New Features for a Pen and Paper-based Exam Scripts Marking System

Nobuhiro Yoshida  
Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology, Japan  
Nobuhiro1986@gmail.com

Kenta Koyama  
Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology, Japan  
preko20@hotmail.com

Kal Ng  
Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology, Japan  
kalngyk@hotmail.com

Wataru Tsukahara  
Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology, Japan  
tsukahara@cc.tuat.ac.jp

Masaki Nakagawa  
Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology, Japan  
nakagawa@cc.tuat.ac.jp

Abstract: This paper presents an exam scripts marking system for answers written using pen and paper devices. Pen and paper devices capture the traces of pen movement (called digital ink), allowing the answers to be reorganized independent from paper format, analyzed and marked semi-automatically under the examiner’s supervision. Based on our earlier work we designed a system which provides new features. The first new feature is to order answers based on the recognized content from the digital ink. While digital ink recognition is still not perfect at present, it will at the very least allow examiners to examine each similar group of answers at a time. Another feature introduced in the new system is the ability to overlay answers given as drawings such as graphs so that examiners can easily find the consensus answer as well as common mistakes. Yet another feature is to analyze the time that each student spent on a question. This analysis is possible because in digital ink, each stroke carries a timestamp. The feature to automatically mark multiple choice questions in our earlier system continues to be available.

1. Introduction

The use of PDAs and tablet based PCs has been increasing rapidly in recent years. In addition to these, new kinds of devices have emerged. Some of these new devices allow users to write on paper instead of a tablet or a computer screen. These devices are sometimes known as pen and paper devices. They are used in almost exactly the same way as one would use pen and paper. Such a device tracks the position and movement of the pen and records this data into memory [1][2]. In general, a computer is not needed at the time of recording. The simple setup and familiar usage makes it immediately usable to even people who has never used a computer. Charts, graphs and illustrations which are typically difficult to input using a keyboard can also be inputted easily.

One of the most promising uses for pen and paper devices is in the marking of examination scripts, a traditionally tedious and burdensome task for examiners. Although there are semi-automated exam scripts marking systems, they typically work by scanning answer sheets into digital images and then performing OCR over the images [3].
However, OCR’s are not the perfect solution. Because of OCR’s lower accuracy on the recognition of the written characters, this kind of system architecture is usually used only for multiple choice question (MCQ) type examinations. Furthermore, the use of scanned images adds the extra burden of separating out the written answers from the other elements (e.g. the questions) already printed on the answer sheets. This mixture potentially reduces further the recognition accuracy. For this reason, such systems typically require the answers to be written on a separate sheet of paper. However, this time examiners have extra burdens of matching their answers to corresponding questions on a separate sheet of paper.

On the contrary, a pen and paper device does not require such a separation since it records only the written answers. It stores recorded answers as digital ink data (the traces of pen movement). The digital ink requires far less storage capacity than images for OCR. Furthermore, it can retrieve additional information such as the timestamp for each stroke (a time sequence of pen-tip coordinates), which is not available with OCR devices. Therefore we base our exam scripts marking system on pen and paper devices.

At present our aim is not towards a fully automated system often emphasized in conventional CAI and e-Learning but rather at a computer-assisted exam script marking system. We consider assisting examiners is more realistic than automatic marking, especially for certain types of questions. For example, questions where it is difficult to define the valid answers (e.g. essay writing), or questions where scores can be given to steps taken in deriving the answer even when the final answer is incorrect (e.g. mathematical questions).

2. New features

There are two main objectives we want to achieve with the system. One is to make script marking efficient, the other is to help understand the examinees’ response to each of the exam questions. We make use of the features of pen and paper devices to achieve these aims.

2.1. Features to enhance efficiency in exam script marking

It is a common practice for examiners to organize the answers to the same question from all the exam scripts and mark them at the same time. This gives the examiner an overall view of the answers, making it easy to compare answers and enabling better overall grading. However, when done with paper scripts, such a method would be inconvenienced with the additional work of organizing papers. In our earlier (pen and paper device based) system, we provided the function of automatically grouping together answers for the same question (Figure 1) [4]. Our test results showed that when the system is used, examiners can typically finish the exam marking in less than half the time needed when the system is not used.

![Figure 1](image-url) Answers for the same question are organized together

Based on this result, we decided to improve support for examiners in our new system so that the system can deal with questions for essay writing and those when the answers should be a certain patterns of graphs or figures.
Our basic idea is to show similar answers in a group and help examiners compare them and take balance in marking. For essay writing questions, answers can be clustered by their content using digital ink recognition (online handwriting recognition) methods and presented to the examiner in a sorted fashion (Figure 2). This would enable more convenient comparison between the answers. While digital ink recognition may make some errors, it will at the very least allow the examiner to examine each similar group of answers at a time.

For questions where the answers are given as graphs, the system overlap the graphs and through the use of different shades and thicknesses in the lines, enable the examiner to see the majority answer as well as common errors (Figure 3). The graphs can be clustered by their shapes and presented to the examiner in a sorted fashion, thus assisting the marking process easier and faster.

In both cases, by collecting similar answers into groups, we can give the examiner the choice to mark the answers in a group collectively.

Finally, in the case multiple-choice questions, markings can be automatically performed, as in our earlier system.

\[\text{Figure 2} \text{ Similar answers are grouped together}\]
2.2. Features to help understand examinees’ response to exam questions

The use of pen and paper devices allows our system to perform a task which was impossible with previous methods such as OCR. Because these devices typically add a time stamp for each recorded stroke, we can use this information to analyze the question answering process of the examinees. More precisely, the use of a pen and paper device allows us to know the following activity during answering:

1. Revisions in the answer given for a question,
2. The time taken in answering a question,
3. The order in which answers are written,
4. Unusually long elapsed time between strokes.

We now analyze each of the above information to make some predictions.

1. Revisions in the answer given for a question
Many people perform routine checks on their answers after finish answering all the exam questions. This allows them to catch mistakes which may be due to carelessness. In the case that a revised answer corrects an original mistake, it is possible that the original mistake was due to carelessness. However, in the case that a revised answer incorrectly modified an originally correct answer, the underlying reason of the revision would be worthy of further investigation. For example, revisions in math exams mean simple calculation mistakes in most of the times. Thus we can suspect when there are several revisions the examinee has a difficulty in calculation. In such cases posing intensive calculation work to the examinee might be effective. This way we can choose adaptive workload using revision information.

Figure 3 Graphs overlapped to help visualization
(2) The time taken in answering a question
The time that the examinee spent on giving an answer can be obtained using the timestamps of the last stroke of the answer and the last stroke of his previous answer (provided that no question has been skipped). This information allows us to estimate how involving the question is to the examinee. In the case that a question requires very short time for most of the examinees, it means that the question is easy to solve. In the case that a question consistently requires a longer time to complete, the examiner may want to consider alternative ways to test the examinee’s understanding of the question’s underlying concepts (e.g. with smaller sub-problems). This modification would result in more balanced questions in terms of the time requirement. This way, examiners can see if the questions were adequate for examinees and improve their questions for the next exam based on the objective duration data.

(3) The order in which answers are written
In the case that the questions are not answered in the order they are given, it is usually the case that the examinee has decided to skip a question for the time being (either giving it up or to return to it later). This could be due to problems in understanding the question, or unfamiliarity with the question. The examinee might also have decided that answering the question would take up too much time and it would be better to take on less involving questions first. If most examinees took such strategy, for a question, examiner can improve the test by giving more time for that question or replace that question to more appropriate one.

(4) Unusually long elapsed time between strokes
An unusually long elapsed time between two strokes while answering a question could indicate that the examinee is stuck at a certain step. This could indicate difficult parts in the subject which may require more emphasis during lectures.

We designed our system based on the above considerations. Our predictions in (1)-(3) would be examined using the system.

The system reports to the examiner the information for (1)–(4). To visualize (4) we colorize the answers to indicate the positions where the strokes involved unusually long elapsed times. In Figure 4 we show an example where the correct answer is the word “puppy”. The colorized position shows the stroke which took longer than a threshold time to be written.

![Figure 4](image-url)
3. System design

3.1. System usage
Figure 5 shows the overall data flow in our system.

The examinees would sit for an examination just as they would normally, except that they would be using pen and paper devices in the place of ordinary pens. After the examination, the devices are connected to a computer and the answers are read into the computer.

Examiners only use PC in marking process. The system reorganizes examiner’s reorganized answers and displays them on a screen. The examiner marks to each question while seeing it. Marking can be completed by using mouse click. Marking result is stored in external files.

Our system is compatible with any pen and paper device which interfaces with the computer using the InkML file format. InkML is an XML compliant file format for recording the coordinates and timestamps of strokes, and is a standard drafted by the W3C.

3.2. A overview of our system design
The design of our examination scripts marking system is as follows (Figure 6).
### 3.3. Grading of examination answers

The system will first need to decide, (1) to which type each question belong, and (2) how should the ink data be interpreted (i.e. whether it is a graph answer, or a multiple choice question answer, etc.). This is done by comparing the coordinates of the strokes with a pre-defined template which specifies the writing area and the type for each question. The processing for each type of answer is treated differently. If the strokes correspond to words, they are given to a handwriting recognition engine to be interpreted. We use our in-house recognition engine for this purpose [5]. The organizing and display of different answer types are also different. Graph answers are overlapped and displayed to the examiner. Similar graphs are also shown clustered to allow collective grading. Sentences are clustered and displayed in a sorted fashion. Answers for multiple choice questions can be set to be automatically graded. After grading is completed for all the questions, the final score for each examinee is calculated and outputted.
3.4. Analysis of the answering process
As discussed in Section 2.2, we use the timestamps which accompany the strokes to analyze the answering process. Strokes which took unusually long time are highlighted in different color. The system also looks for places where revisions have occurred, or where the answers took more time to complete, as well as the answers which are given out of order, and report these occurrences accordingly. We also look for the correlations between these parameters and the internal states of examinees as we discussed in 2.2.

4. Summary
In this paper we described our improved system employing a pen and paper device to assist the marking of examination scripts. The system automatically performs the task of organizing exam answers by the individual questions, as well as group similar answers together, making the grading process more efficient. Furthermore, the system allows examiners to analyze how examinees respond to each of the examination questions. Evaluation remains to be done. We are interested in how the new features are useful for marking exams and how the examinees’ answering process is useful to know how deeply they understood the material and also is useful to improve effectiveness of exams.

Acknowledgement
This work is being partially supported by the R&D fund for "development of pen & paper based user interaction" under Japan Science and Technology Agency.

References