Context-bounded Refinement Filter Algorithm: Improving Recognizer Accuracy of Handwriting in Clock Drawing Test

Hyungsin Kim, Young Suk Cho, and Ellen Yi-Luen Do

GVU Center, School of Interactive Computing, and Health Systems Institute
Georgia Institute of Technology, Atlanta, Georgia USA, 30332
{hyungsin, ycho47, ellendo}@gatech.edu

Abstract

Early detection of cognitive impairment can prevent or delay the progress of cognitive dysfunction. In the field of neurology, the Clock Drawing Test (CDT) is one of the most popular instruments for detecting cognitive impairment. This paper presents the development of the ClockReader system, a computerized Clock Drawing Test. The main function of the system is to automate error handling in handwriting recognition. Since the ClockReader is a screening tool for dementia, it is not desirable to ask the users to fix their input errors in the2 drawing of either numbers or characters. Therefore, we propose a simple machine learning technique, context-bounded refinement filter algorithm. With trial experiments, we prove that this simple algorithm improves the recognizer accuracy of handwriting in clock drawings up to 88%.

Introduction

Handwriting data recognition systems can be easily found in our everyday lives. For example, the United States Postal Service (USPS) began to deploy its first handwritten address-reading prototype in 1997 (Mauk 2007). Now, the large majority of letters are sorted entirely by computers, and the success rates are above 90 % (Srihari, 2007). Another example is banking, we now see ATM systems automatically recognizes our handwritten numbers in check deposits. Both postal-address interpretation and bank-check processing are based on off-line recognition systems, using Optical Character Recognition and Intelligent Character Recognition.

With the advent of pen-based computing, more research efforts are focusing on making online systems for recognizing handwriting data (Pittman 2007, Tappert et al. 1990). Most research has focused on increasing recognizer accuracy, as well as error recovery mechanisms (Shilman, Tan, and Simard 2006). Unlike offline recognition, this system requires real-time character recognition and instant error handling of the inaccurately recognized characters. The most popular error-fixing methods ask users to manually change the erroneous character to the intended one. This tedious work can definitely help users fix their handwriting input. However, sometimes, users are not able to fix the recognition errors due to the purpose of the system.

In this paper, we present our approach so as to improve the recognizer accuracy of handwriting drawings in the ClockReader system. This system is a computerized screening tool for people with dementia. In order to identify people with dementia, the most popular method is to conduct a simple Clock Drawing Test (CDT). By integrating the CDT administration in a computer system, we have focused on the development of drawing recognition of handwritten characters and the automated evaluation. Before we present our proposed method, we will briefly discuss the paper-and-pencil based Clock Drawing Test and the ClockReader system. Then, we will report our preliminary data analysis of 65 handwritten drawings provided by the Emory Alzheimer's Disease Research Center (ADRC). Based on the analysis, we propose a filtering algorithm to increase the accuracy rate and its experimental use. Finally, we will summarize our results and conclude the paper with future directions.

Clock Drawing Test

The Clock Drawing Test, CDT, is a popular cognitive impairment screening tool for people with dementia (Rabins 2004). Different from other dementia screening instruments, CDT reveals the person's visual-spatial, constructional, and higher-order cognitive abilities, including executive aspects (Maruish 1997). It is a complementary approach to the verbally focused dementia-screening tools (such as a three-item recall test) heavily administered (Sunderland 1989). Libon and others suggested that CDT may provide a complementary assessment of other aspects of neuropsychological functioning (Libon et al. 1993).

By simply asking people to draw a clock, it easily identifies people with dementia (Ismail, et al. 2010). Clock drawings from people with dementia frequently show missing or extra numbers, or misplaced clock hands (Freedman, et al. 1994). Figure 1 shows three different clock drawings from three patients (Freedman, et al. 1994). The drawing clearly shows the degradation of the patient's cognition. Interestingly, the patient could not use the space of the clock evenly. Sometimes, due to impairment of part of the brain, patients' drawings are represented by using only one-half of the clock circle (Smith 2009).

Figure 2 shows an example of Allochiria in the clock drawing of a patient with hemi-spatial neglect. The patient omitted the left side of objects when drawing a clock. Even though the patient could verbally express that the clock face has a left side, he or she would fail to notice that the drawing was incomplete. This implies that drawing tasks can play an important role in differentiating the specific impairment of the brain lesion, and not just saying that a patient has dementia.

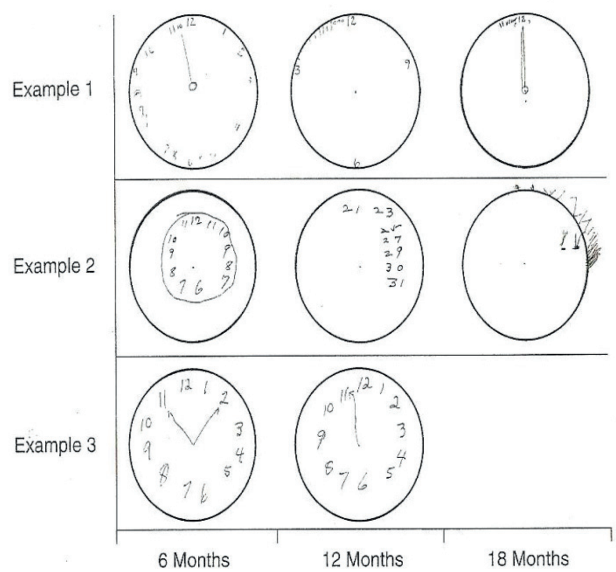


Figure 1: Three examples of clock drawings showing deterioration in dementia

Current practice of the CDT is administered with traditional analog media. In a CDT, patients are asked to use paper and pencil to draw a clock face. Neuropsychologists or neurologists then spend hours to analyze and score the tests. Different scoring systems use slightly different instructions and methodologies for administering the CDT (Pinto and Peters 2009). Patients are asked to draw a clock face in a pre-drawn circle and place all of the numbers on it. Then set the time to 10 past 11. The process is long and tedious. To reduce the tedious efforts of human scoring, and to facilitate a consistent scoring practice and analysis, it is critical to develop a computerized system to conduct this screening process.

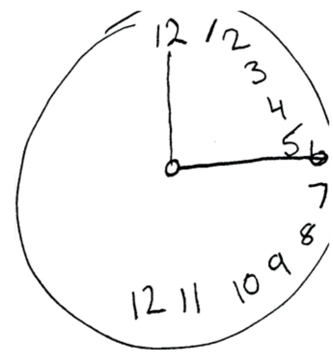


Figure 2: An example of a Clock Drawing from a patient with hemi-spatial neglect

Moreover, with the growing societal phenomenon of our aging population, CDT is becoming popular in hospitals, as well as in retiree communities (Strauss 2000 and Shulman 2006). However, administering CDT by humans is not only time-consuming, but also error-prone. Therefore, a computerized tool will be able to provide more frequent access to testing, while reducing the time the clinical staff will need to perform the analysis. In the next section, we will introduce the ClockReader System and the report of our preliminary data analysis of handwritten drawings from patients.

ClockReader System

The purpose of the ClockReader System is to enable patients to take the Clock Drawing Test without the presence of a human evaluator. The overarching goal of this system is to identify early dementia, delay or prevent the progression of the disease and increase the quality of life for aging people.

The system consists of three main components: data collection, sketch recognition, and data analysis. First, the system should record and recognize a patient's freehand drawing and collect the data. Then, based on the scoring criteria, the system should automatically analyze the drawing and report the score.

Due to the inherent ambiguity of handwritten data, it is much easier to use the context in order to improve a recognizer's accuracy. We name it the context-bounded (specific) recognition approach. Context-bounded recognition relies on the recognizer's processing specifically in a given situation. In the ClockReader system, the context would be bounded for drawing a clock, which means that people would mostly use alphabetical numbers and lines for depicting clock hands. This allows us to develop several specific algorithms.

- Algorithms for recognizing digits and clock hands distinguished from unnecessary strokes

- Algorithms for recognizing the numbers from 1 to 12, together with each number's coordinates: The coordinate helps to distinguish a single digit from double-digit numbers considering the sequence of each number's before-and-after position

- Algorithms for automatically calculating the CDT score results, based on pre-programmed criteria

- Algorithms for excluding unnecessary strokes during the process of evaluation

Understanding Users

When we design and develop a system, especially in clinical settings, there are always challenging factors. One challenge in developing the ClockReader System is that we should take into consideration two different target users. Our main users are patients who need to take the Clock Drawing Test, and clinicians who administer the test and examine the results. Our expectations in system usage considerations for patients and clinical staff are different.

From a patient's perspective, the goal is to offer the affordance of a paper-and-pen environment. Using a stylus on the surface of a Tablet PC is similar in form to using a pen on a piece of paper. From a clinician's perspective, the goal is to offer a well-organized data collection tool, as well as an automated analysis of the results. With a computerized system, doctors or clinical staff may gain easier access and more accurate information about the progress of patients' cognitive impairment. The computerized clock-drawing tests could be performed frequently without requiring the presence of a test administer. Another benefit of a computerized system is to provide a consistent yet customizable scoring for a more general analysis with high inter-rater reliability. Unlike the simple interface designed for patients, the interface for doctors should involve data representation and information visualization to meet a wide variety of needs.

The ClockReader is developed in C# programming language and is supported by "Microsoft Windows XP Tablet PC Edition Software Development Kit 1.7" and "Microsoft Visual Studio 2008. Figure 3 shows a screen shot of our ClockReader system for patients. The Patient User Interface (UI) is very simple. The only pre-drawn circle will be shown in the interface. After clinician's instruction, a patient will construct a clock on the pre-drawn circle with a stylus. Figure 4 shows a screen shot of ClockReader System User Interface (UI) for clinicians. The main purpose of this UI describes the results of Clock Drawing Test. The UI consists of three components: An individual patient's information, their drawing, and the scoring result of the drawings. Current UI for clinicians only implemented one criterion. In the future, we plan to implement several different criteria in the ClockReader System.

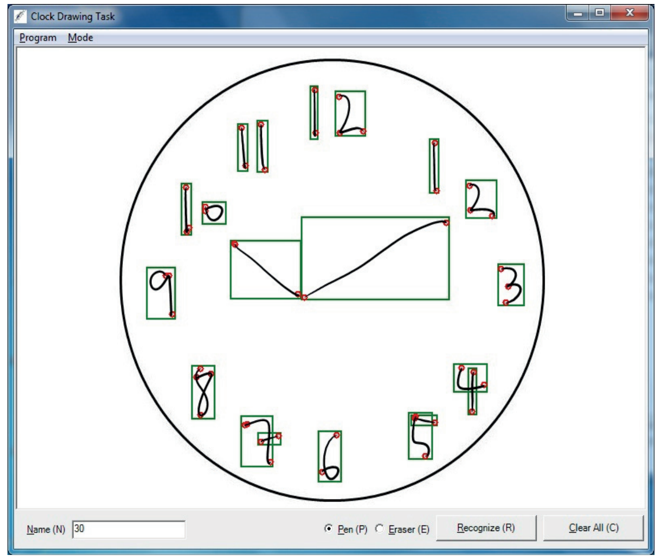


Figure 3. Screen shot of ClockReader System for Patients

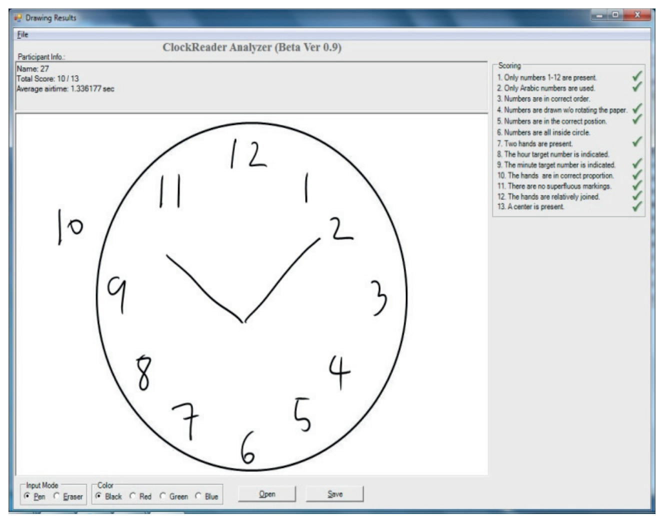


Figure 4. Screen shot of ClockReader Analysis for Clinicians

In the following subsection, we will report the preliminary data analysis, and a different criteria analysis will follow.

Empirical Data Analysis

We collected 65 handwritten clock drawings provided by a local Alzheimer's center. The drawings were randomly chosen from normal aging people to severe dementia patients. Thus, the scoring also varied from 2 to 13.

For this analysis, we chose Freedman et al.'s 13-score criteria (Freedman et al. 1994). Please see the Table 1 for the detailed criteria index. People with a score of 13 means that there is no cognitive impairment and their clock drawings are generally intact. Data sometimes include several drawings per one individual person through multiple years. This practice shows how the CDT visually provides strong evidence to support the progressive degeneration of one's cognition.

Table 1. Example of Evaluation Criteria of CDT

Numbers 1. Only numbers 1 – 12 are present (without adding extra numbers or omitting any) 2. Only Arabic numbers are used (no spelling, e.g., "one, two" no roman numerals) 3. Numbers are in the correct order (regardless of how many numbers there are) 4. Numbers are drawn without rotating the paper 5. Numbers are in the correct position (fairly close to their quadrants & within the pre-drawn circle) 6. Numbers are all inside the circle
Depiction of Time (Hands) 7. Two hands are present (can be wedges or straight lines; Only 2 are present) 8. The hour target number is indicated (somehow indicated, either by hands, arrows, lines, etc) 9. The minute target number is indicated (somehow indicated, either by hands, arrows, lines, etc) 10. The hands are in correct proportion (if subject indicates which one is which after "finishing", have them fix the proportion until they feel they are correct) 11. There are no superfluous markings (extra numbers or errors on the clock that were corrected, but not completely erased, are not superfluous markings) 12. The hands are relatively joined (within 12mm; this does not need to happen in the middle of the circle)
Center 13. A Center (of the pre-drawn circle) is present (drawn or inferred) at the joining of the hand

According to Figure 5, there were 8 people in the range of 1 ~5, 14 people in the range of 6 ~ 10, and finally, 43 people in the range of 10 ~13. In a brief summary, our data consist of drawings from 22 people, somewhat seriously cognitively impaired, and 43 people with mild cognitive impairment due to aging. The x-axis represents the total CDT results, and the y-axis represents the number of people who achieved each score.

Interestingly, we found that even people with dementia wrote the digits with some sequences. However, the sequences frequently missed some digits or added some unnecessary ones. This means that they wrote the numbers with sequences such as 1, 2, 4, 6, 7 or 1, 1, 2, 3, 4, 5, 5, 5, 6. This requires the system to recognize how many digits were written and what were the missing/duplicate numbers.

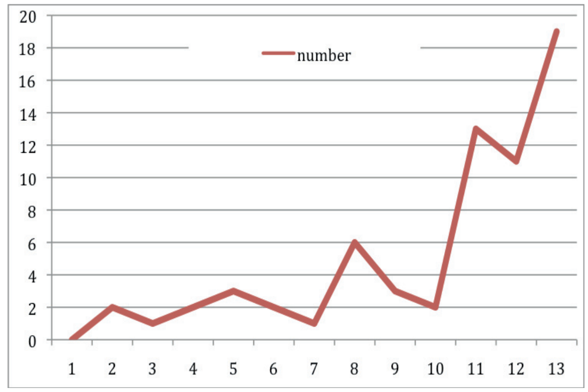
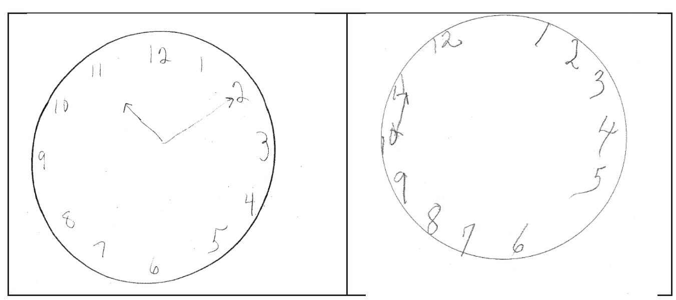


Figure 5. Number of people by Score

From our dataset, the one of the most common error patterns is related to setting the time at ten after eleven (11:10). Figure 6 shows five different clock drawings which incorrectly set the time at 10:11. The first clock drawing in the top left column only shows setting a correct time. The five clocks show different ways to indicate time in an incorrect way.

Patients with Alzheimer's disease frequently set a time at 10 to 11. Freedman et al argues that the time setting requirement places is difficult for people with Alzheimer's disease (Freedman, et al. 1994). Furthermore, research shows that stimulus-bound responses are more common among Alzheimer's disease patients compared to normal elderly and patients with frontal lobe dementia (Cahn et al. 1996 and Blair et al. 2006).

Furthermore, we found that all of them used Arabic numbers rather than spelling them out (2 rather than two). Therefore, we decided that the fundamental way to increase recognizer accuracy was to focus on correcting inaccurate recognition of numbers as characters. In the next section, we will describe the algorithm to shift inaccurately recognized characters into appropriate digits, with the experimental results showing improvement of the recognizer's accuracy rate.



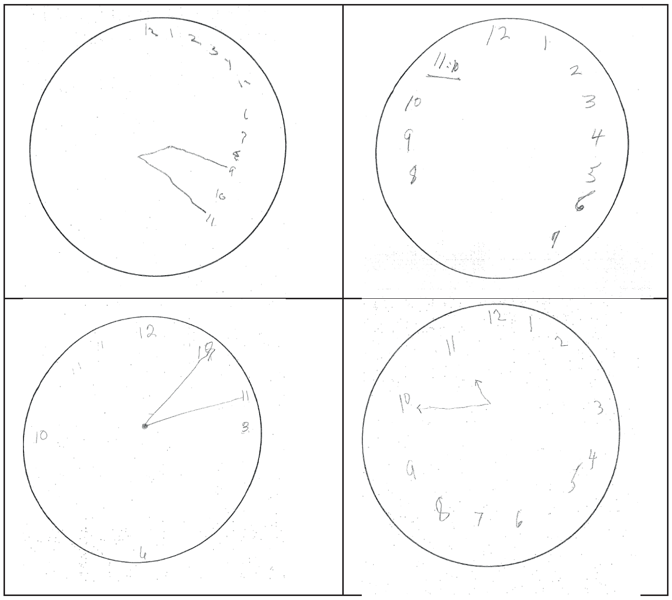


Figure 6. Six Clock Drawings which shows setting a time at 11:10

Improved Recognizer Algorithm

The ClockReader is developed based on the Microsoft Tablet PC recognizer. The current Tablet PC SDK provides character recognition through the stroke level. However, rather than capturing each character based on a stroke, we modified it to capture the character level. Some Arabic numbers include more than two strokes to write. For example, the numbers 4, 5, 7, (and sometimes 8, depending on a person's writing style) requires at least two strokes. Therefore, the recognition process of the ClockReader System starts from setting a rectangular area per character, passing the data from the rectangular area to the Microsoft SDK handwriting recognition engine, and then finally saving the recognized results as a string.

This recognition process can be ideal if there are no recognized errors. We all know that most recognized errors fundamentally come from the system's recognizer engine. The engine excludes the contextual understanding of handwritten data. For example, humans can understand "1," if it is written inside of the clock as an Arabic number however badly it is written. Nevertheless, the system can sometimes understand the number "1" as the letter "l" or the symbol "l." By providing some context, for example, drawing a clock, we can increase the system recognizer's accuracy.

Figure 7 illustrates the overview of how a user's drawing is processed through two sketch recognition engines. When a user begins to draw, the generic recognition engine attempts to recognize. Then, the output will pass to a domain specific filtering algorithm. Thus, the systems recognition rate will improve.

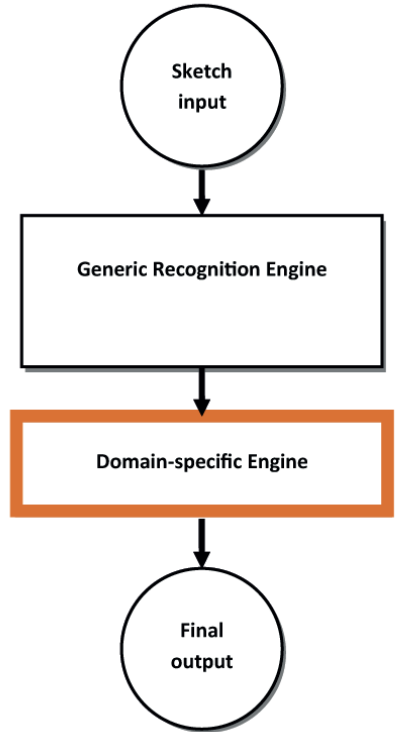


Figure 7. Overview of ClockReader Recognition Engines

The Machine Learning algorithm consists of four processes. The first step is to create two-text-files that are saved in a database. The first-text-files are used to create the results of miss-recognized data per individual digit as "error_pool_<number>.txt" files. Then the second-text-file to create "error_data_<number>.txt" files is based on the error frequency. If a specific error frequently happens more than five times, we program the recognized character to be converted into an appropriate Arabic number.

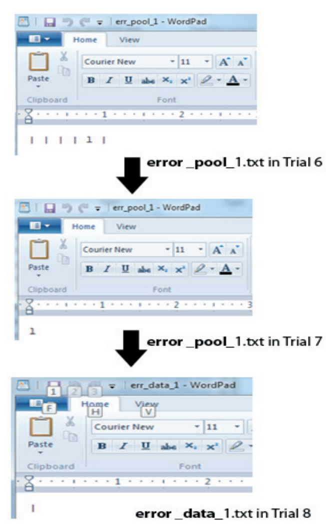


Figure 8. Process to create error data from

In short, the system can improve recognizer accuracy through continuous accumulated error data. In the following section, we will discuss the experimental results of this context-bounded refinement filter in the ClockReader System. Figure 8 demonstrates the process how a misrecognized character has saved in the error database file form the error pool file.

Experiment Results

We ran the program 20 times to conduct the experiment. The goal of the experiment was to see how the recognizer's accuracy rate improved. Table 2 shows the results of the trial analysis. The accuracy rate increased from 53% to 73% in the 7th trial because "l" was saved as "1" in the error database. Similarly, in the 12th trial, "n" had been saved as "7" in the error database. "O" and "p" had been saved as "0" and "8," respectively in the error data after the 15th trial. The Recognition Accuracy graph in Figure 9 also shows the progressive improvement of the accuracy rate by the number of trials.

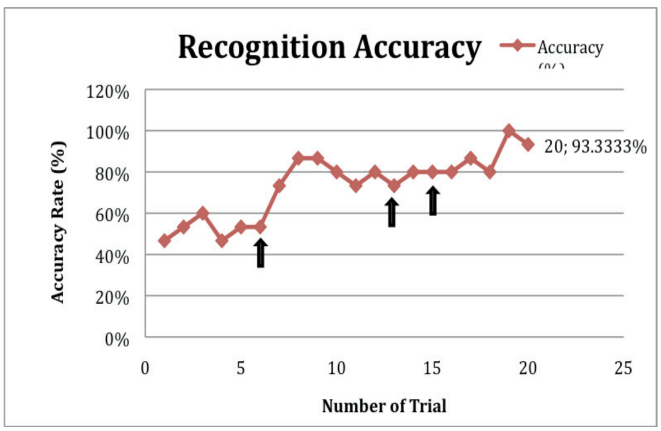


Figure 9. Recognition Accuracy Graph

Table 2. Trial Analysis

Trial #	# of Correct Recognitions	Total # of Written Numbers	Accuracy (%)
1	7	15	46.6667%
2	8	15	53.3333%
3	9	15	60.0000%
4	7	15	46.6667%
5	8	15	53.3333%
6	8	15	53.3333%
7	11	15	73.3333%
8	13	15	86.6667%
9	13	15	86.6667%
10	12	15	80.0000%
11	11	15	73.3333%
12	12	15	80.0000%
13	11	15	73.3333%
14	12	15	80.0000%
15	12	15	80.0000%
16	12	15	80.0000%
17	13	15	86.6667%
18	12	15	80.0000%
19	15	15	100.0000%
20	14	15	93.3333%

Out of running the program 20 times, trials from the 1st to the 6th were in the process of continuously collecting data in the error pool. Thus, there were no data saved in the database, and the filtering process was not initiated. However, after the 6th trail was complete, the frequent error, recognizing the number "1" as the character "l," was saved in the error database because the number of the total error counts was met after the 5th time. Therefore, the 7th trial indicated an improved accuracy rate. It is also interesting to see that by simply adding "l" into the error database, the accuracy rate incredibly improved from 53.33% to 73.33 %. When people construct a clock, they need to write the number "1" five times for five instances: 1, 10, 11, 12.

Table 3 shows accuracy rate changes based on the three critical points where the data are updated. After "1" was added, the most critical errors, there were not many improvements shown. However, the program is based on a learning system: by running it more, more error data can be added, and overall, the accuracy rate is increased. More importantly, the improved accuracy rate keeps a stable status. However, the limitation of this algorithm is that it only filters a number when it is miss-recognized as a character. Thus, we need to improve this recognition engine with other expected errors. One example is that the system can recognize a user's handwritten number as an unintended number.

Table 3. Changes of Accuracy Rate

Section 1 7 to 11 (5 Trials) (Data added on 7th trial)	Section 2 12 to 16 (5 Trials) (Data added on 12th trial)	Section 3 17 to 20 (4 Trials) (Data added on 15th trial)
80.0000%	78.6667%	88.0000%

Handwriting recognition errors in the ClockReader System can be categorized into two cases. The first most common case is that the system would recognize handwritten digits as a character or symbols. Another case could be that the system would recognize a user's handwriting digit as another unintended digit. Adding two domain specific algorithms can solve those two different types of misrecognitions.

The results of the experiment in this section demonstrated that the machine learning algorithm has improved the accuracy rate by filtering the unexpectedly errors. However, it is only an applicable solution for the first error case. The second error case is still a remaining issue to solve. In order to solve the problem which is the system could recognize a user's handwritten number as an unintended number, we plan to add another algorithm after executing the first refinement filter algorithm. The second refinement filter algorithm would be developed by the same way to convert characters to digits. Figure 10 shows the overview of the ClockReader recognition process with enhanced domain specific engine.

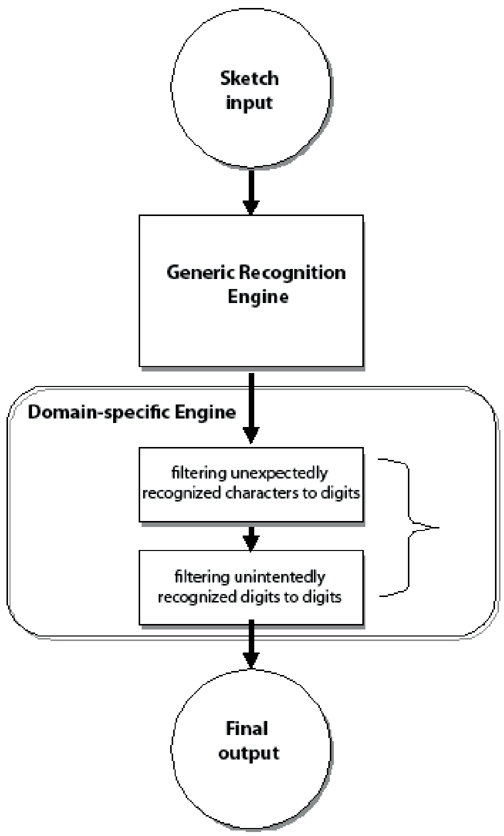


Figure 10. Overview of Revised Recognition Engine for ClockReader System

Conclusion and Future Directions

We have developed the ClockReader system to make a paper-and-pencil based Clock Drawing Test easier, more efficient and effective. In this paper, we proposed an improved recognizer algorithm through a context-bounded refinement filter. With the 20-time-trial experiment, we learned the positive possibility of accuracy improvement to be 80% on average. For the future, in acknowledging the current algorithm limitation, we will plan to improve it with a combination of different algorithms by adopting the strokes' endpoint coordinate information of the characters. After completion of implementing improved algorithm, we will plan to conduct a usability comparison study. The results of the study will enable us to compare the outcomes using computerized CDT test with the outcomes of Pen-and-Paper version of CDT. Ultimately, we establish the validity of the computerized CDT.

Acknowledgments

We would like to thank Health Systems Institute at Georgia Institute of Technology and Emory Alzheimer's Disease Center to their financial support for conducting this study. We also like to appreciate volunteered people for CRIN (Clinical Research in Neurology) to provide us valuable data.

References

Blair, M., Kertesa, Al, McMonagle, P., Davidon, W., & Bodi, N. 2006. Quantitative and qualitative analysis of clock drawing in frontotemporal dementia and Alzheimer's disease. Journal of International Neuropsychological Society, 12, 159 -165.

Cahn, D. A., Salmon, D. P., Monsch, A. U., Butters, N., Widerholt, W. C., & Corey-Bloom, J. 1996. Screening for dementia of the Alzheimer type in the community: The utility of the clock drawing test. Archives of Clinical Neuropsychology, 11, 529 – 539.

Freedman, M., L. Leach, et al. 1994. Clock Drawing: A Neuropsychological Analysis. Oxford, Oxford University Press.

Ismail, Z., T. K. Rajji, et al. 2010. "Brief cognitive screening instruments: an update." Journal of Geriatric Psychiatry 25(2): 111-120

Libon, D., S. RA, et al. 1993. "Clock drawing as an assessment tool for **dementia.**" Archives of Clinical Neuropsychology 8: 405-415.

Maruish ME, ed. 1997. Clinical Neuropsychology: Theoretical Foundations for Practitioners. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mauk, Ben. 2007. LiveScience.com from http://www.lifeslittlemysteries.com/how-do-post-office-machines-read-addresses-0445/

Pittman, James A. 2007. Handwriting Recognition: Tablet PC Text Input. Computer, pp. 49-54, September

Pinto, E. and R. Peters. 2009. Literature Review of the Clock Drawing Test as a Tool for Cognitive Screening.

Dementia and Geriatric Cognitive Disorders. 27(3): p. 201-213.

Rabins, P.V. 2004. Quick cognitive screening for clinicians: Mini mental, clock drawing, and other brief tests. Journal of Clinical Psychiatry. 65(11): p. 1581-1581.

Sargur N. Srihari. 2007. Postal Research. http://www.cedar.buffalo.edu/~srihari/Postal-Research.

Shilman, M., Tan, D. S., and Simard, P. 2006. CueTIP: a mixed-initiative interface for correcting handwriting errors. In *Proceedings of the 19th Annual ACM Symposium on User interface Software and Technology* (Montreux, Switzerland, October 15 - 18, 2006). UIST '06. ACM, New York, NY, 323-332.

Smith, Alastair D. 2009. On the Use of Drawing Tasks in Neuropsychological Assessment. Neuropsychology 23(2):231- 239.

Strauss, E., E. M. S. Sherman, et al., Eds. 2006. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, Oxford University Press.

Shulman, K. 2000. Clock drawing: Is it the ideal cognitive screening test? International Journal of Geriatric Psychiatry. 15(6); 548-561

Sunderland, T. et al. 1989. Clock drawing in Alzheimer 's disease: A novel measure of dementia severity. Journal of the American Geriatric Society. 37:725~729

Tapper, Charles C., Suen, Ching Y., and Wakahara, Toru "The State of the Art in On-line Handwirting Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 12 No 8, August 1990, pp 787-ff, http://users.erols.com/rwservices/pens/biblio90.html#Tappert90c, retrieved May 4, 2010