

Image processing of degraded palm-leaf manuscripts

Lalit Saxena and Ambuja Salgaonkar

University of Mumbai

20 May 2013

Generating a report of the physical condition of degraded manuscripts is an essential but tedious task. Manual processing of manuscripts is prone to error and is harmful to both the health of the conservator and the condition of the folios. The present work describes an implementation of an automatic status report generation (ASRG) system for degraded palm-leaf manuscripts. ASRG employs image processing techniques to generate reports that are more accurate and take less time compared with manual processing systems. Removal of stains (or similar noise) while enhancing the readability of the folio images is an extended aspect of ASRG. A technique to identify a suitable threshold for a binarized form of the manuscript has been developed. The image processing techniques of dilation and erosion are employed to recover minor degradations (due to the insect bites etc.). These techniques generate a cleaner version of the folio image so that the text becomes more readable. The enhanced image is available in digitized form for further processing.

The complex structures of Indian scripts and varied handwriting styles are inherent sources of difficulty in machine-assisted manuscript text processing. Script readers are scarcely available while the number of researchers interested in knowing the texts is relatively large. Manuscript images produced by scanning and recognition software are prone to segmentation errors, and the error rate increases rapidly with degradations. Separating different patterns (characters / words) in the text, mapping these patterns to user-friendly symbols (say a more readable form of the script) and availing oneself of the user-friendly form in future retrieval and processing is one of the objectives of this work.

While this work has been carried out for the Grantha-script manuscripts available in the Library of University of Mumbai, the techniques developed here can be applied to any other variety of manuscripts.