

# Finding the Answer: Techniques for Locating Students' Answers in Handwritten Problem Solutions

Amirali Darvishzadeh, Negin Entezari, Thomas Stahovich  
University of California, Riverside

## ABSTRACT

In many academic subjects, especially science, engineering, and math, paper-based problem solving is an important part of education. However, grading such work can be prohibitively expensive in large university courses. As a remedy, we have developed techniques to support the automated grading of handwritten problem solutions. Students complete their work using Livescribe digital pens and draw boxes around the final answers. We developed techniques that identify the answers by locating the boxes. This problem is challenging as the written work contains a mixture of diagrams and equations, and boxes frequently appear as a part of the diagrams. Additionally, the boxes must be segmented from the remainder of the writing. Thus, a simple shape recognizer is inadequate for this task. Our techniques efficiently locate answer boxes within a complex page of free-form writing. Furthermore, our techniques are designed to be robust to the wide range of variations in the way students write. In a test on 2022 pages of homework problems, our techniques correctly located 95.3% of the 4473 answer boxes. These techniques are an important step towards automated grading of handwritten work because once the answer boxes are located, a variety of handwriting recognition methods can be used to interpret the answers.

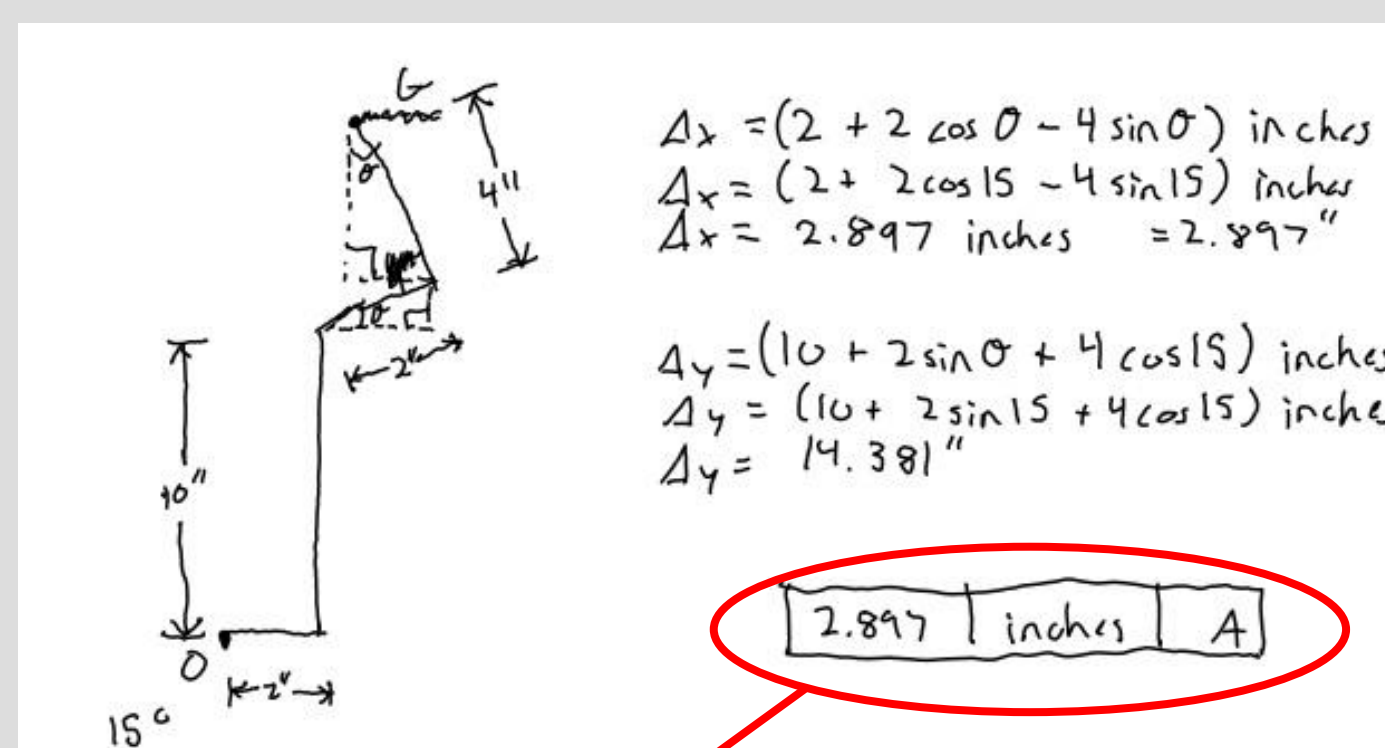
## CONTACT

Amirali Darvishzadeh  
University of California, Riverside  
Email: darvisha@cs.ucr.edu  
Phone: 213-268-9512

## INTRODUCTION

There are many exams and homework assignments, especially in science and engineering courses, which are best assessed by free response questions. The most natural and efficient way for students to answer such questions is to write by hand on a piece of paper. However, grading such handwritten answers has always been a tedious and time-consuming task for teachers especially in large classes. Auto grading even aspects of the handwritten answers can save a huge amount of time for teachers. Locating final answers to questions is an important step toward auto-grading.

**A typical answer to a statics problem contains the value, unit and problem id, which are separated by vertical lines**



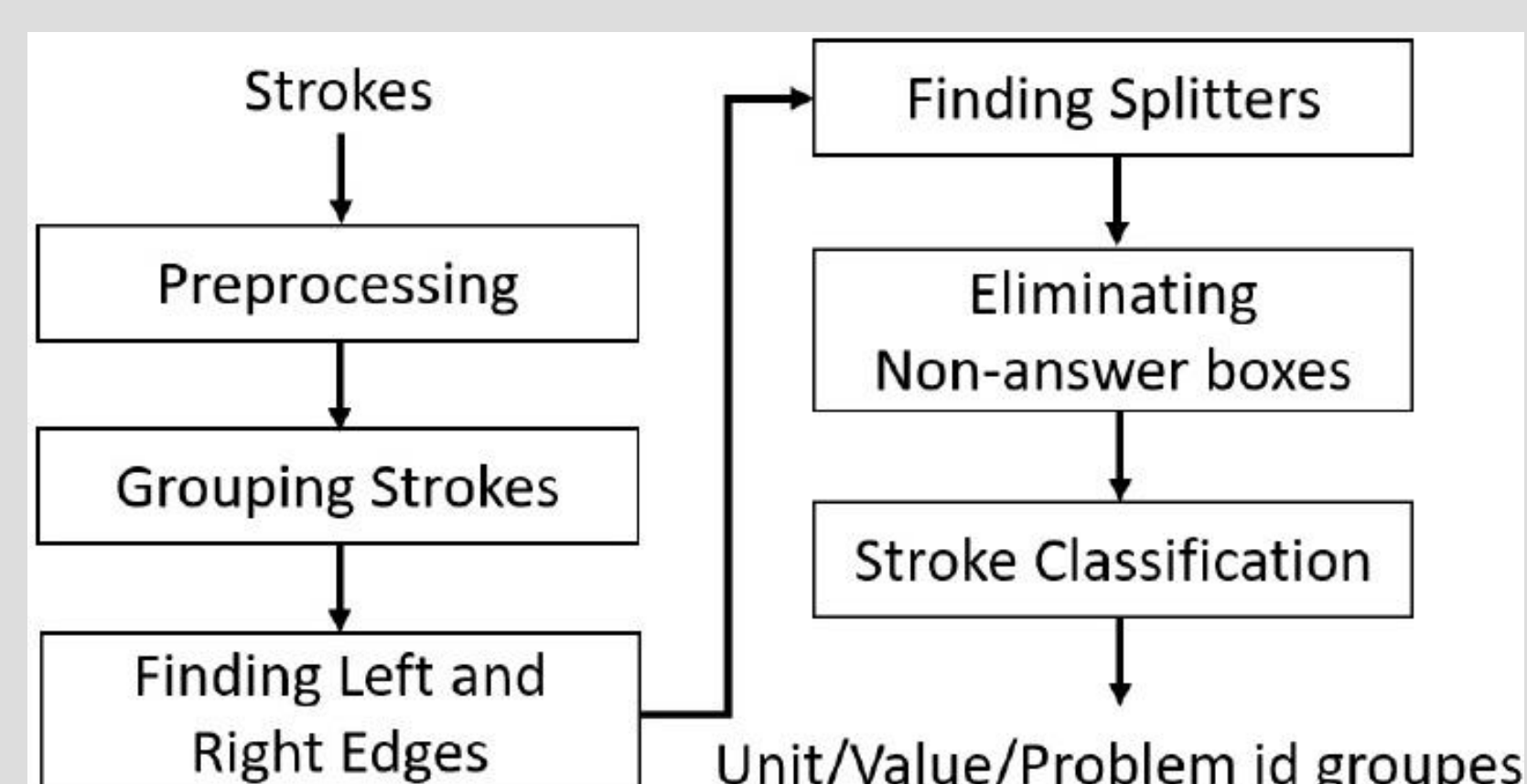
Value	Unit	Id
2.896	inches	A

Difficulties in locating and recognizing answer boxes:

1. Free response answers may involve equations, free body diagram and texts
2. Most answer boxes are sketched like irregular quadrilateral shape.
3. Answer boxes are often drawn with disconnected edges.
4. Presence of other boxes which are not answer boxes

## METHODOLOGY

An overview of the system is shown in following figure:



## METHODOLOGY

### Preprocessing

- Segment strokes at corner points and find all horizontal and vertical segments

### Stroke Grouping

- In the first phase, we group a set of strokes which are intended to be part of a long horizontal line.
- In the second phase, for each group we find best matches to create new groups that potentially form an answer box.

### Finding left and right edges

- First, we find candidate strokes for left and right edges.
- Second, we group strokes that are extent of a line.
- Finally, we utilize a Hausdorff measure to find the best candidate for left and right edge of each group.

### Finding splitters

- For each group, we create a candidate set that contains all vertical segments inside the group.
- We pick two segments that spans most of the space between top and bottom edge of each group.

### Unit/Value/Problem id stroke grouping

- Based on the average x-coordinate of each inner stroke, strokes are grouped into unit, value and problem id groups.

### Answer box recognition

In order to distinguish answer box groups from non-answer boxes, following features are computed for each group:

1. **Group properties:** width and height of a group.
2. **Inner strokes:** number of strokes inside the group's bounding box.
3. **Crossing strokes:** ratio of the crossing strokes to the inner strokes.
4. **Diagonal length:** ratio of the diagonal lengths of transformed bounding box.
5. **Height of strokes:** Ratios of the heights of the of strokes in unit, value and part id to the height of BB.

We train a linear SVM classifier to find actual boxes.

## RESULTS

Performance of grouping methods:

Method	# of candidate groups	TP	Avg time per page (sec)
Ours	12,358	2211	3.6
Method [1]	358,191	1967	95.1
Method [2]	297,118	1834	93.2

Performance of methods in recognition of answer boxes.

Method	Accuracy
Ours	95.3%
Method [3]	88.6%
Method [5]	78.1%
Method [4]	81.5%

## CONCLUSIONS

Locating final answers to free response questions is an important step toward auto grading students' handwritten work. In this paper, we presented methods for locating students' answers. We were able to locate and recognize 95.3% of the 4473 answer boxes.

## REFERENCES

1. A. Delays and K. Lee, "A flexible framework for online document segmentation by pairwise stroke distance learning," *Pattern Recognition*, vol. 48, no. 4, pp. 1197-1210, 2015.
2. F. Pertender, M. Bresler, E.-M. Grossauer, J. Leong, and M. Haller, "cluster: Smart clustering of free-hand sketches on large interactive surfaces," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 2015, pp. 37-46.
3. L. Anthony and J. O. Wobbrock, "\$ n-protractor: A fast and accurate multistroke recognizer," in *Proceedings of Graphics Interface 2012*. Canadian Information Processing Society, 2012, pp. 117-120.
4. B. Paulson and T. Hammond, "Paleosketch: accurate primitive sketch recognition and beautification," in *Proceedings of the 13th international conference on Intelligent user interface*. ACM, 2008.
5. S. Cheema, S. Gulwani, and J. LaViola, "Quickdraw: improving drawing experience for geometric diagrams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM 2012, pp. 10371064