Computing Image Descriptors from Annotations Acquired from External Tools

Jose Carlos Rangel, Miguel Cazorla, Ismael García-Varea, Jesús Martínez-Gómez, Élisa Fromont and Marc Sebban

Abstract Visual descriptors are widely used in several recognition and classification tasks in robotics. The main challenge for these tasks is to find a descriptor that could represent the image content without losing representative information of the image. Nowadays, there exists a wide range of visual descriptors computed with computer vision techniques and different pooling strategies. This paper proposes a novel way for building image descriptors using an external tool, namely: Clarifai. This is a remote web tool that allows to automatically describe an input image using semantic tags, and these tags are used to generate our descriptor. The descriptor generation procedure has been tested in the ViDRILO dataset, where it has been compared and merged with some well-known descriptors. Moreover, subset variable selection techniques have been evaluated. The experimental results show that our descriptor is competitive in classification tasks with the results obtained with other kind of descriptors.

Keywords Descriptor generation · Computer vision · Semantic localization · Robotics

1 Introduction

Representing images in an appropriate way is essential for tasks like image reconstruction, image search or place recognition [17]. Comparisons between image

J.C. Rangel(✉) · M. Cazorla
Computer Science Research Institute, University of Alicante, P.O. Box 99, 03080 Alicante, Spain
e-mail: jcrangel@dccia.ua.es

I. García-Varea · J. Martínez-Gómez
University of Castilla-La Mancha, Albacete, Spain

É. Fromont · M. Sebban
Jean Monnet University, Saint Etienne, France

L.P. Reis et al. (eds.), *Robot 2015: Second Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 418,
DOI: 10.1007/978-3-319-27149-1_52

descriptors can be used to determine the similarity between pairs of images [16]. Moreover, they can also be used for generalization capabilities. This is usually done by learning classification models, where the class corresponds to a desired image category. We can find binary categorization [14], and also multi-nominal proposals [9].

The main goal of an image descriptor is to find a proper representation minimizing the loss of information. Besides, some well-known approaches like histograms of gradients (HoG [4]) or Centrist [19], we can find some novel and interesting approaches. Among these alternative representations, Fei et al. [10] propose the use of an Object Filter Bank (OFB) for scene recognition. OFB contains the responses produced for objects detectors formerly trained. The work presented in [15] includes the generation of a model that simultaneously classifies and obtains a list of annotations from images. Zhou et al. [20] suggest the application of the Super Vector (SV) coding, a non linear method to compute image descriptors. Lampert et al. [8] employ an attribute classification based object recognizer. The proposal relies on semantic attributes like shape or color of an object to perform a high-level description. Banerji et al. [1] construct a descriptor based on the color, shape and texture, through the fusion of two different descriptors using an feature representation technique.

Nowadays, it is very common the use of an external and/or remote tools that provides some functionalities or information. We can find traffic or meteorology information systems [13], but also some technologies offering processing capabilities, like grid and cloud computing [5]. These systems allow the access to a wide range of services and novel capabilities. In this sense, the Clarifai system [1] provides the technology to analyze images and identify descriptive annotations (i.e. tags) related to them. Clarifai offers an Application Programming Interface (API) that obtains the 20 most descriptive annotations from a submitted image.

This article proposes (and analyzes) the use of Clarifai to build an image descriptor based on the labels got by means of this system. To carry out the experimentation, we extract Clarifai descriptors from the visual images included in the ViDRILO dataset [11]. These descriptors are then evaluated and compared with two well-known visual and depth descriptors (GIST [12], and the Ensemble of Shape Functions (ESF [18])) in the scene classification problem, using Support Vector Machines(SVMs [2]) as classifier. Clarifai descriptors are also tested when combined with ESF and GIST. Furthermore, a subset variable selection algorithm is applied to the Clarifai descriptors.

The rest of the paper is organized as follows: Section 2 exposes how the Clarifai annotation system works. Section 3 describes our proposal to build the Clarifai descriptor. The computed descriptors used in this paper are explained in Section 4. Section 5 shows the experimentation and the obtained results. Finally, in Section 6 the conclusions obtained for this work and the future work are presented.

[1] http://www.clarifai.com/