Pattern Recognition Letters 98 (2017) 8–15

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Ensemble of steerable local neighbourhood grey-level information for binarization

F. Kasmin[a,*], A. Abdullah[b], A.S. Prabuwono[c]

[a] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, 76100, Melaka, Malaysia
[b] Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Selangor, Malaysia
[c] Faculty of Computing and Information Technology Rabigh, King Abdul Aziz University, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 March 2016
Available online 31 July 2017

ABSTRACT

Steerable filters are very useful in vision and image processing because the response of the filters at various orientations can be examined and manipulated, which is useful for texture analysis and image denoising. In supervised binarization, a set of grey-level values can be used to represent a particular pixel and to determine whether it belongs to the foreground or background. However, extensive noise may be produced in the resultant images if the chosen grey-level values are insufficient to describe the pixel. This may occur when some of the pixels are wrongly classified. To overcome this problem, the advantages of steerable filters are employed in proposed steerable local neighbourhood (SLN) of grey-level information methods to characterise pixels in images. The proposed methods use a support vector machine to classify each pixel using SLN grey-level information. Sets of normalised intensities of grey-level values are generated according to the orientations at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°. These sets of feature vectors contain the vectors of the local neighbourhood of grey-level information of each pixel. Document and retinal images are used to train and test the accuracy of the proposed classifier. On the basis of the results obtained for the training images, weights are applied to every SLN. Then, the ensemble of steerable local neighbourhood methods, namely the weighted addition rule and the weighted product rule, are used to combine all the SLNs' grey-level information. The results of the proposed methods are promising and clearly show a significant improvement in terms of accuracy compared to that of other methods.

1. Introduction

Image binarization is a key research theme in image processing and a crucial pre-processing method in image recognition. It is an important pre-processing step in many applications, such as document image processing [1–3], automatic visual inspection of defects [4], medical image segmentation [5,6], detection of eye positions [7] and recognition of fingerprint images [8]. Essentially, image binarization converts grey-level images into binary images, which consist of only two classes, namely the foreground and the background. However, this apparently simple procedure has been proven to be a very challenging task, especially when images have variations in contrast and illumination, variations in grey levels within the object and the background, and insufficient contrast and where there is a lack of objective measures to assess the performance of the thresholding algorithms [9]. Nevertheless, in some tasks, image binarization is a prerequisite for the subsequent stages of image recognition.

An effective way to overcome the problems in binarized images, such as noise and artefacts, is to consider the information about each desired object in the grey-level images before the binarization process. This kind of approach is called a supervised approach and uses some preceding information to decide whether a pixel belongs to the foreground or background. A set of grey-level values can be used to determine to which category a particular pixel should be classified to. However, extensive noise is produced in the resultant images if the chosen grey-level values are insufficient to describe the pixel. This problem can occur when some of the pixels are wrongly classified. Furthermore, several low-level features, such as junctions, corners, lines and edges that emerge in images in various orientations [10], cannot be detected if the chosen grey-level values are insufficient to describe the pixel.

In supervised binarization, rules are learnt by classifiers by training them on segmented ground truth images that were prepared by experts [11]. Zhu [12] considered supervised binarization as an evaluation and feedback mechanism and claimed that it

* Corresponding author.
E-mail addresses: fauziah@utem.edu.my, fuzee.kasmin@gmail.com (F. Kasmin).

http://dx.doi.org/10.1016/j.patrec.2017.07.014
0167-8655/

can increase binarization accuracy for the same type of document images and, in particular, stabilise the quality of images. The method consists of two stages, namely learning and binarization, where the outcome of the learning stage is used to optimise the parameters required for binarization. The use of supervised approaches for binarization problems was also investigated by Cheriet et al. [3]. They used a supervised approach for the automated segmentation of document images and proposed a framework for supervised binarization. The framework was divided into two parts, namely learning and application. The optimal binary image was obtained using the optimal parameter values from the learning process. Another research by Ahmadi et al. [1] used the conditional random field (CRF) to binarize degraded document images. The binarization task in this work [1] can be treated as a supervised image segmentation by extracting the informative features for each pixel of the image. The probability distribution of these features was then modelled by the CRF. A training method was then used to estimate the parameters of the model and produce the final image from the binary image that maximised the trained model.

Considering these results, the primary aim of the present work was to use a supervised approach to develop a method that can accurately binarize images. The idea presented herein was motivated by Sampe et al. [13], who used a support vector machine (SVM) [14] to segment mitochondria in fluorescence micrographs. Sampe et al. [13] used four neighbourhood grey-level values for the upper, bottom, left and right sides of a pixel, to describe that pixel. This pixel was then classified as either foreground or background by the SVM. However, the main disadvantage of this method was that it could not detect small foreground objects such as lines, edges, corners and junctions that normally appear in arbitrary orientations. Canny edge detection is one of the most widely used techniques in computer vision. It is good at detecting edges when the actual edges should not be missed, produces good localisation of the edge points and has only a single response to a single edge [15]. However, this technique also has a drawback that it is sensitive to weak edges. To address this concern, Tan et al. [16] used an orientation filter and an orientation constraint in the Canny edge detection algorithm to obtain the foreground strokes in archival documents. From the results, it seems that the orientation filters can help restore such documents effectively.

Hence, to enhance the method presented by Sampe et al. [13], the present work proposes the use of an ensemble of steerable local neighbourhood (ESLN) grey-level information to describe a particular pixel. Using an orientation approach, the best orientation can be determined. Specifically, to improve the description of the pixel, eight steerable filters based on the orientations at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° were used to construct the sets of normalised intensities of grey-level values. These sets of feature vectors contain the vectors of the local neighbourhood grey-level information of each pixel. However, if only one best single steerable local neighbourhood (SLN) of the filters is used to characterise a pixel, the detection of some low-level features could fail. Furthermore, many relevant image features are characterised by two or more orientations. Hence, the weights are assigned to each single oriented set of the feature vectors by examining the accuracy of each orientation.

One way to increase the effectiveness of a classifier is to use the ensemble method [17,18]. This concept was used previously [19], where an averaging was performed using rotating filters. Rotating 3×3 filters in eight possible angles were used to avoid edge blurring by searching the homogeneous area of the current pixel neighbourhood. Another approach as described previously [20] adopts the concept of steerable filters, where the responses of the filters at different orientations are interpolated. Then, the filters from the arbitrary orientations were combined linearly, as given in Eq. (1):

$$f(x, y) = \sum_{j=1}^{k} k_j(\theta) f'(x, y), \tag{1}$$

where M is the number of orientations and $k_j(\theta)$ is the interpolation function. If orientation filtering is applied, the oriented structures can be enlarged, and the noise can be eliminated [20]. The results of the experiments in previous studies [21–23] reveal that the neighbourhood information is important for image segmentation because these images are frequently corrupted by many problems that can cause heterogeneity in intensity. Because of the advantages of the previous works outlined above, two ensemble approaches are proposed in the present work, namely the weighted product rule and the weighted addition rule, to combine steerable filters that contain a set of feature vectors.

Contributions. The originality of this work lies in the following: (1) an ensemble learning approach is used to combine a set of steerable SVM filters for binarization, (2) an ensemble of SVMs is constructed and its performance on four datasets is evaluated, and (3) the effectiveness of the ensemble is demonstrated by comparing it with that of the existing approaches.

The remainder of the paper is organised as follows. Section 2 describes the methods used in the experiments. Section 3 discusses the experimental design, and Section 4 presents the results of the experiments. Section 5 contains a discussion of the findings and concludes the paper.

2. Proposed method

The proposed methods start with a pre-processing stage. Two types of images were used in this experiment, namely retinal and document images. The green channel of the RGB retinal images underwent several pre-processing steps because these images contain extensive information [24]. The pre-processing steps for retinal images are sharpening, contrast enhancement and Gaussian filtering. For document images, the only pre-processing step that was performed was Gaussian filtering.

Before describing the methods proposed in the present work, we review Sampe et al.'s method [13] because the proposed methods are based on their work. To use a supervised approach, a set of features needs to be assigned to a particular pixel. This set of features acts as a descriptor for each particular pixel. In Sampe et al.'s study [13], four neighbourhood grey-level values were used as the features to characterise the hotspot pixel, which was the target pixel to be described. The features were used by an SVM to classify the hotspot pixel as either foreground (1) or background (0). The label for the classifier of the hotspot pixel was extracted from the ground truth images that were manually segmented by an expert. The four neighbourhood grey-level values for the upper, bottom, left and right sides of the hotspot pixel were used to describe the hotspot pixel. During the training session, a model was created by the SVM. This model was then used by the SVM to classify the hotspot pixel as either foreground (1) or background (0). The SVM output contains a set of probability estimations for each of the two classes, namely foreground (1) and background (0). If the probability of being foreground was higher, then the hotspot pixel was classified as foreground and vice versa. The probability estimation is based on the posterior class probability, $\Pr(y = 1|x)$. The probabilities were calculated using the sigmoid function, as given in Eq. (2) [25,26]:

$$\Pr(y = 1|x) \approx P_{A,B}(f) \equiv 1/(1 + \exp(Af + B)), \tag{2}$$

where $f = f(x)$ and A and B are the best parameter settings derived using the maximum likelihood estimation from a training set $(f_i,\ y_i)$. Note that Eq. (2) from Platt [25] was slightly modified in Lin et al. [26] to address some problems.