**COMPUTER SYSTEM OF ANALYSIS OF THE MASS EXAM RESULTS**

 Chingiz Khalilov
*Department of Informatization and Technological Support*
*State Examination Center of the Republic of Azerbaijan*Baku, Azerbaijan
ch.khalilov@dim.gov.az

Rustam Azimov
*Laboratory of recognition, identification and methods of optimal solutions*
*Institute of Control Systems*Baku, Azerbaijan
0000-0002-4554-6985

Elshan Mustafayev
*Laboratory of recognition, identification and methods of optimal solutions*
*Institute of Control Systems*Baku, Azerbaijan
0000-0002-1544-3897

*Abstract*: Pattern recognition is becoming increasingly utilized within extensive information systems. The combination of advancements in image processing theory and the availability of open source libraries allows for the application of innovative solutions to various practical problems. One such problem involves automatically processing the answers from large-scale exams. This paper presents a developed system designed specifically for handling the results of these exams, demonstrating its ability to provide reliable, efficient, and unbiased assessments. To ensure precise template matching, different geometric transformations are used. Simultaneously, the coordinates of the recognized fields' exact positions on the graphic image are determined.

Keywords: information processing, intelligent system, evaluation of results, image processing

# **Introduction**

Optical Marker Recognition (OMR) technologies, utilizing optical scanners, are commonly employed to automatically process the results of large-scale test exams [1]. OMR scanners are favored for their high speed and reasonable accuracy. This approach allows for an objective assessment of a significant number of exam forms within a short timeframe, achieving nearly perfect recognition accuracy [2]. However, OMR technology suffers from several notable drawbacks:

Sensitivity to paper size and quality: OMR scanners are highly sensitive to the size and quality of the paper used, as well as the specific color of the form.

* limitation to specific form sizes: Custom form sizes cannot be accommodated with OMR technology;
* dependence on pristine form condition: OMR processing necessitates flawless form condition, as even slight damage, folding, or crushing can result in rejection;
* inability to process handwritten forms: OMR technology cannot handle forms filled out by hand;
* high cost and maintenance requirements: Implementation and maintenance of OMR technology can be expensive.

To address these issues, a system has been developed to enable independent and anonymous assessment of responses. High-speed scanners are utilized to convert all forms, regardless of size, into graphic format. Using predefined templates, the system identifies forms with questionably filled marker answers.

# **Description of the Processing System**

To begin with, a more detailed explanation of the operational stages of the developed system is provided. All exam materials received from exam sites are directed to the scanning area. The process of recognizing exam forms follows the main steps common to many other industrial systems of form recognition [3], [4]:

A. Generating a set of templates.

B. Determining reference points within the graphical representation of the form.

C. Defining and aligning the template.

D. Recognition.

E. Verification.

## Generation of a set of templates.

During this stage, the system performs scanning and imports the image of the blank form. Subsequently, the system automatically or manually identifies specific elements of the form known as control fields. These control fields can include lines, static text (repeated on all forms), checkboxes, "black squares," and barcodes.

## Definition of reference points in the graphic image of the form

During the printing and scanning process of forms, imperceptible linear distortions often occur along the height and width of the paper. To accurately align the template and compensate for these distortions, reference points are utilized. Typically, four black squares positioned at the corners of the form are employed as reference points (fig. 1).

|  |  |  |
| --- | --- | --- |
|  |  |  |

1. Samples of recognized forms

The identification of the black squares involves the following steps:

* The scanned form is converted to black and white mode.
* In the presumed area where the reference points are located, all connected objects are determined.

## Definition and alignment of the template

Using the coordinates of the reference points identified in the previous stage, the exam template is matched with the graphic image. Depending on the number of black squares detected, different transformation algorithms are employed to achieve the alignment.

In the case of utilizing four squares as reference points, a bilinear transformation is applied, which can be expressed as follows:

$$\left\{\begin{array}{c} x^{'}=a\_{0}+a\_{1}x+a\_{2}y+a\_{3}xy\\y^{'}=b\_{0}+b\_{1}x+b\_{2}y+b\_{3}xy\end{array}\right.$$

where $(x,y)$ и $(x',y')$ – source and converted coordinates, $a\_{i},b\_{i}, i=0..3$, – coefficients of transformation.

To achieve accurate parameter determination, it is sufficient to have the initial and transformed coordinates for four points. This involves solving two systems of linear algebraic equations:

$$\left\{\begin{array}{c}A^{'}\_{x}= a\_{0}+a\_{1}A\_{x}+a\_{2}A\_{y}+a\_{3}A\_{x}A\_{y}\\B^{'}\_{x}= a\_{0}+a\_{1}B\_{x}+a\_{2}B\_{y}+a\_{3}B\_{x}B\_{y}\\C^{'}\_{x}= a\_{0}+a\_{1}C\_{x}+a\_{2}C\_{y}+a\_{3}C\_{x}C\_{y}\\D^{'}\_{x}= a\_{0}+a\_{1}D\_{x}+a\_{2}D\_{y}+a\_{3}D\_{x}D\_{y}\end{array}\right. $$

When three squares are used as reference points for their identification, an affine transformation is applied, which can be described as follows:

$$\left\{\begin{array}{c} x^{'}=a\_{0}+a\_{1}x+a\_{2}y,\\y^{'}=b\_{0}+b\_{1}x+b\_{2}y,\end{array}\right.$$

where $(x,y)$ и $(x',y')$ – source and converted coordinates, $a\_{i},b\_{i}, i=0..2$, – coefficients of transformation.

To accurately calculate the parameters, it is sufficient to have the initial and transformed coordinates for three points. By solving two systems of linear algebraic equations, the coefficients can be determined:

$$\left\{\begin{array}{c}A^{'}\_{x}= a\_{0}+a\_{1}A\_{x}+a\_{2}A\_{y}\\B^{'}\_{x}= a\_{0}+a\_{1}B\_{x}+a\_{2}B\_{y}\\C^{'}\_{x}= a\_{0}+a\_{1}C\_{x}+a\_{2}C\_{y}\end{array}\right.$$

$$\left\{\begin{array}{c}A'\_{x}= b\_{0}+b\_{1}A\_{x}+b\_{2}A\_{y}\\B'\_{x}= b\_{0}+b\_{1}B\_{x}+b\_{2}B\_{y}\\C'\_{x}= b\_{0}+b\_{1}C\_{x}+b\_{2}C\_{y}\end{array}\right.$$

## Recognition

Using the template established in the previous step, the system proceeds to recognize each field within the form. The system supports recognition of various types of fields (fig. 2):

* checkboxes – 🞎,🗷,🗹,🞅,●;
* radio groups – only one mark has to be filled, for example - 🞅●🞅🞅;
* handwritten text - text written in block letters.

|  |  |
| --- | --- |
|  |  |
|  |  |

1. Type of recognized fields

To recognize checkboxes or radio groups, the system employs statistical recognition methods [5]–[8]. Once recognized, the system compares the field's statistics with the template. If the statistics surpass a certain threshold, the system makes a decision regarding whether to mark or leave the item unmarked.

When filling out the form, applicants often make the following mistakes:

* Selecting multiple choices by filling two or more points.
* Incompletely filling the circle.
* Using marks other than filling the circle.

To address these issues and minimize operator intervention, the system incorporates the following threshold values:

* Minimum percentage of filling: This determines the level of completion below which a field will be considered empty.

For handwritten text recognition, a multilevel recognition system based on neural networks (NS) is utilized [9]–[13]. The system primarily focuses on recognizing the Latin alphabet and the Azerbaijani language. However, it can be adapted to other alphabets and languages without significant difficulty.

## Verification

All recognition results are directed to the verification zone, where the following procedures take place:

* Templates of forms that were deferred in the previous step are defined manually.
* Doubtful fields are examined by operators who compare the graphic image with the recognition results on a monitor screen. The operator verifies the recognition and makes corrections if necessary (fig. 3).

Once the recognition and verification stages for the examination forms are completed, the data proceeds to the area of expert assessments for written answers. Independent experts are presented with graphical representations of the written answer areas on their screens.



1. Verification Screen

Once the examination form's results have been fully determined, the data is processed by the system to assess the accuracy of the test questions. The examiner's answers are compared with the correct answers, and specialized algorithms are used to establish the final assessment of the examination form.

In order to ensure transparency and objectivity, all data, including the graphical representation of the examination form, exam results, and correct answers, are made available online in the examinee's personal account.

# **Performance Improvement**

The speed of form processing is an important criterion for form processing systems. In order to achieve high processing speed in an existing system, the following approaches can be utilized:

1. Parallel use of the recognition process on multiple workstations: This involves distributing the workload across multiple workstations, allowing multiple forms to be processed simultaneously.
2. Utilizing the multicore/multithreading capability of modern processors: This approach involves parallelizing processes within a single workstation by taking advantage of the multiple cores or threads available.

In this paper, the focus is on investigating the second approach to increase productivity. Implementing multithreading in the developed system is preferred because it does not require additional equipment and can enhance productivity on existing hardware.

To implement multithreading, the stages of pattern matching and recognition are executed in parallel by separate threads. This approach ensures that these stages can exclude the simultaneous use of the same data by different threads, thereby improving overall efficiency of the recognition process (table 1).

During each experiment, the system processed a total of 1000 forms. To conduct a comparative analysis, the recognition process was executed sequentially using different numbers of threads, ranging from 1 to 8. The following are the results obtained from the experiments (fig. 4).



1. Recognition time (in sec) depending on the thread count
2. Recognition Time (in sec) for each Station Depending on the Thread Count

|  |  |
| --- | --- |
| Cores | Stations |
| N1 | N2 | N3 | N4 | N5 |
| 1 | 343 | 206 | 159 | 163 | 142 |
| 2 | 197 | 121 | 93 | 82 | 79 |
| 3 | 196 | 97 | 66 | 62 | 59 |
| 4 | 193 | 91 | 55 | 55 | 49 |
| 5 | 197 | 89 | 54 | 55 | 45 |
| 6 | 189 | 90 | 56 | 56 | 42 |
| 7 | 197 | 91 | 55 | 60 | 41 |
| 8 | 192 | 91 | 55 | 62 | 39 |

# **Conclusion**

The system that has been developed for processing and objectively evaluating test exams enables the administration of large-scale exams while providing reliable and unbiased assessment. This system can be easily configured to work with various types of forms, eliminating the need for expensive and cumbersome Optical Marker Recognition (OMR) scanners.

##### References

1. Elsa Ajarwati. “What Is OMR and How Different It Is from OCR: A Brief Comparison.” gleematic.com. https://gleematic.com/what-is-optical-mark-recognition-omr-example-of-use-cases/ (accessed June 24, 2023).
2. Karez Abdulwahhab Hamad and Mehmet Kaya, “A Detailed Analysis of Optical Character Recognition Technology,” International Journal of Applied Mathematics, Electronics and Computers, 2016, 4 (Special Issue), pp. 244–249.
3. K. R. Aida-zade and E. E. Mustafayev, “About one hierarchical handwritten recognition system on the bases neural networks,” Transactions of the NAS of Azerbaijan, series of PTMS, №2-3, 2002, pp.94-98. (In Russian).
4. K. R. Aida-zade, S. H. Talybov, and E. E. Mustafayev, “Multilevel recognition system of handwritten forms,” Proceedings of the 11th All-Russian Conference "Mathematical Methods of Pattern Recognition", Moscow, 2003, p.230-233.
5. Parul Himanshu Monga and Mnupreet Kaur. “A Novel Optical Mark Recognition Technique Based on Biogeography Based Optimization,” International Journal of Information Technology and Knowledge Management, July-December 2012, Volume 5, No. 2, pp. 331-333.
6. Rakesh S, Kailash Atal, and Ashish Arora, “Cost Effective Optical Mark Reader,” International Journal of Computer Science and Artificial Intelligence, Jun. 2013, Vol. 3 Iss. 2, pp. 44-49.