

*Editorial***Special Issue on
“Performance Evaluation: Theory, Practice, and Impact”****Guest Editors: Tapas Kanungo, Henry S. Baird, Robert M. Haralick**

The document image analysis research community has been distinguished for over a decade by a serious and sustained commitment to sound methodologies for measuring the performance of algorithms and systems. Objective, quantitative, and standardized performance evaluation methods are essential aids in our attempts to understand the behavior of our systems, predict their future performance, compare rival systems, identify the particular strengths and weaknesses of proposed technologies, and track the progress of our community's research achievements from year to year. We feel that the time is ripe to offer, in a journal special issue, a selection of the strongest papers having, as their principal theme, performance evaluation theory, practice, or impact in a large-scale application.

The manuscripts submitted were reviewed by highly qualified expert referees in a thorough two-stage review procedure. The articles that we have been happy to accept all enjoy a combination of originality, high technical merit, and clear relevance to the topic.

Evaluating geometrical page-layout segmentation algorithms is a challenging task, in part due to the diversity of metrics that have been plausibly proposed for measuring the similarity of two segmentations. J. Hu, R. Kashi, D. Lopresti, and G. Wilfong discuss a methodology for evaluating systems that extract tables from document images in their article “Evaluating the performance of table processing algorithms.” One of their innovations, applicable to a wide variety of layout segmentation tasks, is to probe two “table graphs” at random, counting similarities and dissimilarities, to accumulate a statistical measure of match.

In “Large scale address recognition systems – truing, testing, tools and other evaluation issues,” S. Setlur, A. Lawson, V. Govindaraju, and S. Srihari describe the methodology they used for evaluating, on a stunningly large scale, a USPS postal address recognition system. Their methodology samples a live stream of postal images to create ground-truthed images and evaluates the system using encoding rate and error rate metrics.

When, as not infrequently happens, the number of original ground-truth documents available for experimentation is severely limited, performance evaluation results, however carefully calculated, can be inaccurate and misleading.

In their article “A statistical approach to the generation of a database for evaluating OCR software,” F. S. Brundick, A. E. M. Brodeen, and M. S. Taylor propose a bootstrapping approach to the generation of sufficiently large databases of synthetic ground-truthed documents. These documents can then be printed and scanned to acquire test images each, by construction, corresponding to known ground truth.

Contributing an interesting variation to the large literature on applications of string-matching algorithms to DIA problems, C. Fang, C. Liu, L. Pent, and X. Ding present a specialized algorithm, in “Automatic performance evaluation of printed Chinese recognition systems,” that assists researchers in the evaluation and characterization of character-segmentation errors.

In “An empirical measure of performance of document image segmentation algorithm,” A. K. Das, S. K. Saha, and B. Chanda argue for a new graph-based evaluation metric for page-layout segmentation algorithms, and provide an algorithm to compute the metric.

Statistical classifiers form an integral part of many document image analysis systems. In their article “Performance evaluation of pattern classifiers for handwritten character recognition,” Liu, Sako and Fujisawa compare the performance of well-known statistical classifiers as a function of training sample size, outlier resistance, and ambiguity rejection.

Agreement within the research community on standardized metrics, datasets, and software tools is of course an essential foundation for the most effective use of evaluation methodology. S. Mao and T. Kanungo, in “Software architecture of PSET: A page segmentation evaluation toolkit,” describe in detail the rationale and architecture of public-domain software tools offered for use in the evaluation of a broad class of page-layout segmentation algorithms.

We would like to thank all the authors who submitted manuscripts to this special issue. We regret that only a fraction of the submissions, all of them interesting, could be included. We owe a special debt of gratitude to the many able reviewers who generously commented, often in extraordinary detail, on the submissions. Finally, we would like to acknowledge the good-hearted patience of the authors and of the journal's editors in spite of this special issue's long gestation.