Abstract

Documents are archived and preserved in large quantities worldwide. Electronic scanning is a common approach in handling such materials to facilitate public access to them. But the resulting images are often difficult to read, also have low contrast, and are corrupted by different artifacts. Enhancement of image is normally based on minor deterioration in modern documents to improve optical character recognition. It often ignores cases, such as those typical of historical and other highly degraded documents. Separating background and foreground can make an image readable. The initial approach is using local threshold where an average threshold value is used for sub images. The next approach is using Global thresholding where only single threshold value is used for an entire image. The third approach is combining both the algorithms Local and Global which is Hybrid Binarization where global thresholding is applied and Local thresholding is applied to parts where thresholding is to be done. The last approach being Iterative Global Thresholding which calculates the threshold iteratively and performs thresholding. Filtering is done to remove noises using methods like Smoothing and Sharpening and Peak to signal noise ratio is found before and after to check the degree of enhancement.
A Comparative Study of Binarization Techniques for Enhancement of Degraded Documents

References

- Ergina Kavallieratou and Efstathios Stammatatos, "Improving the Quality of Degraded Document Images," DIAL?06, August 2006
- Abdenour Sehad, Youcef Chibani and Mohamed Cheriet, "Gabor Filters for Degraded Document Image Binarization," 14th International Conference on Frontiers in Handwriting Recognition, 2014

Index Terms

Computer Science

Signal Processing

Keywords

Hybrid binarization  Global thresholding  local thresholding  Iterative Global Thresholding  Peak to Signal Noise ratio.