Pattern Recognition Letters 63 (2015) 36–42

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Document image binarization using a discriminative structural classifier☆

Ehsan Ahmadi [a], Zohreh Azimifar [a,*], Maryam Shams [a], Mahmoud Famouri [a], Mohammad Javad Shafiee [b]

[a] School of Electrical & Computer Engineering, Shiraz University, Iran
[b] Vision & Image Processing Lab, Systems Design Engineering Department, University of Waterloo, Canada

ARTICLE INFO

Article history:
Received 12 November 2014
Available online 18 June 2015

Keywords:
Document image binarization
Graphical model
Structural classifier

ABSTRACT

Binarization is one of the key initial steps in image analysis and system understanding. Different types of document degradations make the binarization a very challenging task. This paper proposes a statistical framework for binarizing degraded document images based on the concept of conditional random fields (CRFs). The CRFs are discriminative graphical models which model conditional distribution and are used in structural classifications. The distribution of binarized images given the degraded ones is modelled with respect to a set of informative features extracted for all sites of the document image. The recent marginal based learning method [5] is used for the estimation of parameters of the model. The proposed graphical framework enables the depending labelling of all the sites of image despite the independent pixel-by-pixel binarization of other methods. The performance of our system is evaluated on different document image datasets and is compared with several well-known binarization methods. Experimental results show comparable performance with respect to other state-of-the-art methods.

1. Introduction

Binarization is one of the most important tasks in image processing and computer vision algorithms. A binary image is a digital image which is represented only by two possible values. Binarization refers to the conversion of a gray-scale image into a binary image. It is an important preprocessing step of most document analysis and system understanding methods. The role of image binarization is crucial in the success of subsequent document analysis tasks such as character segmentation and optical character recognition (OCR). The degradations such as variant illumination, blurring, bleeding-through, and smear make the binarization of document images a challenging task. Fig. 1 shows four examples of such degraded images.

Two broad categories of binarization algorithms have been addressed in the literature. Earlier researches in this field were based on global thresholding. In this category of binarization, a random variable T is considered as a threshold. This threshold is determined by an algorithm which discerns foreground from background. There are many successful approaches on global thresholding; Sahoo et al. [31] compared a significant number of such approaches and showed that the algorithms proposed by Otsu [23], Tsai [37], Johannsen and Bille [12] and Kapur et al. [13] yield good binarization results when the foreground and background are clearly separable. The method proposed by Otsu [23] which is the most famous binarization method determines the global threshold based on a histogram approach. The main drawback of global thresholding methods is that they fail when the original image contains various background patterns or if the background is heterogeneous. In other words, global thresholding methods cannot adapt with illumination varying images and they work poorly in low quality images.

More recent binarization methods have used local thresholding to overcome the limitations of global thresholding. In these methods, a local threshold is determined for each pixel according to its own gray value and the gray values of its neighbouring pixels. Bernsen [2] used a dynamic thresholding for gray-level images based on the minimal and maximal values of local windows. In another work, Gatos et al. [8] first computed the background gray-level using the algorithm of [32], and then the local threshold is determined according to the gray-level results. The local thresholding methods suffer from high computational cost to determine local threshold for each pixel. Moreover, they are highly parametric and need manually-selective parameters; therefore, their performance is highly sensitive to the choice of parameters.

Binarization of historical text documents has attracted a great attention in recent years. The document image binarization contest (DIBCO) [7], [27], [29] and handwritten document image binarization

☆ This paper has been recommended for acceptance by Prof. L. Heutte.

* Corresponding author. Tel./fax: +98 713 647 4605.
E-mail address: azimifar@cse.shirazu.ac.ir, z.azimifar@gmail.com (Z. Azimifar).

http://dx.doi.org/10.1016/j.patrec.2015.06.008
0167-8655/© 2015 Elsevier B.V. All rights reserved.

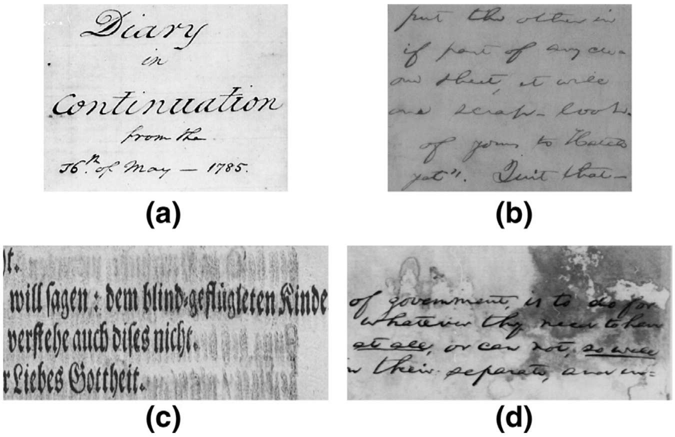

Fig. 1. Four examples of degraded document images.

contest (HDIBCO) [26], [28] are two main competitions which have been held in this field. Novel binarization approaches were proposed in these contests. The algorithm of [18] which won the DIBCO 2009 competition, approximates the background by a polynomial fitting. Then the image contrast is increased using the approximated background and a local threshold is estimated for each pixel using the local edge and contrast information. The winner of HDIBCO 2010 [35] uses the max and min operators and simulates an edge detector. The contrast is normalized by a summation term and the edge pixels are detected using Otsu algorithm. Finally the pixels are thresholded by the statistics of a local sliding window. The algorithm proposed by Howe [11], winner of the HDIBCO 2012, binarizes the image by the information of Laplacian image and Canny edge detector which creates a local energy for each pixel.

In this paper, conditional random field (CRF), a discriminative structured classifier, is used for binarization of degraded document images. The CRF directly models the desired conditional probability distribution (probability of binarized image conditioned on input image) and no independence assumption is made as is done by generative models such as Markov random fields (MRF). The CRF represents a family of undirected graphical models which has been studied for the last decade in many applications. First, Lafferty et al.[16] used CRF for segmenting and labelling sequence data. CRFs have been used in various domains including shallow parsing [33], image labelling [9], object recognition [30] and basic text segmentation [36]. Due to limitations of the Web information extraction with linear-chain CRFs, Zhu et al. [42] utilized a two-dimensional CRF model to automatically extract object information from the Web. CRFs have also been applied for image classification [40], visual tracking [34], statistical synthesis of scientific images [1], video segmentation [41], image segmentation [15] and document image analysis [10,19,20,24,25].

In our proposed framework, the binarization task is considered as a supervised image segmentation or labelling problem. Informative features are extracted for each pixel of the image, then the probability distribution of a binarized field given the input image is directly modelled by a CRF. Because of the large scalability of the problem, the recent marginal based training method [5] is used to estimate the parameters of the model. In the test phase, the most probable binary image which maximizes the trained model is the desired binarization output. The statistical view of the CRF in the task of binarization is one of the main properties of the proposed method. The second main characteristic of the model is the structural nature of the CRF which enables the model to take into account the local dependencies between neighbouring binary labels. On the other hand, constructing the global distribution of binarized output given the input image preserves the global view of the method. The combination of statistical and structural approaches increases the generalization and robustness of our framework compared to other competing methods which usually attempt to extract local specific rules for binarization.

This paper is organized as follows: Section 2 describes the general concepts of CRF. In Section 3, the proposed method is explained. Experimental results are presented and discussed in Sections 4 and 5 bring the paper to the end by concluding and presenting some future remarks.

2. Conditional random fields

CRFs construct a discriminative framework which has been emerged recently in the field of statistical machine learning. This framework directly models the conditional probability distribution rather than the joint probability assumed by MRFs. Moreover, the strong independence assumption of generative models is relaxed by CRFs. The CRFs have also addressed the problem of other discriminative models such as maximum entropy Markov models (MEMMs) which suffers from unexpected bias towards states with fewer successor states (label bias problem [16]).

2.1. CRF definition

Suppose Y indicates a 2-D field of output random variables which are to be predicted according to observed input variables X. Also let $G = (V, E)$ be an undirected graph whose vertices V correspond to the output variables Y and the edges E represent the underlying dependencies between pairs of output variables. Then, (Y, X) is said to be a conditional random field if, when globally conditioned on X, the random variables Y_i obey the Markov property with respect to the graph G. In other words, $P(Y_i|X, Y_{V-\{i\}}) = P(Y_i|X, Y_{N_i})$ where $V - \{i\}$ is the set of all nodes in G except the node i and N_i is the set of neighbours of node i in G. The conditional distribution of the output field Y given the observation X has the form

$$P(Y|X) = \frac{1}{Z(X)} \prod_{c \in C} \psi_c(Y_c, X) \tag{1}$$

where $\psi_c(Y_c, X)$ is a potential function corresponding to the clique $c \in C$[1]. Parameter $Z(X)$ is a normalization constant (partition function) with respect to all configurations Y' of output labels Y and the given observation X

$$Z(X) = \sum_{Y'} \prod_{c \in C} \psi_c(Y'_c, X). \tag{2}$$

The potential function ψ_c is an arbitrary non-negative function of Y_c and X. According to maximum entropy model [14], the potential functions ψ_c are formulated as exponential functions of a linear combination of feature functions

$$\psi_c(Y_c, X) = \exp\left(\sum_{k=1}^{K_c} \sum_{y_c} \lambda_{k,y_c}^{(c)} f_k^{(c)}(Y_c, X) I(Y_c = y_c) \right) \tag{3}$$

where $f_k^{(c)}$ and $\lambda_{k,y_c}^{(c)}$ represent the kth real-valued feature function associated with the clique c with configuration y_c and the corresponding model parameter, respectively. The $I(\cdot)$ is the indicator function and parameter K_c is the number of feature functions defined on the clique c.

2.2. Parameter estimation and inference

The parameter estimation problem is to determine the parameters λ which describe the training data in the best way. Amongst different parameter estimation algorithms, maximum likelihood estimation (MLE) is commonly used for a true estimate of model parameters [38]. Assuming m identically independently distributed (i.i.d)

[1] All nodes of a clique are mutual neighbours of each other.