Engineering Applications of Artificial Intelligence 37 (2015) 71–90

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A learning-based thresholding method customizable to computer vision applications

J.R. Martinez-de Dios *, A. Ollero

Robotics, Vision and Control Group, Universidad de Sevilla Avda, Camino de los Descubrimientos, 41092 Sevilla, Spain

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form
7 July 2014
Accepted 31 August 2014
Available online 27 September 2014

Keywords:
Computer vision
Applications
Thresholding

ABSTRACT

Although a large variety of thresholding techniques have been developed, the selection of a suitable technique for a particular computer vision application is still unsolved and often requires long error and trial procedures analyzing the performance and robustness of different methods. This paper proposes a training-based method that is capable of capturing, learning and imitating thresholding performance from a set of training images allowing ad-hoc adaptation to a given problem. It is applied in two stages: learning and application. In the learning stage a histogram mode object/background classifier is trained with a set of training images and their respective desired threshold values determined by a human. In the application stage, the histogram modes resulting from multi-mode decomposition are classified with the trained classifier and the threshold is computed using a tunable minimum classification error criterion. The presented method can be used in bi-level and multi-level thresholding and requires no settings since all its parameters are determined in the learning step. It has been successfully applied to several problems, some of which are described in the paper.

1. Introduction

Thresholding is a simple and effective segmentation technique. Although global thresholding cannot cope with effects such as local lighting irregularities and non-homogeneities, it has gained high popularity in wide range of computer vision applications in automatic control quality, document analysis, inspection, detection and monitoring, among many others. In fact, global thresholding is by far the most applied segmentation methods for monocromatic images in everyday applications. Image thresholding has been an intense research topic, where a number of successful methods have been developed. However, despite the very wide variety of techniques the practical use of thresholding still has unsolved issues.

Lack of customization to a given problem is one important constraint. Almost all thresholding techniques rely on generic criteria and have low capability – or none at all – to be customized to a problem. Thus, in most cases the selection of a suitable thresholding method requires long trial-and-error procedures analyzing the performance of different methods. Note that the existence of a method suitable for the problem is not guaranteed in many cases. Sensitivity and lack of robustness to the properties of images, such as lighting conditions, noise level, and the target size is another critical issue. It occurs very frequently that methods that perform well under some certain conditions fail if the conditions change. Thus, the selection of a method for a given problem requires exhaustive robustness analyses.

This paper proposes a learning-based thresholding method capable of being customized to a given problem. Selecting suitable threshold values is an easy task for a human but difficult for a machine. The objective of our method is to capture thresholding knowledge from humans and imitate their performance. Like many other techniques, our method assumes that the levels of the objects in the images are distributed in histogram clusters, commonly named modes. The input of our method is a set of training images and their corresponding desired threshold values that have been previously selected by an expert. The method is applied in two stages: learning and application. In the learning stage the sets of training images and threshold values are used to train an object/background mode classifier using an adaptive neuro-fuzzy inference system. In the application stage the histogram of each image is decomposed in its modes, which are classified by the trained classifier. Then, the threshold is computed to separate the object and background modes using a minimum classification error criterion that was customized also in the learning stage.

* Corresponding author.
E-mail addresses: jdedios@us.es (J.R. Martinez-de Dios), aollero@us.es (A. Ollero).

http://dx.doi.org/10.1016/j.engappai.2014.08.015
0952-1976/

The paper starts presenting a version for the bi-level problem and then extends it to multi-level thresholding. The proposed method can be easily customized to a given problem and shows good robustness to the conditions trained in the learning stage. Different user support software tools have been developed to facilitate its use. Our method has been applied to several problems some of which are described in this paper.

The paper is organized as follows. Related work is briefly summarized in the next subsection. Sections 2 and 3 describe respectively the application and learning stages. Section 4 presents its extension to multi-level thresholding. Section 5 illustrates its use in several computer vision problems, shows experimental results of the bi-level and a multi-level thresholding versions, analyzes the sensitivity of the method against the main sources of error and describes implementation details and the user support tools developed. Conclusions is the final section.

1.1. Related work

A large number of successful threshold selection methods have been developed using different criteria. Some are based on analyzing the histogram shape and modes structure using different approaches such as fitting histogram with mixtures of Gaussians as in Ridler and Calvard (1978) and Kittler and Illingworth (1985); searching for peaks and valleys (Sezan, 1990); or performing curvature analysis (Olivo, 1994), among others. Other techniques select the threshold using similarity measures between the original and the binarized version of the image based on edge matching (Venkatesh and Rosin, 1995) or texture (Liu and Srihari, 1994). Others are based on the entropy of the image such as Kapur et al. (1985) or Pal (1996), which minimizes the cross-entropy between the original and the binarized image. Some methods rely on the spatial information such as O'Gorman (1994) and Abutaleb (1989). A wide survey can be found in Sezgin and Sankur (2004).

Automated selection of a suitable algorithm for a given problem is an easy task for an experienced human but it is difficult for a machine. In most practical image segmentation systems, the intervention of a human operator is often needed to choose the algorithm to be used (Olabarriaga and Smeulders, 2001; Udupa et al., 1997). In most cases long trial-and-error procedures are often necessary to select a method with the required accuracy and robustness for the problem. In fact, many works proposing one method include a list of problems where it can be useful. Other works, e.g. Sezgin and Sankur (2004), include a classification of the suitability of methods for a number of different conditions.

Work (Zhang and Luo, 2000) attempted to automate this process by iteratively trying different segmentation methods and evaluating the result of segmentation using heuristic knowledge. The iterations will not stop until a satisfying result is obtained. This brute-force approach does not necessarily involve success since the existence of a suitable method for the problem is not ensured. Also the method cannot be applied in real time.

Work (Xia et al., 2005) describes a procedure to select, among several candidate thresholding methods, which is likely to obtain the best performance. The method extracts a number of features from the image such as mean value, standard deviation, skewness and kurtosis and gives them as input to a previously trained Artificial Neural Network, which determines which method is more likely to be suitable for that image. The paper validates the method using only four candidate methods and does not address how it can be extended to the hundreds of different existing thresholding methods. Again, the existence of a suitable method for the problem is not ensured.

Work (Phillips et al., 2002) proposes a training-based method that learns the region in the color space that corresponds to objects of interest segmented by experts in a number of training images. The method works well as long as conditions do not change but it is very sensitive to changing conditions, e.g. lighting. Also, using training images with different conditions – in order to improve robustness – leads to large regions in the color space, resulting in high false positive rates in image thresholding.

Learning-based schemes have been widely used in many fields such as recognition of gestures (Erol et al., 2007) and human motion (Moeslund et al., 2006), among many others. These methods typically use traditional segmentation methods and then capture and model the object motion or behavior using different tools. This approach is complementary with our method, which could be used instead of the traditional segmentation layers. Other methods learn image background and apply background suppression for segmenting objects in motion. They can only be applied to cases with mobile objects in static images.

This paper proposes a learning-based thresholding method that can be customized to a given problem. It relies on the consistency of histograms representing objects in the images and is particularly useful in partially unstructured environments where representative images of the problem are available. Its main novelties and strengths w.r.t. existing methods are (a) it allows efficiently capturing the know-how from experienced users; (b) it allows combining different criteria in a flexible way under a training-based approach; and (c) it is it capable of learning the influence of conditions and, as result, it is significantly robust to changing conditions.

2. Application stage

This stage is performed after the learning stage but for clarity it is presented before. Fig. 1 shows a simplified scheme of the application stage. It has three main steps. First, a multi-mode decomposition technique is applied to the image histogram. Then, the resulting modes are classified as object or background by a mode classifier, which has been trained in the learning stage. Finally, the threshold between the modes classified as object and as background is computed using a minimum classification error index, also customized in the learning stage.

As many thresholding techniques our method assumes that image histograms can be represented as mixtures of modes originated by the contribution of the different objects in the images:

$$h(z) = \sum_i (P(\omega_i)p(z|\omega_i)) \tag{1}$$

where $P(\omega_i)$ is the a priori probability of histogram mode ω_i and $p(z|\omega_i)$ is its probability density function. Although in some works histogram modes have been modeled as Poisson distributions (Pal and Pal, 1991), Gaussian distributions is the most accepted model. Various techniques have been developed for multi-mode Gaussian histogram decomposition. They often estimate initial values for the Gaussian mode parameters and then perform iterative parameter refinement until a fitting error is satisfied. We adopt the method proposed in Chang et al. (2002) since it simplifies parameter refinement and has lower computer burden.

Fig. 2 shows images from two different applications. Fig. 2-bottom shows the modes resulting of the Gaussian multi-mode histogram decomposition. Image histograms and histogram reconstructions

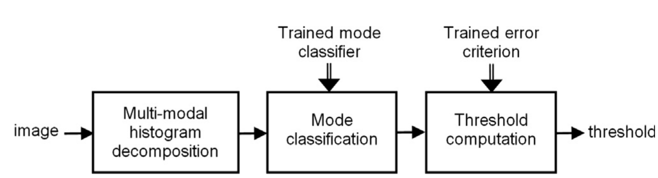

Fig. 1. Basic scheme of the application stage of the proposed method.