samples used	overall						G02-nouns						nouns-matching					
	matches		errors		corr. pct.		matches		errors		corr. pct.		matches		errors		corr. pct.	
1	251	508	9	27	96	95	130	199	2	10	98	95	187	224	6	17	97	92
2	345	585	11	30	97	95	177	255	2	12	99	95	206	283	6	19	97	93
3	393	617	12	34	97	94	199	274	2	14	99	95	241	325	6	21	98	94
4	441	642	12	35	97	95	229	292	2	14	99	95	257	347	8	22	97	93
5	451	667	12	42	98	94	234	306	2	17	99	94	258	363	9	27	97	92
6	459	674	13	45	97	93	248	322	2	18	99	94	272	384	9	31	96	92
7	474	682	16	49	97	93	254	324	3	18	99	94	280	387	11	32	96	92
8	483	685	16	50	97	93	254	324	3	18	99	94	284	387	11	32	96	92
9	498	692	16	51	97	93	261	326	3	18	99	94	288	392	11	32	96	92
10	526	709	22	58	96	92	300	330	4	20	99	94	296	396	12	34	96	91

Table 4. Word selection performance on the original 885 neighborhoods (with 87 percent correct at the top choice) and for the first two levels of the document database graph. The original performance is shown on the left and the graph performance on the right of each column.

matched in G02 at a marginal increase in the error rate. Thus, the document similarity graph can improve performance over that offered by the plain flat database structure.

4. Discussion and Conclusions

This paper presented an adaptation of the vector space model for information retrieval to improving the performance of a word recognition algorithm. The neighborhoods of visually similar words determined by word recognition are matched to a database of documents structured as a graph and a subset of documents with topics that are similar to those of the input image are determined. The vocabulary from those similar documents are used to select the word recognition decisions that have a high probability of being correct.

References

1. H. S. Baird and K. Thompson, "Reading chess," *IEEE Transactions on Pattern Analysis and Machine Intelligence 12* (1990), 552-559.

2. K. W. Church, "A stochastic part of speech program and noun phrase parser for unrestricted text," *Second Conference on Applied Natural Language Processing*, Austin, Texas, 1988.

3. G. DeSilva and J. J. Hull, "Proper noun location in document images," *Pattern Recognition*, 1993. accepted to appear.

4. C. Fox, "A stop list for general text," *SIGIR Forum 24* (Fall/Winter, 1989), 19-35.

5. T. K. Ho, J. J. Hull and S. N. Srihari, "A computational model for recognition of multifont word images," *Machine Vision and Applications, special issue on Document Image Analysis*, Summer, 1992, 157-168.

6. J. J. Hull, "Hypothesis generation in a computational model for visual word recognition," *IEEE Expert 1*, 3 (Fall, 1986), 63-70.

7. J. J. Hull, "Incorporation of a Markov model of language syntax in a text recognition algorithm," *Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, March, 1992, 174-185.

8. J. J. Hull, S. Khoubyari and T. K. Ho, "Visual Global Context: Word Image Matching in a Methodology for Degraded Text Recognition," *Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, March, 1992, 26-39.

9. J. J. Hull and Y. Li, "Word recognition result interpretation using the vector space model for information retrieval," *Second Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, April 26-28, 1993, 147-155.

10. H. Kucera and W. N. Francis, *Computational analysis of present-day American English*, Brown University Press, Providence, Rhode Island, 1967.

11. G. Salton, *Automatic text processing*, Addison Wesley, 1988.