

Adaptation of Hough Transform Algorithm for Text Document Skew Angle Detection

Authors : Kayode A. Olaniyi, Olabanji F. Omotoye, Adeola A. Ogunleye

Abstract : The skew detection and correction form an important part of digital document analysis. This is because uncompensated skew can deteriorate document features and can complicate further document image processing steps. Efficient text document analysis and digitization can rarely be achieved when a document is skewed even at a small angle. Once the documents have been digitized through the scanning system and binarization also achieved, document skew correction is required before further image analysis. Research efforts have been put in this area with algorithms developed to eliminate document skew. Skew angle correction algorithms can be compared based on performance criteria. Most important performance criteria are accuracy of skew angle detection, range of skew angle for detection, speed of processing the image, computational complexity and consequently memory space used. The standard Hough Transform has successfully been implemented for text documentation skew angle estimation application. However, the standard Hough Transform algorithm level of accuracy depends largely on how much fine the step size for the angle used. This consequently consumes more time and memory space for increase accuracy and, especially where number of pixels is considerable large. Whenever the Hough transform is used, there is always a tradeoff between accuracy and speed. So a more efficient solution is needed that optimizes space as well as time. In this paper, an improved Hough transform (HT) technique that optimizes space as well as time to robustly detect document skew is presented. The modified algorithm of Hough Transform presents solution to the contradiction between the memory space, running time and accuracy. Our algorithm starts with the first step of angle estimation accurate up to zero decimal place using the standard Hough Transform algorithm achieving minimal running time and space but lacks relative accuracy. Then to increase accuracy, suppose estimated angle found using the basic Hough algorithm is x degree, we then run again basic algorithm from range between $\pm x$ degrees with accuracy of one decimal place. Same process is iterated till level of desired accuracy is achieved. The procedure of our skew estimation and correction algorithm of text images is implemented using MATLAB. The memory space estimation and process time are also tabulated with skew angle assumption of within 00 and 450. The simulation results which is demonstrated in Matlab show the high performance of our algorithms with less computational time and memory space used in detecting document skew for a variety of documents with different levels of complexity.

Keywords : hough-transform, skew-detection, skew-angle, skew-correction, text-document

Conference Title : ICCSMI 2019 : International Conference on Computer Science and Machine Intelligence

Conference Location : New York, United States

Conference Dates : August 08-09, 2019