Pattern Recognition Letters 38 (2014) 54–62

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Comparison of different approaches to visual terrain classification for outdoor mobile robots ☆

Yuhua Zou [a], Weihai Chen [a,*], Lihua Xie [b], Xingming Wu [a]

[a] School of Automatic Science and Electrical Engineering, Beihang University, Beijing 100191, PR China
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 22 March 2013
Available online 19 November 2013

Keywords:
Image classification
Terrain classification
Image descriptor
Compact composite descriptor
Extreme learning machine

ABSTRACT

In this paper, we present a comparison of multiple approaches to visual terrain classification for outdoor mobile robots based on different color, texture and local features. We introduce and compare three novel composite descriptors called CEDD, FCTH and JCD, with traditional color and texture descriptors, such as LTP, SCD, EHD and a descriptor called CSD–HTD generated by late fusion method. We also test three BOW models based on SIFT, SURF and ORB, respectively. We used two terrain classification datasets of which the images were captured from outdoor moving robots under different weather and ground conditions. Hence some of the images are blurred or unideally exposed. We utilize ELM, SVM and NN for classification to evaluate the performance of different combinations of image descriptors and classifiers. Experiments demonstrate that JCD can represent different terrain images with significant inter-class discrepancies, and ELM has mild optimization constraints and obtains better generalization performance. Results show that the approach based on JCD descriptor and ELM classifier performs best in term of classification effectiveness and it is suitable for real-time outdoor visual terrain classification.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In outdoor environments, a mobile robot typically has to face diverse terrain types, each posing different dangers to the robot and demanding a different motion pattern. Some of them are flat and not slippery, hence the robot can traverse them at relatively high speed. Other ground surfaces are loose, muddy or bumpy, and therefore dangerous. To prevent accidents, the robot has to traverse these regions slowly and carefully. Therefore, the ground surface itself may be a possible hazard to the robot in outdoor environments. Such a hazard is called a *non-geometric hazard* (Wilcox, 1994). The robot should be able to determine the current terrain type, and adjust its motion patterns to cope with the terrain.

There are mainly two ways, i.e., the contact-based way and the range-based way, to determine the terrain type. The former is to directly estimate terrain parameters like slippage or cohesion from sensor measurements when the robot traverses the terrain (Iagnemma and Dubowsky, 2002; Weiss et al., 2006; Iagnemma et al., 2004). Though it is independent from weather conditions and highly reliable, contact-based can not classified terrain type before the robot traverses it. Range-based methods usually learn terrain classes from training data collected by laser scanners or cameras, and then it can predict the class of new terrain data according to the learned model. Ladar-based methods often focus on distinguishing the ground surface from obstacles with geometrical information, rather than estimating the terrain type (Vandapel et al., 2004; Manduchi et al., 2005). Relatively, vision-based methods mainly use visual features (such as color, texture and shape) containing more meaningful information to determine the type of the ground surface (Jansen et al., 2005; Varma and Zisserman, 2003; Zhang et al., 2007; Varma and Zisserman, 2002).

Quite a few approaches for visual terrain classification have been presented in the literature. However, it's still unclear which approach is most appropriate for real-time outdoor robotic applications in terms of classification effectiveness and computation efficiency. The main challenge is the significant intra-class variability in the appearance of natural terrains. Terrain classes which are very similar in appearance but affect the robot mobility differently need to be discriminated correctly. To address such challenge, the terrain classification algorithm has to use more complex representations than simple color histograms and texture descriptors commonly applied in current onboard systems (Dahlkamp et al., 2006; Manduchi, 1999; Upcroft et al., 2008). Numerous authors have studied the problem of representing texture information with filter banks (Varma and Zisserman, 2002, 2005), bag of visual words (BOVW) (Filitchkin and Byl, 2012), co-occurrence matrices

☆ This paper has been recommended for acceptance by J. Laaksonen
* Corresponding author. Tel.: +86 10 82315560.
E-mail address: whchenbuaa@126.com (W. Chen).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.11.004

(Haralick et al., 1973), local patch patterns (Ojala et al., 2002; Khan et al., 2011; Khan et al., 2012), Markov random field models (Manjunath and Chellappa, 1991; Vernaza et al., 2008) and so on. Most of these methods can classify texture images very precisely, but demand massive computation resource. Hence, an effective and efficient image descriptor suitable for an onboard system remains an open question. Compact composite descriptors (CCDs) (Chatzichristofis and Boutalis, 2008a,b; Chatzichristofis et al., 2009) are recently proposed global image features combining color and texture information to describe natural color images in a very compact representation. CCDs' family includes three kind of descriptors, which are color and edge directivity descriptor (CEDD), fuzzy color and texture histogram (FCTH) and joint composite descriptor (JCD). Their qualities have so far been evaluated in retrieval from several benchmarking databases (Chatzichristofis et al., 2010b).

Analogously, there are many image classifiers of varying complexity which achieve different levels of success in classification, such as nearest neighbor (NN), artificial neural network (ANN), support vector machine (SVM), random forests, and fuzzy rule-based system. Nevertheless, a proper classifier with good generalization that is suitable for visual terrain classification remains a hot topic. Hudjakov and Tamre (2011) presented an aerial imagery terrain classification method with ANN. They extracted patterns of 29×29 pixel size from static aerial images and fed them into a neural network containing three hidden layers, to distinguish houses, roads, grass or debris. This approach processes raw patches of the image, rather than extracting features. Zhang et al. (2006) used a NN classifier to find crude clusters of similar classes and then used a more precise classifiers, e.g., SVM, to discriminate among them. Recently, a novel learning algorithm for single hidden layer feed-forward networks (SLFNs), namely, extreme learning machine (ELM), has been proposed by Huang et al. (2004). ELM can be applied for regression and classification problems (Huang et al., 2012). ELM is found to be easy to tune network parameters and fast to learn training samples because ELM randomly chooses the input weights and the hidden layer neurons instead of tuning. Zhang et al. (2008) and Zong and Huang (2011) have successfully applied ELM in the face recognition to improve the accuracy rate.

The main motivation of our paper is to find out an excellent visual descriptor and an effective image classifier for classifying different terrain types on outdoor mobile robots. In this paper, we focus on the combination of CCD descriptor and ELM classifier and evaluate their performance with oter descriptors and classifiers. We extract CCD descriptors (i.e., CEDD, FCTH and JCD) from the acquired real-world images to represent terrain types. To our best knowledge, CCD has not been applied to the domain of terrain identification for mobile robots before. In addition, we investigate four MPEG-7 low level color and texture descriptors including local ternary patterns (LTP), scalable color descriptor (SCD), edge histogram descriptor (EHD) and a descriptor CSD–HTD combining color structure descriptor (CSD) and homogeneous texture descriptor (HTD) in late fusion, as well as three BOVW model based on SIFT, SURF and ORB features, respectively. The extracted visual descriptors are applied to and compared with three different classifiers, which are ELM, SVM and NN. Experimental results show that the method based on JCD descriptor and ELM classifier has higher rate of classification effectiveness, and is robust to illumination changes.

The remainder of this paper is organized as follows. Section 2 summarizes the principles of three kind of adopted image descriptors, including CCDs, BOVW and LTP. These descriptors are then transferred to several classifiers described in Section 3. Results of our experiments are presented and discussed in Section 4. Finally, Section 5 gives some conclusions based on the study.

2. Image descriptors

In our work, we have evaluated and compared a total of 10 different image descriptors. However, in view of length limit, we only introduce the principles of CCDs, BOVW models and LTP briefly in this section. The reader can refer to Salembier et al. (2002) for the details of MPEG-7 features.

2.1. Compact composite descriptor

Compact composite descriptors are global image descriptors simultaneously capturing color and texture features in a very compact representation. CCDs are originally developed for content-based image retrieval (CBIR). In this paper, we use JCD descriptor (Chatzichristofis et al., 2009), which is the combination of CEDD (Chatzichristofis and Boutalis, 2008a) and FCTH (Chatzichristofis and Boutalis, 2008b), for terrain classification. We will briefly summarize the extraction of these features as follows.

The structure of these descriptors consists of n texture areas. The types of texture areas adopted by each descriptor are illustrated in Table 1. In particular, each texture area is separated into 24 sub-regions, with each sub-region describing a color.

CEDD is the extraction of a low level feature that combines color and texture directivity into one histogram. In CEDD feature extraction, the image is separated in a preset number of blocks. Then, color information is extracted in HSV color space using a 10-bins fuzzy linking system, which generates a 10-bins quantized histogram. Each bin of the histogram corresponds to a preset color. The number of blocks assigned to each bin is stored in a feature vector. Then, an extra set of fuzzy rules is applied to a two-input fuzzy system, in order to transform the 10-bins histogram into a 24-bins histogram, thus importing information related to the hue of each color that is presented. Next, 5 digital filters of the MPEG-7 edge histogram descriptor (EHD) are used to characterize the applied image block into corresponding texture types. In other words, these filters can export the information related to the texture of the image by classifying each image block into one or more of the 6 predefined texture regions. Thus shaping a 144-bins histogram. Then, this histogram is nonlinearly quantized via mapping the bin values from the decimal area $[0,1]$ into the integer area $[0,7]$. The final result of CEDD feature extraction is a 144-dimensional vector that gives information about color and edge directivity of an image.

Similar to CEDD, FCTH is also the extraction of a low level feature that combines color and texture. A 24-bins histogram is calculated by applying two sets of fuzzy rules to the image in HSV color space, which follows the same part of color feature extraction in CEDD. Then, each image block is transformed with one-level Haar Wavelet transformation and a set of texture elements are exported. Three energy features in high frequency bands of Haar wavelet transform are used to characterize the applied image block into one or more of eight predefined texture types. Hence, the 24-bins histogram is converted into a 192-bins histogram, importing texture information in the feature. Finally, the histogram is quantized in a similar manner as CEDD.

Note that CEDD and FCTH use the same color information, so the descriptors can be joined by combining texture areas carried by each descriptor. The combined descriptor is called Joint Composite Descriptor (Chatzichristofis et al., 2009). JCD is made up of 7 texture areas, with each area contains 24 color sub-regions, hence becomes a 168-dimensional vector.

For an image $\mathbf{I}$, the elements of the CEDD, FCTH and JCD descriptors can be described as $CEDD(\mathbf{I})_{t_1}^{c}$, $FCTH(\mathbf{I})_{t_2}^{c}$, and $JCD(\mathbf{I})_{t_3}^{c}$, respectively, with $c \in [0,23]$ symbolises color bin index and $t_1 \in [0,5], t_2 \in [0,7]$ and $t_3 \in [0,6]$ symbolise texture area index. For example, $CEDD(\mathbf{I})_3^5$ represents the $3 \times 24 + 5 = 77$th element in CEDD descriptor of image $\mathbf{I}$. Then, JCD can be achieved as follows:

Download English Version:

https://daneshyari.com/en/article/533891

Download Persian Version:

https://daneshyari.com/article/533891

Daneshyari.com