

Vector de características para el reconocimiento de texto manuscrito

Prof. Ing. Marisa R. De Giusti 1 | A.C. Vila, María María | A. C. Villarreal, Gonzalo Luján



Resumen

El reconocimiento de textos manuscritos con una computadora es una tarea compleja que implica la realización de un gran número de pasos que le permitirán obtener un conjunto de características que identifican unívocamente a cada objeto (palabra, carácter, símbolo). Los métodos de extracción de características existentes varían de acuerdo al área de aplicación y desarrollo, existiendo métodos con una clara orientación matemática, algebraica, algorítmica, etcétera. En el presente trabajo se realizó un estudio de algunos de los métodos más significativos, y se obtuvieron conclusiones relativas a la importancia de la información que aportan, al costo de ejecución, y de posibles mejoras utilizando varias técnicas juntas.

Palabras clave: digitalización, manuscritos, topología digital, morfología matemática, recuperación de información, componentes conexas, biblioteca digital, vector de características.

Resumen

A diferencia del ser humano, cuyo sistema de percepción es altamente sofisticado, las computadoras carecen de la capacidad de percibir el mundo y dotarlas de un sistema de visión artificial es una tarea compleja. En este trabajo se analizan varias técnicas de extracción de características de textos manuscritos, entre las que se incluyen Códigos de cadenas, ascendentes y descendentes, ventanas de muestreo y técnicas invariantes.

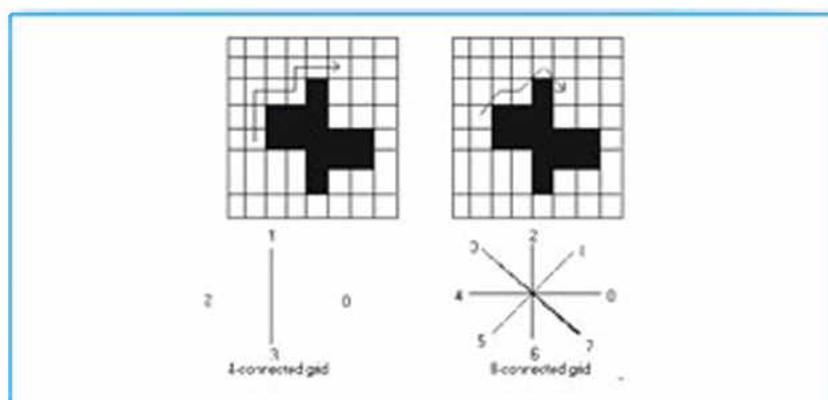


Figura 2: Posibles recorridos con el código de cadenas utilizando 4-vecindad y 8-vecindad.

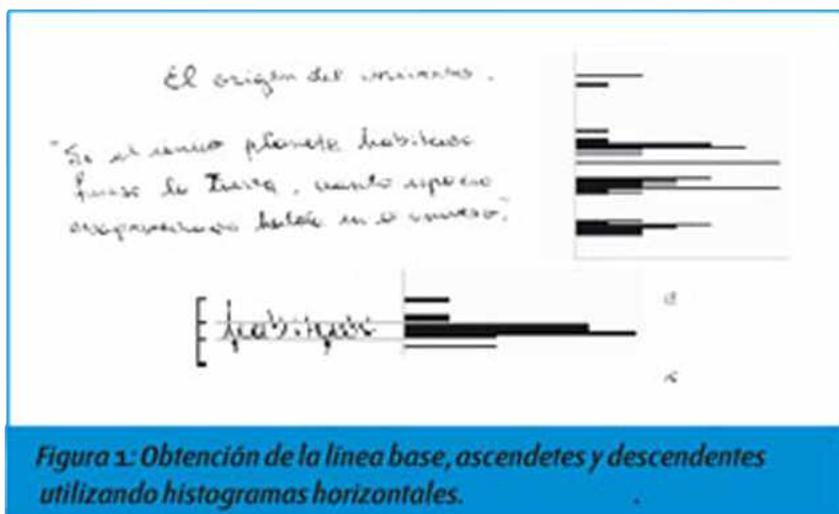


Figura 1: Obtención de la línea base, ascendentes y descendentes utilizando histogramas horizontales.

Resultados Alcanzados

Se han implementado varias técnicas, obteniendo resultados y conclusiones sobre las mismas, y se han logrado mejoras y optimizaciones de las mismas, modificando algunos algoritmos o combinando las técnicas junto con métodos para normalizar, métodos de limpieza de ruido o incluso otras técnicas de obtención de características.