Combination of Decisions by Multiple Document Object Locators

Jung Soh
Visual Information Processing Team
Electronics and Telecommunications Research Institute
Daejeon, Korea
soh@etri.re.kr

Abstract

This paper presents a method for combining multiple document object locators tuned to different object characteristics, with the goal of achieving location performance excelling that of any individual locator. The method includes (i) a scheme for consistent representation of locator outputs regardless of output levels, (ii) the notion of object correspondence and their applications to determining what decisions to combine, (iii) a mechanism for representing knowledge of locators and its use for dynamic locator selection, (iv) functions for combining confidence values of objects. Results from experiments in postal address block location using three locators and 1,100 envelope images are presented.

1 Introduction

Document object location, visually finding the object of interest in a document image, is a significant problem in document image analysis. Traditional approaches use a single location method that is expected to locate as many as possible instances of the object class of interest. However, if the objects in the class possess diverse visual characteristics, it is difficult for the single locator to handle the wide variation. To overcome such problems, a complicated object model is often employed.

An alternative approach is to combine decisions from multiple locators, each of which is suitable only for objects with certain visual characteristics. Rather than relying on one complex locator, this approach uses a collection of relatively simple yet complementary locators. The utility of this approach depends on accurate interpretation and effective combination of the multiple location results available. This paper presents a method for combining the results of multiple document object locators, where each is tuned to different object characteristics. The objective of the combination is to provide integrated location performance that is better than that of any individual locator.

A few address block location systems attempt to combine multiple object location results. Jelinek et al. [4] use red, green, and blue color channels of processing whose destination address candidate lists are combined into a final list of top ten candidates. Palumbo et al. [8] use four segmentation methods whose results are processed independently to derive four different sets of ranked candidate blocks. Only the top choices from each ranking are compared to determine the candidate most likely to be the destination address. Lii et al. [5] employs five segmentation methods resulting in five independent ranking results. This system collects from the initial rankings those candidates showing the address syntax and reranks them. These approaches have limited combination capability in that only the top choices are combined and the same ranking method is used on all segmentation results. No clear guidelines as to reranking of candidates has been suggested either. Research in classifier combination for pattern recognition problems is related to document object locator combination. Much research has been done on combining classifiers [2, 3, 10]. We elaborate on differences between classifier combination and object locator combination in Section 2.4.

This paper is organized as follows. Section 2 presents the proposed method. Section 3 describes experimental results from combining three locators for postal address block location. Finally, Section 4 gives conclusions and discusses future research directions.

2 Proposed Method

2.1 Problem Definition

We define a *true object* as an object belonging to the class of interest, i.e., an object we want to locate. Given a document image, a locator outputs a set of document objects as subimages or regions, as the candidates for the true object. We have multiple such sets, each output by one of the multiple locators available. The problem is to generate a new set of objects that are the candidates for the true object. Each object in the new set (i) is an object output by *at least one* of the locators and (ii) is assigned a *confidence* value such that the confidence value assigned to the true object is as close to