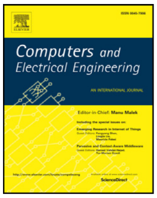
Computers and Electrical Engineering 000 (2018) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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



Nonparametric scene parsing in the images of buildings☆

Mehdi Talebi, Abbas Vafaei*, S. Amirhassan Monadjemi

Faculty of Computer Engineering, University of Isfahan, Hezarjerib Ave., Isfahan, 81746-73441, Iran

ARTICLE INFO

Article history:
Received 13 March 2016
Revised 2 January 2018
Accepted 3 January 2018
Available online xxx

Keywords:
Nonparametric image parsing
Semantic segmentation
Object recognition
Building and door detection

ABSTRACT

In this paper, we present a nonparametric approach to parse an image into regions of building, door, ground, sky, and other possible objects (such as cars, people, and trees). In a nonparametric method, first, similar images to that of the test are retrieved from a labeled training dataset. Then, the labels are transferred from the superpixels of the retrieved images to their corresponding superpixels of the test image. Finally, the conceptual Markov random field model is utilized to increase the superpixel labeling accuracy. In addition, we propose a method to improve door detection accuracy using the line, color, texture, and contextual cues. We have collected 3093 images of 40 different types of buildings from the LabelMe and Sun datasets, consisting of skyscrapers, shops, houses, apartments, churches, and so on. Experimental results on the dataset show the effectiveness of our approach with promising results.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Object recognition is an extremely difficult computer vision problem in such a way that there is not yet a system capable of recognizing objects as a 2-year-old child does [1]. Factors such as a jumble of objects in the real world, the variability intrinsic within a class (e.g., doors), pose, clutter, occlusion, shading, and lighting, all cause the complexity of an object recognition. In some works, only certain objects such as faces, pedestrians, or cars have been detected, while in some other studies, detection of multiple objects have been focused on. Nevertheless, selecting multiple objects reduces detection accuracy and perhaps detection of some objects, such as a balcony or a horse, is not so important when we are not yet able to detecting some more important objects like buildings and doors.

Image segmentation is the task of finding groups of pixels that “go together”. Zhu et al. [2] give an overview of broad segmentation topics including not only the classic unsupervised methods but also the recent weakly-/semi-supervised methods and the fully-supervised methods. Unsupervised methods segment an image into homogeneous meaningful regions with no human intervention. Weakly-/semi-supervised methods allow incorporating a small amount of prior knowledge by using labeled data in a weak or coarse form, e.g. by partially providing labels for a few pixels (e.g. interactive methods) or by manually picking out images containing common objects (e.g. cosegmentation methods). Fully-supervised methods train a segmentation model by using fully annotated data or labeled data in the very fine-grained form (e.g. all pixels in a training image are annotated). Then the segmentation model is used to rank or segment unseen data. Supervised methods achieve state-of-the-art performance for certain tasks, such as scene parsing.

☆ Reviews processed and approved for publication by Editor-in-Chief.
* Corresponding author.
E-mail address: abbas_vafaei@eng.ui.ac.ir (A. Vafaei).

https://doi.org/10.1016/j.compeleceng.2018.01.004
0045-7906/© 2018 Elsevier Ltd. All rights reserved.

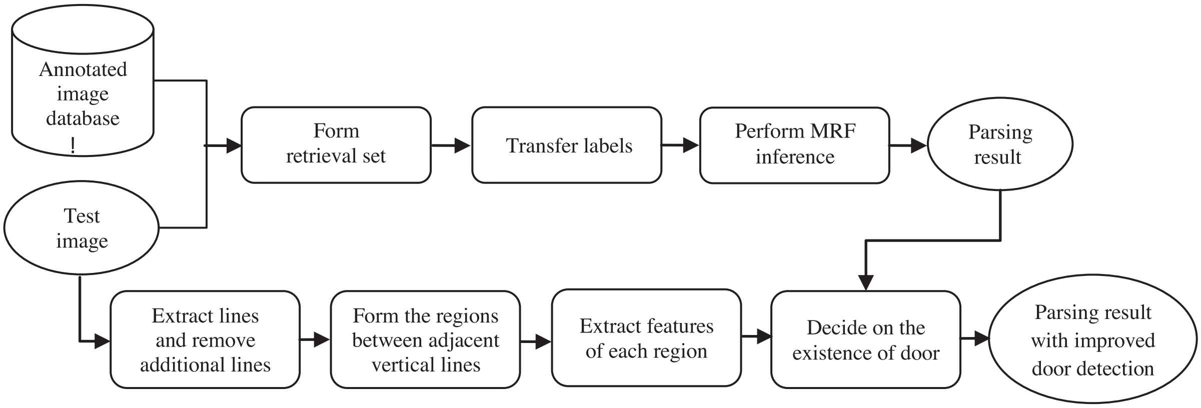

Annotated image database
Test image
Form retrieval set
Transfer labels
Perform MRF inference
Parsing result
Extract lines and remove additional lines
Form the regions between adjacent vertical lines
Extract features of each region
Decide on the existence of door
Parsing result with improved door detection

Fig. 1. Block diagram of image parsing into building, door, ground, sky, and other objects. Upper row shows the steps of nonparametric image parsing (Section 3) and lower row shows the steps of improving door detection (Section 4).

In some recent works, recognition of objects proceeds simultaneously with scene parsing. By scene parsing, we mean the assigning semantic labels such as building, ground, and sky to each pixel of the image. Scene parsing or semantic segmentation is carried out via the two parametric and nonparametric approaches. Parametric approaches are based on learning associated with an estimation of the model parameters in the training phase [3–5]. Zhou and Liu [3] present a semantic method for segmentation of natural scene images. Their method considers low-level global features, local features, and high-level contextual cues simultaneously in the conditional random field (CRF) based inference framework. Kohli et al. [4] propose an algorithm that can compute the solution of the labeling problem (using features based on image segments) in a principled manner. Their method is based on higher order CRF and uses potentials defined on sets of pixels (image segments) generated using unsupervised segmentation algorithms. With the recent advent of deep learning, several works have focused on developing convolutional neural networks (CNNs) to perform semantic segmentation [5]. The disadvantage of parametric approaches is that learning models and their corresponding parameters are updated for new training images once again. Unfortunately, this process is generally very time-consuming; for example, training a state-of-the-art CNN can take several days.

In nonparametric approaches, the knowledge of labeled training images is transferred to the test images [6–14]. To parse an input image, these algorithms first retrieve a small set of similar images and their associated semantic labels from the database and compute classification confidence maps by matching the query with retrieved images in pixels or superpixels. The final semantic labeling is obtained by solving a pairwise Markov random field (MRF) model [15]. For large datasets with hundreds of labels, such approaches have the competitive performance with parametric approaches while do not do any training at all.

In this paper, we have collected 3093 images of 40 different types of buildings from LabelMe [16] and Sun [17] datasets, including skyscrapers, shops, houses, apartments, churches, mosques, temples, hotels, libraries, hospitals, etc., and labeled with 5 classes: building, door, ground, sky, and other objects (such as cars, people, and trees). The mentioned classes and our dataset have been selected for three reasons: first, doors are important objects of entrance and exit from a building for blind people and robots, while the ground represents their path and other objects are obstacles that should be detected to avoid a collision. To segment the entire image, the sky has been further added to the classes consisting of horizontal surfaces (ground) and vertical surfaces (buildings, doors, and other objects). Second, in three-dimensional (3D) city modeling for Google Earth programs, detection of building facades and their elements like doors is a necessary step. Third, we have been unable to find a comprehensive dataset covering a large variety of buildings and doors within.

In Fig. 1, the block diagram of the proposed method is shown. Due to a large number of training images, we use a nonparametric approach to segment our images into the regions of building, door, ground, sky, and other objects (such as cars, people, and trees). This approach consists of 3 main steps: First, for each test image, similar images are retrieved from the training images. In the second step, the corresponding mapping between the regions of retrieval set and those of the test image is obtained and the labels are transferred to the test image regions. In the third step, to improve label transfer accuracy, MRF is utilized. However, in the output segmented image, door detection accuracy is low since doors occupy a small volume of the image space. Thus, we use the line, color, and texture to improve the door detection. After the extraction of lines and removal of additional ones, the regions between the adjacent vertical lines are formed and then the features of each region, including height, width, location, color, texture, and a number of lines inside the region are obtained. Afterward, the context like the door existence on the building close to the ground, its reasonable height and width, and its difference in color and texture from the neighboring region is used for decision making. From our 3093 images, 2893 training images and 200 test images have been considered and satisfactory results have been achieved.

The main contributions of this paper are as follows: (1) We have collected 3093 images of 40 different types of buildings that to the best of our knowledge, is the largest dataset on buildings. (2) We have developed a nonparametric approach for

Please cite this article as: M. Talebi et al., Nonparametric scene parsing in the images of buildings, Computers and Electrical Engineering (2018), https://doi.org/10.1016/j.compeleceng.2018.01.004

Download English Version:

https://daneshyari.com/en/article/9952245

Download Persian Version:

https://daneshyari.com/article/9952245

Daneshyari.com