the noise generated when taking pictures. In MATLAB, you can use the medfilt2 function to implement median filtering:

$$\mathrm{B} = \mathrm{medfilt2}(A, [m, n]) \tag{5}$$

Each of the output pixels contains a median in the $\mathrm{m} \times \mathrm{n}$ neighborhood around the corresponding pixel in the input image. A is a binary image that needs median filtering. B is the median filtered image. The comparison before and after filtering is shown in Fig. 3.

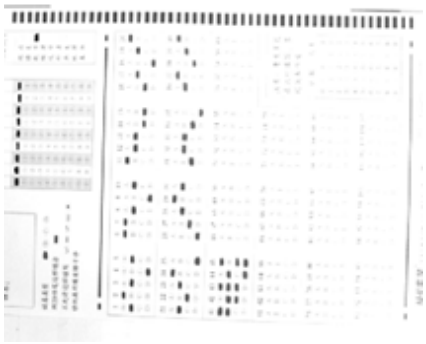

Fig. 1. Grayscale image

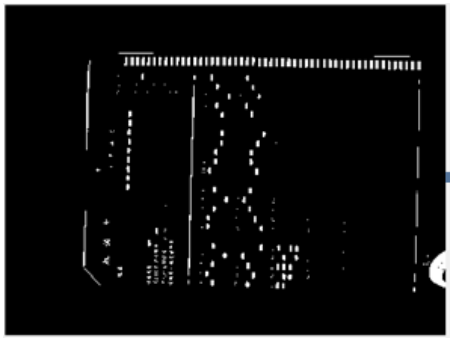

Fig. 2. Binarized image

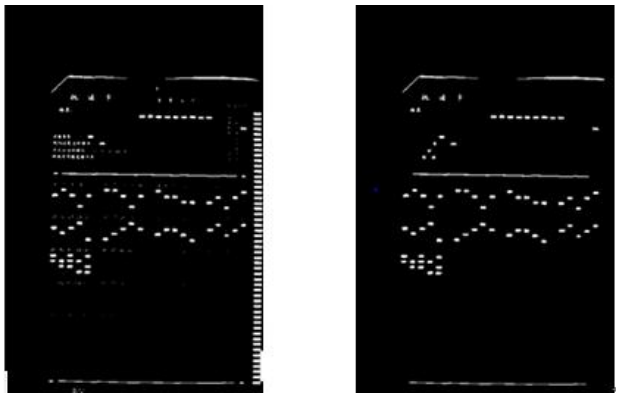

Fig. 3. Comparison chart before and after median filtering

2.4 Image Edge Detection

Different images have different distribution of gray scale and the gray scale pixel span is relatively large place to constitute the image line. The edges of each area have huge gray-scale pixel changes. Based on the knowledge of such a theory, the edges of the image can be basically identified.

Some methods of edge detection:

1. Edge detection using gradient operators. Generally use the edge () function to detect the edge of the image.
2. Laplace operator. It is aimed at the detection of 3 × 3 adjacent regions in the image.
3. Statistical algorithm.
4. Canny edge detection operator. It includes three aspects: good detection effect, accurate edge location and low response times to the same edge.

In this design, Hough transform is used to detect straight lines, and Hough transformation can use polar coordinates and Cartesian coordinates to identify straight lines. The results of edge detection are shown in Fig. 4.

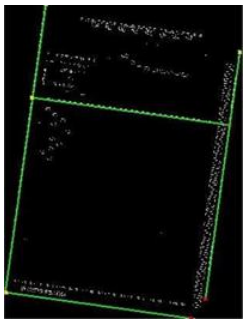

Fig. 4. Edge detection image

3. SYSTEM DESIGN

The scanned answer card images were processed and analyzed in MATLAB. And show the obtained information in the GUI interface. Finally, an answer sheet identification system with intuitive results and convenient search results is realized.

In MATLAB, the answer card identification processing is performed to realize the identification of student numbers, answers and subjects. After the answer card is input into the system, there are many image processing functions available in MATLAB. It involves the functions of image filtering, denoising, grayscale processing, binary processing, image segmentation, image correction and so on [6-7]. It is feasible to comprehensively analyze the design of the answer card recognition system based on MATLAB. The overall flow chart is shown in Fig. 5.

After the preprocessing of the answer sheet image, the system carries on the tilt correction to the answer sheet image, and then carries on the segmentation to the image. The author uses the method of region segmentation, which is realized by the similar characteristics of the pixels in the block region, and separates the picture according to the content of different needs. Generally speaking, region segmentation has the characteristics of oneness, difference and accuracy.

This design is aimed at the answer sheet of 105 filling areas. The upper area is divided into three areas, namely answer sheet type, student number and subject area. The lower area is divided into six horizontal large areas, of which 20 questions filling areas and 15 questions filling

areas have three each. We locate the object in these three areas and calibrate the image, which provides important help for the acquisition of follow-up information.

The mechanism for identifying the student number is the same as identifying the answer. When getting information, the order of scanning small squares is from top to bottom, so when recording, you need to create an array to record the student number and the answer. Finally, the student's score is calculated by the strcmp function in MATLAB. The formula is as follows:

$$J = \mathrm{strcmp}(s1, s2) \tag{6}$$

Due to various reasons, the answer sheet may have errors such as re-selection and miss-selection. After the answer sheet with these errors enters the system, the system will pop up a warning window to remind the reader. The system identifies 45 multiple choice questions, of which 1 to 40 are single choice and 41 to 45 are multiple choice. One to forty questions designed by the system are 2 points for each question and 41 to 45 questions are 4 points for each question, with a full score of 100.

4. GUI INTERFACE DESIGN

The purpose of designing the GUI interface is to reduce the workload of the examiner when it is necessary to identify the answer card and get the score for many times and the design interface is intuitive and simple. The friendliness of the interface directly determines the quality of the design. The design of the GUI is to make it easier to realize the functions of "checking papers" and "reading papers". You can also intuitively see the student number, subject, and score. It plays an indispensable role in the whole system and is the auxiliary function of the whole system [8-10]. The GUI design interface is shown in Fig. 6.

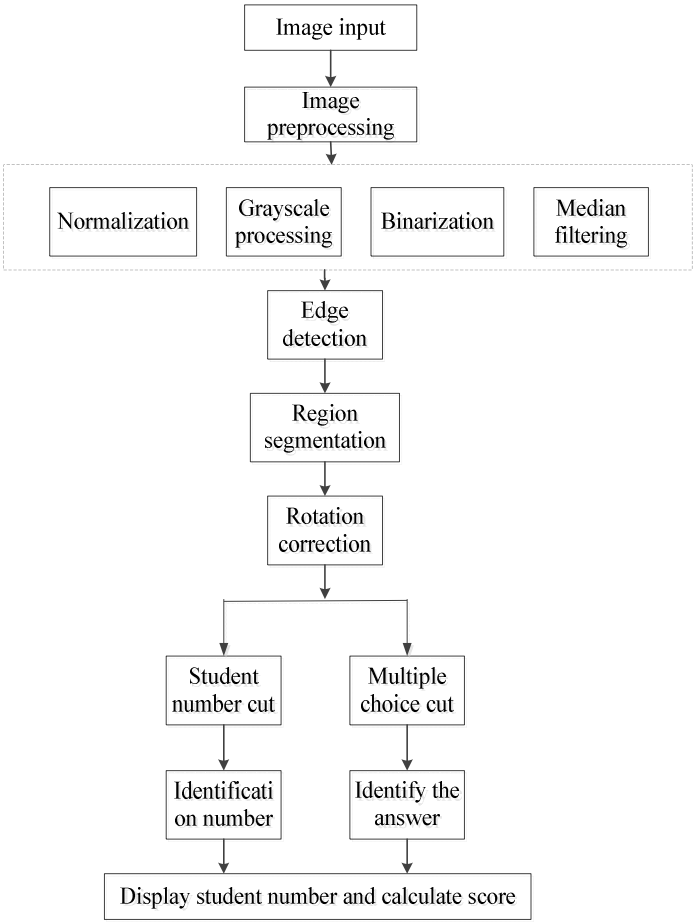

Fig. 5. System flow chart

In the design, it is simple, intuitive and easy to use. The specific functions are as follows:

1. The function of checking paper. Intuitively see the answer sheet filled by the students, which is used to check the filling situation and check the test paper.
2. One-click marking. Students' grades and examination subjects can be viewed intuitively in the GUI interface.
3. The student number shows. After clicking to mark the paper, you can directly see the identified student number and the result of each option.

5. TEST RESULTS

After running the program repeatedly, the author found no errors in the program and all the functions can be realized. In this design system, set the score of 1-40 questions to 2 points, 41-45 questions to 4 points, and a total of 100 points. The design system not only validates the single-selected topics, but also sets up the verification of multiple-selected topics, which increases the scope of application of the design. The accuracy of the answer sheet identification system is verified by the practical test of the answer sheet of several classes.

The test results of one of the classes are shown in Table 1. It can be seen from the test results that there are a few errors between the system score and the actual. The error is caused by the students' failure to fill in the answer card according to the standard (For example, the filled area is inaccurate or too small, etc.). In the future upgrade of the system, we need to focus on this aspect of the problem, to further improve the accuracy of the system. After testing, the recognition system is not very different from other existing recognition systems in terms of running speed, and the accuracy is high.

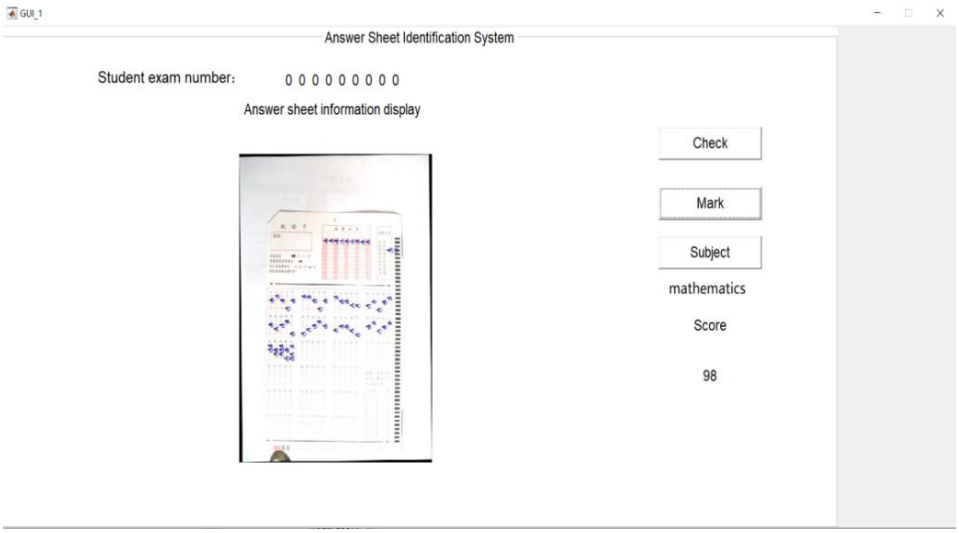

Fig. 6. GUI design interface

Table 1. Test result statistics table

Student ID	Accurate results	Identifying results	Conclusion
201410120	90	90	CORRECT
201410121	96	96	CORRECT
201410122	94	94	CORRECT
201410123	88	88	CORRECT
201410124	98	98	CORRECT
201410125	92	92	CORRECT
201410126	84	82	ERROR
201410127	100	100	CORRECT
201410128	86	86	CORRECT
201410129	82	82	CORRECT

6. CONCLUSION

This design uses MATLAB software to complete the answer card recognition system. The answer card image is processed by grayscale processing, image filtering, edge detection and so on. Identify and calculate the single and multiple choice questions until the final score. The system also designs a GUI interface, which is more conducive to human-computer interaction. Through the test of the system, it is proved that the system is feasible. The identification errors caused by human factors can be solved in the future system optimization to increase the reliability of the system.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Jesus Arias Fisteus, Abelardo Pardo. Grading multiple choice exams with low-cost and portable computer-vision techniques [J]. Journal of Science Education & Technology. 2013;22(4):560-571.
2. Kakani B, Dalal D, Dabhi A. Improved solution on students answer sheet assessment using fuzzy rules[C]//; 2016.
3. Gonzalez RC, Woods RE, Eddins SL. Digital Image Processing Using Matlab [J]. 2010;21(84):197-199.
4. Stephen M, Smith J, Michael Brady. SUSAN — A new approach to Low Level image processing [J]. International Journal of Computer Vision. 1997;23(1):45-78.
5. Florian Westphal, Niklas Lavesson, Hakan Grahn. Document Image Binarization using recurrent neural networks[C]// 2018 13th IAPR International Workshop on Document Analysis Systems (DAS); 2018.
6. Pandey AK, Saroha K, Sharma PD, et al. Development of a simple image processing application that makes abdominopelvic tumor visible on positron emission tomography/computed tomography image [J]. 2017;32(4):330-332.
7. Qiaoli Guo, Junsheng Zheng. An Iris Recognition Algorithm for Identity Authentication[C]. 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE Computer Society; 2018.
8. Ping H, He W, Shi Q, et al. An educational tool design for the course of signal processing based on Matlab GUI [C]; 2015.
9. Shucai Zhu, Ningning Qin, Weili Xiong. Design and simulation for sensor network performance evaluation system based on Matlab GUI[C]. 2015 International Conference on Control, Automation and Information Sciences (ICCAIS). IEEE; 2015.
10. Wang Z, Li H, Huang D. Digital image processing method of gasoline direct injection spray based on matlab/GUI[J]. Transactions of the Chinese Society for Agricultural Machinery. 2013;44(9):19-23.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/55980