Lodz University of Technology
Faculty of Electrical, Electronic, Computer and Control Engineering

Institute of Applied Computer Science

Sebastian Stoliński

Application of image processing algorithms to automatic assessment of handwritten solutions of examination tasks in mathematics

PhD Thesis

ABSTRACT

Supervisors:

dr hab. Szymon Grabowski, prof. PŁ

dr inż. Wojciech Bieniecki

2018

Abstract

The motivation of selecting the topic of PhD thesis was the result of analysing current math exams tasks from final exams in primary and high schools in Poland. There is lack of open tasks, whose part of the solution is drawing a function plot or figure. Moreover, the last audit on the system of external examinations conducted by the Polish Supreme Audit Office showed that 25% of all high school exam works from the years 2009–2014 is incorrectly scored.

The dissertation thesis states that: *image processing algorithms can be used to automatic assessment of handwritten examination tasks in mathematics.*

The dissertation contains an architectural project of an IT system to implement automatic and manual assessment process. The work is focused on preparing and comparison of image processing algorithms, that provides scores of scanned tasks consisting of drawing function plot. Algorithmic solutions are proposed for all phases of the process from scanning to returning final score.

The main part of the thesis contains descriptions of algorithms and conducted experiments.

The dissertation provides comparison of four variants of algorithms for automatic assessment of examination tasks. The first algorithm is based on correlation, the second uses Hough Transform, the third least square fitting. The last one was combination of Hough Transform (classic and general) and least square fitting. Moreover, preprocessing methods were tested and implemented. They contain some steps necessary to prepare scanned images to automatic assessment:

- extracting coordinate system from the scanned sheet,
- removing coordinate system from the picture
- removing noise, thinning,
- removing object crossroads (connections).

The general assessment criteria are comply with the following rules:

- straight lines must intersect the coordinate axes at appropriate points; line segments should be as long as possible;
- extremes and roots of the function must be in the appropriate points;
- in circles, the center and the radius should be preserved;
- domains and ranges of the functions should be preserved;
- the period of a periodic function should be preserved.

An exemplary scanned sheet with three tasks and selected area of each task is in Fig. 1.

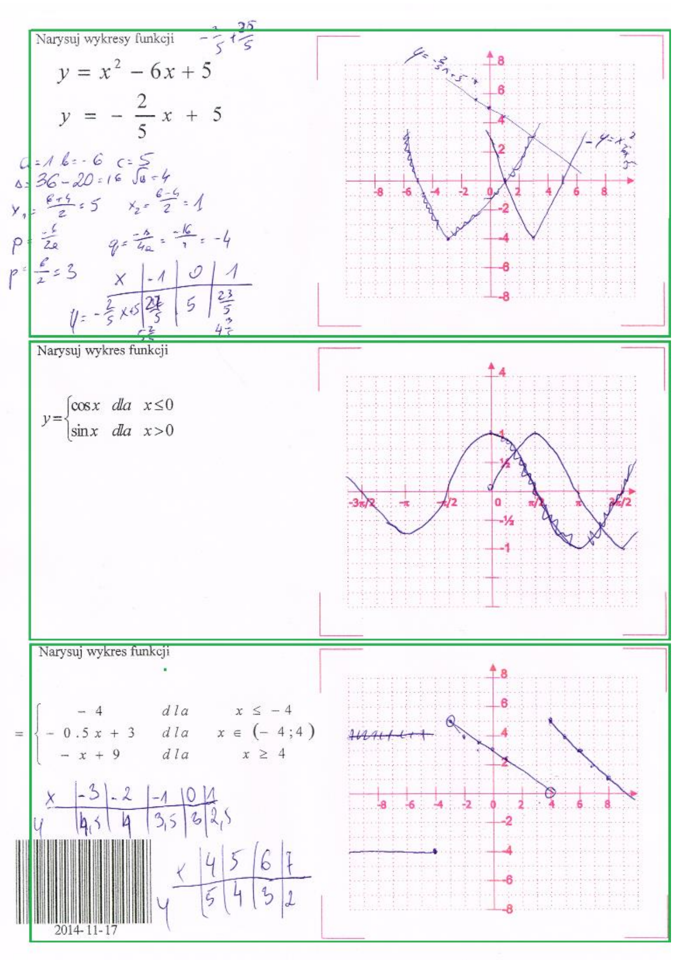

Figure 1 Scanned sheet with marked tasks areas

The algorithms

Correlation algorithm

The correlation based solution uses a set of images as the reference. The correlation between them and a particular area of the assessed task was checked. For the line segment case it was the whole shape, for other tasks verification comprised only a selected part of the images. Based on the correlations level and theirs maximum location the score was returned.

Hough transform

Two different variations of Hough transform were used: classical Hough transform (HT) for recognition of the line segments and the generalized Hough transform (GHT) for recognition of non linear curves like: arc, parabola and trigonometric functions. Based on maximum value of HT and its localization the final task score is calculated. The configuration of HT based solution is quite simple. For the line segment the input values are the line equation coefficients, plus its begin and end coordinates. For GHT, the inputs are images containing the curve to be detected and theirs expected localization in the image.

Least square fitting

The least squares method can be used to identify most of the curves described analytically (with an equation). The difficulty lies only in the transformation of the figure equation to the form of which the iterative process of fitting is convergent. The thesis contains implementation of two phases least squares method. The first phase, algebraic fitting, which produced an initial approximation and the input values for the second phase, which is geometrical fitting. The algorithm output is a matrix containing coefficients an of identified curve. Because some of the curves drawn by student are broken, an algorithm to connect the primitives was implemented. The algorithm is able to assess a partially correct solution with incorrect parameters, e.g. frequency or amplitude for trigonometric functions.

Hough-fit algorithm

To take advantage of the Hough Transform and least squares approximation, both of these methods were combined in a two-pass algorithm. In the first phase, the examined image is processed using Generalized Hough Transform. All objects that HT identified as lines are labeled and remain in the image. In the second phase the approximation is applied according to the description from the previous section.

This algorithm has been applied to the assessed tasks involving sketching parabola, parabola and a line, sine and cosine waves.

The best approach turned out to be a two-pass algorithm combining Fit with HT. In most cases it was better than the other. Also, it was observed that two-pass algorithm is better than their component algorithms – the increase of compliant assessments was of 3–5%.

Voting algorithm

An analysis of results provided by algorithms described above was shown that all of them have correct assessment ratio about 75–85%. However, the incorrect assessments were made by different methods for different tasks, for example, if HT algorithm returns an incorrect mark for specific task solution, usually the remaining algorithm provided results for the solution complying with teacher mark. The conclusion was that voting algorithm may improve the assessment quality.

Two approaches were tested, the first focused on improving marks quality with condition that number of unrecognized solutions should not increase significantly. The second focused on providing the lowest error ratio.

In the first case majority voting is implemented. The mark is returned when not more than one algorithm (correlation, HT, fitting, Hough–fit) marked the solution as unrecognized and the mode for the scores of the tasks exists. In the other case, the solution was marked as unrecognized.

The algorithm was tested, the correct assessment ratio increases by 10% in comparison to correlation, HT, fitting and Hough–fit.

The second approach is unanimous voting. The mark is returned when not more than one algorithm (correlation, HT, fitting, Hough–fit) marked the solution as unrecognized and the assessment provided by all algorithms (except this one that not recognized the solution) are compliant. The experiment results: the unrecognized ratio for the solutions were between 40 and 60 percent, but the quality of the assessment of remaining solutions for all tasks were above 95%.

Unrecognized

All algorithms described above have additional step implemented If the student's solution contains additional objects and a non-zero score was given, the task solution was then marked as “unrecognized”. The task solution was set as “unrecognized”.

The experiment

The following five tasks were used to validation of the algorithms:

1. Draw a figure, made of the central angle with the subtended by an arc formed by the intersection of the circle with the straight line (all primitives are given by equations). Score range from 0 to 3 points.
2. Draw a spline function plot given by an equation. Possible score range from 0 to 3 points.
3. Draw a parabola given by an equation. Possible score range from 0 to 2 points.
4. Draw a graphical presentation of the solution of the system of equations. Possible score range from 0 to 3 points.
5. Drawing the graph of a trigonometric function defined by equation. Possible score range from 0 to 4 points.

The experiment has been carried out on a group of students. Each student received a sheet with an empty coordinate system and had to complete the task. All tasks were assessed by the teacher, his marks were used as the reference. The tested algorithms may return a numerical score or unrecognized status. The last one means the solution containing objects, noise, artefacts cannot be automatically recognized.

Conclusion

The algorithms described above were compared, and it is not possible to select single best algorithm from the set of non-voting variants, which provided the best results for all tasks. Each of the methods has some weak and strong points. In most cases the Hough-fit algorithm results are better than others. The final results of the two-pass algorithm are approximately 3–5% higher than the remaining of non-voting methods.

The hardest type of solution to assess is partially correct task. This happens because the parameter thresholds for the cases: incorrect and partially correct, and partially correct and correct are often a bit vague. The similar errors were detected in some teacher marks used as reference.

The best assessment quality was provided by voting methods, especially unanimous voting was able to get almost 100% of correctly assessed tasks, but in this case unrecognized ratio was very high (approximately half of the tasks).

For all described tasks at least one of algorithms provided a results better than 75% correctly recognized solutions, thus we can conclude that using image processing algorithms to automatic assessment of examination tasks is viable and provides better results than manual assessment.