
Straight line reconstruction for fully materialized table extraction in degraded document images

Héloïse Alhéricière^{*2,1}, Walid Amieur^{*1}, Florence Cloppet^{*1}, Camille Kurtz^{*1}, Jean-Marc Ogier^{*2}, and Nicole Vincent^{*1}

²Laboratoire Informatique, Image et Interaction – Université de La Rochelle : EA2118 – France

¹Laboratoire d'Informatique Paris Descartes – Université Paris Descartes - Paris 5 : EA2517 – France

Abstract

Tables are one of the best ways to synthesize information such as statistical results, key figures in documents. In this article we focus on the extraction of materialized tables in document images, in the particular case where acquisition noise can disrupt the recovering of the table structures. The sequential printings / scannings of a document and its deterioration can lead to "broken" lines among the materialized segments of the tables. We propose a method based on the search for straight line segments in documents, relying on a new image transform that locally defines primitives well suited for pattern recognition and on a proposed theoretical model of lines in order to confirm their presence among a set of confident potential line parts. The extracted straight line segments are then used to reconstruct the table structures. Our approach has been evaluated both from quality and stability points of view.

*Speaker