Abstract

Optical Character Recognition (OCR) is one of the important fields in image processing and pattern recognition domain. OCR with high accuracy finds application in offices, banks, healthcare etc. The accuracy of the OCR is primarily dependent on the quality of the input image. So, to achieve high accuracy OCR we should provide a high quality image, which is free from different types of noises, degradation, skews etc. In this paper, we have made an attempt to remove the noise, which is present in the input image. A novel method based on run length count is proposed to denoise the images. In this approach first the noisy image is binarized. Based on the horizontal and vertical run length count, the noise in the image will be identified and eliminated. The algorithm is tested with noisy epigraphical document images, noisy printed document images. The effectiveness of the algorithm is verified with images having synthetic noise derived from Gaussian, Speckle and Poisson noise models. The experimental results show that the proposed method is efficient for noise elimination.
An Approach based on Run Length Count for Denoising the Kannada Characters

- Mukesh Motwani, Mukesh Gadiya, Rakhi Motwani, and Frederick C. Harris, Jr. A Survey of Image Denoising Techniques, Proceedings of GSPx 2004
An Approach based on Run Length Count for Denoising the Kannada Characters

Index Terms

Computer Science
Pattern Recognition

Keywords

Optical Character Recognition
Image Denoising
Run Length Count
Epigraphical Document
Printed Document

Gaussian Noise
Speckle Noise
Poisson Noise