Pattern Recognition 45 (2012) 4237–4249

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Removal of noise patterns in handwritten images using expectation maximization and fuzzy inference systems

Mehdi Haji *, Tien D. Bui, Ching Y. Suen

Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada

ARTICLE INFO

Article history:
Received 23 November 2011
Received in revised form
23 April 2012
Accepted 16 May 2012
Available online 31 May 2012

Keywords:
Denoising
Handwritten images
Recognition
Fuzzy inference systems
Expectation maximization
Optimization

ABSTRACT

The removal of noise patterns in handwritten images requires careful processing. A noise pattern belongs to a class that we have either seen or not seen before. In the former case, the difficulty lies in the fact that some types of noise patterns look similar to certain characters or parts of characters. In the latter case, we do not know the class of noise in advance which excludes the possibility of using parametric learning methods. In order to address these difficulties, we formulate the noise removal and recognition as a single optimization problem, which can be solved by expectation maximization given that we have a recognition engine that is trained for clean images. We show that the processing time for a noisy input is higher than that of a clean input by a factor of two times the number of connected components of the input image in each iteration of the optimization process. Therefore, in order to speed up the convergence, we propose to use fuzzy inference systems in the initialization step of the optimization process. Fuzzy inference systems are based on linguistic rules that facilitate the definition of some common classes of noise patterns in handwritten images such as impulsive noise and background lines. We analyze the performance of our approach both in terms of recognition rate and speed. Our experimental results on a database of real-world handwritten images corroborate the effectiveness and feasibility of our approach in removing noise patterns and thus improving the recognition performance for noisy images.

1. Introduction

The ability to handle noise is an indispensable part of any real-world image understanding system. The input data that a system is supposed to process are usually mixed with some unwanted data that deteriorate the performance of the system. The extent to which the performance of a system is affected by noise depends on the underlying models and the type of noise. For example consider an Optical Character Recognition (OCR) application where each line of text is segmented into its constituent characters and then the characters are sent to a character recognition engine. If the character recognition engine is only trained for isolated characters and we send a special symbol or a character from another script that may appear in the document, then the output of the engine could be unpredictable. In the OCR application, we consider as noise any pattern that the recognition engine is not supposed to process. Of course, not every type of input noise will result in unpredictable output behavior. For example, if a character '*l*' is broken into two parts due to noise, then the character recognition engine may recognize the image as '*i*' as its first hypothesis, but '*l*' as its second hypothesis.

In order to reduce the chance of unexpected or degraded behavior, it is desirable to remove or reduce the noise as much as possible. The goal of this research is to improve the performance of the IMDS[1] word spotting system for automatic processing of handwritten mails. Therefore, we propose our methodology for the denoising of handwritten images; however, the underlying idea is general and can be applied to similar types of denoising problems.

There are two types of noise that we have to handle when working with handwritten images: low-level and high-level. Low-level noise is the random variation of intensity in document images that is produced by the hardware equipment during the scanning process. High-level noise refers to parts of the image data that are undesirable for the intended application, and as such they can be inherent parts of the input data or artefacts that are produced by the involved hardware equipment or the processing system. Fig. 1 shows samples of handwritten text with high-level noise. Simply, anything other than text is considered as high-level noise. Besides the dot-shaped (impulsive) and line patterns that contaminate the image data in all of these samples, the interfering character strokes from the upper text lines in Fig. 1(g) and (h) are also undesirable for a recognition application. These unwanted

* Corresponding author.
E-mail address: m_haji@encs.concordia.ca (M. Haji).

[1] http://www.imds-world.com.

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.05.013

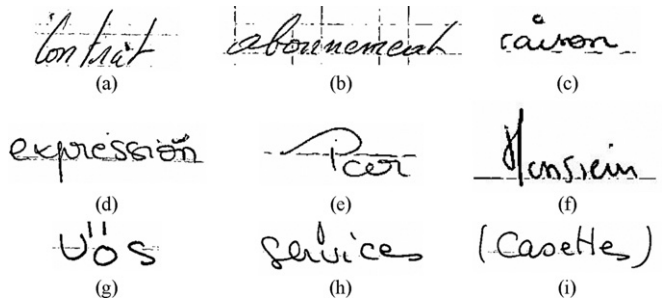

Fig. 1. Samples of handwritten text with high-level noise.

strokes are samples of high-level noise that is introduced to the image data as the result of imperfection in a previous processing step (line/word segmentation). Furthermore, depending on the application, punctuation marks and symbols (Fig. 1(i)) could be considered as high-level noise. They are undesirable parts of the image data in a word spotting application; however, they probably contain useful information in a text-to-speech application.

Low-level noise removal is a well-studied problem in the image processing literature. Recent approaches to low-level noise removal have utilized the state-of-the-art tools in statistics and signal processing [1–4]. These approaches normally address the problem of additive Gaussian noise or impulse noise removal for a general setting where it is often assumed that the image pixels are contaminated by a random process that is independent of the pixel values. However, high-level noise removal depends on the specific application, and obviously the inherent constraints and settings of each problem may call for different treatments. Not surprisingly, the removal of high-level noise in handwritten images has been less studied due to its application dependent nature. For page segmentation applications, a particular type of noise that must be handled is the marginal noise. The marginal noise refers to large black areas around a document image that are normally artefacts produced during the scanning or photocopying process. There are several studies concerning the marginal noise problem [5–7]. However, there has been comparatively less research concerning the detection and removal of other types of noise that appear in document images. In [8], a novel method based on distance transform has been proposed for the detection of removal of clutter in document images, where clutter is defined as unwanted foreground content which is typically *larger* than text. Some common forms of clutter noise in document images are punched holes, ink seeps and ink blots. Another type of noise that especially appears in handwritten images is stroke-like pattern noise, which refers to the background connected components that are similar to character strokes or diacritics. In [9], a classification-based method has been proposed for the detection and removal of stroke-like patterns. The detection of noise patterns is carried out in two phases where the first phase is based on a supervised classification, and the second phase is based on an unsupervised classification technique. The method that we propose in this paper can be considered as an extension of [9] in the sense that our method does not rely on the noise patterns belonging to any particular distribution. Therefore, we formulate the noise removal problem as an unsupervised learning where the optimization criterion is the recognition score for the input image after noise removal. To the best of our knowledge, this work is the first to address the problem of arbitrary noise patterns in handwritten images for recognition applications. We will present an algorithm based on expectation maximization for the unified denoising/recognition optimization problem, and given that prior knowledge about noise is available, we will present a systematic way based on fuzzy logic in order to incorporate that knowledge into the optimization process.

Fuzzy logic is a form of logic derived from fuzzy set theory to deal with variables and reasoning that are approximate. Fuzzy inference systems (FISs) which are rule-based systems based on fuzzy variables have been successfully applied to many fields such as expert systems, data classification, decision making, computer vision and automatic control [10,11]. One main advantage of fuzzy variables and fuzzy rules is that they facilitate the expression of rules and facts that are easily understandable by humans. Furthermore, it is easy to modify a FIS by inserting and deleting rules, meaning that there is no need to create a new system from scratch. In order to train a FIS, it is possible to start with a few rules that are designed by human expert and then fine-tune the parameters of the FIS over a set of training (validation) data.

Recently there has been a great interest in using fuzzy logic for the detection and removal of low-level noise in images [12–16]. In document image processing, fuzzy logic has been applied for the enhancement of low-quality images [17], feature extraction, recognition, etc. [18]. In this paper, we utilize fuzzy logic to incorporate our prior knowledge about some common types of noise patterns into our proposed noise removal algorithm.

2. Problem definition

Let $C_i=\{c_1^i, c_2^i, \ldots, c_{ni}^i\}$ be the set of connected components of a word image W_i. The set of connected components is composed of two disjoint subsets T_i and N_i, where $T_i=\{c_j^i \in Text:\ 1 \le j \le n_i\}$ denotes the subset of connected components that belong to the text, and $N_i=\{c_j^i \notin Text:\ 1 \le j \le n_i\}$ denotes the subset of connected components that do not belong to the text. The text itself is a natural language which is defined over a finite alphabet Σ which is the set of letters, and depending on the application, digits and punctuation marks. A word over the alphabet Σ is defined as a finite sequence of letters. In a natural language not all possible sequences of letters form valid words. Let $V \subset \Sigma^*$ denote the set of valid words, i.e., the vocabulary of the language. The goal is to find the two subsets of text T_i and noise N_i for a word image W_i given the vocabulary V. It should be noted that the vocabulary is application-dependent, and it may be as small as a few 10 of words or it can be as large as tens of thousands of words or even unlimited in which case it must be represented by a set of formation rules or statistical models.

There are two general approaches to find the subsets of text and noise from the image of a word: latent and direct. In the former, we treat the indicator functions associated with the subsets of text and noise as latent variables that have to be inferred from observable variables of the recognition system. In the latter, we either implicitly or explicitly model the likelihood functions of the text and noise based on a priori knowledge. Examples of direct noise removal approaches for handwritten document images are given in the seminal works of Agrawal and Doermann [8,9].

A direct noise removal approach can be formulated as a binary classification problem with the two classes of noise and text. Consequently, we have to make some assumptions about the nature of the patterns belonging to one class in order to be able to distinguish them from the patterns belonging to the other class. The main difficulty here lies in the fact that there could be significant overlaps between certain classes of noise patterns and characters or parts of characters. In such cases, we have to use the context knowledge (i.e., transcription) in order to resolve the ambiguity. Therefore, in this paper, we formulate the noise removal and recognition as a single optimization problem involving latent variables. This makes our approach non-parametric in the sense that it does not make any specific assumption about the nature of noise. However, the non-parametric assumption comes at a price. As we will show in the following section, in general, the