AUTOMATYKA • 2011 • Tom 15 • Zeszyt 3

Sebastian Stoliński*, Wojciech Bieniecki**

The Algorithms for Automatic Evaluation of Selected Examination Tasks from the Geometry

1. Introduction

Electronic marking (e-marking) is also known as Computer Assisted Assessment (CAA) or Computer Based Assessment (CBA) and it is a new idea in the field of teaching. Its aim is to facilitate the laborious process of designing, delivery, collection, scoring and analysis of the assessments [10].

Other advantages of introduction CAA are: easier scheduling and administration of assessments, the speed of results, its increased objectivity and security, the monitoring of student, performance and its suitability for distance learning [17].

Students also appeared to consider CBA as being more promising, credible, objective, fair, interesting, fun, fast and less difficult or stressful, while they stated that they preferred computerized versus written assessment [3, 7]. In [4, 15] it has been shown, that introduction of CAA allows to keep original accuracy of the exam and increase its reliability and even improve exam quality by:

– providing complete anonymity for task assessment;
– suppress the halo effect – a way of solving by one student does not affect the way later tasks are evaluated;
– steadily spreading potential errors in assessing – tasks to assess to the examiners are assigned at random.

* Politechnika Łódzka, Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki, Katedra Informatyki Stosowanej, Łódź, Polska
Stypendysta projektu „Innowacyjna dydaktyka bez ograniczeń-zintegrowany rozwój Politechniki Łódzkiej – zarządzanie Uczelnią, nowoczesna oferta edukacyjna i wzmacnianie zdolności do zatrudniania, także osób niepełnosprawnych” współfinansowanego przez Unię Europejską w ramach Europejskiego Funduszu Społecznego
** Politechnika Łódzka, Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki, Katedra Informatyki Stosowanej, Łódź, Polska

The e-marking method has been widely introduced in Great Britain and the USA. The experience gained by Examination Boards like AQA, OCR and EDEXCEL in Great Britain and ETS in the USA suggests that introducing e-marking improves the quality and reliability of the exams.

Studying the available literature it is possible to select main methods of image processing, analysis and understanding, that appear in CAA solutions. These are:

- Optical Character Recognition / Intelligent Character Recognition (OCR/ICR) [13],
- lexicography analysis [6];
- image understanding techniques [16];
- neural networks to OCR/ICR and text identification algorithms [9, 14];
- Hough transform to object identification [8].

Descriptions of CAA and e-marking systems found in the literature were designed for automatic evaluation of the exams carried out at the computer, which means, they used immediately analytical or lexical form. Dissemination of this form of examination in Poland will not be possible in the next few years, so we deal with somewhat more difficult to implement but easier to deploy method based on the image recognition and understanding. Among the available literature and documentation we have not found any CAA systems that rely on image analysis and could be compared with ours.

In our previous study we worked out the requirements for our own e-marking system [1]. We investigated the types of tasks appear in high school and maturity exams (natural science and mathematics). We chose those, which could be to assess using image processing and analysis techniques.

For tasks of mathematics and natural sciences we consider the following task types:

- write or complete a mathematical expression or a chemical formula;
- draw the structural formula of a chemical compound; write a single word;
- point the object or objects in the figure;
- draw a graph or a geometric figure;
- draw a graphic solution of equation set;
- draw a specific set.

Initially we focused on the image processing algorithms that enabled detection and assessment of plots of simple functions: linear, parabolic [2], discontinuous, combined functions drawn by hand [11, 12].

Using our experience we construct the algorithm for two new examination tasks: draw the triangle and draw a solution of the set of inequalities.

2. The subject of study

First task was to draw a triangle, which sides are designated by lines given by the equations (Fig. 1):

$$
\begin{aligned}
&a) \quad x - y + 1 = 0\\
&b) \quad x + y - 5 = 0\\
&c) \quad x + 5y - 5 = 0
\end{aligned}
$$

Fig. 1. Task 1 – draw the triangle defined by lines

The score for the task was between 0 and 3 (one point for a correctly drawn line segment).

The second task consisted of a graphical solution of the inequality set (Fig. 2):

$$
\begin{cases} y > x^2 - 6x + 5 \\ y < -0,4x + 5 \end{cases}
$$

Fig. 2. Task 2 – draw the solution of the inequality set

The solution was scored from 0 to 3 points (one point for a parabola, one point for a line segment, one point for marking the region).

In Figure 3 exemplary images containing proper solutions are printed.

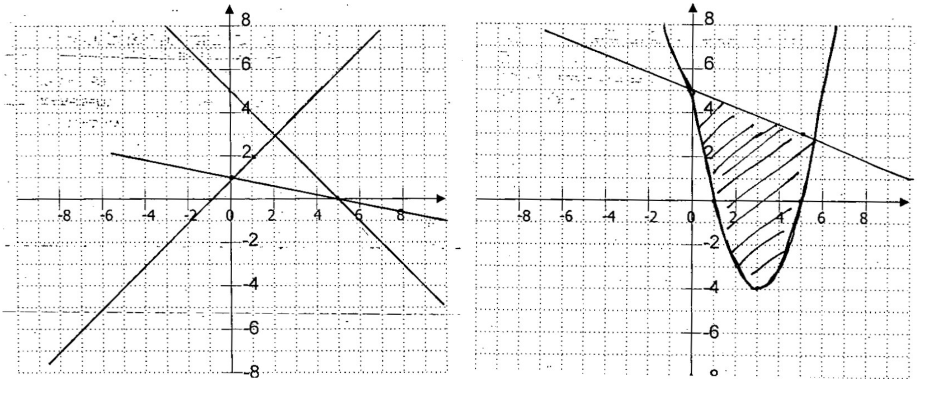

Fig. 3. Exemplary solutions of given tasks

Students should do their drawings on the printed empty coordination set (Fig. 4a).

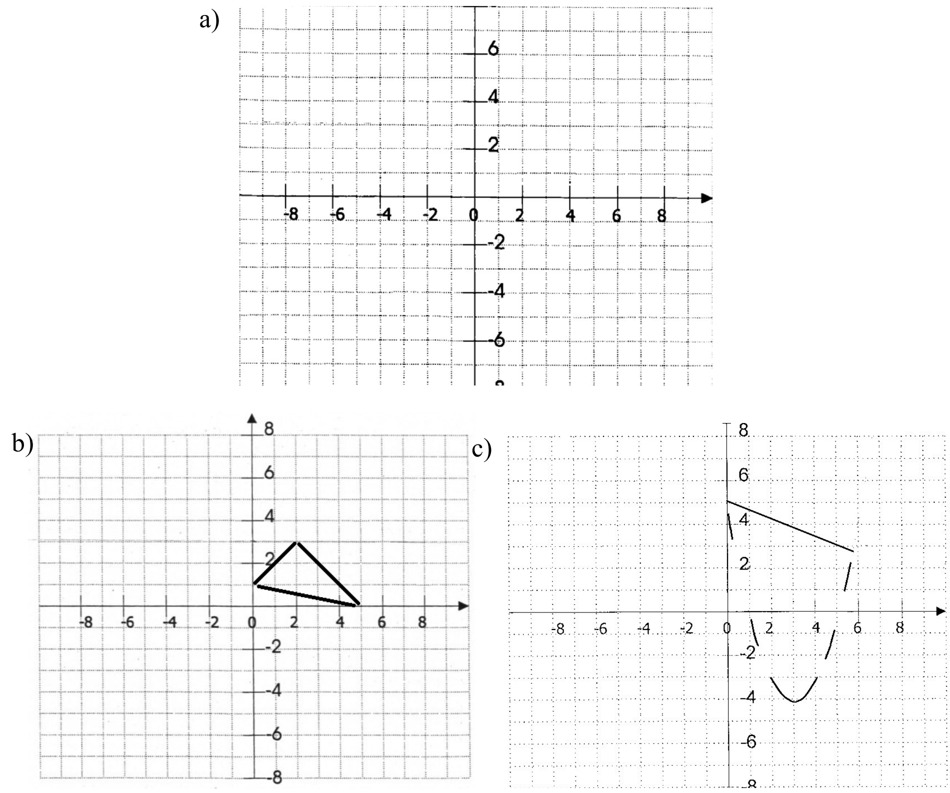

Fig. 4. Empty coordinate system (a), coordinate system with automatically generated sides of the triangle, (c) coordinate system with automatic generated line segment and parabola tested regions

3. Algorithm 1: assessing the triangle

The image processing and analysis algorithm works in several phases. The first phase relies on binarization using Otsu [5] method and extraction of the drawing from the scanned examination sheet and it is identical as presented in [2].

The algorithm for triangle extraction requires five input images:

I_1 – the image of empty coordinate system (Fig. 4a)
I_2 – a scanned image of the student's solution
I_{t1} – automatically generated reference image of first side of the triangle (Fig. 1a)
I_{t2} – automatically generated reference image of second side of the triangle (Fig. 1b)
I_{t3} – automatically generated reference image of third side of the triangle (Fig. 1c)

All the images are binary. In the following Figures black denotes 1 and white denotes 0.

Steps of the algorithm:

With aid of the cross correlation method find the best match of I_1 and I_2 images. Obtain $I_r = I_1 - I_2$. Using mathematical morphology filters, remove the distortions (for instance, fragments of the coordinate system lines remained from imperfect scanner geometry) from I_r and leave only the drawing lines (Fig. 5).

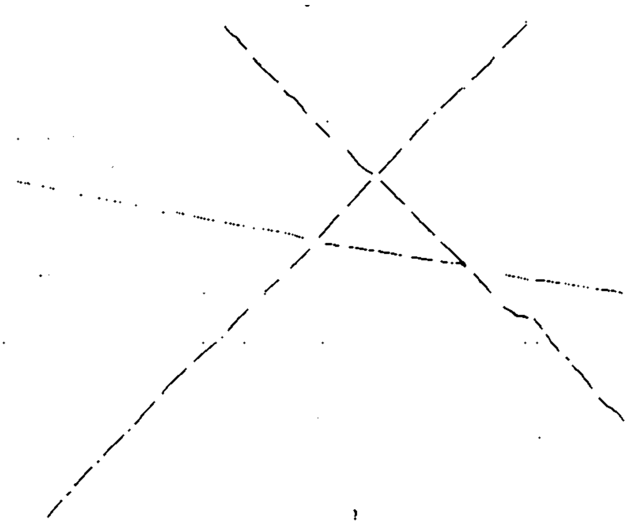

Fig. 5. Plot curves (I_r)

For each side of student's triangle we define (basing on the equations – Fig. 1) a region of interest (ROI), where the solution should be located. Then, the following operations are done:

Best match of each I_{tn} and I_r images (n = 1 ... 3) (Fig. 6) is found using the cross correlation method. Coordinates of point of maximum correlation are used to crop the expected solution image from the query image (I_{2n}).

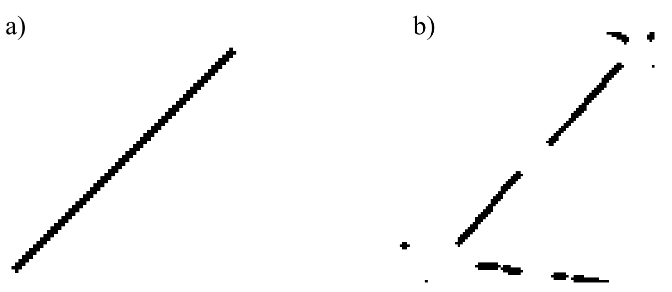

Fig. 6. Reference plot (I_{t3}) (a) and extracted student plot (I_{23}) (b)

For each of I_{tn} do the dilation (the structuring element is a disk $r = 10$ pixels). Assign it as I_{rdn}. Calculate the coefficients:

P_{1n} – number of black pixels in dilated test image ($n = 1..3$),
P_{2n} – number of matching pixels in dilated test and reference images,
P_n – the compliance rate between reference and query lines.

how many pixels from reference plot I_{tn} matches the pixels of test plot I_{rdn}. Using Formula (1).

$$P_{1n} = \left|\{I_{rdn} : I_{rdn} == 1\}\right|$$
$$P_{2n} = \left|\{I_{rdn} \,\&\, I_{tn} : I_{rdn} \,\&\, I_{tn} == 1\}\right| \tag{1}$$
$$P_n = P_{2n} / P_{1n}$$

It may happen, that the number of pixels P_{2n} is very small (the student drew only a few dots, not the whole line). Compare the number of pixels in test and reference dilated plots. We set the condition (2) experimentally.

$$P_{1n} \leq \mathrm{cov}_n \cdot P_{2n} \tag{2}$$

If the condition is not satisfied, the plot is not valid (we put the negative number). In all other cases P_n is in range of (0, 1). For each segment the threshold value cov_n was chosen in the process of training.

Finally, we cut out all found and recognized line segments from I_r (student's solution image) to detect other objects (3), the image I_{e2} containing distortions (black pixels do not belong to any of recognized images).

$$I_{e2} = I_r - I_{t1} - I_{t2} - I_{t3} \tag{3}$$

Solution is classified as unrecognized, when the image I_{e2} has more than R black pixels.

4. Algorithm 2: assessing the system of inequalities

The first phase of the algorithm relies on extraction of the drawing from the scanned examination sheet and it is the same as in Algorithm 1. In Figure 7 the exemplary student's solution of the task is shown (I_r – extracted test image).

Recognition and scoring phases for the system of inequalities include three main steps, related to individual primitives drawn on the plot.

Initially, a square function plot is extracted and assessed. The method of scoring is presented in [2]. The algorithm recognizes the extreme of the parabola and four „crate" points. In order to mark the plot as correct, the extreme and at least three of four crate points have to be recognized as correct. The areas for crate points and the extreme areas are marked in Figure 7. Using formulas: (1) and (2) the number of matching common pixels from I_{tdn} and I_{rn} is calculated.

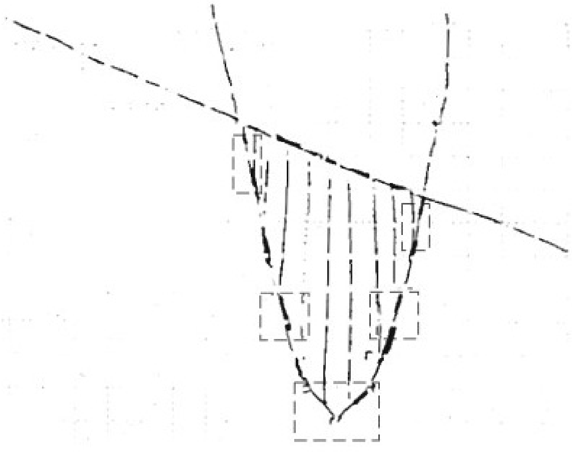

Fig. 7. Extracted drawing (I_r) with the regions of measuring of extreme and crate points

Secondly, the line segment is recognized and scored. Algorithm for finding and scoring of the line segment is equivalent to presented for Task 1.

Thirdly, the selection of a proper area is checked. We found a number of black pixels in the recognized area and if it is above the threshold value, the region is marked as selected. This point is done only if square function and the line segment are scored as correct.

Finally, as in Algorithm 1, the number of extra black pixels is calculated. If it is below threshold, solution is classifies as unrecognized. Before counting black pixels, the plot of the line, which contains checked segment, the plot of a square function and selection area are erased from the solution image.

All thresholds were selected during experiments.

5. The Experiment

The experiment has been carried out on one group of 27 students. Each student received a print-out with two empty coordinate systems and had to complete Task1 and Task 2. All pages were collected and scanned with the resolution 300 DPI with 8-bit grayscale mode. At this resolution, one unit of the coordinate system corresponds to 39 pixels. This value has been used to parameterize the algorithms.

In practice, during the examination, it is assumed that the line is drawn properly, if it does not deviate more than 1/4 units on the coordinate system. Therefore, the image I_{rdn} is formed by dilation with a disk of a diameter 10 pixels structuring element.

The entire set of scans has been randomly divided into two parts: the learning (13) and test (14). The coefficients cov_n $corr_n$ and R presented in Tables 1–3 were chosen in the process of training, so that the training set classification error was minimal. The process of minimizing the error number indicates that the coefficients values are inversely proportional to the size of the test image.

In order to verify the obtained results, they have been compared with a qualified teacher grades. In our algorithm each grade has three possible values: 1 – line segment is correct, 0 – line segment is not correct, U – unrecognized image.

In Tables 1–3 the exemplary comparison of the grades given by the teacher and calculated by the computer system for 14 tested solutions is presented. For Tables 1 and 2 corr1 means the cross-correlation value for the first line segment and cov_1 – the P value (see eq. (1)). In Table 3 the grades (0 or 1) given by the teacher and evaluated by the algorithm are entitled: sqr. – the mark for the parabola, seg. – the mark for a line segment, sel. – the mark for a region selection. R is a percentage of black pixels of I_{e2} image.

For all samples the average error rates have been calculated (Tables 4–5). The error occurs if the teacher's score is not the same as the computer's. In the case of real exam, if the grade is under the pass level the work is evaluated once again. The undesirable situation is when we accept the incorrect task solution (overestimation).

Table 1
Exemplary samples for the algorithm evaluation (Task 1)

Sample code	Grade (teacher) Line segment			Compliance of images							Grade Line segment		
	1	2	3	corr 1	cov 1	corr 2	cov 2	corr 3	cov 3	R [%]	1	2	3
Tresh.				0.4	0.08	0.2	0.07	0.18	0.02	0.7			
118	1	1	1	0.68	0.08	0.51	0.08	0.37	0.09	0.01	1	1	1
119	1	1	0	0.34	0.28	0.44	0.22	0.07	0.14	0.59	0	1	0
120	1	1	0	0.54	0.11	0.57	0.10	0.07	0.04	0.21	1	1	0
121	1	1	1	0.76	0.11	0.56	0.11	0.18	0.03	0.10	1	1	1

Table 2
Exemplary samples for the square function evaluation (Task 2)

Sample code	Compliance of images										Grade
	Extreme		Crate p. 1		Crate p. 2		Crate p. 3		Crate p. 4		
	corr	cov.	corr	cov	corr	cov	corr	cov	corr	cov	
Tresh.	0.2	0.03	0.3	0.04	0.3	0.07	0.3	0.04	0.3	0.07	
117	0.53	0.10	0.46	0.23	0.31	0.28	0.49	0.23	0.43	0.31	1
118	0.00	0.00	0.26	0.11	0.18	0.06	0.23	0.08	0.19	0.01	0
119	0.31	0.15	0.39	0.27	0.26	0.26	0.32	0.30	0.45	0.32	1
120	0.07	0.06	0.29	0.10	0.44	0.06	0.34	0.1	0.25	0.06	0

Table 3
Exemplary samples for Algorithm 2 evaluation

Sample code	Grade (teacher)			Compliance of images				Grade (algorithm)		
	sqr.	seg.	sel.	corr. segm.	cover	selection	R [%]	sqr.	seg.	sel.
Tresh.				0.15	0.005	0.04	1.7			
117	1	1	1	0.39	0.038	0.14	1.1	1	1	1
118	0	0	0	0.04	0.058	0.10	1.2	0	0	1
119	1	1	1	0.35	0.176	0.15	1.2	1	1	1
120	0	1	1	038	0.031	0.06	1.1	0	1	1

Table 4
Rate error for Task 1 (27 samples, 13: learning set, 14: test set)

	Learning set		Test set	
	Samples	Relative error	Samples	Relative error
Overestimated score	0	0%	0	0%
Underestimated score	1	8%	1	7%
Unrecognized	0	0%	0	0%
	Compatible results: 12 (92%)		Compatible results: 13 (93%)	

Table 5
Rate error for Task 2 (27 samples, 13: learning set, 14: test set)

	Learning set		Test set	
	Samples	Relative error	Samples	Relative error
Overestimated score	0	0%	2	15%
Underestimated score	1	8%	1	7%
Unrecognized	0	0%	0	0%
	Compatible results: 12 (92%)		Compatible results: 11 (78%)	

6. Conclusions

The presented algorithms work properly in a case of tested tasks. Some errors occurred for solutions with strike-throughs and amendments. There are no unrecognized solutions, because students did not draw in tested sample any disturbing objects: additional lines, digits.

What should improve the results, is introducing new method for grading the solutions for square function. In this step in order to obtain good results (which is required to conduct extensive tests) it is necessary to choose the best parameters for the algorithms.

The combination of presented algorithms may be helpful for creation of automatic assessment applications for other examination tasks as: function analysis, geometry.

Our future research will include the detection and assessment of other figures as: circles and ellipses, polygons, plots of trigonometric functions.

References

[1] Bieniecki W., Stańdo J., Stoliński S., *Analiza wymagań dla systemu elektronicznego oceniania rozwiązań zadań egzaminacyjnych*. Automatyka (półrocznik AGH), 2009, 1397–1406.

[2] Bieniecki W., Stando J., Stolinski S., *Automatic evaluation of examination tasks in the form of function plot.* Proc. of the 6th IEEE International Conference MEMSTECH 2010: Perspective technologies and methods in mems design, Lviv-Polyana, Ukraine, 140–143.

[3] Croft A.C., Danson M., Dawson B.R., Ward J.P., *Experiences of using computer assisted assessment in engineering mathematics*. Computers and Education, 27, 2001, 53–66.

[4] Fowles D., Adams C., *Does assessment differ when e-marking replaces paper-based marking?* IAEA paper v3(cra+def).doc .2005.

[5] Otsu N., *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-9, Jan. 1979, 62–66.

[6] Prusa D., Hlavac V., *Mathematical Formulae Recognition using 2D Grammars*. Document Analysis and Recognition, ICDAR Parana, 2007, 849–853.

[7] Sambell K., Sambell A., Sexton G., *Student perceptions of the learning benefits of computer-assisted assessment: A case study in electronic engineering*. [in:] S. Brown, P. Race, & J. Bull (Eds.), Computer assisted assessment in higher education (1999*)*. London: Kogan Page.

[8] Sewisy A.A., *Graphical Techniques for Detecting Lines with the Hough Transform*. International Journal of Computer Mathematics, vol. 79(1), 2002, 49–64.

[9] Shashank Araokar, *Visual Character Recognition using Artificial Neural Networks*. eprint arXiv:cs/0505016 2005.

[10] Sim G., Holifield P., Brown M., *Implementation of computer assisted assessment: Lessons from the literature*. Association for Learning Technology Journal (ALT-J), 12(3), 2004, 215–229.

[11] Stoliński S., Bieniecki W., Stańdo J., *Automatic detection and evaluation of the spline function plot* Automatyka (półrocznik AGH), 2010.

[12] Stoliński S., Bieniecki W., *Evaluation of selected examination tasks with aid of image processing and analysis.* Proc. of the SIS Conference, Łódź 2010.

[13] Stoliński S., Bieniecki W., *Application of OCR systems to preprocessing and digitalization of paper documents*. Information Systems in Management: Information and Communication Technologies for e-Business. WULS Press, Warszawa 2011, 102–111.

[14] Strouthopoulos C., Papamarkos N., *Text identification for document image analysis using a neural Network*. Image and Vision Computing, 16, 1998, 879–896.

[15] Szaleniec H., Weziak-Bialowolska D., *Czy e-ocenianie może zapewnić większą rzetelność punktowania?* [in:] Niemierko B., Szmigiel M.K. [red.], Uczenie sie i egzamin w oczach nauczycieli. XIV Krajowa konferencja, „Diagnostyka edukacyjna", PTDE 2008.

[16] Tadeusiewicz R., Ogiela M.R., Szczepaniak P.S., *Notes on a Linguistic Description as the Basis for Automatic Image Understanding*. Int. J. Appl. Math. Comput. Sci., vol. 19, No. 1, 2009, 143–150.

[17] Thelwall M., *Computer-based assessment: A versatile educational tool*. Computers and Education, 34(1), 2000, 37–49.