Efficient Search Strategy in Structural Analysis for Handwritten Mathematical Expression Recognition

Taik Heon Rhee, Jin Hyung Kim
Division of Computer Science,
Korea Advanced Institute of Science and Technology
three@ai.kaist.ac.kr, jkim@cs.kaist.ac.kr

Abstract

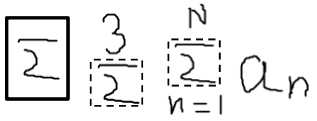
Problems in local ambiguities in handwritten mathematical expressions are often resolved at the global level. For a well performing recognizer, multiple local hypotheses should be kept as long as possible until the ambiguities are resolved by a global analysis. We propose a layered search framework for handwritten mathematical expression (ME) recognition. From given handwritten input strokes, ME structures are constructed through adding a symbol hypothesis one by one, considering every possible symbol identity and spatial relationship with the existing structure. A cost reflecting the likelihood of a structure is estimated for each newly expanded layer so that a best-first search algorithm is applied to seek the most likely structure. The elegance of our method is in that while all the possibilities are examined, the search complexity is made manageable by applying admissible heuristics. Further complexity reduction is achieved by delaying the symbol identity decision. Unless a symbol identity causes structural alternatives for the remaining input strokes, the identity can be determined after the complete structure is fixed. Such a delayed decision reduces undesirable search space expansion. In an implementation targeting high school level MEs, our method achieved high speed with a high level of accuracy which resulted from the system's capacity to examine a large number of possibilities.

Keywords: Handwritten mathematical expression recognition, Structural analysis, Layered search tree, Admissible heuristic, Delayed decision of symbol identity

1. Introduction

Since handwriting is a convenient and natural way to input mathematical expressions (MEs) into computers, and the recognition of handwritten ME has been studied for several decades, it is still an area of research with many challenges. The difficulty mainly comes from the two-dimensional nature of ME structures and the large variations in shape commonly observed in handwriting.

A part of a handwritten ME can yield more than one interpretation if we see the shape in isolation. We call this *local ambiguity*. In the example in Figure 1, the handwriting input in the solid box can be interpreted as either '$\sum$' or a combination of '$-$' (fraction) and '2'. The local ambiguity, however, is easily resolved at a global level. As shown in Figure 1, the part becomes clearer when it is analyzed along with its neighbors. Therefore, the possible interpretations should be kept as local hypotheses until the ambiguity is resolved by a global analysis.



(a) (b)

Figure 1. An example case of local ambiguity. The handwriting input in the solid-lined box would yield more than one interpretation. (a) combination of '$-$' (fraction) and '2'. (b) '$\sum$'

Keeping multiple local hypotheses requires a huge amount of storage and computation even in the case of relatively small MEs. While the recognizer should generate as many local hypotheses as possible in order to cover all the writing variations, it should also generate as small local hypotheses as possible to reduce the requirement. As a result, a compromise needs to be made between complexity and performance.

Previous work has usually divided the ME recognition into three phases: *symbol segmentation*, *symbol recognition* and *structural analysis* [1]. While a few attempts tried to improve the performance of symbol segmentation itself [5, 6], most attempts have assumed that symbols are well segmented from each other [2, 3, 4]. This assumption makes the recognizer's job easy because segmentation reduces the number of local hypotheses. However, it is too strong an assumption because obtaining an acceptable segmentation from natural handwriting samples is a difficult task, and the errors occurring in the segmentation phase cannot be easily recovered in latter ones.

Grammars have been frequently applied in ME recognition study [7, 8, 9]. However, the use of grammar alone

is not effective in reducing the number of local hypotheses in handwritten ME recognition. This is because many alternative interpretations still have to be analyzed even after taking grammars into consideration. Furthermore, irregularity in the symbol writing order in MEs makes the application of the grammar difficult. Therefore, [7] and [8] implicitly assumed a given writing order, while [9] mainly targeted printed MEs.

In this paper, we focus the handwritten ME recognition problem on the search for the most likely structure of a given handwritten set of input strokes. Initially the structure is null. Then the structure is expanded incrementally by adding a symbol hypothesis one by one. The symbol hypothesis is made by consuming the input strokes with respect to every possible symbol identity and their spatial relationship with its parent structure.

Such expansion creates a layered search tree where nodes represent the partial structures of the ME so far constructed. For each newly expanded layer, a cost reflecting the likelihood of the structure is estimated so that the best-first search algorithm [10] may be applied.

The elegance of our method is in that while all the possible structures are examined, the search complexity is made manageable by applying an admissible heuristic which reflects the likelihood of the partial structure to accept unanalyzed remaining input strokes.

Moreover, further search space reduction is achieved by delaying the symbol identity decision. Unless a symbol's identity causes structural alternatives, the identity can be determined after the entire structure is fixed. The delayed decision not only reduces undesirable search space expansion, but also allows the symbol identity decisions to be done at the global level.

The rest of the paper is organized as follows. Section 2 describes the proposed layered search framework for ME recognition. Section 3 explains how to represent a ME interpretation, and section 4 describes the cost function to estimate the likelihood of the interpretation. Section 5 explains strategies for more efficient searches. The experimental results are shown in section 6, and section 7 discusses conclusions and future work.

2. Layered Search Tree for ME Recognition

ME recognition is formulated as searching for the most likely interpretation S^* for the input X. It is equivalent to finding the minimum cost interpretation if we define the cost of an interpretation $C(S)$ as the negative log likelihood of the interpretation S, as shown in Eq. 1. However, enumerating all the possible interpretations $\{S\}$ is infeasible because the number of possible interpretations is huge even for a moderate size ME recognition problem.

$$S^* = \arg \min_{S \in \{S\}} C(S) \tag{1}$$

To avoid the enumeration, we incrementally construct a layered search tree to apply the best-first search algorithm [10]. In the layered search tree, each node N_i holds an interpretation S_i and remaining input strokes $X \setminus S_i$. An interpretation is called *complete* if it has consumed all the input strokes, i.e., the remaining input strokes are empty. Otherwise, it is called *partial*. The edge denotes the expansion of the partial interpretation by adding a symbol. Each edge holds an added symbol identity with the consumed strokes, and the spatial relationship of the symbol to the parent interpretation.

Initially the search tree starts with the null interpretation. Then the interpretation is expanded by adding symbol hypotheses. Several alternative interpretations would be created according to the symbol identity, consumed strokes to make the symbol, and the symbol's spatial relationship to the existing interpretation. Such alternatives create corresponding branches in the search tree. The node expansion is terminated when a node has a complete interpretation.

Figure 2 shows an example of a layered search tree. When handwritten input strokes are given as in Figure 2(a), the layered search tree is constructed as in Figure 2(b). Figure 2(c) shows what a node and an edge hold. As a node is expanded, its interpretation grows, adding a symbol one by one. Correspondingly the remaining input strokes are reduced.

The best-first search algorithm is applied on the search tree. An evaluation function is needed to define the '*best*' for the algorithm. The evaluation function $f(N_i)$ of a node N_i is defined as:

$$f(N_i) = C(S_i) + h(N_i), \tag{2}$$

where $C(S_i)$ is the cost of the partial interpretation S_i which is the negative log likelihood, and $h(N_i)$ is a heuristic function which estimates the additional cost of the partial interpretation S_i to be a complete interpretation.

If the $h(N_i)$ is always lower than the true remaining cost, we call the heuristic *admissible* [10]. With admissible heuristics, the best-first search algorithm guarantees that the goal found first is the minimum cost. In our case, the complete interpretation found first is the minimum cost one, that is, the most plausible interpretation of the given handwritten input strokes.

3. Interpretation Representation

An interpretation of a ME is a structure which holds the component strokes and their relationships. We represent the structure as a symbol relation tree (SRT) [11].

A SRT is formed by a dominant symbol and its subordinated sub-expressions as nodes. The dominant symbol is a prime symbol which dictates the overall structure of the expression. The spatial relationships between

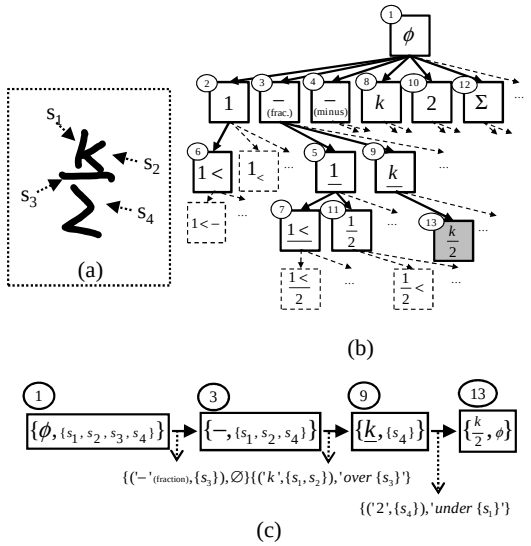


Figure 2. An example of a layered search tree. (a) Handwritten input strokes. (b) Part of the layered search tree. Nodes are labeled with their interpretation. The numbers in the circles are the expansion order by the best-first search algorithm (c) A path from the root node to a complete interpretation node. The nodes hold its interpretation and the remaining input strokes. The edges hold the consumed strokes, the symbol identity, and the spatial relationship with its parent interpretation.

the dominant symbol and its sub-expressions are represented as links. The spatial relationships among the sub-expressions are not represented, because we believe that they have been effectively encoded by the spatial relationships with the dominant symbol. Each of the sub-expressions is also represented recursively as a SRT with its own dominant symbol.

The spatial relationships are categorized into one of six types: *inside*, *over*, *under*, *superscript*, *subscript* and *right*. We believe such categories are sufficient for handling most MEs. An example of SRT representation is shown in Figure 3.

4. Cost of Interpretation

The cost $C(S)$, which is needed for the best-first search algorithm, is defined as the penalty of the input strokes when they are interpreted as SRT S. It is the sum of three negative log likelihoods:

$$C(S) = \alpha C_Y(S) + \beta C_R(S) + \gamma C_G(S). \qquad (3)$$

The cost function $C(S)$ reflects negatively the score of the symbol recognition result $C_Y(S)$, the degree of fitness for the spatial relationship $C_R(S)$, and the contextual validity

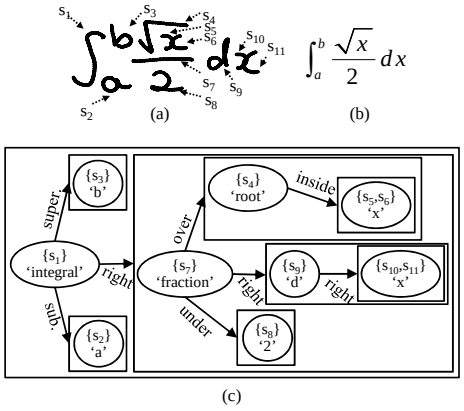


Figure 3. Symbol relation tree (SRT) representation. (a) Handwritten input. s_i denotes stroke. (b) The intended interpretation. (c) SRT representing the intended interpretation. Each rectangle denotes a sub-expression, and the dominant symbol of each sub-expression is denoted as a circle.

$C_G(S)$, where α, β and γ are coefficients for adjusting the relative contribution.

The score of the symbol recognition is defined as the sum of individual symbol likelihoods, which are obtained by a symbol recognizer. In our implementation, the symbol recognizer is developed based on the template matching.

The degree of fitness for the spatial relationship as a structure is difficult to estimate directly. We obtain it by combining all the degree of fitness for the spatial relationship of each symbol as:

$$C_R(S) = \sum_{y_i \in S} C_R(y_i, S). \qquad (4)$$

The degree of fitness for the spatial relationship of a symbol $C_R(y_i, S)$ is determined by analyzing the spatial layout in each sub-expression that contains the symbol. It is formally written as:

$$C_R(y_i, S) = \sum_{S' \in sub(y_i, S)} C_r(d(S'), y_i, t(S')), \qquad (5)$$

where $sub(y_i, S)$ is a set of sub-expressions containing the symbol y_i, $d(S')$ is the dominant symbol and $t(S')$ is the type of the expression S'.

Following the convention of [8], we defined fuzzy membership functions for computing the degree of fitness between a symbol and its dominant symbol under a specific spatial relationship type. The fuzzy membership functions are defined separately according to symbol classes of the dominant and subordinated symbols, and the spatial relationship type between the two symbols, as

follows:

$$C_r(y_i, y_j, t) = F_{class(y_i),class(y_j),t}(y_i, y_j). \quad (6)$$

We can estimate the cost increment contributed by each symbol using the individual symbol likelihood $C_Y(y_i)$ and the degree of fitness for the spatial relationship for the symbol $C_R(y_i, S)$. This will be actively utilized for the heuristic cost estimation in the following section.

Lastly, the contextual validity is to estimate how well the mathematical grammar or the mathematical conventions are satisfied. In the current implementation, we implemented several simple yes-no checking schemes such as the matching of parentheses and the existence of operands in '$-$' (fraction), '$\sum$', '$\sqrt{\ }$' and '$\int$' (integral). Language models and other complicated contextual knowledge could be used, but we retain it as work for the future.

5. Strategies for Efficient Search

In applying the best-first search algorithm, we proposed three strategies for a more efficient search: (1) providing fixed orders of symbol introduction to avoid generating the same interpretation through different paths in the layered search tree, (2) using an admissible heuristic to speed up the search, and (3) delaying symbol identity decisions during the structure search if a symbol's identity does not cause structural alternatives.

5.1. Order of Symbol Introduction

To avoid generating the same interpretation through different paths in the layered search tree, we fix the order of symbol addition to the existing interpretation as follows. Firstly, a dominant symbol should be introduced before any of the symbols in its subordinates. So, the dominant symbol is always located higher than dominated symbol in the upside-down search tree.

Secondly, when a dominant symbol has multiple subordinates, their order is determined by their spatial relationship types. Precedence is set for spatial relationship types as *inside* $>$ *superscript* $>$ *subscript* $>$ *over* $>$ *under* $>$ *right*. With such a fixed order, the same structure is not appears no more than once in the search tree generation. Note that this ordering does not restrict a writer's own pace in handwriting, but it is applied internally only in the search.

5.2. Admissible Heuristic

One of the advantages of the best-first search algorithm is that the problem solving heuristics can be easily incorporated. We proposed an admissible heuristic to estimate the cost increment from current interpretation to a complete interpretation.

The true cost increment from a partial interpretation to the best complete interpretation is difficult to estimate directly. However, we can underestimate the cost increment by estimating it separately for each stroke s_j of the remaining input strokes $X \setminus S_i$ as:

$$h(N_i) = \sum_{s_j \in X \setminus S_i} h(s_j, S_i). \quad (7)$$

The cost increment for each remaining stroke $h(s_j, S_i)$ is computed as follows. Let $Y(s_j, X \setminus S_i)$ denote the symbol hypothesis set which contains s_j among those constructed by the remaining input strokes. Then $h(s_j, S_i)$ is the minimum cost increment among every possible symbol hypothesis in $Y(s_j, X \setminus S_i)$, normalized by the number of strokes for the symbol hypothesis $||y_k||$, as:

$$h(s_j, S_i) = \min_{y_k \in Y(s_j, X \setminus S_i)} \frac{1}{||y_k||} h(y_k, S_i). \quad (8)$$

Let $Exp(S_i, y_k)$ be the set of every expansion using the symbol hypothesis y_k, then the cost increment $h(y_k, S_i)$ contributed by the symbol hypothesis y_k is the minimum among $Exp(S_i, y_k)$ as:

$$h(y_k, S_i) = \min_{S' \in Exp(S_i, y_k)} \alpha C_Y(y_k) + \beta C_R(y_k, S'). \quad (9)$$

The above development offers two nice features. Firstly, we can rank symbol hypotheses in terms of the contribution to the search of the complete interpretation, as we can estimate the cost contributed by each symbol hypothesis as shown in section 4. Secondly, as the heuristic regards only the minimum cost increment for each remaining input stroke, it will always underestimate the true cost increment. Therefore, the heuristic is *admissible* to guarantee to find the least cost complete interpretation first [10].

Since the heuristic utilizes information of not only the symbol recognition score, but also the degree of fitness for the spatial relationship, we call it a *structural heuristic*. Note that the structural heuristic is more informative than a *symbol heuristic* which uses only the symbol recognition score, and, therefore, yields more efficient search.

5.3. Delayed Decision of Symbol Identity

Further reduction of search space can be achieved by delaying the decision of symbol identity. Unlike printed MEs, it is hard to disambiguate the confusing pairs of handwritten symbols in isolation, such as ('C', '(', ('x', '$\times$'), ('S', '$\int$'), ('7', '$>$'), etc. For example, the handwritten stroke s_1 in Figure 3(a) may be recognized as not only an integral '$\int$', but also a capital letter 'S' or a small letter 's'. With such different symbol identities, the search

generates multiple complete interpretations whose structure are the same except for the symbol identity of s_1, as shown below.

$$\left\{ \int_a^b \frac{\sqrt{x}}{2} dx, S_a^b \frac{\sqrt{x}}{2} dx, s_a^b \frac{\sqrt{x}}{2} dx, \cdots \right\} \quad (10)$$

We take advantage of such situations by searching for structures without the symbol identities. The *unlabeled interpretation* is sought by generating structural alternatives. The symbol identity is determined after a complete unlabeled interpretation is found. An admissible heuristic can be developed for the unlabeled interpretation by a similar scheme described in section 5.2. The symbol identities in the complete unlabeled interpretation are found with the *Viterbi decoding algorithm* [12].

6. Experiments

To evaluate the proposed approach, we built a DB consisting of high school level handwritten MEs from 13 writers. Each writer was requested to write 30 different expressions, each of which contained 5-24 symbols. As a result, 390 MEs were collected and used in the experiments.

In collecting handwriting samples, we stressed natural writing. Expressions can be written in the writer's own pace in free writing order. However, two simple restrictions were imposed. A symbol should be written with consecutive strokes in time, and any single symbol has at most of 4 strokes, similarly with the assumption in [6].

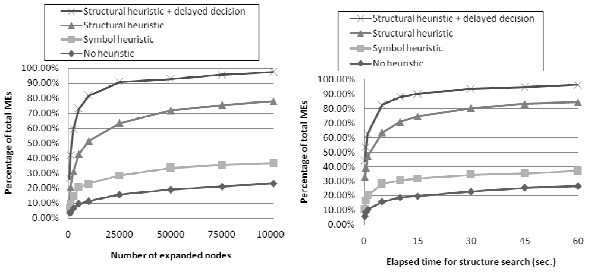
The symbol recognizing module in our system is developed by simple template matching. To handle 98 symbol classes which appear in high school mathematics, we developed 262 templates, considering variability in symbol shapes. Table 1 shows the symbols used in our system.

Table 1. Symbols used in our system

Digit	0, 1, 2, $\cdots$, 9
Alphabets	A,$\cdots$,Z,a,$\cdots$,z (except 'o' and 'O')
Greeks	α, β, θ, π, Δ
Special symbols and operators	$+$, $-$, $\pm$, $\times$, $\div$, $=$, $\neq$, $<$, $>$, $\leq$, $\geq$:, !, $\rightarrow$, ∞ ,$-$(fraction), $\int$, $\sum$
Parentheses	(,), {, }, [,]
math. functions	sin,cos,tan,csc,sec,cot,lim,log,ln

To evaluate the search efficiency of the proposed approach, we compared four search schemes. The first one does not use any heuristics, so it becomes a uniform cost search. The second one uses a simplified heuristic in which the minimum cost increment is modified to reflect only symbol recognition score, not the degree of fitness for the spatial relationship. So we denote it as '**symbol heuristic**'. The third one, denoted as '**structural heuristic**', uses the admissible heuristic proposed in this paper. Finally, the last one, denoted as '**proposed heuristic + delayed decision**', uses the structural heuristic with the delayed decision strategy.

The number of expanded nodes and elapsed time were used as the measure of the comparison of the search efficiency. As shown in Figure 4(a), the proposed approach found a complete interpretation for 90.77% of cases with 25,000 node expansions, whereas the other approaches found it for less than 80% of cases with 100,000 node expansions. Also, as shown in Figure 4(b), the proposed approach found a complete interpretation for 90.26% of the cases within 15 seconds, whereas the other approaches found it for less than 85% of cases even spending more than 60 seconds. These results show that the proposed approach could complete tasks spending a relatively small amount of time and space.



(a) (b)

Figure 4. Search efficiency Comparison: (a) by expanded nodes, and (b) by elapsed time.

Recognition accuracy is the next item of evaluation. The system obtained a recognition result of 385 of 390 cases within one minute. Table 2 shows the symbol level accuracy of the system. The symbol recognition accuracy 89.0% is below our expectation. It is because too many variants of some symbols were tested. However, the system correctly identifies the structure in 93.6%. Such robustness is due to the fact that many local hypotheses are analyzed during the search.

Table 2. Symbol level accuarcy

MEs recognized within 1 minute	385 (98.7%)
Total number of recognized symbols	4,942
Correctly segmented symbols	4,797 (97.1%)
Correctly placed symbols in the constructed structures	4,624 (93.6%)
Correctly labeled symbols	4,396 (89.0%)

The system generates near misses, as well. Figure 5 shows the ME level spatial relationship accuracy. More than 70% of cases were correctly constructed, and more

than 92% of cases had 3 or less errors in constructing whole structures. This shows that our method is robust in constructing structures with spatial relationships though errors of symbol segmentation and symbol recognition still occur in our system. These errors are less critical than those related to spatial relationships because they can be corrected independently at the symbol level.

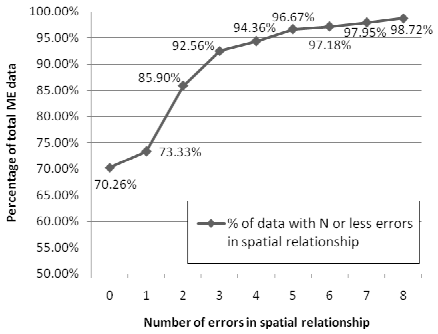
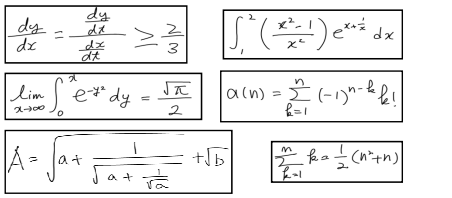
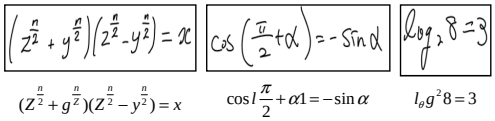


Figure 5. ME level spatial relationship accuracy.

Figure 6 shows examples of the correctly recognized and the misrecognized. As shown in Figure 6(b), the local ambiguity seems hard to be resolved by the search technique only. We learn that the global consistency checking is mandatory to improve the recognition accuracy.



(a)



(b)

Figure 6. Examples used in the experiments. (a) Correct recognition. (b) Misrecognition.

7. Conclusions and Future Work

A layered search framework for handwritten ME recognition is proposed. While all the possibilities are examined through the search, the search complexity is made manageable by applying the admissible heuristic and delaying decisions of symbols' identity. The evaluation experiments showed that our method achieved high speed with a high level of accuracy. We also found that even though the symbol recognition accuracy is still below our expectation, our system robustly construct a correct structure mainly due to the system's capacity to examine a large number of possibilities.

We also found that the overall performance would be improved with global context processing such as grammar checking, incorporation with language models and mathematical conventions. Future work will include incorporation of global text processing with improved symbol recognition and structural analysis for more robust cost estimation.

Acknowledgment

This work is partially supported by the Microsoft Research Asia's grant under the regional theme of Mobile Computing in Education 2007.

References

[1] K. Chan and D. Yeung, "mathematical expression Recognition: A Survey," *Int. Journal on Document Analysis and Recognition*, vol.3, no.1, 2000, pp 3-15.

[2] U. Garain and B. Chaudhuri, "Recognition of Online Handwritten mathematical expressions," *IEEE Trans. on Systems, Man and Cybernetics* (B), vol.34, no.6, 2004, pp 2366-2376.

[3] R. Zanibbi, D. Blostein and J. Cordy, "Recognizing mathematical expressions Using Tree Transformation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.11, 2002, pp 1455-1467.

[4] E. Tapia and R. Rojas, "Recognition of On-line Handwritten mathematical expressions Using a Minimum Spanning Tree Construction and Symbol Dominance," *Graphics Recognition 2003*, LNCS 3088, 2004, pp 329-340.

[5] K. Toyozumi, N. Yamada, T. Kitasaka, K. Mori, Y. Suenaga, K. Mase and T. Takahashi, "A Study of Symbol Segmentation Method for Handwritten Mathematical Formula Recognition Using Mathematical Structure Information," *Int. Conf. on Pattern Recognition*, vol.2, 2004, pp 630-633.

[6] Y. Shi, H. Li and F. Soong, "A Unified Framework for Symbol Segmentation and Recognition of Handwritten mathematical expressions," *Int. Conf. on Document Analysis and Recognition*, vol.2, 2007, pp 854-858.

[7] R. Yamamoto, S. Sako, T. Nishimoto and S. Sagayama, "On-Line Recognition of Handwritten mathematical expressions Based on Stroke-Based Stochastic Context-Free Grammar," *Int. Workshop on Frontiers in Handwriting Recognition*, 2006, pp 249-254.

[8] J. Fitzgerald, F. Geiselbrechtinger and T. Kechadi, "Mathpad: A Fuzzy Logic-Based Recognition System for Handwritten Mathematics," *Int. Conf. on Document Analysis and Recognition*, vol.2, 2007, pp 694-698.

[9] E. Miller and P. Viola, "Ambiguity and Constraint in mathematical expression Recognition," *National Conf. on Artificial Intelligence* (AAAI), 1998, pp 784-791.

[10] S. Russell and P. Norvig, "*Artificial Intelligence: A Modern Approach*," Prentice Hall, 1995.

[11] H. Lee and J. Wang, "Design of a mathematical expression Understanding System,", *Pattern Recognition Letters*, vol.18, no.3, 1997, pp 289-298.

[12] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. on Information Theory*, vol.13, no.2, 1967, pp 260-269.