

OPTICAL CHARACTER RECOGNITION SYSTEMS AND METHODS FOR PERSONAL DATA EXTRACTION

ABSTRACT

[00152] Methods and systems for extracting personal data from a sensitive document are provided. The system includes a document prediction module, a cropping module, a denoising module, and an optical character recognition (OCR) module. The document prediction module predicts type of document of the sensitive document using a keypoint matching-based approach and the cropping module extracts document shape and extracts one or more fields comprising text or pictures from the sensitive document. The denoising module prepares the one or more fields for optical character recognition, and the OCR module performs optical character recognition on the denoised one or more fields to detect characters in the one or more fields.

FIG. 4