THE USE OF GLOBAL CONTEXT IN TEXT RECOGNITION

Jonathan J. Hull

State University of New York at Buffalo
Department of Computer Science
Buffalo, New York 14260

ABSTRACT

An approach for using context between words (called global context) in an algorithm for reading text is discussed. This approach uses a small number of features extracted from a word image to determine a subset of words in a dictionary called a *neighborhood*. Global context is then used to reduce the size of the neighborhood and the smaller neighborhood is used to direct further detailed analysis of the input. The result of this process is a match of the input image to one of the words in the neighborhood. Results are reported in this paper on the performance and cost of two forms of global context. One uses neighborhood-to-word transitions and the other uses word-to-word transitions. The potential of these forms of global context to reduce the number of words in a neighborhood is explored in a set of statistical experiments. Image processing experiments demonstrate the feasibility of the neighborhood calculation. The use of general purpose global context such as syntactic and semantic categories in a similar framework is also discussed.

1. INTRODUCTION

The development of computer reading algorithms that possess the same versatility as a human reader is a long standing goal of pattern recognition research. One approach that has been followed in the pursuit of this goal is to determine how humans read and to convert useful portions of this knowledge into algorithms. This method was followed by Shillman[7] in his work on isolated character recognition. He determined the parameters of the human character recognition process and developed algorithms that used these parameters.

The complete, fluent human reading process involves much more than just isolated character recognition. A human reader routinely employs knowledge about the syntactic, semantic, and featural dependencies between words [5]. Although the use of this type of knowledge source in an algorithm for reading text has been suggested [6], no such technique is known.

The use of knowledge about dependencies between words in an algorithm for reading text is explored in this paper. This knowledge is represented in two ways: either as a table of allowable transitions between words, or as a table of allowable transitions between a group of words (called a *neighborhood*) and another word. These representations describe different forms of featural dependency. Word-to-word transitions represent knowledge of the words that can follow a given word under the assumption that it can be accurately recognized. Neighborhood-to-word transitions represent knowledge about legal successor words under the assumption that the predecessor word can only be approximately recognized as one of the words in a neighborhood.

These representations are most suitable for use in a situation where the words that could follow a given word are predictable with a high degree of confidence. An example of this is the last line of a postal address where the first few words make up a city name that should come from a fixed set of possibilities. A generalization of this technique to free format running text would require a less specific form of global context such as syntactic categories (e.g., noun, verb, adjective, etc.). This information could be used in a manner similar to that discussed here.

2. READING ALGORITHM

An outline of the reading algorithm used to evaluate the potential usefulness of global context is presented in Figure 1. This algorithm assumes that the input is a sequence of word images w_i, i = 1,2,...,n. These images are given to a **hypothesis generation** routine that uses a few gross features of the image to determine a neighborhood of words N_i in a dictionary that share those features. This neighborhood is then input to the **global contextual analysis** routine that produces a reduced neighborhood N_i^*. The **hypothesis testing** routine uses the contents of N_i^* to direct further selective analysis of the input image, resulting in a recognition decision. The hypothesis testing process is discussed in [4] and will not be analyzed in this paper. Rather, the focus will be on the usefulness of global context in reducing the number of words in a neighborhood, thereby making the job of hypothesis testing that much easier.

The hypothesis generation procedure is similar to that discussed in [1]. The method used here (more fully described and analyzed in [2]) is designed as a generator of candidate matches rather than as a complete method for word recognition. This technique is based on the detection of a small

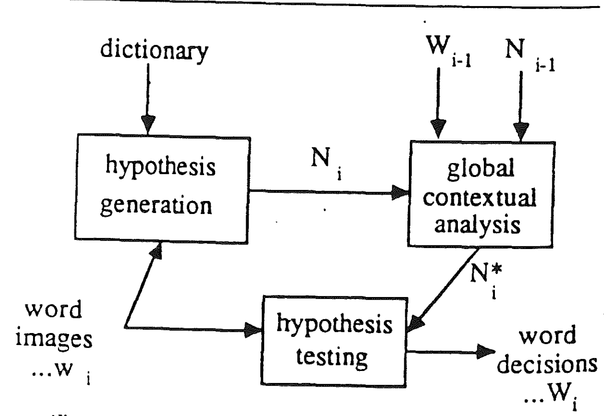

Figure 1. An outline of the reading algorithm.

Eighth IAPR Int. Conf on Pattern Recognition, Paris, France, Oct. 28-31, 1986, 1218-1220.

number of features in the image of a word. Six features (numbered 0 through 6) were used in the experiments discussed in this paper, which assumed that all text was printed entirely in lower case. Features 1 to 5 were different vertical bars (short, ascender, descender, short with dot on top, and long with dot on top). These features can be detected by convolving an image of a word with a vertical bar mask and thresholding the output. Feature 0 was a significant horizontal space that did not contain a vertical bar. (Note that these features are specialized for lower case text; however, the technique can be extended to mixed case and upper case input by definition of a suitable feature set). The neighborhood of an input word image is determined by finding the dictionary words that contain the same left-to-right sequence of features as the input word. The neighborhood is denoted by the feature-numbers associated with this sequence of features.

Two procedures are now discussed that use precompiled tables of allowable transitions to reduce the size of the neighborhood of an input word. In the first method, transitions between two words are used; in the second method, transitions between neighborhoods and words are used.

The representations for global context are used under the assumption that either the preceding word in the text was perfectly recognized or just its neighborhood was correctly computed. Elements are removed from the neighborhood of the current word that are not contained in the appropriate table of allowable transitions. Figure 2 shows a portion of the data structure needed for both these techniques. If word-to-word context is used, a word dictionary is augmented by pointers from a given word to all the words that can follow it. If neighborhood-to-word context is used, a neighborhood dictionary is provided that includes neighborhood-to-word pointers to entries in the word dictionary. In reference to Figure 2, if word-to-word context is used when recognizing the word "shape" in the phrase "... visual shape ..." and the only word that can follow "visual" in the word-to-word transition table is "shape", then the neighborhood of "shape", which is {"shape", "slope"}, can be reduced to {"shape"}, i.e., a perfect recognition. If neighborhood-to-word context is used and the sequence of features in "visual" is determined to be "041112", the neighborhood of "shape" cannot be reduced.

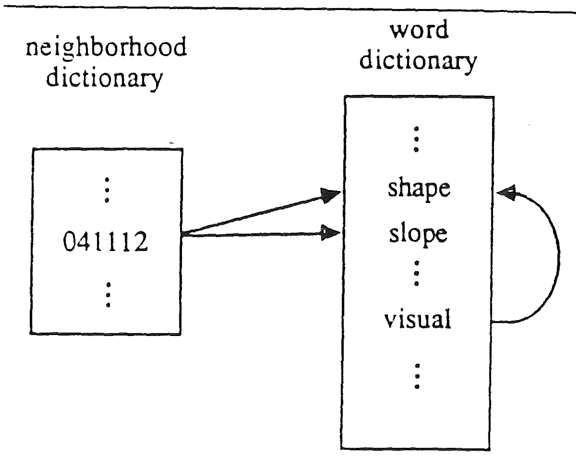

Figure 2. Example of data structures for two forms of global context.

3. STATISTICAL EFFECTS OF GLOBAL CONTEXT

The statistical effect of both forms of global context can be measured by their effect on two statistics: the percentage of text that is uniquely specified by shape (PER_UNIQ), i.e., the percentage of text that has a neighborhood containing a single word, and the average neighborhood size per text word (ANS_t). Given that $ns(tw_i)$ returns that number of words in the neighborhood of the i^{th} word of a given text,

$$ANS_t = \frac{1}{N_t}\sum_{i=1}^{N_t} ns(tw_i)$$

where N_t is the number of words in the given text. The average neighborhood size per text word thus measures the average number of words that would be output by the hypothesis generation procedure when it read the given text.

The potential influence of global context on the reading algorithm is measured by its effect on PER_UNIQ and ANS_t. Since the global contextual analysis produces a reduced neighborhood size, it should increase PER_UNIQ and reduce ANS_t from what they are when no global context is used. In the best case, PER_UNIQ would be 100% and ANS_t would be 1.0. This would correspond to a case where complete recognition would be provided by the combination of hypothesis generation and global contextual analysis.

A study was carried out in which these statistics were calculated from a database of all the city names that are associated with a postal code in the United States. (This database was extracted from the Directory of Post Offices (Dopo) file maintained by the U.S. Postal Service.) Altogether there were 48,349 entries in this database, each of which contained from 1 to 9 words. The dictionary of unique words contained 18,745 words. The tables of word-to-word and neighborhood-to-word transition probabilities were computed from this database. It is reasonable to apply the techniques for global context discussed here in this situation, since the words in the city name of an address are highly constrained and are determined by the entries in the Dopo file. It is assumed for the purposes of this study that the words on either side of a city name have already been recognized and provide known boundaries. It is also assumed that the city name is printed completely in lower case.

The values of PER_UNIQ and ANS_t were computed in three ways: 1. using no global context; 2. using neighborhood-to-word context; 3. using word-to-word context. The results of this study are shown in Table 1. It is seen that ANS_t is reduced from 18.1 to 9.5 when both forms of global context are used. PER_UNIQ increases from 19.0% to 43.7% when neighborhood-to-word transitions are used, and from 19.0% to 44.6% when word-to-word transitions are used. This says that almost half the words in the database could be uniquely recognized by only computing their neighborhoods. No additional feature analysis would be needed for these words.

	no context	neigh-word	word-word
ANS_t	18.1	9.5	9.5
PER_UNIQ	19.0	43.7	44.6

Table 1. A comparison of the effect of three types of global context on ANS_t and PER_UNIQ.

Eighth IAPR Int. Conf on Pattern Recognition, Paris, France, Oct. 28-31, 1986, 1218-1220.

This study assumed a population of addresses in which each city name occurs once and only once. This is unreasonable for an address reading machine. However, the trend of significantly improved performance shown in this study should also occur if the entries in the database are weighted by their frequency of occurrence. Some evidence for this are the results discussed in [3] where PER_UNIQ and ANS_t were calculated with the forms of global context used here. A running text of over 1,000,000 words was used that presumably exhibits some of the frequency characteristics not accounted for in the study discussed above. PER_UNIQ increased from 9% to 36% and ANS_t decreased from 38.4 to 4.2, on the average, when global context was used.

4. EXPERIMENTAL RESULTS

Two experiments were conducted to demonstrate the feasibility of this methodology. The first experiment showed that the neighborhood of the image of a word can be accurately computed under a variety of conditions that include several different fonts and words with characters touching or overlapping. The second experiment explored the storage required by the transition tables for both representations of global context. These experiments are summarized here. More complete details are discussed in [3].

In the first experiment images of 100 words were generated in five different fonts under three conditions of segmentation: 1. All characters topologically distinct; 2. Every character touching the character to its left and the character to its right; 3. Every character overlapping the characters to its left and right by two pixels. The neighborhoods were correctly calculated, on the average, in over 99.5% of the cases under condition 1, over 96% of the cases under condition 2, and over 91% of the cases under condition 3.

In the second experiment the storage required for both forms of global context was analyzed. It was shown that the storage needed for the transitions was only about twice that needed for a full-text representation of the dictionary. It was also shown that the storage requirement increased linearly with the number of words in the dictionary.

5. EXTENSION TO SYNTAX AND SEMANTICS

The extension of the technique presented here to syntactic and semantic categories is straightforward. This requires the use of a dictionary that associates one or more syntactic or semantic categories with each word. The allowable transitions between the categories would have to be determined, for example, by experimentation with a large corpus. (The transitions could also be dynamically determined by an inference process such as parsing.) This augmented dictionary would be used in the same way the present structure is. For example, given that a word has been recognized as an adjective and it is highly likely that a noun will follow it, all the words that are not nouns could be eliminated from the neighborhood of the next word. Such a methodology would be more flexible than the one discussed in this paper and would be suitable for recognizing free format running text.

6. CONCLUSIONS

Two methods for incorporating knowledge about a source text beyond the word level (referred to as global context) in an algorithm for reading text were explored. The reading algorithm includes a step in which a small number of features are extracted from the image of a word and a group of visually similar words called a *neighborhood* is retrieved from a dictionary. The neighborhood is then used to control further processing that determines which word matches the input image. Global context is used to reduce the number of entries in the neighborhood, thus improving the performance of the matching process.

The effect of two representations for global context on reducing the number of words in the neighborhood of an input word was explored in a set of statistical experiments on a database of city names. A significant reduction in neighborhood size was achieved with both representations. Image processing experiments demonstrated the feasibility of the basic operations needed for this method and an empirical study of storage costs showed that both representations for global context could be realized at a relatively modest cost. Future work in this area will include the development of a more general purpose version of using global context that incorporates syntactic and semantic information.

Acknowledgements

The author gratefully acknowledges the helpful comments of Jeffery Zucker, Sargur Srihari, Stuart Shapiro, William Rapaport and Paul Palumbo. The acquisition of the Dopo database was made possible by Martin Sack of the United States Postal Service. The USPS also provided financial support for this work under the BOA program.

References

1. L. D. Earnest, "Machine recognition of cursive writing", *Information Processing 1962, Proceedings of IFIP Congress 62*, Munich, Germany, August 27 - September 1, 1962, 462-466.

2. J. J. Hull, "Word shape analysis in a knowledge-based system for reading text", *The Second IEEE Conference on Artificial Intelligence Applications*, Miami Beach, Florida, December 11-13, 1985, 114-119.

3. J. J. Hull, "Inter-word constraints in visual word recognition", *Proceedings of the Conference of the Canadian Society for Computational Studies of Intelligence*, Montreal, Canada, May 21-23, 1986, 134-138.

4. J. J. Hull and S. N. Srihari, "A computational approach to visual word recognition: hypothesis generation and testing", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, June 22-26, 1986, 156-161.

5. K. Rayner, M. Carlson and L. Frazier, "The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences", *Journal of Verbal Learning and Verbal Behavior 22* (1983), 358-374.

6. J. Schurmann, "Reading machines", *Proceedings of the 6th International Conference on Pattern Recognition*, Munich, West Germany, October 19-22, 1982, 1031-1044.

7. R. J. Shillman, *Character recognition based on phenomenological attributes: theory and methods*, Massachusetts Institute of Technology, August, 1974. Ph.D. Dissertation.

Eighth IAPR Int. Conf on Pattern Recognition, Paris, France, Oct. 28-31, 1986, 1218-1220.