Abstract

Documents containing mixed types of text content (printed and handwritten) are proliferating in business and academic environments. They result frequently from annotating printed documents such as bills, administrative forms, birth-certificates, letters, etc. A document can be annotated in many ways. Annotations can be handwritten text, underlines, cuts, marks, special symbols of irregular shapes, and handmade drawings. The annotated text can be multi-oriented and multi-script. Several methods for extracting annotations are outlined in the literature. Most of these systems extract annotations in controlled scenario wherein the layout is predictable, like bank checks, postal address, drafts, forms etc. Extracting handwritten annotations with non-predictable layouts in a real environment remains a difficult task because annotations can be complex due to multi-oriented handwritten text and marks which may overlap with printed text. This thesis is aimed at developing methods for localizing complex annotations in non-predictable layouts and identifying the writer for the handwritten words.

We develop methods for localizing annotations, categorizing them as textual and symbolic annotations. We further sub-categorize symbolic annotations as underlines and encirclements; and textual annotations as marginal text and inline text. We apply our methods to localize annotations written on documents such as conference papers, articles, books, office documents etc. We use statistical spectral partitioning to segment out annotations from printed text. In this approach, we work with a reduced feature set to efficiently extract the annotations on a cluttered background. We develop a new feature called Envelope Straightness to enhance the feature set. This has improved performance over the state-of-the-art features. We then investigate the use of two top-down visual saliency models for categorizing annotations. The first model makes use of supervised learning in the form of conditional random fields with a sparse encoding of feature vectors. The second model makes use of a weakly supervised learning formulation for discriminant saliency. The experimental results corroborate our hypothesis that our attention gets directed towards annotated regions in an image, and therefore, top-down saliency models can be learned to give high saliency values for the annotated regions. Along with supervision, these models take advantage of the structure and context of the annotations.

For scenarios where multiple writers annotate on the same page we develop a method to identify the writers for the handwritten words. A sliding window technique is used to extract allographic features for sub-word portions. We formulate a supervised framework and exploit the discriminative properties of the features that belong to the same cluster. We propose a new technique for separating ascenders and descenders of hand-written words from its core-region. We use the structural properties of ascenders and descenders to identify the writers of the handwritten words. The work also contributes towards dataset creation and ground truth generation for the various problems addressed in this thesis.

…