

Technology Profile Fact Sheet

Title: Method of Distinguishing Handwritten and Machine-Printed Images

Aliases: Print Identification

Technical Challenge: The proliferation of economical computers, scanners, and fax machines has resulted in the use of document images as a common practice in nearly every office around the world today. Reliable algorithms to process all types and varieties of document images automatically are needed, but current automated techniques are limited in scope. This technology addresses these deficiencies by discriminating between handwritten and machine-printed document images for a diverse set of documents.

Description: Document images can contain an incredible variety of information, including tables, graphs, drawings, charts, and text characters. These documents may be written in a large variety of languages and may contain either handwritten or machine-printed information. Automated techniques have been developed to process large volumes of document images of a given type, but flexibility tends to be limited. For example, a character recognition algorithm may recognize only machine-printed English characters.

This technology discriminates between handwritten and machine-printed document images for a diverse set of documents. When a document image is input the system automatically determines whether the document was written by hand or was produced by some type of electronic printer or typewriter. Unlike many existing techniques, this technology is independent of the language used or the character font style. When handwritten images are processed by character recognition, the resulting text typically contains numerous errors, greatly reducing the performance of information retrieval applications. However, this technology could be applied prior to the character recognition process so as to allow only the machine-printed documents to be passed to the character recognition engine. This would not only increase the accuracy of information retrieval, but also would reduce the required computational requirements.

Demonstration Capability: This capability can be easily demonstrated.

Potential Commercial Application(s): Companies involved in automated document processing should be interested in this technology. Additionally, companies that develop character recognition engines for document images may be interested as well.

Patent Status: Patent 7,072,514 have been issued by the USPTO.

Reference Number: 1272