

M. Sarfraz, **S. A. Mahmoud** and Z. Rasheed,  
“On Skew Estimation and Correction of Text “;“,  
2<sup>nd</sup> International Conference Geometric Modelling and Imaging GMAI'07, 4-6 July,  
2007, SWITZERLAND.

**Key words:** Image processing, skew estimation, tilt correction, character recognition,  
connected component.

## **On Skew Estimation and Correction of Text**

M. Sarfraz, S.A. Mahmoud and Z. Rasheed  
*Department of Information and Computer Science*  
*King Fahd University of Petroleum and Minerals*  
*KFUPM # 1510, Dhahran 31261, Saudi Arabia*  
*Email: {sarfraz; smasaad; hasham@kfupm.edu.sa}*

### **Abstract**

*This paper, in addition to reporting some existing techniques, proposes some new techniques for skew correction. It includes two novel document skew detection algorithms based on histogram statistics and connected component analysis. The histogram based algorithm works much efficient for detecting skew angle in which we analyze the lines as peaks and valleys on histogram. The connected component analysis is based on finding the connected components within a single line and considers them as one blob to estimate skew angle. The two methods have been experimented on various text documents and very promising results have been achieved given more than 99% accuracy. A comparative study has been reported to provide a detailed analysis of the proposed methods.*

**Key words:** Image processing, skew estimation, tilt correction, character recognition,  
connected component.

## On Skew Estimation and Correction of Text

M. Sarfraz, S.A. Mahmoud and Z. Rasheed  
*Department of Information and Computer Science*  
*King Fahd University of Petroleum and Minerals*  
*KFUPM # 1510, Dhahran 31261, Saudi Arabia*  
*Email: {sarfraz; smasaad; hasham@kfupm.edu.sa}*

### Abstract

*This paper, in addition to reporting some existing techniques, proposes some new techniques for skew correction. It includes two novel document skew detection algorithms based on histogram statistics and connected component analysis. The histogram based algorithm works much efficient for detecting skew angle in which we analyze the lines as peaks and valleys on histogram. The connected component analysis is based on finding the connected components within a single line and considers them as one blob to estimate skew angle. The two methods have been experimented on various text documents and very promising results have been achieved given more than 99% accuracy. A comparative study has been reported to provide a detailed analysis of the proposed methods.*

**Keywords:** Image processing, skew estimation, tilt correction, character recognition, connected component.

OCR systems typically assume that the documents are printed with a single direction of the text and that the acquisition process did not introduce a relevant skew. The advent of flat bed scanners and the need to process large amounts of documents at high rates, made the above assumption unreliable and the introduction of the skew estimation phase has become mandatory. In fact, a little skewing of the page is often introduced during processes such as copying or scanning. Today, documents are ever more free styled. Moreover, text aligned along different directions is not an uncommon feature. The subsequent stages of OCR systems, mainly depend upon the accuracy of preprocessing stage. For instance, if OCR system under estimates or over estimates skew angle, then OCR system which is utilizing projection based technique will fail miserably.

The paper has been organized in such a way that, in the next section, we provide brief literature survey of skew estimation algorithms in which we discuss some consequences related to them. Two newly proposed algorithms are explained in Section 3, some